

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТРАНСПОРТНЫХ ПОТОКОВ



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

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТРАНСПОРТНЫХ ПОТОКОВ

Под редакцией А. В. Гасникова

*Допущено
Учебно-методическим объединением
высших учебных заведений Российской Федерации
по образованию в области прикладных математики и физики
в качестве учебного пособия для студентов вузов
по направлению «Прикладные математика и физика»*

МОСКВА
МФТИ
2010

УДК 519.1:519.2:519.6:519.8(075)

ББК 22.1я73

Б24

Рецензенты:

Лаборатория волновых процессов, механико-математического факультета
МГУ им. М.В. Ломоносова (зав. лаб. проф. Н. Н. Смирнов – зам. декана мехмата МГУ)

к. ф.-м. н. В. И. Швецов (Институт системного анализа РАН)

Научный консультант академик А. А. Петров

Авторы:

А.В. Гасников, С.Л. Кленов, Е.А. Нурминский, Я.А. Холодов, Н.Б. Шамрай;
Приложения: М.Л. Бланк, Е.В. Гасникова, А.А. Замятин и В.А. Малышев,
А.В. Колесников, А.М. Райгородский

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. — 362 с.

ISBN 978-5-7417-0334-2

Излагается математический аппарат и некоторые *физические концепции*, которые могут пригодиться при создании (модернизации) комплексной интеллектуальной транспортной системы (КИТС).

Предназначено для студентов старших курсов и аспирантов физико-математических специальностей (МФТИ, МГУ, Независимого московского университета). Рекомендуется научным работникам, интересующимся вопросами математического моделирования.

ISBN 978-5-7417-0334-2

© 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., 2010

© Государственное образовательное учреждение
высшего профессионального образования
«Московский физико-технический институт
(государственный университет)», 2010

Оглавление

Предисловие.....	4
Введение.....	11
Глава 1. Моделирование транспортных потоков на основе теории равновесия.....	17
Глава 2. Математические модели транспортных потоков.....	69
Глава 3. Теория Кернера трех фаз в транспортном потоке – новый теоретический базис для интеллектуальных транспортных технологий.....	168
Приложение М. Л. Бланка. Процессы с запретами в моделях транспортных потоков.....	203
Приложение Е. В. Гасниковой. О возможной динамике в модели расчета матрицы корреспонденций.....	225
Приложение А. А. Замятина, В. А. Малышева. Введение в стохастические модели транспортных потоков.....	247
Приложение А. В. Колесникова. Транспортная задача и концентрация.....	288
Приложение А. М. Райгородского. Модели случайных графов и их применения.....	300
Задачи.....	326
Используемые сокращения.....	361

Предисловие

Идея написания этого пособия принадлежит декану факультета управления и прикладной математики (ФУПМ) МФТИ проф. Александру Алексеевичу Шананину. Более двух лет назад он предложил начать читать на Физтехе курс по выбору «Введение в математическое моделирование транспортных потоков», некоторые детали которого (глава 2) к тому моменту уже обсуждались в течение нескольких лет на семинаре «Квазилинейные уравнения и обратные задачи» в ВЦ РАН под его руководством.¹ Основная цель курса заключалась в том, чтобы познакомить заинтересованных студентов старшекурсников и аспирантов физико-математических специальностей с математикой, необходимой для решения, например, таких задач:

- эволюция затора (как будет распространяться информация о заторе по транспортному потоку),
- задача о светофоре (при каких условиях перед светофором не будет скапливаться очередь),
- задача о выборе оптимальной топологии транспортной сети (где и какую дорогу «лучше» строить),
- расчет матрицы корреспонденций и распределения потоков,
- задача о надежности графа транспортной сети.

Курс содержал дополнительные главы следующих дисциплин:

- уравнений математической физики (обобщенные решения законов сохранения, групповой анализ, автомодельная редукция, принципы максимума для квазилинейных параболических уравнений);
- теории вероятностей и случайных процессов (аппарат производящих функций, системы массового обслуживания, концентрация меры, исследование асимптотик с помощью метода перевала);

¹ Здесь также хотелось бы обратить внимание на ту огромную роль, которую сыграл зав. каф. вычислительной математики МФТИ чл.-корр. РАН А. С. Холодов во внедрении этой тематики в образовательный процесс на Физтехе. За несколько лет до того, как был поставлен упомянутый курс, Александр Сергеевич уже читал лекцию по гидродинамическим моделям транспортного потока в рамках своего семестрового курса для студентов ФУПМ МФТИ «Нелинейные вычислительные процессы». Энтузиазм Александра Сергеевича «зажег» тогда многих (и не только студентов). Отметим также, что с 2005 года В. И. Швецов ведёт курс «Математические модели транспортных потоков» для студентов ФУПМ МФТИ на базовой кафедре Института системного анализа РАН.

- функционального анализа (сжимающие отображения, монотонные операторы, конусные методы);
- теории динамических систем (методы функционалов Ляпунова) и эргодической теории (концентрация инвариантной меры, элементы статистической физики);
- кинетической теории (уравнения Колмогорова, социодинамика, динамика систем с мотивацией, самоорганизация);
- теории игр (эволюционные игры: равновесие Нэша как устойчивое положение равновесия динамической системы локального «нащупывания» наилучших ответов);
- оптимизации в конечномерных и бесконечномерных пространствах (принцип Лагранжа, двойственность, отделимость, принцип Беллмана, элементы теории управления);
- дискретной математики (задачи на графах и эффективные (приближенные, вероятностные) алгоритмы их решения);
- численных методов выпуклой оптимизации (полиномиальные алгоритмы внутренней точки Нестерова–Немировского, стохастические квазиградиентные методы для задач огромной размерности).

Настоящее пособие представляет собой попытку связно преподнести как материалы прочитанных курсов, так и в целом математический аппарат и некоторые «физические концепции», которые могут пригодиться при создании (модернизации) комплексной интеллектуальной транспортной системы (КИТС). О важности такой системы в «борьбе с проблемами» (в Москве) было много сказано за последнее время.

По сути, речь идет о том, как оптимальным образом использовать имеющуюся информацию. Например, в Москве сейчас установлено (в основном на крупных перекрестках) в общей сложности более 500 видеокамер. Порядка $10^4 - 10^5$ автомобилей,² курсирующих по Москве и области, оснащены GPS-навигаторами, что позволяет получать треки (пути следования) автомобилей, с информацией о скоростях движения вдоль этих треков. Заметим, что всего в Москве ежедневно бывает более $3 \cdot 10^6$ автомобилей.

Создание КИТС на основе имеющихся данных предполагает выполнение следующих действий.

² Если говорить о сечении по времени, то таких автомобилей на дорогах будет на порядок меньше.

- Выработку адекватной (имеющимся данным³) математической модели, описывающей транспортный поток. Например, можно в не-плохом приближении уподоблять транспортный поток сжимаемой жидкости с мотивацией и использовать гидродинамические модели (или модели клеточных автоматов). Калибровка таких моделей на прямолинейных участках дороги (ребрах графа транспортной сети) довольно просто осуществляется исходя из исторической информации.
- Для постановки начально-краевых условий: описание характеристик источников и стоков автомобилей, узлов графа транспортной сети (перекрестки, въезды, съезды и т.п.) – также требуется работа с историческими данными. В результате такой работы получается матрица корреспонденций, на основе этой матрицы рассчитываются распределения потоков, а затем и матрицы перемешивания в узлах графа транспортной сети. На наш взгляд, адекватная постановка начально-краевых условий – это одна из самых сложных текущих задач. И наметки приведенного здесь пути – далеко не единственный способ получения краевых условий.
- Откалиброванная модель может использоваться для локального (по времени) управления на основе текущей информации, например, светофорной сигнализацией (въездами на крупные магистрали). Возникающие здесь задачи связаны с управлением сложными (гибридными) динамическими системами в условиях неопределенности. Такое локальное управление позволит несколько разгрузить складывающуюся на дорогах в данный момент ситуацию (правильным образом распределяя ресурсы транспортной сети между ее пользователями). Например, в Калифорнии коллектив, работающий в Беркли и возглавляемый П. Варая и А. Б. Куржанским, предложил несколько лет назад способ локального управления въездами на основные магистрали. Это привело к тому, что в результате управления время в пути среднестатистического водителя уменьшилось на 30%.
- Помимо задач локального управления имеются задачи долгосрочного управления. Где и какую дорогу следует построить при заданных бюджетных ограничениях? Каким образом (в каком размере) взимать плату за проезд с трасс в центре Москвы?⁴ На каких трассах

³ Важно подобрать модель, адекватную имеющимся данным, дабы «не забивать микроскопом гвозди».

⁴ Нужна некая «золотая середина», с одной стороны, плата за проезд должна разгрузить эти трассы, с другой желательно, чтобы пропускные ресурсы трасс использовались по максимуму.

стоит в первую очередь увеличивать число полос? Где стоит в первую очередь переделывать развязки – делать новые (в частности, решать вопрос: а выгодно ли увеличивать степень непланарности графа транспортной сети)? Эти задачи так же, как и задачи предыдущего пункта, должны решаться для всей сети в целом (не локально по пространству!). Иначе говоря, сумматорные функционалы качества критериев должны в себя включать всех участников дорожного движения. Понятно, что для решения этих задач достаточно просматривать различные сценарии (в том числе предложенные руководством города) с помощью выработанной и откалиброванной модели на предмет их состоятельности. Путем (разумного) перебора выбирать лучшие предложения по разгрузке ситуации на дорогах.



Рис. 1. Транспортный затор на одной из улиц Москвы

Некоторые рассмотренные в пособии задачи имеют также и коммерческий выход. Например, актуальной в последнее время задачей⁵ является *задача маршрутизации*: выбор оптимального (кратчайшего) маршрута следования. Понятно, что если считать веса ребер графа транспортной сети известными и не меняющимися со временем, то эта задача довольно эффективно решается. Но на практике далеко не всегда все нужные веса ребер бывают известными. В зависимости от времени суток ситуация на дорогах может кардинально меняться, поэтому воз-

⁵ Решение которой может быть интересно, например, НИС ГЛОНАСС, ЗАО «Российские навигационные технологии», различным интернет-службам, следящим за пробками на дорогах, и компаниям, производящим КПК-навигаторы (с выходом в интернет) для автомобилей.

никает необходимость прогнозирования загрузки элементов сети. Примером таких изменений служит образование заторов в «часы пик» – за короткий промежуток времени движение может быть практически парализовано даже на многополосных магистралях (рис. 1).

Несмотря на отмеченную актуальность приведенных выше задач, еще раз подчеркнем, что в пособии изложен в основном лишь математический аппарат и некоторые физические концепции, которые могут пригодиться для их решения. Важно также заметить, что формат пособия не предполагал включения технически сложных вещей, обремененных большим количеством деталей. Тем не менее по возможности мы старались хотя бы на концептуальном уровне разъяснить практически все основные нюансы. Как следствие, пособие вобрало в себя довольно много материала (который не удавалось рассказать студентам меньше чем за год), и в ходе его подготовки было использовано более 500 литературных источников, многие из которых впоследствии было решено привести в пособии. Последнее обстоятельство также не характерно для учебных пособий, но при выбранном уровне детализации и объеме излагаемого материала вполне уместно.

Во многом определяющим моментом в создании этого пособия стало принятие (в начале 2010 года) приглашения участвовать в его написании рядом ведущих специалистов в своих областях. Так, глава 1 пособия написана проф. Е. А. Нурминским и доц. Н. Б. Шамрай (ИАПУ ДВО РАН) и посвящена применению теории бескоалиционного равновесия для расчета транспортной сети при условии стационарности потоков и моделям построения матрицы корреспонденций. В написании главы 2, посвященной математическим моделям транспортных потоков, приняли также участие доц. С. Л. Кленов (кафедра общей физики МФТИ) и доц. Я. А. Холодов (кафедра вычислительной математики МФТИ). Глава 3, посвященная теории трех фаз Кернера транспортного потока, всецело написана С. Л. Кленовым (коллегой Б. С. Кернера) и содержит как упомянутые выше физические концепции, так и примеры эмпирических (измеренных) пространственно-временных структур плотного потока на скоростных автомагистралях. Как показала обратная связь от студентов, слушавших упомянутый выше курс по выбору, вос требованными оказались «стохастические» приложения доц. А. А. Замятиной и проф. В. А. Малышева (кафедра теории вероятностей мехмата МГУ) и проф. А. М. Райгородского (кафедра математической статистики и случайных процессов мехмата МГУ, кафедра анализа данных Яндекс МФТИ). Важную роль в пособии играют эргодические приложения проф. М. Л. Бланка (лаборатория Р. Л. Добрушина ИППИ РАН), асп.

Е. В. Гасниковой (кафедра анализа систем и решений ФУПМ МФТИ) и приложение доц. А. В. Колесникова (Независимый московский университет), посвященное связи задачи Монжа–Канторовича о перемещении масс и явления концентрации меры. Эти три приложения, помимо того что представляют самостоятельную ценность, также завязывают (математически) между собой многие темы этого пособия. Другими словами, знакомство с ними желательно для формирования целостного восприятия.

В конце учебного пособия приведены задачи, часть из которых в разное время предлагалась студентам. При подготовке задач большую помощь оказали молодые ученые, работающие в близких направлениях. В пособии имеется целый раздел задач (написанный асс. кафедры МОУ ФУПМ МФТИ Е. Г. Молчановым), посвященный задачам на графах и, по сути, восполняющий нехватку в пособии раздела «Транспортные потоки и Computer Science». В этом же разделе приводится довольно интересная задача, пришедшая из практических приложений от службы Яндекс-пробки.

Конечно, представленный в пособии материал далеко не полон и не может рассматриваться как *state of the art* рассматриваемой области знаний. Причина проста – колossalный объем накопившегося на данный момент материала, посвященного транспортной проблематике. Достаточно сказать, что сейчас в мире существуют десятки реферируемых научных журналов, в которых регулярно публикуются материалы на транспортную тематику. Упомянем лишь некоторые из них: Transportation Research B, Physical Review E, Review of modern physics, Transportation Science. Не говоря уже об электронных ресурсах, таких, например, как <http://arxiv.org/>. Раз в два года проводится крупнейшая в транспортном сообществе конференция по математическому моделированию транспортных потоков и смежным вопросам: «Traffic and granular flow», труды которой публикует известное немецкое издательство Springer. Кстати, в 2011 г. эта конференция впервые пройдет в Москве. В Москве ежемесячно проводится семинар «Научно-практические задачи развития автомобильно-дорожного комплекса в России» под руководством вице-президента РАН акад. В. В. Козлова. На приводимых ниже электронных ресурсах <http://kozlov-traffic-ras.ru/>, <http://wtran.dvo.ru/> можно познакомиться с работами участников этого семинара. Более того, четвертый номер журнала Труды МФТИ (ФУПМ) за 2010 г. под ред. В. В. Козлова всецело посвящен транспортной проблематике. Однако, несмотря на вышеуказанное, мы все же постарались собрать, поскольку считаем это

полезным для читателя, наиболее базовые (математически) вещи и привести соответствующий (математический) *state of the art*.

В заключение хотелось бы выразить благодарность проф. А. А. Шананину, акад. В. В. Козлову, акад. А. А. Петрову, чл.-корр. А. С. Холдову. Общение и участие в мероприятиях, к которым они имели отношения, всегда доставляло большое удовольствие и иногда вдохновляло на улучшение материала данной книги. Ценную обратную связь при подготовке этого пособия получал от проф. А. П. Буслаева и доц. О. С. Розановой, а также от всех коллег, принимавших участие в его написании. Много полезных замечаний по всему тексту сделал самый активный слушатель курса – Ю. В. Дорн (студент 6-го курса ФАКИ МФТИ). Ряд ценных замечаний по темам, изложенным в пособии, сделали: В. И. Аркин, Л. Г. Афанасьева, П. П. Бобрик, А. С. Бугаев, В. В. Веденяпин, И. Е. Виноградов, К. А. Волосов, А. И. Голиков, А. Н. Дарьин, К. Дафермос, А. В. Дмитрук, В. А. Дружинина, В. Г. Жадан, А. В. Казейкина, Б. С. Кернер, В. Ф. Колчин, Н. С. Кукушкин, А. Г. Куликовский, Г. Л. Литвинов, И. А. Лубашевский, И. С. Меньшиков, В. Д. Мильман, И. И. Морозов, А. И. Назаров, Е. Ю. Панов, Н. С. Петросян, С. А. Пирогов, В. М. Полтерович, Ю. С. Попков, И. Г. Поспелов, В. В. Пухначёв, В. Н. Разжевайкин, И. В. Рублев, В. Ж. Сакбаев, А. Ю. Семёнов, Д. Серр, Н. Н. Смирнов, А. Н. Соболевский, Е. О. Степанов, Н. Н. Субботина, В. Н. Тарасов, С. П. Тарасов, Г. М. Хенкин, Б. Н. Четверушкин, А. П. Чугайнова, С. В. Чуканов, А. Х. Шенъ, В. И. Швецов, М. В. Яшина. Ценным было общение с акад. А. Б. Куржанским.

Особо хотелось бы поблагодарить зав. кафедрой математических основ управления (МОУ) МФТИ доц. С. А. Гуза и зам. зав. каф. МОУ доц. О. С. Федько, создавших идеальные условия как для проведения занятий, так и для создания пособия, регулярно стимулировавших весь процесс написания и внимательно относившихся ко всем особенностям работы.

Работа над книгой была проведена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (мероприятие 1.2.1, НК-15П, П949; мероприятие 1.3.1, НК-215П, П1490; НОЦ № 14.740.11.0397) и частично поддержана грантами РФФИ № 08-07-00158-а, 10-07-00620-а, РГНФ № 08-02-00347, ПФИ ОМН РАН № 3, ПФИ Президиум РАН П-14.

A. B. Гасников (avgasnikov@gmail.com)
доц. кафедры МОУ ФУПМ МФТИ
12 ноября 2010 г.

Введение⁶

В 50-е годы прошлого века, в связи с исследованиями процессов, возникающих при взрыве бомбы, наблюдалось бурное развитие газовой динамики (обобщенные решения законов сохранения, устойчивые разностные схемы расчета решений). Тогда же появились первые макроскопические (гидродинамические) модели, в которых транспортный поток уподобляется потоку «мотивированной» сжимаемой жидкости (М. Лайтхилл и Дж. Уизем, П. Ричардс), и первые микроскопические модели (следования за лидером), в которых явно выписывается уравнение движения каждого автомобиля (А. Рёшель, Л. Пайпс и др.). В модели Лайтхилла–Уизема (–Ричардса) (1955) транспортный поток уподобляется потоку сжимаемой жидкости и описывается законом сохранения количества (погонной плотности) автомобилей. При этом в модели постулируется существование функциональной зависимости (уравнения состояния) между величиной потока автомобилей (= скорость \times плотность) и плотностью. Эту зависимость часто называют *фундаментальнойной диаграммой* (как правило, вогнутая функция). Собственно, в эту зависимость и «зашита» мотивация в простейших моделях.

В последующие годы класс микро- и макромоделей был значительно расширен. В современном макроскопическом подходе (А. Эу и М. Раскль, 2000) транспортный поток часто описывается нелинейной системой гиперболических уравнений (для плотности и скорости потока) с диффузией (Х. Пэйн, Р. Кюне, Б. Кернер и П. Конхойзер). При этом уравнение состояния «зашивается» во второе уравнение этой системы, как стремление водителей двигаться с желаемой скоростью.

В современном микроскопическом подходе преобладают модели типа «разумного водителя», в которых ускорение автомобиля описывается некоторой функцией от скорости этого автомобиля, расстояния до впереди идущего автомобиля (лидера) и скорости относительно лидера (М. Трайбер, 1999). При этом в таких моделях и время может течь дискретно, и сама динамика движения автомобилей может быть стохастической (марковской). Как правило, тогда такие модели называют *моделями клеточных автоматов*. В приложении М. Л. Бланка продемонстрирован один из способов того, как с помощью простейших моделей

⁶ При написании настоящего введения, носящего обзорный характер, многое было заимствовано из текста введения акад. В. В. Козлова к специальному выпуску (№ 4, 2010 г.) журнала Труды МФТИ (ФУПМ).

клеточных автоматов можно получать (математически строго) правдоподобные макроскопические уравнения состояния транспортного потока (например, треугольную фундаментальную диаграмму).

Продолжая аналогию с газовой динамикой, И. Пригожин полвека назад (а затем С. Павери-Фонтана, Д. Хельбинг и др.) предложил описывать транспортный поток кинетическим уравнением (типа Больцмана с «интегралом взаимодействия автомобилей» вместо «интеграла столкновения частиц газа»). При таком подходе макроскопическая модель получается из кинетической подобно тому, как система уравнений Эйлера получается из уравнения Больцмана.

Отметим ввиду вышесказанного, что задача математически строго обоснования кинетической модели, исходя из микроскопической, так же, как и задача обоснования макроскопической модели, исходя из кинетической, является открытой. Более того, в режимах, соответствующих «фазовому переходу» в транспортном потоке, такое обоснование, по-видимому, принципиально невозможно: нельзя осуществить соответствующий скейлинг, нельзя перейти к динамике средних, нельзя пользоваться эргодичностью системы (инвариантная мера не единственна), не ясно как обрывать (замыкать) моментную цепочку зацепляющихся уравнений. В таких режимах можно лишь нестрого говорить о похожести моделей.

Несмотря на то, что с момента появления первых фундаментальных работ прошло более полувека, по мнению ряда известных специалистов в области математического моделирования дорожного движения (К. Нагель, Х. Махмасани, М. Шрекенберг и др.), проблема образования предзаторных и заторных ситуаций еще до конца не изучена (и сродни проблеме описания турбулентных течений). Используя терминологию, предложенную Б. С. Кернером, можно сказать, что на данный момент нет общепринятого подхода, описывающего поведение движения автотранспорта в области синхронизированного потока. Иначе говоря, если автомобильный поток уподобляется жидкости, то наиболее сложная для моделирования ситуация – это замерзающая жидкость. Подтверждением вышесказанному может служить тот факт, что разные коллективы, занимающиеся моделированием транспортных потоков, как правило, используют разные модели: начиная от модели Лайтхилла–Уизема (А. А. Куржанский и др.), заканчивая моделями, в которых каждый водитель характеризуется своим вариационным принципом (И. А. Лубашевский и др.). Отметим здесь главу 3, в которой приводится «эмпирический базис», т.е. даются свойства реальных пространственно-временных структур, возникающих в плотном транспортном потоке вблизи «узкого мес-

та»,⁷ для анализа различных подходов к описанию транспортного потока. Важным атрибутом многих современных зарубежных работ, в которых предлагаются математические модели транспортного потока, является проверка предложенных моделей на возможность описания ими трех фаз Кернера транспортного потока, наблюдавшихся в многочисленных эмпирических (измеренных) данных.

Математическая теория управления транспортными потоками, как уже упоминалось выше, сейчас активно развивается в работах калифорнийской школы, возглавляемой П. Варая и А. Б. Куржанским. Исходя из модели клеточных автоматов К. Даганзо (1994) = схема Годунова + модель Лайтхилла–Уизема + треугольная фундаментальная диаграмма, предлагается способ оптимального управления светофорами и въездами на магистралях в Калифорнии. Здесь стоит обратить внимание на соизмеримость грубости выбранной модели, качества имеющихся данных <http://pems.eecs.berkeley.edu> и простоты работы с этой моделью. Поясним основную идею того, как следует управлять. Из фундаментальной диаграммы следует, что одному и тому же значению потока автомобилей соответствуют разные (как правило, две) плотности и, как следствие, разные скорости. Очевидно, что более выгодным режимом является режим с большей скоростью. Задача управления (скажем, светофорами или въездами на основные магистрали) заключается в том, чтобы большую часть времени среднестатистический водитель проводил именно в таких режимах.

Подробнее об изложенном выше можно прочитать в главах 2 и 3.

Из-за сильной неустойчивости решений уравнений (при достаточно больших плотностях), описывающих транспортные потоки, задача получения достоверного прогноза загрузки транспортной сети по имеющимся данным на час вперед сродни задаче получения достоверного прогноза погоды на неделю вперед. При этом вычислительные мощности современных высокопроизводительных кластеров (триллион и выше операций типа умножения чисел с плавающей точкой в секунду) позволяют просчитывать реальную ситуацию по Москве (в которой, напомним, порядка трех миллионов автомобилей) со значительным опережением реального времени. Другими словами, основной проблемой при моделировании транспортных потоков является не ограничение

⁷ Заметим, что, как правило, исследователи ограничиваются изучением транспортного потока на отдельном прямолинейном участке транспортной сети с простейшими начально-краевыми условиями. В то время как причиной заторов (согласно К. Даганзо) часто являются «узкие места» (перекрестки, въезды).

по вычислительным мощностям (ресурсам памяти), а большая чувствительность описываемой реальной транспортной системы к входным данным (характеристики источников и стоков автомобилей) и невозможность собрать достаточно полную информацию о входных данных.

Одним из возможных выходов из этого является рассмотрение в некотором смысле (например, в смысле теории систем массового обслуживания) усредненных показателей транспортной системы. Обгоны на многополосной дороге, очереди перед светофорами и многое другое можно описывать таким образом – о чем говорится в приложении А. А. Замятиной и В. А. Малышева (в основе моделей этого приложения лежат эргодические марковские процессы). При таком подходе исследователь следит лишь за трендом и «не обращает внимание» на высокочастотные случайные колебания (флуктуации), возможно большой амплитуды, вокруг этого тренда. В связи с упоминанием словосочетания *усредненные показатели* отметим здесь также приложение А. М. Райгородского, в котором исследуются различные свойства случайных графов (транспортных графов, web-графов). Например, такое важное свойство, как надежность графа транспортной сети к случайным отказам ребер (отказ ребра означает, что на ребре образовалась пробка). В этих приложениях наблюдается плотная концентрация исследуемых макропоказателей (макропоказателей) в маленьких окрестностях своих математических ожиданий. Однако если в приложении Замятиной–Малышева мера, которая концентрируется, порождается (как финальная = стационарная) эргодической марковской динамикой, то в приложении А. М. Райгородского она задается непосредственно, скажем, из соображений независимости и однородности (модель случайного графа Эрдёша–Рényи).

В зависимости от того, какая конкретная задача поставлена, следует отдавать предпочтение тому или иному подходу или даже какому-то их сочетанию.

Вернемся теперь к тому, как все-таки ставить начально-краевые условия для целостного описания транспортного потока на полном графике транспортной сети. Будем считать, что есть лишь информация о том, сколько людей живет в том или ином районе и сколько рабочих мест есть в том или ином районе. В главе 1 и приложении Е. В. Гасниковой приведены различные способы обоснования известной на практике энтропийной (гравитационной) модели А. Дж. Вильсона (1967) расчета (исходя из указанных выше данных) матрицы корреспонденций (сколько людей, проживающих в районе i , работают в районе j). По сути, матрица корреспонденций определяется как наиболее вероятное макросостояние, в окрестности которого и будет плотная концентрация, стацио-

нарной меры «разумной» эргодической марковской динамики, порождающей изучаемую макросистему. Точнее говоря, эргодическая марковская динамика приводит на больших временах к стационарной (инвариантной) пуассоновской (сложной) мере (прямое произведение распределений Пуассона) на пространстве макросостояний. Эта мера экспоненциально быстро концентрируется с ростом числа агентов в окрестности наиболее вероятного макросостояния, которое и принимается за положение равновесия макросистемы. Задача поиска наиболее вероятного макросостояния асимптотически (по числу агентов) эквивалентна задаче максимизации энтропийного функционала (воспользовались формулой Стирлинга) на множестве (как правило, аффинной структуры), заданном ограничениями – законами сохранения. Приятной особенностью такого класса задач является явная (легко выписываемая) зависимость решения прямой задачи через двойственные переменные. Поскольку число ограничений, как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы базируются на решении двойственной задачи: минимизации выпуклой функции. Отметим, что описанная здесь задача энтропийно линейного программирования имеет много общего с обычной транспортной задачей.

Далее, исходя из известных потребностей (корреспонденций), водители начинают «нащупывать» некую равновесную конфигурацию потоков (конкурентное равновесие, равновесие Нэша–Вардропа (1952)). Понятно, что корреспонденция не определяет, вообще говоря, однозначно путь следования. Скажем, из Физтеха можно добираться до МГУ разными способами. И если ситуация равновесная, то никому не должно быть выгодно менять свой путь следования – стратегию (ситуация равновесия по Нэшу). Это означает, что времена движения по всем путям, которые хоть кто-нибудь выбрал, соответствующим данной корреспонденции, должны быть одинаковыми. О том, как происходит «нащупывание» равновесия, какие есть обобщения у этой модели и какие есть численные способы решения возникающих по ходу задач оптимизации, написано в главе 1 и в задаче, предложенной Е. В. Гасниковой и Ю. В. Дорном (в конце пособия). К счастью, популярный сейчас формат данных о транспортной системе в виде GPS-треков автомобилей позволяет контролировать (и тем самым постоянно «подкручивать») выводы рассмотренных моделей и некоторых их важных обобщений.

Имея информацию о том, как распределяются потоки, уже можно получать оценки матриц перемешивания в узлах графа транспортной сети, тем самым замыкать целостную модель. К сожалению, такой спо-

соб крайне чувствителен к точности (полноте) входных (обучающих) данных.

В приложении А. В. Колесникова рассматривается задача Монжа–Канторовича о перемещении масс⁸ (эквивалентная, при весьма общих условиях, задаче Монжа). Оптимальный план перевозок (точнее потенциал этого отображения) удовлетворяет уравнению (в частных производных) Монжа–Ампера и порождает метрику Канторовича (–Рубинштейна). С помощью этой метрики устанавливаются довольно тонкие функциональные неравенства о *концентрации меры* (М. Громов, М. Талагран, К. Мартон, М. Леду и др.). Сам термин *концентрации меры*, по-видимому, был впервые предложен В. Д. Мильманом, внесшим значительный вклад в эту область. Геометрически этот принцип можно довольно просто пояснить задачей из конца пособия (Пуанкаре–Леви): площадь многомерной сферы (с выделенным северным и южным полюсом) практически полностью сосредоточена в маленькой полоске вокруг экватора. Этот принцип нашел широкие применения, например, в теории вероятностей (нелинейные законы больших чисел – концентрация значений липшицевых функций в окрестности медианы), асимптотической комбинаторике (например, при исследовании плотной концентрации различных функций, типа числа независимости, на случайных графах; см. приложение А. М. Райгородского). В понимании ряда материалов пособия явление концентрации меры играет немаловажную роль: концепция равновесия макросистем (модель расчета матрицы корреспонденций), исследования надежности графа транспортной сети (по модели Эрдёша–Рényи), оценка скорости сходимости к равновесию (неравенства Пуанкаре, Чигера) и др.

⁸ Сильно связанная с транспортной задачей, о которой написано в главе 1, приложении Е. В. Гасниковой, и фигурирующей также в задачах, приведенных в конце пособия.

Глава 1. Моделирование транспортных потоков на основе теории равновесия¹

1.1.	Задача транспортного равновесия.....	18
1.1.1.	Моделирование транспортных потоков как задача принятия решений	18
1.1.2.	Постановка задачи.....	20
1.1.3.	Сведение к вариационному неравенству	22
1.1.4.	Построение функций транспортных затрат	30
1.1.5.	Численные методы решения задач транспортного равновесия.....	33
1.1.6.	Соотношение между системным оптимумом и конкурентным равновесием.....	36
1.2.	Построение матрицы корреспонденций.....	44
1.2.1.	Гравитационная модель.....	44
1.2.2.	Энтропийная модель.....	46
1.2.3.	Связь между гравитационной и энтропийной моделями	51
1.3.	Парадоксы транспортного равновесия.....	53
1.3.1.	Парадокс Брайеса	53
1.3.2.	Транспортно-экологические парадоксы.....	56
1.4.	Практическая работа	61
	Литература.....	65

¹Эта глава написана Е. А. Нурминским и Н. Б. Шамрай.

1.1. Задача транспортного равновесия

Моделирование и исследование транспортных потоков часто проводится с помощью теории конкурентного бескоалиционного равновесия, описывающего достаточно адекватный механизм функционирования автомобильных улично-дорожных сетей (УДС). Рассматриваемые модели позволяют получить прогнозные оценки по загрузке элементов транспортной сети. Подобные задачи интересны в частности тем, что являются одним из инструментов для объективной оценки эффективности проектов по модификации УДС с точки зрения разгрузки наиболее проблемных участков дорог и уменьшения общих затрат на передвижение пользователей сети.

1.1.1. Моделирование транспортных потоков как задача принятия решений

Для определения объемов загрузки УДС в первую очередь необходимо выявить правила, по которым водители выбирают тот или иной маршрут следования. Поведенческие принципы пользователей транспортной сети окончательно были сформулированы в работе [48], где постулировались следующие две возможные ситуации.

1. Пользователи сети независимо друг от друга выбирают маршруты следования, соответствующие их минимальным транспортным расходам.
2. Пользователи сети выбирают маршруты следования исходя из минимизации общих транспортных расходов в сети.

С тех пор в транспортной науке приведенные поведенческие принципы получили названия соответственно *первого и второго принципов Вардрона*.

Распределение транспортных потоков согласно первому принципу Вардрона соответствует конкурентному бескоалиционному равновесию, предполагающему совершенный эгоизм участников дорожного движения — каждый стремится достигнуть конечного пункта своей поездки как можно быстрее и из имеющихся возможных вариантов следования выбирает тот маршрут, по которому будет нести минимальные затраты (временные, финансовые, моральные и т.п.) на проезд. Поэтому данный принцип также называют *оптимизацией пользователей* (user optimization).

Стоит отдельно отметить, что первый поведенческий принцип предполагает определенные допущения. Во-первых, это совершенная информированность участников движения о ситуации на дорогах — каждый знает затраты на передвижения по тем или иным маршрутам. Конечно, на данный момент такое предположение выглядит изрядной идеализацией, но развитие автоматизированных автонавигаторов и интеллектуальных транспортных систем приближает нас к такой ситуации. Во-вторых, предполагается ничтожно малое влияние отдельного участника движения на затраты по всем маршрутам. Хотя такое предположение и заведомо неверно для крупногабаритных транспортных средств, для легковых автомобилей оно представляется достаточно разумным, исключая случаи аварийных ситуаций или неопытных водителей за рулем.

Второй принцип Вардропа предполагает централизованное управление движением в сети. Соответствующее ему распределение транспортных потоков называют *системным оптимумом* (system optimization). Примером пользователей, передвигающихся согласно второму принципу, служат водители маршрутизированного транспорта.

Несмотря на то, что принципы потокового равновесия широко цитируются как принципы Вардропа, на самом деле чуть ранее их сформулировали Ф. Найт [36] и А. Пигу [45], утверждая, что все участники движения, направляющиеся из некоторого узла сети в другой, распределяются по различным маршрутам таким образом, чтобы удельная (в расчете на один автомобиль) стоимость поездки была одна и та же для всех.

В ситуации массовой автомобилизации, имеющей место практически во всех странах, подавляющее большинство участников дорожного движения любого города составляют легковые автомобили и минивэны, совершающие преимущественно маятниковые поездки: место проживания — место работы и обратно. Именно такие поездки создают пиковые нагрузки на УДС, вызывают основные потери времени и других ресурсов, повышают аварийность и усложняют социально-экономическую ситуацию. Поэтому при исследовании загрузки сети рассмотрим транспортные потоки, порождаемые именно легковым частным транспортом. Как правило, водители такого транспорта преследуют исключительно личные цели и стремятся выбрать путь следования с наименьшими затратами. Такое поведение явно соответствует первому принципу Вардропа.

Все результаты настоящей главы получены в конечномерном евклидовом пространстве \mathbb{R}^n со скалярным произведением xy и нормой $\|x\| = \sqrt{xx}$, $x, y \in \mathbb{R}^n$. Элементами пространства являются векторы-столбцы, однако знак транспонирования и дополнительные скобки при скалярном умножении будем опускать, чтобы не загромождать запись формул.

1.1.2. Постановка задачи

Исходя из приведенных соображений, построим экономико-математическую модель транспортных потоков в УДС, соответствующую первому поведенческому принципу Вардропа.

Транспортную сеть опишем в виде ориентированного графа $\Gamma(V, E)$, где V — множество вершин, E — множество дуг сети. Каждая дуга соответствует реальному участку автодороги без перекрестков. Каждая вершина представляет узел, разделяющий участки дорог. Направление дуги определяет ход следования автотранспорта. Магистрали с двусторонним движением соответственно имеют парные противоположно ориентированные дуги.

При исследовании потокообразующих факторов в множестве вершин V выделим два подмножества: первое $S \subseteq V$ содержит пункты, порождающие потоки, элементы множества S назовем *источниками*; второе $D \subseteq V$ содержит пункты, поглощающие потоки, элементы множества D назовем *стоками*. Применимально к задаче моделирования потоков, порождаемых ежедневной трудовой миграцией для утренних часов пик, источниками являются спальные районы и пригороды, стоками — деловые районы города. Множество всех потокообразующих пар представим в виде декартова произведения:

$$W = \{w = (i, j) : i \in S, j \in D\}.$$

Каждой паре источник–сток $w = (i, j) \in W$ соответствует свой спрос на перевозку ρ_w — общий объем пользователей, которые из пункта i должны прибыть в пункт j . Набор $\{\rho_w : w \in W\}$ называется *матрицей корреспонденций*. Объемы корреспонденций ρ_w могут иметь фиксированные значения или являться функциями от затрат на передвижения в сети, то есть $\rho_w = \rho(u_w)$, где u_w — минимальные транспортные затраты на проезд для пары w , зависящие в свою очередь от загрузки сети. В первом случае говорят о задаче транспорт-

ного равновесия с фиксированным спросом, во втором — о задаче с эластичным спросом.

Путем (маршрутом) в сети Γ , соединяющим вершины i и j , назовем последовательность дуг $e_1 = (i \rightarrow k_1), e_2 = (k_1 \rightarrow k_2), \dots, e_l = (k_{l-1} \rightarrow k_l), e_{l+1} = (k_l \rightarrow j)$, где $e_t \in E$ при всех $t = 1, \dots, l+1$. Предполагается отсутствие петель и циклов в маршрутах. Обозначим через P_w множество альтернативных маршрутов, следуя которым для каждой пары $w = (i, j) \in W$ исходящий из источника i поток достигает стока j . Совокупность всех путей в сети Γ обозначим через $P = \bigcup_{w \in W} P_w$.

Пусть x_p — это величина потока, идущего по пути $p \in P$. Традиционно для транспортных задач потоковые переменные должны быть неотрицательными и удовлетворять балансовым ограничениям. Поэтому для каждой пары w потоки $x_p, p \in P_w$, должны принадлежать множеству

$$X_w = \{x_p \geq 0 : p \in P_w, \sum_{p \in P_w} x_p = \rho_w\}.$$

Объединим величины x_p в вектор $x = (x_p : p \in P)$. Тогда допустимой областью для вектора x является множество, образованное как декартово произведение всех X_w :

$$X = \prod_{w \in W} X_w = \{x \geq 0 : \sum_{p \in P_w} x_p = \rho_w, w \in W\}. \quad (1)$$

Преодоление каждого из путей $p \in P$ сопровождается некоторыми затратами (время, топливо, деньги, амортизация автомобиля, износ дороги и т.п.). Количественная характеристика таких затрат зависит от интенсивности и плотности движения в сети. Как правило, в моделях рассматриваются временные или финансовые затраты. Обозначим через G_p удельные затраты пользователей на проезд по пути p . Поскольку на затраты по одному маршруту может влиять загрузка других путей УДС, то в общем случае G_p представляют собой функции от загрузки всей сети, то есть $G_p = G_p(x)$.

Во введенных обозначениях первый принцип Вардропа можно формализовать следующим образом. Водители выбирают путь с наименьшими транспортными расходами, поэтому если по пути $p \in P_w$ для пары w идет ненулевой поток, то затраты по этому пути мини-

мальны, то есть

$$\text{если } x_p^\dagger > 0, \text{ то } G_p(x^\dagger) = \min_{q \in P_w} G_q(x^\dagger) = u_w(x^\dagger), \quad (2)$$

где $u_w(x^\dagger)$ — минимальные транспортные затраты по маршрутам, соединяющим пару $w \in W$, при загрузке сети, определяемой вектором x^\dagger . Потоки $x^\dagger \in X$, удовлетворяющие условию (2), называются *равновесными*. Проблема поиска равновесных потоков $x^\dagger \in X$ называется задачей *транспортного (потокового) равновесия*.

1.1.3. Сведение к вариационному неравенству

Основной подход к решению задачи транспортного равновесия состоит в сведении условия (2) к вариационному неравенству или задаче дополнительности [17, 30, 38, 40], а в частном случае — к оптимизационной задаче [3, 16, 20, 26], и дальнейшем применении существующих численных методов.

Для компактности последующего изложения объединим компоненты $G_p(x)$ в вектор-функцию $G(x) = (G_p(x) : p \in P)$. Отдельно рассмотрим случаи эластичного и неэластичного спроса.

1.1.3.1. Транспортная задача с фиксированным спросом

Пусть для каждой пары $w \in W$ объемы корреспонденций ρ_w заданы и имеют фиксированные значения. Справедлив следующий результат.

Теорема 1. *Вектор $x^\dagger \in X$ удовлетворяет условию равновесия (2) тогда и только тогда, когда является решением вариационного неравенства*

$$G(x^\dagger)(x - x^\dagger) \geq 0, \quad \forall x \in X. \quad (3)$$

Доказательство. Пусть вектор $x^\dagger = (x_p^\dagger : p \in P) \in X$ является решением вариационного неравенства (3). Покажем, что в x^\dagger выполнено условие (2). Предположим противное, а именно, что для пары w существует путь $\bar{p} \in P_w$, такой, что $x_{\bar{p}}^\dagger > 0$ и $G_{\bar{p}}(x^\dagger) > G_{\bar{q}}(x^\dagger)$ для некоторого $\bar{q} \in P_w$, $\bar{q} \neq \bar{p}$. Рассмотрим вектор $x^\varepsilon = (x_p^\varepsilon : p \in P)$, такой, что

$$x_p^\varepsilon = \begin{cases} x_p^\dagger, & p \neq \bar{p}, p \neq \bar{q}, \\ x_{\bar{p}}^\dagger - \varepsilon, & p = \bar{p}, \\ x_{\bar{q}}^\dagger + \varepsilon, & p = \bar{q}, \end{cases}$$

где $\varepsilon > 0$ достаточно мало и не нарушает условия неотрицательности $x^\varepsilon \geq 0$. Нетрудно видеть, что $x^\varepsilon \in X$, при этом

$$\begin{aligned} G(x^\dagger)(x^\varepsilon - x^\dagger) &= G_{\bar{p}}(x^\dagger)(x_{\bar{p}}^\varepsilon - x_{\bar{p}}^\dagger) + G_{\bar{q}}(x^\dagger)(x_{\bar{q}}^\varepsilon - x_{\bar{q}}^\dagger) = \\ &= \varepsilon(G_{\bar{q}}(x^\dagger) - G_{\bar{p}}(x^\dagger)) < 0, \end{aligned}$$

что противоречит тому, что x^\dagger — решение вариационного неравенства (3). Следовательно, в точке x^\dagger условие (2) всегда выполнено.

Пусть вектор $x^\dagger \in X$ удовлетворяет условию (2). При этом для всех $p \in P_w$ и $w \in W$ выполнено

$$\begin{aligned} G_p(x^\dagger) - u_w(x^\dagger) &\geq 0, & (G_p(x^\dagger) - u_w(x^\dagger))x_p^\dagger &= 0, \\ (G_p(x^\dagger) - u_w(x^\dagger))x_p &\geq 0, & \forall x \in X, \end{aligned}$$

следовательно, имеет место оценка

$$\begin{aligned} 0 &\leq \sum_{w \in W} \sum_{p \in P_w} (G_p(x^\dagger) - u_w(x^\dagger))(x_p - x_p^\dagger) = \\ &= \sum_{w \in W} \sum_{p \in P_w} G_p(x^\dagger)(x_p - x_p^\dagger) - \sum_{w \in W} \sum_{p \in P_w} u_w(x^\dagger)(x_p - x_p^\dagger) = \\ &= G(x^\dagger)(x - x^\dagger) - \sum_{w \in W} u_w(x^\dagger) \left(\sum_{p \in P_w} x_p - \sum_{p \in P_w} x_p^\dagger \right) = G(x^\dagger)(x - x^\dagger), \end{aligned}$$

то есть x^\dagger — решение вариационного неравенства (3). \square

К настоящему времени теория и методы решения вариационных неравенств уже достаточно хорошо разработаны (см., например, монографии [32, 37, 38, 40]). Одним из наиболее известных является проективный метод и его модификации, по мнению авторов, особенно подходящие для решения транспортных задач. В связи с этим далее понадобятся следующие результаты.

Определение 1. Проекцией точки $y \in \mathbb{R}^n$ на множество $X \subset \mathbb{R}^n$ называется точка $\pi_X(y) = \operatorname{argmin}\{\|y - x\| : x \in X\}$.

Критерием проверки, является ли вектор p проекцией точки $y \in \mathbb{R}^n$ на множество X , служит выполнение условия

$$(p - y)(x - p) \geq 0, \quad \forall x \in X. \tag{4}$$

Решение вариационного неравенства (3) тесно связано с поиском неподвижных точек проективного отображения $H(x) = \pi_X(x - \lambda G(x))$, где $\lambda > 0$ — некоторое фиксированное число.

Утверждение 1. *Множество решений $X^\dagger \subseteq X$ вариационного неравенства (3) совпадает с множеством неподвижных точек отображения $H(x)$, то есть $X^\dagger = \{x^\dagger \in X : x^\dagger = H(x^\dagger)\}$.*

Доказательство. Пусть $x^\dagger \in X^\dagger$ и $\lambda > 0$, тогда выполнено неравенство

$$(x^\dagger - (x^\dagger - \lambda G(x^\dagger)))(x - x^\dagger) \geq 0, \quad \forall x \in X,$$

следовательно, в силу свойства (4) имеем $x^\dagger = \pi_X(x^\dagger - \lambda G(x^\dagger)) = H(x^\dagger)$.

Пусть $x^\dagger = H(x^\dagger)$, тогда по свойству (4) для любого $x \in X$ выполнено условие

$$0 \leq (x - x^\dagger)(x^\dagger - (x^\dagger - \lambda G(x^\dagger))) = G(x^\dagger)(x - x^\dagger),$$

то есть $x^\dagger \in X^\dagger$. □

Теорема 2. *Пусть вектор-функция G непрерывна по каждой компоненте, множество X непусто, выпукло и замкнуто. Если X ограничено, то вариационное неравенство (3) разрешимо.*

Доказательство. Для выпуклого множества X отображение $H(x) : X \rightarrow X$ является непрерывным и однозначным. Множество X по условию теоремы компактно, следовательно, по теореме Брауэра (см., например, [1, 8, 18]) у $H(x)$ существует неподвижная точка $x^\dagger = H(x^\dagger)$, которая в силу утверждения 1 одновременно является решением вариационного неравенства (3). □

Нетрудно видеть, что при фиксированных неотрицательных корреспонденциях ρ_w допустимая область X , определенная в (1), является непустым полиэдральным ограниченным множеством. Следовательно, если затраты на передвижение $G_p(x)$ являются непрерывными функциями от потоков $x \in X$, то задача транспортного равновесия с фиксированным спросом всегда разрешима. Единственность решения гарантирует свойство строгой монотонности вектор-функции $G(x)$.

Определение 2. *Вектор-функция $G : X \rightarrow \mathbb{R}^n$ называется строго монотонной на X , если для любых $x, y \in X$, таких, что $x \neq y$, выполнено $(G(x) - G(y))(x - y) > 0$.*

Теорема 3. *Если вектор-функция $G(x)$ строго монотонна, то вариационное неравенство (3) может иметь не более одного решения.*

Доказательство. Предположим противное, а именно, что существуют два различных решения $x^1, x^2 \in X$, $x^1 \neq x^2$, задачи (3). Очевидно, при этом выполнены неравенства

$$G(x^1)(x^2 - x^1) \geq 0, \quad G(x^2)(x^1 - x^2) \geq 0,$$

складывая которые, получаем

$$(G(x^1) - G(x^2))(x^2 - x^1) \geq 0,$$

что противоречит свойству строгой монотонности $G(x)$. \square

С точки зрения задачи транспортного равновесия утверждение теоремы 3 означает, что если транспортные затраты возрастают с увеличением загрузки сети, а это весьма естественное предположение, то существует единственное равновесное распределение транспортных потоков.

1.1.3.2. Транспортное равновесие с эластичным спросом

Рассмотрим случай, когда спрос на перевозки зависит от транспортных затрат, то есть $\rho_w = \rho_w(u_w)$. Предположим, что для каждого маршрута $p \in P$ транспортные затраты $G_p(x)$ строго положительны, а для всех пар $w \in W$ функция спроса $\rho_w(u_w)$ принимает только неотрицательные значения.

Объединим величины u_w в вектор $u = (u_w : w \in W)$, функции $\rho_w(u_w)$ — в вектор $\rho(u) = (\rho_w(u_w) : w \in W)$. Построим векторы

$$z = \begin{pmatrix} x \\ u \end{pmatrix}, \quad F(z) = \begin{pmatrix} G(x) - \Xi u \\ \Xi^T x - \rho(u) \end{pmatrix},$$

где $\Xi = (\xi_{pw} : p \in P, w \in W)$ — матрица инцидентности путей и пар источник–сток:

$$\xi_{pw} = \begin{cases} 1, & \text{если путь } p \text{ соединяет пару } w, \\ 0 & \text{— в противном случае.} \end{cases}$$

Допустимым множеством для вектора z будет неотрицательный ортант $Z = \{z : z \geq 0\}$.

Утверждение 2. Вектор $z^\dagger = (x^\dagger, u^\dagger) \geq 0$ является решением вариационного неравенства

$$F(z^\dagger)(z - z^\dagger) \geq 0, \quad \forall z \in Z, \quad (5)$$

тогда и только тогда, когда x^\dagger — решение вариационного неравенства (3) с допустимой областью $X = X(u^\dagger) = \{x \geq 0 : \Xi^T x = \rho(u^\dagger)\}$.

Доказательство. Пусть $x^\dagger \in X$ — решение вариационного неравенства (3) и $u^\dagger = (u_w(x^\dagger) : w \in W)$, где $u_w(x^\dagger) = \min_{q \in P_w} G_q(x^\dagger) > 0$. Тогда для любых $x \geq 0$ и $u \geq 0$ выполнены условия

$$\begin{aligned} (G(x^\dagger) - \Xi u^\dagger)x^\dagger &= 0, & (G(x^\dagger) - \Xi u^\dagger)x &\geq 0, \\ (\Xi^T x^\dagger - \rho(u^\dagger))u^\dagger &= 0, & (\Xi^T x^\dagger - \rho(u^\dagger))u &= 0. \end{aligned}$$

Откуда следует, что

$$0 \leq (G(x^\dagger) - \Xi u^\dagger)(x - x^\dagger) + (\Xi^T x^\dagger - \rho(u^\dagger))(u - u^\dagger) = F(z^\dagger)(z - z^\dagger).$$

Покажем обратное. Пусть $z^\dagger = (x^\dagger, u^\dagger) \geq 0$ — решение вариационного неравенства (5), то есть для любых $z \geq 0$ выполнено

$$F(z^\dagger)z^\dagger \leq F(z^\dagger)z.$$

Рассмотрим точки $z^\sigma = \sigma z^\dagger \geq 0$ для всех $\sigma \geq 0$. Имеем

$$F(z^\dagger)z^\dagger \begin{cases} \leq 0 & \text{при } \sigma = 0, \\ \geq \frac{F(z^\dagger)z^\dagger}{\sigma} \rightarrow 0 & \text{при } \sigma \rightarrow +\infty. \end{cases}$$

Следовательно, $F(z^\dagger)z^\dagger = 0$ и $F(z^\dagger)z \geq 0$.

Если предположить существование индекса l , такого, что соответствующий ему элемент вектора $F(z^\dagger)$ отрицательный, $F_l(z^\dagger) < 0$, то, выбирая $z_l \rightarrow +\infty$, получаем нарушение неравенства $F(z^\dagger)z \geq 0$. Отсюда $F(z^\dagger) \geq 0$. Таким образом, точка $z^\dagger = (x^\dagger, u^\dagger) \geq 0$ удовлетворяет системе

$$F(z^\dagger) \geq 0, \quad z^\dagger \geq 0, \quad F(z^\dagger)z^\dagger = 0, \quad (6)$$

известной в литературе как *нелинейная задача дополнительности* (см., например, [17, 32]).

Перепишем условия (6) в виде

$$G(x^\dagger) - \Xi u^\dagger \geq 0, \quad x^\dagger \geq 0, \quad (G(x^\dagger) - \Xi u^\dagger)x^\dagger = 0, \quad (7)$$

$$\Xi^T x^\dagger - \rho(u^\dagger) \geq 0, \quad u^\dagger \geq 0, \quad (\Xi^T x^\dagger - \rho(u^\dagger))u^\dagger = 0. \quad (8)$$

Система (7) показывает, что вектор u^\dagger соответствует минимальным транспортным затратам в сети при загрузке, определяемой потоками x^\dagger . При условии положительности транспортных затрат из (8) следует, что $\Xi^T x^\dagger - \rho(u^\dagger) = 0$, тогда неравенство (5) можно переписать в виде

$$G(x^\dagger)(x - x^\dagger) \geq u^\dagger(\Xi^T x - \rho(u^\dagger)) = 0, \quad \forall x \in X(u^\dagger).$$

□

Из утверждения 2 следует, что решение задачи транспортного равновесия с эластичным спросом сводится к решению вариационного неравенства (5), которое в свою очередь эквивалентно нелинейной задаче дополнительности (6). Допустимая область Z вариационного неравенства (5) неограничена, поэтому для существования решения z^\dagger помимо непрерывности необходимы дополнительные предположения о свойствах вектора-функции $F(z)$.

В случаях неограниченного допустимого множества вводят дополнительные предположения о свойствах задачи, например, ограниченность потенциального множества решений, коэрцитивность, монотонность и прочие. Общая идея выявления таких свойств состоит в следующем. Выберем радиус $R > 0$, такой, что пересечение замкнутого шара $B = \{z : \|z\| \leq R\}$ с выпуклым замкнутым множеством Z непусто, положим $Z_R = Z \cap B \neq \emptyset$. По теореме 2 существует точка $z_R^\dagger \in Z_R$, такая, что

$$F(z_R^\dagger)(z - z_R^\dagger) \geq 0, \quad \forall z \in Z_R. \quad (9)$$

Теорема 4. Пусть вектор-функция F непрерывна по каждой компоненте, множество Z непусто, выпукло и замкнуто. Если существует радиус $R > 0$, такой, что $Z_R \neq \emptyset$, и решение $z_R^\dagger \in Z_R$ вариационного неравенства (9) удовлетворяет условию $\|z_R^\dagger\| < R$, то вариационное неравенство (3) разрешимо.

Доказательство. Для произвольного $z \in Z$ выберем $\lambda \in (0, 1]$, такое, что точка $\bar{z} = z_R^\dagger + \lambda(z - z_R^\dagger) \in Z_R$. Имеем

$$0 \leq F(z_R^\dagger)(\bar{z} - z_R^\dagger) = F(z_R^\dagger)(z_R^\dagger + \lambda(z - z_R^\dagger) - z_R^\dagger) = \lambda F(z_R^\dagger)(z - z_R^\dagger),$$

то есть z_R^\dagger одновременно является решением вариационного неравенства (3). \square

Из теоремы 4 выводится ряд следствий (см., например, [17]).

Следствие 1. *Пусть вектор-функция F непрерывна по каждой компоненте, множество Z непусто, выпукло и замкнуто. Если вектор-функция $F(z)$ коэрцитивна относительно Z , то есть*

$$\lim_{\|z\| \rightarrow \infty, z \in Z} \frac{F(z)(z - \bar{z})}{\|z\|} \rightarrow \infty \quad \text{для некоторого } \bar{z} \in Z, \quad (10)$$

то вариационное неравенство (5) разрешимо.

Доказательство. Условие коэрцитивности (10) позволяет для каждого фиксированного $C > 0$ подобрать достаточно большое $R_C > 0$, такое, что

$$F(z)(z - \bar{z}) \geq C\|z\|, \quad \forall z \in Z, \quad \|z\| = R_C,$$

для какого-то $\bar{z} \in Z_{R_C}$, не зависящего от C и R_C .

В силу теоремы 2 разрешимо вариационное неравенство

$$F(z_{R_C}^\dagger)(z - z_{R_C}^\dagger) \geq 0, \quad \forall z \in Z_{R_C}.$$

Если $\|z_{R_C}^\dagger\| < R_C$, то по теореме 4 точка $z_{R_C}^\dagger$ является решением исходного вариационного неравенства (5).

Если $\|z_{R_C}^\dagger\| = R_C$, то получаем

$$F(z_{R_C}^\dagger)(\bar{z} - z_{R_C}^\dagger) \leq -C\|z_{R_C}^\dagger\| = -CR < 0,$$

что противоречит определению $z_{R_C}^\dagger$. \square

В случае разрешимости вариационного неравенства (5) строгая монотонность вектор-функции $F(z)$ гарантирует существование не более одного решения (теорема 3).

1.1.3.3. Симметричные задачи транспортного равновесия

Из теории оптимизации известно, что условие

$$\nabla f(x^\dagger)(x - x^\dagger) \geq 0, \quad \forall x \in X, \quad (11)$$

представляет необходимый критерий оптимальности в задаче

$$f(x) \rightarrow \min, \quad x \in X, \quad (12)$$

с дифференцируемой целевой функцией $f(x)$ и выпуклым замкнутым допустимым множеством X .

В самом деле, пусть $x^\dagger = \operatorname{argmin}\{f(x) : x \in X\}$. Рассмотрим точку $x_\lambda = x^\dagger + \lambda(x - x^\dagger) \in X$, где $\lambda \in (0, 1)$ достаточно мало. Имеет место следующая оценка:

$$0 \leq \frac{f(x_\lambda) - f(x^\dagger)}{\lambda} = \frac{f(x^\dagger) + \lambda \nabla f(x^\dagger)(x - x^\dagger) + o(\lambda) - f(x^\dagger)}{\lambda}.$$

Переходя к пределу при $\lambda \rightarrow 0$, получаем (11).

Таким образом, если предположить существование дифференцируемой функции $f : X \rightarrow \mathbb{R}$, такой, что $\nabla f(x) = G(x)$ (вектор-функция G в таком случае называется потенциальной), то вариационное неравенство (3) эквивалентно оптимизационной задаче (12).

В общем случае считается, что решить оптимизационную задачу намного проще, чем вариационное неравенство [42]. Теория методов оптимизации богата разнообразными алгоритмами. Кроме того, существует множество программных пакетов для решения этого класса задач, чего нельзя сказать о вариационных неравенствах. Однако основная трудность состоит в построении функции $f(x)$. Для потенциальных отображений такую функцию можно построить проведя следующие рассуждения.

Рассмотрим кривую L , зафиксируем на ней точку x^0 и вычислим интеграл $G(x)$ вдоль этой кривой до некоторой точки $x \in L$.

Пусть кривая L задана параметрически $L = \{x(t) : t \in [0, 1]\}$, где $x(t)$ — гладкая вектор-функция, при этом $x(0) = x^0$, $x(1) = x$. Имеем

$$\begin{aligned} \mathcal{I} &= \oint_{x^0}^x G(x(t)) d(x(t)) = \int_0^1 G(x(t)) x'_t(t) dt = \int_0^1 \nabla f(x(t)) x'_t(t) dt = \\ &= \int_0^1 df(x(t)) = f(x(t)) \Big|_0^1 = f(x(1)) - f(x(0)) = f(x) - f(x^0). \end{aligned}$$

Видим, что значение интеграла \mathcal{I} не зависит от параметрического задания кривой L . Рассмотрим простейший пример такого задания:

$x(t) = x^0 + t(x - x^0)$, тогда при $G(x) = \nabla f(x)$ вариационное неравенство (3) эквивалентно следующей оптимизационной задаче:

$$f(x) = f(x^0) + \int_0^1 G(x^0 + t(x - x^0))(x - x^0)dt \rightarrow \min, \quad x \in X. \quad (13)$$

Отметим, что признаком потенциальности вектор-функции $G : X \rightarrow \mathbb{R}^n$ является симметричность матрицы Якоби $\nabla G(x) = \left(\frac{\partial G_p(x)}{\partial x_q} : p, q \in P \right)$ для всех $x \in X$. Поэтому задачи транспортного равновесия (2), которые можно свести к оптимизационной задаче (13), называют *симметричными*.

На практике для получения численных значений равновесного распределения потоков в сети предварительно необходимо решить ряд проблем, связанных с получением исходных данных задачи, чему и посвящены следующие разделы. К таким проблемам относятся: построение функций транспортных затрат $G_p(x)$, формирование множеств всех возможных маршрутов P_w , определение объемов корреспонденций ρ_w .

1.1.4. Построение функций транспортных затрат

Сложность численного решения задачи транспортного равновесия во многом зависит от аналитического задания функций $G_p(x)$. Интуитивно вполне очевидно, что на транспортные затраты при проезде из источника в сток в первую очередь влияют издержки на дугах, составляющих маршрут следования. В литературе, посвященной изучению проблем моделирования транспортных потоков, рассматриваются разные формы такой зависимости.

Обозначим через y_e величину потока по дуге $e \in E$. Зная распределение потоков по путям, можно рассчитать загрузку каждой дуги по следующей формуле:

$$y_e = \sum_{p \in P} \theta_{ep} x_p, \quad \text{где } \theta_{ep} = \begin{cases} 1, & \text{если путь } p \text{ проходит через дугу } a; \\ 0 & \text{— в противном случае.} \end{cases} \quad (14)$$

Определим $\Theta = (\theta_{ep} : e \in E, p \in P)$ — матрицу инцидентности дуг и путей, $y = (y_e : e \in E)$ — вектор, описывающий загрузку дуг

сети Γ . В матричной форме взаимосвязь потоков по путям и дугам описывается уравнением $y = \Theta x$.

В ряде случаев рассматриваются транспортные задачи в терминах только потоковых переменных по дугам. Отметим, что в множестве X , определенном в (1), от потоковых переменных по путям x можно легко перейти к вектору y , обратный переход неоднозначен.

Удельные затраты на прохождение дуги e обозначим через τ_e . В общем случае значение τ_e зависит не только от величины потока y_e , но и от потоков по другим дугам сети. Характерным примером тому служат нерегулируемые перекрестки, где порядок движения определяет приоритетность дорог, регулируемые перекрестки с дополнительной стрелкой сигнала светофора — движение в так называемом режиме «просачивания» и т.п. Поэтому правильно предположить, что $\tau_e = \tau_e(y)$. Сформируем вектор-функцию $\tau(y) = (\tau_e(y) : e \in E)$.

Самым распространенным и простым предположением о свойствах функций транспортных затрат является аддитивная зависимость $G(x)$ от $\tau(y)$, означающая, что транспортные затраты на прохождение каждого пути $p \in P$ складываются только из затрат на проезд по дугам, составляющим этот путь [29, 30, 40]:

$$G_p(x) = \sum_{e \in E} \theta_{ep} \tau_e(y). \quad (15)$$

В результате вектор-функция $G(x)$ вариационного неравенства (3) имеет вид

$$G(x) = \Theta^T \tau(y), \quad y = \Theta x. \quad (16)$$

Рассмотрим частный случай, когда затраты на проезд по дуге $\tau_e(y)$ зависят только от объема идущего по ней потока y_e , то есть $\tau_e(y) \equiv \tau_e(y_e)$. В этом случае для любых $p, q \in P$, $p \neq q$, имеем

$$\frac{\partial G_p}{\partial x_q} = \sum_{e \in E} \theta_{ep} \frac{\partial \tau_e}{\partial y_e} \frac{\partial y_e}{\partial x_q} = \sum_{e \in E} \theta_{ep} \theta_{eq} \frac{\partial \tau_e}{\partial y_e} = \frac{\partial G_q}{\partial x_p}.$$

Следовательно, матрица Якоби $\nabla G(x)$ симметрична для любых $x \in X$, то есть вектор-функция $G(x)$ потенциальна и равновесные транспортные потоки можно определить как решение оптимизационной задачи (13). Учитывая соотношения (16), вид целевой функции

$f(x)$ определяется как

$$\begin{aligned}
f(x) &= \int_0^1 \sum_{p \in P} G_p(x^0 + t(x - x^0))(x_p - x_p^0)dt = \\
&= \int_0^1 \sum_{p \in P} \left(\sum_{e \in E} \theta_{ep} \tau_e(y_e^0 + t(y_e - y_e^0)) \right) (x_p - x_p^0) dt = \\
&= \int_0^1 \sum_{p \in P} \sum_{e \in E} \theta_{ep} (x_p - x_p^0) \tau_e(y_e^0 + t(y_e - y_e^0)) dt = \\
&= \int_0^1 \sum_{e \in E} \tau_e(y_e^0 + t(y_e - y_e^0)) \sum_{p \in P} \theta_{ep} (x_p - x_p^0) dt = \\
&= \int_0^1 \sum_{e \in E} \tau_e(y_e^0 + t(y_e - y_e^0)) (y_e - y_e^0) dt = \\
&= \int_0^1 \sum_{e \in E} \tau_e(y_e^0 + t(y_e - y_e^0)) d(y_e^0 + t(y_e - y_e^0)) = \\
&= \sum_{e \in E} \int_{y_e^0}^{y_e} \tau_e(z) dz.
\end{aligned}$$

Таким образом, при $\tau_e(y) \equiv \tau_e(y_e)$ задача (13) перепишется в виде

$$\sum_{e \in E} \int_0^{y_e} \tau_e(z) dz \rightarrow \min, \quad y = \Theta x, \quad x \in X. \quad (17)$$

Одной из широко использующихся форм функции затрат $\tau_e(y)$ является так называемая BPR-функция (Bureau of Public Road), описывающая временные затраты на проезд:

$$\tau_e(y) = \tau_e^0 (1 + \mu (y_e/c_e)^n),$$

где τ_e^0 — задержки на передвижение по пустой дуге e , c_e — пропускная способность дуги e , μ и n — некоторые положительные константы. При использовании BPR-функции задача транспортного равенства сводится к оптимизационной задаче (17).

В общем случае построение функции затрат $\tau_e(y)$ является задачей, требующей отдельных исследований. Здесь окажутся полезными как натурные замеры потоков и соответствующих им задержек

в реальных УДС, так и результаты компьютерного моделирования, например, при помощи специальных программ для агентного моделирования, так активно развивающихся в последние годы.

Существуют ситуации, когда предположение об аддитивности функций $G_p(x)$ не подходит для описания транспортных затрат. Стремление к более адекватному моделированию автомобильных потоков привело к новым формам аналитического описания затрат [23, 34, 39]. Неаддитивные транспортные затраты возникают, например, в случаях, когда при моделировании одновременно учитываются и временные, и финансовые расходы. Так, в работе [34] предложена функция, характеризующая финансовые затраты, на которые в свою очередь влияют временные задержки:

$$G_p(x) = \Phi_p \left(\sum_{e \in E} \theta_{ep} \tau_e(y) \right) + \Psi_p(x) + \eta \sum_{e \in E} \theta_{ep} \tau_e(y),$$

где $\tau_e(\cdot)$ — время, потраченное на прохождение дуги e , $\Phi_p(\cdot)$ — функция, преобразующая временные задержки для пути p в финансовые затраты, $\Psi_p(\cdot)$ — финансовые затраты, характеризующие маршрут p , которые могут меняться в зависимости от загрузки сети, $\eta > 0$ — эксплуатационные расходы в единицу времени. В работе [23] предложен более общий вид неаддитивной функции затрат:

$$G_p(x) = U_w \left(\sum_{e \in E} \theta_{ep} \tau_e(y) + g_p(\Psi_p) \right), \quad p \in P_w,$$

где Ψ_p — фиксированные финансовые затраты, характеризующие маршрут p , $g_p(\cdot)$ — функция, преобразующая финансовые затраты во временные задержки, $U_w(\cdot)$ — функция потерь (отрицательной полезности) для пары $w \in W$.

С одной стороны, неаддитивные затраты более реалистично могут описать функционирование транспортной системы, с другой — вариационное неравенство (3) (а тем более (5)) при сложных функциях $G_p(x)$ весьма сложно для анализа и решения.

1.1.5. Численные методы решения задач транспортного равновесия

Эквивалентность задачи транспортного равновесия вариационному неравенству, а в частном случае — оптимизационной задаче, позволя-

ет адаптировать численные методы решения последних для поиска равновесных потоков.

К настоящему времени можно выделить два основных подхода к построению алгоритмических схем. В первом случае задачу транспортного равновесия моделируют только через потоковые переменные по дугам и, соответственно, поиск равновесия ведется по дугам сети (дуговые алгоритмы). Во втором — основной переменной задачи является поток по пути, соответственно, итерирование ведется по допустимым маршрутам (маршрутные алгоритмы). И в первом, и во втором случаях основная трудность при численных расчетах состоит в большой размерности решаемых задач, особенно на реальных транспортных сетях.

Богатый практический опыт накоплен для частного случая, когда транспортное равновесие ищется как решение оптимизационной задачи (17).

Наиболее распространенным дуговым алгоритмом является метод Франка—Вульфа [33], несмотря на то, что этот алгоритм имеет довольно медленную сходимость, существенно замедляющуюся при приближении к равновесию и весьма чувствительному к размерности задачи. Причиной такого поведения является как практически неизбежно вырождающийся характер вспомогательной задачи линейного программирования, так и неравномерная сходимость потоков к равновесным значениям или так называемый эффект «застрекающих потоков» [21, 28, 35] — в процессе решения формируется некоторый набор дуг, по которым потоки сильно отличаются от равновесных, и такая ситуация не меняется при последующем итерировании.

Маршрутные алгоритмы распределяют корреспонденции непосредственно по множеству альтернативных путей, причем это множество, как правило, формируется в процессе решения [25, 44]. Перераспределение потоков не по дугам, а сразу по маршрутам, позволяет своевременно уйти от «застрекающих потоков», поэтому алгоритмы данного класса не обладают отмеченным недостатком метода Франка—Вульфа и сходятся равномерно, однако и здесь есть свои проблемы. Основная идея алгоритмов состоит в последовательной балансировке потоков между альтернативными маршрутами для каждой пары источник–сток. Поскольку перераспределение одного потока между маршрутами изменяет транспортные затраты во всей сети и тем самым влияет на распределение других корреспонденций,

то возникает необходимость многократного просмотра всех потоков, образующих пар и повторения перераспределения потоков. Отсутствие необходимости априорного задания всех допустимых маршрутов для каждой пары источник–сток, с одной стороны, делает алгоритмы поиска равновесия по путям привлекательными для использования, с другой, — как показала вычислительная практика [21, 25], такие алгоритмы сводят к минимуму количество используемых путей, то есть теряется возможность равномерного расщепления корреспонденции по множеству привлекательных маршрутов.

Поиск транспортного равновесия как решения вариационного неравенства (3) исследован в основном теоретически, несмотря на то, что более адекватное моделирование транспортных потоков все-таки требует рассмотрения общего случая непотенциальной и/или неаддитивной функции затрат $G(x) = (G_p(x) : p \in P)$. В целом алгоритмический аппарат для вариационных неравенств разработан достаточно хорошо, о чем свидетельствуют монографии [32, 37], но большое количество переменных и сложное описание вектор-функции $G(x)$ на практике делают задачу (3) труднорешаемой.

Среди существующих методов решения вариационных неравенств отдельно можно выделить проективные методы, отличающиеся простотой своих итерационных схем:

$$x^{k+1} = \pi_X(x^k - \lambda_k G(x^k)), \quad \lambda_k > 0, \quad k = 0, 1, 2, \dots, \quad (18)$$

где $\pi_X(y) = \operatorname{argmin}\{|y - x| : x \in X\}$ — проекция точки y на множество X . Для простых множеств (гиперплоскость, полупространство, шар, брус и т.п.) операция проектирования вычисляется аналитически, в общем случае требуется решать задачу квадратичного программирования, что значительно усложняет общий процесс. При выборе шага $\lambda_k \rightarrow 0$, $\sum \lambda_k = \infty$, проективный метод сходится к равновесному распределению при весьма общих предположениях о свойствах задачи, однако на практике такой выбор ведет к очень медленной скорости сходимости.

Для декомпозиции и ускорения сходимости процесса (18) предполагается применить подходы, основанные на теории фейеровских процессов с малыми возмущениями [9, 10] с использованием адаптивной регулировки шага [11]. Основная идея этого подхода заключается в следующем. Допустимое множество X из (1) можно представить в виде пересечения конечного числа гиперплоскостей H_w и

неотрицательного ортантта H^+ :

$$X = \bigcap_{w \in W} H_w \bigcap H^+,$$

где $H_w = \{x_p : \sum_{p \in P_w} x_p = \rho_w\}$, $H^+ = \{x_p \geq 0 : p \in P\}$. Объединим супермножества H_w и H^+ в семейство множеств $\mathcal{H} = \{H^+, H_w : w \in W\} = \{\mathcal{H}_l : l = 1, 2, \dots, |W| + 1\}$. Операция проектирования $\pi_{\mathcal{H}_l}(\cdot)$ для любого элемента \mathcal{H}_l вычисляется аналитически. Поэтому для численных расчетов транспортных потоков использовалась следующая модификация процесса (18), получившая название *метода последовательных проекций* [15]:

$$\begin{aligned} x^{k+1} &= x^k + \lambda_k v^k, \quad v^k = (\pi_{\mathcal{H}_l}(\bar{x}^k) - x^k) / \lambda_k, \\ \mathcal{H}_l \in \mathcal{H}, \quad \bar{x}^k &= x^k - \lambda_k G(x^k) \notin \mathcal{H}_l, \quad \lambda_k > 0, \quad k = 0, 1, 2, \dots \end{aligned} \quad (19)$$

Значительного ускорения сходимости процесса (19) к равновесному решению удается достичь за счет аддитивного выбора шагового множителя λ_k . Обозначим $V(k, m) = \text{conv}\{v^k, v^{k+1}, \dots, v^m\}$ — выпуклую оболочку векторов v^k, v^{k+1}, \dots, v^m через $B = \{x : \|x\| \leq 1\}$ — единичный шар. Для заданной последовательности $\theta_t \rightarrow +0$ при $t \rightarrow \infty$ определим последовательность индексов $\{k_t\}$. Шаговые множители λ_k определялись по следующим правилам.

1. При $t = 0$ полагается $k_t = 0$, $\lambda_0 > 0$ — произвольное, $q \in (0, 1)$.
2. Для данных t и k_t определяется индекс k_{t+1} , такой, что $0 \notin V(k_t, s) + \theta_t B$, $\lambda_s = \lambda_{k_t}$, $k_t \leq s < k_{t+1}$, $0 \in V(k_t, k_{t+1}) + \theta_t B$.
3. Положить $\lambda_{k_{t+1}} = q\lambda_{k_t}$.
4. Увеличить номер итерации $t = t + 1$ и повторить вычисление (19) для текущего значения λ_k .

Другими словами, по условию п. 2 первый переход к шагу k_{t+1} после k_t осуществляется тогда, когда $0 \in \text{conv}\{v^{k_t}, v^{k_t+1}, \dots, v^{k_{t+1}}\} + \theta_t B$, при этом шаговый множитель уменьшается в $q < 1$ раз (п. 3).

1.1.6. Соотношение между системным оптимумом и конкурентным равновесием

Очевидно, что общие затраты при системной оптимизации не могут превышать общих затрат при пользовательской оптимизации. По-

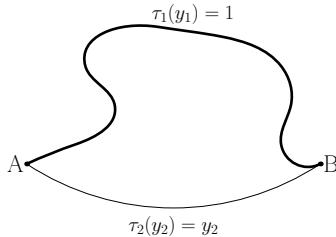


Рис. 1. Пример транспортной сети Пигу

этому разность между совокупными транспортными затратами, которые несут пользователи сети, перемещаясь согласно либо только первому, либо только второму поведенческим принципам Вардропа, можно рассматривать как цену анархии, и существуют примеры, когда эта цена составляет существенную долю от общих расходов.

На принципиальную разницу между конкурентным транспортным равновесием и системным оптимумом одним из первых обратил внимание А. Пигу [45], рассмотрев простейшую транспортную сеть, состоящую из двух дуг, соединяющих два пункта, скажем, спальный район A и бизнес-зону B (см. рис. 1). Жители пункта A вольны выбирать, по какой из двух дорог им лучше добираться до работы. Обозначим через y_1 и y_2 доли общего объема трудового потока, едущего по первой и второй дорогам соответственно. Дороги в рассматриваемой сети неравноценны. Первая представляет магистральное шоссе, которое способно принять весь поток автомобилей из пункта A в пункт B без всякого замедления движения. Однако эта дорога достаточно длинная и проезд по ней требует определенного времени τ_1 , которое будем считать равным, например, одному часу, то есть $\tau_1(y_1) = 1$. По второй дороге путь существенно короче, но это дорога узкая и движение сильно замедляется при наличии на ней потока автомобилей. Чтобы подчеркнуть суть примера, будем считать, что время проезда по второй дороге τ_2 линейно зависит от потока по этой дороге y_2 и задается соотношением $\tau_2(y_2) = y_2$. Тогда в соответствии с первым принципом Вардропа (Пигу—Найта—Вардропа) равновесному состоянию будет соответствовать такое распределение потоков $(y_1^\dagger, y_2^\dagger)$, что

$$\tau_1(y_1^\dagger) = \tau_2(y_2^\dagger), \quad y_1^\dagger + y_2^\dagger = 1, \quad y_1^\dagger, y_2^\dagger \geq 0,$$

откуда немедленно следует, что $y_1^\dagger = 0$, $y_2^\dagger = 1$, при этом системные затраты $c(y_1^\dagger, y_2^\dagger) = 1 \cdot y_1^\dagger + y_2^\dagger \cdot y_2^\dagger = 1$.

Распределение потоков в соответствии со вторым принципом Вардропа (системный оптимум) определяется как решение оптимизационной задачи:

$$\min y_1 + y_2^2, \quad y_1 + y_2 = 1, \quad y_1, y_2 \geq 0, \quad (20)$$

минимум которой достигается в точке $y_1^* = y_2^* = 0.5$, минимальные затраты $c(y_1^*, y_2^*) = 0.75$, что на 25% уменьшает системные издержки в сети.

Приведенный пример Пигу показывает, что суммарные затраты в конкурентном равновесии могут составлять $4/3$ от суммарных затрат системного оптимума. Оказывается, это соотношение представляет собой неулучшаемую оценку сверху для конкурентного потокового равновесия с аффинными функциями затрат и не зависит от топологии сети. Для подробного изложения этого результата установим некоторые полезные соотношения, характеризующие равновесные и оптимальные потоки.

Условие равновесия (2) для вектора $x^\dagger \in X$ можно переписать следующим образом: для каждой пары источник–сток $w \in W$ в сети Γ выполнены условия:

$$\text{если } x_p^\dagger > 0, \text{ то } G_p(x^\dagger) = u_w^\dagger \leq G_q(x^\dagger) \text{ для всех } q \in P_w.$$

Отсюда, очевидно, следует, что если $x_p^\dagger > 0$ и $x_q^\dagger > 0$ для путей $p, q \in P_w$, то $G_p(x^\dagger) = G_q(x^\dagger) = u_w^\dagger$, что, собственно говоря, и представляет собой математическую запись первого поведенческого принципа Вардропа.

Поскольку вклад в суммарные системные затраты (обозначим их $c(x^\dagger)$) при равновесном распределении x^\dagger вносят только ненулевые потоки $x_p^\dagger > 0$, а для них все удельные затраты в пределах одной пары w одинаковы и равны u_w^\dagger , то значение $c(x^\dagger)$ можно рассчитать следующим образом:

$$\begin{aligned} c(x^\dagger) &= \sum_{p \in P} G_p(x^\dagger) x_p^\dagger = \sum_{w \in W} \sum_{p \in P_w} G_p(x^\dagger) x_p^\dagger = \\ &= \sum_{w \in W} u_w^\dagger \sum_{p \in P_w} x_p^\dagger = \sum_{w \in W} u_w^\dagger \rho_w. \end{aligned} \quad (21)$$

Распределение потоков по второму принципу Вардропа x^* и системный оптимум $c(x^*)$ соответствуют решению оптимизационной задачи:

$$c(x) = \sum_{p \in P} G_p(x)x_p \rightarrow \min, \quad x \in X. \quad (22)$$

Положим $c_p(x) = G_p(x)x_p$ и будем предполагать, что для всех $p \in P$ функции $c_p(x)$ являются выпуклыми и непрерывно дифференцируемыми. Справедлива следующая теорема.

Теорема 5. *Пусть x^* — решения задачи (22), то есть оптимальное распределение потоков в сети. Тогда для всякой пары $w \in W$, если $x_p^* > 0$, $p \in P_w$, то $\frac{\partial c(x^*)}{\partial x_p} \leq \frac{\partial c(x^*)}{\partial x_q}$ для всех $q \in P_w$.*

Доказательство. Предположим противное, а именно, что для пары w существует путь $\bar{p} \in P_w$, такой, что $x_{\bar{p}}^* > 0$ и $\frac{\partial c(x^*)}{\partial x_{\bar{p}}} > \frac{\partial c(x^*)}{\partial x_q}$ для некоторого $\bar{q} \in P_w$, $\bar{q} \neq \bar{p}$. Рассмотрим вектор $x^\varepsilon = (x_p^\varepsilon : p \in P)$, такой, что

$$x_p^\varepsilon = \begin{cases} x_p^*, & p \neq \bar{p}, p \neq \bar{q}, \\ x_{\bar{p}}^* - \varepsilon, & p = \bar{p}, \\ x_{\bar{q}}^* + \varepsilon, & p = \bar{q}, \end{cases}$$

где $\varepsilon > 0$ достаточно мало и не нарушает условия неотрицательности $x^\varepsilon \geq 0$. Нетрудно видеть, что $x^\varepsilon \in X$, при этом в силу выпуклости функции $c(x)$ имеем оценку

$$c(x^\varepsilon) - c(x^*) \leq \nabla c(x^\varepsilon)(x^\varepsilon - x^*) = \varepsilon \left(\frac{\partial c(x^\varepsilon)}{\partial x_{\bar{q}}} - \frac{\partial c(x^\varepsilon)}{\partial x_{\bar{p}}} \right) < 0,$$

что противоречит оптимальности x^* .

□

В частном случае, когда $G_p(x) \equiv G_p(x_p)$, из теоремы 5 непосредственно следует, что в оптимальном распределении потоков x^* для всякой пары $w \in W$, если $x_p^* > 0$, $p \in P_w$, то $\frac{\partial c_p(x^*)}{\partial x_p} \leq \frac{\partial c_q(x^*)}{\partial x_q}$ для всех $q \in P_w$. При этом, как и в случае равновесных потоков, для оптимальных потоков справедливы равенства $\frac{\partial c_p(x^*)}{\partial x_p} = \frac{\partial c_q(x^*)}{\partial x_q}$ для тех $p, q \in P_w$, для которых $x_p^* > 0$, $x_q^* > 0$. Эти условия известны также как условия Гиббса (см., например, [7]).

Рассмотрим случай, когда транспортные затраты на прохождение каждого пути $p \in P$ складываются только из затрат на проезд по дугам, составляющим этот путь, то есть $G_p(x)$ определяются по формуле (15), при этом затраты по дугам $\tau_e(y)$ описываются аффинными функциями $\tau_e(y) = a_e y_e + b_e$, где a_e и b_e — неотрицательные коэффициенты для всех $e \in E$. При этом функция системных затрат и ее частные производные определяются как

$$c(x) = \sum_{p \in P} \sum_{e \in E} \theta_{ep} \tau_e(y) x_p = \sum_{e \in E} \tau_e(y) y_e = \sum_{e \in E} (a_e y_e^2 + b_e y_e),$$

$$\frac{\partial c(x)}{\partial x_p} = \sum_{e \in E} \theta_{ep} (2a_e y_e + b_e).$$

Обозначим через $y^\dagger = (y_e^\dagger : e \in E)$ и $y^* = (y_e^* : e \in E)$ — загрузку дуг сети, порожденную потоками x^\dagger и x^* соответственно. Опираясь на приведенные выше результаты для равновесных x^\dagger и оптимальных x^* потоков, выполнены следующие условия:

равновесие: если $x_p^\dagger > 0$, $p \in P_w$, то для любого $q \in P_w$ выполнено

$$\sum_{e \in E} \theta_{ep} (a_e y_e^\dagger + b_e) \leq \sum_{e \in E} \theta_{eq} (a_e y_e^\dagger + b_e); \quad (23)$$

оптимальность: если $x_p^* > 0$, $p \in P_w$, то для любого $q \in P_w$ выполнено

$$\sum_{e \in E} \theta_{ep} (2a_e y_e^* + b_e) \leq \sum_{e \in E} \theta_{eq} (2a_e y_e^* + b_e). \quad (24)$$

Любопытно, что при линейных функциях задержек $\tau_e(y) = a_e y_e$ из неравенств (23) и (24) следует совпадение равновесных и оптимальных потоков.

Следуя работе [46], через тройку $[\Gamma, \rho, G(x)]$ обозначим транспортную модель, определенную на сети Γ , с матрицей корреспонденций $\rho = (\rho_w : w \in W)$ и затратами $G(x) = (G_p(x) : p \in P)$. Везде далее будем полагать

$$G_p(x) = \sum_{e \in E} \theta_{ep} \tau_e(y) = \sum_{e \in E} \theta_{ep} (a_e y_e + b_e). \quad (25)$$

Имеет место следующий результат.

Лемма 1. Пусть x^\dagger — решение задачи транспортного равновесия для модели $[\Gamma, \rho, G(x)]$. Тогда вектор $\frac{1}{2}x^\dagger$ является решением оптимизационной задачи для модели $[\Gamma, \frac{1}{2}\rho, G(x)]$.

Доказательство. Если x^\dagger является допустимым решением равновесной модели $[\Gamma, \rho, G(x)]$, то, очевидно, $\frac{1}{2}x^\dagger$ — допустимое решение оптимизационной модели $[\Gamma, \frac{1}{2}\rho, G(x)]$, при этом неравенства (24) для $\frac{1}{2}x^\dagger$ переходят в (23). \square

Более того, для каждого маршрута $p \in P_w$, такого, что $x_p^\dagger > 0$, выполнено

$$\frac{\partial c(\frac{1}{2}x^\dagger)}{\partial x_p} = \sum_{e \in E} \theta_{ep} \{a_e y_e^\dagger + b_e\} = u_w(x^\dagger).$$

Лемма 2. Пусть x^* — оптимальное распределение потоков, отвечающее транспортной модели $[(\Gamma, \rho, G(x))]$. Тогда для любого допустимого потока x^δ в модели $[(\Gamma, (1+\delta)\rho, G(x))]$ справедлива оценка

$$c(x^\delta) \geq c(x^*) + \delta \sum_{w \in W} v_w(x^*) \rho_w, \quad (26)$$

$$\text{зде } \delta \geq 0, v_w(x^*) = \min_{p \in P_w} \frac{\partial c(x^*)}{\partial x_p}.$$

Доказательство. Рассмотрим допустимые относительно модели $[(\Gamma, (1+\delta)\rho, G(x))]$ потоки x^δ . При затратах $G_p(x)$, определенных в (25), где все коэффициенты $a_e \geq 0$, функция $c(x)$ выпукла, отсюда

$$\begin{aligned} c(x^\delta) &\geq c(x^*) + \frac{\partial c(x^*)}{\partial x_p} (x^\delta - x^*) = c(x^*) + \sum_{p \in P} \frac{\partial c(x^*)}{\partial x_p} (x_p^\delta - x_p^*) = \\ &= c(x^*) + \sum_{w \in W} \sum_{p \in P_w} \frac{\partial c(x^*)}{\partial x_p} (x_p^\delta - x_p^*) = \\ &= c(x^*) + \sum_{w \in W} \left(\sum_{p \in P_w} \frac{\partial c(x^*)}{\partial x_p} x_p^\delta - \sum_{p \in P_w} \frac{\partial c(x^*)}{\partial x_p} x_p^* \right). \end{aligned}$$

Поскольку для p , таких, что $x_p^* > 0$, производная $\frac{\partial c(x^*)}{\partial x_p}$ принимает минимальное значение:

$$\frac{\partial c(x^*)}{\partial x_p} = \min_{q \in P_w} \frac{\partial c(x^*)}{\partial x_q} = v_w(x^*),$$

то

$$\sum_{p \in P_W} \frac{\partial c(x^*)}{\partial x_p} x_p^* = \sum_{p \in P_w} v_w(x^*) x_p^*.$$

Следовательно, продолжая оценку снизу для $c(x^\delta)$, получаем

$$\begin{aligned} c(x^\delta) &\geq c(x^*) + \sum_{w \in W} \left(\sum_{p \in P_w} v_w(x^*) x_p^\delta - \sum_{p \in P_w} v_w(x^*) x_p^* \right) = \\ &= c(x^*) + \sum_{w \in W} v_w(x^*) \left(\sum_{p \in P_w} x_p^\delta - \sum_{p \in P_w} x_p^* \right) = \\ &= c(x^*) + \sum_{w \in W} v_w(x^*) ((1 + \delta) \rho_w - \rho_w) = c(x^*) + \delta \sum_{w \in W} v_w(x^*) \rho_w. \end{aligned}$$

□

Итоговый результат текущего раздела устанавливает следующая теорема.

Теорема 6. Для транспортной модели $[\Gamma, \rho, G(x)]$ с аффинными функциями задержек (25) для оптимального x^* и равновесного x^\dagger распределений потоков выполняется соотношение

$$c(x^\dagger)/c(x^*) \leq 4/3.$$

Доказательство. Согласно (21) системные затраты для равновесного распределения x^\dagger рассчитываются как

$$c(x^\dagger) = \sum_{w \in W} u_w(x^\dagger) \rho_w,$$

по лемме 1 поток $\frac{1}{2}x^\dagger$ оптимален для транспортной модели $[\Gamma, \frac{1}{2}\rho, G(x)]$, при этом $v_w(\frac{1}{2}x^\dagger) = u_w(x^\dagger)$.

Положим $\delta = 1$ в оценке (26). Тогда для произвольного потока x , допустимого в модели $[\Gamma, 2\frac{1}{2}\rho, G(x)] = [\Gamma, \rho, G(x)]$, имеем

$$\begin{aligned} c(x) &\geq c\left(\frac{1}{2}x^\dagger\right) + \sum_{w \in W} \frac{1}{2}v_w\left(\frac{1}{2}x^\dagger\right) \rho_w = \\ &= c\left(\frac{1}{2}x^\dagger\right) + \frac{1}{2} \sum_{w \in W} u_w(x^\dagger) \rho_w = c\left(\frac{1}{2}x^\dagger\right) + \frac{1}{2}c(x^\dagger). \end{aligned} \tag{27}$$

Осталось получить оценку снизу $c(\frac{1}{2}x^\dagger)$ в терминах $c(x^\dagger)$, что легко сделать, учитывая вид функций задержки:

$$c\left(\frac{1}{2}x^\dagger\right) = \sum_{e \in E} \frac{1}{2}y_e^\dagger \left(\frac{1}{2}a_e y_e^\dagger + b_e\right) \geq \frac{1}{4} \sum_{e \in E} y_e^\dagger (a_e y_e^\dagger + b_e) = \frac{1}{4}c(x^\dagger),$$

где для промежуточных вычислений использовались потоки по дугам ($y_e^\dagger, e \in E$), индуцированные равновесными потоками по маршрутам x^\dagger . Очевидно, что при этом потоки $\frac{1}{2}x^\dagger$ будут индуцировать загрузку дуг ($\frac{1}{2}y_e^\dagger, e \in E$).

В результате, продолжая оценку (27), получим

$$c(x) \geq \frac{1}{4}c(x^\dagger) + \frac{1}{2}c(x^\dagger) = \frac{3}{4}c(x^\dagger).$$

Вычисляя в последнем неравенстве минимум левой части по всем x , допустимым в модели $[\Gamma, \rho, G(x)]$, получаем

$$c(x^\dagger)/c(x^*) \leq 4/3.$$

□

1.2. Построение матрицы корреспонденций

В задаче транспортного равновесия с фиксированным спросом корреспонденция ρ_w , $w = (i, j)$ рассматривается как средний поток пользователей, который из источника $i \in S$ должен прибыть в сток $j \in D$. В данном разделе вместо ρ_w будем использовать обозначение ρ_{ij} , чтобы выделять характеристики источников i и стоков j .

Существуют разные методики для вычисления элементов матрицы $\rho = (\rho_{ij} : i \in S, j \in D)$, в том числе с применением математических моделей. Рассмотрим наиболее часто используемые, а именно, гравитационную и энтропийную модели построения матрицы корреспонденций. Описание указанных моделей для транспортных сетей можно найти, например, в работах [3–5; 13, 14, 20, 31].

1.2.1. Гравитационная модель

Идею к построению гравитационной модели дал всемирный закон тяготения, утверждающий, что *все тела притягиваются друг к другу с силой, прямо пропорциональной произведению масс этих тел и обратно пропорциональной квадрату расстояния между ними*. Применительно к транспортной системе в качестве тел выступают пункты, порождающие/поглощающие потоки, за массу тела принимается суммарный объем выезжающих/въезжающего потока, физическое расстояние можно заменить на любые другие затраты, связанные с передвижением. В самом простой форме гравитационная модель имеет вид

$$\rho_{ij} = \kappa \frac{s_i d_j}{c_{ij}^2}, \quad i \in S, j \in D. \quad (28)$$

где s_i — общий объем выезжающих из пункта $i \in S$, d_j — общий объем въезжающих в пункт $j \in D$, c_{ij} — удельные затраты на передвижение из i в j , $\kappa > 0$ — калибровочный коэффициент.

Система (28) обладает существенным недостатком. Нетрудно видеть, что при увеличении объемов s_i и d_j , например, в два раза, модель (28) приведет к увеличению корреспонденции ρ_{ij} в четыре раза, что совершенно нелогично. Поэтому вместо классической гравитационной модели (28) на практике используют ее модификацию, в которой к условию (28) добавляют дополнительные условия, например, балансовые ограничения на выезд и въезд. Кроме того, квадрат расстояния (затрат) c_{ij}^2 заменяют на так называемую функцию тя-

готения $f(c_{ij})$, характеризующую предпочтения индивидуумов при выборе пары источник–сток (i, j) для передвижения. В результате модифицированная гравитационная модель имеет вид

$$\rho_{ij} = \frac{s_i d_j}{f(c_{ij})}, \quad \sum_{j=1}^n \rho_{ij} = s_i, \quad \sum_{i=1}^m \rho_{ij} = d_j, \quad \rho_{ij} \geq 0, \quad i \in S, j \in D,$$

или, что то же:

$$\rho_{ij} = \alpha_i \beta_j s_i d_j f(c_{ij}), \quad i \in S, j \in D, \quad (29)$$

где калибровочные коэффициенты α_i и β_j определяются из системы

$$\alpha_i = \left[\sum_{j \in D} \beta_j d_j f(c_{ij}) \right]^{-1}, \quad \beta_j = \left[\sum_{i \in S} \alpha_i s_i f(c_{ij}) \right]^{-1}. \quad (30)$$

Очевидно, что система будет совместной только тогда, когда суммарные объемы по выезду и въезду равны $\sum_{i \in S} s_i = \sum_{j \in D} d_j$.

Выбор функции тяготения f осуществляется либо в процессе калибровки модели на основе сопоставления расчетных данных по модели и эмпирических наблюдений, либо на основе некоторых соображений о предпочтениях при выборе пары источник–сток. Одна из аппроксимаций функции имеет следующий вид: $f(c_{ij}) = \exp(-\gamma c_{ij}^\theta)$, где при расчете корреспонденций трудовых миграций полагают $\gamma \approx 0.065$, $\theta \approx 1$ (см., например, [20] и ссылки там).

Важно отметить, что величины α_i и β_j зависят от всего набора s_i и d_j , а следовательно, и объемы корреспонденций ρ_{ij} зависят от загрузки всей системы.

Численные значения α_i и β_j определяют по специальной итеративной процедуре. В отечественной литературе такая процедура известна как метод балансировки Шацкого–Шелейховского [19, 22]. В зарубежной литературе метод балансировки имеет свою независимую историю развития. Например, в работе [24] описана следующая процедура: начиная с матрицы

$$\rho_{ij}^0 = s_i d_j f(c_{ij}) \left[\sum_{l \in D} d_l f(c_{il}) \right]^{-1},$$

каждая итерация метода состоит из последовательности операций:

$$\begin{aligned} \varrho_{ij}^k &= \begin{cases} \rho_{ij}^k d_j \left[\sum_{i \in S} \rho_{ij}^k \right]^{-1}, & \text{если } \sum_{i \in S} \rho_{ij}^k > d_j, \\ \rho_{ij}^k & \quad \text{— в противном случае} \end{cases} \\ q_i &= s_i - \sum_{j \in D} \varrho_{ij}^k, \quad r_j = d_j - \sum_{i \in S} \varrho_{ij}^k; \\ \rho_{ij}^{k+1} &= \varrho_{ij}^k + q_i r_j f(c_{ij}) \left[\sum_{l \in D} r_l f(c_{il}) \right]^{-1}. \end{aligned} \quad (31)$$

Вычислительные эксперименты по расчету трудовых корреспонденций на примере УДС г. Владивостока [12] (строилась матрица размерности 638×638) показали высокую скорость сходимости процесса (31) к искомой матрице корреспонденций — сбалансированная матрица была получена всего за 4 итерации.

1.2.2. Энтропийная модель

Как и в случае гравитационного подхода, идею построения энтропийной модели подсказала физика, а именно второй закон термодинамики, утверждающий, что любая замкнутая физическая система стремится достичь устойчивого равновесного состояния, которое характеризуется максимумом энтропии этой системы. Впервые концепция энтропии для определения матрицы корреспонденций была использована в работе [47].

Транспортную систему как систему передвижения индивидуумов по УДС города объединяет с физической наличием очень большого числа неуправляемых элементов. При определенных допущениях, например, таких как неизменность затрат на проезд по маршрутам, неизменность топологии УДС (исключаются реконструкция, введение новых, закрытие старых дорог) и т.п., транспортную систему можно считать замкнутой. Таким образом, проблему определения корреспонденций ρ_{ij} можно ставить как задачу максимизации энтропии в транспортной системе.

Пусть задано фиксированное пространственное распределение населения по зонам, порождающим потоки, как и ранее, назовем такие зоны *источниками* и объединим их в множество S , и по зонам, порождающим потоки, назовем их *стоками* и объединим в множество

D. Источниками, например, могут служить районы жилых массивов, стоками — места приложения труда. Индивидуумы в транспортной системе перемещаются от источников к стокам. Предположим, что все индивидуумы имеют уникальный идентификатор, например, номер паспорта. Состояние транспортной системы определяется распределением «помеченных» индивидуумов между парами источник–сток.

При определении объемов корреспонденций значимым является только общее количество индивидуумов без детализации по составу их идентификаторов. Поэтому каждой паре источник–сток соответствует величина корреспонденции ρ_{ij} — количество индивидуумов, выезжающих из источника $i \in S$ и прибывающих в сток $j \in D$. Очевидно, что существует множество состояний, приводящих к одной и той же матрице корреспонденций $\rho = \{\rho_{ij} : i \in S, j \in D\}$. Следуя принципу максимизации энтропии, будем искать значения ρ_{ij} , доставляющие максимум функции $P(\rho)$, определяющей вероятность реализации состояния системы, соответствующего матрице корреспонденций ρ .

Обозначим через $\nu(\rho)$ — вероятность каждой реализации матрицы ρ , через $Q(\rho)$ — количество состояний системы, соответствующих ρ . Тогда

$$P(\rho) = \nu(\rho)Q(\rho). \quad (32)$$

Пусть в системе имеется n источников и m стоков. Обозначим через $\mathcal{R} = \sum_{i=1}^n \sum_{j=1}^m \rho_{ij}$ — общее количество индивидуумов в системе, через $\nu_{ij} > 0$ — вероятность выбора индивидуумом коммуникации ρ_{ij} .

По аналогии со схемой Бернулли значение $\nu(\rho)$ определяется формулой

$$\nu(\rho) = \nu_{11}^{\rho_{11}} \cdot \nu_{12}^{\rho_{12}} \cdot \dots \cdot \nu_{nm}^{\rho_{nm}} = \prod_{i=1}^m \prod_{j=1}^n \nu_{ij}^{\rho_{ij}}.$$

Вычислим количество состояний $Q(\rho)$. Если объем корреспонденции из источника 1 в сток 1 равен ρ_{11} , то количество способов достижения этого объема равно $C_{\mathcal{R}}^{\rho_{11}}$. Далее, из оставшейся части индивидуумов количество способов достижения объема корреспонденции ρ_{12} равно $C_{\mathcal{R}-\rho_{11}}^{\rho_{12}}$, корреспонденции ρ_{13} равно $C_{\mathcal{R}-\rho_{11}-\rho_{12}}^{\rho_{13}}$ и так далее.

В итоге получаем следующую формулу для $Q(\rho)$:

$$\begin{aligned}
Q(\rho) &= C_{\mathcal{R}}^{\rho_{11}} \cdot C_{\mathcal{R}-\rho_{11}}^{\rho_{12}} \cdot C_{\mathcal{R}-\rho_{11}-\rho_{12}}^{\rho_{13}} \cdot \dots \cdot C_{\mathcal{R}-\sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \rho_{ij}}^{\rho_{nm}} = \\
&= \frac{\mathcal{R}!}{(\mathcal{R}-\rho_{11})! \rho_{11}!} \cdot \frac{(\mathcal{R}-\rho_{11})!}{(\mathcal{R}-\rho_{11}-\rho_{12})! \rho_{12}!} \cdot \\
&\quad \frac{(\mathcal{R}-\rho_{11}-\rho_{12})!}{(\mathcal{R}-\rho_{11}-\rho_{12}-\rho_{13})! \rho_{13}!} \cdot \dots \cdot \frac{\left(\mathcal{R} - \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \rho_{ij} \right)!}{\rho_{mn}!} = \\
&= \frac{\mathcal{R}!}{\prod_{i=1}^m \prod_{j=1}^n \rho_{ij}!}.
\end{aligned}$$

Очевидно, что результат не зависит от того, в каком порядке берутся корреспонденции ρ_{ij} для вычисления количества способов распределения индивидуумов в системе.

Подставив рассчитанные значения $\nu(\rho)$ и $Q(\rho)$ в формулу (32), получаем критерий выбора наиболее вероятного состояния системы:

$$P(\rho) = \mathcal{R}! \prod_{i=1}^m \prod_{j=1}^n \frac{\nu_{ij}^{\rho_{ij}}}{\rho_{ij}!} \rightarrow \max. \quad (33)$$

Помимо требования максимизации вероятности $P(\rho)$ на значения ρ_{ij} , как правило, накладываются дополнительные условия. Самыми естественными из них являются балансовые ограничения и условия неотрицательности. Пусть в каждой зоне–источнике $i \in S$ задан общий объем выезжающих s_i , в каждой зоне–стоке $j \in D$ — общий объем въезжающих d_j . Рассмотрим только те корреспонденции ρ_{ij} , которые удовлетворяют следующим условиям:

$$\sum_{j=1}^n \rho_{ij} = s_i, \quad \sum_{i=1}^m \rho_{ij} = d_j, \quad \rho_{ij} \geq 0, \quad i \in S, j \in D. \quad (34)$$

Очевидно, для совместности системы суммарный объем выезжающих должен быть равен суммарному объему въезжающих:

$$\sum_{i=1}^m s_i = \sum_{j=1}^n d_j = \mathcal{R}. \quad (35)$$

Дополнительно к условиям баланса (34) введем ограничение на общие затраты при проезде:

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} \rho_{ij} = C, \quad (36)$$

где c_{ij} — удельные затраты на передвижения из источника i в сток j , C — полные затраты в транспортной системе.

Таким образом, проблема построения матрицы корреспонденции $\rho = (\rho_{ij} : i \in S, j \in D)$ сводится к задаче условной оптимизации (33), (34), (36).

Нет сомнений, что в заданной форме (33) функция $P(\rho)$ весьма неприятна для оптимизации. Для удобства максимизации можно воздействовать на $P(\rho)$ любым монотонным оператором, например, прологарифмировать $P(\rho)$ и вместо (33) использовать критерий

$$\ln P(\rho) = \ln \mathcal{R}! + \sum_{i=1}^m \sum_{j=1}^n (\rho_{ij} \ln \nu_{ij} - \ln \rho_{ij}!) \rightarrow \max. \quad (37)$$

Проводя параллель между физической и транспортной системами, было отмечено наличие большого количества неуправляемых элементов, что позволяет предположить, что значения ρ_{ij} достаточно велики. Поэтому вполне правомерно для дальнейшего преобразования критерия (37) использовать формулу Стирлинга $\ln z! = z \ln z - z$, которая справедлива при больших z . Имеем

$$\ln P(\rho) \approx \mathcal{R} \ln \mathcal{R} + \sum_{i=1}^m \sum_{j=1}^n \rho_{ij} \ln \frac{\nu_{ij}}{\rho_{ij}}.$$

При фиксированных объемах выездов s_i и въездов d_j и выполнении равенства (35) величина $\mathcal{R} \ln \mathcal{R}$ постоянна и может быть исключена из критерия.

В результате проведенных преобразований наиболее вероятное состояние транспортной системы будет соответствовать такой матрице корреспонденций ρ , элементы которой удовлетворяют условиям (34), (36) и критерию

$$\sum_{i=1}^m \sum_{j=1}^n \rho_{ij} \ln \frac{\nu_{ij}}{\rho_{ij}} \rightarrow \max. \quad (38)$$

При построении энтропийной модели (34), (36), (38) предполагалось, что известна априорная информация о предпочтении индивидуумом одной коммуникации другой. Если же любое состояние системы принимает с равной вероятностью, то есть для любых пар (i, j) значение ν_{ij} постоянно и определяется как $\nu_{ij} = \frac{1}{mn}$, то вместо критерия (38) рассматривают

$$\sum_{i=1}^m \sum_{j=1}^n \rho_{ij} \ln \frac{1}{\rho_{ij}} \rightarrow \max. \quad (39)$$

Допустимая область, задаваемая условиями (34), (36), образует полиэдральное множество. Целевая функция критерия (38) на допустимой области является строго вогнутой. В самом деле, матрица Гессе для (38) имеет вид диагональной матрицы размерности $mn \times mn$ с элементами на главной диагонали $\{-\frac{1}{x_{ij}}\}$. Такая матрица отрицательно определена для любых m, n и $x_{ij} \geq 0$. Таким образом, задача (34), (36), (38) относится к классу задач выпуклой гладкой оптимизации. Строгая вогнутость целевой функции гарантирует единственность ее решения. Несмотря на свои хорошие свойства, для реальных транспортных сетей задача (34), (36), (38) имеет большую размерность, что в свою очередь серьезно усложняет применение на практике стандартных для этого класса задач численных методов. Так, например, для расчета трудовых корреспонденций в УДС г. Владивостока [12] территория города была поделена на зоны 800×800 метров. В результате получилась сетка 22×29 квадратов, каждый из которых одновременно являлся зоной–источником и зоной–стоком, при этом размерность задачи (34), (36), (38) составила 407 044 переменных, 1277 ограничений равенств (34), (36).

Для решения задачи (34), (36), (38) разработана простая итерационная схема [19, 22]: начиная с матрицы $\{\rho_{ij} = i\nu_{ij}\}$ на каждой итерации метода попарно достигается выполнение балансовых ограничений для выездов и въездов:

$$\varrho_{ij}^k = \rho_{ij}^k s_i \left[\sum_{j \in D} \rho_{ij}^k \right]^{-1}, \quad \rho_{ij}^{k+1} = \varrho_{ij}^k d_j \left[\sum_{i \in S} \varrho_{ij}^k \right]^{-1}. \quad (40)$$

В работе [2] доказана сходимость процесса (40) к оптимальному решению задачи (34), (36), (38). Существуют и другие подходы к решению энтропийных моделей (см., например, [6, 31]).

Подробнее генезис и феноменология энтропийных моделей для поиска равновесного состояния макросистем, в том числе транспортных, рассмотрена в приложении Е. В. Гасниковой настоящего пособия. Особый интерес тут представляет связь энтропийного критерия с динамикой достижения равновесного состояния.

1.2.3. Связь между гравитационной и энтропийной моделями

Количество переменных в задаче (34), (36), (38), как правило, во много раз превышает число ограничений. Традиционно в такой ситуации вместо исходной решается двойственная задача, которая в данном случае заключается в максимизации функции Лагранжа:

$$L(\rho, \lambda, \mu, \gamma) = \sum_{i=1}^m \sum_{j=1}^n \left[\rho_{ij} \ln \frac{\nu_{ij}}{\rho_{ij}} + \lambda_i (s_i - \rho_{ij}) + \mu_j (d_j - \rho_{ij}) + \gamma (C - c_{ij} \rho_{ij}) \right],$$

где $\lambda = (\lambda_i : i \in S)$ — вектор двойственных переменных, соответствующих балансовым ограничениям (34) для источников, $\mu = (\mu_j : j \in D)$ — вектор двойственных переменных, соответствующих балансовым ограничениям (34) для стоков, γ — двойственная переменная, соответствующая ограничению по затратам (36).

Точка максимума для $L(\rho, \lambda, \mu, \gamma)$ должна удовлетворять условиям (34), (36) и системе уравнений

$$\ln \frac{\nu_{ij}}{\rho_{ij}} - 1 - \lambda_i - \mu_j - \gamma c_{ij} = 0, \quad i \in S, j \in D. \quad (41)$$

Из системы (41) можно выразить корреспонденции

$$\rho_{ij} = \nu_{ij} \exp(-1 - \lambda_i - \mu_j - \gamma c_{ij}). \quad (42)$$

Видим, что для $\nu_{ij} \geq 0$ условие неотрицательности корреспонденций ρ_{ij} выполнено автоматически, поэтому может не учитываться при построении двойственной задачи и применении к ней численных методов. Однако заметим, что случай, когда $\nu_{ij} = 0$ означает отсутствие корреспонденции между парой (i, j) , следовательно, $\rho_{ij} = 0$ и

максимизация функции Лагранжа должна рассматриваться в пространстве меньшей размерности.

Введем обозначения

$$\alpha_i = \frac{\exp(-1 - \lambda_i)}{s_i}, \quad \beta_j = \frac{\exp(-\mu)}{d_j}.$$

Тогда выражение (42) перепишется в виде

$$\rho_{ij} = \alpha_i \beta_j s_i d_j \nu_{ij} \exp(-\gamma c_{ij}). \quad (43)$$

При подстановке (43) в балансовые ограничения (34) определяются параметры α_i и β_j :

$$\alpha_i = \left[\sum_{j \in D} \beta_j d_j \nu_{ij} \exp(-\gamma c_{ij}) \right]^{-1}, \quad \beta_j = \left[\sum_{i \in S} \alpha_i s_i \nu_{ij} \exp(-\gamma c_{ij}) \right]^{-1}.$$

Отметим, что величина C на практике, как правило, не известна, поэтому лагранжевый множитель γ нельзя определить из решения уравнения (36). Значение γ определяется обычными методами калибровки.

Сравнивая выражение (43) с гравитационной моделью (29), видим, что отличие между ними состоит только в аналитическом задании функции тяготения $f(c_{ij})$. При $f(c_{ij}) = \nu_{ij} \exp(-\gamma c_{ij})$ гравитационная (29) и энтропийная (34), (36), (38) модели эквивалентны. Таким образом, при однородной цели поездок, при заданных объемах выездов s_i , въездов d_j , затратах на передвижение c_{ij} , при фиксированных полных затратах C существует наиболее вероятное распределение поездок между зонами (i, j) и это распределение совпадает с тем, которое задается гравитационной моделью с экспоненциальной функцией притяжения.

1.3. Парадоксы транспортного равновесия

В данном разделе рассматривается ряд антиинтуитивных примеров транспортных ситуаций, в которых применение принципа равновесия приводит к неожиданным решениям.

1.3.1. Парадокс Брайеса

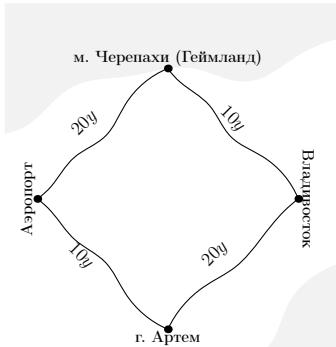
Пример Пигу (см. раздел 1.1.6) заставляет усомниться в эффективности «невидимой руки рынка» Адама Смита, которая, направляя эгоистичные действия пользователей сети, позволяет достичь общественного блага. Последующий пример Брайеса показывает, что конкурентное бескоалиционное равновесие может не только отклоняться от системного оптимума, но и ухудшать ситуацию для всех участников движения.

Рассмотрим появление парадокса Брайеса в результате последовательных весьма вероятных трансформаций транспортной сети окрестностей г. Владивостока, которые представлены в серии рис. 2. Будем рассматривать ситуацию с точки зрения перевозок Аэропорт – Владивосток, с общей потребностью в перевозках 6 условных единиц. На рис. 2 изображены воображаемые этапы изменения участка транспортной сети в окрестности Владивостока, которые могут быть связаны с созданием игровой зоны на мысе Черепахи.

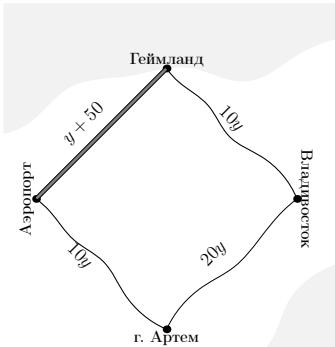
Начальное состояние а): Аэропорт и Владивосток соединены двумя дорогами, одна из которых проходит через г. Артем, а другая через мыс Черепахи — место, где будет построена игровая зона Геймланд. В начальный момент обе дороги невысокого качества и, как показано на рис. 2 а), время проезда по ним сильно зависит от нагрузки y .

Очевидно, что в силу симметрии равновесные потоки распределятся поровну между двумя маршрутами Аэропорт – м. Черепахи – Владивосток и Аэропорт – Артем – Владивосток с соответствующими потоками $y = 3$. Пользовательские затраты на проезд — 90, системные — 540.

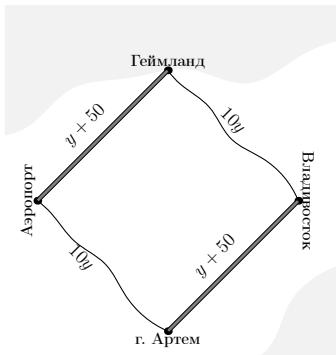
Построено шоссе Аэропорт – Игровая зона б): Снизилась зависимость времени проезда из Аэропорта в Геймланд, в затратах на



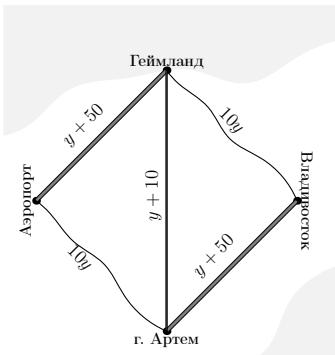
а) Исходное состояние сети



б) Строительство шоссе
Аэропорт – Игровая зона



в) Строительство шоссе
Артем – Владивосток



г) Строительство шоссе
Артем – Игровая зона

Рис. 2. Парадокс Брайеса

проезд появилась постоянная составляющая, которая может представлять собой время проезда по пустой дороге. Часть равновесного трафика переместилась на направление Аэропорт – м. Черепахи – Владивосток (3.17), соответственно поток по другому маршруту Аэропорт – Артем – Владивосток упал до 2.83. Пользовательские затраты на проезд составили 84.88, системные — $84.88 \cdot 6 = 509.28$.

Построено шоссе Артем – Владивосток в): Полученные доходы от игорного бизнеса позволили модернизировать часть одного из маршрутов Владивосток – Артем – Аэропорт, в результате чего равновесные потоки снова стали симметричными и затраты пользователей составили 83, а системные — 498, дальнейшее снижение.

Построено шоссе Артем – Игровая зона г): Поскольку игорная зона обслуживается в основном жителями Артема, был поставлен и положительно решен вопрос о строительстве дороги Артем – Игровая зона. Затраты на проезд соответствовали классу уже построенных дорог, а постоянное слагаемое уменьшилось в силу территориальной близости.

Равновесные потоки теперь распределяются по трем маршрутам: Владивосток – Геймланд – Артем – Аэропорт, Владивосток – Геймланд – Аэропорт, Владивосток – Артем – Аэропорт, причем нагрузка на каждый из них будет составлять 2 единицы трафика, а пользовательские затраты на проезд неожиданно возросли до 92, а системные – до 552 !

В результате как системные, так и личные затраты превзошли даже первоначальный уровень а) несмотря на то, что на каждом предыдущем этапе мы улучшали транспортную ситуацию как с пользовательской, так и с системной точки зрения.

Причиной этого эффекта является то, что постройка на этапе г) шоссе Артем – Игровая зона создала оппортунистическую возможность проехать по маршруту Аэропорт – Артем – Игровая зона Геймланд – Владивосток. При нулевом потоке на маршруте Артем – Игровая зона Геймланд временные затраты составляют 70 и провоцируют водителей на выбор именно этого маршрута. Однако когда эта идея овладеет массами, то поток по сегменту Артем – Игровая зона Геймланд будет уже ненулевой, что увеличит соответствующие общие затраты. Равновесная ситуация установится при одинаковых

затратах (временах) по всем маршрутам, что и приводит к этому парадокльному результату.

1.3.2. Транспортно-экологические парадоксы

Существует ряд парадоксов [41], связанных с транспортными ситуациями, в которых помимо времени проезда учитываются и дополнительные критерии. Одним из таких критериев, важных в настоящее время, является загрязнение окружающей среды (ЗОС). ЗОС является сложным многокомпонентным понятием, включающим различные виды ущерба для окружающей среды: газовое и тепловое загрязнение, разрушение сложившихся природных ландшафтов, мест обитания редких животных и пр. В данном случае будем все же считать, что ЗОС измеряется некоторым универсальным показателем, связанным с данным участком дороги и зависящим, вообще говоря, от потока транспорта по этой дороге. ЗОС от различных участков дороги суммируются, образуя итоговый ЗОС либо от маршрутов, либо от всей транспортной сети в целом.

1.3.2.1. Экологический парадокс Брайеса

В своей схематической форме сеть, реализующая парадокс Брайеса, представлена на рис. 3, где на каждом ребре показаны как временные затраты $\tau(y)$, так и экологический ущерб $e(y)$ как функции потоков по этим ребрам y . Взяв для предполагаемого потока те же данные, что в предыдущем примере, оценим ЗОС до и после строительства новой дороги. Для исходного состояния сети ЗОС оценивается как

$$E = 2 \cdot 3 \cdot (0.2 + 0.1) = 1.8 .$$

После строительства новой дороги с нулевым экологическим ущербом ЗОС становится равным

$$E = 4 \cdot 0.2 + 2 \cdot 0.1 + 2 \cdot 0 + 4 \cdot 0.2 + 2 \cdot 0.1 = 2!$$

Как нетрудно понять, в данном случае парадокс вызван тем, что в результате перераспределения потоков увеличились потоки именно по тем дугам, которые имеют максимальные удельные приращения ЗОС.

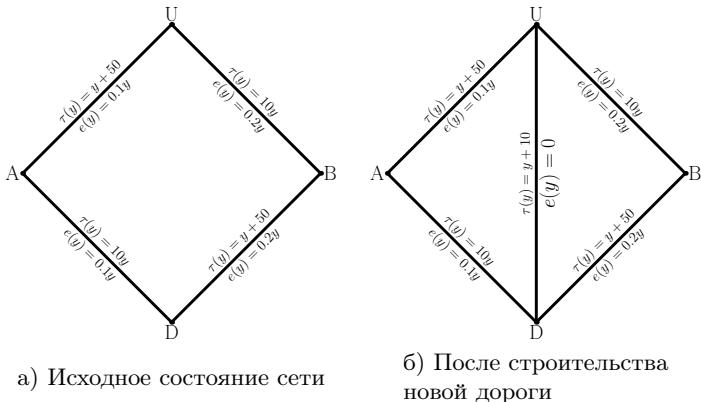


Рис. 3. Парадокс Брайеса

1.3.2.2. Экологический треугольник

Рассмотрим теперь еще более простую транспортную сеть, представленную на рис. 4. Так же, как и ранее, на дугах этой сети при-

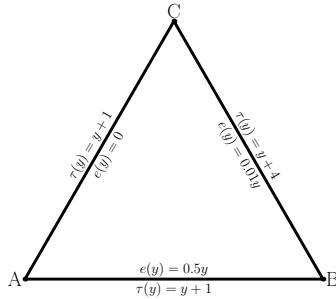


Рис. 4. Уменьшение перевозок вызывает увеличение ЗОС

ведены формулы, описывающие временные затраты $\tau(y)$ и экологический ущерб $e(y)$ как функции потока y по этой дуге. Пусть требуется перевезти 2 единицы груза из C в B и одну единицу из C в A . Очевидно, что достаточно рассмотреть 3 маршрута: $p_1 = C \rightarrow A$, $p_2 = C \rightarrow A \rightarrow B$ и $p_3 = C \rightarrow B$. Условия равновесия совмест-

но с условиями удовлетворения спроса на перевозки дают систему уравнений

$$\begin{aligned}(x_1 + x_2) + 1 + x_2 + 1 &= x_3 + 4, \\ x_2 + x_3 &= 2, \quad x_1 = 1,\end{aligned}$$

где $x_i, i = 1, 2, 3$ — это потоки по маршрутам $p_i, i = 1, 2, 3$. Решение этой системы дает равновесные потоки $x_1^\dagger = x_2^\dagger = x_3^\dagger = 1$ с общим экологическим ущербом $E = 0.51$.

Теперь предположим, что спрос на перевозки по маршруту $C \rightarrow \rightarrow A$ упал до $1/2$. Тогда решение аналогичной системы дает $x_1^\dagger = 1/2$, $x_2^\dagger = 7/6 > 1$, $x_3^\dagger = 5/6 < 1$ с общим экологическим ущербом $E = 0.5 \cdot 7/6 + 0.01 \cdot 5/6 \approx 0.591 > 0.51$.

Заметим, что увеличение экологического ущерба в этом случае вызвано уменьшением нагрузки на экологически чистую дугу сети. С одной стороны, это вызвало переход части трафика с маршрута p_3 , не проходящего через дугу $C \rightarrow A$ на маршрут p_2 , проходящий через эту дугу, однако, с другой стороны, это вызвало увеличение трафика по дуге $A \rightarrow B$ с высоким экологическим ущербом и суммарный эффект оказался негативным.

1.3.2.3. Рокадная экология

Последующий пример показывает, как такая популярная мера, как строительство рокадной дороги улучшенного качества, может на самом деле ухудшить экологическую ситуацию.

Рассмотрим дорожную сеть, изображенную на рис. 5, часть а). Предположим, что эта сеть предназначена для перемещения автомо-

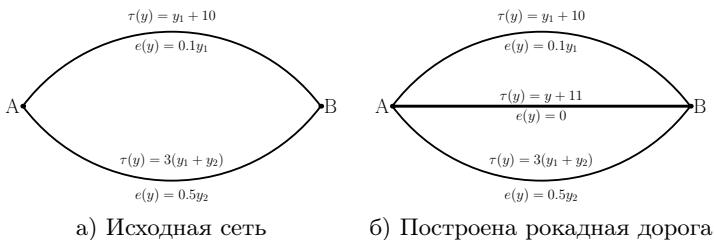


Рис. 5. Новая рокадная дорога вызывает увеличение ЗОС

бильного трафика в объеме 5 условных единиц из точки A в точку

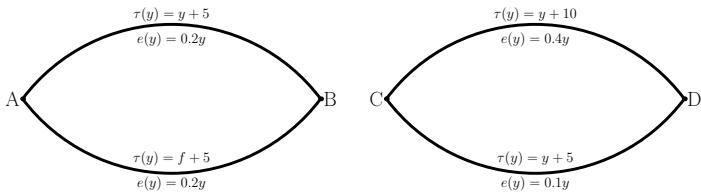


Рис. 6. Сети автомобильных дорог и трамвайных линий

B по двум дублирующим дорогам различного качества, временные затраты по которым задаются соотношениями на соответствующих дугах. В этих соотношениях y_1 означает поток по верхней дуге, y_2 — по нижней. Зависимость времени проезда по нижней дуге от потока по верхней может быть вызвана указанием приоритета на соответствующем перекрестке.

Определяющая система уравнений для равновесных потоков имеет вид

$$y_1 + y_2 = 5, \quad y_1 + 10 = 3(y_1 + y_2),$$

откуда $y_1^\dagger = 5$, $y_2^\dagger = 0$. Затраты пользователей на проезд составляют при этом $\tau^\dagger = 15$, а экологический ущерб составляет 0.5.

Если, как показано на правой части рис. 5, построена новая дорога с нулевым ущербом для окружающей среды и временными характеристиками $\tau(y) = y + 11$, то новая определяющая система будет иметь вид

$$y_1 + y_2 + y = 5, \quad y_1 + 10 = 3(y_1 + y_2) = y + 11,$$

и ее решение: $y_1^\dagger = y_2^\dagger = 2$, $y^\dagger = 1$. Затраты пользователей на проезд при этом уменьшились до $\tau_\dagger = 12$, а экологический ущерб возрос до 1.2 !

1.3.2.4. Перераспределение спроса

Еще один пример показывает, что перенос части пассажиров с транспортного средства экологически более вредного (скажем автобус) на менее вредный (например, трамвай) может в действительности увеличить ЗОС. В этой модели рассматривается сеть, состоящая

из двух компонент, отражающих различные виды транспорта. В этих компонентах необходимо перевезти 10 единиц потока из A в B и 5 единиц потока из C в D . Пункты отправления A и C , а также пункты прибытия C и B будем считать близкими, так что между ними возможен обмен перевозками.

В такой транспортной схеме появляются 4 маршрута, каждый из которых состоит из одной дуги, которые мы будем считать про-нумерованными сверху вниз. Затраты на перевозку по маршрутам совпадают, конечно, с затратами по дугам и обозначены на рисунке.

При этих начальных данных решение задачи равновесия имеет вид:

Автомобильная сеть. В силу симметрии потоки равны и составляют $y_1^\dagger = y_2^\dagger = 5$. Экологический ущерб $E = 2$.

Трамвайная сеть. $y_3^\dagger = 0$, $y_4^\dagger = 5$. Экологический ущерб $E = 0.5$.

Суммарный экологический ущерб составляет 2.5.

Предположим теперь, что 2.5 единицы трафика перенесены из сильно загрязняющей автомобильной сети в менее загрязняющую трамвайную с целью уменьшить суммарный ЗОС. Новое равновесное решение будет иметь вид:

Автомобильная сеть. В силу симметрии потоки равны и составляют $y_1^\dagger = y_2^\dagger = 3.75$. Экологический ущерб $E = 1.5$.

Трамвайная сеть. $y_3^\dagger = 0$, $y_4^\dagger = 5$. Экологический ущерб $E = 1.125$.

Суммарный экологический ущерб составляет 2.625, что превосходит (!) ущерб в предыдущем варианте распределения нагрузки для двух видов трафика.

То, что перенос трафика осуществляется в размере 2.5 единиц на самом деле несущественно, любое уменьшение объема перевозок по автомобильной сети вызывает появление ненулевого объема перевозок по экологически затратной дуге 3 в трамвайной сети, что не компенсируется уменьшением объемов перевозок по дугам 1, 2. Увеличение объемов перевозок по автомобильной сети также не приводит к уменьшению ЗОС, так как вызывает в два раза больший экологический ущерб, чем уменьшение ЗОС от трамвайного трафика, который будет продолжать концентрироваться на дуге 4.

1.4. Практическая работа

Упражнение 1. На входе в Великий Федеральный Университет есть две двери, через которые входят и выходят студенты. При подходе к ним перед каждым студентом возникает проблема, какой дверью воспользоваться, если он заинтересован в скорейшем входе или выходе. Найти равновесное распределение потоков в двух случаях:

- 1) задержка в дверях одинакова для входящих и выходящих и пропорциональна произведению интенсивностей входящего и выходящего потоков;
- 2) задержка в дверях пропорциональна интенсивности потока студентов, двигающихся в том же направлении плюс задержка, пропорциональная произведению интенсивностей входящего и выходящего потоков.

Общее количество входящих и выходящих за единицу времени студентов считать одинаковым.

Упражнение 2. Рассматривается транспортная сеть $\Gamma = (V, E)$ (пример сети взят из работы [27]), состоящая из 25 вершин ($|V| = 25$) и 40 ориентированных дуг ($|E| = 40$). Топология сети с направлением дуг представлена на рис. 7.

Дороги (дуги) транспортной сети Γ поделены на четыре категории:

- 1) магистрали $E_h = \{(6 \rightarrow 7), (8 \rightarrow 9), (10 \rightarrow 11), (12 \rightarrow 13), (14 \rightarrow 15), (17 \rightarrow 18), (19 \rightarrow 20), (21 \rightarrow 22), (23 \rightarrow 24), (25 \rightarrow 16)\};$
- 2) выезды $E_{ex} = \{(16 \rightarrow 1), (15 \rightarrow 1), (24 \rightarrow 2), (7 \rightarrow 2), (22 \rightarrow 3), (9 \rightarrow 3), (20 \rightarrow 4), (11 \rightarrow 4), (18 \rightarrow 5), (13 \rightarrow 5)\};$
- 3) въезды $E_{en} = \{(1 \rightarrow 6), (1 \rightarrow 17), (2 \rightarrow 25), (2 \rightarrow 8), (3 \rightarrow 23), (3 \rightarrow 10), (4 \rightarrow 21), (4 \rightarrow 12), (5 \rightarrow 19), (5 \rightarrow 14)\};$
- 4) второстепенные дороги $E_s = \{(15 \rightarrow 6), (7 \rightarrow 8), (9 \rightarrow 10), (11 \rightarrow 12), (13 \rightarrow 14), (16 \rightarrow 17), (18 \rightarrow 19), (20 \rightarrow 21), (22 \rightarrow 23), (24 \rightarrow 25)\}.$

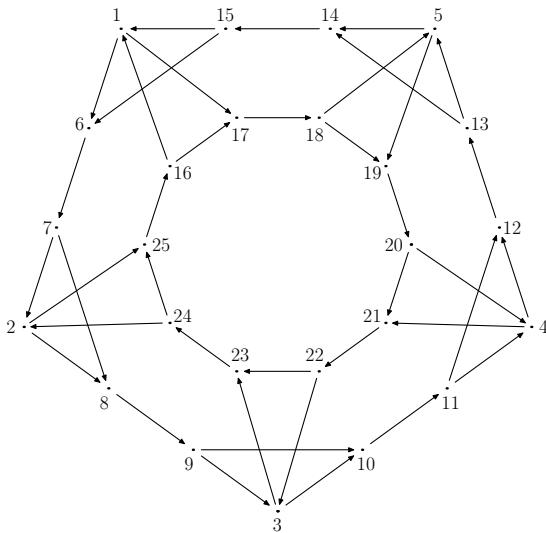


Рис. 7. Транспортная сеть $\Gamma = (V, E)$

Пропускная способность магистралей равна 100 единицам потока, остальных дуг — 50 ед.

Категория дороги влияет на затраты при передвижении. Минимальные транспортные затраты τ_e^0 по каждому из участков $e \in E$ определяются по формуле $\tau_e^0 = \kappa_e l_e$, где l_e — длина дуги e , данные приведены в таблице 1, $\kappa_e > 0$ — коэффициент, зависящий от категории, которой принадлежит дуга e :

$$\kappa_e = \begin{cases} 0.011, & e \in E_h, \\ 0.025, & e \in E_{ex} \cup E_{en}, \\ 0.033, & e \in E_s. \end{cases}$$

В сети Γ выделено пять вершин-источников $S = \{1, 2, 3, 4, 5\}$ с заданными объемами выходящего трафика $s = (69, 90, 10, 100, 53)$ и пять вершин-стоков $D = \{17, 19, 21, 23, 25\}$ с заданными объемами входящего трафиков $d = (128, 59, 34, 61, 40)$.

Используя аппарат математического моделирования и численные методы, рассчитать объемы корреспонденций ρ_w для всех пар источник–сток $w \in W = S \times D$.

Таблица 1. Длины дуг сети $\Gamma = (V, E)$

Дуга	Длина	Дуга	Длина	Дуга	Длина
$6 \rightarrow 7$	4	$16 \rightarrow 1$	6	$1 \rightarrow 6$	3
$8 \rightarrow 9$	10	$15 \rightarrow 1$	9	$1 \rightarrow 17$	7
$10 \rightarrow 11$	3	$24 \rightarrow 2$	3	$2 \rightarrow 25$	6
$12 \rightarrow 13$	3	$7 \rightarrow 2$	8	$2 \rightarrow 8$	2
$14 \rightarrow 15$	5	$22 \rightarrow 3$	5	$3 \rightarrow 23$	6
$17 \rightarrow 18$	1	$9 \rightarrow 3$	1	$3 \rightarrow 10$	5
$19 \rightarrow 20$	2	$20 \rightarrow 4$	10	$4 \rightarrow 21$	6
$21 \rightarrow 22$	6	$11 \rightarrow 4$	8	$4 \rightarrow 12$	8
$23 \rightarrow 24$	9	$18 \rightarrow 5$	5	$5 \rightarrow 19$	7
$25 \rightarrow 16$	2	$13 \rightarrow 5$	3	$5 \rightarrow 14$	4
$15 \rightarrow 6$	1	$7 \rightarrow 8$	6	$9 \rightarrow 10$	4
$11 \rightarrow 12$	3	$13 \rightarrow 14$	9	$16 \rightarrow 17$	10
$18 \rightarrow 19$	4	$20 \rightarrow 21$	6	$22 \rightarrow 23$	10
$24 \rightarrow 25$	1				

При моделировании необходимо учесть, что предпочтения при выборе пары w определяются функцией $f(c_w) = \exp(-0.065c_w)$, где c_w — минимальные транспортные затраты на передвижение для пары w . Предполагается, что величина c_w характеризует длину кратчайшего пути для каждой пары w и определяется из соотношений: $c_w = 0.05$, если источник и сток совпадают, в противном случае

$$c_w = \min_{p \in P_w} \sum_{e \in E} \theta_{ep} \tau_e^0, \quad \theta_{ep} = \begin{cases} 1, & \text{если путь } p \text{ проходит} \\ & \text{через дугу } e; \\ 0, & \text{в противном случае,} \end{cases}$$

где через P_w обозначено множество всех допустимых маршрутов передвижения для пары w .

Для рассчитанных корреспонденций сравнить загрузку транспортной сети, индивидуальные и общие транспортные затраты при нормативном и дескриптивном распределении потоков. Под нормативным распределением понимается централизованное управление движением, имеющее своей целью минимизацию совокупных транспортных затрат (*второй поведенческий принцип Вардрона*). Под дескриптивным — отсутствие централизованного управления, каждый

пользователь выбирает маршрут следования исходя из минимизации собственных транспортных затрат (*первый поведенческий принцип Вардрона*).

Удельные транспортные затраты для каждой пары источник–сток w складываются из затрат по дугам τ_e , входящим в маршрут следования из источника в сток. В свою очередь на значение τ_e влияет величина потока, проходящего по дуге $e \in E$, такая зависимость описывается функцией $\tau_e(y_e) = \tau_e^0(1 + (y_e/c_e)^4)$, где c_e — пропускная способность дуги e .

Литература

1. Ашманов С. А. Математические модели и методы в экономике. М.: Изд-во МГУ, 1980.
2. Брэгман Л. Д. Доказательство сходимости метода Шелейховского для задачи с транспортными ограничениями // ЖВМ и МФ. 1967. № 1. С. 147–156.
3. Васильева Е. М., Левит Б. Ю., Лившиц В. Н. Нелинейные транспортные задачи на сетях. М.: Финансы и статистика, 1981.
4. Васильева Е. М., Игудин Р. В., Лившиц В. Н. Оптимизация планирования и управления транспортными системами. М.: Транспорт, 1987.
5. Вильсон А. Дж. Энтропийные методы моделирования сложных систем. М.: Наука, 1978.
6. Гасникова Е. В. Двойственные мультиплективные алгоритмы для задачи энтропийно-линейного программирования // ЖВМ и МФ. 2009. Т. 49. № 3. С. 453–464.
7. Данскин Дж. Теория максимина и ее приложение к задачам распределения вооружений М.: Сов. радио, 1970.
8. Никайдо Х. Выпуклые структуры и математическая экономика. М.: Мир, 1972.
9. Нурминский Е. А. Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов // ЖВМ и МФ. 2008. Т. 48. № 12. С. 2121–2128.
10. Нурминский Е. А. Фейеровские процессы с малыми возмущениями // Доклады АН. 2008. Т. 422. № 5. С. 601–605.
11. Nurminski E. A. Envelope stepsize control for iterative algorithms based on Fejer processes with attractants // Optimization Methods and Software. 2010. V. 25. № 1. P. 97-108.
12. Нурминский Е. А., Шамрай Н. Б. Моделирование транспортных потоков г. Владивостока на основе теории равновесия // Sisteme de transport si logistica: Materialele Conf. Int.,

Chisinau, 22-23 octombrie 2009; red. Resp. Dumitru Solomon; Acad. de Transporturi, Informatica si Comunicatii. Ch.: Evrica, 2009. P. 334–348.

13. *Попков Ю. С., Порохин М. В., Гутнов А. Э., Шмульян Б. Л.* Системный анализ и проблемы развития городов. М.: Наука, 1983.
14. *Попков Ю. С.* Макросистемные модели пространственной экономики. М.: КомКнига, 2008.
15. *Шамрай Н. Б.* Решение задач транспортного равновесия с декомпозицией по ограничениям // Труды всероссийской конференции "Равновесные модели в экономике и энергетике". – Иркутск: Изд-во ИСЭМ СО РАН. 2008. С. 618–624.
16. *Стенбринк П. А.* Оптимизация транспортных сетей. М.: Транспорт, 1961.
17. *Попов Л. Д.* Введение в теорию, методы и экономические приложения задач о дополнительности. Екатеринбург: Изд-во Урал. ун-та, 2001.
18. *Тодд М. Дж.* Вычисление неподвижных точек и приложения к экономике. М.: Наука, 1983.
19. *Шацкий Ю. А.* Расчет схемы расселения и трудовых корреспонденций при разработке генерального плана города // Развитие системы городского транспорта. Киев. 1971. № 4. С. 3–14.
20. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.
21. *Швецов В. И.* Алгоритмы распределения транспортных потоков // Автоматика и телемеханика. 2009. № 10. С. 148–157.
22. *Шелайховский Г. В.* Транспортные основания композиции городского плана. Л., 1936.
23. *Agdeppa R. P., Yamashita N., Fukushima M.* The traffic equilibrium problem with nonadditive costs and its monotone mixed complementarity problem formulation // Transportation Research Part B. 2007. № 41. P. 862–874.

24. *Arrowsmith G. A.* A behavioural approach to obtaining a doubly constrained trip distribution model // Operational Research Quarterly. 1973. V. 24. № 1. P. 101–111.
25. *Bar-Gera H.* Origin-based algorithm for the traffic assignment problem // Transportation Science. 2002. V. 36. № 4. P. 398–417.
26. *Beckmann M., McGuire C. B., Winsten C. B.* Studies in the economics of transportation. RM-1488. Santa Monica: RAND Corporation, 1955.
27. *Bertsekas D., Gafni E.* Projection methods for variational inequalities with application to the traffic assignment problem // Mathematical Programming Study. 1982. № 17. P. 139-159.
28. *Boyce D., Ralevic-Dekic B., Bar-Gera H.* Convergence of traffic assignments: how much is enough? // Journal Transport Engineer. 2004. V. 130. № 1. P. 49–55.
29. *Chen M., Bernstein D. H., Chien S. I. J., Mouskos K.* Simplified formulation of toll design problem // Transportation Research Record. 1999. № 1667. P. 88–95.
30. *Dafermos S.* Traffic equilibrium and variational inequalities // Transportation Science. 1980. V. 14. № 1. P. 42-54.
31. *Fang S.-C., Rajasekera J. R., Tsao H.-S.J.* Entropy optimization and mathematical programming. Kluwer Academic Publisher, 1997.
32. *Facchinei F., Pang J.-S.* Finite-Dimensional Variational Inequalities and Complementarity Problems (V. I, II). Springer, 2003.
33. *Frank M., Wolfe P.* An algorithm for quadratic programming // Naval Research Logistics Quarterly. 1956. V. 3. P. 95–110.
34. *Gabriel S. A., Bernstein D.* The traffic equilibrium problem with nonadditive path costs // Transportation Science. 1997. V. 31. № 4. P. 337–348.
35. *Janson B., Zozaya-Gorostiza C.* The problem of cyclic flows in traffic assignment // Transportation Research Part B. 1987. V. 21. № 4. P. 299–310.

36. *Knight F. H.* Some fallacies in the interpretation of social cost // The Quarterly Journal of Economics. 1924. V. 38. № 4. P. 582–606.
37. *Konnov I. V.* Combined relaxation methods for variational inequalities. Berlin: Springer, 2001.
38. *Konnov I. V.* Equilibrium Models and Variational Inequalities. Elsevier Science, 2007.
39. *Lo H. K., Chen A.* Traffic equilibrium problem with rout-specific costs: formulation and algorithms // Transportation Research Part B. 2000. V. 34. № 6. P. 493–513.
40. *Nagurney A.* Network Economics: A Variational Inequality Approach. Dordrecht: Kluwer Academic Publishers, 1999.
41. *Nagurney A., Dong J.* Paradoxes in networks with zero emission links: implications for telecommunications versus transportation // Transportation Research Part D. 2001. V. 6. № 4. P. 283–296.
42. *Nemirovsky A., Yudin D.* Informational complexity and efficient methods for solution of convex extremal problems. N.Y.: Wiley, 1983.
43. *Nesterov Y., de Palma A.* Stationary dynamic solutions in congested transportation networks: summary and perspectives // Networks and spatial economics. 2003. № 3. P. 371–395.
44. *Patriksson M.* The traffic assignment problem — models and methods. Utrecht, Netherlands: VSP. 1994.
45. *Piugou A. C.* The economics of welfare, London: MacMillan, 1932, 4-th edition. (Русский перевод: Пигу А.С. Экономическая теория благосостояния Т. 1–2, Сер. Экономическая мысль Запада, М.: Прогресс, 1985).
46. *Roughgarden T., Tardos E.* How bad is selfish routing? // Journal of the ACM. 2002. V. 49. № 2. P. 236–259.
47. *Wilson A. G.* A statistical theory of spatial distribution models // Transportation Research. 1967. V. 1. P. 253–270.
48. *Wardrop J.* Some Theoretical Aspects of Road Traffic Research // Proceedings of the Institute of Civil Engineers, 1952.

Глава 2. Математические модели транспортных потоков¹

2.1. Макроскопические модели	70
2.1.1. Модель Лайтхилла–Уизема–Ричардса (LWR).....	70
2.1.2. Модель Танака	82
2.1.3. Модель Уизема	83
2.1.4. Модель Пэйна и её обобщения	92
2.1.5. Кинетические модели	97
2.1.6. Практические приложения моделей	99
2.2. Микроскопические модели	102
2.2.1. Модель оптимальной скорости Ньюэлла	102
2.2.2. Модель следования за лидером Дженерал Моторс	107
2.2.3. Модель «разумного водителя» Трайбера	109
2.2.4. Модели клеточных автоматов	112
2.3. Модельные задачи	119
2.3.1. Эволюции глобального затора в транспортном потоке, описываемом моделями LWR и Уизема	119
2.3.2. Эволюции локального затора в транспортном потоке, описываемом моделями LWR и Уизема	136
2.3.3. Задача о светофоре (при каких условиях перед светофором не будет скапливаться очередь)	143
2.4. Теория Кернера–Конхойзера движущихся локальных кластеров в моделях Дженерал Моторс класса	146
2.4.1. Фундаментальные эмпирические свойства перехода от свободного транспортного потока к плотному и модели транспортного потока	147
2.4.2. Характеристические параметры широкого движущегося кластера	151
2.4.3 Линия J Кернера	154
Литература.....	157

¹ Глава 2 написана А. В. Гасниковым при участии С. Л. Кленова и Я. А. Холодова. Огромную помощь в написании окказал А. А. Шананин – по сути, идеяный вдохновитель этой главы. Рисунки к пп. 2.1–2.3 этой главы подготовили Андрей Ярошенко (МАДИ) и Игорь Виноградов (ВМиК, МГУ; Институт машиноведения РАН).

2.1. МАКРОСКОПИЧЕСКИЕ МОДЕЛИ

В пункте 2.1 приводятся основные (с исторической точки зрения, и с точки зрения возможных приложений) *макроскопические модели транспортных потоков*. Много внимания уделяется гидродинамическим аналогиям. *Транспортный поток* уподобляется *сжимаемой жидкости с мотивацией*, которая присутствует, например, в *уравнении состояния транспортного потока* (зависимости *скорости потока от плотности*). Ключевым понятием этого пункта является *обобщенное решение* начальной задачи Коши для закона сохранения, описывающего транспортный поток. Так, например, разрывы обобщенного решения интерпретируются как границы заторов (переход от свободного движения к заторному).

2.1.1. Модель Лайтхилла–Уизема–Ричардса (LWR)

Во второй половине 40-х годов и в 50-е 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 уравнений) типа закона сохранения. В это же время наблюдался и рост приложений, в которых встречаются схожие уравнения [3, 4]. Так в 1955 г. независимо в работах [5, 6] (см. также [7]) была предложена, по-видимому, первая макроскопическая (гидродинамическая) модель однополосного² транспортного потока, названная впоследствии *моделью Лайтхилла–Уизема (Уитема)–Ричардса (LWR)*, в которой поток АТС (вместо термина «автомобиль» и тем более «машина» в транспортной литературе принято использовать 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).

² Полоса бесконечная в обе стороны, движение происходит слева направо (для определенности), нет источников и стоков *автомобильных транспортных (автоматранспортных) средств (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 *ATC (автомотранспортных 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, [5, 8]), а как функции³, получающиеся при переходе от *микроскопического описания* (в том числе и описания с помощью *клеточных автоматов* (см. п. 2.2)) 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 $\rho(t, x)$, $v(t, x)$ 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 (ввиду 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

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

³ Отметим в связи с употреблением здесь слова «функция», что В. И. Арнольд неоднократно во время различных выступлений делал докладчикам следующее замечание: «Запись $f(x)$ означает не функцию, а значение функции в точке x , а функцию следует обозначать просто f ». Тем не менее, здесь и далее в этой главе было решено оставить более привычную для студентов запись.

– интенсивность потока АТС (количество АТС, проходящих в единицу времени через заданное сечение). Зависимость $Q(\rho)$ часто называют **фундаментальной (или основной) диаграммой**. Отметим также, что и зависимость $V(\rho)$ иногда называют фундаментальной диаграммой (см., например, приложение М. Л. Бланка). Для однополосного потока принято считать [7]:

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

Это условие 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. Однако если агрегировать несколько полос в одну (иначе говоря, заменить несколько полос одной агрегированной, на которой уже использовать рассматриваемую модель), то, как показывают наблюдения за реальными транспортными потоками, от вогнутости функции $Q(\rho)$, вообще говоря, придется отказаться.

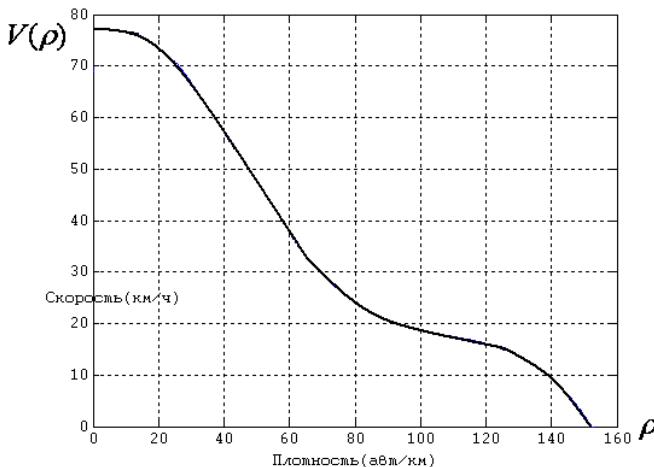


Рис. 1. Уравнение состояния транспортного потока

Так на рис. 1, 2 отображены экспериментальные данные «Центра исследования транспортной инфраструктуры» г. Москвы, собранные (в

течение одного дня в 2005 г.) по четырем полосам на участке третьего транспортного кольца от Автозаводской улицы до Варшавского шоссе, и сагрегированные на одну полосу. Заметим, что в действительности измерялась зависимость $V(Q)$.

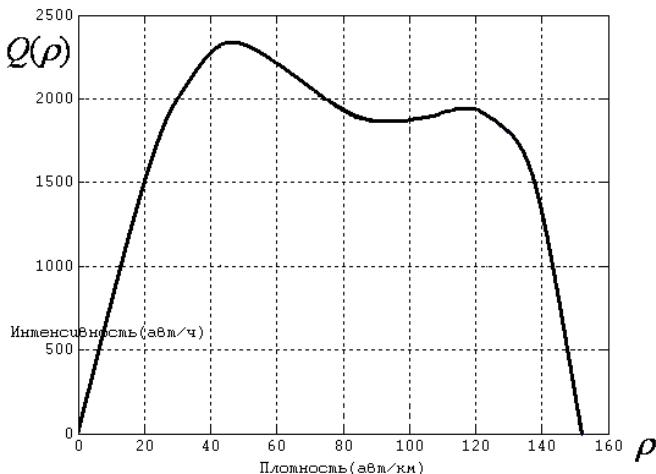


Рис. 2. Фундаментальная диаграмма

Объяснить небольшой провал интенсивности потока $Q(\rho)$ при плотностях $\rho \sim 60-115$ АТС/км можно, по-видимому, тем, что при этих плотностях существенное влияние на интенсивность потока оказывают перемещения АТС с одной полосы на другую. Перестраивания АТС из одной полосы в другую полосу при этих плотностях, снижают интенсивность потока. С одной стороны, за счет перемещения из полосы в полосу можно двигаться быстрее (так оно и происходит при плотностях $\rho \sim 30-50$ АТС/км), но, с другой стороны, в среднем такие перемещения при $\rho \sim 50-120$ АТС/км приводят к дополнительным затратам на само перестраивание, и замедление тех АТС, перед которыми встраивается новое АТС [9]. Другое объяснение этого наблюдения имеется в главе 3 (см. также [10]) и связано с тем, что при $\rho \sim 50-120$ АТС/км само понятие «фундаментальная диаграмма» не совсем корректно. Иначе говоря, при этих плотностях нет четкой зависимости величины потока (скорости) от плотности. Одному

значению плотности соответствует целый промежуток возможных значений потока (скорости).

Второе предположение выразим законом сохранения

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

Отсюда следует, что для любого прямоугольного контура Γ в полу-плоскости $t \geq 0$, $x \in \mathbb{R}$, 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 $\rho(t, x)$:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(v\rho)}{\partial x} = \frac{\partial \rho}{\partial t} + \frac{\partial(V(\rho)\rho)}{\partial x},$$

т.е.

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

Поставим начальное условие вида (*условие типа Римана*)

$$\rho(0, x) = \begin{cases} \rho_-, & x < x_-, \\ \rho_0(x), & x_- \leq x < x_+, \\ \rho_+, & x \geq x_+. \end{cases} \quad (5)$$

Задача Коши (4), (5) возникает, например, при описании распространения затора (пробки): пусть

$$\rho'_0(x) \geq 0, \quad \rho_+ = \rho_{\max},$$

где ρ_{\max} – максимально возможная плотность (ситуация «бампер к бамперу»). Требуется определить, как по транспортному потоку будет распространяться информация о заторе впереди. Решение этой задачи позволит ответить, например, на следующий вопрос: если движение АТС с утра на Дмитровском шоссе в сторону Москвы «встало» в районе г. Долгопрудный, то через какое время затор дойдет до г. Дмитрова? Ряд интересных модельных задач (задача о светофоре, об эволюции локаль-

ногого затора и др.) для закона сохранения (4) рассмотрен во 2-й и 3-й главе книги [7] (см. также п. 2.3).

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. 3. 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. 3).

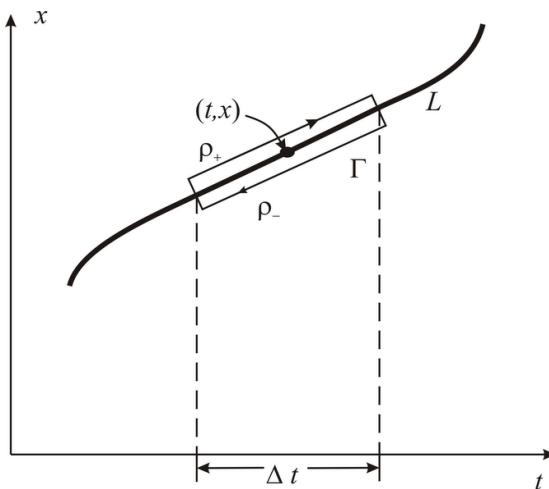


Рис. 3. RH-условие на разрыве

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 [11]) условием Римана–Ранкина–Гюгонио:

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

Оказывается, что уравнение (4) всегда имеет слабое (удовлетворяющее соотношению (3) и начальному условию (5) в слабом смысле) решение (см. [12]), но, как показывает следующий пример, оно может иметь бесконечно много решений, т.е. нет единственности.

Пример (О. А. Олейник [13]). Рассмотрим уравнение Хопфа (см. [14]):

$$\frac{\partial \rho}{\partial t} + \rho \frac{\partial \rho}{\partial x} = 0$$

и начальное условие Римана:

$$\rho(0, x) = \begin{cases} 1, & x \leq 0, \\ -1, & x > 0. \end{cases}$$

О важности (исторической, и не только) именно этого уравнения для моделирования транспортных потоков см. в п. 2.1.3. Линейной заменой переменных и неизвестной функции к уравнению Хопфа можно свести довольно популярный частный случай модели LWR, в котором $Q(\rho)$ – вогнутая парабола (фундаментальная диаграмма Гриншилдса).

При любом $q \geq 1$ определенная в точках полуплоскости $t \geq 0$ функция

$$\rho_q(t, x) = \begin{cases} 1, & x \leq \frac{1-q}{2}t, \\ -q, & \frac{1-q}{2}t < x \leq 0, \\ q, & 0 < x \leq \frac{q-1}{2}t, \\ -1, & \frac{q-1}{2}t < x \end{cases}$$

удовлетворяет при $t > 0$ уравнению (4) в смысле (3) (достаточно проверить, что на разрывах выполняется условие RRH) и начальному условию (5) (см. рис. 4). ■

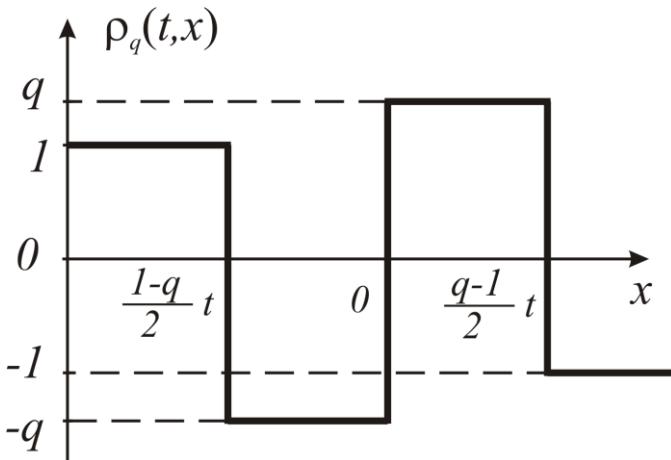


Рис. 4. Пример О. А. Олейник (1957)

Единственное решение выбирает условие отбора, по-видимому, впервые предложенное О. А. Олейник в 1958 г. [15, 16] (в частном случае – выпуклой (вогнутой) функции $Q(\rho)$ см. О. А. Олейник (1957) [13], а также П. Лакс (1957) [11]) и И. М. Гельфандом в 1959 г. [17]⁴ как *условие устойчивости (допустимости) разрыва*.

На разрыве, помимо RRH-условия, также должно выполняться E-условие:

$$\forall \rho \in (\rho_-, \rho_+) \rightarrow \sigma(\rho_-, \rho_+) \leq \sigma(\rho_-, \rho), \text{ если } \rho_- < \rho_+;$$

$$\forall \rho \in (\rho_+, \rho_-) \rightarrow \sigma(\rho_-, \rho_+) \geq \sigma(\rho_-, \rho), \text{ если } \rho_- > \rho_+.$$

Это условие также называют *энтропийным условием*, *энтропийным условием О. А. Олейник*, *E-условием О. А. Олейник*. Объяснение (основанное на методе исчезающей вязкости, см. п. 2.1.3) того, откуда возникло E-условие (и RRH-условие) будет приведено в п. 2.3.1. Заметим также, что в классе кусочно-постоянных начальных условий (аппроксимирую-

⁴ Работа [17] представляет собой запись курса лекций, сыгравшего важную роль в популяризации теории квазилинейных уравнений и законов сохранения, которые И. М. Гельфанд читал на мехмате МГУ в 1957–1958 гг.

щих класс ограниченных измеримых начальных условий) добавление Е-условия как условия отбора возможных разрывов к соотношению (3) однозначно и конструктивно определяет динамику $\rho(t, x)$ (нужно также оговориться, что кусочно-гладкая функция $Q(\rho)$ не имеет точек сгущения нулей второй производной). Отмеченная конструктивность, активно использовалась в 70-х и 80-х годах XX века, например, при исследовании модельных задач п. 2.3.

Е-условие имеет наглядную (см. рис. 5) геометрическую интерпретацию (для определенности считаем $\rho_- < \rho_+$): график функции $Q(\rho)$ при $\rho \in (\rho_-, \rho_+)$ лежит не ниже прямой, проходящей через точки $(\rho_-, Q(\rho_-))$ и $(\rho_+, Q(\rho_+))$. При этом скорость движения разрыва c равна наклону этой прямой.

Для случая вогнутой функции $Q(\rho)$ разрыв устойчив тогда и только тогда, когда дав точкам разрыва двигаться вдоль характеристик, мы сразу же получим многозначное решение [11; 17–19] (разрыв будет опрокидываться подобно морской волне).

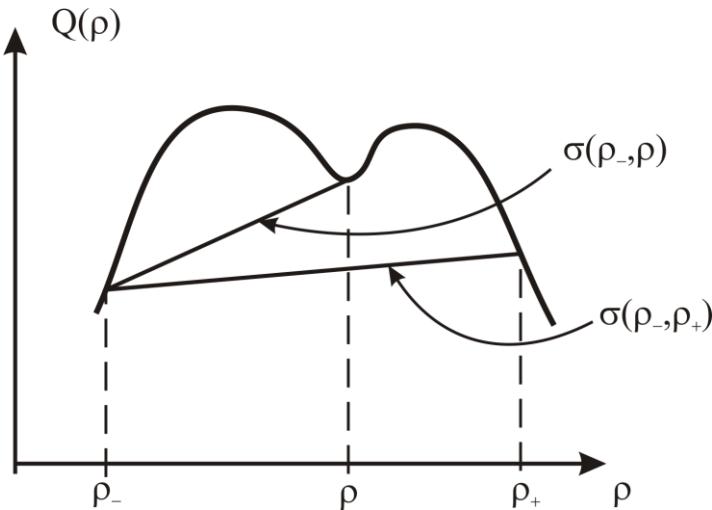


Рис. 5. Е-условие

Пример (см. [18]). Снова возьмем уравнение Хопфа и начальное условие

$$\rho(0, x) = \begin{cases} -1, & x \leq 0, \\ 1, & x > 0. \end{cases}$$

Возможны следующие слабые решения задачи Коши (4), (5):

$$\rho_1(t, x) = \begin{cases} -1, & x \leq -t, \\ \frac{x}{t}, & -t < x \leq t, \\ 1, & x > t, \end{cases} \quad \rho_2(t, x) = \begin{cases} -1, & x \leq 0, \\ 1, & x > 0. \end{cases}$$

«Размазывая» разрыв начальных данных в точке $x=0$, т.е. вводя $\rho^\delta(0, x)$ – монотонно возрастающую непрерывную функцию, совпадающую вне отрезка $|x| \leq \delta$ с $\rho(0, x)$, мы увидим (используя, например, классический метод характеристик (см. рис. 6), работоспособность которого в данной ситуации обеспечивается отсутствием пересечений у характеристик), что

$$\lim_{\delta \rightarrow 0^+} \rho^\delta(t, x) = \rho_1(t, x).$$

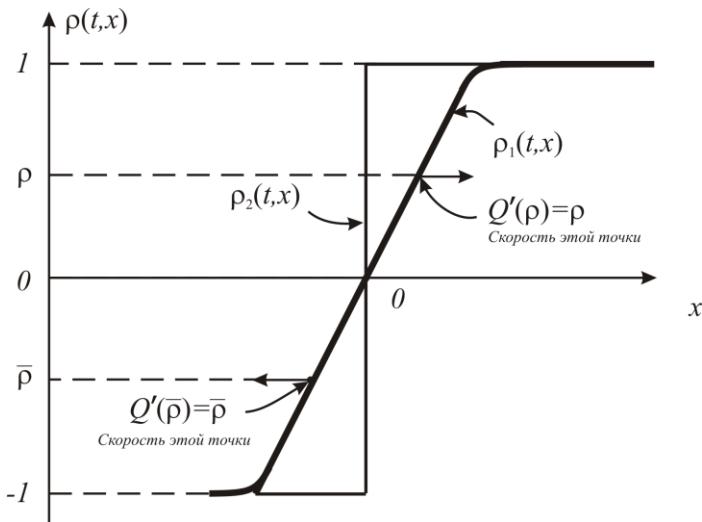


Рис. 6

Т.е. неклассическое решение $\rho_2(t, x)$ не является устойчивым решением. Несложно проверить, что на разрыве решения $\rho_2(t, x)$ не выполня-

ется Е-условие, поэтому $\rho_2(t, x)$ не является решением. Если посмотреть на поведение характеристик системы, то можно заметить, что для решения $\rho_2(t, x)$ разрыв надуман – он не вызван пересечением характеристик. Вместо того чтобы пересекаться на разрыве, характеристики «расходятся» от разрыва.

В заключение этого примера заметим, что уравнения семейства характеристик на плоскости $(t; x)$ имеют вид

$$\frac{dx}{dt} = Q'(\rho(t, x)).$$

Поэтому полная производная по времени от функции $\rho(t, x)$ вдоль характеристики есть

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \frac{\partial\rho}{\partial x} \frac{dx}{dt} = \frac{\partial\rho}{\partial t} + Q'(\rho) \frac{\partial\rho}{\partial x} = \frac{\partial\rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = 0,$$

т.е. $\rho(t, x) \equiv \text{const}$ вдоль характеристики. ■

Отметим, что классический метод характеристик для решения уравнений в частных производных первого порядка может использоваться лишь локально для уравнения (4), т.к. по прошествии некоторого времени характеристики могут начать пересекаться и возникнет неоднозначность: одной точке (t, x) будут соответствовать несколько, вообще говоря, разных значений ρ («принесенных» по характеристикам). Собственно, там, где характеристики начинают пересекаться, и возникает разрыв у решения уравнения (5) [19]. Метод характеристик вкупе с условиями на разрыве был одним из первых методов исследования задачи Коши (4), (5).

Заметим также, что процесс, описываемый разрывным решением (4), необратим во времени (см., например, [18, 19] и п. 2.3.2). Причем условие разрывности процесса существенно для необратимости. Так в примере О. А. Олейник $\rho_q(t, x)$ при $q=1$ является разрывным решением (4), (5), для которого выполняется Е-условие (как строгое неравенство). Если решать (4), (5) в попятном времени, то неравенство в Е-условии поменяется на противоположное, поэтому если функция $\rho_q(t, x)$ при $q=1$ удовлетворяла «прямому» Е-условию, то она точно не может удовлетворять «попятному» Е-условию. Заметим, однако, что уравнение (4) выглядит симметричным относительно обращения времени ($t \rightarrow -t$), поскольку при попятном течении времени величина потока

$Q(\rho)$ изменяет знак на противоположный. Однако, как уже отмечалось, уравнение (4), понимаемое в слабом смысле, определяет эволюцию системы, вообще говоря, не единственным образом. Выделение единственного решения является необратимой по времени процедурой.

Приведем в заключение еще один пример, показывающий отличие закона сохранения (4) от уравнений, описывающих только гладкие решения.

Пример (И. М. Гельфанд [17]). Снова возьмем уравнение Хопфа, и запишем его в двух дивергентных формах (задав $\rho(0, x) > 0$, мы можем быть уверены в том, что везде в дальнейшем $\rho(t, x) > 0$):

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial(\rho^2/2)}{\partial x} &= 0, \\ \frac{\partial(\rho^2/2)}{\partial t} + \frac{\partial(\rho^3/3)}{\partial x} &= 0. \end{aligned}$$

Второе уравнение получается умножением уравнения Хопфа на $\rho > 0$ и приведением полученного уравнения к дивергентному виду. Написав для каждого из этих уравнений условие типа RRH на разрыве:

$$\begin{aligned} C_1 &= \frac{\rho_+^2/2 - \rho_-^2/2}{\rho_+ - \rho_-} = \frac{1}{2}(\rho_+ + \rho_-) = \\ &= \frac{2}{3} \frac{\rho_+^2 + \rho_+ \rho_- + \rho_-^2}{\rho_+ + \rho_-} = \frac{\rho_+^3/3 - \rho_-^3/3}{\rho_+^2/2 - \rho_-^2/2} = C_2, \end{aligned}$$

легко убеждаемся в неэквивалентности этих дивергентных форм (если рассматривать обобщенные решения). Мы написали «условие типа RRH», потому что пользуемся обобщением этого условия:

$$c = \frac{q(\rho_+) - q(\rho_-)}{\eta(\rho_+) - \eta(\rho_-)}$$

на разрыве решения уравнения

$$\frac{\partial \eta(\rho)}{\partial t} + \frac{\partial q(\rho)}{\partial x} = 0,$$

с $\eta'(\rho) > 0$. ■

Замечание (см. [20]). Пусть

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

Тогда для любого k :

$$Q'(\rho_-) < k < Q'(\rho_+) \quad (Q'(\rho_-) > k > Q'(\rho_+))$$

существует такая положительная гладкая функция $\psi(\rho)$, что уравнение (4), умноженное на $\psi(\rho)$, будет иметь скорость разрыва k .

2.1.2. Модель Танака

Приведем один из способов определения зависимости $V(\rho)$, предложенный в 1963 г. Танака и др. [8, 21] (по-видимому, этот способ был известен значительно раньше).

Рассматривается однополосный поток АТС. Пусть скорость АТС не может превышать v_{\max} . Плотность

$$\rho(v) = \frac{1}{d(v)},$$

где

$$d(v) = L + c_1 v + c_2 v^2$$

– среднее (безопасное) расстояние между АТС при заданной скорости v движения потока,⁵ L – средняя длина АТС, c_1 – время, характеризующее реакцию водителей, c_2 – коэффициент пропорциональности тормозному пути (см. также п. 2.2.3). Из зависимости $d(v)$ можно получить зависимость (1) $V(\rho)$, удовлетворяющую условию (2).

Коэффициент c_2 , вообще говоря, зависит от дорожных условий. Так при нормальных условиях [21]:

$$d(v)[m] = 5,7[m] + 0,504[c] \cdot v[m/c] + 0,0285[c^2/m] \cdot v^2[m^2/c^2],$$

для мокрого асфальта [22]

$$d(v)[m] = 5,7[m] + 0,504[c] \cdot v[m/c] + 0,0570[c^2/m] \cdot v^2[m^2/c^2],$$

а для обледенелой дороги [22]

$$d(v)[m] = 5,7[m] + 0,504[c] \cdot v[m/c] + 0,1650[c^2/m] \cdot v^2[m^2/c^2].$$

⁵ $d(v)$ также называют *динамическим габаритом* или *дистанцией видимости*, и понимают под $d(v)$ часть полосы, содержащую АТС вместе с дистанцией экстренного торможения.

Моделью Танака называют LWR модель, в которой уравнение состояния $V(\rho)$ определяется так, как описано выше. Несмотря на свою простоту, модель Танака играет очень важную роль в современных исследованиях транспортных потоков [8].

2.1.3. Модель Уизема

Следующим шагом (упомянутым еще в 1955 г. и окончательно предложенным в 1974 г. Дж. Уиземом [7]) был учет «дальнозоркости» водителей:

$$v(t, x) = V(\rho(t, x)) - \frac{D(\rho(t, x))}{\rho(t, x)} \frac{\partial \rho(t, x)}{\partial x}, \quad D(\rho) > 0.$$

Откуда, с учётом закона сохранения количества АТС:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(v\rho)}{\partial x} = 0,$$

получим *уравнение типа Бюргерса* (закон сохранения с нелинейной дивергентной диффузией):

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \frac{\partial}{\partial x} \left(D(\rho) \frac{\partial \rho}{\partial x} \right). \quad (6)$$

Появившиеся в правых частях новые (по сравнению с (1) и (4)) диффузионные слагаемые соответствуют тому факту, что водители снижают скорость при увеличении плотности потока АТС впереди и увеличивают при уменьшении. Гидродинамическая (макроскопическая) модель (2), (5), (6) называется **моделью Уизема**.

В случае (Б. Гриншилдс (1934)), когда $Q(\rho)$ – парабола (вогнутая),

$$D(\rho) \equiv \varepsilon \quad (\varepsilon > 0)$$

можно показать, что уравнение (4) сводится с помощью линейной замены переменных и неизвестной функции: $t \rightarrow \tilde{t}$, $x \rightarrow \tilde{x}$, $\rho \rightarrow \tilde{\rho}$ к *уравнению (Бэйтмена–) Бюргерса* (играющему важную роль в гидродинамике (см., например, работы Х. Бэйтмена (1915), Ж. Лерэ (1934), Дж. Бюргерса (1940), Э. Хопфа (1950)):

$$\frac{\partial \tilde{\rho}}{\partial \tilde{t}} + \tilde{\rho} \frac{\partial \tilde{\rho}}{\partial \tilde{x}} = \varepsilon \frac{\partial^2 \tilde{\rho}}{\partial \tilde{x}^2}, \quad \varepsilon > 0.$$

С помощью замены (Форсайта–) Флорина–Хопфа–Коула [4, 7, 19] (см. также замечание в конце этого пункта):

$$\tilde{\rho} = -2\varepsilon \frac{\partial}{\partial \tilde{x}} (\ln w) = -2\varepsilon \frac{w_{\tilde{x}}}{w};$$

задача (4), (5) (для уравнения Бюргерса) сводится к задаче Коши для уравнения теплопроводности:⁶

$$\frac{\partial w}{\partial t} = \varepsilon \frac{\partial^2 w}{\partial \tilde{x}^2}, \quad w(0, \tilde{x}) = \exp\left(-\frac{1}{2\varepsilon} \int_0^{\tilde{x}} \tilde{\rho}(0, \xi) d\xi\right).$$

Используя этот факт, Э. Хопф в 1950 г. изучал поведение решения начальной задачи Коши для уравнения Бюргерса [7, 14]. Так, например, им был обоснован предельный переход (получивший название *метода исчезающей вязкости*⁷) при $\varepsilon \rightarrow 0+$ от уравнения Бюргерса к уравнению Хопфа:

$$\frac{\partial \tilde{\rho}}{\partial t} + \tilde{\rho} \frac{\partial \tilde{\rho}}{\partial \tilde{x}} = 0.$$

В связи с вышесказанным напомним, что для нелинейного закона сохранения (4) гладкое решение задачи Коши (4), (5) существует, как правило, только в малой окрестности линии, где заданы начальные условия. По разрывным начальным условиям решение задачи Коши для нелинейных уравнений, вообще говоря, не определяется однозначно даже в сколь угодно малой окрестности линии, где заданы начальные условия. Для того чтобы задача Коши для нелинейных уравнений с гладкими или разрывными начальными условиями была однозначно разрешима в большей области, необходимо рассматривать разрывные решения уравнения и по-новому ставить задачу Коши. Казалось бы, что достаточно (следуя идеям Н. М. Гюнтера, С. Л. Соболева, Л. Шварца в линейном случае) равенства (4), (5) понимать в слабом смысле (понимать (4) в смысле соотношения (3)). Однако (см. пример О. А. Олейник из п. 2.1.1) такое определение решения не обеспечивает его единствен-

⁶ Причина, по которой уравнение Бюргерса линеаризуется, объясняется в [23] и связана с тем, что уравнение Бюргерса достаточно симметрично (допускает бесконечномерную алгебру Ли (группу преобразований)). Интересно заметить, что есть определенная техника, позволяющая по заданному эволюционному уравнению определять, линеаризуется ли оно или нет. Для более подробного ознакомления с групповым анализом дифференциальных уравнений можно рекомендовать монографии [24] и [25]. Заметим также, что при определенных (специально подобранных) начальных условиях могут быть получены точные формулы для решений ряда важных в приложениях существенно нелинейных уравнений параболического (и не только) типа (см. работы К. А. Волосова и его учеников [26]).

⁷ Хотя в рассматриваемом нами случае речь идет скорее о диффузии, чем о вязкости.

ности. Корректный способ заключается в том, чтобы понимать решения задачи Коши (4), (5) $\rho(t, x)$ как предел (почти всюду по x при любом фиксированном значении $t > 0$) при⁸

$$\varepsilon \rightarrow 0+, D(\rho) := \varepsilon D(\rho), D(\rho) > 0$$

решений задач Коши (6), (5) $\rho_\varepsilon(t, x)$:

$$\|\rho(t, \cdot) - \rho_\varepsilon(t, \cdot)\|_{L_1(\mathbb{R})} = O(\sqrt{\varepsilon t}) \text{ (оценка Н. Н. Кузнецова (1975)).}$$

При этом $\rho(t, x)$ – ограниченная измеримая функция, не зависящая от $D(\rho_\varepsilon) > 0$, слабо удовлетворяющая закону сохранения (4) и начальному условию (5). Так определённую функцию $\rho(t, x)$ часто называют энтропийным решением задачи Коши (4), (5) (см. также замечание в конце п. 2.3.2).

Это представляется естественным. Ведь оба уравнения (4) и (6) возникли (на разных уровнях детализации) при описании одного явления. Обоснованием метода исчезающей вязкости интенсивно занимались в 50-е годы XX века (Э. Хопф, О. А. Олейник, А. Н. Тихонов и А. А. Самарский, П. Лакс, О. А. Ладыженская, И. М. Гельфанд и др.). Наиболее общие результаты получил С. Н. Кружков в конце 1960-х гг. (подробности см. в [11; 27–36], а также в конце п. 2.3.2).

Заметим, что уравнение (6) параболического типа, следовательно, решение можно понимать в обычном (классическом) смысле даже при разрывных начальных условиях. Для таких нелинейных уравнений развит достаточно эффективный аппарат. Прежде всего это различные варианты принципа максимума и основанные на них методы априорных оценок старших производных [37, 38] (С. Н. Бернштейн, И. Г. Петровский, О. А. Олейник, О. А. Ладыженская и др.), позволяющие довольно тонко исследовать различные свойства решений начально-краевых за-

⁸ Подчеркнем, что независимость предела от вида диффузионного слагаемого обусловлена тем, что диффузия входит дивергентным образом в правую часть уравнения (6). Если бы, скажем, уравнение (6) имело вид

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \varepsilon D(\rho) \frac{\partial^2 \rho}{\partial x^2},$$

то предел, вообще говоря, уже зависел бы от вида $D(\rho)$. В чем легко убедиться, рассмотрев пример И. М. Гельфанда (см. п. 3.1.1). Отметим тем не менее, что если в пределе при $D(\rho) \equiv 1$ получается функция без разрывов, то предел по-прежнему не будет зависеть от $D(\rho) > 0$.

дач для уравнений параболического типа. Напомним, в чем заключается принцип максимума: решение уравнения параболического типа достигает максимального и минимального значений на параболической границе (основание и боковые стороны) области, в которой рассматривается это решение (вместо уравнений можно рассматривать и неравенства). Так, например, решение начальной задачи Коши для уравнения параболического типа ограничено теми же постоянными, что и начальное условие. Установив различные свойства решения задачи Коши (6), (5) и осуществив предельный переход при $\varepsilon \rightarrow 0+$, можно получать разнообразные свойства решения задачи Коши (4), (5).

Посредством «хорошего» уравнения (6) попытаемся теперь установить связь между законом сохранения и *уравнением Гамильтона–Якоби* (Γ – \mathcal{Y}). Для этого рассмотрим две приводимые ниже задачи Коши:

$$\begin{cases} \frac{\partial \rho_\varepsilon}{\partial t} + \frac{\partial Q(\rho_\varepsilon)}{\partial x} = \varepsilon \frac{\partial^2 \rho_\varepsilon}{\partial x^2}, \\ \rho_\varepsilon(0, x) = \rho(0, x) \end{cases} \quad (7)$$

$$\begin{cases} \frac{\partial U_\varepsilon}{\partial t} + Q\left(\frac{\partial U_\varepsilon}{\partial x}\right) = \varepsilon \frac{\partial^2 U_\varepsilon}{\partial x^2}, \\ U_\varepsilon(0, x) = \int_0^x \rho(0, y) dy. \end{cases} \quad (8)$$

Как уже отмечалось, для следующей задачи Коши (в частности, для задачи Коши (7)):

$$\begin{aligned} \frac{\partial \rho_\varepsilon}{\partial t} + \frac{\partial Q(\rho_\varepsilon)}{\partial x} &= \varepsilon \frac{\partial}{\partial x} \left(D(\rho_\varepsilon) \frac{\partial \rho_\varepsilon}{\partial x} \right), \quad D(\rho_\varepsilon) > 0, \\ \rho_\varepsilon(0, x) &= \rho(0, x) \end{aligned}$$

существует почти всюду по x при любом фиксированном значении $t > 0$:

$$\lim_{\varepsilon \rightarrow 0+} \rho_\varepsilon(t, x) = \rho(t, x).$$

Используя схожую технику, можно показать, что

$$\lim_{\varepsilon \rightarrow 0+} U_\varepsilon(t, x) = U(t, x)$$

равномерно на любом компакте в полуплоскости $t > 0$, $x \in \mathbb{R}$. Причём $U(t, x)$ – ограниченная непрерывная функция, которая слабо удовлетворяет *уравнению Γ – \mathcal{Y}* :

$$\frac{\partial U}{\partial t} + Q\left(\frac{\partial U}{\partial x}\right) = 0 \quad (9)$$

и начальному условию

$$U(0, x) = U_0(x) = \int_0^x \rho(0, y) dy. \quad (10)$$

Так определенную функцию $U(t, x)$ называют *вязкостным решением* (Крэндалла–Лионса) задачи Коши (9), (10).⁹ Поскольку решения задач (7) и (8) классические, то имеет место следующая формула:

$$\rho_\varepsilon(t, x) = \frac{\partial U_\varepsilon(t, x)}{\partial x}.$$

Из того, что почти всюду по x при любом фиксированном значении $t > 0$:

$$\lim_{\varepsilon \rightarrow 0^+} \rho_\varepsilon(t, x) = \rho(t, x)$$

следует, что почти всюду по x при любом фиксированном значении $t > 0$:

$$\lim_{\varepsilon \rightarrow 0^+} \frac{\partial U_\varepsilon(t, x)}{\partial x} = \rho(t, x).$$

⁹ Рассмотрим следующую каноническую задачу (Лагранжа) оптимального управления Понтрягинского типа (см., например, [39, 40]):

$$\begin{aligned} J(t, x; u(\cdot)) &= \int_{t_0}^t L(t, x(\tau), u(\tau)) d\tau + \varphi(t_0, x_0), \\ \dot{x}(\tau) &= f(\tau, x(\tau), u(\tau)), \quad u(\tau) \in M, \quad x(t) = x. \end{aligned}$$

Положим «функцию цены» $U(t, x)$ равной

$$U(t, x) = \inf_{u(\cdot) \in M} J(t, x; u(\cdot)), \quad U(t_0, x) = \varphi(t_0, x).$$

Тогда с некоторыми оговорками справедлива следующая теорема (см., например, [41, 42]).

Теорема (принцип Беллмана). $U(t, x)$ является *вязкостным решением начальной задачи Коши для прямого уравнения Гамильтона–Якоби–Беллмана (Г–Я–Б)*:

$$\frac{\partial U}{\partial t} + \sup_{u \in M} \left\{ \left\langle \frac{\partial U}{\partial x}, f(t, x, u) \right\rangle - L(t, x, u) \right\} = \frac{\partial U}{\partial t} + Q\left(\frac{\partial U}{\partial x}\right) = 0, \quad U(t_0, x) = \varphi(t_0, x).$$

Приведенная основная теорема динамического программирования наряду с принципом максимума Понтрягина является базой теории оптимального управления. Причём из этой теоремы мы получаем управление в форме синтеза $u(t, x)$, а не в программном виде $u(\tau)$, как в принципе максимума, что более ценно для приложений. Отметим также, что динамическим программированием, как правило, пользуются лишь при небольших размерностях фазового пространства управляемой динамической системы.

В случае выпуклой (вогнутой) функции $Q(\rho)$ или $U_0(x)$ установлено, что почти всюду по x при любом фиксированном $t > 0$:

$$\lim_{\varepsilon \rightarrow 0+} \frac{\partial U_\varepsilon(t, x)}{\partial x} = \frac{\partial U(t, x)}{\partial x} = \rho(t, x). \quad (11)$$

Используя теорему о дифференцировании (по направлению) под знаком супремума¹⁰, можно проверить, что для вязкостного решения $U(t, x)$ справедливы следующие формулы, встречавшиеся в работах Э. Хопфа (1965), которые принято называть *формулами Хопфа* (–Лакса):¹¹

1) Пусть $U_0(x)$ – выпуклая функция, тогда

$$U(t, x) = \sup_{s \in \mathbb{R}} [sx - Q(s)t - U_0^*(s)],$$

где

$$U_0^*(s) = \sup_{x \in \mathbb{R}} [sx - U_0(x)];$$

2) Пусть $Q(\rho)$ – вогнутая функция, тогда

$$U(t, x) = \inf_{f \in \mathbb{R}} [U_0(x - tf) - Q^*(f)t],$$

где

$$Q^*(f) = \sup_{s \in \mathbb{R}} [Q(s) - sf].$$

Аналогичные формулы можно выписать и для выпуклой функции $Q(\rho)$ или вогнутой функции $U_0(x)$.

Проверим, например, формулу из 1). Для этого сразу заметим (ввиду выпуклости $U_0(x)$), что

¹⁰ См. [43, 44] (формулы Хопфа см. там же [44]). Необходимые для понимания факты выпуклого анализа имеются, например, в книгах [45, 46]. Теорему о дифференцировании (по направлению) под знаком супремума иногда называют *теоремой Демьянова–Данскина*, поскольку оба автора независимо пришли к аналогичному утверждению в конце 60-х годов XX века [43]. Однако эта теорема была известна и раньше. Так, в середине 60-х годов XX века А. Я. Дубовицкий и А. А. Милютин при получении принципа максимума для задач с фазовыми ограничениями доказали и фактически уже использовали теорему о дифференцировании (по направлению) под знаком супремума [39, 47].

¹¹ Аналогичные формулы возникали ранее, например, в работах О. А. Олейник и П. Лакса; приблизительно в то же время, что и в работах Э. Хопфа, эти формулы использовал С. Н. Кружков; близкие идеи “о сведении решения задачи Коши для обыкновенного нелинейного дифференциального уравнения или для нелинейного уравнения в частных производных к решению задачи-ач оптимизации” предлагались в 1965 г. (и ранее) Р. Беллманом и Р. Калабой [48]. Упомянем здесь также работу Б. Н. Пшеничного и М. И. Сагайдака [49].

$$U(0, x) = U_0^{**}(x) = U_0(x) \text{ (теорема Фенхеля–Моро [45]).}$$

По теореме о дифференцировании (по направлению) под знаком супремума

$$\partial U(t, x)/\partial t = -Q(s(t, x)), \quad \partial U(t, x)/\partial x = \rho(t, x),$$

где

$$s(t, x) = \arg \sup_{s \in \mathbb{R}} [sx - Q(s)t - U_0^*(s)]$$

— точка как функция параметров (t, x) , в которой достигается максимум по s выражения $sx - Q(s)t - U_0^*(s)$. Отсюда следует, что

$$\frac{\partial U(t, x)}{\partial t} + Q\left(\frac{\partial U(t, x)}{\partial x}\right) = -Q(s(t, x)) + Q(s(t, x)) = 0.$$

Для наглядности считаем, что супремум достигается в одной единственной точке (поэтому написано \arg , а не Arg). Можно показать, что супремум достигается не в одной точке тогда и только тогда, когда в этой точке функция

$$\rho(t, x) = s(t, x)$$

терпит разрыв. Собственно, формула

$$\rho(t, x) = s(t, x) = \operatorname{Arg} \sup_{s \in \mathbb{R}} [sx - Q(s)t - U_0^*(s)]$$

дает энтропийное решение задачи (4), (5) в случае неубывающей начальной функции $\rho(0, x)$, если под $\rho(t, x)$ договориться понимать многозначную функцию, которая принимает в точках разрыва всевозможные значения из отрезка, соответствующего скачку разрыва.

Используя далее теорему о дифференцировании под знаком супремума и соотношение (11), можно также получить и формулы для энтропийного решения задачи Коши (4), (5) $\rho(t, x)$ (заметим, что формула, полученная из пункта 2), называется *формулой Лакса–Олейник* [11, 42]). Упомянутые здесь формулы использовались, например, при исследованиях задач п. 2.3. В частности, С. Н. Кружков и Н. С. Петросян в 1982 г. получили доказательство утверждения, очень близкого к теореме 1 п. 2.3.1, исходя из формулы 1).

Замечание (идемпотентный принцип соответствия Литвиnova–Маслова [50–52]). Рассмотрим уравнение теплопроводности

$$\frac{\partial w_\varepsilon}{\partial t} = \varepsilon \frac{\partial^2 w_\varepsilon}{\partial x^2}$$

и уравнение Гамильтона–Якоби–Хопфа:

$$\frac{\partial U_\varepsilon}{\partial t} + \frac{1}{2} \left(\frac{\partial U_\varepsilon}{\partial x} \right)^2 = \varepsilon \frac{\partial^2 U_\varepsilon}{\partial x^2}.$$

Несложно проверить, что замена (схожую замену также делал Э. Шредингер)

$$U_\varepsilon = -2\varepsilon \ln w_\varepsilon$$

переводит одно уравнение в другое. Таким образом, кстати сказать, и была получена замена Флорина–Хопфа–Коула русским механиком В. А. Флориным в 1948 г. (на два года раньше Э. Хопфа и на три года раньше С. Коула). Используя при $\varepsilon \rightarrow 0+$ замену

$$w_\varepsilon \rightarrow -2\varepsilon \ln w_\varepsilon = U_\varepsilon, \text{ т.е.}$$

$$\lim_{\varepsilon \rightarrow 0+} w_\varepsilon \stackrel{\text{def}}{=} w \rightarrow \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln w_\varepsilon \stackrel{\text{def}}{=} U,$$

В. П. Маслов и В. П. Белавкин [50] в конце 80-х годов XX века предложили ввести *идемпотентное* (также говорят, *тропическое*) полуполе (от обычного поля полуполе отличается тем, что в нем отсутствует операция, обратная сложению (вычитание)) с операциями сложения и умножения \oplus, \odot , которые определяются следующим образом:

$$\begin{aligned} w^1 &\rightarrow U^1, \quad w^2 \rightarrow U^2; \\ w^1 + w^2 &\rightarrow \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^1/(2\varepsilon)} + e^{-U^2/(2\varepsilon)} \right) = \\ &= \min \left\{ \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^1/(2\varepsilon)} \right), \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^2/(2\varepsilon)} \right) \right\} = \\ &= \min \left\{ U^1, U^2 \right\} \stackrel{\text{def}}{=} U^1 \oplus U^2; \\ w^1 \cdot w^2 &\rightarrow \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^1/(2\varepsilon)} \cdot e^{-U^2/(2\varepsilon)} \right) = \\ &= \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^1/(2\varepsilon)} \right) + \lim_{\varepsilon \rightarrow 0+} -2\varepsilon \ln \left(e^{-U^2/(2\varepsilon)} \right) = \\ &= U^1 + U^2 \stackrel{\text{def}}{=} U^1 \odot U^2. \end{aligned}$$

Далее строился функциональный анализ над идемпотентным полуполем подобно тому, как строился обычный функциональный анализ над полем действительных или комплексных чисел. Построение такого анализа выявляет много связей (иногда называемых *принципами соответствия* (Литвинова–Маслова)) между классическими понятиями. Скажем, идемпотентным аналогом преобразования Фурье будет преобразование Юнга–Фенхеля–Лежандра [45], а вариационные принципы

механики – это идемпотентный вариант фейнмановского подхода к квантовой механике через интегралы по траекториям. Кстати сказать, все это тесно связано с формулами типа Лакса–Олейник. Однако здесь мы хотим обратить внимание прежде всего на то, что для некоторых нелинейных уравнений (Γ –Я, Γ –Я–Б) справедлив принцип суперпозиции (В. П. Маслова), правда, не над обычными полями действительных или комплексных чисел, а над идемпотентным полуполем: если U^1, U^2 – решения, то для любых действительных чисел λ^1, λ^2

$$U = \lambda^1 \odot U^1 \oplus \lambda^2 \odot U^2$$

также является решением.

В заключение этого пункта, в котором мы пояснили, что понимается под решением задачи Коши (4), (5), приведем, следуя монографии [33], оценку устойчивости решения

$$\rho(t, x; \rho(0, x), Q(\rho))$$

задачи Коши для закона сохранения (4) с начальной функцией $\rho(0, x)$ и с функцией потока $Q(\rho)$ по $\rho(0, \cdot)$ и $Q(\cdot)$:

$$\begin{aligned} & \left\| \rho(t, \cdot; \rho_1(0, x), Q_1(\rho)) - \rho(t, \cdot; \rho_2(0, x), Q_2(\rho)) \right\|_{L_1(\mathbb{R})} \leq \\ & \leq \left\| \rho_1(0, \cdot) - \rho_2(0, \cdot) \right\|_{L_1(\mathbb{R})} + \\ & + t \min \left\{ \text{T.V.}(\rho_1(0, \cdot)), \text{T.V.}(\rho_2(0, \cdot)) \right\} \|Q_1(\cdot) - Q_2(\cdot)\|_{Lip}, \end{aligned}$$

где T.V. – total variation (полная вариация), а

$$\|Q_1(\cdot) - Q_2(\cdot)\|_{Lip} \stackrel{\text{def}}{=} \sup_{\tilde{\rho} \neq \rho} \left| \frac{Q_1(\tilde{\rho}) - Q_2(\rho)}{\tilde{\rho} - \rho} \right|.$$

Отметим, что приводимая выше оценка при

$$Q_1(\rho) = Q_2(\rho)$$

следует из аналогичной оценки устойчивости по начальным данным решения задачи Коши для уравнения (6), которая в свою очередь является следствием принципа максимума для параболических уравнений. Заметим, что эта оценка (для уравнения (6)) обеспечивает единственность (с точностью до почти всюду по x при любом значении $t \geq 0$) энтропийного решения задачи Коши (4), (5). Обратим также внимание на равномерность по времени оценки устойчивости по начальным данным (для уравнений (4) и (6)).

2.1.4. Модель Пэйна и её обобщения¹²

Следующим важным шагом стала **модель Пэйна** (1971) [7, 53] (модель Пэйна–Уизема (см. п. 2.1.3)). Эту модель можно понимать как закон сохранения

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0,$$

в котором уже не предполагается зависимость скорости от плотности (уже не предполагается, что желаемая скорость устанавливается мгновенно). А для скорости выписывается уравнение¹³

$$\frac{d}{dt} v = \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = -\frac{1}{\tau} \left(v - \left(V(\rho) - \frac{D(\rho)}{\rho} \frac{\partial \rho}{\partial x} \right) \right)$$

стремления реальной скорости v к желаемой скорости

$$V(\rho) - \frac{D(\rho)}{\rho} \frac{\partial \rho}{\partial x},$$

причём τ ($\tau \sim 1$ с) характеризует скорость стремления (в электротехнической терминологии τ – время релаксации; если же уподоблять транспортный поток сжимаемой неньютоновской (максвелловской) жидкости, то параметр τ характеризует максвелловское затухание [54]). Полученную систему уравнений запишем в виде

$$\frac{\partial}{\partial t} \begin{pmatrix} \rho \\ v \end{pmatrix} + \begin{pmatrix} v & \rho \\ D/(\tau\rho) & v \end{pmatrix} \cdot \frac{\partial}{\partial x} \begin{pmatrix} \rho \\ v \end{pmatrix} = \frac{1}{\tau} \begin{pmatrix} 0 \\ V - v \end{pmatrix}, \quad (12)$$

из которого легко следует строгая гиперболичность этой системы (матрица при $\partial/\partial x$ имеет различные вещественные собственные значения).

Интересно заметить следующий, достаточно общий, факт: основное отличие гидродинамических моделей транспортных потоков от соответствующих гидродинамических аналогов заключается в правых частях, возникающих, как правило, гиперболических (строго) систем уравнений и их диффузионных аналогов. Действительно, первое уравнение системы Пэйна есть просто «закон сохранения массы» (в дивергентной форме), а второе уравнение – «закон сохранения (изменения) импульса», приведенное к дивергентной форме, имеет вид

¹² Этот пункт написан совместно с Я. А. Холдовым.

¹³ В литературе принято называть моделью Пэйна частный случай описанной модели:

$$D(\rho) = \tau c_0^2 > 0.$$

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho v^2 + P(\rho))}{\partial x} = -\frac{1}{\tau}(\rho v - \rho V(\rho)),$$

где «давление»

$$P(\rho) = \frac{1}{\tau} \int_0^\rho D(\tilde{\rho}) d\tilde{\rho}.$$

Отмеченное обстоятельство представляется естественным. Ведь «транспортная жидкость» – это жидкость с мотивацией (стремление двигаться с желаемой скоростью), которая и присутствует в правой части. Это замечание позволяет использовать в расчетах по гидродинамическим моделям транспортных потоков хорошо разработанные более чем за полвека вычислительные алгоритмы (например, схемы П. Лакса, С. К. Годунова, сеточно-характеристический метод (Магомедова–Холодова) и др.), см., например, [4, 11, 18, 33; 54–56] и цитированную там литературу.¹⁴

¹⁴ Хорошим тестом на устойчивость выбранной разностной схемы является разложение конечных разностей по пространственной переменной в ряд Тейлора до второго порядка включительно и исследование матрицы при вторых производных (*аппроксимационной вязкости*) на положительную определенность [54, 55].

Напомним также вкратце (нам это понадобится в п. 3.2.4), следуя книге П. Лакса [11] (см. также [54, 55]), в чём заключается метод численного решения начально-краевой задачи для закона сохранения, предложенный С. К. Годуновым в конце 50-х 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 (см. п. 3.3.1). «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 $t = \delta/(2c_{\max})$ – шаг по времени (такой выбор шага иногда называют *условием Лакса*, причем автоматически выполняется необходимое условие Куранта–Фридрихса–Леви [2] $\tau/\delta \leq (c_{\max})^{-1}$) сходимости разностных схем при численном решении гиперболических уравнений), 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 и систем типа Пэнна.

Заметим также [19, 29], что уже для системы двух законов сохранения (система Х. Пэйна (12) как раз представляет пример такой системы, причём имеется ещё и нелинейная правая часть) в общем случае неизвестно, как корректно определять глобальное по времени обобщенное решение. Метод исчезающей вязкости для систем оказывается уже чувствительным к выбору положительно определённой матрицы $D(\rho)$ в правой части (проблема неединственности решения) [18]. Тем не менее для строго гиперболической системы законов сохранения с одной пространственной переменной за последние 15 лет был достигнут определённый прогресс (см. [11; 29–36]): в общем случае построена глобальная теория существования, единственности и устойчивости по начальным данным.¹⁵ Отметим, что теория была построена разными способами, в том числе и с помощью метода исчезающей вязкости (так построенное обобщенное решение часто называют *энтропийным*):

$$\varepsilon \rightarrow 0+, D(\rho) := \varepsilon I, \text{ где } I = \text{diag}\{1, \dots, 1\}$$

Тем не менее строгое доказательство устойчивости схемы Годунова имеется, насколько нам известно, лишь для конкретных систем. Но (поскольку всего одна пространственная переменная) если схема Годунова сходится, то непременно к энтропийному решению [34] (см. следующий абзац основного текста). Отметим сильную «качественную» связь (которую, впрочем, можно обосновать и теоретически) между описанной схемой Годунова, схемами бегущего счета, схемой потенциального сглаживания [18] и сеточно-характеристическим методом [57]. Отметим также в чём-то схожий «front tracking» метод [33, 34], в котором не решение аппроксимируется (кусочно-постоянной функцией), а вектор-функция потока (в скалярном случае $Q(\rho)$) аппроксимируется кусочно-линейной функцией. С помощью этого метода (так же как в свое время с помощью метода Годунова) недавно были получены продвижения в вопросах корректности начальной задачи Коши для системы законов сохранения [33, 34]. В заключение заметим, что схема Годунова для LWR модели может быть содержательно проинтерпретирована (см. п. 3.2.4). 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* (см. п. 3.2)). 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.

¹⁵ Особо отметим в этой связи работу Д. Глимма (1965) [11, 34], предложившего стохастическую модификацию метода Годунова ($\rho_k = \rho(0, k\delta + \xi\rho)$, где $\xi \in R[0, 1]$ – равномерно распределенная на отрезке $[0, 1]$ случайная величина), с помощью которой была установлена теорема существования для начальных данных, близких к константе (имеющих малую полную вариацию). Отметим также в этой связи, что в методе Годунова «закон(ы) сохранения» выполняется(-ются) точно (и эта «консервативность» очень важна, как отмечал в 2004 г. во время доклада в МИАН РАН С. К. Годунов, иначе довольно быстро могут накапливаться ошибки), а в методе Глимма точно «в среднем». Выступление С. К. Годунова можно посмотреть здесь http://www.mathnet.ru/php/presentation.phtml?option_lang=rus.

– единичная матрица (с единицами на диагонали и нулями на других позициях).

Желание «размазать разрывы решений» привело от модели LWR к модели Уизема. Это же желание мотивировало и введение в модель Пэйна диффузионные поправки. Заметим, что, вводя в правую часть системы диффузионные (дисперсионные) поправки, мы, как правило, решаем вопрос о том, что мы понимаем под решением [54]. Иначе говоря, для корректной (и адекватной «физике процесса») постановки начальной (начально-краевой) задачи Коши для системы законов сохранения важно знать, как эта система «была приготовлена». Откуда и как она возникла: что огрубили, чем пренебрегли и т.д. На таком пути получаются новые модели: Р. Кюна (1993), Кернера–Конхойзера (1994) (см. п. 2.4) и др. В [58] приведен достаточно подробный обзор работ (более 100 моделей). Здесь мы также упомянем некоторые подходы российских ученых (в которых обобщается подход Х. Пэйна): Н. Н. Смирнова, А. Б. Киселёва и др. (МГУ) [59–61]; А. С. Холодов и др. (МФТИ) [62].

Несколько недостатков модели Пэйна (и многих впоследствии предложенных моделей, в том числе с диффузионными поправками) были указаны К. Даганзо (Даганцо) (1995) [56, 58, 63] (см. также критику Б. С. Кернера в [10], отчасти приведенную в п. 2.4 и главе 3). В частности, было показано, что при сильных пространственных неоднородностях начальных условий могут возникать отрицательные значения скоростей (затор «рассасывается назад» как результат действия вязкости). При определенных значениях параметров могут возникать плотности, превышающие максимально допустимые («бампер к бамперу»). Также, согласно этим моделям, на движение АТС заметное влияние оказывают АТС, находящиеся сзади, что в случае одной полосы вряд ли возможно в реальном транспортном потоке. В начале XXI века А. Эу и М. Раскль [64], Дж. М. Гринберг [65], Х. М. Чзан [66] показали, как можно устранить недостатки, отмеченные К. Даганзо. Основная идея заключается в изменении второго уравнения в системе Пэйна:

$$\frac{d}{dt}(v + p(\rho)) = \frac{\partial(v + p(\rho))}{\partial t} + v \frac{\partial(v + p(\rho))}{\partial x} = 0.$$

При этом требуется, чтобы $p'(\rho) > 0$. В частности, для «давления» $p(\rho)$ были предложены следующие формулы:

$p(\rho) = \rho^\gamma$, $\gamma > 0$	А. Эу и М. Раскль (2000)
$p(\rho) = -V(\rho)$	Дж. М. Гринберг (2001), Х. М. Чзан (2002)

В конце статьи [64] указано, что можно оставить релаксационное слагаемое в правой части второго уравнения системы типа Пэйна:

$$\frac{d}{dt}(v + p(\rho)) = -\frac{1}{\tau}(v - V(\rho)).$$

При этом все положительные приобретения сохраняются, но добавляются и новые.

За последние десять лет появилось большое количество статей (А. Эу, А. Клар, П. Гоатэн, Р. Коломбо, М. Гаравелло, Б. Пикколи, Ф. Сиебель и В. Маузер, Д. Хельбинг (Хельбин) и др., см. сайт <http://arxiv.org/>), в которых модель Эу–Раскля обобщалась в различных направлениях. Например, в [67], для того, чтобы объяснить экспериментально обнаруженные Б. С. Кернером (1996) [10] три фазы транспортного потока (см. также главу 3): «газ, жидкость (свободное движение) – жидкость, замерзающая жидкость (синхронизированный режим движения) – замерзающая жидкость, лед (широко движущиеся кластеры)». Эти исследования, по-видимому, мотивированы желанием так обобщить (в классе систем из двух уравнений гиперболического типа) гидродинамическую модель Эу–Раскля, чтобы предложенная модель объясняла все основные наблюдаемые свойства транспортного потока.

Естественно теперь задаться вопросом: «переходят» ли модели типа Пэйна при $\tau \rightarrow 0+$ (предел нулевой релаксации) в модель Уизема или модель LWR? Интуиция подсказывает, что должны переходить (так же, как и в п. 2.1.3, можно ссыльаться на то, что описывается одно и то же явление на разных уровнях детализации). Однако в общем случае, насколько нам известно, нет строгого обоснования такого перехода. Ссылки на различные успешно исследованные случаи имеются, например, в работах [30, 64].

В заключение этого пункта упомянем модель третьего порядка Хельбинга–Эйлера–Навье–Стокса (1995) [58, 68], в которой к системе уравнений Пэйна добавляется третье уравнение («закон сохранения энергии») для вариации (дисперсии) скорости θ (характеризующей «разброс скоростей» относительно среднего значения). При этом во второе уравнение (которое понимается как уравнение для среднего значения скорости) вводится дополнительное слагаемое, зависящее от θ . Заметим при этом, что для системы уравнений Навье–Стокса, описывающей движение вязкой ньютоновской жидкости, не известно в общем случае, как поставить начальную (начально–краевую) задачу Коши, чтобы глобальное решение (определенное при всех значениях времени) было единственным. За решение этой проблемы математический институт Клея в 2000 г. назначил премию в один миллион долларов. Естеств-

венно считать, что если уравнение описывает реальный процесс, то оно обязано решаться (притом единственным образом). Однако выбранные уравнения – это лишь некоторая модель описываемого явления (возможно не всегда вполне адекватная), и «механизм», выбирающий единственное решение, мог быть «огрублен» при выводе уравнений. Особенno, если описываемый процесс чувствителен к малейшим возмущениям, флуктуациям. Интересные взгляды на проблему имеются в статьях О. А. Ладыженской [69] и В. И. Юдовича [70]. Так, в статье [69] проблема единственности переформулирована следующим образом: «Дают ли уравнения Навье–Стокса с начальными и краевыми условиями детерминистское описание динамики несжимаемой жидкости или не дают?» Сложности, возникающие при описании транспортного потока, во многом схожи со сложностями, возникающими при описании турбулентного движения жидкостей (см. статью А. Майда в сборнике [30], а также [71]).

2.1.5. Кинетические модели

Продолжая аналогию с газовой динамикой, будущий нобелевский лауреат по химии И. Пригожин (при участии Ф. Эндрюса и Р. Хермана) в 1960 г. предложил описывать транспортный поток кинетическим уравнением (типа Больцмана с «интегралом взаимодействия АТС» вместо «интеграла столкновения частиц газа») [56, 58, 72]. Подход И. Пригожина был впоследствии развит в работах С. Павери-Фонтана (1975), Д. Хельбинга (1995) и др. [56, 58].

Из кинетических моделей транспортного потока (в основном многополосного) можно получать макроскопические (гидродинамические) модели подобно тому, как в кинетической теории получаются уравнения газовой динамики (гидродинамики), т.е. с помощью умножения на различные функции от скорости и последующего интегрирования по скоростям кинетического уравнения для плотности в расширенном (на скорости) фазовом пространстве $(t; x; v)$. При этом, вообще говоря, будет получаться цепочка зацепляющихся уравнений.¹⁶ Так, если

¹⁶ Заметим, что современное понимание классической неравновесной статистической механики основывается на во многом схожей теории цепочек уравнений Боголюбова–Борна–Грина–Киркуда–Ивона (ББГКИ) [73]. Впрочем, в последнее время появился новый интересный подход (см. работы В. В. Козлова с коллегами [74]), восходящий к работам А. Пуанкаре и Дж. Гиббса. Много интересных и разнообразных связей собрано в записях курса лекций В. В. Веденяпина, прочитанного несколько лет назад студентам МФТИ и МГОУ и посвященного кинетической теории [75].

умножить кинетическое уравнение на единицу и проинтегрировать, получим уравнение для плотности («закон сохранения массы»), в которое будет входить средняя скорость. Если умножить кинетическое уравнение на скорость и проинтегрировать, получим уравнение для средней скорости («закон сохранения (изменения) импульса»), в которое будет входить вариация скорости θ (по сути, определяющаяся средним значением квадрата скорости). Если умножить кинетическое уравнение на квадрат скорости и проинтегрировать, получим уравнение для среднего значения квадрата скорости (откуда можно получить уравнение для вариации скорости), в которое будет входить среднее значение куба скорости, и т.д. Приходится в какой-то момент обрывать (замыкать) цепочку, привлекая, как правило, дополнительные «физические» соображения (гипотезы). Например, постулировать (на основе экспериментов или другим способом) для замыкания моментной цепочки некоторые соотношения (так называемые определяющие уравнения) между величинами, входящими в эти уравнения. Так, для газа в зависимости от этих соотношений получается модель идеального газа или модель Навье–Стокса–Фурье (вязкий теплопроводный газ) [36].

В связи с вышесказанным уместно заметить, что классической задачей статистической физики, восходящей к работам Максвелла [36], [75], является исследование перехода от уравнения Больцмана к уравнениям газодинамики (гидродинамики). Центральным местом здесь является проблема замыкания моментной цепочки для решения уравнения Больцмана. Однако не менее важным является изучение перехода от стохастической марковской динамики (например, транспортных потоков), лежащей в основе движения (см. п. 2.2.4), к кинетической динамике. При этом стохастическая марковская динамика (заданная, как правило, линейной полугруппой) порождает за счет скейлинга или перехода к «динамике средних» нелинейные кинетические уравнения (например, типа Больцмана–Пригожина), которые в свою очередь порождают нелинейные гидродинамические уравнения. Важно заметить, что без понимания этих «переходов» невозможно, на наш взгляд, правильно объяснить экспериментальные данные: три фазы транспортного потока.

В заключение отметим, что имеются также модели, промежуточные между кинетическими и гидродинамическими моделями, так называемые *мезоскопические*. Такой моделью двухполосного движения пользуется, например, коллектив, возглавляемый Б. Н. Четверушкиным (см. [76, 77]).

2.1.6. Практические приложения моделей

Несмотря на элементарность, модель LWR (а также ее дифференциально-разностные и разностные аналоги) достаточно популярна в прикладных расчетах. 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 графике транспортной сети. Основные сложности при этом возникают при постановке краевых условий в узлах графа транспортной сети (см., например, [78, 79]). Модель LWR (точнее ее разностные аналоги) хорошо подходит и для управления транспортными потоками. В подтверждение этих слов приведем некоторые идеи, использованные, например, в подходе Берклиевской группы (А. Б. Куржанский, А. А. Куржанский, П. Варайя, Р. Хоровитц и др. [80, 81]) к управлению дорожным движением.

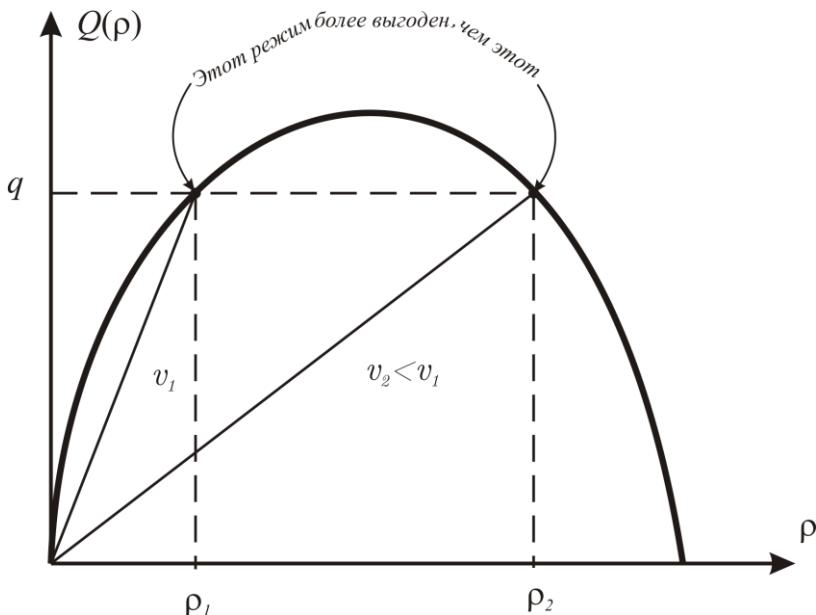


Рис. 7

Из фундаментальной диаграммы следует, что одному и тому же значению потока АТС соответствуют разные (как правило, две) плотности и, как следствие, разные скорости.¹⁷ Очевидно, что более выгодным режимом является режим с большей скоростью (см. рис. 7): 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).

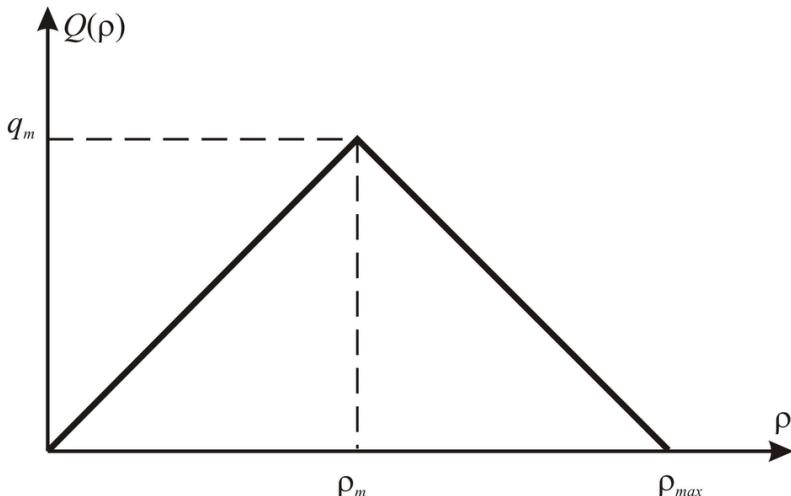


Рис. 8. Треугольная фундаментальная диаграмма

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 [82, 83] (CTM – Cell Transmission Model, см. п. 2.2.4) К. Даганзо = схема Годунова (1959) для (4) + треугольная фундаментальная диаграмма¹⁸ (см. рис. 8), был предложен

¹⁷ Это обстоятельство является также причиной сложностей, возникающих при постановке краевых условий в узлах (вершинах) графа транспортной сети [34, 78, 79]. Знание характеристик источников и стоков и матриц перемешивания в узлах (матриц, характеризующих расщепление потоков в узлах) недостаточно для корректной постановки начально-краевой задачи.

¹⁸ Также часто используется трапециoidalная фундаментальная диаграмма. Например, диаграмму на рис. 2 более естественно аппроксимировать именно трапециoidalной диаграммой, а не треугольной.

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 [80, 81].

Несмотря на то, что с момента появления первых фундаментальных работ прошло более полувека, по мнению ряда ведущих специалистов в области математического моделирования дорожного движения (К. Нагель, Х. Махмасани, М. Шрекенберг и др.), проблема образования предзаторных и заторных ситуаций еще до конца не изучена. Используя терминологию, предложенную Б. С. Кернером [10] (см. также главу 3), можно сказать, что на текущий момент нет общепринятого подхода, описывающего поведение АТС в области «синхронизированного потока». Иначе говоря, если поток АТС уподобляется жидкости, то наиболее сложная для моделирования ситуация – это «замерзающая жидкость». Подтверждением вышесказанному может служить тот факт, что разные коллективы, занимающиеся моделированием транспортных потоков, как правило, используют разные модели: начиная от модели LWR (М. Гаравелло и Б. Пикколи [34, 78]; А. А. Куржанский и др. [80, 81]), заканчивая моделями, в которых каждый водитель описывается своим вариационным принципом (И. А. Лубашевский и др. [86, 87]). Отметим также, 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 [63]) часто являются «узкие места» (перекрестки, въезды). Поэтому особенно важно (для приложений) создать целостную модель транспортных потоков, адекватную имеющимся данным, включающую описание источников, стиков АТС и поведение АТС в вершинах графа транспортной сети (перекрестки, въезды, выезды и т.п.).

2.2. МИКРОСКОПИЧЕСКИЕ МОДЕЛИ

В пункте 2.2 будет рассказано о некоторых подходах к микроскопическому моделированию движения в основном однополосных транспортных потоков. В основе подходов лежит концепция «о желании придерживаться при движении безопасной дистанции до лидера». Также будет рассказано о связях, имеющихся между микроскопическими и макроскопическими моделями. Прежде всего будут описаны модели *оптимальной скорости* и *следования за лидером*. Так же будет описана одна из наиболее популярных моделей (за последние десять лет) – модель *разумного водителя Трайбера*. В заключение этого пункта будут приведены модели *клеточных автоматов* (которые часто являются, по сути, разностными аналогами определенных макроскопических моделей), в том числе востребованные в приложениях.

2.2.1. Модель оптимальной скорости Ньюэлла

Пусть АТС в однополосном потоке пронумерованы слева направо. Обозначим через $s_n(t)$ – координату центра n -го АТС в момент времени $t \geq 0$. Положим

$$h_n(t) = s_{n+1}(t) - s_n(t), \quad v_n(t) = s'_n(t).$$

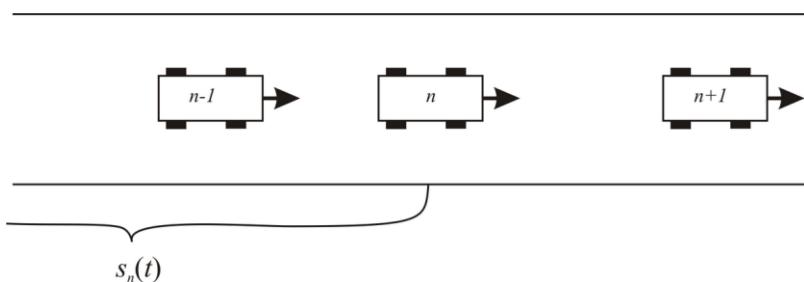


Рис. 9. Микроскопическая модель

В микроскопической *модели Ньюэлла* (этот модель была предложена в 1961 г. и является одной из первых нелинейных моделей *оптимальной*

скорости [7, 88]) постулируется, что (для каждого водителя существует «безопасная» скорость движения, зависящая от дистанции до лидера):

$$v_n(t+\tau) = V(1/h_n(t)),$$

где τ – время, характеризующее реакцию водителей, и

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

Заметим, что по зависимости интенсивности потока

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

от плотности ρ в окрестности ρ_{\max} (максимально возможное значение плотности также часто обозначается как ρ_j (см., например, [7])) можно определить τ , если известна средняя длина АТС L [7] ($L \sim 5,7$ м):

$$\tau = -L/Q'(\rho_{\max}) \quad (\tau \sim 0,4 \text{ с для рис. 2}).$$

Действительно, путь $V(\rho)\tau$, пройденный АТС за время τ , 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 ρ_{\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 (из рис. 2 следует, что $d \sim 6,5$ м). Приведенные в этом абзаце формулы активно используются при исследовании роста затора [10] (см. также п. 2.4). Отметим также, что если известна средняя длина АТС, время, характеризующее реакцию водителей, и желаемая скорость свободного потока (определяет наклон левой ветви фундаментальной диаграммы), то треугольная фундаментальная диаграмма однозначно строится (рис. 8).

Вернемся к модели. Введём функции двух переменных

$$h(t, x), \rho(t, x) = 1/h(t, x), v(t, x),$$

задав их значения в полуплоскости $t \geq 0, x \in \mathbb{R}$, на счётом наборе кривых согласно формулам

$$h\left(t, \frac{s_n(t) + s_{n+1}(t)}{2}\right) = h_n(t), \quad v(t, s_n(t)) = v_n(t).$$

Считая $h_n(t)$ и τ малыми величинами и учитывая, что

$$v(t + \tau, s_n(t + \tau)) = V(\rho(t, s_n(t)) + 1/2 h_n(t)),$$

$$\frac{d}{dt} h(t, s_n(t) + 1/2 h_n(t)) = v(t, s_{n+1}(t)) - v(t, s_n(t)),$$

получим

$$\begin{aligned} & v(t, s_n(t)) + (v_t(t, s_n(t)) + v(t, s_n(t)) v_x(t, s_n(t))) \tau = \\ & = V(\rho(t, s_n(t))) + V'(\rho(t, s_n(t))) \rho_x(t, s_n(t)) (1/2) h(t, s_n(t)), \\ & h_t(t, s_n(t)) + v(t, s_n(t)) h_x(t, s_n(t)) = v_x(t, s_n(t)) h(t, s_n(t)). \end{aligned}$$

Умножая второе уравнение на $-\rho^2$ и опуская у функций аргументы (продолжая «по непрерывности» $\rho(t, x)$ и $v(t, x)$ со счетного набора близких кривых на полуплоскость $t \geq 0$), придем к системе

$$v + (v_t + v v_x) \tau = V(\rho) + \frac{V'(\rho)}{2\rho} \rho_x,$$

$$\rho_t + (v \rho)_x = 0.$$

Таким образом, мы «вывели», следуя [7], модель Пэйна, получив новую интерпретацию для τ и

$$D(\rho) = -V'(\rho)/2.$$

Если бы мы с самого начала полагали $\tau = 0$, то в результате выполнения указанных выше операций пришли бы к модели Уизема. А если бы еще пренебрегли слагаемым $(1/2)V'(\rho)\rho_x h$ в сравнении с $V(\rho)$ (напомним, что мы считаем h – малым), то получили бы модель LWR.

Использованный выше прием называют *автомодельной редукцией*. Хорошим примером автомодельной редукции является вывод Д’Аламбером (1780) волнового уравнения исходя из модели колебания струны (с закрепленными концами), состоящей из множества одинаковых шариков (известной массы), соединенных одинаковыми пружинками (известной длины и жесткости), имеющими нулевую

массу. Поведение каждого шарика описывается вторым законом Ньютона (для отклонения шарика от положения равновесия u (см. рис. 10)) и зависит (посредством закона Гука) только от положений соседних шариков.

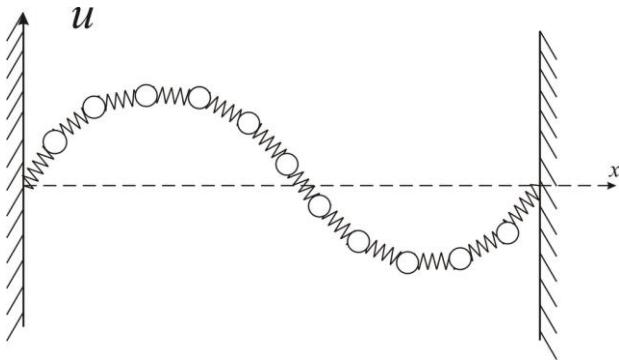


Рис. 10. Колебания струны, Д'Аламбер (1780)

Таким образом, получается система обыкновенных дифференциальных уравнений, из которой с помощью предельного перехода (шарики выбираются все меньше и меньше, пружинки, соединяющие шарики, становятся все меньше и меньше) получается одно уравнение в частных производных – волновое уравнение, описывающее колебание струны. Другой, более современный (середина 50-х годов XX века), пример автомодельной редукции: задача Ферми–Паста–Улама (вывод уравнения Кортевега–де Фриза (1895)) – см. [89] и цитированную там литературу. Следует заметить, что приём автомодельной редукции (достаточно популярный в математической физике) является эвристическим (для приведенных выше выкладок это особенно очевидно, если вспомнить, что, например, плотность может терпеть разрывы – и ни о каких оценках малости отбрасываемых членов ряда Тейлора не может идти и речи). В данной задаче при $\tau = 0$ частичная обоснованность (сходство, причём в некотором смысле равномерная по времени, в поведении решений (начальных задач), полученных по исходной модели Ньюэлла, и по моделям LWR и Уизема) автомодельной редукции следует из результатов [18, 33; 90–111] (см. также п. 2.3.1).

Пусть поток АТС однороден и стационарен:

$$\rho(t, x) \equiv \rho, v(t, x) \equiv V(\rho).$$

Приведем, следуя [7], условие устойчивости в линейном приближении этого режима движения, считая, что транспортный поток описывается 1) моделью Ньюэлла; 2) моделью Пэйна¹⁹ с

$$D(\rho) = -V'(\rho)/2.$$

В обоих случаях ответ одинаков:

$$2\rho^2 |V'(\rho)|\tau < 1.$$

¹⁹ Рассмотрим систему Пэйна. Будем искать решение в виде

$$\rho(t, x) = \rho + r(t, x), \quad v(t, x) = V(\rho) + w(t, x),$$

считая $r(t, x)$, $w(t, x)$ вместе с частными производными, входящими в систему, малыми. Линеаризуем систему Пэйна: разложим нелинейные функции в ряды Тейлора (до первого порядка включительно) по степеням r и w . Оставим в полученной системе только те слагаемые, которые содержат только первые степени $r(t, x)$, $w(t, x)$ и их производных (так, например, слагаемые, содержащие r_r или rw , будут отброшены). Будем далее искать решение линеаризованной системы Пэйна в виде

$$\bar{z} \exp(ikx - i\omega t).$$

Подставляя этого кандидата в линеаризованную систему Пэйна, получим систему двух линейных уравнений

$$A(k, \omega) \bar{z} \exp(ikx - i\omega t) = \bar{0}.$$

Из условия разрешимости системы

$$\det A(k, \omega) = 0$$

получим дисперсионное соотношение (используя аналогию с распространением электромагнитных волн в дисперсионных средах, можно сказать: “получим зависимость частоты ω от волнового числа $k = 2\pi/\lambda$ (или длины волны λ)”):

$$\Lambda: \tau \cdot (\omega - V(\rho)k)^2 + i \cdot (\omega - Q'(\rho)k) - D(\rho)k^2 = 0.$$

С некоторыми оговорками любое решение начальной задачи Коши для линеаризованной системы может быть представлено в виде интеграла по многообразию $(k, \omega) \in \Lambda \subset \mathbb{R} \otimes \mathbb{C}$ (*теорема Эренпрайса–Паламодова*)

$$\int_{\Lambda} \bar{z}(k, \omega) \exp(ikx - i\omega t) d\mu(k, \omega),$$

где $\mu(k, \omega)$ и $\bar{z}(k, \omega)$ подбираются так, чтобы выполнялось начальное условие (см., например, фундаментальный труд Л. Хёрмандера [112]). Из такого представления следует, что для устойчивости стационарного решения (возмущения стационарного решения со временем затухают) достаточно (а с незначительными уточнениями и необходимо) потребовать, чтобы многообразие Λ , которое можно представить в виде двух кривых (корней квадратного уравнения – дисперсионного соотношения):

$$\omega_1(k), \quad \omega_2(k), \quad k \in \mathbb{R},$$

лежало в полупространстве $\operatorname{Im} \omega < 0$. Приходим к выписанному далее условию устойчивости.

Эта формула объясняет экспериментально установленный факт: при больших значениях плотности поток АТС становится неустойчивым. По этой причине его трудно адекватно описывать.

Обратим также внимание на одно интересное обстоятельство. Оказывается, модель Ньюэлла при $\tau = 0$ соответствует хорошо известной экономистам (и подробно изученной) модели Полтеровича–Хенкина, описывающей распространение новых технологий [97–101; 104–106]. Для обнаружения соответствия достаточно выписать, исходя из модели Ньюэлла, цепочку дифференциальных уравнений для скоростей $v_n(t)$. Если интерпретировать скорость $v_n(t)$ как функцию распределения $F_k(t)$ предприятий в отрасли производства по уровням эффективности $k = -n$, то получим цепочку уравнений Полтеровича–Хенкина.

Заметим, что модель Ньюэлла тесно связана со схемами бегущего счёта и дивергентными схемами С. К. Годунова (см., например, [18, 33, 54, 55], а также п. 2.1.4). Но в отличие от разностных схем в модели Ньюэлла время течёт непрерывно. Если же дискретизировать время по схеме Эйлера в этой модели, то получится разностная схема, принадлежащая упомянутым выше классам разностных схем, устойчиво аппроксимирующая закон сохранения.

Заметим также, что модель Ньюэлла с $\tau = 0$ тесно связана с моделью Танака (см. п. 2.1.2). Для того чтобы установить связь, нужно разрешить (и выбрать физически осмысленное решение) относительно $v_n(t)$ следующее уравнение:

$$s_{n+1}(t) - s_n(t) = L + c_1 v_n(t) + c_2 (v_n(t))^2.$$

2.2.2. Модель следования за лидером Дженерал Моторс

Другим важным классом микроскопических моделей (наряду с моделями оптимальной скорости) являются *модели следования за лидером* [8, 10, 56, 58, 113].

В 1959 г. сотрудники концерна Дженерал Моторс Д. Газис, Р. Херман, Р. Потс [58, 113] предложили одну из первых (первыми, по-видимому, были модели А. Рёшеля (1950) и Л. Пайпса (1953)) нетри-виальных микроскопических моделей однополосного транспортного потока (обозначения те же, что и выше), с помощью которой можно

получить фундаментальную диаграмму. Простейшим вариантом предложенной модели является следующая модель:

$$s_n''(t + \tau) = \alpha \frac{s_{n+1}'(t) - s_n'(t)}{s_{n+1}(t) - s_n(t)}, \quad \alpha > 0.$$

Ускорение n -го АТС $s_n''(t)$ прямо пропорционально разности скоростей:

$$\Delta v_n(t) = s_{n+1}'(t) - s_n'(t)$$

(если $\Delta v_n(t) > 0$, то $s_n''(t) > 0$ – ускорение n -го АТС; $\Delta v_n(t) < 0$ – торможение; $\Delta v_n(t) = 0$ – стационарный режим (ускорение равно нулю)) с коэффициентом пропорциональности (чувствительности) обратно пропорциональным расстоянию до впереди идущего АТС:

$$h_n(t) = s_{n+1}(t) - s_n(t).$$

Перепишем эту модель следующим образом:

$$\frac{dv_n(t)}{dt} = \alpha \frac{d}{dt} \ln h_n(t),$$

или

$$v_n(t + \tau) - v_n(\tau) = \alpha \ln \frac{h_n(t)}{h_n(0)}.$$

Положим

$$v_n(\tau) = 0, \quad h_n(0) = 1/\rho_{\max}.$$

Тогда (в обозначениях предыдущего пункта)

$$V(\rho) = \alpha \ln \left(\frac{\rho_{\max}}{\rho} \right), \quad \rho > \frac{1}{L}.$$

Эта зависимость была экспериментально обнаружена Х. Гринбергом также в 1959 г. по данным для туннеля Линкольна в Нью-Йорке.

В 1961 г. Д. Газис, Р. Херман и Р. Розэри предложили следующую модель [56, 113]:

$$s_n''(t + \tau) = \beta \frac{\left(s_n'(t + \tau) \right)^{m_1}}{\left(s_{n+1}(t) - s_n(t) \right)^{m_2}} \left(s_{n+1}'(t) - s_n'(t) \right), \quad \beta > 0,$$

где $m_1 (< 1)$, $m_2 (> 1)$ – эмпирически подбираемые константы ($m_1 \approx 0,8$, $m_2 \approx 2,8$). Исходя из этой микроскопической модели (путём интегрирования), несложно получить уравнение состояния транспортного потока:

$$V(\rho) = V^0 \cdot \left[1 - \left(\frac{\rho}{\rho_{\max}} \right)^{m_2-1} \right]^{\frac{1}{1-m_1}},$$

где V^0 – скорость свободного движения (желаемая скорость, максимально возможная скорость). При $m_1 = 0$, $m_2 = 2$ получим уравнение Гриншилдса состояния транспортного потока (см. п. 2.1.3).

Заметим, что время реакции $\tau > 0$ вводится в модели следования за лидером по той же причине, что и в модель Ньюэлла: для неустойчивости в линейном приближении стационарного режима движения при больших значениях плотности.

2.2.3. Модель «разумного водителя» Трайбера

Модели оптимальной скорости и следования за лидером можно объединить в одну общую *микроскопическую модель разумного водителя* (*Intelligent Driver Model (IDM)*):

$$s_n''(t) = F(s_{n+1}(t) - s_n(t), s_{n+1}'(t) - s_n'(t), s_n'(t)).$$

Как показали численные эксперименты, наиболее «удачной» моделью этого класса является²⁰ модель *M. Трайбера (Трайба)* (1999) [10, 56, 58, 114, 115]:

$$s_n''(t) = a_n \cdot \left[1 - \left(\frac{s_n'(t)}{V_n^0} \right)^\delta - \left(\frac{d_n^*(s_n'(t), s_{n+1}'(t) - s_n'(t))}{s_{n+1}(t) - s_n(t)} \right)^2 \right].$$

Первое слагаемое

$$a_n \cdot \left[1 - \left(\frac{s_n'(t)}{V_n^0} \right)^\delta \right]$$

в правой части этого соотношения описывает динамику ускорения АТС на свободной дороге, в то время как второе слагаемое описывает торможение из-за взаимодействия с лидером (впереди идущим АТС). Собственно модель

²⁰ Калибровка и численные эксперименты с этой моделью показали, что ее свойства устойчивы к вариации параметров; модель демонстрирует реалистическое поведение при разгоне и торможении и воспроизводит основные наблюдаемые свойства однополосного транспортного потока.

$$s_n''(t) = a_n \cdot \left[1 - \left(\frac{s_n'(t)}{V_n^0} \right)^\delta \right]$$

даже более естественно называть моделью оптимальной скорости, чем, скажем, модель Ньюэлла, которая, скорее, ближе к моделям следования за лидером, а модель

$$s_n''(t) = a_n \cdot \left[1 - \left(\frac{d_n^*(s_n'(t), s_{n+1}'(t) - s_n'(t))}{s_{n+1}(t) - s_n(t)} \right)^2 \right]$$

естественно называть моделью следования за лидером.

Очевидно, что параметр δ отвечает за поведение при разгоне (при $\delta=1$ имеет место экспоненциальный по времени разгон, в пределе при $\delta \rightarrow \infty$ разгон происходит с постоянным «комфортным» ускорением a_n вплоть до достижения желаемой скорости V_n^0). Тормозящий член определяется отношением желаемой дистанции d_n^* (безопасным расстоянием) к фактической дистанции:

$$h_n(t) = s_{n+1}(t) - s_n(t),$$

причем желаемая дистанция определяется следующим образом:

$$d_n^*(s_n'(t), s_{n+1}'(t) - s_n'(t)) = d_n + T_n s_n'(t) - \frac{s_n'(t)(s_{n+1}'(t) - s_n'(t))}{2\sqrt{a_n b_n}},$$

где d_n – расстояние между АТС (n -м и $(n+1)$ -м) в заторе (естественно, что $d_n \geq L$, где $L \sim 5,7$ м – средняя длина АТС, и, действительно, принято считать, что $d_n \sim 7,5$ м), b_n – ускорение «комфортного» торможения ($a_n \sim b_n \sim 2 \text{ м}/\text{с}^2$), T_n – аналог времени реакции водителя.

Поясним предложенную для безопасного расстояния формулу. Пока водитель n -го АТС реагирует на изменение ситуации впереди, он проедет путь $T_n s_n'(t)$. Потом, «поняв», что надо, скажем, тормозить» ($s_{n+1}'(t) < s_n'(t)$), он успеет выровнять свою скорость со скоростью впереди идущего АТС (двигаясь с ускорением торможения b_n) до момента, когда достигнет $(n+1)$ -е АТС, только если расстояние в момент, когда «пришло понимание» между n -м и $(n+1)$ -м АТС, было не меньше

$$-\frac{s_n'(t)\left(s_{n+1}'(t) - s_n'(t)\right)}{2\sqrt{a_n b_n}}.$$

Ситуация, когда надо ускоряться ($s_{n+1}'(t) > s_n'(t)$), рассматривается аналогичным образом. Собственно, из-за желания охватить «одной формулой» две довольно разные ситуации (ускорение и торможение) и возник знаменатель $2\sqrt{a_n b_n}$.

Заметим, что в *правилах дорожного движения (ПДД)* некоторых стран величина T_n достаточно жестко регламентирована (ограничена снизу ПДД). Так, например, в США от водителя требуют увеличивать безопасное расстояние (считается, что впереди идущее АТС имеет ту же скорость) на длину АТС L при увеличении скорости на 5 м/с (т.е. на 18 км/ч). Таким образом,

$$T_n \sim 5,7 \text{ [м]} / 5 \text{ [м/с]} \sim 1,1 \text{ с},$$

что хорошо согласуется с оценками этой величины, приведенными (см. п. 2.1.4) и полученными (см. п. 2.2.1) ранее.

В равновесном потоке одинаковых АТС, когда

$$\begin{aligned} s_n''(t) &\equiv 0, \quad s_{n+1}'(t) - s_n'(t) \equiv 0, \quad s_n'(t) \equiv V : \\ d(V) &\stackrel{\text{def}}{=} s_{n+1}(t) - s_n(t) = d^*(V, 0) \left[1 - \left(V/V^0 \right)^\delta \right]^{-1/2}. \end{aligned}$$

Из этого соотношения (считая, что $\rho(v) = 1/d(V)$) можно сначала построить уравнение состояния транспортного потока – зависимость $V(\rho)$, а потом фундаментальную диаграмму $Q(\rho)$. В пределе при $\delta \rightarrow \infty$ так построенная фундаментальная диаграмма будет стремиться к треугольной (см. рис. 8):

$$Q(\rho) = \min \left\{ \rho V^0, \frac{1-d\rho}{T} \right\}.$$

В этой формуле (а также и в других формулах этого пункта) вместо среднего расстояния между соседними АТС в заторе $d \sim 7,5$ м пишут также среднюю длину АТС $L \sim 5,7$ м (см., например, п. 2.2.1). В частном случае, когда $\delta = 1$, $d_n \sim 0$ м (отметим, что в этом случае адекватность модели в значительной степени сохраняется), можно найти и аналитическое выражение для равновесной скорости $V(\rho)$.

2.2.4. Модели клеточных автоматов²¹

В моделях клеточных автоматов (Cellular Automata (CA)) дорога разбивается на клетки, дискретным считается и время. Часто (но далеко не всегда [80–83], см. также ниже в этом пункте) считается, что в клетке может находиться не больше одного АТС. Таким образом, получаются разностные аналоги рассматриваемых ранее макроскопических уравнений. Заметим также, что часто и множество возможных значений скорости АТС считают дискретным в таких моделях.

Концепция клеточных автоматов была введена Дж. фон Нейманом (Нойманом) в 50-е годы XX века [116] в связи с разработкой теории самовоспроизводящихся машин. Применять клеточные автоматы для моделирования транспортных потоков предлагалось в работе [117]. Однако активное использование этой концепции началось только после работы К. Нагеля и М. Шрекенберга [118] (подробности см. в обзорах [119, 120]).

Опишем вкратце модель Нагеля–Шрекенберга (1992). В СА-модели на каждом шаге $m \rightarrow m+1$ состояние всех АТС в системе обновляется в соответствии со следующими правилами.

Шаг 1. Ускорение (отражает тенденцию двигаться как можно быстрее, не превышая максимальную допустимую скорость):

$$v_n(m+1) = \min \{v_n(m)+1, v_{\max}\}.$$

Шаг 2. Торможение (гарантирует отсутствие столкновений с впереди идущими АТС):

$$v_n(m+1) = \min \{v_n(m), s_{n+1}(m) - s_n(m) - d\},$$

где $d \sim 7,5$ м (см. п. 2.2.3).

Шаг 3. Случайные возмущения (учитывают различия в поведении АТС):

$$v_n(m+1) = \begin{cases} \max \{v_n(m)-1, 0\}, & p, \\ v_n(m), & 1-p. \end{cases}$$

Шаг 4. Движение:

$$s_n(m+1) = s_n(m) + v_n(m).$$

Все четыре приведенных шага необходимы для воспроизведения основных свойств транспортного потока. Так, например, шаг 3 обуслав-

²¹ Этот пункт написан совместно с Я. А. Холодовым.

ливают неустойчивость транспортного потока при достаточно больших плотностях.

В работах [121, 122] описан переход от моделей типа Нагеля–Шрекенберга к гидродинамическим моделям (типа Бюргерса, Хопфа) – гидродинамический предельный переход. А в работе [123] описан обратный переход – ультраметрический предельный переход.

Продемонстрируем, следуя М. Л. Бланку [124–126] (см. также приложение М. Л. Бланка и цитированную там литературу), один из способов вывода уравнения состояния $V(\rho)$, исходя из довольно простой СА-модели.

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» (это понятие пришло из статистической физики, см., например, [127] и цитированную там литературу) – 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:

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

Тогда по эргодической теореме для марковских процессов (см. приложение Е. В. Гасниковой) при $0 < p < 1$ средняя временная скорость каждого АТС (при $p = 1$ см. [124])

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

Зависимость «средней» скорости V от плотности ρ изображена на рис. 11, из которого становится ясно, что фундаментальная диаграмма, соответствующая рис. 11, будет треугольного типа. Ввиду простоты модели несложно качественно объяснить приведенную на рис. 11 зависимость.

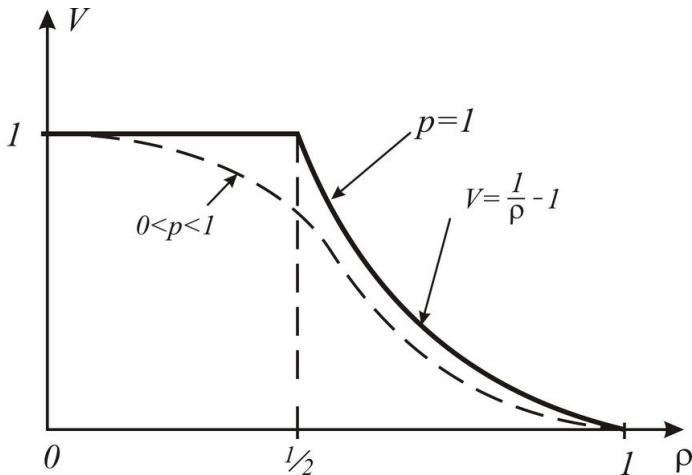


Рис. 11. «Фазовый переход» при $\rho = 1/2$

Для дальнейшего знакомства с СА-моделями см. [128]. Оказывается, что простые модификации рассматриваемых здесь моделей демонстрируют [119, 126, 128], что одной плотности соответствует целый набор средних скоростей, или последнее понятие может не быть корректно определено. С точки зрения фазовых переходов описанное поведение соответствует возникновению новой «гистерезисной» фазы, которая, по-видимому, соответствует режиму синхронизированного движения (см. [10] или главу 3).

Интересные идеи исследования многополосности (которые также объясняют экспериментальную «размазанность» фундаментальной диаграммы) недавно были предложены А. П. Буслаевым и др. (кафедра высшей математики МАДИ) [129–131]. Для описания транспортного потока использовалась модель Танака (см. п. 2.1.2), в которую помимо плотности и скорости еще вводился «параметр регулярности», с помощью которого “часть водителей АТС «разрыхляет регулярное движение» для того, чтобы осуществить относительное перемещение внутри потока”. Отметим также, что с помощью моделей такого типа удалось теоретически объяснить невогнутость фундаментальной диаграммы (рис. 2) в случае многополосного движения.

Перейдем теперь к другому типу клеточных моделей, которые, по сути, являются разностными аналогами рассматриваемых ранее уравнений

ний. В п. 2.1.6 мы уже упоминали одну из таких моделей (СТМ модель). Приведем схему, в которую «ложатся» многие модели этого класса. Ограничимся ситуацией магистрали (вообще говоря, многополосной) с въездами и выездами. 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 выезд (см. рис. 12). Тогда

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

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

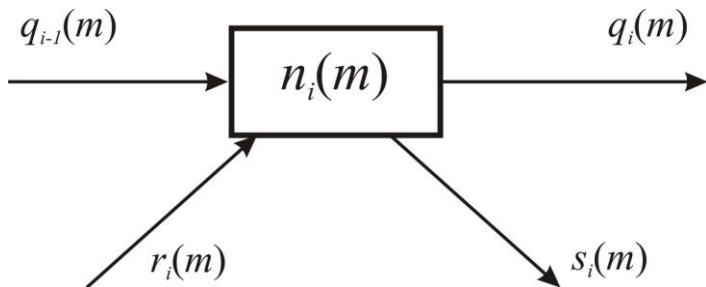


Рис. 12

Так, в работах [84, 85] (рассматривалась кольцевая топология транспортной сети без въездов и выездов) 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 следующих «неправдоподобных» ситуаций (считаем $\alpha_i \equiv 0$ и $r_i(m) \equiv 0$). Предположим, что есть две клетки (без въездов и выездов). 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. Предположим теперь, что первая клетка загружена, например, наполовину, а в следующей по ходу движения клетки – стоячая пробка. Тогда $q_i(m) \equiv 0$, а модель говорит об обратном. Тем не менее исследование

«до критических режимов» – «левая (возрастающая) ветка» фундаментальной диаграммы с помощью этой модели вполне корректно.

В работах [80–83] (рассматривался случай треугольной фундаментальной диаграммы (см. рис. 8), «вершина» которой имеет координаты $(\tilde{n}_i, Q_i^{\max})$) полагали

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

– аналог схемы Годунова (см. п. 2.1.4) для модели LWR с треугольной фундаментальной диаграммой (СТМ-модель К. Даганзо (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» (знания матрицы перемешивания может оказаться не достаточно). Например, таких, как следующая: пусть имеется перекресток. Две трети водителей с входящей в перекресток дороги A хотят повернуть на выходящую с перекрестка дорогу C , и две трети водителей с входящей дороги B также хотят повернуть на дорогу C . Будем считать, что двухполосная дорога C может пропустить 4000 АТС/час (максимальная пропускная способность). Дороги A и B также двухполосные, и на обеих из них поток АТС, входящий в перекресток, по 4000 АТС/км. Очевидно, что ситуация не доопределена. Для понимания того, что бу-

дет 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 «беря на вооружение» правило работы светофора, можно получить упомянутое обобщение рассмотренных моделей на графы транспортной сети общего вида. Аналогично можно рассмотреть и более сложные развязки. Здесь также можно упомянуть, что еще в 1936 году молодой профессор Московского университета А. Н. Колмогоров в письме в журнал «Строительство Москвы», по сути, обсуждал вопрос, связанный с правильной организацией перекрестка [132]. Кстати сказать, проходившее в августе 2010 года обследование различных развязок г. Москвы специалистами из Японии, приглашенными руководством г. Москвы, показало «не оптимальность» ряда важных развязок.

Как уже отмечалось, несмотря на свою относительную простоту, СТМ-модель является одной из наиболее востребованных в приложениях (см., например, [80–83] и цитированную там литературу).

Что касается изучения аналитических свойств этих моделей (полученных с помощью разностных схем для LWR-модели), то, по-видимому, проще исследовать все-таки дифференциально-разностные аналоги LWR-моделей, т.е. считать, что время течет непрерывно [84, 85]. Разностные схемы (в том числе и упомянутые выше) очевидным образом переделываются в дифференциально-разностные, при этом необходимое условие Куранта–Фридрихса–Леви (см. п. 2.1.4) 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. [84] для графов специальной структуры (кольцевой, «цветочной» и т.п.). В 2006 г. А. И. Назаров обобщил результаты статьи [85] 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 [80] исследовалась асимптотическая устойчивость (глобаль-

ная) 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 систему обыкновенных дифференциальных уравнений (СОДУ) на графе транспортной сети,²² используя лишенную деталей модификацию правила «пропорциональных приоритетов» [81, 83]:

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

здесь каждое ребро ориентированного графа транспортной сети пронумеровано, ρ_i – плотность потока на i -м 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, $\sum_{k: i \rightarrow k} \beta_k^i(t, \vec{\rho}) < 1$. Имеются и другие «правила обработки» узлов графа транспортной сети (см., например, [34, 78, 79]).

Упражнение ** (мотивированное работами [80, 84, 85]). 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 треугольных фундаментальных диаграмм.

²² Для простоты считаем, что нет источников и стоков АТС, в противном случае их следовало бы учитывать, например, подобно тому, как это было сделано выше.

2.3. МОДЕЛЬНЫЕ ЗАДАЧИ

В пункте 2.3 будут приведены решения ряда *модельных задач*. Основная задача – это *задача об эволюции затора* (локального, глобального) в предположении, что транспортный поток описывается моделью LWR, моделью Уизема, моделью Ньюэлла, моделью Пэйна. Будет определена *скорость распространения затора*. Заметим, что, как будет показано в этом пункте, информация о заторе, вообще говоря, может распространяться по транспортному потоку не только в виде одной (*бегущей, ударной*) волны, но и в виде *системы таких волн*, в которую могут также входить и *волны разрежения*. В заключение будет рассмотрена (путем решения ряда модельных задач об эволюции затора по модели LWR) *задача Лайтхилла–Уизема о светофоре*: при каком соотношении между временами горения красного и зеленого сигналов светофора перед ним не будет скапливаться очередь?

2.3.1. Эволюции глобального затора в транспортном потоке, описываемом моделями LWR и Уизема

Для простоты изложения будем везде в дальнейшем в этом пункте считать, что $Q(\rho)$ – кусочно-гладкая функция, имеющая кусочно-гладкие производные до четвертого порядка включительно и не имеющая точек сгущения нулей второй производной.

Напомним необходимые соотношения (по модели LWR):

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = 0, \quad (13)$$

$$\rho(0, x) = \begin{cases} \rho_-, & x < 0 \\ \rho_+, & x \geq 0 \end{cases} \quad (\text{начальное условие Римана}), \quad (14)$$

$$\rho(0, x) = \begin{cases} \rho_-, & x < x_- \\ \rho_0(x), & x_- \leq x < x_+ \\ \rho_+, & x \geq x_+ \end{cases} \quad (\text{начальное условие типа Римана}), \quad (15)$$

где $\rho_0(x)$ – ограниченная измеримая функция (см. рис. 13, 14). Для определенности будем считать в формулах (14), (15) $\rho_- < \rho_+$. Случай $\rho_- > \rho_+$ рассматривается аналогично и может быть заменой $\rho \rightarrow -\rho$.

сведен к случаю $\rho_- < \rho_+$ (с $Q(\rho) := Q(-\rho)$). Задача Коши (13), (14) называется задачей (Римана) о распаде (произвольного) разрыва.

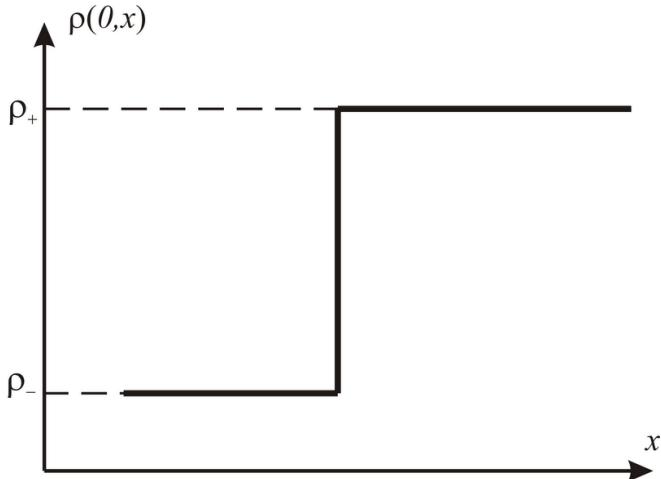


Рис. 13. Начальное условие Римана

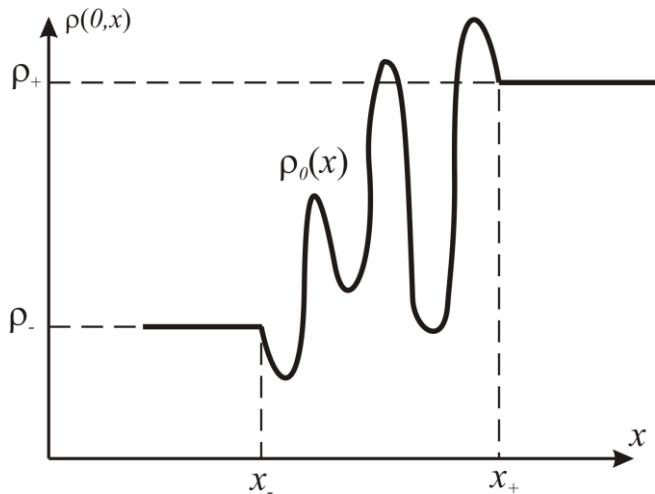


Рис. 14. Начальное условие типа Римана

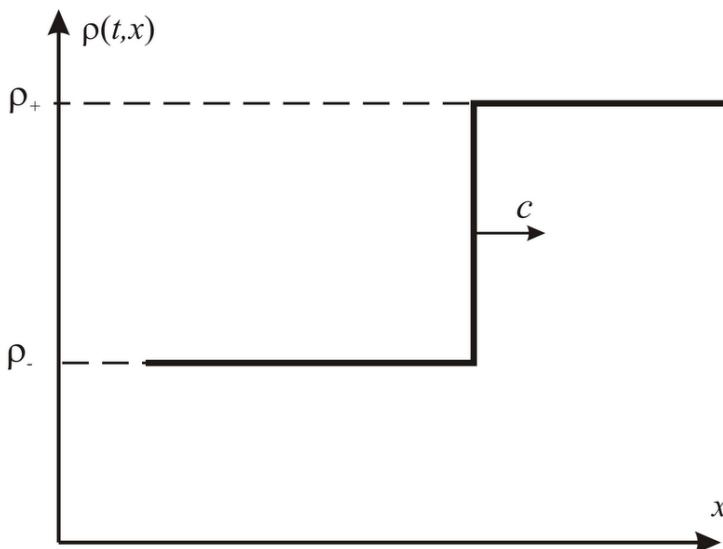


Рис. 15 Ударная волна

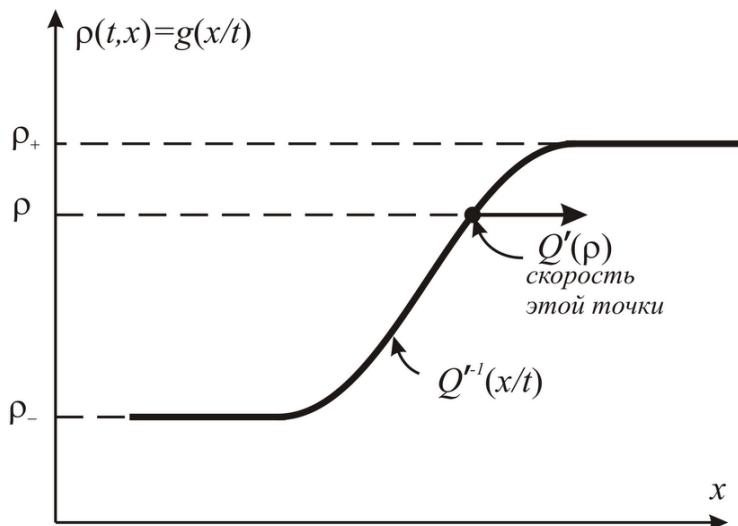


Рис. 16. Волна разрежения

Для того чтобы понять, какое решение будет иметь задача Коши (13), (14) (впрочем, частично ответ на этот вопрос мы уже можем дать, основываясь на примерах, разобранных в п. 2.1.1), заметим сначала, что:

- закон сохранения (13) имеет однопараметрическое ($x_0 \in \mathbb{R}$ – параметр) семейство автомодельных решений вида *ударной волны* (см. рис. 15):

$$\tilde{\rho}(x-ct) = \begin{cases} \rho_-, & x < x_0 + ct, \\ \rho_+, & x \geq x_0 + ct \end{cases}$$

тогда и только тогда, когда на разрыве этой ударной волны выполняется RRH-условие и E-условие;

- закон сохранения (13) имеет двухпараметрическое ($t \rightarrow t + \tau$; $x \rightarrow x + a$) семейство автомодельных решений вида *волны разрежения* (см. рис. 16):

$$g\left(\frac{x}{t}\right) = \begin{cases} \rho_-, & x < Q'(\rho_-)t, \\ Q'^{-1}\left(\frac{x}{t}\right), & Q'(\rho_-)t \leq x < Q'(\rho_+)t, \\ \rho_+, & x \geq Q'(\rho_+)t \end{cases} \quad (16)$$

($Q'^{-1}(\cdot)$ означает обратную функцию к функции $Q'(\cdot)$)

тогда и только тогда, когда $Q''(\rho) > 0$ при $\rho \in (\rho_-, \rho_+)$, за исключением, быть может, конечного числа точек, в которых имеет место равенство (напомним, что $\rho_- < \rho_+$).

Замечание. По поводу автомодельных решений см., например, [1, 133, 134], а также более ранние книги этих авторов. Грубо говоря, автомодельные решения – это решения рассматриваемого уравнения в частных производных, которые могут быть представлены как функции (в данном случае) одного аргумента, зависящего от пространственных и временных переменных (т.е. от t и x). Групповой анализ указывает на то, что если автомодельные решения существуют, то прежде всего их следует искать в виде неизвестной функции от инвариантов (сконструированных из аргументов неизвестной функции (т.е. от t и x)) группы преобразований, допускаемой рассматриваемым уравнением. Поясним сказанное немного подробнее. Хорошо известно, что сфера, например, допускает группу ортогональных преобразований (всевозможных поворотов), т.е. сфера будет переходить сама в себя при таких преобразованиях.

ниях. С другой стороны, на сферу можно смотреть, как на многообразие, заданное хорошо известным уравнением. Групповой анализ предлагает смотреть на уравнение в частных производных, как на многообразие, заданное этим уравнением в продолженном (на различные частные производные) пространстве. Правда, в отличие от обычных многообразий, на класс возможных групп преобразований налагается условие, по сути, определяющее действие группы на продолженных переменных, через действие группы на неизвестную функцию и её аргументы.

В качестве примера укажем, что закон сохранения (13) допускает группу трансляций по времени и по координате, группу подобных преобразований временной и пространственной переменной. Это означает, что вид уравнения (13) не поменяется при заменах:

$$t \rightarrow t + \tau; \quad x \rightarrow x + a; \quad t \rightarrow kt, \quad x \rightarrow kx.$$

Отсюда, поскольку $x - ct$ (при любом $c \in \mathbb{R}$) является инвариантом группы трансляций:

$$t \rightarrow t + \tau, \quad x \rightarrow x + c\tau,$$

а $\xi = x/t$ является инвариантом группы растяжений:

$$t \rightarrow kt, \quad x \rightarrow kx,$$

в частности, следует, что решение уравнения (13) прежде всего следует попробовать искать в виде

$$\tilde{\rho}(x - ct) \text{ и } g\left(\frac{x}{t}\right).$$

Роль автомодельных решений (параметрических семейств таких решений) в теории эволюционных (эволюция по времени) уравнений в частных производных часто аналогична роли неподвижных точек в теории обыкновенных дифференциальных уравнений [1; 89–97; 100–111; 133–144]. Параметрическое семейство автомодельных решений (или семейство, полученное путем «склеивания» автомодельных решений) часто является, например, притягивающим (другие решения) семейством (см. теоремы 1, 2 в п. 2.3.1).

Зная приведенный выше (в этом пункте) материал и тот материал, который был изложен в п. 2.1.1, И. М. Гельфанд построил в конце 50-х годов XX века (см. [17, 28]) решение задачи о распаде произвольного разрыва, т.е. решение задачи Коши (13), (14). Можно было также добавить (см. п. 2.1.3): «построил с помощью метода исчезающей вязкости». Поскольку, как будет показано ниже в этом пункте, выполнение RRH-условия и E-условия на разрывах обобщенных решений уравнения (13) есть прямое следствие этого метода.

Приведем способ построения решения. Строгое обоснование см. в оригинальных, и отчасти монографических, записях курса лекций [28], которые С. Н. Кружков читал в конце 60-х и начале 70-х годов XX века на мехмате МГУ. Для этого введем функцию

$$\tilde{Q}(\rho) = F^{**}(\rho), \quad F(\rho) = Q(\rho) + I_{[\rho_-, \rho_+] }(\rho),$$

$$I_{[\rho_-, \rho_+]}(\rho) = \begin{cases} 0, & \rho \in [\rho_-, \rho_+], \\ \infty, & \rho \notin [\rho_-, \rho_+], \end{cases}$$

где

$$F^*(x) = \sup_{y \in \mathbb{R}} (xy - F(y))$$

– функция, сопряженная (по Юнгу–Фенхелю–Лежандру) к $F(y)$ [45].

Таким образом, $\tilde{Q}(\rho)$ – нижняя граница выпуклой оболочки множества

$$\{(\rho, q) : \rho \in [\rho_-, \rho_+], q \geq Q(\rho)\}.$$

Будем искать автомодельное решение уравнения

$$\frac{\partial \rho}{\partial t} + \frac{\partial \tilde{Q}(\rho)}{\partial x} = 0 \quad (17)$$

вида

$$\rho(t, x) := \rho\left(\frac{x}{t}\right)$$

(заметим, что $\xi = x/t$ является инвариантом группы растяжений: $x' \rightarrow kx$, $t' \rightarrow kt$, допускаемой законом сохранения (17)), удовлетворяющее начальному условию (14). Подстановка $\rho(x/t)$ в (17) приводит к обыкновенному дифференциальному уравнению

$$\rho'(\xi)(\xi - \tilde{Q}'(\rho)) = 0.$$

Откуда с учетом начального условия (14) следует, что

$$\rho(t, x) = \tilde{Q}'^{-1}\left(\frac{x}{t}\right) (\tilde{Q}'^{-1}(\cdot) – обратная функция к $\tilde{Q}'(\cdot)$).$$

Для уточнений и большей наглядности представим решение немного в другом виде. Положим (см. рис. 17)

$$\{\rho \in [\rho_-, \rho_+] : Q(\rho) > \tilde{Q}(\rho)\} = (\alpha_0, \beta_0) \cup (\alpha_1, \beta_1) \cup \dots \cup (\alpha_n, \beta_n),$$

$c_0 = \tilde{Q}'(\alpha_0) = Q'(\beta_0)$, $c_n = Q'(\alpha_n) = \tilde{Q}'(\beta_n)$,
 $c_k = Q'(\alpha_k) = Q'(\beta_k)$ (очевидно, что $c_{k-1} \leq c_k$), $k = 1, \dots, n-1$,
где $\rho_- = \alpha_0 \leq \beta_0 \leq \alpha_1 < \beta_1 \leq \alpha_2 < \dots < \beta_{n-1} \leq \alpha_n \leq \beta_n = \rho_+$.

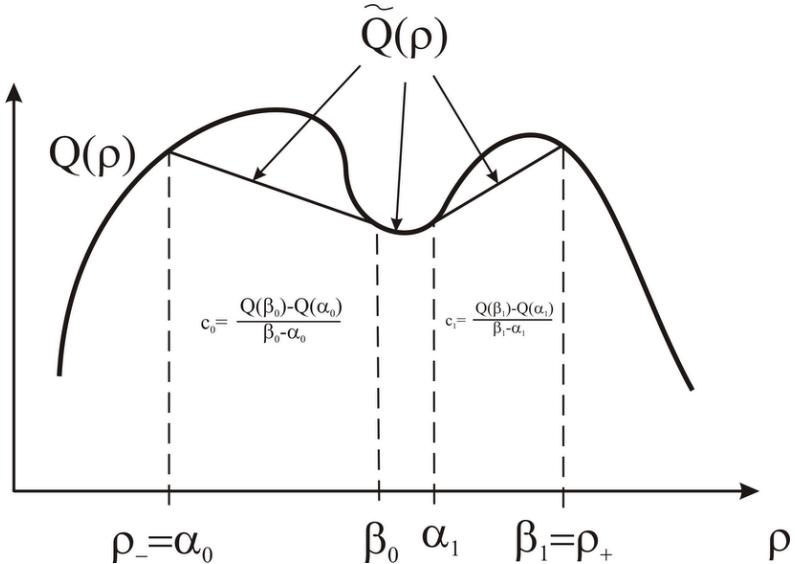


Рис. 17. Построение И. М. Гельфандома (1958)

И будем считать (для простоты и наглядности формулировок), что выполняются условия «отсутствия прилипания» [106]:

$$Q''(\rho) > 0 \text{ при } \rho \in [\beta_k, \alpha_{k+1}], \quad k = 0, \dots, n-1;$$

$$Q''(\alpha_0) > 0, \text{ если } \alpha_0 < \beta_0 \text{ и } Q'(\alpha_0) = c_0;$$

$$Q''(\beta_n) > 0, \text{ если } \alpha_n < \beta_n \text{ и } Q'(\beta_n) = c_n.$$

Тогда (см. рис. 18 с $d_k = 0$)

$$\rho(t, x) = \begin{cases} \rho_-, & x < c_0 t, \\ Q'^{-1}(x/t), & c_{k-1} t \leq x < c_k t, \quad k = 1, \dots, n, \\ \rho_+, & x \geq c_n t. \end{cases} \quad (18)$$

Заметим, что по построению мы можем быть уверены лишь в нестрогих неравенствах:

$Q''(\rho) \geq 0$ при $\rho \in [\beta_k, \alpha_{k+1}]$, $k = 0, \dots, n-1$;

$Q''(\alpha_0) \geq 0$, если $\alpha_0 < \beta_0$ и $Q'(\alpha_0) = c_0$;

$Q''(\beta_n) \geq 0$, если $\alpha_n < \beta_n$ и $Q'(\beta_n) = c_n$.

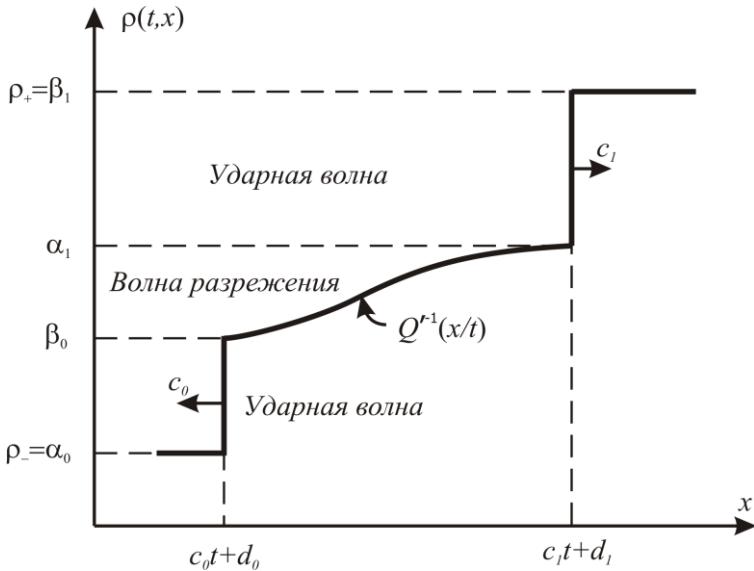


Рис. 18. Система волн

Если же условия «отсутствия прилипания» не выполняются, то все приведенные ниже результаты остаются в принципе верными (с некоторыми несущественными для приложений оговорками), но оценки будут менее точными (см. [110, 111] и цитированную там литературу).

Заметим также, что $\rho(t, x)$ определяется с точностью до почти всюду по x при любом фиксированном значении $t \geq 0$ (см. п. 2.1.3). Поэтому в формуле (18) можно не доопределять значения $\rho(t, x)$ на разрывах. Естественно называть разрывы решения (18) *ударными волнами*, а

$$Q'^{-1}(x/t), c_{k-1}t < x < c_k t, k = 1, \dots, n$$

— *волнами разрежения* (волна разрежения исчезает, если $c_{k-1} = c_k$, поскольку в этом случае $\beta_{k-1} = \alpha_k$).

Естественно ожидать, что если немного «размазать разрыв» в начальных условиях (14), то решение уравнения (13) будет иметь схожий вид. Иначе говоря, мы ожидаем, что решение задачи Коши (13), (15) будет структурно похоже на (18). Действительно, имеет место следующий факт [14, 90; 93–96; 106, 111] (см. рис. 18).

Теорема 1. *Существует такой набор $\{d_k\}_{k=0}^n$ (причем если $c_{k-1} = c_k$, то $d_{k-1} \leq d_k$), что решение задачи Коши (13), (15) сходится при $t \rightarrow \infty$ в $L_1(\mathbb{R}_x)$ к системе волн*

$$\tilde{\rho}(t, x; \{d_k\}_{k=0}^n) = \begin{cases} \rho_-, & x < c_0 t + d_0, \\ Q^{t^{-1}}(x/t), & c_{k-1} t + d_{k-1} \leq x < c_k t + d_k, \quad k = 1, \dots, n, \\ \rho_+, & x \geq c_n t + d_n. \end{cases} \quad (19)$$

Заметим, что в работах [90, 95, 96] также были найдены формулы для расчета d_k .

Рассмотрим теперь модель Уизема:

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \varepsilon \frac{\partial}{\partial x} \left(D(\rho) \frac{\partial \rho}{\partial x} \right), \quad D(\rho) > 0, \quad \varepsilon > 0. \quad (20)$$

«Построим» решение задачи Коши (20), (15). Поскольку задачи Коши (13), (15) и (20), (15) описывают одно и то же явление (на разных уровнях детализации) – эволюцию («размазанного») глобального затора в транспортном потоке, то мы вправе ожидать, что решение задачи Коши (20), (15) будет структурно похоже на решение задачи Коши (13), (15) и, стало быть, будет вести себя на больших временах подобно системе волн (19).

Действуя по тому же плану, что и при исследовании асимптотики (по времени) задачи Коши (13), (15), найдем сначала автомодельные решения уравнения (20). Согласно работам [17, 90]:

- **закон сохранения с диффузией (20) имеет однопараметрическое ($x_0 \in \mathbb{R}$ – параметр) семейство автомодельных решений вида бегущей волны (см. рис. 19):**

$$\tilde{\rho}_\varepsilon(x - ct + x_0),$$

таких, что

$$\lim_{s \rightarrow -\infty} \tilde{\rho}_\varepsilon(s) = \rho_-, \quad \lim_{s \rightarrow \infty} \tilde{\rho}_\varepsilon(s) = \rho_+,$$

тогда и только тогда, когда выполняется RRH-условие

$$c = \frac{Q(\rho_+) - Q(\rho_-)}{\rho_+ - \rho_-}$$

и строгое E-условие (напомним, что $\rho_- < \rho_+$)

$$\forall \rho \in (\rho_-, \rho_+) \rightarrow \sigma(\rho_-, \rho_+) < \sigma(\rho_-, \rho);$$

- закон сохранения с диффузией (20) имеет двухпараметрическое семейство асимптотически автомодельных решений (при подстановке такого автомодельного «решения» в уравнение (20) почти всюду возникает невязка (в нашем случае $O(1/t^2)$), которая стремится к нулю с ростом времени) вида *волны разрежения* (16) тогда и только тогда, когда $Q''(\rho) > 0$ при $\rho \in (\rho_-, \rho_+)$, за исключением, быть может, конечного числа точек, в которых имеет место равенство.

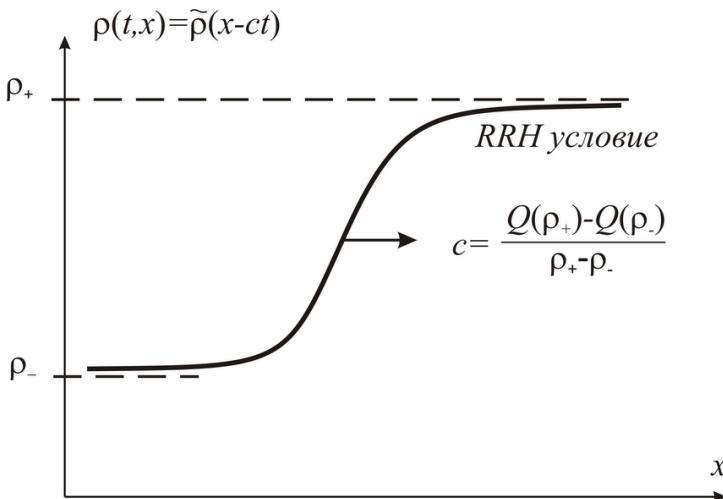


Рис. 19. Бегущая волна

Заметим, что для доказательства всех приведенных утверждений относительно бегущей волны достаточно рассмотреть краевую задачу для обыкновенного дифференциального уравнения с разделяющимися переменными, которое получается при подстановке

$$\tilde{\rho}_\varepsilon(x - ct + x_0)$$

в (20) и интегрировании возникшего уравнения с учетом одного из краевых условий $\lim_{s \rightarrow -\infty} \tilde{\rho}_\varepsilon(s) = \rho_-$:

$$\varepsilon \frac{d\tilde{\rho}_\varepsilon(s)}{ds} = (Q(\tilde{\rho}_\varepsilon(s)) - c\tilde{\rho}_\varepsilon(s) - (Q(\rho_-) - c\rho_-)) / D(\tilde{\rho}_\varepsilon(s)).$$

Заметим также, что RRH-условие получается из оставшегося краевого условия

$$\lim_{s \rightarrow \infty} \tilde{\rho}_\varepsilon(s) = \rho_+.$$

А строгое E-условие равносильно требованию отсутствия (за исключением $s = \pm\infty$) особых точек (точек, в которых $\tilde{\rho}'_\varepsilon(s) = 0$) у этого обыкновенного дифференциального уравнения, что в свою очередь равносильно условию

$$\tilde{\rho}'(s) > 0 \text{ (поскольку } \rho_- < \rho_+).$$

Если же допустить существование особой точки

$$s^* \in (-\infty, \infty),$$

то интеграл

$$\int (Q(\tilde{\rho}_\varepsilon) - c\tilde{\rho}_\varepsilon - (Q(\rho_-) - c\rho_-))^{-1} d\tilde{\rho}_\varepsilon$$

будет иметь неинтегрируемую особенность в точке

$$\tilde{\rho}_\varepsilon = \tilde{\rho}_\varepsilon(s^*).$$

Отсюда следует, что

$$s^* = \infty \text{ или } s^* = -\infty.$$

Пришли к противоречию с предположением

$$s^* \in (-\infty, \infty).$$

Теперь можно объяснить уже анонсированную возможность обоснования RRH-условия и E-условия на разрывах обобщенных решений уравнения (13) с помощью метода исчезающей вязкости:

$$\tilde{\rho}_\varepsilon(x - ct + x_0) = \tilde{\rho}_1\left(\frac{x - ct + x_0}{\varepsilon}\right) \xrightarrow[\varepsilon \rightarrow 0^+]{} \begin{cases} \rho_-, & x \leq ct - x_0, \\ \rho_+, & x > ct - x_0. \end{cases}$$

Упражнение* (см. [19, 28]). Почему строгое E-условие «переходит в пределе» при $\varepsilon \rightarrow 0^+$ просто в E-условие?

Замечание (модель Лайтхилла–Уизема–Кортевега–де Фриза–Бюргерса). Если, подобно модели Уизема (см. п. 2.1.3), положить

$$v(t, x) = V(\rho(t, x)) - \frac{1}{\rho(t, x)} \left(\varepsilon \frac{\partial \rho(t, x)}{\partial x} + \mu \frac{\partial^2 \rho(t, x)}{\partial x^2} \right),$$

где $V(\rho(t, x))$ – желаемая скорость при данной плотности, а выражение в скобке описывает « дальновидность » водителей (поэтому $\varepsilon > 0$, $\mu > 0$), то получим, с учётом закона сохранения количества АТС:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(v\rho)}{\partial x} = 0,$$

уравнение типа Кортевега–де Фриза (Вриза)–Бюргерса (КдФБ):

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \varepsilon \frac{\partial^2 \rho}{\partial x^2} + \mu \frac{\partial^3 \rho}{\partial x^3}. \quad (21)$$

Поскольку предложенная модель описывает то же самое явление, что и модель Уизема, то мы вправе надеяться на структурную схожесть автомодельных решений уравнений (20) и (21). Действительно, предположим, что выполняется строгое Е-условие, а скорость c определяется формулой РРН. Тогда (см. [138]) если²³

$$4 \sup_{\rho \in (\rho_-, \rho_+)} \left(c - \frac{Q(\rho) - Q(\rho_+)}{\rho - \rho_+} \right) \leq \frac{\varepsilon^2}{\mu},$$

то существует однопараметрическое семейство строго возрастающих автомодельных решений уравнения (21) вида бегущей волны $\tilde{\rho}_{\varepsilon, \mu}(x - ct + x_0)$. Легко показать, что уравнение (21) при $Q''(\rho) > 0$, $\rho \in (\rho_-, \rho_+)$ имеет и асимптотическое автомодельное решение вида волны разрежения (16). Интересно при этом заметить, что в отличие от уравнения типа Бюргерса для уравнения типа КдФБ обоснование метода исчезающей вязкости и дисперсии – по-прежнему открытая задача (см. статью П. Лакса в сборнике [30], а также [34, 36]).

Имеет место следующий факт (для простоты формулировки (общий случай изложен в статье [110]) также считаем, что $c_{k-1} < c_k$,

²³ Если это условие нарушается, то сначала (когда не сильно нарушается) теряется свойство монотонности бегущей волны (возникают затухающие осцилляции типа $(\sin x)/x$ при стремлении решения к предельным значениям на краях), а затем автомодельные решения вида бегущей волны вообще перестают существовать. Численное подтверждение этих фактов имеется в работах А. Г. Куликовского с коллегами (см., например, [54, 145] и цитированную там литературу). Аналитические исследования, подтверждающие сказанное выше, были недавно проведены А. В. Казейкиной (Вестник МГУ, серия 15. Вычислительная математика и математическая кибернетика. 2011. № 1 (в печати)).

$k = 1, \dots, n$ (или, что то же самое, $\beta_{k-1} < \alpha_k$, $k = 1, \dots, n$) [14, 90, 92; 102–110] (см. рис. 20).

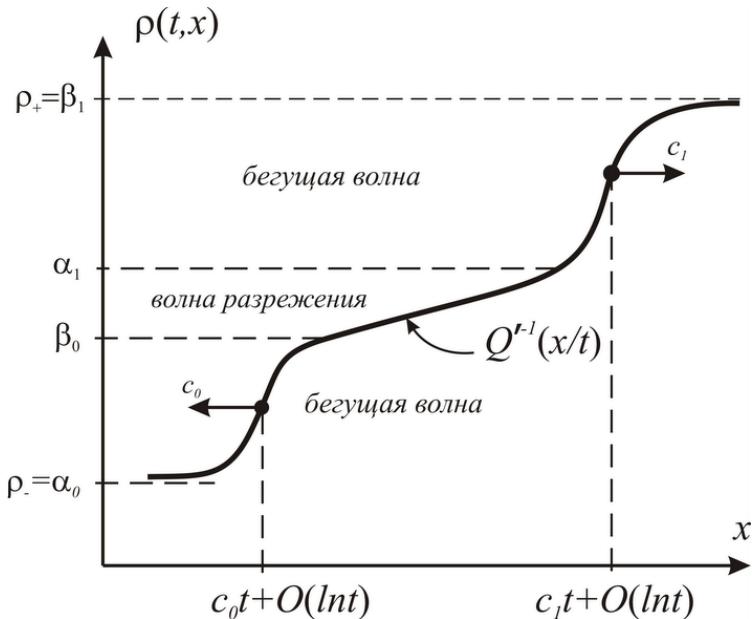


Рис. 20. Система волн

Теорема 2. Существует такой набор функций

$$\{d_k(t)\}_{k=0}^n = O(\varepsilon \ln t) + O(1),$$

что решение задачи Коши (20), (15) сходится при $t \rightarrow \infty$ в $L_\infty(\mathbb{R}_x)$ (равномерно по $x \in \mathbb{R}$) к системе волн

$$\tilde{\rho}(t, x; \{d_k(t)\}_{k=0}^n) = \begin{cases} \rho_-, & x < c_0 t - \sqrt{t}, \\ \tilde{\rho}^k(x - c_k t + d_k(t)), & c_k t - \sqrt{t} \leq x < c_k t + \sqrt{t}, \\ Q'^{-1}(x/t), & c_{k-1} t + \sqrt{t} \leq x < c_k t - \sqrt{t}, \\ \rho_+, & x \geq c_n t + \sqrt{t}, \end{cases} \quad k = 0, \dots, n,$$

зде $\tilde{\rho}^k(x - c_k t)$ – решение уравнения (20) вида бегущей волны с

$$\lim_{s \rightarrow -\infty} \tilde{\rho}^k(s) = \alpha_k, \quad \lim_{s \rightarrow \infty} \tilde{\rho}^k(s) = \beta_k.$$

Замечание (А. М. Ильин, О. А. Олейник (1960)). Можно показать (см. [90, 91, 107, 110]), что если

$$\rho_- = \alpha_0 < \beta_0 = \rho_+ \text{ и } Q'(\alpha_0) \neq c_0, Q'(\beta_0) \neq c_0,$$

то

$$\lim_{t \rightarrow \infty} d_0(t) = d_0,$$

где d_0 находится из первого интеграла (20) (при $t = 0$):

$$I(t; d_0) = \int_{-\infty}^{\infty} \left\{ \rho(t, x) - \tilde{\rho}^0(x - c_0 t + d_0) \right\} dx \equiv 0.$$

Полезной находкой А. М. Ильина и О. А. Олейник стало введение функции

$$w(t, x) = \int_{-\infty}^x \left\{ \rho(t, s) - \tilde{\rho}^0(s - c_0 t + d_0) \right\} ds,$$

которая так же, как и $\rho(t, x)$, удовлетворяет уравнению параболического типа, но при этом $\lim_{x \rightarrow \pm\infty} w(t, x) \equiv 0$.

Замечание, аналогичное замечанию Ильина–Олейник, имеет место и для теоремы 1.

Замечание (Г. М. Хенкин, А. А. Шананин (2004)). Можно показать (см. [104–106]), что асимптотические оценки сдвигов фаз

$$\{d_k(t)\}_{k=0}^n = O(\varepsilon \ln t) + O(1)$$

«разумно» определять из предложенных в работе [104] «локализованных законов сохранения (первых интегралов)»:

$$I_k(t; d_k(t)) = \int_{c_k t - \sqrt{t}}^{c_k t + \sqrt{t}} \left\{ \rho(t, x) - \tilde{\rho}^k(x - c_k t + d_k(t)) \right\} dx \equiv 0.$$

Если приравнять нулю полную производную по времени от $I_k(t; d_k(t))$, используя уравнение (20), то получим оценку $d_k'(t)$ (приняв априорно за гипотезу, которую апостериорно проверим, что $d_k'(t) = o(t^{-1/2})$). Точность оценки определяется значениями разностей:

$$\rho(t, c_k t \pm \sqrt{t}) - \tilde{\rho}^k(\pm \sqrt{t} + d_k(t)) \text{ и } \rho_x(t, c_k t \pm \sqrt{t}) - \tilde{\rho}_x^k(\pm \sqrt{t} + d_k(t)).$$

Эти значения в свою очередь определяются исходя из исследования сходимости решения к системе волн на участках, соответствующих поведению «волна разрежения». Оказывается, что выбор зависимости \sqrt{t} в определении системы волн и в определении локализованных законов сохранения даёт наилучшую по точности оценку сдвигов фаз, исходя из имеющихся способов оценки разностей. В заключение отметим, что при $D(\rho) \equiv 1$ в работах [103; 105–106] были найдены первые (логарифмические) члены асимптотических рядов для сдвигов фаз. Недавно Г. М. Хенкиным была высказана гипотеза о том, что следующие члены асимптотических рядов будут ограниченными функциями. Пока, насколько нам известно, не удалось ни доказать, ни опровергнуть эту гипотезу.

Замечание. Можно показать (см. [108–110]), что в теореме 2 скорость сходимости есть²⁴ $O(1/\sqrt[3]{t})$, т.е.

$$\rho(t, x) = \tilde{\rho}\left(t, x; \left\{d_k(t)\right\}_{k=0}^n\right) + O\left(\frac{1}{\sqrt[3]{t}}\right).$$

В теореме 1 (см. [35, 90, 95, 96]) скорость сходимости в равномерной норме (за вычетом сколь угодно малых, но фиксированных на момент рассмотрения окрестностей ударных волн решения) есть $O(1/\sqrt{t})$, причём эта оценка неулучшаема [90] и достигается на асимптотике вида волны разрежения.

Упражнение (см. п. 2.1.3).** При каких условиях уравнение (20) линеаризуется? Попробуйте построить асимптотические ряды для $d_k(t)$ в этих (частных) случаях. Подобно тому, как это было сделано в работе [14] (см. также [7]) для уравнения Бюргерса (в случае общих начальных условий см. [103]).

Замечание. Ввиду замечания, касающегося модели Лайтхилла–Уизема–Кортевега–де Фриза–Бюргерса (см. выше), можно ожидать, что результат, аналогичный теореме 2, имеет место и для задачи Коши (21), (15), если только μ/ε^2 достаточно мало (насколько мало, зависит как от функции $Q(\rho)$, так и от начального условия (15)). Для случая, когда

²⁴ Несложно показать (см. [90]), рассмотрев случай сходимости к одной волне разрежения, что имеет место сходимость не быстрее $\sim 1/\sqrt{t}$. Однако пока в общем случае (при отсутствии прилипания) не доказано (и не опровергнуто), что скорость сходимости в теореме 2 оценивается как $O(1/\sqrt{t})$.

$Q(\rho)$ – вогнутая парабола, локальная сходимость к одному из представителей однопараметрического семейства бегущих волн (параметр – сдвиг фазы, определяется аналогично замечанию Ильина–Олейник) была установлена П. И. Наумкиным и И. А. Шишмарёвым (1991) [139, 140]. Этот результат недавно был перенесен А. В. Казейкиной (при участии А. А. Шананина) на произвольную вогнутую функцию [141]. Для случая, когда $Q(\rho)$ – выпуклая функция, глобальная сходимость к волне разрежения была установлена Р. Дуанем и Х. Чжао (2007) [142].

Замечание. Интересно также заметить, что все сказанное выше относительно модели Уизема в точности (та же структура асимптотики, те же самые скорости бегущих волн, та же асимптотика у *сдвигов фаз* $d_k(t) = O(\ln t)$) переносится и на модель Ньюэлла (подробности см. в [97–107]). Для модели Пэйна получено (см., например, [7]) условие существования автомодельного решения вида «бегущей – ударной волны». Оказывается, что скорость такой волны будет удовлетворять формуле RRH.

Замечание (Т.-П. Лю и др. (1997)). Результаты о сходимости к бегущей волне и волне разрежения решения начальной задачи Коши (20), (15) могут быть «перенесены» на начально-краевую задачу для уравнения (20) (подробности см., например, в [143, 144]).

Приведем в заключение этого пункта схему доказательства теорем 1, 2 и аналогичного утверждения для модели Ньюэлла. Доказательство основывается на следствиях из *принципа максимума* для линейных параболических уравнений (с коэффициентами, зависящими от (t, x)) [37, 38, 136]: на *принципе сравнения*

$$\rho^1(0, x) \leq \rho^2(0, x) \Rightarrow \rho^1(t, x) \leq \rho^2(t, x), \quad t \geq 0$$

и на *принципе сравнения на фазовой плоскости*

$$\{\rho^1(0, x) < (>) \rho^2(0, x) \Rightarrow \rho^1(0, x') \leq (\geq) \rho^2(0, x'), \quad x' \geq (\leq) x\} \Rightarrow$$

$$\Rightarrow \{\rho^1(t, x) < (>) \rho^2(t, x) \Rightarrow \rho^1(t, x') \leq (\geq) \rho^2(t, x'), \quad x' \geq (\leq) x\}, \quad t \geq 0.$$

Эти принципы переносятся и на нелинейные параболические уравнения, а также на некоторые их дифференциально-разностные (и просто разностные) аналоги. Идея такого перенесения достаточно простая (и широко используется в теории параболических уравнений [37, 38], например, для оценки старших производных неизвестной функции) – нелинейные коэффициенты при частных производных объявляются некоторыми функциями независимых переменных (в нашем случае

t и x). Далее делаются «априорные» предположения относительно неизвестной функции (как правило, предположения равномерной ограниченности неизвестной функции или(и) существования у нее (равномерно ограниченных) производных). В этих предположениях коэффициенты при частных производных оказываются «настолько хорошими» функциями, что к возникшему уравнению (в котором нелинейные коэффициенты, зависящие от неизвестной функции, интерпретируются просто как некоторые функции независимых переменных) применимы принципы сравнения. Затем, уже с помощью этих принципов, проверяется, что сделанные априорно предположения выполняются. В этой связи также заметим, что иногда априорные предположения выбирают единственное решение из множества возможных. Так, для обычного уравнения теплопроводности априорное предположение, «что решение начальной задачи Коши (с равномерно ограниченной начальной функцией) будем искать в классе равномерно ограниченных функций», приводит начальную задачу Коши в класс корректных (по Адамару), т.е. имеющих и притом единственное решение, устойчивое по начальным данным. Но, как показывает пример А. Н. Тихонова, если не накладывать ограничение на рост решения с увеличением времени, то построенное решение уже не будет единственным [146].

Приведенные принципы сравнения также переносятся и на некоторые уравнения (например, на закон сохранения), которые получаются путем предельного перехода (например, с помощью метода исчезающей вязкости) из нелинейных параболических уравнений.

В доказательстве часто в качестве сравниваемых функций выбираются: решение рассматриваемой задачи и специальным образом подобранное автомодельное решение (асимптотически автомодельное решение) либо специальным образом «склеенная» функция из таких решений. При этом значения на $x = \pm\infty$ (а также ординаты «склеек», в случае если таковые имелись) у таких автомодельных решений выбираются из множества точек

$$\{\alpha_0, \beta_0, \alpha_1, \dots, \beta_{n-1}, \alpha_n, \beta_n\}$$

либо из малых окрестностей этих точек, а сдвиги фаз подбираются так, чтобы в начальный момент времени выполнялись условия используемого варианта принципа сравнения.

Заметим, что основные идеи описанного подхода к доказательству теорем 1, 2 применялись ранее для исследования асимптотики (по времени) решения начальной задачи Коши для уравнения теплопровод-

ности с нелинейным источником (такого рода уравнения возникают, например, при описании распространения генных волн и пламени). Так, еще в 1937 г. в пионерской работе А. Н. Колмогорова, И. Г. Петровского, Н. С. Пискунова [137] была исследована сходимость к одной бегущей волне. Современное состояние дел отражено в работах [136, 138] и цитированной там литературе.

Следуя монографии [136], можно называть сходимость в теореме 2 *сходимостью по форме*. А согласно терминологии работы [133], системы волн, возникающие в теоремах 1, 2, следует называть *промежуточными асимптотиками* (заметим (см. теорему 2), что промежуточная асимптотика (система волн), в общем случае сама не обязана являться решением уравнения).

2.3.2. Эволюции локального затора в транспортном потоке, описываемом моделями LWR и Уизема

Предположим, что в однородном стационарном транспортном потоке (одна бесконечная полоса без въездов и выездов) образовалось локальное увеличение плотности (локальный затор). Проследим (не строго!) за эволюцией этого «бугорка» (за тем, как будет рассасываться это локальное увеличение плотности). Для этого введем $h(t)$ АТС/км – «высоту» бугорка и $l(t)$ км – «длину» бугорка в момент времени t (см. рис. 21). Поскольку площадь бугорка сохраняется (что как раз и отражает закон сохранения количества АТС), то

$$h(t)l(t) \approx 2S_0,$$

где S_0 – площадь (размерность площади – АТС) под бугорком. Здесь стоит приближенное равенство из-за того, что бугорок, вообще говоря, не будет в точности иметь вид прямоугольного треугольника (гипотенуза, в общем случае, будет кривой линией и лишь на больших временах достаточно хорошо приближается прямой). Сам факт, превращения (с ростом времени) произвольного бугорка в криволинейный прямоугольный треугольник (так же, как на рис. 7, считаем, что $Q''(\rho_0) < 0$) несложно устанавливается с помощью метода характеристик.

Перейдем теперь в систему координат, движущуюся со скоростью

$$c_0 = Q'(\rho_0),$$

где ρ_0 – значение плотности вдали от бугорка. В этой системе координат «скорость разбегания бугорка» можно посчитать по формуле RRH:

$$l'(t) = \left| \frac{Q(\tilde{\rho}(t)) - Q(\rho_0)}{\tilde{\rho}(t) - \rho_0} - c_0 \right| \approx \frac{1}{2} |Q''(\rho_0)| h(t),$$

где

$$\tilde{\rho}(t) = \rho_0 + h(t)$$

– максимальное значение плотности в момент времени t . Таким образом,

$$l'(t) \sim \frac{|Q''(\rho_0)| S_0}{l(t)} \Rightarrow l(t) \sim \sqrt{2|Q''(\rho_0)| S_0} \cdot \sqrt{t} \Rightarrow h(t) \sim \sqrt{\frac{2S_0}{|Q''(\rho_0)|}} \frac{1}{\sqrt{t}}.$$

Сделанные качественные выводы неплохо согласуются с практикой.

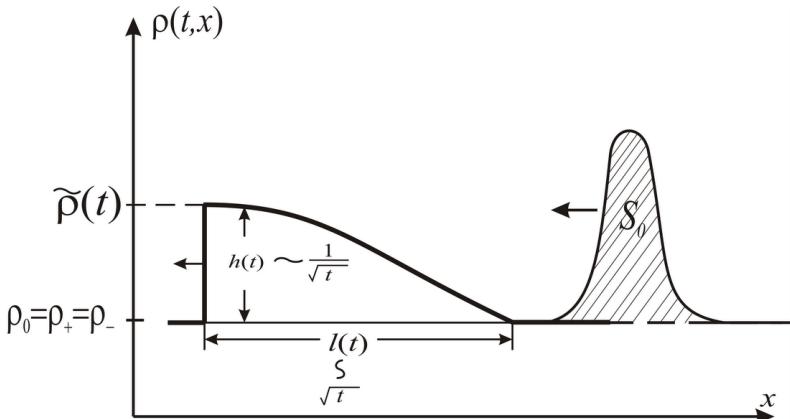


Рис. 21. Эволюция локального затора

Приведенные выше рассуждения являются частным случаем более общего и более точного утверждения «об асимптотике вида N -волн» (см. работы П. Лакса, Т.-П. Лю, Р. Ди Перна, К. Дафтермоса, С. Н. Кружкова и Н. С. Петросян [2, 7, 11, 35, 93, 96]).

Упражнение (см. [35, 42, 96]). Пусть, как и в приведенных выше рассуждениях,

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

Используя формулу Лакса–Олейник (см. п. 2.1.3), строго обоснуйте приведенный выше результат. Выкладки будут особенно простыми, если дополнительно известно, что зависимость $Q(\rho)$ – параболическая.

Для того чтобы описать поведение локального затора согласно модели Уизема, умножим уравнение (20) на

$$\rho(t, x) - \rho_0 \in L_1(\mathbb{R}_x) \cap L_\infty(\mathbb{R}_x)$$

и проинтегрируем (по частям)²⁵ по x от $-\infty$ до ∞ :

$$\begin{aligned} \frac{d}{dt} \int_{-\infty}^{\infty} \frac{1}{2} (\rho(t, x) - \rho_0)^2 dx &= - \int_{-\infty}^{\infty} \left(D(\rho(t, x)) \frac{\partial \rho(t, x)}{\partial x} \right)^2 dx \leq \\ &\leq -D_{\min} \int_{-\infty}^{\infty} \left(\frac{\partial \rho(t, x)}{\partial x} \right)^2 dx < 0, \end{aligned}$$

где

$$D_{\min} = \min_{\rho \in [0, \rho_{\max}]} D^2(\rho) > 0.$$

Полученное неравенство, которое обычно называют *энергетическим неравенством* [19, 107], отражает тот факт, что процесс эволюции локального затора происходит с диссилиацией «энергии». Иначе говоря, «энергия»

$$V(\rho(t, \cdot)) = \int_{-\infty}^{\infty} \frac{1}{2} (\rho(t, x) - \rho_0)^2 dx$$

есть *функционал Ляпунова* [135] для уравнения (20) на многообразии таких решений $\rho(t, x)$, что $\rho(t, x) - \rho_0 \in L_1(\mathbb{R}_x) \cap L_\infty(\mathbb{R}_x)$. Действительно,

$$\begin{aligned} \frac{d}{dt} V(\rho(t, \cdot)) &< 0 \text{ при } \rho(t, x) \stackrel{L_2(\mathbb{R}_x)}{\neq} \rho_0 \text{ и } \frac{d}{dt} V(\rho_0) = 0; \\ V(\rho(t, \cdot)) &> 0 \text{ при } \rho(t, x) \stackrel{L_2(\mathbb{R}_x)}{\neq} \rho_0 \text{ и } V(\rho_0) = 0. \end{aligned}$$

Кроме того, из энергетического неравенства и из неравенства Харди–Литлвуда–Поля (частный случай неравенства для производных колмогоровского типа см. замечание ниже).

²⁵ Заметим, что при интегрировании по частям мы пользовались различными следствиями из принципа максимума (см. пп. 3.1.3, 3.3.1). В частности, следующими двумя:

$$\bullet \quad \lim_{x \rightarrow \pm\infty} \rho(0, x) = \rho_{\pm} \Rightarrow \forall t \geq 0 \rightarrow \lim_{x \rightarrow \pm\infty} \rho(t, x) = \rho_{\pm};$$

- частные производные $\rho(t, x)$, входящие в уравнение (20), равномерно ограничены.

Причём последнее свойство, как правило, используется вместе с неравенствами для производных колмогоровского типа (см. замечание ниже).

городовского типа с $n = 2$, $k = 1$, $p = q = r = 2$ (см. замечание ниже)) следует существование такой константы $\tilde{K} > 0$, что

$$\begin{aligned} \frac{dV(\rho(t, \cdot))}{dt} &\leq -\tilde{K} V(\rho(t, \cdot))^2 \Rightarrow \\ \Rightarrow \|\rho(t, \cdot) - \rho_0\|_{L_2(\mathbb{R})} &= \sqrt{2V(\rho(t, \cdot))} = O\left(\frac{1}{\sqrt{t}}\right). \end{aligned}$$

Немного более аккуратные рассуждения (см, например, [107, 110]) позволяют установить также сходимость в $L_1(\mathbb{R})$ (сходимость в среднем)²⁶ и в $L_\infty(\mathbb{R})$ (равномерную сходимость), причём так же, как и для модели LWR:

$$\|\rho(t, \cdot) - \rho_0\|_{L_\infty(\mathbb{R})} = O\left(\frac{1}{\sqrt{t}}\right).$$

Замечание. Неравенствами для производных колмогоровского типа будем называть неравенства следующего вида:²⁷

²⁶ Из сходимости в среднем следует, что площадь бугорка (см. рис. 21) со временем уменьшается и стремится к нулю.

²⁷ Такого рода неравенства возникали в исследованиях А. Н. Колмогорова конца 30-х годов XX века. Для отыскания не улучшаемых значений K приходилось решать, например, задачи вариационного исчисления, оптимального управления (в том числе и с фазовыми ограничениями). В то время как из приложений (в основном военных, и связанных с движением ракет) задачи оптимального управления активно стали приходить в 40-е годы прошлого века. Общие же способы решения задач оптимального управления (без смешанных ограничений) появились только в середине 1950-х годов в школах Л. С. Понтрягина (принцип максимума Понтрягина) и Р. Беллмана (принцип динамического программирования). В цикле работ, начавшихся в 1960-е годы, А. Я. Дубовицкий и А. А. Миллопин предложили наиболее общую (из известных на данный момент) схему получения необходимых условий экстремума [39, 47], которая в задачах оптимального управления (в том числе и со смешанными ограничениями) позволяет получить принцип максимума (не путать с принципами максимума для решений уравнений параболического типа). В основе схемы лежит простая идея: пересечение конуса направлений убывания, если задача на минимум, и возрастания, если на максимум, функционала с конусами возможных (допустимых) направлений ограничений должно быть пустым в точке оптимума. В предположении выпуклости рассматриваемых, конусов это условие представляется как *уравнение Эйлера–Лагранжа*: существует нетривиальный набор элементов из сопряженных конусов, сумма которых равняется нулю. Другой подход (см., например, [45]) состоит в использовании «гладко-выпуклой» структуры задачи. При этом для «борьбы» с гладкой частью задачи (для получения необходимых условий локальной оптимальности) используется принцип Ферма (оптимум следует искать среди экстремальных значений функционала), а для «борьбы» с выпуклой частью – *принцип Лагранжа* (возможность «перенесения» ограничений (необязательно всех) задачи оптимизации в функционал с должным образом определенными множителями Лагранжа). Для того чтобы была понятна связь принципа Лагранжа с принципом Ферма, можно представить задачу оптимизации в виде

$$\exists K > 0 : \forall z(x) \in W_{p,r}^n(\mathbb{R}) \rightarrow \|z^{(k)}(\cdot)\|_{L_q(\mathbb{R})} \leq K \|z(\cdot)\|_{L_p(\mathbb{R})}^\alpha \|z^{(n)}(\cdot)\|_{L_r(\mathbb{R})}^\beta,$$

где $0 \leq k < n$ – целые числа, $0 < p, q, r \leq \infty$, $\alpha, \beta \geq 0$, $W_{p,r}^n(\mathbb{R})$ – пространство таких функций $z(x) \in L_p(\mathbb{R})$, у которых $(n-1)$ -я производная локально абсолютно непрерывна на \mathbb{R} и n -я производная $z^{(n)}(x) \in L_r(\mathbb{R})$. Константа K зависит от пяти параметров n, k, p, q, r . Величины α и β однозначно ими определяются:

$$\alpha = \frac{n-k-1/r+1/q}{n-1/r+1/p}, \quad \beta = 1-\alpha.$$

В работах В. Н. Габушина [147, 148] было показано, что если $q \neq p$ при $k=0$, то

$$K(n, k, p, q, r) < \infty \Leftrightarrow (n-k)/p + k/r \geq n/q, \quad r \geq 1.$$

Более подробно о тех случаях, в которых удалось вычислить наилучшие (точные) значения константы K , написано, например, в работах [45, 149].

В п. 2.1.3 обобщенные энтропийные решения уравнения LWR получались как пределы решений соответствующих уравнений Уизема. Поэтому можно ожидать, что и на предельных решениях уравнений Уизема энергия также убывает. И действительно, если предельное (энтропийное) решение имеет разрывы (и только в этом случае), то на каждом

Гранжа с выпуклостью задачи, заметим, что этот принцип есть, по сути, теорема об отдельности (одно из следствий теоремы Хана–Банаха): в вещественном векторном пространстве граничные точки «телесного» (имеющего внутренние точки) выпуклого множества отделимы гиперплоскостью от самого множества, т.е. найдется такая гиперплоскость, проходящая через выбранную граничную точку, что рассматриваемое выпуклое множество будет «лежать» в одном из двух полупространств, на которые гиперплоскость делит пространство. Множители Лагранжа как раз и задают уравнение этой гиперплоскости. При этом выпуклое множество строится по самой задаче (функционалу и каждому ограничению, которое хотим «внести» в функционал, соответствуют свои «компоненты»), а точка, которую нужно отделять, соответствует решению задачи оптимизации. Как следствие, множители Лагранжа являются элементами (алгебраически) сопряженных пространств к пространствам, в которых «живут» ограничения. Элементы сопряженного пространства для некоторых важных пространств (например, таких, как пространство измеримых ограниченных функций) довольно трудно описать. Это является, пожалуй, одним из основных сдерживающих факторов в развитии необходимых условий оптимальности. Обратим внимание на то, что для получения принципов максимума в задачах оптимального управления со смешанными ограничениями очень сложно (если вообще возможно) напрямую использовать описанный выше гладко-выпуклый формализм. В основном при получении необходимых условий оптимальности в таких задачах удобно использовать схему Дубовицкого–Милнотина.

разрыве (ударной волне) происходит потеря энергии в соответствии с формулой (для определенности $\rho_- < \rho_+$):

$$\frac{d}{dt}V(\rho(t, \cdot)) = \frac{Q(\rho_+) + Q(\rho_-)}{2}(\rho_+ - \rho_-) - \int_{\rho_-}^{\rho_+} Q(\rho(t, x)) dx < 0.$$

Этот интересный и довольно простой факт доказан, например, в пособии [19].

Таким образом, мы получили еще одно объяснение необратимости по времени процесса, описываемого моделью LWR (при наличии ударных волн в решении) как свойство, «унаследованное» от модели более высокого уровня – модели Уизема.

Замечание (С. Н. Кружков, Е. Ю. Панов (1969, 1989, 1994)). Пусть $\Phi(\rho)$ – произвольная дважды гладкая выпуклая функция, $f(t, x) \geq 0$ – произвольная дважды гладкая финитная в полу平面ости $\pi = \{(t, x) : t > 0, x \in \mathbb{R}\}$ пробная функция. Умножим теперь уравнение (20) на $\Phi'(\rho(t, x))f(t, x)$ и проинтегрируем в π , перебрасывая производные на пробную функцию (для упрощения выкладок будем считать $D(\rho) \equiv 1$ (в общем случае см., например, [20, 35])):

$$0 \leq \varepsilon \iint_{\pi} \Phi''(\rho(t, x)) \rho_x(t, x)^2 f(t, x) dt dx = \iint_{\pi} \left\{ f_t(t, x) \int_k^{\rho} \Phi'(\tilde{\rho}) d\tilde{\rho} + f_x(t, x) \int_k^{\rho} \Phi'(\tilde{\rho}) Q'(\tilde{\rho}) d\tilde{\rho} + \varepsilon f_{xx}(t, x) \Phi(\rho(t, x)) \right\} dt dx,$$

где k – произвольное действительное число. Устремив $\varepsilon \rightarrow 0+$, получим

$$\iint_{\pi} \left\{ f_t(t, x) \int_k^{\rho} \Phi'(\tilde{\rho}) d\tilde{\rho} + f_x(t, x) \int_k^{\rho} \Phi'(\tilde{\rho}) Q'(\tilde{\rho}) d\tilde{\rho} \right\} dt dx \geq 0,$$

которое справедливо для любого k . Заметим, что в случае, когда $\Phi(\rho)$ – линейная функция, вместо неравенства можно писать равенство. Введём обозначения:

$$\eta(\rho) = \int_k^{\rho} \Phi'(\tilde{\rho}) d\tilde{\rho} – \text{энтропия},$$

$$q(\rho) = \int_k^{\rho} \Phi'(\tilde{\rho}) Q'(\tilde{\rho}) d\tilde{\rho} \quad \text{– поток энтропии.}$$

Полученное неравенство можно переписать как *энтропийное неравенство* (перебрасывая обратно производные с пробной функции) [29, 35]:

$$\frac{\partial \eta(\rho)}{\partial t} + \frac{\partial q(\rho)}{\partial x} \leq 0,$$

которое следует понимать в слабом смысле (причём с неотрицательной финитной пробной функцией). Энтропийные неравенства для уравнений газовой динамики, по-видимому, впервые рассматривал Э. Жуге в начале XX века (см., например, [29, 54]). Напомним (см. примеры О. А. Олейник и И. М. Гельфанд из п. 2.1.1), что начальная задача Коши для уравнения (13), понимаемого в слабом смысле, имеет, вообще говоря, не единственное решение и что закон сохранения (13) и закон сохранения (13), умноженный, скажем, на $\Phi'(\rho) > 0$, вообще говоря, имеют разные решения. Однако было подмечено [150], что если к закону сохранения (13) добавить энтропийное неравенство²⁸, то оно может отобрать единственное решение. В конце 1960-х С. Н. Кружков с помощью энтропийных неравенств построил (по сути, используя метод исчезающей вязкости (см. п. 2.1.3)) вполне законченную теорию обобщенных решений начальной задачи Коши для закона сохранения (13) [27].²⁹ В начале

²⁸ Функция, удовлетворяющая (в слабом смысле) уравнению (13) и энтропийному неравенству, полученному исходя из функции $\Phi(\rho)$, называют Φ -решением.

²⁹ Основная идея заключалась в следующем. Назовём энтропийным решением функцию, которая является Φ -решением, для любой дважды гладкой выпуклой функции $\Phi(\rho)$. По построению, решение, полученное с помощью метода исчезающей вязкости, необходимо является энтропийным решением. Причём в качестве функций $\Phi(\rho)$ можно брать лишь всевозможные линейные функции и функции вида $\{|\rho - k|\}_{k \in \mathbb{R}}$, поскольку любая дважды гладкая выпуклая функция раскладывается по этому «базису». Оказывается, что энтропийное решение всегда единственно. Для того чтобы это понять, заметим, что энтропийное решение в точках гладкости удовлетворяет соотношению (13) в классическом смысле, а в точках разрыва удовлетворяет RRH-условию и E-условию (простое доказательство этих фактов имеется, например, в пособии [19]). В конце 1980-х, в контексте вышенаписанного, С. Н. Кружковым был поставлен вопрос [151]: когда Φ -решение единственны? Ответ на этот вопрос был получен Е. Ю. Пановым в работе [152]. Оказывается, что в случае, когда $Q(\rho)$ – строго вогнутая (выпуклая) функция, Φ -решение единственны ($\Phi(\rho)$ – произвольная дважды гладкая строго выпуклая функция). Если $Q(\rho)$ не является строго вогнутой (выпуклой) функцией, то Φ -решение задачи Коши при подходящем выборе начальных данных не единственны [152].

1970-х годов исследованием энтропийных неравенств (для систем уравнений) занимался П. Лакс [11, 29, 35].

2.3.3. Задача о светофоре (при каких условиях перед светофором не будет скапливаться очередь)

В 1955 г. в работе [5] (см. также [7]) М. Лайтхиллом и Дж. Уиземом была поставлена и решена (на основе модели LWR) следующая задача:

Найти такое число $k > 0$, что перед светофором (работающим в двух режимах: зеленый и красный) не будет скапливаться очередь, если

$$T_{\text{зел}} / T_{\text{кр}} \geq k.$$

Считать, что транспортный поток вдали от светофора имеет плотность $\rho_i < \rho_m$ (значение потока q_i), где ρ_m – плотность, при которой значение потока максимально.

Пусть загорелся красный цвет (рис. 22). Тогда от светофора на встречу транспортному потоку пойдет ударная волна со скоростью³⁰

$$c_{\text{кр}} = \left| \frac{Q(\rho_{\max}) - Q(\rho_i)}{\rho_{\max} - \rho_i} \right| = \frac{q_i}{\rho_{\max} - \rho_i}.$$

«Излишек» АТС, скопившихся перед светофором за время горения красного цвета, равен

$$(\rho_{\max} - \rho_i) \frac{q_i}{\rho_{\max} - \rho_i} T_{\text{кр}} = q_i T_{\text{кр}}.$$

Пусть теперь загорелся зеленый цвет (рис. 23, 24). Тогда до тех пор, пока весь излишек не пройдет через светофор, поток АТС через светофор будет максимальным и равным q_m . Это не очень очевидное утверждение может быть установлено с помощью решения соответствующей задачи о распаде произвольного разрыва (см. п. 2.3.1). Таким образом, перед светофором не будет скапливаться очередь, если

$$(q_m - q_i) T_{\text{зел}} \geq q_i T_{\text{кр}} \Rightarrow k = \frac{q_i}{q_m - q_i}.$$

³⁰ Возникшую краевую задачу удобно понимать, как начальную, если договориться считать, что за светофором плотность АТС максимальная ρ_{\max} , т.е. за светофором движения нет.

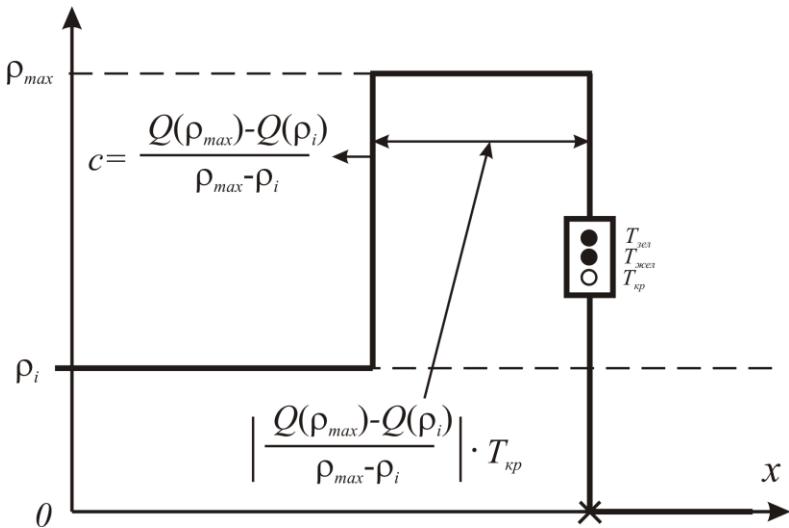


Рис. 22. Момент окончания горения красного цвета

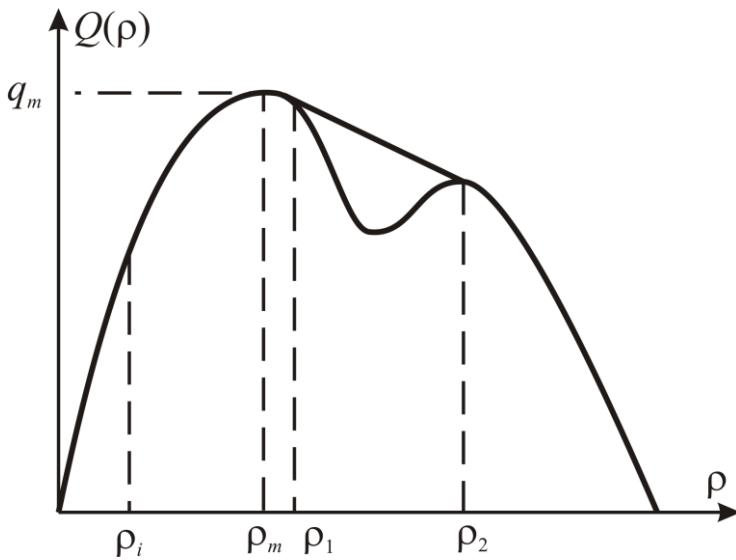


Рис. 23

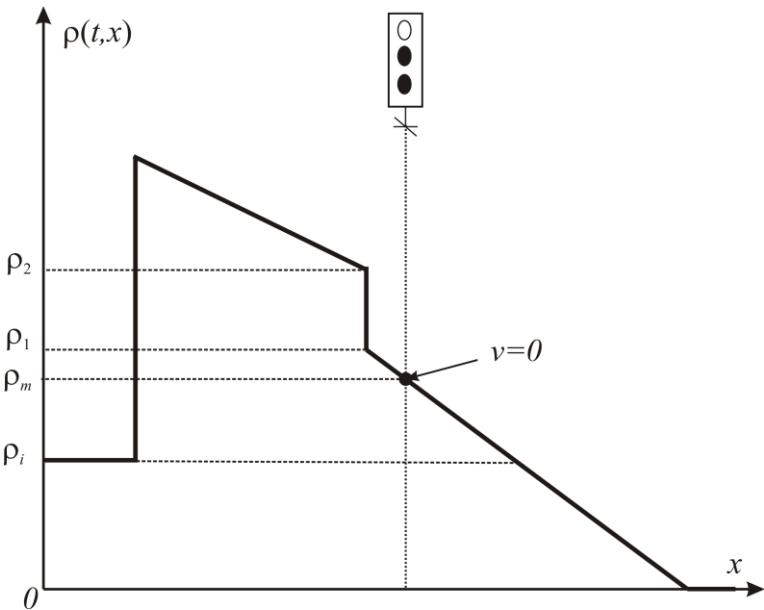


Рис. 24. Горит зеленый цвет

Заметим, что полученное соотношение достаточно хорошо согла-
ется с интуитивными представлениями. Действительно, если принять,
что когда горит красный цвет, тогда поток АТС через светофор равен
нулю, а когда горит зеленый, тогда поток максимальный (в течение все-
го времени горения зеленого), то получим условие: излишек АТС, скопившийся за время горения красного $q_i T_{kp}$, должен быть не больше, чем
так «добавка», которую получаем за время горения зеленого:

$$(q_m - q_i)T_{zel}.$$

«Добавка» же в свою очередь обусловлена тем, что при наличии свето-
фора интенсивность потока АТС через светофор во время горения зеле-
ного q_m превышает интенсивность потока АТС, подъезжающих к све-
тофору q_i .

2.4. ТЕОРИЯ КЕРНЕРА–КОНХОЙЗЕРА ДВИЖУЩИХСЯ ЛОКАЛЬНЫХ КЛАСТЕРОВ В МОДЕЛЯХ ДЖЕНЕРАЛ МОТОРС КЛАССА³¹

В пунктах 2.1–2.3 были рассмотрены в основном модели транспортного потока, в которых стационарные состояния (в стационарных состояниях АТС движутся с постоянной скоростью и плотностью) «отвечают» фундаментальной диаграмме транспортного потока. С точки зрения образования пространственно-временных структур плотного потока детальная классификация этих моделей была дана, например, в главе 10 книги [10]. Там же имеется и критическое сравнение этих модельных решений с фундаментальными эмпирическими свойствами перехода к плотному транспортному потоку [153]. В этом пункте мы рассматриваем нелинейное решение, которое возникает в результате неустойчивости в моделях Дженерал Моторс (ДМ) класса. Чтобы понять термин *нелинейное решение*, нужно напомнить сначала смысл термина *неустойчивость* исходно однородного состояния транспортного потока в этих моделях. Неустойчивость означает нарастание во времени очень малых неоднородных возмущений. Окончательный результат этого нарастания приводит к структурам транспортного потока конечной и в некоторых случаях очень большой амплитуды. В последнем случае для нахождения и математического описания этих нелинейных решений уже нельзя пользоваться математическим аппаратом анализа неустойчивостей модели, который для множества моделей ДМ-класса детально рассмотрен в обзорах [58, 119]. Именно рассмотрение нелинейных пространственно-временных решений большой амплитуды, впервые найденных в 1994 г. Б. С. Кернером и П. Конхойзером [157, 158] в моделях ДМ-класса, и посвящен этот пункт данной главы. Однако прежде чем мы рассмотрим эти нелинейные решения в п. 2.4.2, необходимо коротко изложить фундаментальные эмпирические свойства перехода от свободного к плотному транспортному потоку и выяснить, могут ли модели транспортного потока, рассмотренные выше, показать эти эмпирические свойства. Более подробно об этом будет написано в главе 3.

³¹ Этот пункт написан С. Л. Клёновым. Автор выражает благодарность Б. С. Кернеру за предоставление оригинальных рисунков из его книг.

2.4.1. Фундаментальные эмпирические свойства перехода от свободного транспортного потока к плотному и модели транспортного потока

Основные эмпирические свойства перехода к плотному транспортному потоку следующие [153]:

1. Переход к плотному транспортному потоку (traffic breakdown) является $F \rightarrow S$ переходом (буква F соответствует «free flow», т.е. свободному потоку, буква S обозначает фазу синхронизованного потока, в английской литературе «synchronized flow»).

2. Вероятность спонтанного $F \rightarrow S$ перехода является растущей функцией величины потока АТС.

3. Может быть как спонтанный, так и индуцированный (т.е. вызванный внешним возмущением большой амплитуды) $F \rightarrow S$ переход около одного и того же узкого места на дороге (bottleneck в англоязычной литературе).

Как показано в главе 10 книги [10], с точки зрения перехода к плотному транспортному потоку многие модели пп. 2.1–2.3 можно разделить на два больших класса:

(а) *Модели типа Лайтхилла–Уизема–Ричардса (LWR)*, в которых переход к плотному транспортному потоку возникает не в результате неустойчивости, а за счет существования точки, в которой достигается максимум функции потока на фундаментальной диаграмме.

(б) *Модели Дженерал Моторс класса (ДМ)*, в которых переход к плотному потоку связан с неустойчивостью модельных решений начиная с некоторой критической плотности транспортного потока.

Модели LWR-типа (теория кинематических и ударных волн в транспортном потоке) не могут описывать пункты 2 и 3 фундаментальных эмпирических свойств перехода к плотному транспортному потоку. Теория, основанная на моделях типа ДМ, не может описывать никаких фундаментальных эмпирических свойств перехода к плотному транспортному потоку (пункты 1–3). Тем не менее эти модели представляют большой интерес для анализа. Одна из причин такого вывода рассматривается в этом пункте, в котором показано, что модели ДМ-класса описывают характеристические параметры стационарного движения широкого движущегося кластера по дороге, наблюдаемого в эмпирических данных.

Неустойчивость однородного свободного транспортного потока в моделях ДМ-класса³² детально рассмотрена в огромном количестве работ, обзор которых можно найти в статьях Д. Кроудери с соавторами [119] и Д. Хельбинга [58]. Однако нарастание неоднородных возмущений в этих моделях, происходящее в результате упомянутой неустойчивости, приводило к волнам нереалистично большой амплитуды (в макроскопических моделях) или к столкновению АТС (в микроскопических моделях).

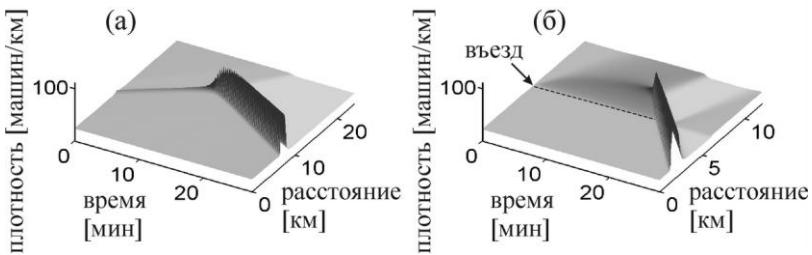


Рис. 25. Эффект «бумеранга» для нарастающего возмущения в свободном потоке, метастабильном относительно возникновения широких движущихся кластеров. Плотность АТС в пространстве и времени. Расчет на основе макроскопической модели Пэйн-класса (22), (23). (а) Однородная автодорога, (б) автодорога с въездом. Взято из [10]

Какие же типы структур плотного транспортного потока возникают в результате такой неустойчивости в этих и других моделях ДМ-класса? Ответ на этот вопрос был дан Б. С. Кернером и П. Конхайзером в 1993–1994 годах [157, 158] на основе численного расчета (рис. 25 и 26) и версии модели Пэйна (см. п. 2.1.4):

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0, \quad (22)$$

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \frac{V(\rho) - v}{\tau} - \frac{c_0^2}{\rho} \frac{\partial \rho}{\partial x} + \frac{\mu}{\rho} \frac{\partial^2 v}{\partial x^2}, \quad (23)$$

³² К моделям этого класса относятся, например, модель Пэйна [53], оптимальной скорости Ньюэлла [88], Бандо [154], разумного водителя Трайбера [58, 115], клеточных автоматов Нагеля–Шрекенберга [118, 155], модели Видемана [156] и многие другие модели [10] (см. также главу 2).

где τ , c_0 и μ являются константами. По сути, по сравнению с обычной моделью Пэйна произошло всего одно изменение: добавление диффузионного слагаемого $(\mu/\rho)\partial^2 v/\partial x^2$ в правую часть уравнения (23).

Чтобы найти реалистичные решения большой амплитуды, возникающие в результате неустойчивости, Б. С. Кернер и П. Конхойзер использовали при расчетах в модели типа Пэйна специальную форму фундаментальной диаграммы (рис. 26 (а)). Особенность этой фундаментальной диаграммы состоит в том, что, начиная уже с не очень больших плотностей, скорость АТС на фундаментальной диаграмме экспоненциально стремилась к нулю. Это искусственная форма фундаментальной диаграммы, с помощью которой удалось в рамках макроскопической модели смоделировать задержку водителей $\tau_{del,jam}^{(a)}$ при их ускорении один за другим из состояния с нулевой скоростью на заднем фронте широкого движущегося кластера (см. ниже). Эта задержка отличается от задержки реакции водителя в других ситуациях, которая дается величиной τ в формуле (23). Идея Кернера–Конхойзера моделирования задержки при ускорении из состояния с нулевой скоростью с помощью формы фундаментальной диаграммы математически равносильна правилу slow-to-start в микроскопическом моделировании (см. работы группы М. Шрекенберга в 1998 г. [155]).

В статьях [157, 158] было получено, что в результате неустойчивости в моделях ДМ-класса образуется широкий движущийся кластер большой амплитуды (локальный движущийся затор), внутри которого плотность АТС высока, а скорость их движения близка к нулю, в то время как впереди и позади такого кластера существует свободный поток малой плотности. В англоязычной литературе для обозначения такого рода неоднородных локальных состояний большой плотности используется название «wide moving jam» (в дальнейшем широкий движущийся кластер большой амплитуды будет обозначаться буквой J). Могут возникать как отдельные движущиеся кластеры, так и последовательности таких кластеров (рис. 25 и 26).

Фазовый переход в моделях с неустойчивостью однородного состояния, по-видимому, является фазовым переходом первого рода от свободного однородного потока к широкому движущемуся кластеру (jam) и обозначается как F→J переход (напомним, что буква F соответствует «free flow», т.е. свободному потоку, а буква J соответствует «wide moving jam»). Последующие исследования показали (см. ссылки в обзоре Д. Хельбинга [58]), что этот результат является общим для всех моделей ДМ-класса, рассмотренных выше в пп. 2.1.3, 2.1.4, 2.2.

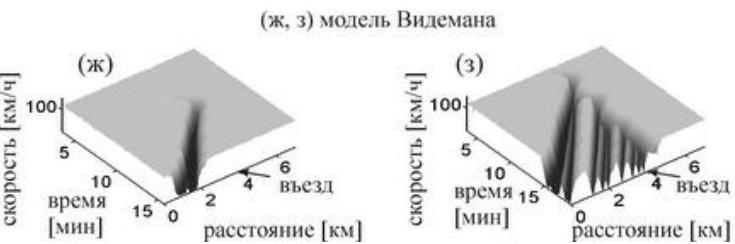
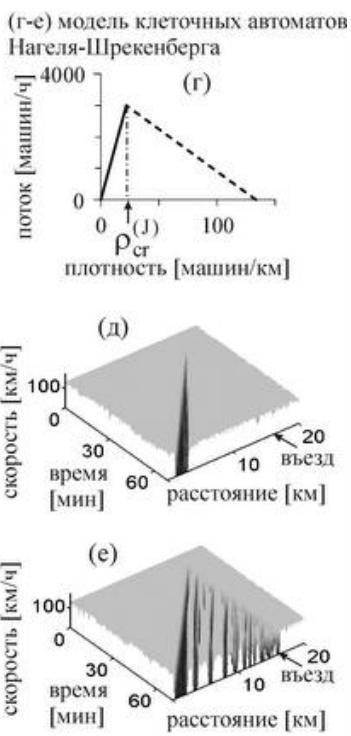
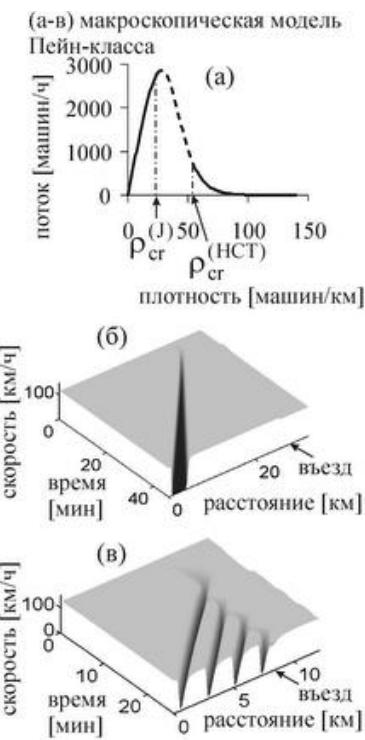


Рис. 26. Возникновение плотного потока на автодороге с въездом в моделях ДМ-класса, показывающих неустойчивость свободного потока при повышении плотности: возникновение широких движущихся кластеров в макроскопической модели (10), (11) Пэйн-класса (а–в), в модели клеточных автоматов Нагеля–Шрекенберга (г–е), и в модели Видемана (ж–з). (а, г) Фундаментальная диаграмма в макроскопической модели Пэйн-класса (а) и в модели Нагеля–Шрекенберга (г); пунктирная часть диаграмм отвечает неустойчивым состояниям. (б, в, д–з) Средняя скорость АТС в пространстве и времени. Взято из [10]

Позже Б. С. Кернеру стало ясно, что этот результат,³³ относящийся ко всем моделям ДМ-класса, противоречит фундаментальным эмпирическим свойствам, перечисленным в пунктах 1–3 выше (см. также главу 3 о теории трех фаз Кернера). Как уже было указано в пункте 1, переход от свободного к плотному транспортному потоку в эмпирических данных связан не с $F \rightarrow J$, а с $F \rightarrow S$ переходом (напомним, что буква S обозначает фазу синхронизованного потока, в английской литературе «synchronized flow»).

Хотя неустойчивость свободного потока в моделях ДМ-класса неправильно описывает переход от свободного к плотному потоку в реальном транспортном потоке, тем не менее результат нелинейного решения моделей ДМ-класса [158] (рис. 25 и 26) – стационарное движение широкого кластера по дороге – полностью соответствует эмпирическим данным и, следовательно, является важным результатом этих моделей, который остается также и в теории трех фаз Кернера (см. главу 3). Как объяснено выше, под результатом нелинейного решения здесь понимается конечное пространственно-временное распределение параметров потока, возникающее в результате неустойчивости исходно однородного решения модели транспортного потока.

Стационарное движение кластеров обладает определенными нелинейными свойствами, которые Б. С. Кернер и П. Конхойзер назвали характеристическими свойствами движения широких кластеров [158]. В дальнейшем эти характеристические свойства движения широкого кластера были подтверждены при исследовании всех других моделей класса ДМ [58]. Эти характеристические свойства движения широкого кластера рассмотрены в следующем пункте.

2.4.2. Характеристические параметры широкого движущегося кластера

Свойства стационарного движения широких кластеров состоят в следующем [158]: существуют характеристические параметры широкого движущегося кластера, которые при заданных внешних условиях движения транспорта (погода, день недели, процент грузовых АТС и т.п.) не зависят от параметров потока впереди широкого кластера и остаются неизменными в процессе движения кластера по дороге. Характеристические параметры являются одинаковыми для различных широ-

³³ Состоящий в том, что переход от свободного к плотному потоку связан с $F \rightarrow J$ переходом.

ких движущихся кластеров. Такими характеристическими параметрами широкого движущегося кластера являются:

- 1) Средняя скорость движения заднего (по направлению потока) фронта движущегося кластера, обозначаемая как v_g .
- 2) Величина потока q_{out} , плотность ρ_{min} и средняя скорость АТС v_{max} в выходном потоке из широкого движущегося кластера. Эти величины являются характеристическими параметрами только при условии, что выходной поток из широкого движущегося кластера отвечает свободному потоку.
- 3) Средняя плотность АТС внутри широкого движущегося кластера обозначается как ρ_{max} . Здесь и далее ρ_{max} совсем не обязательно совпадает с максимально возможной плотностью («бампер к бамперу»).

Характеристические параметры широкого движущегося кластера качественно проиллюстрированы на рис. 27, на котором для заданного момента времени приведены распределения скорости, потока и плотности вдоль дороги, связанные с распространением широкого движущегося кластера в исходно однородном транспортном потоке. Поток q_h и плотность ρ_h в исходно однородном свободном потоке выбраны больше, а соответственно скорость v_h меньше, чем соответствующие характеристические параметры в выходном потоке широкого движущегося кластера: $q_h > q_{out}$, $\rho_h > \rho_{min}$, $v_h < v_{max}$. Ясно, что впереди от широкого движущегося кластера остается исходный однородный транспортный поток. Однако этого не происходит позади широкого движущегося кластера из-за того, что по мере движения широкого кластера АТС, покидающие задний (в направлении движения) фронт кластера, формируют новый свободный поток с величиной потока q_{out} , плотностью ρ_{min} и средней скоростью v_{max} .

Чтобы пояснить термин *выходной поток из широкого движущегося кластера*, давайте более детально посмотрим на рис. 27. Внутри широкого кластера скорость АТС равна нулю, а плотность равна ρ_{max} . На переднем фронте широкого кластера АТС должны резко тормозить вплоть до их остановки. На заднем фронте широкого кластера АТС ускоряются. В результате ускорения АТС из широкого кластера на его заднем фронте образуется выходной транспортный поток. Это объясняет термин *выходной поток из широкого кластера*, использованный вы-

ше. В случае, когда этот поток отвечает свободному потоку, он обозначен буквой q_{out} и показан на рис. 27.

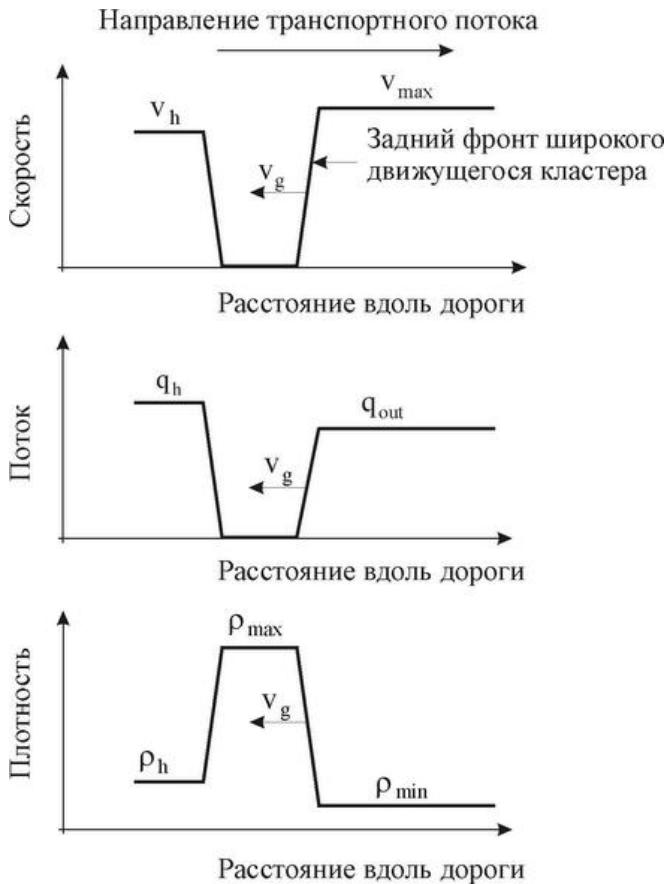


Рис. 27. Качественная иллюстрация характеристических параметров широкого движущегося кластера. Схематическое представление кластера в фиксированный момент времени. Пространственные распределения скорости АТС v , потока q , и плотности ρ в широком движущемся кластере, который распространяется в исходно однородном свободном потоке, имеющем скорость v_h , величину

потока q_h , и плотность ρ_h . Взято из [10]

Кроме многочисленных численных исследований характеристического движения широких кластеров, опубликованных в огромном количестве работ различных авторов, существует асимптотическая теория движущихся широких кластеров [159], основанная на математической теории сингулярных возмущений. Заинтересованный читатель может найти обзор численных исследований широких кластеров в обзоре [58], а асимптотическая теория широких кластеров подробно разбирается в оригинальной статье [159]. В остальной части этого пункта будет рассмотрено одно из важнейших характеристических свойств движущегося широкого кластера – линия J.

2.4.3 Линия J Кернера

Характеристические параметры широкого движущегося кластера могут быть проиллюстрированы линией на плоскости поток–плотность (рис. 28). Эта линия использовалась Б.С. Кернером и была названа им линией J [10, 153].

Необходимо подчеркнуть, что линия J Кернера не имеет никакого отношения к линии для плотного потока на треугольной диаграмме Даганзо 1994 г. [82] (или же к любым другим фундаментальным диаграммам плотного транспортного потока). Действительно, наклон линии J Кернера определяется средней скоростью заднего фронта широкого кластера, а левая координата этой линии отвечает потоку, вытекающему из широкого кластера. Другими словами, линия J – это характеристика не плотного потока, а равномерного распространения заднего фронта широкого кластера. Эта линия фактически соединяет две точки на плоскости поток–плотность: одну точку, отвечающую потоку и плотности внутри широкого движущегося кластера, и вторую точку, отвечающую потоку и плотности в выходном потоке после широкого кластера.

Поясним этот важный вопрос более подробно. Каждая точка на линии для плотного потока в треугольной фундаментальной диаграмме Даганзо [82] отвечает соотношению между плотностью и величиной потока в однородном плотном потоке. В то же время линия J описывает стационарное распространение заднего фронта широкого кластера, а не зависимость потока от плотности, которая отвечает фундаментальной диаграмме плотного потока. Чтобы понять это важное качество линии J и ее отличие от фундаментальной диаграммы плотного потока, можно дополнитель но обратиться к рис. 9 в главе 3. На этом рисунке можно видеть, что плотный поток в теории трех фаз Кернера отвечает не какой-то кривой или линии на плоскости поток–плотность, а двумерной

области. При этом линия J разбивает эту двумерную область на две части. Таким образом, в теории трех фаз Кернера вместо прямой линии треугольной фундаментальной диаграммы Даганзо [82] для плотного потока постулируется двумерная область состояний, т.е. одному значению плотности отвечает не одно значение потока, как в диаграмме Даганзо, а бесконечное количество значений потока. В свою очередь линия J не описывает связь между потоком и плотностью в плотном потоке АТС, а как уже отмечалось, линия J – это характеристика равномерного распространения заднего фронта широкого кластера.

Чтобы объяснить линию J, рассмотрим среднюю скорость движения заднего фронта широкого кластера, где происходит ускорение АТС одного за другим. Каждое АТС, стоящее в кластере, может начать ускоряться на заднем фронте этого кластера, только если выполнены следующие условия:

- предыдущее АТС уже начало двигаться из кластера;
- в результате движения предыдущего АТС спустя некоторое время дистанция между двумя АТС превысила некоторое безопасное расстояние, которое удовлетворяет условию безопасного ускорения.

Таким образом, существует некоторая временная задержка в ускорении АТС на заднем фронте широкого движущегося кластера. Среднее время этой задержки ускорения АТС на заднем фронте широкого движущегося кластера обозначим как $\tau_{del, jam}^{(a)}$. Согласно эмпирическим данным, $\tau_{del, jam}^{(a)} \sim 1.5 - 2$ с. Движение заднего фронта широкого движущегося кластера связано с последовательным ускорением АТС, стоящих внутри кластера, на заднем фронте этого кластера. Поскольку среднее расстояние между АТС внутри кластера, включая среднюю длину АТС, равно $1/\rho_{max}$, средняя скорость заднего фронта широкого движущегося кластера равна

$$v_g = -\frac{1}{\rho_{max} \tau_{del, jam}^{(a)}}. \quad (24)$$

В эмпирических данных характеристическая средняя скорость заднего фронта широкого движущегося кластера по порядку величиныдается формулой $v_g \approx -15$ км/ч. Наклон линии J равен характеристической скорости v_g . В случае, когда свободный транспортный поток формируется после кластера, характеристические величины потока q_{out} и плот-

ности ρ_{\min} определяют левую координату (ρ_{\min}, q_{out}) линии J. Правая координата $(\rho_{\max}, 0)$ линии J отвечает плотности и потоку внутри широкого движущегося кластера; на рис. 28 предполагается, что средняя скорость и, следовательно, поток АТС внутри кластера равны нулю. В результате величина выходного потока q_{out} из широкого движущегося кластера равна

$$q_{out} = |v_g|(\rho_{\max} - \rho_{\min}),$$

или, используя соотношение (24):

$$q_{out} = -\frac{1}{\tau_{del, jam}^{(a)}} \left(1 - \frac{\rho_{\min}}{\rho_{\max}} \right).$$

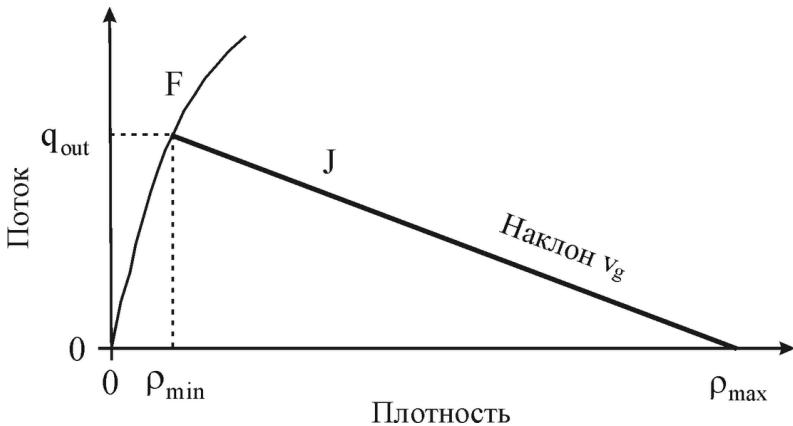


Рис. 28. Качественное представление фундаментальной диаграммы свободного потока (F) вместе с линией J Кернера, наклон которой равен средней скорости движения v_g заднего фронта широкого движущегося кластера. Взято из [10]

Необходимо еще раз подчеркнуть, что, несмотря на важность моделей ДМ-класса, переход от свободного к плотному потоку в этих моделях связан с переходом к широким движущимся кластерам ($F \rightarrow J$ переход). Напротив, во всех реальных данных этот переход связан с переходом к синхронизованному потоку, т.е. с $F \rightarrow S$ переходом [153]. Последний вопрос более подробно разбирается в главе 3 о теории трех фаз Кернера.

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. Крайко А. Н. Краткий курс теоретической газовой динамики. М.: МФТИ, 2007.
4. Гордин В. А. Математика, компьютер, прогноз погоды и другие сценарии математической физики. М.: Физматлит, 2010.
5. Lighthill M. J., Whitham G. B. On kinematic waves: II. Theory of traffic flow on long crowded roads // Proc. R. Soc. London, Ser. A. 1955. V. 229. P. 281–345.
6. Richards P. I. Shock Waves on the Highway // Oper. Res. 1956. V. 4. P. 42–51.
7. Узем Дж. Линейные и нелинейные волны. М.: Мир, 1977.
8. Traffic flow theory: A state-of-the-art report. Editors N.H. Gartner, C. J. Messer, A. K. Rathi. Washington DC: Transportation Research Board, 2001.
9. Луканин В. Н., Буслаев А. П., Трофимов Ю. В., Яшина М. В. Авто-транспортные потоки и окружающая среда. М.: ИНФРА-М, Ч. 1, 2. 1998, 2001.
10. Kerner B. S. Introduction to modern traffic flow theory and control. The long road to three – phase traffic theory. Springer, 2009.
11. Лакс П. Д. Гиперболические дифференциальные уравнения в частных производных. М.–Ижевск: НИЦ «РХД», ИКИ, 2010.
12. Ballou D. P. Solution to nonlinear hyperbolic Cauchy problems without convexity condition // Trans. Amer. Math. Soc. 1970. V. 152. № 2. P. 441–460.
13. Олейник О. А. Разрывные решения нелинейных дифференциальных уравнений// УМН. 1957. Т. 12. № 3(75). С. 3–73.
14. Hopf E. The partial differential equation $u_t + uu_x = \mu u_{xx}$ // Comm. Pure Appl. Math. 1950. V. 3. № 3. P. 201–230.
15. Олейник О. А. Об одном классе разрывных решений квазилинейных уравнений первого порядка // Научные доклады высшей школы. Физико-математические науки. 1958. № 3. С. 91–98.
16. Олейник О. А. О единственности и устойчивости обобщенного решения задачи Коши для квазилинейного уравнения // УМН. 1959. Т. 14. № 2(86). С. 165–170.

17. Гельфанд И. М. Некоторые задачи теории квазилинейных уравнений // УМН. 1959. Т. 14. № 2(86). С. 87–158.
18. Рождественский Б. Л., Яненко Н. Н. Системы квазилинейных уравнений и их приложения к газовой динамике. М.: Наука, 1978.
19. Горицкий А. Ю., Кружков С. Н., Чечкин Г. А. Уравнения с частными производными первого порядка: учебное пособие. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 1999.
20. Гасников А. В. Сравнение определений обобщенного решения задачи Коши для квазилинейного уравнения. М.: ВЦ РАН, 2006.
21. Иносэ Х., Хамада Т. Управление дорожным движением. М.: Транспорт, 1983.
22. Бабков В. Ф. Дорожные условия и безопасность дорожного движения. М.: Транспорт, 1982.
23. Kumei S., Bluman G. W. When nonlinear differential equations are equivalent to linear differential equations // SIAM J. Appl. Math. 1982. V. 42. № 5. P. 1157–1173.
24. Олвер П. Приложения групп Ли к дифференциальным уравнениям. М.: Мир, 1989.
25. Овсянников Л. В. Групповой анализ дифференциальных уравнений. М.: Наука, 1993.
26. Волосов К. А., Вдовина Е. К., Волосова А. К. Новые точные решения уравнений с частными производными параболического типа: учебное пособие. М.: МИИТ, 2010.
27. Кружков С. Н. Квазилинейные уравнения первого порядка со многими независимыми переменными // Матем. сб. 1970. Т. 81(123). № 2. С. 228–255.
28. Кружков С. Н. Нелинейные уравнения с частными производными (Лекции). Ч. 2. Уравнения первого порядка. М.: Изд-во МГУ, 1970.
29. Serre D. System of conservation laws: A challenge for the XXIst century, in: B. Enquist, W. Schmid (Eds.), Mathematics Unlimited – 2001 and Beyond. Berlin, New York: Springer-Verlag, 2001. P. 1061–1080.
30. Лионс П.-Л. (Lions P.-L.) О некоторых интригующих проблемах нелинейных уравнений в частных производных, в книге: «Математика: границы и перспективы». М.: ФАЗИС, 2005. С. 193–211.
31. Тупчев В. А. Обобщенные решения законов сохранения. М.: Наука, 2006.
32. Эванс Л. К. Методы слабой сходимости для нелинейных уравнений с частными производными. Новосибирск: Тамара Рожковская (Белая серия в математике и физике; Т. 2), 2006.

33. *Holden H., Risebro N. H.* Front tracking for hyperbolic conservation laws. Springer, 2007.
34. Nonlinear conservation laws and applications. University of Minnesota, July 13–31, 2009.
<http://www.ima.umn.edu/2008-2009/SP7.13-31.09/index.html#schedule>
35. *Dafermos C. M.* Hyperbolic conservation laws in continuum physics. Springer, 2010. (в издательстве РХД готовится перевод этой книги на русский язык)
36. *Галкин В. А.* Анализ математических моделей: системы законов сохранения, уравнения Больцмана и Смолуховского. М.: БИНОМ. Лаборатория знаний, 2009.
37. *Ладыженская О. А., Солонников В. А., Уральцева Н. Н.* Линейные и квазилинейные уравнения параболического типа. М.: Наука, 1967.
38. *Крылов Н. В.* Лекции по эллиптическим и параболическим уравнениям в пространствах Гельдера: учебное пособие. Новосибирск: Научная книга (Университетская серия; Т. 2), 1998.
39. *Милютин А. А., Дмитрук А. В., Осмоловский Н. П.* Принцип максимума в оптимальном управлении. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2004.
<http://www.milyutin.ru/papers.html>
40. Оптимальное управление / под ред. Н. П. Осмоловского и В. М. Тихомирова. М.: МЦНМО, 2009.
41. *Красовский Н. Н., Субботин А. И.* Позиционные дифференциальные игры. М.: Наука, 1974.
42. *Эванс Л. К.* Уравнения с частными производными. Новосибирск: Рожковская (Университетская серия; Т. 7), 2003.
43. *Демьянов В. Ф.* Минимакс, дифференцируемость по направлениям. Л.: Изд-во ЛГУ, 1974.
44. *Субботин А. И.* Обобщенные решения уравнений в частных производных. Перспективы динамической оптимизации. М.-Ижевск: НИЦ «РХД», ИКИ, 2003.
45. *Магарил-Ильяев Г. Г., Тихомиров В. М.* Выпуклый анализ и его приложения. М.: УРСС, 2003.
46. *Пшеничный Б. Н.* Выпуклый анализ и экстремальные задачи. М.: Наука, 1980.
47. *Дубовицкий А. Я., Милютин А. А.* Задачи на экстремум при наличии ограничений // ЖВМ и МФ. 1965. Т. 5. № 3. С. 395–453.
<http://www.milyutin.ru/papers.html>
48. *Беллман Р., Калаба Р.* Квазилинеаризация и нелинейные краевые задачи. М.: Мир, 1968.

49. Пшеничный Б. Н., Сагайдак М. И. О дифференциальных играх с фиксированным временем // Кибернетика. 1970. Т. 2. С. 54–63.
50. Maslov V. P., Belavkin V. P. Design of the optimal Dynamic Analyzer: Mathematical Aspects of Sound and Visual Pattern Recognition, in Mathematical Aspects of Computer Engineering. Edited by V. P. Maslov, K. A. Volosov. M.: MIR, 1988. P. 146–237.
51. Kolokoltsov V. N., Maslov V. P. Idempotent analysis and applications. Dordrecht: Kluwer Acad. Publ., 1997. (имеется схожая книга тех же авторов на русском языке (1994))
52. Litvinov G. L. Tropical mathematics, idempotent analysis, classical mechanics and geometry. AMS, Contemp. Math., 2010. [arXiv:1005.1247v1](https://arxiv.org/abs/1005.1247v1) (Семинар «Глобус», 2009. вып. 4)
53. Payne H. J. Models of freeway traffic and control, in: Simulation Council Proc. 28, Mathematical Models of Public Systems. Edited by G. A. Bekey. 1971. V. 1. P. 51–61.
54. Куликовский А. Г., Погорелов Н. В., Семенов А. Ю. Математические вопросы численного решения гиперболических систем уравнений. М.: Физматлит, 2001.
55. Петров И. Б., Лобанов А. И. Лекции по вычислительной математике. М.: Бином, 2006.
56. Швецов В. И. Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.
57. Чарахчьян А. А. Об алгоритмах расчета распада разрыва для схемы С. К. Годунова // ЖВМ и МФ. 2000. Т. 40. № 5. С. 782–796.
58. Helbing D. Traffic and related self-driven many particle systems // Reviews of modern physics. 2001. V. 73. № 4. P. 1067–1141.
[arXiv:cond-mat/0012229](https://arxiv.org/abs/cond-mat/0012229)
59. Смирнов Н. Н., Киселёв А. Б., Никитин В. Ф., Юмашев М. В. Неустановившиеся движения автотранспорта на кольцевой магистрали // Прикладная математика и механика. 2000. Т. 64. № 4. С. 651–658.
60. Смирнов Н. Н., Киселев А. Б., Никитин В. Ф., Юмашев М. В. Математическое моделирование автомобильных потоков на магистралях // Вестник Московского университета. Математика. Механика. 2000. № 4. С. 39–44.
61. Киселев А. Б., Кокорева А. В., Никитин В. Ф., Смирнов Н. Н. Математическое моделирование автотранспортных потоков на регулируемых дорогах // Прикладная математика и механика. 2004. Т. 68. № 6. С. 1047–1054.
62. Холодов Я. А., Холодов А. С., Гасников А. В., Морозов И. И., Тарасов В. Н. Моделирование транспортных потоков – актуальные проблемы

- мы и пути их решения // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)
63. *Daganzo C. F.* Fundamentals of transportation and traffic operations. New-York: Elsevier Science inc., 1997.
 64. *Aw A., Rascle M.* Resurrection of “second order” models of traffic flow // SIAM Journal of Applied Mathematics. 2000. V. 60. P. 916–938.
 65. *Greenberg J. M.* Extensions and amplifications of a traffic model of Aw and Rascle // SIAM J. Appl. Math. 2001. V. 62. № 3. P. 729–745.
 66. *Zhang H. M.* A non-equilibrium traffic model devoid of gas-like behavior // Transp. Res. B. 2002. V. 36. P. 275–290.
 67. *Siebel F., Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // e-print [arXiv:physics/0509124v2](https://arxiv.org/abs/physics/0509124v2), 2006.
 68. *Helbing D.* Improved fluid-dynamic model for vehicular traffic // Phys. Rev. E. 1995. V. 51. P. 3163–3169.
 69. Ладыженская О. А. Шестая проблема тысячелетия: уравнение Навье–Стокса, существование и гладкость // УМН. 2003. Т. 58. № 2(350). С. 45–78.
 70. Юдович В.И. Глобальная разрешимость — против коллапса в динамике несжимаемой жидкости, в книге: «Математические события XX века». М.: Фазис, 2003. С. 519–548.
 71. Проблемы турбулентности. Сборник работ. М.–Ижевск: НИЦ «РХД», ИКИ, 2006.
 72. *Prigogine I., Herman R.* Kinetic theory of vehicular traffic. N.Y.: Elsevier, 1971.
 73. Кац М. Вероятность и смежные вопросы в физике. М.: Мир, 1965.
 74. Козлов В. В. Ансамбли Гиббса и неравновесная статистическая механика. М.–Ижевск: НИЦ «РХД», ИКИ, 2008.
 75. Веденягин В. В. Кинетическая теория по Максвеллу, Больцману и Власову. М.: Изд-во МГОУ, 2005.
 76. Карамзин Ю. Н., Трапезникова М. А., Четверушкин Б. Н., Чубарова Н. Г. Двумерная модель автомобильных потоков // Матем. мод. 2006. Т. 18. № 6. С. 85–95.
 77. Сухинова А. Б., Трапезникова М. А., Четверушкин Б. Н., Чубарова Н. Г. Двумерная макроскопическая модель транспортных потоков // Матем. мод. 2009. Т. 21. № 2. С. 118–126.
 78. *Garavello M., Piccoli B.* Traffic Flow on Networks. Volume 1 of AIMS Series on Applied Mathematics. AIMS, 2006.

79. 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. C.I.M.E. Courses, 2009.
<http://php.math.unifi.it/users/cime/Courses/2009/01/200914-Notes.pdf>
80. Kurzhanskiy 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://lihodeev.com/pubs.html>
<http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-148.pdf>;
81. Куржанский А. А., Куржанский А. Б., Варайя П. Роль макро-моделирования в активном управлении транспортной сетью // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)
82. Daganzo C. F. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // Transp. Res. B. 1994. V. 28. № 4. P. 269–287.
83. Daganzo C. F. The cell transmission model, Part II: Network traffic // Transp. Res. B. 1995. V. 29. № 2. P. 79–93.
84. Буслаев А. П., Таташев А. Г., Яшина М. В. О свойствах решений одного класса систем нелинейных дифференциальных уравнений на графах // Владикавказкий матем. журн., ВНЦ РАН. 2004. Т. 6. № 4. С. 4–18.
85. Назаров А. И. Об устойчивости стационарных режимов в одной системе ОДУ, возникающей при моделировании автотранспортных потоков // Вестник СПбГУ. Серия 1. Математика. Механика. Астрономия. 2006. № 3. С. 35–43.
86. Lubashevsky I., Kalenkov S., Mahnke R. Towards a variational principle for motivated vehicle motion // Phys. Rev. E. 2002. V. 65. P. 1–5.
87. Lubashevsky I., Wagner P., Mahnke R. Towards the fundamentals of car following theory // e-print [arXiv:cond-mat/0212382v2](https://arxiv.org/abs/cond-mat/0212382v2), 2003.
88. Newell G. F. Nonlinear effects in the dynamics of car – following // Oper. Res. 1961. V. 9. P. 209–229.
89. Новокшенов В. Ю. Введение в теорию солитонов. М.–Ижевск: НИЦ «РХД», ИКИ, 2002.
90. Ильин А. М., Олейник О. А. Асимптотическое поведение решений задачи Коши для некоторых квазилинейных уравнений при больших значениях времени // Матем. сб. 1960. Т. 51(93). № 2. С. 191–216.

91. *Osher S., Ralston J.* L^1 stability of traveling waves with application to convective porous media flow // Comm. Pure Appl. Math. 1982. V. 35. P. 737–749.
92. *Weinberger H. F.* Long-time behavior for regularized scalar conservation law in absence of genuine nonlinearity // Ann. Inst. H. Poincaré, Anal. Non Linéaire. V. 7. 1990. P. 407–425.
93. *Liu T.-P.* Admissible solutions of hyperbolic conservation laws // Mem. Amer. Math. Soc. 1981. V. 30. № 240. P. 1–78.
94. *Cheng K.-S.* Asymptotic behavior of solution of a conservation law without convexity condition // J. Diff. Equat. 1981. V. 40. № 3. P. 343–376.
95. *Петросян Н. С.* Об асимптотике решения задачи Коши для квазилинейного уравнения первого порядка с невыпуклой функцией состояния // УМН. 1983. Т. 38. № 2(230). С. 213–214.
96. *Кружков С. Н., Петросян Н. С.* Асимптотическое поведение решений задачи Коши для нелинейных уравнений первого порядка // УМН. 1987. Т. 42. № 5(257). С. 3–40.
97. *Jennings G.* Discrete shocks // Comm. Pure Appl. Math. 1974. V. 27. P. 25–37.
98. *Harten A., Hyman J. M., Lax P.D.* On finite-difference approximations and entropy conditions for shocks // Comm. Pure Appl. Math. 1976. V. 29. P. 297–322.
99. *Engquist B., Osher S.* One-sided difference approximations for nonlinear conservation laws // Math. Comp. 1981. V. 36. P. 321–351.
100. *Henkin G. M., Polterovich V. M.* Shumpetarian dynamics as non-linear wave theory // J. Math. Econom. 1991. V. 20. P. 551–590.
101. *Henkin G. M., Polterovich V. M.* A difference-differential analogue of the Burgers equation and some models of economic development // Discrete and continuous dynamic systems. 1999. V. 5. № 4. P. 697–728.
102. *Mejai M., Volpert Vit.* Convergence to systems of waves for viscous scalar conservation laws // Asymptotic Analysis. 1999. V. 20. P. 351–366.
103. *Engelberg S., Schochet S.* Nonintegrable perturbation of scalar viscous shock profiles // Asymptotic Analysis. 2006. V. 48. P. 121–140.
104. *Henkin G. M., Shanarin A. A.* Asymptotic behavior of solutions of the Cauchy problem for Burgers type equations // J. Math. Purés Appl. 2004. V. 83. P. 1457–1500.
105. *Henkin G. M., Shanarin A. A., Tumanov A. E.* Estimates for solution of Burgers type equations and some applications // J. Math. Purés Appl. 2005. V. 84. P. 717–752.

106. *Henkin G. M.* Asymptotic structure for solutions of the Cauchy problem for Burgers type equations // J. fixed point theory appl. 2007. V. 1. № 2. P. 239–291.
107. *Serre D.* L^1 stability of shock waves in scalar conservation laws, in: Evolutionary Equations // Handbook of Differential Equations, North-Holland, Amsterdam. 2004. V. 1. P. 473–553.
108. *Гасников А. В.* О промежуточной асимптотике решения задачи Коши для квазилинейного уравнения параболического типа с монотонным начальным условием // Известия РАН. Теория и системы управления. 2008. № 3. С. 154–163.
109. *Гасников А. В.* Сходимость по форме решения задачи Коши для квазилинейного уравнения параболического типа с монотонным начальным условием к системе волн // ЖВМ и МФ. 2008. Т. 48. № 8. С. 1458–1487.
110. *Гасников А. В.* Асимптотическое по времени поведение решения начальной задачи Коши для закона сохранения с нелинейной дивергентной вязкостью // Известия РАН. Серия математическая. Т. 76. 2009. № 6. С. 39–76.
111. *Гасников А. В.* Асимптотика по времени решения задачи о распаде “размазанного разрыва” для закона сохранения // Труды МФТИ (специальный выпуск, посвященный юбилею ФУПМа). 2009. Т. 1. № 4. С. 120–125. <http://mipt.ru/nauka/trudy/N4.html>
112. *Хёрмандер Л.* Анализ линейных дифференциальных операторов с частными производными. Т. 1 – Т. 4. М.: Мир, 1986–1988.
113. *Gazis D. C.* Traffic science. N.Y.: Wiley, 1974.
114. *Treiber M., Helbing D.* Explanation of observed features of self-organization in traffic flow // e-print [arXiv:cond-mat/9901239](https://arxiv.org/abs/cond-mat/9901239), 1999.
115. *Treiber M., Hennecke A., Helbing D.* Congested traffic states in empirical observations and microscopic simulation // Phys. Rev. E. 2000. V. 62. P. 1805–1824.
116. *Фон Нейман Дж.* Теория самовоспроизводящихся автоматов. М.: УРСС, 2010.
117. *Cremer M., Ludwig J.* A fast simulation model for traffic flow on the basis of Boolean operations // Math. Comp. Simul. 1986. V. 28. P. 297–303.
118. *Nagel K., Schreckenberg M.* A cellular automation model for freeway traffic // Phys. I France. 1992. V. 2. P. 2221–2229.
119. *Chowdhury D., Santen L., Schadschneider A.* Statistical physics of vehicular traffic and some related systems // Phys. Rep. 2000. V. 329. P. 199–329. [arXiv:cond-mat/0007053v1](https://arxiv.org/abs/cond-mat/0007053v1)

120. *Nagatani T.* The physics of traffic jams // Reports on Progress in Physics. 2002. V. 65. P. 1331–1386.
121. *Benassi A., Fouque J.-P.* Hydrodynamic limit for the asymmetric simple exclusion process // Ann. of Probability, V. 15. № 2. 1987. P. 546–560.
122. *Kipnis C., Olla S., Varadhan S. R. S.* Hydrodynamics and large deviation for simple exclusion processes // Comm. on Pure and Applied Mathematics. 1989. V. 42. P. 115–137.
123. *Nishinari K., Matsukidaira J., Takahashi D.* Two-dimensional Burgers cellular automaton // e-print [arXiv:nlin/0102027v1](https://arxiv.org/abs/nlin/0102027v1), 2001.
124. *Бланк М. Л.* Точный анализ динамических систем, возникающих в моделях транспортных потоков // УМН. 2000. Т. 55(333). № 3. С. 167–168.
125. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111. № 3–4. P. 903–930.
[arXiv:math.DS/0206194](https://arxiv.org/abs/math/0206194)
126. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120. № 3–4. P. 627–658. [arXiv:math/0408240](https://arxiv.org/abs/math/0408240)
127. *Минлос Р. А.* Введение в математическую статистическую физику. М.: МЦНМО, 2002.
128. *Maerivoet S., De Moor B.* Cellular automata models of road traffic // Physics Reports 2005. V. 419. № 1. P. 1–64. [arXiv:physics/0509082](https://arxiv.org/abs/physics/0509082)
129. *Буслاءев А. П., Новиков А. В., Приходько В. М., Таташев А. Г., Яшина М. В.* Вероятностные и имитационные подходы к оптимизации автодорожного движения. М.: Мир, 2003.
130. *Buslaev A. P., Prikhodko V. M., Tatashev A. G., Yashina M. V.* The deterministic – stochastic flow model // e-print [arXiv:physics/0504139v1](https://arxiv.org/abs/physics/0504139v1), 2005.
131. *Buslaev A. P., Gasnikov A. V., Yashina M. V.* Selected mathematical problems of traffic flow theory // International Journal of Computer Mathematics. Published By: Taylor & Francis. 2010. (in print)
132. Явление чрезвычайное. Книга о А. Н. Колмогорове. М.: ФАЗИС, МИРОС, 1999. С. 236–237.
133. *Баренблatt Г. И.* Автомодельные явления – анализ размерностей и скейлинг. Долгопрудный: Издательский дом «Интеллект», 2009.
134. *Ибрагимов Н. Х.* Практический курс дифференциальных уравнений и математического моделирования. Нижний Новгород: Издательство Нижегородского университета, 2007.
135. *Шестаков А. А.* Обобщенный прямой метод Ляпунова для систем с распределенными параметрами. М.: КомКнига, 2007.

136. Volpert A. I., Volpert Vit. A., Volpert Vl. A. Traveling waves solutions of parabolic system // Translations of Mathematical Monographs. 2000. V. 140. P. 1–455.
137. Колмогоров А. Н., Петровский И. Г., Пискунов Н. С. Исследование уравнения диффузии, соединенной с возрастанием количества вещества, и его применение к одной биологической проблеме // Бюл. МГУ. Математика и механика. 1937. № 6. Т. 1. С. 1–26.
138. Разжевайкин В. Н. Решения типа бегущей волны для уравнения реакции – нелинейной диффузии // Труды МФТИ (специальный выпуск, посвященный юбилею ФУПМа). 2009. Т. 1. № 4. С. 99–119.
<http://mipt.ru/nauka/trudy/N4.html>
139. Наумкин П. И., Шишимарев И. А. Задача о распаде ступеньки для уравнения Кортевега–де Фриза–Бюргерса // Функц. анализ и его прил. 1991. Т. 25. № 1. С. 21–32.
140. Наумкин П. И., Шишимарев И. А. О распаде ступеньки для уравнения Кортевега–де Фриза–Бюргерса // Функц. анализ и его прил. 1991. Т. 26. № 2. С. 88–93.
141. Казейкина А. В. Устойчивость решения задачи Коши вида бегущей волны для уравнения Кортевега–де Фриза–Бюргерса // ЖВМ и МФ. 2010. Т. 50. № 4. С. 1–21.
142. Duan R., Zhao H. Global stability of strong rarefaction waves for the generalized KdV–Burgers equation // Nonlinear Anal. 2007. V. 66. P. 1100–1117.
143. Liu T.-P., Nishihara K. Asymptotic behavior for scalar viscous conservation laws with boundary effect // Journal of differential equations. 1997. V. 133. P. 296–320.
144. Liu T.-P., Matsumura A., Nishihara K. Behaviors of solutions for the Burgers equation with boundary corresponding to rarefaction waves // SIAM J. Math. Anal. 1998. V. 29. № 2. P. 293–308.
145. Куликовский А. Г., Чугайнова А. П. Классические и неклассические разрывы в решениях уравнений нелинейной теории упругости // УМН. 2008. Т. 63. № 2(380). С. 85–152.
<http://www.mi.ras.ru/spm/pdf/007.pdf>
<http://www.mi.ras.ru/noc/lectures/16kulikovskii.pdf>
146. Шубин М. А. Лекции об уравнениях математической физики. М.: МЦНМО, 2003.
147. Габушин В. Н. Неравенства между производными в метриках L_p при $0 < p \leq \infty$ // Известия АН СССР. Серия математическая. 1976. Т. 40. № 4. С. 869–892.

148. Габушин В. Н. Неравенства для производных решений обыкновенных дифференциальных уравнений в метриках L_p ($0 < p \leq \infty$) // Дифференциальные уравнения. 1988. Т. 24. № 10. С. 1662–1670.
149. Арестов В. В. Приближение неограниченных операторов ограниченными и родственные экстремальные задачи // УМН. 1996. Т. 51. № 6(312). С. 89–124.
150. Годунов С. К. Проблема обобщенного решения в теории квазилинейных уравнений и в газовой динамике // УМН. 1962. Т. 17. № 3(105). С. 147–158.
151. Арнольд В. И., Виишук М. И., Ильяшенко Ю. С., Калашников А. С., Кондратьев В. А., Кружков С. Н., Ландис Е. М., Милиончиков В. М., Олейник О. А., Филиппов А. Ф., Шубин М. А. Некоторые нерешенные задачи теории дифференциальных уравнений и математической физики // УМН. 1989. Т. 44. № 4(268). С. 191–202.
152. Панов Е. Ю. О единственности решения задачи Коши для квазилинейного уравнения первого порядка с одной допустимой строго выпуклой энтропией // Матем. заметки. 1994. Т. 55. № 5. С. 116–129.
153. Kerner B. S. The Physics of Traffic. Berlin: Springer, 2004.
154. Bando M., Hasebe K., Nakayama A., Shibata A., Sugiyama Y. Dynamical model of traffic congestion and numerical simulation // Phys. Rev. E. 1995. V. 51. P. 1035–1042.
155. Barlovic R., Santen L., Schadschneider A., Schreckenberg M. Metastable states in cellular automata for traffic flow // Eur. Phys. J. B. 1998. V. 5. P. 793.
156. Wiedemann R. Simulation des Verkehrsflusses. Karlsruhe: University of Karlsruhe, 1974.
157. Kerner B. S., Konhäuser P. Cluster effect in initially homogeneous traffic flow // Phys. Rev. E. 1993. V. 48. P. 2335–2338.
158. Kerner B. S., Konhäuser P. Structure and parameters of clusters in traffic flow // Phys. Rev. E. 1994. V. 50. P. 54–83.
159. Kerner B. S., Klenov S. L., Konhäuser P. Asymptotic theory of traffic jams // Phys. Rev. E. 1997. V. 56. P. 4200–4216.

Глава 3. Теория Кернера трех фаз в транспортном потоке – новый теоретический базис для интеллектуальных транспортных технологий¹

3.1. Три фазы транспортного потока.....	171
3.2. Свободный транспортный поток – фаза F	172
3.3. Плотный транспортный поток.....	172
3.4. Определение фаз J и S в плотном транспортном потоке	173
3.5. Возникновение плотного потока (traffic breakdown) – F→S фазовый переход	175
3.6. Бесконечное число пропускных способностей скоростной автомагистрали.....	180
3.7. Широкие движущиеся кластеры (локальные движущиеся заторы) – фаза J	182
3.8. Синхронизованный транспортный поток – фаза S	184
3.9. S→J фазовый переход	185
3.10. Неоднородные пространственно-временные структуры транспортного потока, состоящие из фаз S и J	186
3.11. Стохастическая модель в рамках теории трех фаз Кернера	188
3.12. Применение теории трех фаз Кернера для интеллектуальных транспортных технологий	197
Литература.....	199

¹ Эта глава написана С. Л. Клёновым. Автор выражает благодарность Б. С. Кернеру за предоставление оригинальных рисунков из его книг.

В 1996–2002 годах Б. С. Кернер с сотрудниками концерна Даймлер провели детальные исследования эмпирических данных, измеренных с помощью датчиков на многочисленных скоростных автомагистралях мира (в Германии, Голландии, Англии, США). Главный результат этих исследований был сформулирован в предисловии к книге [1]:

«Теории транспортного потока и математические модели, которые доминируют в настоящее время в научных журналах и учебных курсах большинства университетов, не могут объяснить ни сам переход от свободного к плотному потоку (traffic breakdown), ни основные свойства возникающих в результате этого перехода структур транспортного потока».

По этой причине Б. С. Кернер предложил и разработал альтернативную теорию транспортных потоков, названную *теорией трех фаз*, которая может предсказать и объяснить эмпирические свойства перехода к плотному потоку (traffic breakdown) и результирующих пространственно-временных структур в транспортном потоке. Как достижения, так и критика предшествующих классических подходов к теории транспортных потоков представлены в главе 10 книги [1]. В настоящей главе кратко излагаются основные положения теории трех фаз Кернера в соответствии с [1, 2].

Цель этой главы состоит в том, чтобы дать читателю определенное представление об эмпирическом базисе и основных идеях теории трех фаз. Тем не менее эта глава не может заменить книг Б. С. Кернера, чтение которых необходимо тем, кто хочет разобраться в данной теории. По этой причине основное внимание уделяется не математическому обоснованию тех или иных положений теории трех фаз, а в основном качественному описанию. Единственное исключение составляет п. 3.11, в котором кратко описана стохастическая трехфазная модель в рамках теории трех фаз и некоторые ее решения. Для более подробного ознакомления с математическими результатами теории трех фаз Кернера рекомендуется прочитать главу 11 книги [1] и часть III книги [2].

Прежде чем перейти к изложению некоторых положений теории трех фаз Кернера, важно отметить имевшиеся на тот момент фундаментальные достижения (см. пп. 2.1–2.3) по математической формулировке моделей транспортного потока, а также многочисленных эффектов взаимодействия между водителями [3–25]. К ним относится, в первую очередь, введенная в моделях ДМ-класса задержка водителей [10–13], приводящая к переторможению как реакции на замедление АТС впереди [13, 19, 20]; формулировка задержки водителей через модельные флуктуации, введенные в моделях Нагеля–Шрекенберга и А. Шашнай-

дера (см. обзоры [19–22]); микроскопическое описание slow-to-start rule в работах группы М. Шрекенберга (см. обзоры [19–22]). Эти и другие математические формулировки в поведения водителя являются также важными элементами математических моделей трех фаз (см. главу 11 книги [1]).

Несмотря на эти достижения, как было сказано выше, предшествующие модели не могут объяснить ни сам переход от свободного к плотному потоку, ни основные свойства возникающих пространственно-временных структур, наблюдавшихся в эмпирических данных. Детальное объяснение этого «парадокса» дано в главе 10 книги [1]. Этот парадокс объясняется очень просто: анализ эмпирических данных, который позволил выявить фундаментальные эмпирические свойства перехода от свободного к плотному и основные свойства результирующих пространственно-временных структур, стал возможным только в конце 90-х годов, когда стало доступным огромное количество данных измерений со скоростных магистралей Германии и Голландии. Иными словами, выдающиеся ученые, которые создали многочисленные модели транспортного потока, перечисленные выше и многие другие, просто не могли знать, какими же реальными свойствами обладает переход от свободного к плотному транспортному потоку.

Эти фундаментальные эмпирические пространственно-временные свойства перехода от свободного к плотному потоку, а также другие фундаментальные эмпирические свойства фазовых переходов в транспортном потоке детально описаны в п. 2.4 и части II книги [2]. В рамках данной главы нет возможности остановиться на рассмотрении всех этих эмпирических свойств транспортного потока. Тем не менее повторим здесь еще раз фундаментальные эмпирические свойства перехода к плотному транспортному потоку и эмпирические свойства фазовых переходов в транспортном потоке (см. п. 2.4).

Фундаментальные эмпирические свойства перехода к плотному транспортному потоку следующие:

1. Переход к плотному транспортному потоку (*traffic breakdown*) является F→S переходом (буква F соответствует *free flow*, т.е. свободному потоку, буква S обозначает фазу синхронизованного потока, в английской литературе *synchronized flow*).

2. Вероятность спонтанного F→S перехода является растущей функцией величины потока ATC.

3. Может быть как спонтанный, так и индуцированный F→S переход около одного и того же узкого места на дороге (*bottleneck*).

LWR-модель кинематических и ударных волн в транспортном потоке, дискретной версией которой является СТМ-модель Даганзо [9], не могут описывать пункты 2 и 3, а модели, относящиеся к ДМ-классу, не могут описывать пункты 1–3 (см. главу 2).

Фундаментальные эмпирические свойства фазовых переходов в транспортном потоке следующие:

а) В соответствии со свойствами 1–3, указанными выше, переход от свободного к плотному транспортному потоку (traffic breakdown) является $F \rightarrow S$ фазовым переходом I рода.

б) Широкие движущиеся кластеры (wide moving jams, обозначается ниже буквой J) возникают спонтанно только в синхронизованном потоке, т.е. в результате последовательности $F \rightarrow S \rightarrow J$ фазовых переходов.

в) $S \rightarrow J$ фазовый переход происходит позднее и часто совсем в другом месте, чем $F \rightarrow S$ фазовый переход.

LWR-теория кинематических волн не может описывать пункты б) и в), а теория, основанная на моделях ДМ-класса, не может описывать пункты а)–в).

3.1. Три фазы транспортного потока

Теория трех фаз фокусируется главным образом на физике плотного транспортного потока на скоростных автомагистралях. И описывает три фазы транспортного потока, в то время как классические теории, базирующиеся на фундаментальной диаграмме транспортного потока, рассматривают две фазы: *свободный поток* и так называемый *плотный поток* (congested traffic в английской терминологии). В плотном потоке выделяются две фазы, соответственно существуют три фазы транспортного потока:

- 1.** Свободный поток – фаза F.
- 2.** Синхронизированный поток – фаза S.
- 3.** Широкий движущийся кластер (локальный движущийся затор, в английской литературе wide moving jam) – фаза J.

Фаза определяется как некоторое *состояние* транспортного потока, рассматриваемое в *пространстве и времени*.

Следует подчеркнуть, что в теории трех фаз разделение на свободный и плотный поток точно такое же, как и в классических теориях Лайтхилла–Уизема и Дженерал Моторс (см. п. 3.3 ниже). Фундаментальное отличие теории Б. С. Кернера состоит в том, что он

выделяет две фазы в плотном потоке на основе общих эмпирических пространственно-временных свойств транспортного потока, которые за все годы измерений остаются одни и те же на разных автодорогах мира. Другими словами, как определение фаз транспортного потока, так и остальные положения теории, приведенные в пп. 3.1–3.10, основаны исключительно на эмпирических данных.

3.2. Свободный транспортный поток – фаза F

В свободном транспортном потоке достаточно малой плотности водители могут практически свободно установить желаемую для них скорость. Эмпирические данные, относящиеся к свободному потоку, показывают положительную корреляцию между величиной потока q , измеряемой в количестве АТС в единицу времени (проходящих через данное сечение дороги), и плотностью ρ , измеряемой в количестве АТС на единицу длины дороги [14, 15]. Зависимость потока q от плотности ρ для свободного потока ограничена максимальным значением величины потока $q = q_{\max}$ и соответствующим критическим значением плотности $\rho = \rho_{crit}$ (рис. 1), которые могут быть достигнуты в свободном потоке.

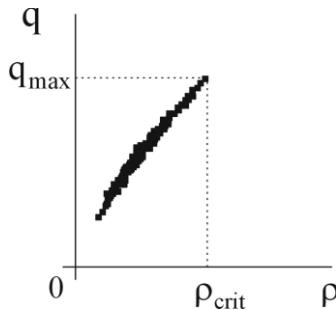


Рис. 1. Зависимость потока q от плотности АТС ρ в свободном потоке [14, 20]

3.3. Плотный транспортный поток

В плотном транспортном потоке, который определяется так же, как и в классических теориях Лайтхилла–Уизема и Дженирал Моторс (см. главу 2), скорость АТС меньше, чем минимально возможная ско-

рость АТС в свободном потоке. Это означает, что прямая с наклоном, равным минимальной скорости в транспортном потоке (штриховая линия на рис. 2), разделяет все эмпирические данные (точки) на плоскости поток–плотность на две области: слева от этой прямой находятся данные, относящиеся к свободному потоку, а справа – данные, относящиеся к плотному потоку.

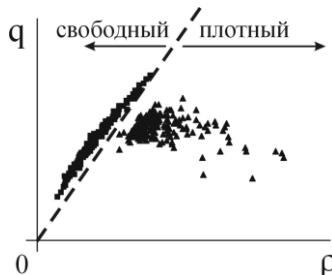


Рис. 2. Зависимость потока от плотности АТС
в свободном и плотном потоке [14, 20]

Как следует из данных измерений, возникновение плотного потока обычно происходит вблизи неоднородности на автомагистрали, вызванной въездом на автомагистраль, съездом с нее, изменением числа полос, сужением дороги, подъемом и т.п. Такого типа неоднородность, вблизи которой может происходить переход к плотному транспортному потоку, в дальнейшем будем называть *узким местом* или «бутылочным горлом» [14, 20].

3.4. Определение фаз J и S в плотном транспортном потоке

Б. С. Кернер показал, что фундаментальная диаграмма и ее применения в том виде, как они используются в классических теориях транспортного потока, неадекватным образом описывают сложную динамику в плотном транспортном потоке. Он выделяет, таким образом, в плотном транспортном потоке фазу S *синхронизованного потока*, в англоязычной литературе «*synchronized flow*», и фазу J *широкого движущегося кластера* (локальный движущийся затвор, «*wide moving jam*»). Определение фаз [J] и [S] в плотном потоке является результатом общих пространственно-временных свойств реальных данных, полученных в результате ежедневных измерений параметров транспортного потока во многих странах на различных скоростных автодорогах в течение многих лет. Б. С. Кернер определил фазы J и S следующим образом.

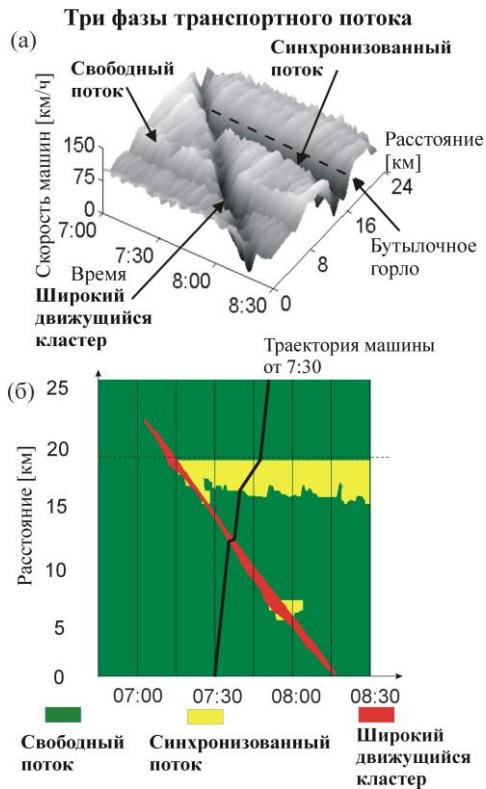


Рис. 3. Данные измерений скорости АТС в пространстве и времени (а) и их представление на координатно-временной плоскости (б). Взято из [1]

Определение фазы [L] широкого движущегося кластера:

Задний по направлению движения фронт широкого движущегося кластера (локального движущегося затора), где АТС, выезжающие из кластера, ускоряются вплоть до свободного или до синхронизированного потока, движется против потока с постоянной средней скоростью v_g , проходя через все узкие места на скоростной автомагистрали. Это характеристическое свойство широкого движущегося кластера.

Определение фазы [S] синхронизованного потока:

Задний по направлению движения фронт области синхронизированного потока, где АТС ускоряются вплоть до свободного потока, НЕ обладает характеристическим свойством широкого движущегося кла-

стера. В частности, задний фронт синхронизированного потока часто фиксирован вблизи узкого места на скоростной автомагистрали.

Необходимо подчеркнуть, что определение фаз [J] и [S] вытекает из эмпирических пространственно-временных свойств плотного потока, т.е. исходно не имеет никакого отношения к какой-либо математике. Теоретический смысл этих определений можно понять, прочтя раздел 6.1 книги [1].

Данные измерений средней скорости АТС (рис. 3 (а)) иллюстрируют определения [J] и [S]. На рис. 3 (а) имеются две пространственно-временные структуры плотного потока с низкой скоростью АТС. Одна из них распространяется против потока с почти постоянной скоростью заднего фронта через все узкие места на скоростной автомагистрали. Согласно определению [J], эта область плотного потока относится к фазе «широкого движущегося кластера». Напротив, задний фронт другой области плотного потока фиксирован вблизи места съезда с автомагистрали. Согласно определению [S], эта область плотного потока относится к фазе «синхронизированного потока» (рис. 3 (а) и (б)).

3.5. Возникновение плотного потока (traffic breakdown) – F→S фазовый переход

Переход от свободного к плотному потоку в англоязычной литературе известен как traffic breakdown. В теории трех фаз такой переход объясняется возникновением фазы синхронизированного потока, т.е. F→S фазовым переходом. Такое объяснение основывается на имеющихся данных измерений, которые показывают, что после возникновения плотного потока вблизи узкого места на автомагистрали задний фронт возникшего плотного потока фиксирован вблизи этого узкого места. Таким образом, возникший плотный поток удовлетворяет определению [S] фазы синхронизированного потока. В самом деле, типичный пример перехода из свободного в синхронизированный поток вблизи въезда показан на рис. 4. Из рисунка видно, что в то время как скорость АТС резко уменьшается в процессе перехода (рис 4 (а)), поток меняется мало (рис. 4 (б)). С скачок скорости при мало меняющемся потоке особенно наглядно виден на рис. 4 (в). В течение всего времени после перехода задний фронт между плотным и свободным потоками фиксирован на въезде дороге. По этой причине плотный поток соответствует определению фазы синхронизированного потока, поэтому весь плотный поток относится к фазе синхронизированного потока.

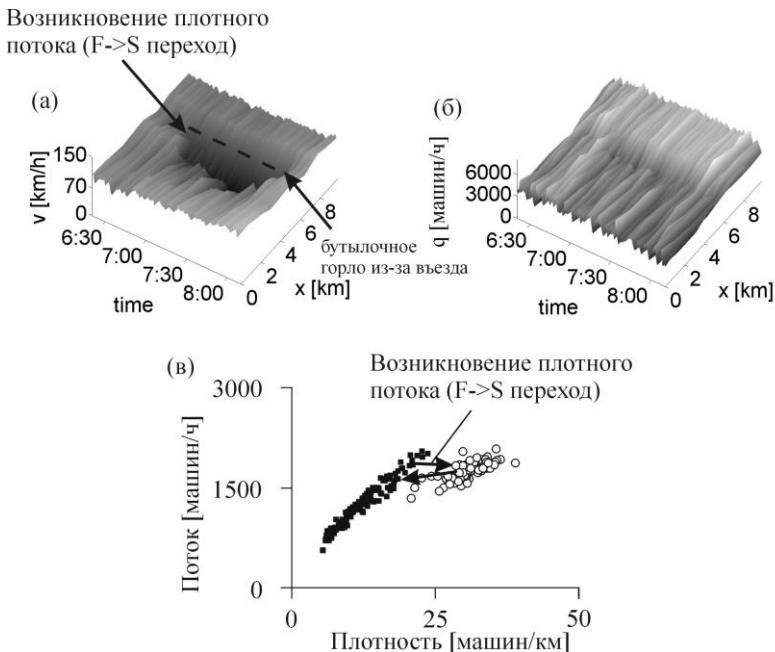


Рис. 4. Эмпирический пример возникновения плотного потока и эффект гистерезиса у бутылочного горла из-за въезда на автодорогу. (а, б) Средняя скорость (а) и поток (б) на автодороге в пространстве и времени (увеличение потока после въезда на (б) связано с потоком въезжающих на дорогу АТС). (в) Эффект гистерезиса в плоскости поток-плотность, обозначенный двумя стрелками, показывающими переход к плотному потоку и обратный переход к свободному потоку. 1-минутные данные. Взято из [1]

Образование плотного потока примерно в 6:30 (показанное стрелкой слева направо на рис. 4 (в)) и его исчезновение примерно в 7:45 (показанное стрелкой справа налево на рис. 4 (в)) сопровождается гистерезисом, хорошо известным в теории фазовых переходов 1 рода, наблюдаемых в широком классе неравновесных физических, химических и биологических систем. Это свойство F \rightarrow S фазового перехода является общим свойством реального (эмпирического) транспортного потока, который также представляет собой сложную сильно неравновесную систему.

Второй эмпирический пример перехода к плотному потоку показан на рис. 5. На этом примере можно видеть, как в реальном транс-

портном потоке образуются широкие движущиеся кластеры (рис. 5 (б)), координата дороги $x = 0$ км). Как можно видеть, в результате перехода к плотному потоку на узком месте, связанным с въездом на скоростную магистраль, сначала образуется фаза S синхронизованного потока.

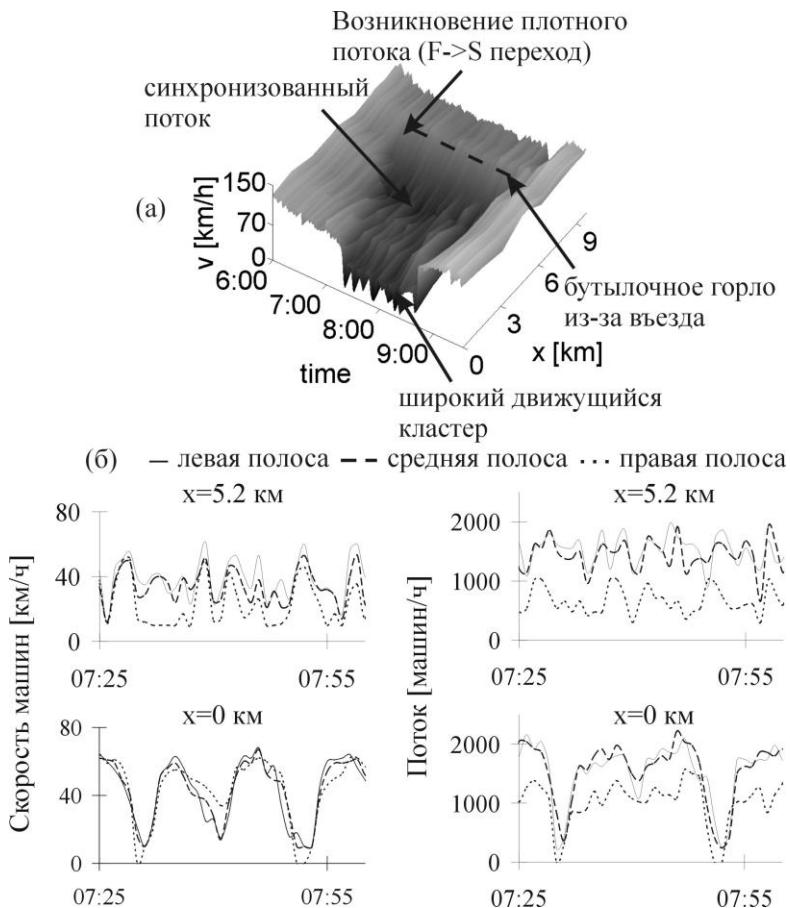


Рис. 5. Эмпирический пример возникновения широких движущихся кластеров в синхронизированном потоке: (а) Скорость АТС в пространстве и времени. (б) Скорость (слева) и поток (справа) на трех полосах дороги в области синхронизированного потока ($x = 5.2$ км) и в области широких движущихся кластеров ($x = 0$ км). Взято из [1]

Действительно, в течение всего времени существования плотного потока на этом узком месте задний фронт плотного потока, на котором АТС ускоряются из плотного потока до свободного потока, фиксирован на этом узком месте. Поэтому по определению фаз в теории Б. С. Кернера в результате перехода к плотному потоку образуется фаза синхронизованного потока. Другими словами, плотный поток образуется в результате $F \rightarrow S$ перехода. Напротив, широкие движущиеся кластеры возникают позднее уже внутри фазы синхронизованного потока. Этот $S \rightarrow J$ фазовый переход будет рассмотрен ниже в п. 3.9.

Таким образом, переход от свободного к плотному потоку в эмпирических данных есть $F \rightarrow S$ переход первого рода. Это эмпирическое свойство есть общее свойство реальных транспортных потоков на скоростных магистралях. Напротив, в моделях ДМ-класса, как было объяснено в п. 2.4, переход от свободного к плотному потоку связан с возникновением широких движущихся кластеров.

Исходя из эмпирических данных, был сделан вывод, что синхронизованный поток может возникать в свободном потоке спонтанно (спонтанный $F \rightarrow S$ переход) или индуцированным образом (индуцированный $F \rightarrow S$ переход). Спонтанный $F \rightarrow S$ переход означает, что переход к синхронизованному потоку происходит в случае, когда до момента перехода в окрестности узкого места существует свободный поток, а сам фазовый переход происходит в результате роста внутреннего возмущения транспортного потока. В противоположность этому индуцированный $F \rightarrow S$ переход происходит из-за возмущения транспортного потока, которое первоначально возникает на некотором удалении от положения узкого места, и затем по мере распространения достигает окрестности узкого места. Обычно индуцированный $F \rightarrow S$ переход связан с распространением в направлении против потока области синхронизованного потока или же широкого движущегося кластера, которые первоначально возникли вблизи следующего в направлении потока узкого места. Эмпирический пример индуцированного фазового перехода, приводящего к возникновению синхронизованного потока, показан на рис. 3: синхронизованный поток возникает благодаря распространению против потока широкого движущегося кластера.

Природу $F \rightarrow S$ фазового перехода можно объяснить с помощью «соревнования» во времени и пространстве двух противоположных процессов: ускорения АТС при обгоне более медленного АТС впереди, названном «переускорением», и в случае, когда обгон невозможен, торможения АТС до скорости более медленного АТС, названном «адаптация скорости». «Переускорение» поддерживает дальнейшее существование

вание свободного потока. Напротив, «адаптация скорости» ведет к синхронизованному потоку. Было постулировано, что вероятность обгона, которая совпадает с вероятностью «переускорения», является разрывной функцией плотности (рис. 6): при данной плотности АТС вероятность обгона в свободном потоке много больше, чем в синхронизированном потоке.

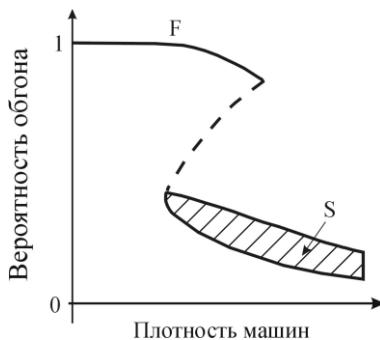


Рис. 6. Объяснение фазового перехода к плотному потоку (traffic breakdown) на основе Z-образной нелинейной функции вероятности обгона (вероятности «переускорения») в теории Б. С. Кернера. Пунктирная линия описывает критическое значение вероятности обгона как функцию плотности АТС (из [1])

Разрывная функция вероятности обгона является одной и той же как для спонтанного, так и для индуцированного $F \rightarrow S$ фазового перехода: термины *спонтанный* и *индуцированный* отличаются только источником возмущения, приводящего к $F \rightarrow S$ фазовому переходу. $F \rightarrow S$ переход происходит при условии, что вероятность обгона внутри возмущения в свободном потоке меньше, чем критическая вероятность. Эта критическая вероятность показана пунктирной линией на рис. 6. Другими словами, не имеет значения, будет ли это критическое значение вероятности обгона достигнуто благодаря возмущению в свободном потоке (спонтанный переход) или благодаря распространению до узкого места некоторого возмущения, возникшего ранее в другой области дороги (индуцированный переход).

Отметим, что $F \rightarrow S$ фазовый переход и обратный $S \rightarrow F$ фазовый переход сопровождаются гистерезисом. Этот гистерезис не имеет никакого отношения к хорошо известному гистерезису в математических моделях ДМ-класса, по-видимому, впервые найденному в теории Кер-

нера–Конхойзера. Этот известный гистерезис, описанный в огромном количестве математических работ (смотри работы [19–22; 26, 27] и ссылки в них), описывает $F \rightarrow J$ и обратный $J \rightarrow F$ фазовый переход. Как уже несколько раз отмечалось, спонтанный $F \rightarrow J$ переход не наблюдается в реальном транспортном потоке.

3.6. Бесконечное число пропускных способностей скоростной автомагистрали

Спонтанное образование плотного потока, т.е. спонтанный $F \rightarrow S$ фазовый переход может произойти в широком диапазоне значений величины потока q в свободном транспортном потоке. Основываясь на эмпирических данных измерений, был сделан вывод, что существует бесконечное число значений пропускной способности автомагистрали в свободном потоке. Это бесконечное число значений пропускной способности находится в диапазоне между минимальным q_{th} и максимальным q_{max} значениями пропускной способности (см. рис. 7). Если величина потока близка к максимальному значению пропускной способности q_{max} , то уже достаточно малое возмущение в свободном потоке вблизи узкого места приведет к спонтанному $F \rightarrow S$ фазовому переходу. С другой стороны, если величина потока близка к минимальному значению пропускной способности q_{th} , то только возмущение очень большой амплитуды способно привести к спонтанному $F \rightarrow S$ фазовому переходу.

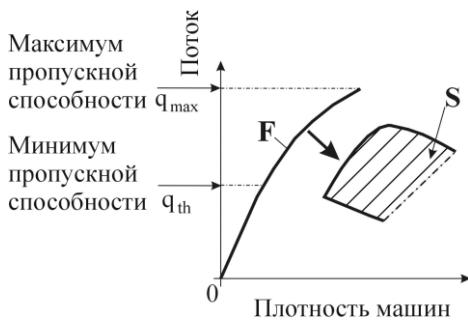


Рис. 7. Максимум и минимум пропускной способности скоростной автомагистрали в теории трех фаз. Взято из [1]

Вероятность возникновения малых возмущений в свободном транспортном потоке много выше, чем вероятность возникновения возмущений большой амплитуды. По этой причине, чем выше величина потока q в свободном потоке вблизи узкого места, тем выше вероятность спонтанного F→S фазового перехода.

Если величина потока q меньше, чем минимальная пропускная способность q_{th} , то возникновение плотного потока (F→S переход) невозможно.

Бесконечное число значений пропускной способности автомагистрали вблизи узкого места может быть объяснено тем, что свободный поток при значениях величины потока q в диапазоне

$$q_{th} \leq q \leq q_{\max}$$

является метастабильным. Это означает, что при возникновении малых возмущений свободный поток сохраняется, т.е. является устойчивым относительно малых возмущений. Однако для больших возмущений свободный поток оказывается неустойчивым и происходит F→S фазовый переход к синхронизованному потоку.

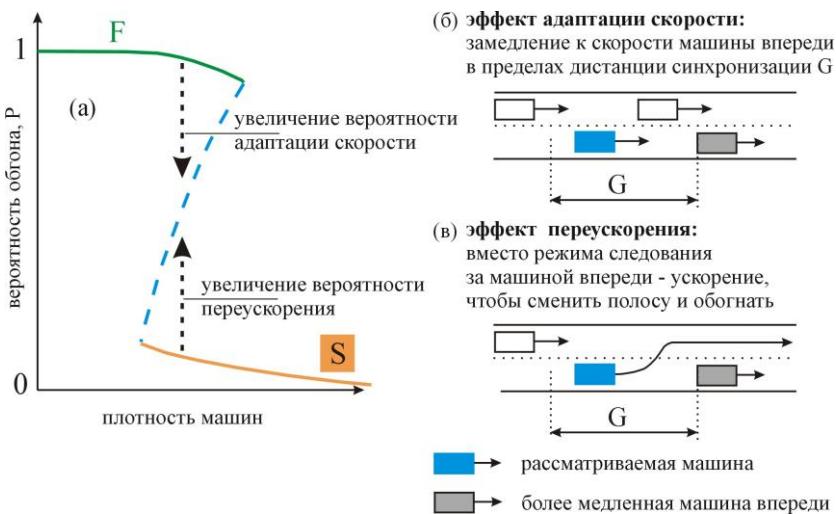


Рис. 8. Пояснение соревнования между переускорением и адаптацией скорости, которое объясняет бесконечное число значений пропускной способности автомагистрали. Взято из [1]

Как уже упоминалось в п. 3.5, природу $F \rightarrow S$ фазового перехода можно объяснить с помощью «соревнования» во времени и пространстве двух противоположных процессов: ускорения АТС при обгоне более медленного АТС впереди, названный «переускорением», и в случае, когда обгон невозможен, торможения АТС до скорости более медленного АТС, названном «адаптация скорости». На рис. 8 поясняется это соревнование более детально. На рис. 8 (а), который соответствует рис. 6, стрелочка вниз означает, что если в свободном потоке вблизи узкого места возникает локальное уменьшение скорости АТС, то вероятность обгона внутри этого возмущения падает. Если это уменьшение вероятности обгона становится меньше, чем критическая величина вероятности обгона, показанная пунктирной линией на рис. 8 (а), то $F \rightarrow S$ фазовый переход происходит внутри этого возмущения; в противоположном случае возмущение затухает и свободный поток остается на узком месте. Стрелочка вверх на рис. 8 (а) означает, что если исходное состояние отвечает фазе синхронизированного потока и в этом состоянии возникает случай локального увеличения скорости АТС, то внутри данного возмущения вероятность обгона возрастает. Если это возрастание вероятности обгона превышает критическое значение (как выше отмечалось, эта критическая вероятность обгона отвечает пунктирной кривой на рис. 8 (а)), то в области возмущения происходит $S \rightarrow F$ переход; в противоположном случае возмущение затухает и остается синхронизированный поток.

Бесконечное число значений пропускной способности автомагистрали вблизи узкого места в теории трех фаз фундаментально противоречит классическим теориям транспортного потока (а также методам управления и автоматического регулирования транспортными потоками), которые предполагают существование в любой момент времени некоторой (фиксированной или случайной) пропускной способности.

3.7. Широкие движущиеся кластеры (локальные движущиеся заторы) – фаза J

Широкий движущийся кластер может быть назван широким только при условии, что его ширина (вдоль дороги) заметно превышает ширину фронтов кластера. Средняя скорость движения АТС внутри широкого движущегося кластера много меньше, чем скорость АТС в свободном потоке. На заднем (в направлении потока) фронте кластера АТС могут ускоряться вплоть до свободного потока. На переднем

фронте кластера АТС, подъезжающие к фронту, должны сильно уменьшать свою скорость. Согласно определению [J], широкий движущийся кластер обычно сохраняет среднюю скорость заднего фронта v_g , даже если кластер проходит через другие фазы транспортного потока и узкие места. Величина потока сильно падает внутри широкого движущегося кластера.

Как отмечалось в п. 2.4, эмпирические результаты показывают, что характеристические параметры широких движущихся кластеров не зависят от величины потока на дороге и особенностей узкого места (где и когда кластер возник). Однако эти характеристические параметры зависят от погоды, дорожных условий, конструктивных характеристик АТС, процента длинных машин и т.п. Скорость заднего фронта широкого движущегося кластера v_g в противоположном потоку направлении является характеристическим параметром, так же как и величина выходного потока q_{out} из кластера в случае, когда свободный поток формируется после кластера (рис. 9).

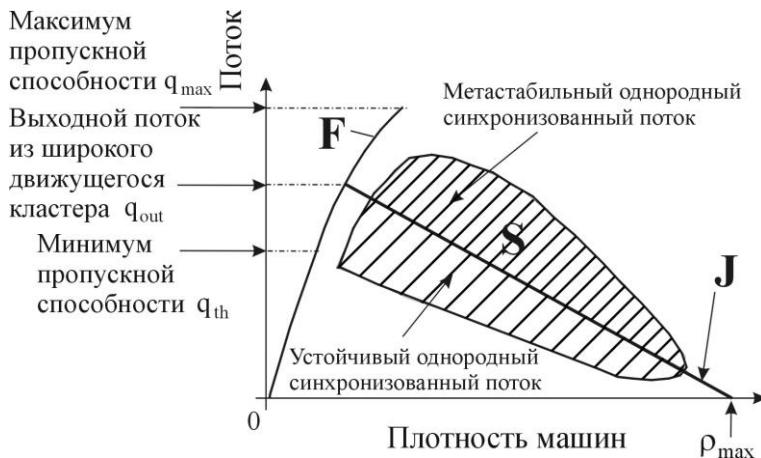


Рис. 9. Три фазы транспортного потока в плоскости поток-плотность в теории трех фаз Кернера. Взято из [1]

Это означает, что разные широкие движущиеся кластеры имеют одинаковые параметры при одинаковых условиях. Благодаря этому эти параметры могут быть предсказаны. Движение заднего фронта широкого

движущегося кластера может быть показано на плоскости поток–плотность с помощью прямой, называемой линия J Кернера (рис. 9). Наклон линии J Кернера равен скорости заднего фронта v_g , в то время как координата пересечения линии J Кернера с осью абсцисс (при нулевом потоке) отвечает плотности АТС ρ_{\max} в широком движущемся кластере (о линии J Кернера более подробно смотри в п. 2.4.3).

Подчеркнем, что минимум пропускной способности q_{th} и величина выходного потока из широкого движущегося кластера q_{out} описывают два *качественно различных свойства* свободного транспортного потока. Минимум пропускной способности q_{th} характеризует F→S фазовый переход вблизи узкого места, т.е. возникновение плотного потока (traffic breakdown). В свою очередь величина выходного потока из широкого движущегося кластера q_{out} характеризует условия существования таких кластеров, т.е. фазы J. В зависимости от внешних условий, таких как погода, процент длинных машин в потоке и т.п., а также от характеристик узкого места, вблизи которого может произойти F→S фазовый переход, минимум пропускной способности q_{th} может быть как меньше (рис. 9), так и больше, чем величина выходного потока q_{out} . Важно, что величина выходного потока из широкого движущегося кластера q_{out} оказывается меньше, чем максимальный поток q_{\max} в свободном потоке перед кластером. Это означает, что в свободном потоке водители могут выбирать более короткую временную дистанцию до АТС впереди, чем та дистанция, которую они принимают, ускоряясь на заднем фронте широкого движущегося кластера.

3.8. Синхронизованный транспортный поток – фаза S

В отличие от широких движущихся кластеров в синхронизованном потоке как величина потока q , так и скорость АТС могут меняться заметным образом. Задний по направлению потока фронт синхронизованного потока часто фиксирован в пространстве (см. определение [S]), обычно вблизи расположения узкого места. Величина потока q в фазе синхронизованного потока может оставаться почти такой же, как и в свободном потоке, даже если скорость АТС сильно уменьшается.

Поскольку синхронизованный поток не имеет характеристического свойства [J] фазы широкого движущегося кластера J, в теории

трех фаз предполагается, что гипотетические однородные состояния синхронизованного потока покрывают двумерную область в плоскости поток–плотность (см. заштрихованную область на рис. 9).

3.9. S→J фазовый переход

Широкие движущиеся кластеры не возникают в свободном потоке, но они могут возникать в области синхронизированного потока. Этот фазовый переход называется S→J фазовый переход. Эмпирический пример S→J перехода показан на рис. 10. Таким образом, образование широких движущихся кластеров в свободном потоке наблюдается в результате каскада F→S→J фазовых переходов: сначала, область синхронизированного потока возникает внутри свободного потока.

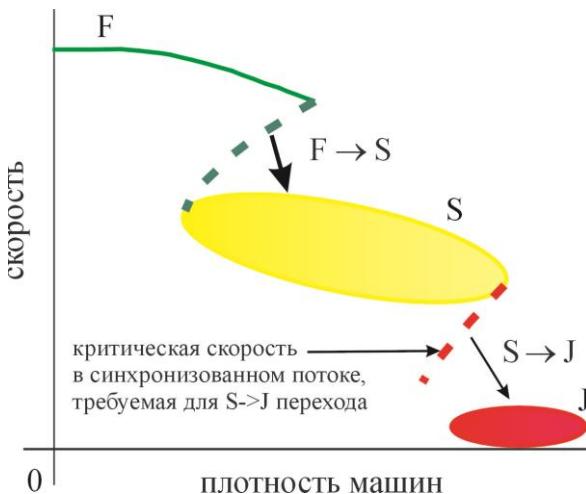


Рис. 10. Двойная Z-характеристика в теории трех фаз, поясняющая каскад F→S→J фазовых переходов. Взято из [1]

Как было объяснено выше, такой F→S фазовый переход происходит в большинстве случаев вблизи узкого места. Далее внутри синхронизированного потока происходит «сжатие» потока, т.е. плотность АТС возрастает, в то время как их скорость падает. Это сжатие называется «пинч» эффект. В области синхронизированного потока, где происходит пинч эффект, возникают узкие движущиеся кластеры. Было пока-

зано, что частота возникновения узких движущихся кластеров тем выше, чем выше плотность в синхронизованном потоке. По мере того как эти узкие движущиеся кластеры нарастают, некоторые из них трансформируются в широкие движущиеся кластеры, другие же исчезают. Широкие движущиеся кластеры в дальнейшем распространяются против потока, проходя через все области синхронизированного потока и через все узкие места.

Чтобы детальнее проиллюстрировать $S \rightarrow J$ фазовый переход, следует заметить, что в теории трех фаз линия J делит все однородные состояния синхронизированного потока на две области (рис. 9). Состояния выше линии J Кернера являются метастабильными относительно образования широких движущихся кластеров, в то время как состояния ниже линии J Кернера являются устойчивыми. Метастабильные состояния синхронизированного потока означают, что относительно малых возникающих возмущений состояние потока остается устойчивым, однако при больших возмущениях в синхронизированном потоке происходит $S \rightarrow J$ фазовый переход.

Каскад $F \rightarrow S \rightarrow J$ фазовых переходов можно пояснить на основе двойной Z характеристики в теории трех фаз (рис. 10). Пунктирная линия между фазой F и фазой S качественно соответствует критической скорости внутри локального возмущения свободного потока, при которой происходит $F \rightarrow S$ фазовый переход. Другими словами, $F \rightarrow S$ фазовый переход происходит внутри локального возмущения свободного потока, в котором скорость меньше, чем критическая скорость (символически этот фазовый переход показан стрелкой между фазой F и S на рис. 10). Пунктирная линия между фазой S и фазой J качественно соответствует критической скорости внутри локального возмущения синхронизированного потока, при которой происходит $S \rightarrow J$ фазовый переход. Другими словами, $S \rightarrow J$ фазовый переход происходит внутри локального возмущения синхронизированного потока, в котором скорость меньше, чем критическая скорость (символически этот фазовый переход показан стрелкой между фазой S и J на рис. 10).

3.10. Неоднородные пространственно-временные структуры транспортного потока, состоящие из фаз S и J

В эмпирических данных можно наблюдать очень сложные пространственно-временные структуры в плотном транспортном потоке, образовавшиеся в результате $F \rightarrow S$ и $S \rightarrow J$ фазовых переходов.

Неоднородная пространственно-временная структура, которая состоит только из синхронизованного потока, называется *структурой синхронизированного потока* (СП). Когда задний фронт СП фиксирован вблизи узкого места на дороге, а передний фронт не распространяется против потока, такая СП называется *локализованной структурой синхронизированного потока* (ЛСП). Однако во многих случаях передний фронт структуры синхронизированного потока распространяется в направлении против потока. Если при этом задний фронт по-прежнему остается фиксированным вблизи узкого места, то ширина области синхронизированного потока увеличивается. Такая структура называется расширяющейся структурой синхронизированного потока (РСП). Возможна также ситуация, когда задний фронт синхронизированного потока уже не фиксирован вблизи узкого места, а оба фронта синхронизированного потока движутся в направлении против потока. Такая структура называется бегущей, или мигрирующей структурой синхронизированного потока (МСП).

Разница между пространственно-временными структурами, состоящими из только синхронизированного потока, и широкими движущимися кластерами становится особенно ясной, когда РСП или МСП достигают следующего узкого места, расположенного вверх по течению транспортного потока. В этом случае структура синхронизированного потока «захватывается» на этом узком месте (так называемый «*catch-effect*» в английской терминологии), и возникает новая пространственно-временная структура в транспортном потоке. Напротив, широкий движущийся кластер не захватывается вблизи узкого места, а распространяется дальше против потока, т.е. пробегая через узкое место на дороге. Кроме того, в отличие от широкого движущегося кластера структура синхронизированного потока, даже если она распространяется в виде МСП, не имеет характеристических параметров. В результате скорость заднего фронта МСП может заметно меняться в процессе распространения, и эта скорость может быть разной у разных МСП. Данные особенности структур синхронизированного потока и широких движущихся кластеров вытекают из определения фаз [S] и [J]. Наиболее типичная пространственно-временная структура плотного транспортного потока состоит из обеих фаз S и J. Такая структура называется *общей структурой плотного потока* (ОП).

На многих скоростных автомагистралях узкие места, связанные с въездами/выездами, располагаются очень близко друг к другу. Пространственно-временная структура, в которой синхронизированный поток охватывает два и более узких места, называется *единой структурой плотного потока* (ЕП). ЕП может состоять только из синхронизирован-

го потока, тогда она называется ЕСП (единая структура синхронизованного потока). Однако обычно широкие движущиеся кластеры возникают в синхронизованном потоке. В этом случае ЕП называется ЕОП (*единая общая структура плотного потока*) (см. рис. 11).

В данной главе были рассмотрены основные качественные положения теории трех фаз Кернера. Эти качественные положения, начиная с 2002 года, были использованы как теоретический базис при создании целого ряда микроскопических и макроскопических трехфазных моделей транспортного потока (см. ссылки на оригинальные работы в главе 11 книги [1]). В следующем пункте этой главы дается краткий обзор стохастической микроскопической модели в рамках теории трех фаз и приведены некоторые результаты численных расчетов.

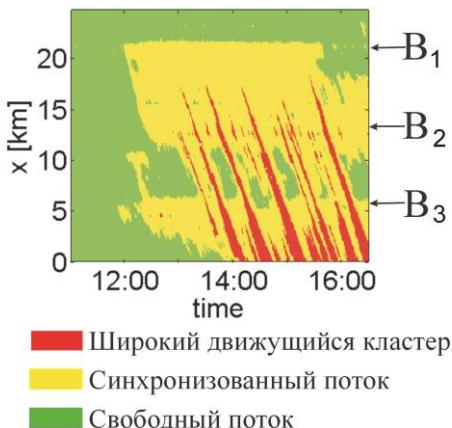


Рис. 11. ЕОП измеренная на магистрали с тремя узкими местами B_1 , B_2 , B_3 . Взято из [1]

3.11. Стохастическая модель в рамках теории трех фаз Кернера

Теория трех фаз Кернера является качественной теорией. Различные математические трехфазные модели транспортных потоков были разработаны в последние годы в рамках теории трех фаз. Впервые микроскопическая трехфазная модель, которая может воспроизводить эм-

тические свойства перехода к плотному потоку (traffic breakdown) и результирующих пространственно-временных структур, была разработана Б. С. Кернером и С. Л. Кленовым в 2002 году [28]. Несколькоими месяцами позже Б. С. Кернер, С. Л. Кленов и Д. Вольф предложили трехфазную модель на основе клеточных автоматов (ККВ-модель) [29]. Позднее были также разработаны другие модели транспортного потока в рамках теории трех фаз: Л. Дэвис [30], а также Б. С. Кернер и С. Л. Кленов [31] предложили детерминистические микроскопические модели трех фаз; Х. Ли и М. Шрекенберг с соавторами [32], Р. Янг и К. Ву [33] и К. Гао с соавторами [34] разработали различные трехфазные модели клеточных автоматов (КА); Дж. Лаваль [35] и С. Хугендорн с соавторами [36] разработали макроскопические модели трех фаз. Последние результаты моделирования транспортного потока, проведенные различными научными группами в США, Германии, Голландии, Китае, Южной Корее и Японии в рамках теории трех фаз, можно найти в [37–60].

В данном пункте кратко рассматривается стохастическая микроскопическая трехфазная модель транспортного потока, предложенная Б. С. Кернером и С. Л. Кленовым в [28, 61]. В этой модели [61] использовалось дискретное время с шагом τ , в то время как перемещение в пространстве предполагалось непрерывным. Ниже будет рассмотрена дискретная версия модели [61], сформулированная в [58], в которой используется достаточно мелкая дискретизация пространства с шагом δx (см. также формулировку дискретной версии модели в [59]). При этом в приведенных ниже формулах координата измеряется в единицах δx , в то время как скорость АТС и ее ускорение измеряются соответственно в единицах $\delta v = \delta x / \tau$ и $\delta a = \delta v / \tau$, где временной шаг $\tau = 1$ с.

Уравнения движения АТС в дискретной версии [58, 59] стохастической трехфазной модели [61] транспортного потока на скоростной 2-полосной автодороге в приближении идентичных АТС задаются следующими формулами:

$$v_{n+1} = \max(0, \min(v_{free}, \tilde{v}_{n+1} + \xi_n, v_n + a\tau, v_{s,n})), \quad x_{n+1} = x_n + v_{n+1}\tau, \quad (1)$$

$$\tilde{v}_{n+1} = \max(0, \min(v_{free}, v_{s,n}, v_{c,n})), \quad (2)$$

$$v_{c,n} = \begin{cases} v_n + \Delta_n, & \text{если } g_n \leq G_n, \\ v_n + a_n \tau, & \text{если } g_n > G_n, \end{cases} \quad (3)$$

$$\Delta_n = \max(-b_n \tau, \min(a_n \tau, v_{\ell,n} - v_n)), \quad (4)$$

где индекс n отвечает дискретному времени $t = n\tau$, $n = 0, 1, 2, \dots$; x_n и v_n – координата и скорость АТС на временном шаге n ; v_{free} – макси-

мальная скорость АТС в свободном потоке; $g_n = x_{\ell,n} - x_n - d$ – расстояние до АТС впереди, индекс ℓ относится ко всем переменным и функциям, описывающим АТС впереди, d – длина АТС, которая предполагается одинаковой и включает в себя также среднее расстояние между АТС, когда они стоят внутри широкого движущегося кластера; \tilde{v}_n – величина скорости АТС без шумовой компоненты ξ_n ; $v_{s,n}$ – безопасная скорость, определенная ниже; $a_n \geq 0$ и $b_n \geq 0$; a – максимальное ускорение; G_n – максимальное расстояние, на котором водитель синхронизирует свою скорость со скоростью АТС впереди (так называемая дистанция «синхронизации скорости»);

$$G_n = G(v_n, v_{\ell,n}), \quad (5)$$

где функция (5):

$$G(u, w) = \max \left(0, \left\lfloor ktu + (u-w)ua^{-1} \right\rfloor \right), \quad (6)$$

константа $k > 1$, $\lfloor z \rfloor$ – означает целую часть действительного числа z .

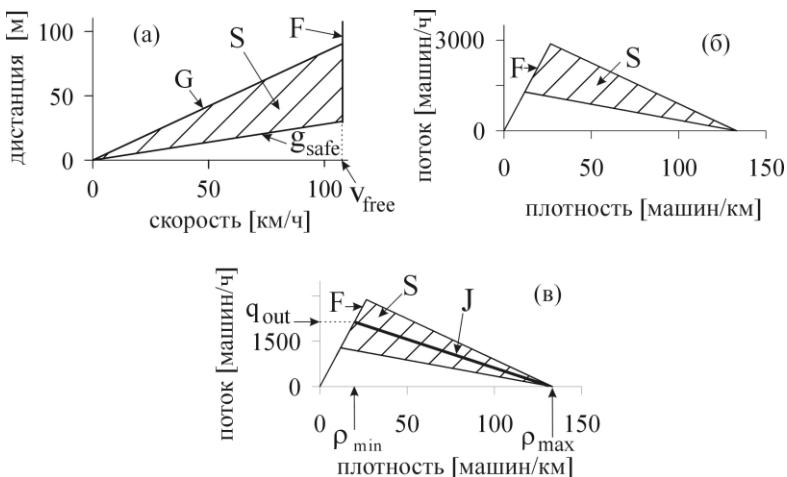


Рис. 12. Стационарные однородные состояния и линия J Кернера в стохастической трехфазной модели: (а, б) свободный поток (F) и синхронизованный поток (заштрихованная двумерная область, обозначенная буквой S) на плоскости расстояние между АТС–скорость (а) и на плоскости поток–плотность (б). На (в) показано соответствие между линией J и однородными стационарными состояниями модели на плоскости поток–плотность, взятыми из (б)

Решения модели (1) – (6) при стационарном и однородном движении АТС (которое отвечает потоку, где АТС находятся на одинаковом расстоянии друг от друга и движутся с постоянной скоростью) показаны на рис. 12 (а) и (б). В соответствии с фундаментальной гипотезой теории трех фаз (рис. 9) фаза синхронизованного потока (буква S) как на плоскости расстояние–скорость (рис. 12 (а)), так и на плоскости поток–плотность (рис. 12 (б)) покрывает двумерную область (заштрихованная область на рис. 12). Для того чтобы пояснить смысл этой двумерной области для стационарных состояний синхронизированного потока, рассмотрим более подробно рисунок 12 (а). Видно, что при некоторой заданной величине скорости $v < v_{free}$ в стационарном состоянии существует бесконечное количество расстояний между АТС в диапазоне $g_{safe} \leq g \leq G$, где g – расстояние между АТС, $G = G(v, v)$ – дистанция синхронизации скорости (5), взятая при одинаковых скоростях АТС, g_{safe} – безопасное расстояние между АТС, которое является решением уравнения $v = v_{safe}(g_{safe})$, где безопасная скорость v_s в стационарном однородном состоянии определяется как $v_s(g) = g/\tau$. Таким образом, стационарные однородные состояния стохастической трехфазной модели (рис. 12) принципиально отличаются от таких же состояний большинства предшествующих моделей, в которых заданной скорости отвечает одно единственное расстояние между АТС, соответствующее точке на фундаментальной диаграмме.

Из рис. 12 (в) также можно видеть, что линия J Кернера не имеет никакого отношения к фундаментальной диаграмме стационарного однородного плотного потока в моделях, основанных на фундаментальной диаграмме. Действительно, в трехфазной стохастической модели линия J, которая находится из стационарного движения широкого кластера в этой модели, разделяет двумерную область стационарных однородных состояний синхронизованного потока на две части. Оказывается, что в соответствии с теорией трех фаз стационарные однородные состояния выше линии J являются метастабильными к образованию широких движущихся кластеров, а состояния ниже линии J являются устойчивыми к образованию широких кластеров.

Чтобы моделировать случайное время задержки водителя как при ускорении, так и при замедлении в различных транспортных ситуациях величины a_n и b_n в (4) и (5) задаются как случайные функции:

$$a_n = a\theta(P_0 - r_1), \quad b_n = a\theta(P_1 - r_1),$$

$$P_0 = \begin{cases} p_0, & \text{если } S_n \neq 1, \\ 1, & \text{если } S_n = 1, \end{cases} \quad P_1 = \begin{cases} p_1, & \text{если } S_n \neq -1, \\ p_2, & \text{если } S_n = -1, \end{cases}$$

где величины $1-P_0$ и $1-P_1$ представляют собой вероятности для времени задержки водителя соответственно при ускорении и при замедлении АТС; $p_0(v)$ и $p_2(v)$ являются функциями скорости АТС, p_1 – константа; $r = \text{rand}(0,1)$ представляет собой случайную величину, равномерно распределенную в интервале от 0 до 1.

Случайная компонента ξ_n в формуле (1) описывает случайное замедление или ускорение и применяется в зависимости от того, тормозит ли АТС, или ускоряется, или не меняет свою скорость:

$$\xi_n = \begin{cases} -\xi_b, & \text{если } S_{n+1} = -1, \\ \xi_a, & \text{если } S_{n+1} = 1, \\ \xi^{(0)}, & \text{если } S_{n+1} = 0, \end{cases}$$

где S – это состояние движения АТС в отсутствие случайной компоненты ξ_n ,

$$S_{n+1} = \begin{cases} -1, & \text{если } \tilde{v}_{n+1} < v_n, \\ 1, & \text{если } \tilde{v}_{n+1} > v_n, \\ 0, & \text{если } \tilde{v}_{n+1} = v_n, \end{cases}$$

ξ_a , ξ_b являются случайными источниками соответственно для ускорения и замедления АТС:

$$\xi_a = a^{(a)} \tau \theta(p_a - r), \quad \xi_b = a \tau \theta(p_b - r),$$

а случайный источник

$$\xi^{(0)} = a^{(0)} \tau \begin{cases} -1, & \text{если } r < p^{(0)}, \\ 1, & \text{если } p^{(0)} \leq r < 2p^{(0)} \text{ и } v_n > 0, \\ 0 & \text{в остальных случаях} \end{cases}$$

применяется в отсутствие ускорения или замедления и связан с невозможностью точно поддерживать заданную скорость. Величины p_a и p_b являются вероятностями соответственно случайного ускорения или торможения АТС, $p^{(0)}$ и $a^{(0)}$ – константы, $r = \text{rand}(0,1)$, ступенчатая функция (Хевисайда) $\theta(z)$ определяется как

$$\theta(z) = 0 \text{ при } z < 0 \text{ и } \theta(z) = 1 \text{ при } z \geq 0.$$

Безопасная скорость $v_{s,n}$ определяется следующим образом:

$$v_{s,n} = \min\left(v_n^{(safe)}, v_\ell^{(a)} + \frac{g_n}{\tau}\right), \text{ где}$$

$$v_\ell^{(a)} = \max\left(0, \min\left(v_{\ell,n}^{(safe)}, v_{\ell,n}, g_{\ell,n}/\tau\right) - a\tau\right)$$

– это так называемая «ожидаемая» (прогнозируемая) скорость АТС впереди, функция $v_n^{(safe)} = \lfloor v^{(safe)}(g_n, v_{\ell,n}) \rfloor$ задается безопасной скоростью $v^{(safe)}(g_n, v_{\ell,n})$, которая была предложена в модели С. Краусса с соавторами [62] в 1997 году и которая в свою очередь является решением уравнения П. Гиппса [63]:

$$v^{(safe)}\tau + X_d(v^{(safe)}) = g_n + X_d(v_{\ell,n}),$$

где $X_d(u)$ – тормозной путь, проходимый АТС, движущимся с первоначальной скоростью u и тормозящим с постоянным ускорением b вплоть до полной остановки. В модели с дискретным временем этот путь дается формулой $X_d(u) = bt^2 \cdot (\alpha\beta + 0.5(\alpha-1)\alpha)$, α и β – соответственно целая и дробная части величины u/bt .

В рассматриваемой модели автодороги с двумя полосами смена полосы АТС происходит независимо от того, находятся ли эти АТС вдали или вблизи неоднородностей дороги, связанной с въездом–выездом. АТС меняет полосу, если некоторые необходимые условия для перехода с правой полосы на левую ($R \rightarrow L$) или с левой полосы на правую ($L \rightarrow R$) выполняются совместно с условиями безопасности при смене полосы. Необходимые для смены полосы условия имеют вид

$$R \rightarrow L: v_n^+ \geq v_{\ell,n} + \delta_1 \text{ и } v_n \geq v_{\ell,n},$$

$$L \rightarrow R: v_n^+ > v_{\ell,n} + \delta_1 \text{ или } v_n^+ > v_n + \delta_1.$$

Условия безопасности при смене полосы имеют вид

$$g_n^+ > \min(v_n\tau, G_n^+), \quad g_n^- > \min(v_n^-\tau, G_n^-), \text{ где} \quad (7)$$

$$G_n^+ = G(v_n, v_n^+), \quad G_n^- = G(v_n^-, v_n),$$

верхние индексы "+" и "-" относятся соответственно к АТС впереди и позади на соседней полосе.

Если условия (7) не выполняются, то используются более жесткие условия для «вдавливания» АТС на соседнюю полосу:

$$x_n^+ - x_n^- - d > g_{target}^{(min)}, \text{ где } g_{target}^{(min)} = \lfloor \lambda v_n^+ + d \rfloor. \quad (8)$$

В дополнение к (8) используется условие, что АТС проходит среднюю точку $x_n^{(m)} = \lfloor (x_n^+ + x_n^-)/2 \rfloor$ между двумя соседними АТС на соседней полосе, т.е. следующие условия выполняются:

$$x_{n-1} < x_{n-1}^{(m)} \text{ и } x_n \geq x_n^{(m)} \text{ или } x_{n-1} \geq x_{n-1}^{(m)} \text{ и } x_n < x_n^{(m)}.$$

Если условия для смены полосы выполняются, АТС меняет полосу с вероятностью $p_c < 1$ на текущем шаге.

После смены полосы скорость v_n устанавливается равной $\hat{v}_n = \min(v_n^+, v_n^- + \Delta v^{(1)})$, что описывает изменение скорости после маневра по смене полосы. После смены полосы координата АТС не меняется, если выполнены условия (7), и она устанавливается равной $x_n = x_n^{(m)}$, если выполнены условия (8). Величины $\Delta v^{(1)}$, p_c , δ_1 и λ являются константами. Более подробно об использованных параметрах модели, условиях смены полос, и модели поведения водителя на въезде/выезде со скоростной автодороги можно прочитать в главе 11 книги [1].

На рис. 13 приведены результаты расчета пространственно-временных структур, возникающих вблизи въезда на автодорогу и областей их существования (диаграммы) на плоскости с координатами поток по основной дороге q_{in} , приведенный на одну из двух полос дороги, и поток со стороны въезда на дорогу q_{on} . На рис. 13 (а) показана диаграмма этих пространственно-временных структур. По оси абсцисс этой диаграммы откладывается поток АТС q_{on} , по оси ординат указан поток АТС по основной дороге q_{in} . Граница $F_s^{(B)}$ на этой диаграмме разделяет свободный поток влево от границы от структур плотного потока, возникающих вблизи въезда АТС. На границе $F_s^{(B)}$ пространственно-временные структуры плотного потока возникают спонтанно. Граница $S_j^{(B)}$ разделяет структуры синхронизованного потока (СП) от общей структуры плотного потока (ОП). Это означает, что между границами $F_s^{(B)}$ и $S_j^{(B)}$ возникают различного типа СП. При больших q_{in} и маленьких q_{on} возникает *мигрирующая структура синхронизованного потока – МСП* (рис. 13 (г)). При увеличении q_{on} МСП превращается в *расширяющуюся структуру синхронизованного потока – РСП* (рис. 13 (б)).

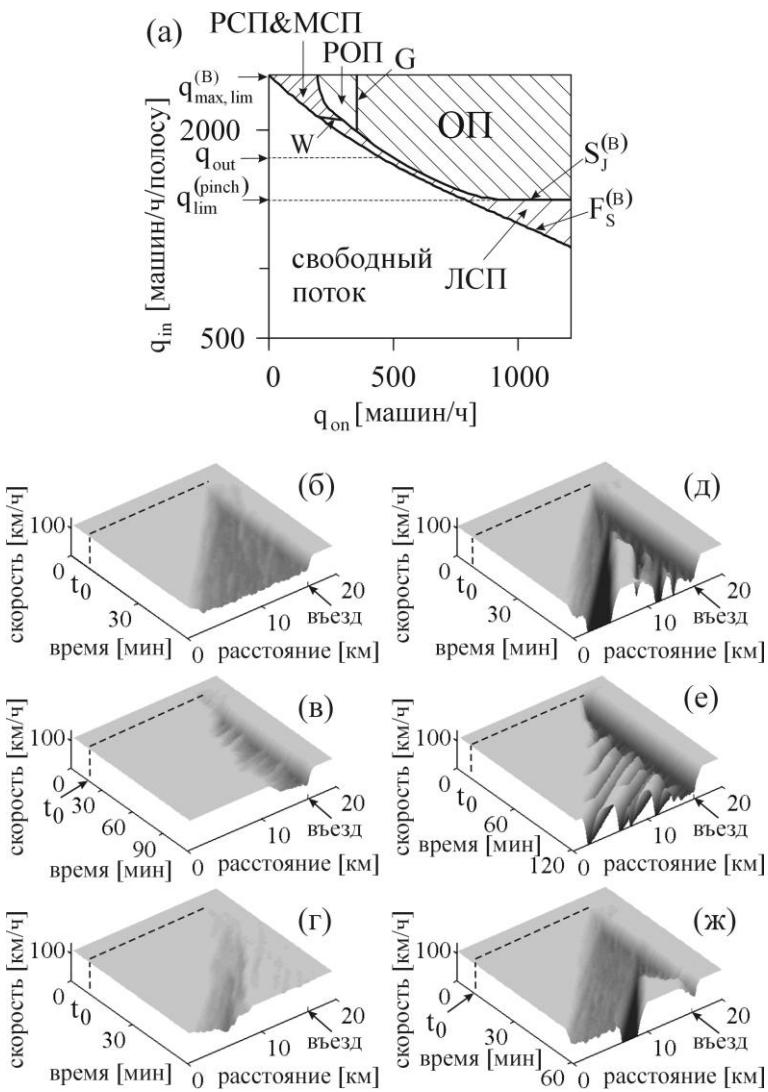


Рис. 13. Диаграмма пространственно-временных структур транспортного потока близи изолированного въезда (а) и соответствующие типы пространственно-временных структур (б–ж), относящиеся к диаграмме (а): СП (б–г) и ОП (д–ж). Взято из [2]

Если уменьшать q_{in} и увеличивать q_{on} , то на границе, обозначенной буквой W, РСП превращается в *локализованную структуру синхронизованного потока* – ЛСП (рис. 13 (в)).

Укажем еще некоторые особенности общей структуры плотного потока – ОП, которая возникает выше границы $S_j^{(B)}$ на диаграмме (рис. 13 (а)). Существует граница, обозначенная буквой G внутри области структур ОП. Слева от этой границы, после того как широкий кластер возникает в синхронизованном потоке, новые широкие кластеры больше не формируются (рис. 13 (ж)), и возникает *рассасывающаяся общая структура плотного потока* (РОП). Правее от границы G возникает общая структура, в которой непрерывно рождаются новые кластеры внутри синхронизированного потока (рис. 13 (д)). Однако если уменьшать поток по дороге q_{in} , оставаясь в области ОП, и перейти к значениям q_{in} меньше, чем q_{out} (смысл потока q_{out} объяснен в п. 2.4.3), то широкие движущиеся кластеры, спонтанно возникающие в синхронизованном потоке ОП, постепенно рассасываются, распространяясь против течения. В результате возникает ОП, показанная на рис. 13 (е). Другая особенность общей структуры состоит в том, что если поток из въезда q_{on} дальше увеличивать, то возникает некоторый эффект насыщения, связанный с возникновением плотного потока на дороге, ведущей к въезду АТС. Этот эффект насыщения связан с тем, что поток внутри области синхронизированного потока в ОП достигает своего предельного минимального значения, обозначенного на рис. 13 (а) как $q_{lim}^{(pinch)}$.

Распределение скорости и потока внутри ОП структуры на рис. 13 (д) как функция времени при разных координатах вдоль дороги показано на рис. 14. Можно видеть, что сначала на 16.5 км возникает F→S фазовый переход из свободного в синхронизированный поток. Через 2 км против течения ($x = 14.5$ км) можно видеть развивающиеся узкие кластеры внутри синхронизированного потока. По мере распространения этих кластеров против течения амплитуда кластеров увеличивается и их ширина возрастает. В результате нарастания кластеров возникает последовательность широких движущихся кластеров ($x = 8$ км), которые обладают характеристическими параметрами, описанными в п. 2.4.2. В частности, распространение заднего фронта этих кластеров соответствует линии J Кернера, описанной в п. 2.4.3.

Более подробно свойства пространственно-временных структур, полученные в результате моделирования, рассматриваются в [1] и [2].

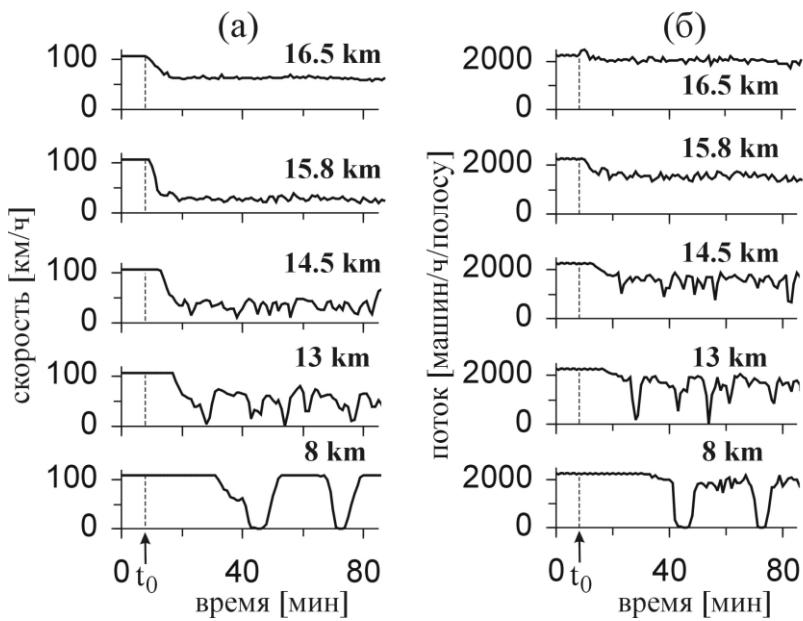


Рис. 14. Одноминутные данные виртуальных детекторов, отвечающие ОП на рис. 13 (д). Въезд на скоростную дорогу отвечает координате 16 км. Взято из [2]

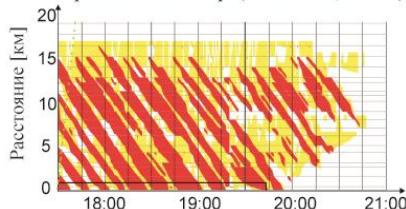
3.12. Применение теории трех фаз Кернера для интеллектуальных транспортных технологий

Б. С. Кернер с сотрудниками предложил и частично внедрил в эксплуатацию целый ряд новых методов интеллектуальных транспортных технологий. Одним из внедренных и уже установленных на скоростных автодорогах применений теории трех фаз является метод ASDA/FOTO. Метод ASDA/FOTO функционирует в работающей онлайн системе регулирования транспортных потоков, где на основе измерений выделяются фазы S и J в плотном транспортном потоке. Распознавание, отслеживание и прогнозирование положений фаз S и J осуществляется на основе методов теории трех фаз. Метод ASDA/FOTO реализован в компьютерной системе, способной быстро и эффективно обрабатывать большие объемы данных, измеренных датчиками в сети скоростных автомагистралей (см. примеры из трех стран на рис. 15).

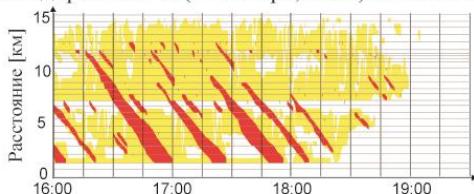
Дальнейшее развитие приложений теории трех фаз Кернера связано с разработкой и усовершенствованием моделей для транспортных симуляторов, методов регулирования въездного потока на автомагистраль (ANCONA), методов коллективного регулирования транспортных потоков, системы автоматического ассистента водителя и методов детектирования состояния транспортного потока, описанных в монографиях [1, 2].

Единая общая структура плотного потока (ЕОП)

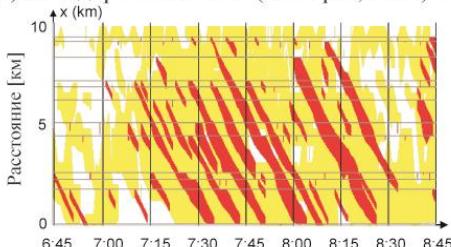
(а) Автодорога A5-Север (14 июня, 2006) в Германии



(б) Автодорога M-42 (11 января, 2008) в Великобритании



(в) Автодорога I405-IОг (04 марта, 2003) в США



■ Широкий движущийся кластер
■ Синхронизированный поток

Рис. 15. Пространственно-временная структура транспортного потока, полученная методом ASDA/FOTO в трех странах. Взято из [1]

Литература

1. *Kerner B. S.* Introduction to Modern Traffic Flow Theory and Control. Berlin: Springer, 2009.
2. *Kerner B. S.* The Physics of Traffic. Berlin: Springer, 2004.
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. 1955. V. 229. P. 281–345.
4. *Richards P. I.* Shock Waves on the Highway // Oper. Res. 1956. V. 4. P. 42–51.
5. Узем Дж. Линейные и нелинейные волны. М.: Мир, 1977.
6. *Newell G. F.* Applications of Queuing Theory. London: Chapman and Hall, 1982.
7. *Newell G. F.* Nonlinear effects in the dynamics of car-following // Oper. Res. 1961. V. 9. P. 209–229.
8. *Newell G. F.* A moving bottleneck // Transp. Res. B. 1998. V. 32. P. 531–537.
9. *Daganzo C. F.* The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // Transp. Res. B. 1994. V. 28. № 4. P. 269–287.
10. *Herman R., Montroll E. W., Potts R. B., Rothery R. W.* Traffic dynamics: studies in car following // Oper. Res. 1959. V. 7. P. 86–106.
11. *Gazis D. C., Herman R., Potts R. B.* Car following theory of steady state traffic flow // Oper. Res. 1959. V. 7. P. 499–505.
12. *Gazis D. C., Herman R., Rothery R. W.* Nonlinear follow the leader models of traffic flow // Oper. Res. 1961. V. 9. P. 545–567.
13. *Gazis D. C.* Traffic Theory. Berlin: Springer, 2002.
14. *May A. D.* Traffic Flow Fundamentals. Englewood Cliffs, NJ: Prentice-Hall, 1990.
15. *Leutzbach W.* Introduction to the Theory of Traffic Flow. Berlin: Springer, 1988.
16. *Daganzo C. F.* Fundamentals of Transportation and Traffic Operations. New York: Elsevier Science Inc., 1997.
17. *Muñoz J. C., Daganzo C. F.* Traffic and Transportation Theory. Editor M. A. P. Taylor. Oxford: Pergamon, 2002. P. 441–462.
18. *Gartner N. H., Messer C. J., Rathi A.* (editors) Traffic Flow Theory. Washington, DC: Transportation Research Board, 2001.

19. Chowdhury D., Santen L., Schadschneider A. Statistical physics of vehicular traffic and some related systems // Phys. Rep. 2000. V. 329. P. 199–329.
20. Helbing D. Traffic and related self-driven many particle systems // Rev. Mod. Phys. 2001. V. 73. P. 1067–1141.
21. Nagatani T. The physics of traffic jams // Rep. Prog. Phys. 2002. V. 65. P. 1331–1386.
22. Nagel K., Wagner P., Woesler R. Still flowing: Approaches to traffic flow and traffic jam modelling // Oper. Res., 2003. V. 51. P. 681–716.
23. Mahnke R., Kaupužs J., Lubashevsky I. Probabilistic description of traffic flow // Phys. Rep. 2005. V. 408. P. 1–130.
24. Rakha H., Pasumarthy P., Adjerid S. A simplified behavioral vehicle longitudinal motion model // Transp. Lett. 2009. V. 1. P. 95–110.
25. Delitala M., Tosin A. Mathematical modelling of vehicular traffic: A discrete kinetic theory approach // Math. Models Methods Appl. Sci. 2007. V. 17. P. 901–932.
26. Blank M. Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120. № 3-4. P. 627–658.
27. Maerivoet S., De Moor B. Cellular automata models of road traffic // Phys. Rep. 2005. V. 419. № 1. P. 1–64.
28. Kerner B. S., Klenov S. L. A microscopic model for phase transitions in traffic flow // J. Phys. A: Math. Gen. 2002. V. 35. P. L31–L43.
29. Kerner B. S., Klenov S. L., Wolf D. E. Cellular automata approach to three-phase traffic theory // J. Phys. A: Math. Gen. 2002. V. 35. P. 9971–10013.
30. Davis L. C. Multilane simulations of traffic phases // Phys. Rev. E, 2004. V. 69. 016108.
31. Kerner B. S., Klenov S. L. Deterministic microscopic three-phase traffic flow models // J. Phys. A: Math. Gen. 2006. V. 39. P. 1775–1809.
32. Lee H. K., Barlović R., Schreckenberg M., Kim D. Mechanical Restriction versus Human Overreaction Triggering Congested Traffic States // Phys. Rev. Lett. 2004. V. 92. 238702.
33. Jiang R., Wu Q. S. Spatial-temporal patterns at an isolated on-ramp in a new cellular automata model based on three-phase traffic theory // J. Phys. A: Math. Gen. 2004. V. 37. P. 8197–8213.
34. Gao K., Jiang R., Hu S.-X., Wang B.-H., Wu Q. S. Cellular-automaton model with velocity adaptation in the framework of Kerner's three-phase traffic theory. Phys. Rev. E. 2007. V. 76. 026105.

35. *Laval J. A.* Linking synchronized flow and kinematic waves // In: Traffic and Granular Flow' 05. Editors A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, D. E. Wolf. 2007. P. 521–526.
36. *Hoogendoorn S., van Lint H., Knoop V. L.* Macroscopic Modeling Framework Unifying Kinematic Wave Modeling and Three-Phase Traffic Theory // Trans. Res. Rec. 2008. V. 2088, P. 102–108.
37. *Davis L. C.* Controlling traffic flow near the transition to the synchronous flow phase // Physica A. 2006. V. 368. P. 541–550.
38. *Davis L. C.* Effect of cooperative merging on the synchronous flow phase of traffic // Physica A. 2006. V. 361. P. 606–618.
39. *Davis L. C.* Driver Choice Compared to Controlled Diversion for a Freeway Double On-Ramp in the Framework of Three-Phase Traffic Theory // Physica A. 2006. V. 379. P. 274–290.
40. *Jiang R., Hua M.-B., Wang R., Wu Q.-S.* Spatiotemporal congested traffic patterns in macroscopic version of the Kerner–Klenov speed adaptation model // Phys. Lett. A. 2007. V. 365. P. 6–9.
41. *Jiang R., Wu Q.-S.* Toward an improvement over Kerner–Klenov–Wolf three-phase cellular automaton model // Phys. Rev. E. 2005. V. 72. 067103.
42. *Jiang R., Wu Q.-S.* Dangerous situations in a synchronized flow model // Physica A. 2007. V. 377. P. 633–640.
43. *Li X. G., Gao Z. Y., Li K. P., Zhao X. M.* Relationship between microscopic dynamics in traffic flow and complexity in networks // Phys. Rev. E. 2007. V. 76. 016110.
44. *Pottmeier A., Thiemann C., Schadschneider A., Schreckenberg M.* Mechanical Restriction Versus Human Overreaction: Accident Avoidance and Two-Lane Traffic Simulations // In: Traffic and Granular Flow'05. Editors A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, D. E. Wolf. Berlin: Springer, 2007. P. 503–508.
45. *Siebel F., Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // Phys. Rev. E. 2006. V. 73. 066108.
46. *Wang R., Jiang R., Wu Q.-S., Liu M.* Synchronized flow and phase separations in single-lane mixed traffic flow // Physica A. 2007. V. 378. P. 475–484.
47. *Kerner B. S., Klenov S. L., Hiller A., Rehborn H.* Microscopic features of moving traffic jams // Phys. Rev. E. 2007. V. 73. 046107.
48. *Kerner B. S., Klenov S. L., Hiller A.* Criterion for traffic phases in single vehicle data and empirical test of a microscopic three-phase traffic theory // J. Phys. A: Math. Gen. 2006. V. 39. P. 2001–2020.

49. Kerner B. S., Klenov S. L., Hiller A. Empirical test of a microscopic three-phase traffic theory // Non. Dyn. 2007. V. 49. P. 525—553.
50. Kerner B. S. A theory of traffic congestion at heavy bottlenecks // J. Phys. A: Math. Theor. 2008. V. 41.
51. Davis L. C. Driver Choice Compared to Controlled Diversion for a Free-way Double On-Ramp in the Framework of Three-Phase Traffic Theory // Physica A. 2008. V. 387. P. 6395–6410.
52. Davis L. C. Realizing Wardrop equilibria with real-time traffic information // Physica A. 2009. V. 388. P. 4459–4474.
53. Davis L. C. Predicting travel time to limit congestion at a highway bottleneck // Physica A. 2010. V. 389. P. 3588–3599.
54. Gao K., Jiang R., Wang B.-H., Wu Q. S. Discontinuous transition from free flow to synchronized flow induced by short-range interaction between vehicles in a three-phase traffic flow model // Physica A. 2009. V. 388. P. 3233–3243.
55. Wu J. J., Sun H. J., Gao Z. Y. Long-range correlations of density fluctuations in the Kerner–Klenov–Wolf cellular automata three-phase traffic flow model // Phys. Rev. E. 2008. V. 78. 036103.
56. Jia B., Li X.-G., Chen T., Jiang R., Gao Z.-Y. Cellular automaton model with time gap dependent randomisation under Kerner's three-phase traffic theory // Transportmetrica. 2009. 1944–0987.
DOI: 10.1080/18128600903312789.
57. Tian J.-F., Jia B., Li X.-G., Jiang R., Zhao X.-M., Gao Z.-Y. Synchronized traffic flow simulating with cellular automata model // Physica A. 2009. V. 388. P. 4827–4837.
58. Kerner B. S., Klenov S. L. Phase transitions in traffic flow on multi-lane roads // Phys. Rev. E. 2009. V. 80. 056101.
59. Kerner B. S., Klenov S. L. A theory of traffic congestion on moving bottlenecks // J. Phys. A: Math. Theor. 2010. V. 43. 42510.
60. Kokubo S., Tanimoto J., Hagishima A. A new cellular automata model including a decelerating damping effect to reproduce Kerner's three-phase theory // Physica A. 2011. V. 390. P. 561–568.
61. Kerner B. S., Klenov S. L. Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks // Phys. Rev. E. 2003. V. 68. 036130.
62. Krauß S., Wagner P., Gawron C. Metastable states in a microscopic model of traffic flow // Phys. Rev. E. 1997. V. 55. P. 5597–5602.
63. Gipps P. G. A behavioural car-following model for computer simulation // Transportation Research B. 1981. V. 15. P. 105–111.

Процессы с запретами в моделях транспортных потоков¹

Изучено несколько простых моделей транспортных потоков в виде процессов с запретами (как решеточных, так и с непрерывным пространством) и установлены явные формулы для некоторых связанных с ними статистик. В частности, получена так называемая фундаментальная диаграмма, выражающая зависимость средней скорости движения от плотности частиц.

1. Введение

Одним из естественных способов математического моделирования транспортных потоков является их реализация в виде процессов с запретами (Exclusion Processes). Последние представляют собой системы частиц, совершающих случайные блуждания и взаимодействующих по закону «исключенного объема» (hard core). Впервые простейшую решеточную модель этого типа предложил Ф. Спитцер в 1970 г., и с тех пор подобные, на первый взгляд примитивные, модели нашли весьма широкое применение в самых различных областях, начиная с моделей транспортных потоков [2, 3, 10, 12, 16], синтеза протеинов и молекулярных моторов в биологии, роста случайных поверхностей в физике (см. [7, 17]) и до анализа диаграмм Юнга в теории представлений [9].

Качественно с точки зрения порядка взаимодействий частиц имеется два типа процессов с запретами: асинхронные и синхронные. В первом случае не более одной частицы может сдвинуться в данный момент времени, а во втором — все частицы двигаются одновременно. Последний вариант приводит к необходимости анализа кратных одновременных взаимодействий, однако с точки зрения приложений представляется более естественным.

¹Работа была частично поддержана грантами РФФИ.

В работе будет изучено несколько простых моделей транспортных потоков этого типа (как решеточных, так и с непрерывным пространством) и установлены явные формулы для некоторых связанных с ними статистик. В частности, получена так называемая фундаментальная диаграмма, выражающая зависимость средней скорости движения от плотности частиц. Читатель может найти серьезные обзоры по «физическим» постановкам задач по этой тематике в [8, 12].

Мой собственный интерес к такого рода задачам возник в связи со следующим практическим наблюдением. Иногда быстрее идти против движения в медленно двигающейся толпе людей (например, в переходе метро), чем по движению. Стандартная вероятностная модель диффузии частицы по или против потока явно противоречит этому наблюдению, что указывает на специальную (неслучайную) внутреннюю структуру потока в рассматриваемом случае. Одной из целей настоящей работы является обсуждение того, как подобная структура возникает из произвольных (случайных) начальных конфигураций частиц.

Мы начнем исследование с простейшей детерминированной модели на целочисленной решетке (описанной в разделе 2). Эта модель может быть полностью изучена элементарными средствами, и мы приводим ее в основном в педагогических целях. Однако даже в рамках такой модели отбрасывание условия регулярности начальной конфигурации частиц приводит к необходимости значительно более сложного математического анализа (см. [2, 3]). В дальнейшем оказалось, что с точки зрения математики проще изучать, на первый взгляд, значительно более сложную модель – процесс с запретами в непрерывном пространстве. В разделе 3 мы опишем эту модель и покажем, что ее ограничение на множество конфигураций, расположенных в целочисленных точках, инвариантно. Поэтому все полученные результаты применимы также к модели с дискретным пространством.

Все известные подходы к анализу решеточных систем существенно используют комбинаторную структуру пространства конфигураций частиц (отметим, например, идею двойственности «частица–пустая позиция», используемую в разделе 2). Никаких аналогов подобных комбинаторных структур в непрерывном пространстве нет, что приводит к необходимости разработки фундаментально нового подхода.

Процессы с запретами в непрерывном пространстве новы не толь-

ко как модели транспортных потоков, но и с чисто математической точки зрения. Первые результаты на эту тему получены недавно в [5], где была разработана оригинальная техника, позволяющая изучать эргодические (статистические) свойства таких процессов. Главной технической новинкой здесь является *метод динамического каплинга* (описанный в разделе 6). Отметим, что этот метод не только нов, но и используется нестандартно: вместо доказательства (обычного в теории каплинга) существования «успешного склеивания» (которого может и не быть в наших условиях) мы используем его наличие или отсутствие в качестве диагностического средства.

2. Простейшая модель на целочисленной решетке

Начнем с простейшей одномерной модели транспортного потока, введенной в [16]. Эта модель описывается следующей динамической системой с дискретным временем и дискретным фазовым пространством – целочисленной решеткой \mathbb{Z} , на которой расположены частицы. В следующий момент времени каждая частица либо передвигается вперед на одну позицию, если она свободна, либо остается на месте в противном случае. В случае конечной решетки с периодическими граничными условиями анализу (в основном численному) этой модели и некоторых ее обобщений в последнее время было посвящено большое число публикаций (см. [8; 10–12; 15, 16] и дальнейшие ссылки в них). Наиболее интересным явлением, обнаруженным в данных работах, является нетривиальная зависимость средней скорости движения частиц от их плотности $V(\rho)$, равная 1 при $\rho \in [0, 1/2]$ и $\frac{1}{\rho} - 1$ при $\rho \in (1/2, 1]$. Ниже мы выведем этот результат при помощи техники двойственных отображений для произвольных конечных и бесконечных решеток и начальных конфигураций. Кроме того, мы дадим полное описание предельных множеств (соответствующих стационарным транспортным потокам) и точную оценку длины переходного периода. Физическая интерпретация описанного результата – это наличие фазового перехода «газ–жидкость» от свободного движения частиц (при малой плотности) к постоянному наличию транспортных пробок (при большой плотности).

С точки зрения теории динамических систем описанная выше модель может быть представлена следующим образом. Пусть

$X = \{0, 1\}^{\mathbb{Z}}$ – множество всех возможных конфигураций – бинарных последовательностей $x = x(i)$, $i \in \mathbb{Z}$, единицы в которой соответствуют частицам, а нули – незанятым позициям на решетке. Рассмотрим отображение $T : X \rightarrow X$:

$$Tx(i) := \begin{cases} 1, & \text{если } x(i) = 0, x(i-1) = 1 \text{ или } x(i) = x(i+1) = 1, \\ 0 & \text{– в остальных случаях.} \end{cases}$$

Группу из (более одной) последовательно стоящих частиц мы назовем *кластером*; а частицу в позиции i (т.е. $x(i) = 1$), позиция после которой не занята (т.е. $x(i+1) = 0$), назовем *свободной*. Будем называть конфигурацию $x \in X$ регулярной, если имеется число $\rho = \rho(X)$ (плотность частиц) и монотонная функция $\varphi(N) \rightarrow 0$ при $N \rightarrow \infty$, такие, что для любого N число частиц с координатами от $n + 1$ до $n + N$ отличается от $N\rho$ не более чем на $N\varphi(N)$ для любого n . Заметим, что конфигурация на конечной решетке длины n с периодическими граничными условиями соответствует n -периодической конфигурации на бесконечной решетке, которая удовлетворяет условию регулярности с $\varphi(N) = n\rho(1 - \rho)/N$. Под средней (по пространству) скоростью (частиц) $V(x)$ понимается среднее значение (если оно корректно определено) перемещения частиц в конфигурации x во время следующей итерации отображения T . Отметим, что в разделе 4 будет введено и изучено более тонкое понятие средней скорости индивидуальной частицы.

Теорема 2.1. Для любой регулярной начальной конфигурации $x \in X$ с плотностью $\rho \neq 1/2$ через не более чем $t_c(x) = \frac{1}{2}\varphi^{-1}(|\frac{1}{2} - \rho(x)|)$ итераций отображения T средняя скорость станет равна $V = \min(1, \frac{1}{\rho} - 1)$, и при любом $t \geq t_c(x)$ выполняется следующая альтернатива: конфигурация $T^t x$ состоит либо только из свободных частиц, либо не имеет кластеров незанятых позиций. Более того, $\forall n$ для n -периодических начальных конфигураций ограничение $\rho \neq 1/2$ снимается, для $t_c(x)$ справедлива лучшая оценка $t_c(x) = \min(\rho(x)N, N - \rho(x)N)$, и при $t \geq t_c(x)$ последовательность $\{T^t x\}_t$ становится n -периодической по t .

Доказательство этой теоремы и ряда других результатов настоящей работы основано на идее введения двойственной динамической системы (T^*, X^*) , описывающей динамику незанятых позиций на решетке под действием основного отображения T . Здесь для конфигурации $x \in X$ двойственная конфигурация x^* определяется соот-

напшением $x_i^* = 1 - x_i$ для всех i . Можно показать, что $(Tx)^* = T^*x^*$ при всех $x \in X$. Для рассматриваемой модели отображение T^* отличается от T только направлением движения частиц, что сводит анализ к конфигурациям низкой плотности $\rho \in [0, 1/2]$, поскольку большая плотность соответствует плотности незанятых позиций меньшей $1/2$. Это наблюдение резко упрощает задачу, поскольку динамика в случае высокой плотности частиц нетривиальна и трудно поддается непосредственному анализу. Далее показывая, что длина любого кластера частиц не может возрастать (т.е. в этой модели не могут возникать транспортные пробки), а число свободных частиц убывать, мы приходим к описанной в формулировке теоремы альтернативе, что и приводит к требуемым оценкам.

Рассмотрим теперь модель движения со сверхбыстрыми частицами, отличающуюся от предыдущей тем, что на каждом шагу частица сдвигается вперед до следующей занятой позиции.

Теорема 2.2. *Для любой начальной конфигурации x , удовлетворяющей закону больших чисел с плотностью $\rho(x) \notin \{0, 1\}$, средняя скорость частиц не зависит от времени и равна $\frac{1}{\rho(x)} - 1$.*

Качественно динамика этой модели богаче, чем в модели с медленными частицами, например, транспортные пробки типичны даже для конфигураций малой плотности. С другой стороны, несмотря на это, средняя скорость движения частиц $V(\rho)$ для этой модели совпадает с предыдущим случаем при высокой плотности и аналитически продолжает ее при малой плотности.

Сделаем несколько замечаний о простейших обобщениях и приложениях описанных моделей. Во-первых, часто рассматривается вероятностная постановка, при которой частица переходит на незанятую позицию с заданной вероятностью p (случай $p = 1$ возвращает нас к описанной детерминированной задаче). Как показывает численный анализ и качественные рассуждения (см. [8, 9, 15, 16]), результаты для детерминированного случая выглядят очень похоже и для стохастической версии при p достаточно близко к 1. Полный математический анализ здесь к настоящему времени проведен только для модели движения конечного набора частиц по окружности, а не бесконечной решетке (см. [11, 12]). Частичный ответ в общем случае получен также при анализе динамики в непрерывном пространстве (см. следующий раздел и [5]).

Важным представляется вопрос о возможности описания мно-

гополосного движения в рамках процессов с запретами. Одной из возможностей здесь является изменение условия о том, что не более одной частицы может находиться в одной позиции на решетке, на условие о максимальном числе $M > 1$ частиц. В случае $M = 2$ эта модель в точности соответствует двухполосному движению, а при $M > 2$ представляет собой некоторое упрощение. Математический анализ детерминированной постановки данной задачи проведен в [2].

До сих пор мы обсуждали только модели, при которых оказывается справедливой точная зависимость между средней скоростью движения частиц и их плотностью. Как известно, экспериментальные данные показывают, что в общем случае одной плотности частиц может соответствовать целый набор средних скоростей или последнее понятие может не быть корректно определено. Оказывается, что простые модификации рассматриваемых нами моделей демонстрируют подобное поведение (см. [3–5; 8, 9, 15]). С точки зрения фазовых переходов описанное поведение соответствует возникновению новой «гистерезисной» фазы.

Дадим теперь математическое описание наблюдения о движении пассивной быстрой частицы (имитирующей поведение спешащего прохожего) в медленном транспортном потоке, который мы сформулировали в начале данного раздела. Упрощая ситуацию, мы будем полагать (как обычно делают в гидродинамике), что движение нашей быстрой частицы не влияет на транспортный поток и описывается следующим образом. Положим $\tau_x^+(y) := \min(i : y < i \text{ и } x(i) = 1)$, $\tau_x^-(y) := \max(i : y > i \text{ и } x(i) = 1)$. Тогда совместная динамика T_\pm конфигурации частиц $x \in X$ и положения быстрой частицы $y \in \mathbb{Z}$ определяется косым произведением отображения T и одного из отображений τ^\pm (знак соответствует движению по/против потока), т.е. $T_\pm(x, y) := (Tx, \tau_x^\pm(y))$.

Под скоростью в момент t пассивной частицы будем понимать суммарное расстояние (со знаком), пройденное ею к этому моменту времени, деленное на t . Опираясь на полученное полное описание предельных множеств модели медленных частиц, мы получаем следующий результат.

Теорема 2.3. *Для любой регулярной начальной конфигурации с плотностью $\rho(X) \notin \{0, 1/2, 1\}$ в случае неограниченной решетки средняя скорость быстрой частицы стремится (по t) к 1 при $\rho \leq 1/2$ и движении вперед (по потоку), и к $-\max(1, 1/\rho(x) - 1)$ при движении назад (против потока).*

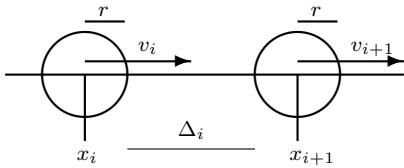


Рис. 1. Процесс с запретами в непрерывном пространстве

3. Процессы с запретами в непрерывном пространстве

Перейдем теперь к изучению более общего класса процессов с запретами в непрерывном пространстве с синхронными взаимодействиями (т.е. все частицы пытаются двигаться одновременно).

В непрерывном пространстве координатное представление конфигураций (принятое в предыдущем разделе) неудобно, и вместо этого предлагается следующее. Под конфигурацией частиц $x := \{x_i\}_{i \in \mathbb{Z}}$ будем понимать бесконечную (в обе стороны) последовательность действительных чисел $x_i \in \mathbb{R}$, которые можно интерпретировать как центры шаров заданного радиуса $r \geq 0$ (см. рис. 1). Предполагается, что упорядочивание по индексу соответствует естественному порядку позиций центров шаров, т.е. $\dots \leq x_{-1} \leq x_0 \leq x_1 \leq \dots$. Чтобы отметить зависимость от радиуса шара $r \geq 0$, мы используем обозначение $x(r)$ и только в предельном случае $r = 0$ не отмечаем этой зависимости, т.е. $x \equiv x(0)$. Будем говорить, что конфигурация $x(r)$ допустима, если

$$x_i(r) + r \leq x_{i+1}(r) - r \quad \forall i \in \mathbb{Z}$$

(соответствующие шары не пересекаются и могут только касаться), и обозначим через $X(r)$ пространство допустимых конфигураций.

Динамика в пространстве конфигураций определяется следующим образом. Начнем с тривиальной конфигурации, состоящей из единственной частицы, находящейся в момент времени $t \geq 0$ в точке $x_0^t \in \mathbb{R}$ (т.е. $x^t \equiv \{x_0^t\}$). В этом случае полагаем

$$x_0^{t+1} := x_0^t + v_0^t,$$

где $\{v_0^t\}$ – заданная последовательность (случайных) величин. Значения v_0^t естественно рассматривать как локальные *скорости* частицы в момент времени t . Таким образом, полученный процесс – это

простое случайное блуждание в \mathbb{R} . Обобщая эту тривиальную постановку на случай бесконечной конфигурации $x(r) \in X$ и вновь интерпретируя (бесконечную в обе стороны по $i \in \mathbb{Z}$) последовательность $\{v_i^t\}_{i,t}$ как локальные скорости частиц в конфигурации $x^t(r)$ в момент t , получаем бесконечный набор случайных блужданий, ограниченных условиями сохранения порядка и законом исключенного объема (hard core exclusion rule).

Для упрощения изложения мы ограничимся только случаем неотрицательных локальных скоростей, собственно только эта ситуация осмысlena в задачах транспортного моделирования. В общем случае при анализе локальных скоростей обоих знаков определения становится существенно сложнее, но как результаты, так и доказательства почти не изменяются (см. [5]).

Для неотрицательных локальных скоростей рассматриваемые на- ми запреты означают, что условие допустимости нарушается для i -й частицы в момент $t \in \mathbb{Z}_+$ тогда и только тогда, когда неравенство

$$x_i^t(r) + v_i^t + r \leq x_{i+1}^t(r) - r$$

перестает выполняться. В последнем случае мы будем говорить о *конфликте* между частицами i и $i+1$, для разрешения которого применяется конструкция *нормализации*:

$$v_i^t \rightarrow \mathcal{N}(v_i^t, x^t(r)).$$

Позиции частиц в момент времени $t+1$ вычисляются по правилу:

$$x_i^{t+1}(r) := x_i^t(r) + \mathcal{N}(v_i^t, x^t(r)) \quad \forall i.$$

Нормализация может быть проведена различными способами (что приводит к существенно разным статистическим свойствам). В настоящей работе мы рассмотрим только *слабую нормализацию* (другие возможности изучены в [5]), при которой в случае конфликта локальная скорость меняется так, чтобы соответствующая частица могла продвинуться вперед на максимально возможное расстояние. В терминах *зазоров*

$$\Delta_i(x^t) \equiv \Delta_i^t := x_{i+1}^t - x_i^t - 2r$$

между частицами в конфигурации x^t нормализация записывается следующим образом:

$$\mathcal{N}(v_i^t, x^t) := \begin{cases} v_i^t, & \text{если } v_i^t \leq \Delta_i^t; \\ \Delta_i^t & \text{— в противном случае.} \end{cases}$$

Здесь важно отметить, что между любыми двумя конфигурациями частиц $x(r)$, $\dot{x}(\dot{r})$ с общей последовательностью зазоров $\Delta := \{\Delta_i\}$ имеется взаимно однозначное соответствие φ :

$$\dot{x}_i(\dot{r}) = \varphi(x_i(r)) := x_i(r) - 2i(r - \dot{r}) \quad \forall i \in \mathbb{Z}.$$

Поскольку нормализация зависит только от зазоров между частицами, достаточно провести анализ случая частиц нулевого радиуса ($r = 0$). Статистика в общем случае $r > 0$ пересчитывается при помощи замены переменных φ . С другой стороны, полагая $r = 1/2$, $x_i^0(r) \in \mathbb{Z} \quad \forall i \in \mathbb{Z}$ и $v_i^t \in \mathbb{Z} \quad \forall i \in \mathbb{Z}, t \geq 0$, мы получаем, что $x_i^t(r) \in \mathbb{Z} \quad \forall i \in \mathbb{Z}, t \geq 0$. Последнее означает, что системы на целочисленной решетке инвариантны относительно введенной динамики. Поэтому наши результаты приводят к принципиально новому подходу для анализа решеточных систем. Заметим все же, что в случае $r = 0$ условие допустимости разрешает наличие произвольного (и даже бесконечного) числа частиц в одной точке пространства, что запрещено для решеточной системы.

Естественно, без специальных предположений о структуре локальных скоростей $\{v_i^t\}_{i,t}$ никакие содержательные результаты о динамике подобных систем невозможны. Будем полагать, что $v_i^t \in [0, v] \quad \forall i \in \mathbb{Z}, t \in \mathbb{Z}_0 := \mathbb{Z}_+ \cup \{0\}$ и выполнено одно из следующих (на первый взгляд противоположных) предположений:

- (a) $v_i^t \equiv v_0^t \quad \forall i \in \mathbb{Z}, t \in \mathbb{Z}_0$ и $\exists \bar{v}(\gamma) := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \min(v_0^s, \gamma) \quad \forall \gamma > 0$
почти наверное (п.н.);
- (b) $\{v_i^t\}$ являются независимыми одинаково распределенными (н.о.п.), как по i , так и по t , случайными величинами (с.в.).

Пересечение между множествами локальных скоростей, удовлетворяющих предположениям (a) или (b), не пусто и содержит принципиально важный случай чисто детерминированных скоростей: $v_i^t \equiv v \quad \forall i \in \mathbb{Z}, t \in \mathbb{Z}_0$. Как мы покажем, свойства всех систем, удовлетворяющих условию (a), близки к чисто детерминированному случаю. Поэтому в дальнейшем мы будем говорить о постановке (a) как о *детерминированной*,² а о постановке (b) как о *стохастической*.

²Заметим, что $\{v_0^t\}$ может быть траекторией детерминированного хаотического отображения $f : [0, 1] \rightarrow [0, 1]$, т.е. $v_0^{t+1} := vf^t(v_0^t/v)$, как и (несмотря на название) реализацией настоящей стохастической цепи Маркова.

Заметим, что кажущаяся простейшей чисто детерминированная постановка $v_i^t \equiv v \quad \forall i \in \mathbb{Z}, t \in \mathbb{Z}_0$ приводит к чрезвычайно сложной динамике частиц. Это видно например из того, что детерминированная динамическая система, описывающая динамику конфигураций частиц, в этом случае оказывается хаотической, и, более того, топологическая энтропия этой системы бесконечна (теорема 5.3).

Обычно математический анализ систем взаимодействующих частиц начинают с изучения инвариантных распределений на них и, выбрав удачное инвариантное распределение, переходят к анализу его статистических характеристик. В нашем случае этот подход не работает. Дело в том, что у рассматриваемых нами систем может быть как бесконечно много инвариантных распределений, так и ни одного (напомним тривиальный пример одной частицы, совершающей асимметричное случайное блуждание). Несмотря на отсутствие инвариантного распределения, последний пример демонстрирует, что здесь имеется другая важная статистическая характеристика – средняя скорость движения частиц, легко вычисляемая в этом примере.

4. Элементарные свойства

В этом разделе мы изучим вопросы, связанные с определениями понятий плотности и средней скорости частиц для процессов в непрерывном пространстве.

Под *плотностью* $\rho(x, I)$ конфигурации $x \in X$ в ограниченном сегменте $I = [a, b] \in \mathbb{R}$ будем понимать число частиц из x , центры которых x_i находятся в I , деленное на длину $|I| > 0$ сегмента I . Если для любой последовательности *вложенных* ограниченных сегментов $\{I_n\}$ с $|I_n| \xrightarrow{n \rightarrow \infty} 0$ предел

$$\rho(x) := \lim_{n \rightarrow \infty} \rho(x, I_n)$$

корректно определен, то этот предел назовем *плотностью* конфигурации $x \in X$. В противном случае рассматриваются верхняя и нижняя (по отношению ко всем возможным коллекциям вложенных ограниченных сегментов $\{I_n\}$) плотности частиц $\rho_{\pm}(x)$.

Замечание 4.1. (а) Если $\rho(x) < \infty$, то $|x_n - x_m| \xrightarrow{|n-m| \rightarrow \infty} 1/\rho(x)$.

- (b) Пусть конфигурации $x(r) \in X(r)$, $r > 0$ и $x \in X$ имеют общую последовательность зазоров $\{\Delta_i\}$. Тогда $\rho_{\pm}(x(r)) = \frac{\rho_{\pm}(x)}{1+2r\rho_{\pm}(x)}$.

Лемма 4.2. *Верхняя/нижняя плотности $\rho_{\pm}(x^t)$ инвариантны относительно динамики, т.е. $\rho_{\pm}(x^t) = \rho_{\pm}(x^{t+1}) \quad \forall t$.*

Под (средней по времени) *скоростью* i -й частицы в конфигурации $x \in X$ в момент $t > 0$ будем понимать

$$V(x, i, t) := \frac{1}{t} \sum_{s=0}^{t-1} \mathcal{N}(v_i^s, x^s) \equiv (x_i^t - x_i^0)/t.$$

Если предел

$$V(x, i) := \lim_{t \rightarrow \infty} V(x, i, t)$$

корректно определен, назовем его (средней по времени) *скоростью* i -й частицы. В противном случае рассматриваются нижняя и верхняя скорости частицы $V_{\pm}(x, i)$.

Лемма 4.3. *Для любой конфигурации $x \in X$ выполнено*

$$|V(x, j, t) - V(x, i, t)| \xrightarrow{t \rightarrow \infty} 0 \text{ n.h. } \forall i, j \in \mathbb{Z}.$$

Следствие 4.4. *Нижняя и верхняя скорости i -й частицы $V_{\pm}(x, i)$ не зависят от индекса i .*

Доказательство этого результата кроме всего прочего демонстрирует тот факт, что в детерминированной постановке зазоры между последовательными частицами не могут существенно увеличиваться. Следующее утверждение показывает, что при некоторых слабых технических предположениях (заведомо выполняемых при высокой плотности частиц) большие зазоры со временем исчезают.

Лемма 4.5. *Пусть $x \in X$ и рассматривается только чисто детерминированная постановка (т.е. $v_i^t \equiv v$). Предположим, что $\forall t \exists j > t : \Delta_j(x^t) < v$. Тогда $\forall i \exists t_i < \infty : \Delta_i(x^t) < 2v \quad \forall t \geq t_i$.*

5. Эргодические свойства

Сформулируем теперь основные результаты для процессов с запретами в непрерывном пространстве.

Теорема 5.1. Пусть плотность $\rho(x)$ конфигурации $x \in X$ корректно определена. Тогда множество предельных точек при $t \rightarrow \infty$ последовательности $\{V(x, t)\}_{t \in \mathbb{Z}_0}$ зависит только от $\rho(x)$.

Теорема 5.2 (фундаментальная диаграмма). В детерминированной постановке

$$V(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \min(1/\rho, v_0^s) = \begin{cases} v, & \text{если } \rho(x) \leq 1/v, \\ 1/\rho(x) - 2r & \text{— в противном случае,} \end{cases}$$

если $v_0^t \equiv v$.

Следствие 5.1. Пусть для конфигурации $x(r) \in X(r)$, $r > 0$ плотность $\rho(x(r))$ корректно определена и пусть $\forall i$, t $v_i^t \equiv v$. Тогда

$$V(x(r)) = \begin{cases} v, & \text{если } \rho(x) \leq \frac{1}{v+2r}, \\ 1/\rho(x(r)) - 2r & \text{— в противном случае.} \end{cases}$$

В частности, для версии процесса на целочисленной решетке получаем

$$V(x(1/2)) = \begin{cases} v, & \text{если } \rho(x) \leq \frac{1}{v+1}, \\ 1/\rho(x(1/2)) - 1 & \text{— в противном случае.} \end{cases}$$

Замечание 5.2. Последний результат совпадает с соответствующим утверждением о процессе на решетке, описанном в теореме 2.1 (см. также [2, 16]). Несмотря на это сходство, в решеточном случае имеется важное качественное отличие динамики: при высокой плотности частицы неминуемо образуют плотные кластеры (статические транспортные пробки). Доказательство же теоремы 5 в действительности показывает, что «типичное» поведение конфигураций высокой плотности качественно отлично: они также образуют кластеры частиц (т.е. наборы последовательных частиц, расстояния между которыми строго меньше v), но эти кластеры не стоят на месте, а передвигаются с постоянной скоростью как «эшелон». Интересно отметить, что ранее был разработан целый ряд весьма сложных решеточных моделей для имитации подобного поведения.

В чисто детерминированной постановке (т.е. $v_i^t \equiv v \quad \forall i, t$) рассматриваемая система описывается детерминированным отображением $T_v : X \rightarrow X$ из пространства допустимых конфигураций в

себя. Покажем, что это отображение сильно хаотическое в том смысле, что его топологическая энтропия бесконечна.³ Читатель может найти детальное описание конструкций, связанных с энтропией динамической системы и ее свойств, например, в [13]. Чтобы обойти сложности, связанные с некомпактностью фазового пространства, мы определим топологическую энтропию отображения T_v (означение $h_{\text{top}}(T_v)$) как супремум по всем метрическим энтропиям этого отображения относительно его вероятностных инвариантных мер.

Теорема 5.3. *Топологическая энтропия чисто детерминированного процесса с запретами в непрерывном пространстве бесконечна.*

Доказательство этого результата основано на аналогичном утверждении для действия отображения сдвига $\sigma_v : X \rightarrow X$ в непрерывном пространстве:

$$(\sigma_v x)_i := x_i + v \quad i \in \mathbb{Z}, \quad x \in X.$$

Лемма 5.3. *Топологическая энтропия отображения сдвига σ_v в непрерывном пространстве бесконечна.*

Идея здесь состоит в том, чтобы построить инвариантное подмножество пространства конфигураций X , на котором отображение σ_v изоморфно полному отображению сдвига в пространстве последовательностей со счетным алфавитом. Замечая теперь, что топологическая энтропия полного отображения сдвига в пространстве последовательностей с алфавитом из n элементов равна $\ln n$, получаем наше утверждение.

6. Каплинг

Одной из основных технических новаций в доказательстве результатов, сформулированных в предыдущем разделе, является конструкция «динамического» каплинга.

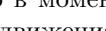
Напомним, что под каплингом двух марковских процессов x^t и y^t , действующих на пространстве X , понимается представление этой

³Обычно говорят, что отображение хаотическое, если его топологическая энтропия положительна, поэтому бесконечное значение энтропии говорит об очень высоком уровне хаотичности.

пары процессов на общем вероятностном пространстве. Иными словами, каплинг — это процесс пар (x^t, y^t) , определенный на пространстве прямого произведения $X \times X$, удовлетворяющий следующим условиям:

$$P((x^t, y^t) \in A \times X) = P(x^t \in A), \quad P((x^t, y^t) \in X \times A) = P(y^t \in A),$$

т.е. проекции нового процесса пар ведут себя точно так же, как исходные процессы.

Обсудим теперь конструкцию динамического каплинга между двумя копиями x^t, \dot{x}^t рассматриваемого нами марковского процесса. Обычно при анализе систем взаимодействующих частиц на решетке с асинхронными взаимодействиями используется так называемый «равный» каплинг (см., например, [14]). В этом случае каплинг состоит в спаривании частиц процессов x^t, \dot{x}^t , занимающих одинаковые позиции. После спаривания все выборы случайных скоростей для элементов одной пары предполагаются одинаковыми. В рассматриваемом нами случае систем с синхронными взаимодействиями этот подход не работает. Действительно, произвольное число частиц может сдвинуться одновременно, что приводит к ситуации, когда частицы, принадлежащие процессам x^t, \dot{x}^t , обгоняют друг друга, но при этом ни в какой момент времени не занимают одинаковые позиции. Кроме того, имеется и более важное препятствие: может оказаться, что движение только одной частицы из пары заблокировано в момент t неспаренной частицей. В результате одновременного движения всех этих частиц получаем следующую диаграмму: . Как видим, старая пара уничтожается, но при равном каплинге новая пара не образуется. Здесь и далее мы используем диаграммное представление для конфигураций при каплинге: спаренные частицы обозначаются черными кружками, а неспаренные — белыми, при этом верхняя строка диаграммы показывает x -частицы (т.е. частицы x -процесса), а нижняя строка соответствует \dot{x} -частицам.

Чтобы обойти это препятствие, мы и вводим *динамический*⁴ каплинг, описанный в [5, 6]. Отметим для сравнения идейно близкую конструкцию каплинга в [1, 11], предложенную для случая решеточных систем с асинхронными взаимодействиями. Важным преимуще-

⁴Слово «динамический» используется для того, чтобы подчеркнуть то, что взаимное положение частиц в одной паре меняется со временем, в отличие от равного каплинга.

ством динамического каплинга по отношению к этим конструкциям является ограниченность расстояний между элементами одной пары (в конструкциях [1, 11] эти расстояния могут становиться бесконечно большими).

Под *динамическим каплингом* процессов x^t, \dot{x}^t понимается последовательное спаривание достаточно близко расположенных частиц в разных процессах, удовлетворяющее следующим условиям.

- (A1) В момент $t = 0$ все частицы предполагаются неспаренными. Локальные скорости взаимно спаренных частиц всегда однаковы.
- (A2) Однажды созданная пара частиц никогда не исчезает; при этом частицы, образующие данную пару, могут меняться.
- (A3) Частица, обгоняющая под действием динамики за один шаг времени некоторые неспаренные частицы, становится спаренной с одной из них.

Согласно (A1) – (A3) частицы, принадлежащие одной паре, движутся синхронно до тех пор, пока либо нарушается условие допустимости для одной из них (т.е. ее движение заблокировано другой частицей), либо одна из частиц в паре заменяется неспаренной частицей, принадлежащей тому же процессу (см. рис. 2). Удобно представлять результат каплинга наших процессов как «газ», состоящий из ординарных (неспаренных) частиц и «гантелей» (пар). Спаренная прежде частица (элемент гантели) может наследовать роль ординарной от одного из своих соседей. Для того чтобы удобнее отслеживать позиции неспаренных частиц, мы будем называть их *x - и \dot{x} -дефектами* в зависимости от процесса, к которому они принадлежат.

Практически динамический каплинг может быть реализован самыми разными способами (в частности, используя только идею обгона частиц). Чтобы продемонстрировать гибкость конструкции, мы опишем другой подход. Отметим, что в дальнейшем только свойства (A1) – (A3) используются в доказательствах.

Под *x -тройкой* ($\bullet^\circ\bullet$ или $\bullet^\circ\bullet$) в процессе пар (x^t, \dot{x}^t) понимается две взаимно спаренные частицы и x -дефект, находящийся между ними, индекс которого отличается на единицу от индекса спаренной x -частицы. *\dot{x} -тройка* ($\bullet^\circ\bullet$ или $\bullet\bullet^\circ$) определяется аналогично.

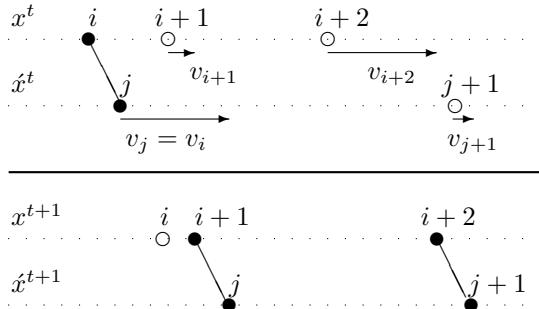


Рис. 2. Спаривание частиц. Взаимно спаренные частицы обозначены черными кружками и соединены прямыми линиями, а дефекты — белыми кружками. В момент t частицы i и j спарены, тогда как в момент $t + 1$ x -частица i становится неспаренной, а \dot{x} -частица j спаривается с x -частицей $i + 1$. Неспаренные в момент t частицы $i + 2$ и $j + 1$ спариваются в момент $t + 1$

Говорят, что две пары частиц *пересекают* друг друга, если прямые линии, соединяющие позиции частиц из одной пары, пересекаются, т.е. $\bullet^*\bullet$ (здесь взаимно спаренные частицы обозначены одинаковыми символами).

x -дефект в x_i^t вместе с ближайшим к нему⁵ \dot{x} -дефектом в \dot{x}_j^t (\circ или \circ) назовем *d-парой*, если $|x_i^t - \dot{x}_j^t| < v$, эта пара дефектов не пересекается с другими взаимно спаренными частицами, и интервал (x_i^t, \dot{x}_j^t) не содержит других дефектов. Будем говорить, что *d-пара* (i, j) *меньше* чем *d-пара* (n, m) , если $|i| < |n|$, или $i < n$ — в случае $|i| = |n|$. Заметим, что ситуация $i = n$ и $j \neq m$ может произойти, в отличие от ситуации $i \neq n$ и $j = m$.

Приведем два примера. В наборе $\circ\bullet\bullet\circ$ две первых x -частицы вместе с первой \dot{x} -частицей образуют x -тройку, несмотря на наличие дополнительной спаренной частицы в интервале между ними. С другой стороны, набор $\circ\bullet\circ$ не содержит ни троек, ни *d-пар*.

Пара конфигураций (x^t, \dot{x}^t) называется *правильной*, если она не содержит x - или \dot{x} -троек, *d-пар* и пересекающихся взаимно спаренных частиц.

⁵Если имеется несколько ближайших \dot{x} -дефектов, то выбирается тот, который имеет минимальный индекс.

Правильность пары конфигураций (x^t, \dot{x}^t) в момент t в общем случае не препятствует тому, что под действием динамики в момент $(t+1)$ пара (x^{t+1}, \dot{x}^{t+1}) может оказаться неправильной. В частности, могут возникнуть тройки обоих типов и d -пары, например $\bullet^\circ \circ \rightarrow \bullet^\circ \circ$ или $^\circ \circ \circ \rightarrow ^\circ \circ$. Здесь важно, что ввиду сохранения порядка частиц пересекающиеся взаимно спаренные частицы не могут появиться.

Лемма 6.1. *Пусть пара конфигураций (x^t, \dot{x}^t) не имеет пересекающихся взаимно спаренных частиц. Тогда тройки одного типа не могут иметь общих элементов.*

Поэтому все тройки одного типа могут быть устраниены одновременно. Под *устранением* x - или \dot{x} -тройки будем понимать следующее: бывший дефект спаривается с частицей из другого процесса, а спаренная ранее частица становится неспаренной:

$$^\circ \bullet \rightarrow \bullet^\circ .$$

Устранение d -пары еще проще: дефекты «аннигилируют» друг друга, образуя спаренную пару частиц: $^\circ \circ \rightarrow \bullet^\circ \bullet$. Во всех случаях позиции частиц сохраняются, а меняются только их «роли».

В результате конструкция динамического каплинга состоит из следующих шагов.

1. Каждая x -тройка рекурсивно устраняется: $\bullet^\circ \bullet \rightarrow \bullet^\circ \circ$.
2. Каждая \dot{x} -тройка рекурсивно устраняется: $^\circ \bullet \bullet \rightarrow \bullet^\circ \bullet$.
3. Наименьшая⁶ d -пара рекурсивно устраняется: $^\circ \circ \rightarrow \bullet^\circ \bullet$.

Лемма 6.2. *Описанная процедура каплинга корректно определена, приводит к марковскому процессу пар и удовлетворяет условиям (A1) – (A3).*

Чтобы объяснить необходимость рекурсий, заметим, что пространственные сегменты, на которых расположены спаренные частицы, могут пересекаться. Поэтому устранение x - или \dot{x} -тройки может привести к созданию новой тройки того же типа:

$$\bullet^\circ \bullet^\circ \bullet \rightarrow \bullet^\circ \bullet \circ^\circ \bullet \rightarrow \bullet^\circ \bullet \bullet^\circ \circ.$$

⁶Порядок d -пар может меняться после каждой процедуры рекурсии.

Заметим теперь, что при рекурсиях в процедуре каплинга дефект может сдвинуться на произвольно большое расстояние от его начальной позиции:

$$\bullet \circ \bullet \cdots \bullet \circ \bullet \rightarrow \bullet \circ \bullet \cdots \bullet \circ \bullet \circ.$$

Обозначим через $\rho_u(x, I)$ плотность x -дефектов в конечном сегменте I , а через $\rho_u(x) := \rho_u(x, \mathbb{R})$ — верхний предел величин $\rho_u(x, I_n)$, взятый по всем возможным наборам вложенных конечных сегментов I_n , длины которых стремятся к бесконечности.

Говорят, что каплинг двух марковских процессов x^t, \dot{x}^t *почти удачен*, если верхняя плотность x -дефектов $\rho_u(x)$ стремится к нулю по времени почти наверное. Это определение существенно отличается от принятого определения удачного каплинга (см., например, [14]), которое, грубо говоря, означает, что рассматриваемые процессы со временем стремятся друг к другу.

Применяя понятие почти удачного каплинга к рассматриваемым процессам с запретами, получаем следующий условный результат.

Лемма 6.3. *Пусть $x, \dot{x} \in X$ с $\rho(x) = \rho(\dot{x}) > 0$, и предположим, что имеет место почти удачный каплинг (x^t, \dot{x}^t) , удовлетворяющий дополнительному условию о том, что расстояния между взаимно спаренными частицами равномерно ограничены сверху величиной $\gamma(t) = o(t)$. Тогда*

$$|V(x, 0, t) - V(\dot{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0.$$

7. Схема доказательства основных эргодических результатов

Начнем с двух технических результатов.

Лемма 7.1. *Супремум $|W_{ij}^t| := x_i^t - \dot{x}_j^t$ по всем взаимно спаренным частицам при динамическом каплинге (см. раздел 6) процессов (x^t, \dot{x}^t) равномерно ограничен величиной v для любого $t \in \mathbb{Z}_0$.*

Лемма 7.2. *Пусть $\rho(x) = \rho(\dot{x})$ и пусть при каплинге $\forall i, j$ найдется такой (случайный) момент времени $t_{ij} < \infty$, что $x_i^t > \dot{x}_j^t$ для любого $t \geq t_{ij}$. Тогда каплинг почти успешен.*

При наших предположениях (стандартный) удачный каплинг⁷ может не существовать (например, в случае чисто детерминированной постановки с двумя равнораспределенными начальными конфигурациями, сдвинутыми друг относительно друга). Поэтому лемма 7.1 не может быть применена непосредственно для сравнения скоростей частиц. Тем не менее оказывается, что это не является серьезным препятствием и даже отсутствие удачного каплинга может быть использовано в качестве диагностического средства.

Идея доказательства теоремы 5.1 состоит в следующем. Рассмотрим две произвольных допустимых конфигурации x, \dot{x} одинаковой плотности $\rho > 0$. Если предположить, что имеется почти удачный динамический каплинг процесса пар с начальными условиями x, \dot{x} , то по лемме 7.1 выполняются условия леммы 6.3, откуда следует, что $|V(x, 0, t) - V(\dot{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0$. Применение леммы 4.3 доказывает наше утверждение.

Предположим теперь, что нет почти удачного динамического каплинга. Определим новые случайные величины:

$$W_{ij}^t := x_i^t - \dot{x}_j^t, \quad i, j \in \mathbb{Z}, \quad t \in \mathbb{Z}_0.$$

Тогда

$$V(x, i, t) - V(\dot{x}, j, t) = W_{ij}^t / t - W_{ij}^0 / t.$$

Согласно лемме 4.3 средние скорости разных частиц в одной конфигурации стремятся друг к другу со временем. Поэтому достаточно рассмотреть случай $i = j = 0$. Для W_{00}^t возможны три следующих ситуации.

- (a) $\lim_{t \rightarrow \infty} W_{00}^t / t = 0$. Тогда $|V(x, 0, t) - V(\dot{x}, 0, t)| \leq |W_{00}^t|/t + |W_{00}^0|/t \xrightarrow{t \rightarrow \infty} 0$, что по следствию 4.4 влечет совпадение средних скоростей.
- (b) $\limsup_{t \rightarrow \infty} W_{00}^t / t > 0$. Тогда $\forall i \in \mathbb{Z}$ i -я частица x -процесса со временем обгонит любую частицу \dot{x} -процесса, исходно расположенную правее точки x_i^0 . Это наблюдение вместе с условием равенства плотностей позволяет применить лемму 7.2, согласно которой каплинг почти успешен. С другой стороны, по

⁷Удачный каплинг — когда почти все частицы со временем оказываются спаренными.

лемме 7.1 расстояния между взаимно спаренными частицами не могут превысить величины v . Поэтому по лемме 6.3 имеем $|V(x, 0, t) - V(\dot{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0$, что противоречит предположению (b).

- (c) $\limsup_{t \rightarrow \infty} W_{00}^t/t < 0$. Меняя роли процессов x^t, \dot{x}^t , возвращаемся к случаю (b).

Поэтому только предположение (a) может иметь место.

Идея доказательства теоремы 5 состоит в построении для каждого значения плотности специального семейства конфигураций, остающегося инвариантным под действием динамики. Показывается, что для любой конфигурации из этого семейства все частицы двигаются с постоянной скоростью, которая явно вычисляется. Применяя теперь результат теоремы 5.1 о том, что конфигурации одинаковой плотности имеют одинаковые средние скорости, получаем требуемое утверждение.

Литература

1. *Angel O.* The Stationary Measure of a 2-type Totally Asymmetric Exclusion Process // J. Combin. Theory Ser. A. 2006. V. 113. N. 4. P. 625–635. [arXiv/0501005 math.CO].
2. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111. N. 3–4. P. 903–930.
3. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120. N. 3–4. P. 627–658. [math.DS/0408240].
4. *Blank M.* Travelling with/against the Flow. Deterministic Diffusive Driven Systems // J. Stat. Phys. 2008. V. 133. N. 4. P. 773–796. [arXiv:0810.2205 math.DS].
5. *Blank M.* Metric properties of discrete time exclusion type processes in continuum // J. Stat. Phys. 2010. V. 140. N. 1. P. 170–197. [arXiv:0904.4585 math.DS].
6. *Бланк М. Л., Пирогов С. А.* О квазиуспешном каллинге марковских процессов // Пробл. передачи информ. 2007. Т. 43. № 4. С. 51–67.
7. *Borodin A., Ferrari P. L., Sasamoto T.* Large time asymptotics of growth models on space-like paths II: PNG and parallel TASEP // [arXiv:0707.4207 math-ph], 2007.
8. *Chowdhury D., Santen L., Schadschneider A.* Statistical physics of vehicular traffic and some related systems // Physics Reports. 2000. V. 329. P. 199–329. [arXiv:0007053 cond-mat].
9. *Comtet A., Majumdar S. N., Ouvry S., Sabhapandit S.* Integer partitions and exclusion statistics: limit shapes and the largest parts of Young diagrams // J. Stat. Mech. 2007. P10001. [arXiv:0707.2312].
10. *Evans M. R., Rajewsky N., Speer E. R.* Exact solution of a cellular automaton for traffic // J. Stat. Phys. 1999. V. 95. P. 45–98.
11. *Evans M. R., Ferrari P. A., Mallick K.* Matrix representation of the stationary measure for the multispecies TASEP // [arXiv:0807.0327 math.PR], 2008.

12. *Gray L., Griffeath D.* The ergodic theory of traffic jams // J. Stat. Phys. 2001. V. 105. N. 3–4. P. 413–452.
<http://psoup.math.wisc.edu/traffic/>
13. *Корнфельд И. П., Синай Я. Г., Фомин С. В.* Эргодическая теория. М.: Наука, 1980.
14. *Liggett T. M.* Interacting particle systems. N.Y.: Springer-Verlag, 1985.
15. *Maerivoet S., De Moor B.* Cellular Automata Models of Road Traffic // Physics Reports. 2005. V. 419. N. 1 P. 1–64.
[arXiv:physics/0509082].
16. *Nagel K., Schreckenberg M.* A cellular automaton model for freeway traffic // J. Physique I. 1992. V. 2. P. 2221–2229.
17. *Penrose M. D.* Existence and spatial limit theorems for lattice and continuum particle systems // Probab. Surveys. 2008. V. 5. P. 1–36.

О возможной динамике в модели расчета матрицы корреспонденций

1. ВВЕДЕНИЕ

После работ ряда крупных биологов первой половины прошлого века, а также работ Е. Т. Джейнса (конца 50-х годов XX века) [1], А. Дж. Вильсона (конца 60-х годов XX века) [2], И. Пригожина с коллегами, Г. Хакена (70-е годы XX века) [3, 4] в литературе достаточноочно прочно укрепилась концепция о плодотворности перенесения термодинамического формализма (см., например, [5–14] и цитированную там литературу) на различные макросистемы (в частности, встречающиеся в экономике, биологии, социальной сфере [2–4; 15–24]). В России систематические исследования в этом направлении были предприняты Л. Н. Розоноэром в начале 1970-х [25] (см. также [26–34] и цитированную там литературу). Упомянутая концепция часто используется для нахождения равновесия макросистемы. А именно, по аналогии с феноменологической термодинамикой, вводится вероятностное распределение на множестве состояний, в которых может пребывать макросистема. Такое распределение может, например, совпадать с *инвариантной мерой эргодической динамической системы*, порождающей рассматриваемую макросистему [11], или с *финальным* (равным *стационарному*) распределением эргодического (например, марковского) случайного процесса, порождающего рассматриваемую макросистему [35–40]. Если размерность макросистемы увеличивается, то, как правило, распределение сосредотачивается в окрестности наиболее вероятного макросостояния¹. Таким образом, с ростом времени наблюдения за макросистемой и размерности макросистемы следует ожидать нахождения макросистемы с большой вероятностью в малой окрестности наиболее вероятного макросостояния вне зависи-

¹ Заметим, что отмеченное обстоятельство (концентрация) может быть по-разному обосновано (как правило, достаточно элементарных комбинаторных соображений и формулы Стирлинга [1, 2, 30, 33].

симости от того, в каком состоянии макросистема находилась сначала (иначе говоря, большую часть времени (иногда, и просто, на больших временах) макросистема будет пребывать в малой окрестности наиболее вероятного макросостояния). Естественно поэтому под *равновесием макросистемы* понимать наиболее вероятное макросостояние. Задача нахождения наиболее вероятного макросостояния часто сводится (асимптотически по размерности системы) к задаче *максимизации энтропийно подобного функционала* при ограничениях (в термодинамике таким образом можно получить статистики Больцмана, Ферми–Дирака, Бозе–Эйнштейна [1, 5]). Подробнее о приложениях этой концепции см., например, [1, 2, 30, 33; 41–45].²

2. ВОЗМОЖНАЯ ДИНАМИКА, ПРИВОДЯЩАЯ В АСИМПТОТИКЕ (ПО ВРЕМЕНИ) К СТАТИЧЕСКОЙ МОДЕЛИ А. ДЖ. ВИЛЬСОНА РАСЧЁТА МАТРИЦЫ КОРРЕСПОНДЕНЦИЙ

Рассмотрим для начала более простой пример, иллюстрирующий формализм, описанный во введении.

Пример 1 (кинетика социального неравенства [23, 47]). В городе живет $N \gg 1$ (например, 10 000) пронумерованных жителей. У каждого i -го жителя есть в начальный (нулевой) момент времени целое (неотрицательное) количество рублей $s_i(0)$ (монетами, 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

² Укажем некоторые часто встречающиеся в приложениях [41–45] формализмы, также приводящие к задачам оптимизации энтропийно подобных функционалов: принцип наибольшего правдоподобия (при оценке неизвестных параметров по простой выборке); принцип максимума апостериорной вероятности; наименьшее отклонение в смысле расстояния Кульбака–Лейблера (энтропийное расстояние) [46]; принцип наименьшей неопределенности (энтропия – мера неопределенности) в теории информации (рассуждения опираются в ряде случаев на теорему Шеннона–МакМиллана–Бреймана). Важно также отметить, что энтропийно подобный функционал часто является *функцией Ляпунова* для динамической системы (например, системы обыкновенных дифференциальных уравнений, итерационного процесса, уравнения(-ий) в частных производных эволюционного типа и т.п.), порождающей рассматриваемую макросистему [12; 17–20]. Пожалуй, наиболее ярким примером этого тезиса является кинетическая теория (Л. Больцман [8]).

номерами r и l ($1 \leq r < l \leq N$) 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 $p_{k:m}(t) \equiv \kappa N^{-1}$ ($\kappa > 0$). При этом «в среднем» в единицу времени осуществляется $\kappa N/2$ встреч. Т.е., скажем, при $\kappa = 1$ в единицу времени каждый житель с вероятностью, большей $1/2$, с кем-то должен встретиться. Приблизительно такую постановку задачи в конце XVIII века предложил известный итальянский экономист Вильфредо Парето, чтобы объяснить социальное неравенство.

Пусть $c_s(t)$ – доля жителей города, имеющих ровно s рублей в момент времени t (заметим, что $c_s(t)$ – случайная величина). Пусть

$$S = \sum_{i=1}^N s_i(0), \quad \bar{s} = \frac{S}{N}.$$

Тогда по эргодической теореме для конечных однородных марковских цепей (см. [18, 19; 35–38] и упражнение ниже):³

$$\forall q = 0, \dots, S \exists \lambda_q > 0, T_q = O(N): \quad \forall t \geq T_q$$

$$P\left(\left|\frac{c_s(t)}{Ce^{-s/\bar{s}}} - 1\right| \leq \frac{\lambda_q}{\sqrt{N}}, s = 0, \dots, q\right) \geq 0.999,$$

где C определяется из условия нормировки:

$$\sum_{s=0}^S Ce^{-s/\bar{s}} = 1, \text{ т.е. } C \approx \frac{1}{\bar{s}}.$$

Скорость сходимости оценивается сверху, исходя из оценок в доказательстве эргодической теоремы для однородных марковских цепей с

³ Эргодическая теорема используется для нахождения распределения случайных величин $c_s(t)$ на больших временах. Далее используются законы больших чисел или, другими словами, явление концентрации инвариантной (стационарной) меры, о котором мы подробнее поговорим в следующем примере. Точнее не само это явление, а его следствие о том, что «хорошие» (например, липшицевы) функции на «хороших» компактных пространствах с мерой большого числа измерений почти везде близки к медиане [48].

конечным числом состояний.⁴ Как показывают численные эксперименты, оценка $O(N)$ 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». Заметим, что описанный выше случайный процесс обратим во времени. Однако наблюдается необратимая динамика $c_s(t)$. 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.

Приведем отчасти схожую постановку задачи (та же самая мера будет концентрироваться), восходящую к В. П. Маслову [33]. Ниже приведен фрагмент его интервью 2009 года, посвященного объяснению финансового кризиса 2008 г.

В. П. Маслов: *Поясню на знаменитом триюке Коровьева–Фагота – помните булгаковского героя, который разбрасывал в варьте червонцы? Понятно, что кому-то досталось больше купюр, кому-то меньше, а кто-то вообще остался ни с чем. Вопрос: если купюр миллион, то сколько должно быть зрителей, чтобы ни один не остался без банкноты? Вроде очень неопределенная задача, не*

⁴ Необходимо (для простоты формулировок, в рамках этой сноски считаем время дискретным) асимптотически (по размеру макросистемы) оценить второе по величине модуля собственное значение матрицы переходных вероятностей – инфинитезимальной матрицы (первое собственное значение, которое для неотрицательных матриц часто называют *числом Фробениуса–Персона*, равно единице, поскольку матрица стохастическая – все элементы неотрицательны и сумма всех элементов в любой строке равна единице), определяющее основание геометрической прогрессии, мажорируемой последовательность норм отклонений текущего состояния от стационарного в различные моменты времени [35–38]. Здесь нельзя не упомянуть о том, что в этом направлении за последние несколько десятков лет произошла определенная революция [49], которую можно пояснить рассмотренным примером 1. Несложно проверить, что число (макро-) состояний марковской цепи в этом примере растет быстрее, чем экспоненциально с ростом N . В то время как по прошествии всего лишь $O(N)$ тактов распределение цепи будет уже довольно близко к финальному (предельному) = стационарному (инвариантному). Таким образом, если возникает потребность быстро сгенерировать дискретные случайные величины, которые могут принимать огромное число значений, то в ряде случаев удается подобрать такую марковскую цепь, которая быстро «выйдет» на стационарный режим, соответствующий желаемому распределению. Несколько интересных примеров в этом направлении (модель Изинга и др.) собрано, например, в современном курсе марковских случайных процессов [38]. Заметим, что при оценках второго по величине модуля собственного значения активно используется уже упоминавшийся принцип концентрации меры (см. [49, 50] и цитированную там литературу, а также приложение А. В. Колесникова).

имеющая однозначного решения. И тем не менее ответ есть: примерно квадратный корень из миллиона, то есть тысяча зрителей.

Точнее говоря, как следует из выше написанного, с вероятностью, близкой к 1, доля банкротов будет равна примерно $1/\bar{s} \sim 0.001$, поскольку по условию $\bar{s}N = S = 10^6$ и $N \sim \sqrt{S}$. Поэтому количество банкротов с вероятностью, близкой к 1, незначительно отличается от $N/\bar{s} \sim 1$.

Упражнение (принцип сжимающих отображений и фокусирующие операторы, эргодическая теорема для конечных однородных марковских цепей [51])*.

a) Покажите, что 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) Пусть $X = \mathbb{PR}_{++}^n$ – множество лучей пространства \mathbb{R}^n , лежащих во внутренности неотрицательного ортантта, на котором введена метрика Биркгофа:

$$\rho(x, y) = \ln \min \left\{ \frac{\beta}{\alpha}; \alpha x \leq y \leq \beta x \right\} = \ln \min_{j, k=1, \dots, n} \frac{x_j y_k}{x_k y_j}.$$

Покажите, что X – полное метрическое пространство. Покажите, что если линейный оператор $A: X \rightarrow X$ ($A = \|a_{ij}\|_{i, j=1}^n$ – 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 Фробениус–Перрона, или эргодическая теорема для конечных однородных дискретных марковских цепей (д.м.ц.)). Следующие условия для стохастической матрицы $n \times n$ P равносильны:

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

$$2) P \text{ – эргодическая матрица, т.е. } \exists \vec{p}^* > \vec{0}: \lim_{m \rightarrow \infty} P^m = \underbrace{\left[\vec{p}^*, \vec{p}^*, \dots, \vec{p}^* \right]}_n^T,$$

причём \vec{p}^* является единственным решением системы:

$$\vec{p}^{*T} = \vec{p}^{*T} P, \sum_{k=1}^n p_k^* = 1. \text{⁵} \quad (\text{S})$$

Пример 2 (модель расчета матрицы корреспонденций [2]). В некотором городе имеется n районов, $L_i > 0$ – число жителей i -го района, $W_j > 0$ – число работающих в j -м районе (число рабочих мест), $x_{ij}(t) \geq 0$ – число жителей, живущих в i -м районе и работающих в j -м в момент времени $t \geq 0$. Со временем пронумерованные жители (количество которых не меняется⁶ и равно $N = \sum_{i=1}^n L_i = \sum_{j=1}^n W_j$) меняют места жительства (квартиры). Считается, что отмеченные изменения могут происходить только за счёт обмена квартирами, т.е.

⁵ Заметим, что из вида матрицы $\lim_{m \rightarrow \infty} P^m = \underbrace{\left[\vec{p}^*, \vec{p}^*, \dots, \vec{p}^* \right]^T}_n$ следует, что

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

Это условие означает равенство финального распределения ($\lim_{m \rightarrow \infty} \vec{p}(m)$) стационарному \vec{p}^* ($\vec{p}^* = P^T \vec{p}^*$) вне зависимости от начального распределения $\vec{p}(0)$. Заметим также, что условия 1), 2) равносильны следующим требованиям: конечная однородная д.м.ц. с матрицей переходных вероятностей P – неразложимая = неприводимая (т.е. из произвольного состояния «можно прийти» в любое, наперед заданное) и непериодическая (Н.О.Д. $\left\{ k : \left[P^k \right]_{11} > 0 \right\} = 1$). Если убрать условие непериодичности, то

$$\exists! \vec{p}^* > \vec{0} \quad (\vec{p}^* \in S): \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N P^m = \underbrace{\left[\vec{p}^*, \vec{p}^*, \dots, \vec{p}^* \right]^T}_n \quad (\text{сходимость по Чезаро}).$$

Если перейти к непрерывному времени, осуществляя соответствующий скейлинг, то легко показать, что необходимость в условии непериодичности исчезает. Если же цепь разложима, то система (S), вообще говоря, уже будет разрешима не единственным образом. Финальное распределение существует, но уже может зависеть от того, с какого распределения стартуем. Соответствующий вариант эргодической теоремы приведен, например, в [37]. Доказательство в [37] также базируются на принципе сжимающих отображений. Другое, более вероятностное, доказательство эргодической теоремы для конечных однородных д.м.ц. имеется, например, в [35, 38] и базируется на каплинге (см. также приложение М. Л. Бланка).

⁶ Поскольку мы будем следить за системой на больших временах, то сделанное предположение кажется неестественным. Заметим, однако, что если «номер жителя» передаётся его потомкам (номер папы передаётся сыну, номер мамы – дочери), то предположение о постоянстве состава жителей выглядит разумным в первом приближении. Здесь мы также пренебрегаем миграционными потоками (город изолирован).

$$x_{ij}(t) \geq 0, \sum_{j=1}^n x_{ij}(t) \equiv L_i, \sum_{i=1}^n x_{ij}(t) \equiv W_j, i, j = 1, \dots, n. \quad (\text{A})$$

Пусть в момент времени $t \geq 0$ r -й житель живет в k -м районе и работает в m -м, а l -й житель живет в p -м районе и работает в q -м. Тогда $p_{k,m; p,q}^L(t)\Delta t + o(\Delta t)$ – есть вероятность того, что жители с номерами r и l ($1 \leq r < l \leq N$) поменяются квартирами в промежутке времени $(t, t + \Delta t)$. Естественно считать, что вероятность в единицу времени обмена местами жительства зависит только от мест проживания и работы обменивающихся.

Например, можно считать, что время, потраченное в пути от района i до района j , есть $t_{ij} \geq 0$ (вместо t_{ij} в приводимую ниже формулу также осмысленно подставлять $l_{ij} \geq 0$ – расстояние от района i до района j), а

$$p_{k,m; p,q}^L(t) \equiv p^L N^{-1} \exp\left(\gamma \cdot (t_{km} + t_{pq} - (t_{pm} + t_{qa})) / 2\right) > 0,$$

где $p^L > 0$, $\gamma > 0$. Тогда по эргодической теореме для конечных однородных марковских цепей (см. [18, 19; 35–38]):

$$\begin{aligned} \forall \left\{x_{ij}\right\}_{i=1, j=1}^{n, n} \in (\text{A}) \rightarrow \lim_{t \rightarrow \infty} P(x_{ij}(t) = x_{ij}, i, j = 1, \dots, n) = \\ = Z^{-1} \prod_{i=1, j=1}^{n, n} \exp(-\gamma t_{ij} x_{ij}) \cdot (x_{ij} !)^{-1} \stackrel{\text{def}}{=} p\left(\left\{x_{ij}\right\}_{i=1, j=1}^{n, n}\right), \end{aligned}$$

где статсумма Z находится из условия нормировки получившейся пуассоновской вероятностной меры [52]. Распределение $p\left(\left\{x_{ij}\right\}_{i=1, j=1}^{n, n}\right)$ на множестве (A) сконцентрировано при $N \gg 1$ (см. ниже) в окрестности наиболее вероятного значения $\left\{x_{ij}^*\right\}_{i=1, j=1}^{n, n}$, которое находится, как решение задачи энтропийно линейного программирования:⁷

⁷ Нетрудно заметить, что задача поиска наиболее вероятного состояния асимптотически (по n) эквивалентна задаче максимизации энтропийного функционала (воспользовавшись формулой Стирлинга):

$$\ln p\left(\left\{x_{ij}\right\}_{i=1, j=1}^{n, n}\right) \sim - \sum_{i=1, j=1}^{n, n} x_{ij} \ln(x_{ij}/e) - \gamma \sum_{i=1, j=1}^{n, n} t_{ij} x_{ij} + \text{const}_n$$

$$\sum_{i=1, j=1}^{n, n} x_{ij} \ln(x_{ij}/e) + \gamma \sum_{i=1, j=1}^{n, n} t_{ij} x_{ij} \rightarrow \min_{\{x_{ij}\}_{i=1, j=1}^{n, n} \in (A)} . \quad (1)$$

Решение задачи (1) можно представить как

$$x_{ij} = \exp(-1 - \lambda_i^L - \lambda_j^W - \gamma t_{ij}),$$

на множестве **(A)**. Поскольку функционал строго вогнутый и решение задачи максимизации, без предположения целочисленности $\{x_{ij}\}_{i=1, j=1}^{n, n}$, считаем таким, что $x_{ij}^* \gg 1$ (так как $n \gg 1$), то ограничение «целочисленности» является для данной задачи несущественным, и применение асимптотической формулы Стирлинга было законным. Обратим внимание также на то, что задача максимизации энтропийного функционала на множестве **(A)**, т.е. по *принципу Лагранжа* [53] (балансовые ограничения **(A)** перенесли в функционал) задача

$$\begin{aligned} L\left(\{x_{ij}\}_{i=1, j=1}^{n, n}; \bar{\lambda}^L, \bar{\lambda}^W\right) = & - \sum_{i=1, j=1}^{n, n} x_{ij} \ln(x_{ij}/e) - \gamma \sum_{i=1, j=1}^{n, n} t_{ij} x_{ij} + \\ & + \sum_{i=1}^n \lambda_i^L \cdot \left(\sum_{j=1}^n x_{ij} - L_i \right) + \sum_{j=1}^n \lambda_j^W \cdot \left(\sum_{i=1}^n x_{ij} - W_j \right) \rightarrow \sup_{x_{ij} \geq 0, i, j=1, \dots, n} \end{aligned}$$

имеет и притом единственное решение $\{x_{ij}^*\}_{i=1, j=1}^{n, n}$, которое определяется из системы:

$$\partial L\left(\{x_{ij}\}_{i=1, j=1}^{n, n}; \bar{\lambda}^L, \bar{\lambda}^W\right) / \partial x_{ij} = -\ln x_{ij} + \lambda_i^L + \lambda_j^W - \gamma t_{ij} = 0, \quad i, j = 1, \dots, n.$$

Приятной особенностью класса задач энтропийно линейного программирования (задач максимизации энтропийно подобных функционалов, при наличии линейных ограничений в виде равенств и неравенств на неотрицательном ортанте) является явная (легко выписываемая) зависимость решения прямой задачи, через *двойственные переменные*. Поскольку количество ограничений (количество двойственных переменных – *множителей Лагранжа*), как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы отыскания решений базируются на решении *двойственной задачи*, представляющей собой задачу минимизации выпуклой функции на неотрицательном ортанте [30, 45, 54]. Можно показать, см. [54], что многие популярные в литературе [30, 45] численные методы решения этой двойственной задачи являются барьерно-мультиплекативными аналогами квазиградиентных итерационных методов. В частности, к ним относится один из первых методов: «метод балансировки», заключающийся в подстановке прямых переменных, как функций двойственных, в систему ограничений (в методе предполагается, что есть ограничения только в виде равенств) и разрешение полученной системы (размерность которой как раз равна числу двойственных переменных) относительно двойственных переменных. Для рассматриваемого далее частного случая $\forall i, j = 1, \dots, n \rightarrow t_{ij} = \tau > 0$, все это можно сделать аналитически, и в результате получить формулу (2). Отметим здесь также эффективность сепарельных (функционал декомпозируется в аддитивную сумму функций одного аргумента) алгоритмов типа Нестерова–Немировского для задач энтропийного программирования, возникающих при нахождении равновесий макросистем [55].

где множители Лагранжа (двойственные переменные) $\{\lambda_i^L\}_{i=1}^n$ и $\{\lambda_j^W\}_{j=1}^n$ определяются из равенств системы (A). На практике мы имеем информацию о $\{L_i, W_i\}_{i=1}^n$ и $\{t_{ij}\}_{i=1, j=1}^{n, n}$. Решив задачу (A), мы найдем

$$\left\{x_{km} \left(\{L_i, W_i\}_{i=1}^n; \{t_{ij}\}_{i=1, j=1}^{n, n} \right) \right\}_{k=1, m=1}^{n, n}.$$

Такой способ расчета матрицы корреспонденций в литературе часто называют *гравитационной моделью*:

$$x_{ij} = A_i B_j L_i W_j \exp(-\gamma t_{ij}),$$

где $\{A_i\}_{i=1}^n$ и $\{B_j\}_{j=1}^n$ определяются из соотношений [2, 30, 32]:

$$A_i = \left(\sum_{j=1}^n B_j W_j \exp(-\gamma t_{ij}) \right)^{-1}, \quad B_j = \left(\sum_{i=1}^n A_i L_i \exp(-\gamma t_{ij}) \right)^{-1}.$$

Описанная модель (точнее говоря, рассчитанная по этой модели матрица корреспонденций) активно используется в теоретико-игровых моделях (например, базирующихся на *принципах Дж. Г. Вардрона*) равновесного распределения потоков [32] (см. также [56] и главу 1). Подробнее об этих моделях речь идет в главе 1 (Е. А. Нурминского и Н. Б. Шамрай). Один из возможных способов определения времени в пути, в зависимости от загрузки ребра, предложен в приложении М. Л. Бланка. Заметим также, что задача (1) по своим свойствам очень похожа на транспортную задачу (см. приложение А. В. Колесникова).

Перейдем теперь к исследованию полученного стационарного распределения вероятностей $p\left(\{x_{ij}\}_{i=1, j=1}^{n, n}\right)$ на макросостояниях

$$\{x_{ij}\}_{i=1, j=1}^{n, n} \in (A).$$

$$N \sim n^{\alpha+1}, \quad L_i, W_j \sim n^\alpha \quad (\alpha > 1), \quad n \gg 1, \quad \forall i, j = 1, \dots, n \rightarrow t_{ij} = \tau > 0,$$

⁸ Отметим, что хотя в этом случае динамика рассматриваемой нами макросистемы обратима по времени (так же, как и в примере 1), макросистема (в каком бы состоянии она не находилась в нулевой момент времени) по прошествии достаточно большого времени окажется в малой окрестности равновесного макросостояния (характеризующегося наибольшим из возможных значением энтропии) и будет в дальнейшем пребывать в этой окрестности подавляющую часть времени. Схожая ситуация имеет место и в статистической физике (см., например, [8, 11, 14, 18, 19, 27]).

то распределение вероятностей $p\left(\{x_{ij}\}_{i=1, j=1}^{n, n}\right)$ на множестве **(A)** сконцентрировано в $O(n^{(\alpha-1)/2})$ окрестности (почему именно в такой окрестности, будет показано ниже) наиболее вероятного значения

$$x_{ij}^* \approx L_i W_j / N \sim n^{\alpha-1}, \quad i, j = 1, \dots, n, \quad (2)$$

которое ищется с помощью *метода множителей Лагранжа* [2, 53]. Формулу (2) можно интерпретировать как наличие у районов «потенциалов притяжения» [2]:

$$L_i / \sqrt{N} \sim \exp(\lambda_i^L), \quad i = 1, \dots, n \text{ и } W_j / \sqrt{N} \sim \exp(\lambda_i^W), \quad j = 1, \dots, n,$$

произведение которых даёт пассажиропоток x_{ij}^* , $i, j = 1, \dots, n$ (асимптотически совпадающий с математическим ожиданием и медианой).

Исследуем теперь, следуя [1, 2, 30], явление концентрации стационарного распределения $p\left(\{x_{ij}\}_{i=1, j=1}^{n, n}\right)$ в окрестности наиболее вероятного значения $\{x_{ij}^*\}_{i=1, j=1}^{n, n}$. Для этого прежде всего заметим, что из определения $\{x_{ij}^*\}_{i=1, j=1}^{n, n}$ (см. также сноска 7) следует

$$\forall \{x_{ij}\}_{i=1, j=1}^{n, n} \in (\text{A}) \rightarrow \sum_{i=1, j=1}^{n, n} \frac{\partial \ln p\left(\{x_{ij}^*\}_{i=1, j=1}^{n, n}\right)}{\partial x_{ij}} \cdot (x_{ij} - x_{ij}^*) \leq 0.$$

Поэтому

$$\begin{aligned} & \forall \{x_{ij}\}_{i=1, j=1}^{n, n} \in (\text{A}) \exists \theta \in [0, 1]: \\ & \ln p\left(\{x_{ij}\}_{i=1, j=1}^{n, n}\right) \leq \ln p\left(\{x_{ij}^*\}_{i=1, j=1}^{n, n}\right) + \\ & + \sum_{i=1, j=1}^{n, n} \frac{\partial^2 \ln p\left(\{x_{ij}^* \theta + x_{ij} \cdot (1-\theta)\}_{i=1, j=1}^{n, n}\right)}{\partial x_{ij}^2} \cdot \frac{(x_{ij} - x_{ij}^*)^2}{2}. \end{aligned}$$

Но

$$\frac{\partial^2}{\partial x_{ij}^2} \left(\ln p\left(\{x_{ij}\}_{i=1, j=1}^{n, n}\right) \right) = \frac{\partial^2}{\partial x_{ij}^2} \left(- \sum_{i=1, j=1}^{n, n} x_{ij} \ln x_{ij} \right) = -\frac{1}{x_{ij}}.$$

Следовательно, приходим к «неравенству о концентрации меры»:

$$\forall M > 0, \forall \{x_{ij}\}_{i=1, j=1}^{n, n} \in (\mathbf{A}): \sum_{i=1, j=1}^{n, n} \frac{(x_{ij} - x_{ij}^*)^2}{2 \max\{x_{ij}, x_{ij}^*\}} \geq M$$

$$P\left(\left\{x_{ij}\right\}_{i=1, j=1}^{n, n}\right) \leq e^{-M} P\left(\left\{x_{ij}^*\right\}_{i=1, j=1}^{n, n}\right).$$

Из этого неравенства имеем результат о *концентрации распределения* $P\left(\left\{x_{ij}\right\}_{i=1, j=1}^{n, n}\right)$ на множестве **(A)** в $O(n^{(\alpha-1)/2})$ окрестности наиболее вероятного значения $\{x_{ij}^*\}_{i=1, j=1}^{n, n}$:

$$\exists \lambda > 0: \lim_{t \rightarrow \infty} P\left(\forall i, j = 1, \dots, n \rightarrow |x_{ij}(t)/x_{ij}^* - 1| \leq \lambda/n^{(\alpha-1)/2}\right) \geq 0.999.$$

Упражнение (М. С. Ишманов, 2010). а)** Верно ли, что скорость сходимости в этом соотношении оценивается как $O(n^{\alpha+1})$?

б) Верно ли, что при фиксированном n и $N \rightarrow \infty$ скорость сходимости (в том же смысле, что и в п. а)) оценивается как $O(N)$?

Замечание (о других возможных подходах к исследованию концентрации стационарного распределения). Один из способов восходит к методу вычисления математических ожиданий Дарвина–Фаулера [2, 5, 9] (метод производящих функций и анализ их асимптотического поведения методом перевала) – в этом случае концентрация наблюдается в окрестности математического ожидания (интересные приложения этого метода в комбинаторике имеются, например, в [57], см. также приложение А. А. Замятиной и В. А. Малышева в этой книге и конец следующего пункта). Исследование концентрации в окрестности математического ожидания можно также приводить, например, используя предельные меры [58], метод канонических ансамблей [59] или обобщённую схему размещения [60], нашедшие применения к задачам асимптотической перечислительной комбинаторики,⁹ к исследованию случайных матриц и уравнений, к изучению статистических свойств группы перестановок с приложениями к теории разбиений (диаграммам Юнга) и асимптотической теории чисел, а также к теории предельных форм. К методу производящих функций также тесно примыкают метод моментов [60], метод пуассоновской и гауссовской аппроксимации (метод локальной предельной теоремы) [7, 60]. Друг-

⁹ В частности, к теории случайных графов (компьютерные, транспортные сети – вопросы надёжности и т.п. [24, 39], см. также приложение А. М. Райгородского в этой книге).

гой способ восходит к *принципу концентрации* А. Пуанкаре и П. Леви, получившему дальнейшее развитие в работах В. Д. Мильмана и др. [61] – в этом случае концентрация наблюдается в окрестности медианы.¹⁰ В заключение краткого обзора методов исследования концентрации меры упомянем теоремы тауберова типа [63] и мартингальные неравенства [62]. Ввиду всего вышесказанного важно отметить, что поиск наиболее вероятного распределения – это особенность работ, в которых изучаются равновесия макросистем. Но также важно заметить, что наиболее вероятное распределение в содержательных задачах асимптотически (по размеру системы) эквивалентно математическому ожиданию (в работе [2] соответствующие выкладки проделаны на примере модели расчета матрицы корреспонденций А. Дж. Вильсона, см. также следующий пункт) и медиане [61].

Замечание (модели Д. Бернулли–Лапласа, П. и Т. Эренфестов). Важно обратить внимание на то, что описанный способ изучения равновесных состояний макросистем применим к достаточно широкому классу макросистем (модель Д. Бернулли–Лапласа, модель Эренфестов, круговая модель М. Каца и др. (см., например, [8, 27, 64] и цитированную там литературу)), в том числе встречающихся в экономике, биологии, социальной сфере [3, 4; 15–24].

3. ОБЩАЯ СХЕМА ИССЛЕДОВАНИЯ РАВНОВЕСИЙ МАКРОСИСТЕМ

Ниже приводится (во многом под влиянием работ [18, 19]) общая схема, в которую «ложатся» примеры 1 и 2.

Предположим, что некоторая макросистема может находиться в различных состояниях, характеризуемых вектором \vec{n} с неотрицательными целочисленными компонентами (скажем, в модели «кинетика социального неравенства» $n_i(t)$ – количество жителей города, имеющих в момент времени $t \geq 0$ i рублей). 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

¹⁰ Этот принцип также нашёл широкие применения в асимптотической перечислительной комбинаторике; в качестве достаточно известного примера можно упомянуть неравенство М. Талаграна и его приложения к изучению макросвойств (связность и т.п.) случайных графов [62] (см. также приложения А. М. Райгородского и А. В. Колесникова в этой книге).

возможные типы реакций, при этом $(\vec{\alpha}, \vec{\beta}) \in J \Rightarrow (\vec{\beta}, \vec{\alpha}) \in J$.

Введем, следуя Березину–Клесову (1978) [19], *интенсивность реакции*:

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

где $K_{\vec{\beta}}^{\vec{\alpha}} \geq 0$ – константы реакции (в химической кинетике – постоянные, а в социодинамике (В. Вайдлих, 2000 [20]) – необязательно); при этом часто считают $\sum_i n_i \equiv M$. Т.е. $\lambda_{(\vec{\alpha}, \vec{\beta})}(\vec{n})$ –

вероятность осуществления в единицу времени перехода $\vec{n} \rightarrow \vec{n} - \vec{\alpha} + \vec{\beta}$: в единицу времени равновероятно выбираются («приближение среднего поля») любые два жителя города и в зависимости от того, в каких состояниях они находились, «случайно» переводятся (разыгрывают один рубль) в новые состояния. На макроуровне все это соответствует принципам химической кинетики (закон действующих масс Гульдберга–Вааге, 1864 [18]).

Предположим теперь, что множество J не зависит от M , и в начальный момент времени для любого i существует предел

$$c_i(0) = \lim_{M \rightarrow \infty} n_i(0)/M.$$

Тогда (Малышев–Пирогов–Рыбко, 2004 [19, 65]) 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)

$$c_i(t) = \lim_{M \rightarrow \infty} n_i(t)/M.$$

Описанный выше приём называется *каноническим скейлингом*. В результате такого скейлинга приходим к «динамике квазисредних» (терминология В. Вайдлиха [20]):

$$\frac{dc_i}{dt} = \sum_{(\vec{\alpha}, \vec{\beta}) \in J} (\beta_i - \alpha_i) K_{\vec{\beta}}^{\vec{\alpha}}(\vec{c}) \prod_j c_j^{\alpha_j}. \quad (3)$$

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 (1) 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 $n_i(t)$.¹¹ Пусть существует хотя бы одно положительное решение $\vec{\xi}$ системы (*условие динамического равновесия*, В. В. Веденяпин (2001) [18], Малышев–Пирогов–Рыбко (2004) [19, 65], Батищева–Веденяпин [66]):

$$\forall \vec{\alpha} \rightarrow \sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = \sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j} \quad (K_{\vec{\beta}}^{\vec{\alpha}}(\vec{c}) \equiv K_{\vec{\beta}}^{\vec{\alpha}}). \quad (4)$$

Частным, но часто встречающимся в приложениях, случаем условия (4) является условие детального равновесия:

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

Несложно проверить, что $\vec{\xi}$, удовлетворяющее (4), является неподвижной точкой системы (3). Более того, если существует хотя бы одно положительное решение системы (4), то тогда все неподвижные точки системы (3) удовлетворяют условию (4), также называемому *условием унитарности*, или *условием Штюкельберга–Батищевой–Пирогова*. Покажем, во многом следуя В. В. Веденяпину [18], что в этом случае (существования хотя бы одного положительного решения (4)) траектория (3) сходится к неподвижной точке (какой именно, зависит, вообще говоря, от «точки старта»).¹² 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 (3). Посчитаем полную производную H в силу системы (3):

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

¹¹ Заметим, что этот переход и возможность его использования в правой части равенства (3) нуждаются в строгом обосновании (и далеко не всегда правомочны). В качестве примера, укажем популярный в литературе [3; 16–22] марковский процесс «рождения–гибели» (приводящий к системе уравнений «хищник–жертва»), для которого «флуктуации играют решающую роль, качественно меняя выводы макроскопического анализа».

¹² Стоит заметить, что аттрактор системы (3) с постоянными коэффициентами реакции, по-видимому, в общем случае может быть сколь угодно сложным множеством [19, 67].

$$+ \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} \cdot \sum_i (\beta_i - \alpha_i) = \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} \cdot \ln \prod_i y_i^{\beta_i - \alpha_i} ,$$

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 $u = 1$.

Оказывается (Малышев–Пирогов–Рыбко [19, 65]), что условие (4) можно проинтерпретировать как условие инвариантности пуассоновской меры [52]:¹³

$$\mu(\vec{n}) = \prod_i \lambda_i^{n_i} e^{-\lambda_i} / n_i ! ,$$

где $\lambda_i = \xi_i^* M$, а $\vec{\xi}^*$ – произвольное решение (4); относительно предложенной стохастической марковской динамики. Эта мера экспоненциально быстро концентрируется, с ростом n_i , в окрестности наиболее вероятного состояния (также удовлетворяющего условию (4)), 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):

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

на множестве, заданном (как правило, линейными) ограничениями – законами сохранения (интегралами движения для (3)).

Действительно, будем считать, что ограничения (законы сохранения) задаются СЛАУ $A\vec{n} = \vec{d}$, где $A = \|A_{ki}\|$ – матрица макси-

¹³ Пуассоновская мера также часто встречается (при наиболее естественных условиях) в качестве притягивающей инвариантной меры, если осуществляется термодинамический предельный переход (число агентов и состояний стремятся, сохраняя пропорции, к бесконечности) [68].

мального ранга ($k = 1, \dots, m$). Обозначим через \mathbf{A} множество неотрицательных целочисленных векторов \vec{n} , удовлетворяющих $A\vec{n} = \vec{d}$. Тогда равновесие \vec{n}^* находится как решение задачи $E(\vec{n}) \rightarrow \max_{\vec{n} \in \mathbf{A}}$ (поскольку функционал строго вогнутый и считаем $n_i^* \gg 1$, то целочисленностью переменных можно пренебречь). Используя принцип Лагранжа, можно показать, что решение этой задачи представляется в виде

$$n_i(\vec{y}^*) = \lambda_i \exp\left(\sum_k A_{ki} y_k^*\right),$$

где двойственные переменные (множители Лагранжа) \vec{y}^* определяются из системы $A\vec{n}(\vec{y}) = \vec{d}$.

Приведем, во многом следуя [2], другой путь (восходящий к Дарвину–Фаулеру), по которому можно прийти к аналогичным формулам.

Для этого введем производящую функцию

$$\begin{aligned} F(\vec{z}; A) &= \sum_{\vec{n} \geq 0} \mu(\vec{n}) \prod_k z_k^{\sum A_{kl} n_l} = \prod_i \sum_{n_i \geq 0} \frac{\left(\lambda_i \prod_k z_k^{A_{ki}} \right)^{n_i}}{n_i!} e^{-\lambda_i} = \\ &= \prod_i \exp\left(\lambda_i \cdot \left(\left(\prod_k z_k^{A_{ki}} \right) - 1 \right) \right). \end{aligned}$$

Тогда по формуле Коши:

$$\begin{aligned} E\left[n_{r_1}^{p_1} \cdots n_{r_Q}^{p_Q}\right] &= \frac{1}{Z} \frac{1}{(2\pi i)^m} \oint dz_1 \cdots dz_m \times \\ &\times \left[\left\{ \prod_k z_k^{-d_k-1} \right\} \cdot \left\{ \prod_q \left(\frac{1}{\ln z_1} \frac{\partial}{\partial A_{1r_q}} \right)^{p_q} F(\vec{z}; A) \right\} \right], \\ Z &= \frac{1}{(2\pi i)^m} \oint dz_1 \cdots dz_m \cdot \left[\left\{ \prod_k z_k^{-d_k-1} \right\} \cdot F(\vec{z}; A) \right]. \end{aligned}$$

Здесь математическое ожидание $E\left[n_{r_1}^{p_1} \cdots n_{r_Q}^{p_Q}\right]$ считается по вероятностной мере, порожденной мерой Пуассона $\mu(\vec{n})$, а интегралы по

dz_k берутся в комплексной плоскости по замкнутым контурам, охватывающим точку ноль. Используя метод перевала [69], асимптотически оценим математическое ожидание:

$$E[n_{r_1}^{p_1} \cdot \dots \cdot n_{r_q}^{p_q}] \approx \frac{1}{F(\bar{z}^*; A)} \left(\ln z_1^* \right)^{-\sum_q p_q} \cdot \left\{ \prod_q \left(\frac{\partial}{\partial A_{1r_q}} \right)^{p_q} F(\bar{z}^*; A) \right\},$$

где «точка перевала» \bar{z}^* определяется как решение системы:

$$z_k \cdot (\partial F(\bar{z}; A) / \partial z_k) \approx d_k \cdot F(\bar{z}; A), \quad k = 1, \dots, m.$$

В частности,¹⁴

$$E[n_i] \approx \lambda_i \prod_k (z_k^*)^{A_{ki}}, \quad D[n_i] \approx \lambda_i \prod_k (z_k^*)^{A_{ki}},$$

где \bar{z}^* определяется как решение системы уравнений:

$$\sum_i A_{ki} \cdot \left\{ \lambda_i \prod_k z_k^{A_{ki}} \right\} = d_k, \quad k = 1, \dots, m.$$

Очевидна связь «точки перевала» \bar{z}^* с двойственными переменными $\bar{y}^* : z_k^* = \exp(y_k^*)$.

4. ЗАКЛЮЧЕНИЕ

В приложении обсуждается концепция равновесия макросистемы. Приводятся различные подходы к обоснованию следующего принципа: *равновесие = наиболее вероятное макросостояние инвариантной (стационарной) меры динамической системы (марковского процесса), порождающей исследуемую макросистему*. Рассматриваются примеры конкретных макросистем. В частности, один из примеров «объясняет» популярную в приложениях модель А. Дж. Вильсона расчета матрицы корреспонденций.

Повторим в заключение описанную в приложении схему.

1. Макросистема состоит из огромного числа пронумерованных агентов, каждый из которых может находиться в одном из возможных состояний. Число состояний, как минимум, на несколько по-

¹⁴ Обратим внимание на то, что получилось $E[n_i] \approx D[n_i] \gg 1$ – это означает концентрацию распределения случайной величины n_i в $\sqrt{D[n_i]}$ окрестности своего математического ожидания $E[n_i]$.

рядков меньше числа агентов (иногда можно и без этого требования). Распределение агентов (с учетом их номеров) по состояниям будем называть микросостоянием, а без учета номеров – макросостоянием.

2. Задана марковская динамика распределения агентов по состояниям, в основу которой на микроуровне положена равноправность агентов одного типа (в приближении среднего поля) и заранее прописанные возможности случайных превращений (переходов) агентов (химические реакции): равновероятно выбирается агент и в зависимости от того, в каком состоянии он находится, «случайно» переводится в новое состояние. Аналогично рассматриваются парные взаимодействия и взаимодействия, в которых участвует большее число агентов. На макроуровне это соответствует принципам химической кинетики (Гульдберг–Вааг, 1864).

Предполагается, что из любого возможного макросостояния можно перейти согласно такой динамике в любое другое (характерное время такого перехода определяет скорость сходимости к равновесию). Также считается, что описанная динамика имеет макрозаконы сохранения. Соотношения (как правило, линейные) между макропараметрами, которые не меняются со временем.

Пусть выполняется условие: динамика задана линейной полугруппой (однородность), динамика «обратима» (детальный баланс, условие динамического равновесия).

Тогда эргодическая марковская динамика приводит на больших временах к стационарной (инвариантной) пуассоновской (сложной) мере (прямое произведение распределений Пуассона) на пространстве макросостояний. Эта мера экспоненциально быстро концентрируется, с ростом числа агентов, в окрестности наиболее вероятного макросостояния, которое и принимается за положение равновесия макросистемы. Задача поиска наиболее вероятного макросостояния асимптотически (по числу агентов) эквивалентна задаче максимизации энтропийного функционала (воспользовались формулой Стирлинга) на множестве (как правило, аффинной структуры), заданном ограничениями – законами сохранения. Этот же энтропийный функционал (взятый со знаком минус) возникает как функция Ляпунова динамики, полученной в результате канонического скейлинга исходной марковской динамики. Отыскание предельной неподвижной точки (этой динамики), в которую придет система, сводится к решению той же самой задачи энтропийно линейного программирования.

ЛИТЕРАТУРА

1. *Jaynes E. T.* Probability theory. The logic of science. Cambridge: Cambridge University Press, 2003; Papers on probability, statistics and statistical physics. Dordrecht: Kluwer Academic Publisher, 1989.
2. *Вильсон А. Дж.* Энтропийные методы моделирования сложных систем. М.: Наука, 1978.
3. *Николис Г., Пригожин И.* Самоорганизация в неравновесных системах. М.: Мир, 1979.
4. *Хакен Г.* Информация и самоорганизация. Макроскопический подход к сложным системам. М.: УРСС, 2005.
5. *Шредингер Э.* Статистическая термодинамика. М.: ИЛ, 1948.
6. *Крылов Н. С.* Работы по обоснованию статистической физики. М.-Л.: Издательство АН СССР, 1950.
7. *Хинчин А. Я.* Математические основания статистической механики. М.-Ижевск: НИЦ «РХД», ИКИ, 2003.
8. *Кац М.* Вероятность и смежные вопросы в физике. М.: Мир, 1965.
9. *Хуанг К.* Статистическая механика. М.: Мир, 1966.
10. *Рюэль Д.* Статистическая механика. Строгие результаты. М.: Мир, 1971.
11. *Корнфельд И. П., Синай Я. Г., Фомин С. В.* Эргодическая теория. М.: Наука, 1980.
12. *Evans L. C.* Entropy and partial differential equations. Department of mathematics, UC Berkeley, 2003. <http://math.berkeley.edu/~evans/>
13. *Минлос Р. А.* Введение в математическую статистическую физику. М.: МЦНМО, 2002.
14. *Козлов В. В.* Ансамбли Гиббса и неравновесная статистическая механика. М.-Ижевск: НИЦ «РХД», ИКИ, 2008.
15. *Марри Дж.* Нелинейные дифференциальные уравнения в биологии. М.: Мир, 1983.
16. *Смирежев Ю. М.* Нелинейные волны, диссипативные структуры и катастрофы в экологии. М.: Наука, 1987.
17. *Гардинер К. В.* Стохастические методы в естественных науках. М.: Мир, 1986.
18. *Веденяпин В. В.* Кинетическая уравнения Больцмана и Власова. М.: Физматлит, 2001.

19. *Малышев В. А., Пирогов С. А.* Обратимость и необратимость в стохастической химической кинетике // УМН. 2008. Т. 63. № 1. С. 3–36.
20. *Вайдлих В.* Социодинамика: системный подход к математическому моделированию в социальных науках. М.: УРСС, 2010.
21. *Castellano C., Fortunato S., Loreto V.* Statistical physics of social behavior // Review of modern physics. 2009. V. 81. P. 591–646.
[arXiv:0710.3256v2](https://arxiv.org/abs/0710.3256v2)
22. *Занг В.-Б.* Синергетическая экономика: время и перемены в нелинейной экономической теории. М.: Мир, 1999.
23. *Drăgulescu A., Yakovenko V. M.* Statistical mechanics of money // The European Physical Journal B. 2000. V. 17. P. 723–729.
[arXiv:cond-mat/0001432v4](https://arxiv.org/abs/cond-mat/0001432v4)
24. *Baldi P., Frasconi P., Smyth P.* Modeling the Internet and the Web: Probabilistic methods and algorithms. Published by John Wiley & Sons, 2003.
25. *Розоноэр Л. И.* Обмен и распределение ресурсов (обобщенный термодинамический подход) I, II, III // Автоматика и телемеханика. 1973. № 5, № 6, № 8.
26. *Горбань А. Н.* Обход равновесия. Новосибирск: Наука, 1984.
27. *Опойцев В. И.* Нелинейная системостатистика. М.: Наука, 1986.
28. *Малишевский А. В.* Качественные модели в теории сложных систем. М.: Наука, 1998.
29. *Сергеев В. М.* Пределы рациональности. М.: Фазис, 1999.
30. *Попков Ю. С.* Теория макросистем: равновесные модели. М.: УРСС, 1999.
31. *Цирлин А. М.* Методы оптимизации в необратимой термодинамике и микроэкономике. М.: Физматлит, 2003.
32. *Швецов В. И., Алиев А. С.* Математическое моделирование загрузки транспортных сетей. М.: УРСС, 2003.
33. *Маслов В. П.* Кvantовая экономика. М.: Наука, 2006.
34. *Олемской А. И.* Синергетика сложных систем: Феноменология и статистическая теория. М.: КРАСАНД, 2009.
35. *Веретенников А. Ю.* Параметрическое и непараметрическое оценивание для цепей Маркова. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2000.
36. *Боровков А. А.* Эргодичность и устойчивость случайных процессов. М.: УРСС, 1999.

37. Булинский А. В., Ширяев А. Н. Теория случайных процессов. М.: Физматлит; Лаборатория базовых знаний, 2003.
38. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. 2. М.: МЦНМО, 2010.
39. Вишневский В. М. Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003.
40. Ивицкий В. А. Теория сетей массового обслуживания. М.: Физматлит, 2004.
41. The maximum entropy formalism, ed. by R. D. Levin, M. Tribus. Conf. Mass. Inst. Tech., Cambridge 1978. MIT Press, 1979.
42. International workshops on Bayesian inference and maximum entropy methods in science and engineering. AIP Conf. Proceedings (holds every year from 1980).
43. Kapur J. N. Maximum – entropy models in science and engineering. John Wiley & Sons, Inc., 1989.
44. Golan A., Judge G., Miller D. Maximum entropy econometrics: Robust estimation with limited data. Chichester, Wiley, 1996.
45. Fang S.-C., Rajasekera J. R., Tsao H.-S. J. Entropy optimization and mathematical programming. Kluwer's International Series, 1997.
46. Маслов В. П., Черный А. С. О минимизации и максимизации энтропии в различных дисциплинах // ТВП. 2003. Т. 48. № 3. С. 466-486.
47. Богданов К. Ю. Прогулки с физикой. Библиотечка «Квант» В. 98. М.: Бюро Квантум, 2006 (глава 18).
48. Зорич В. А. Математический анализ задач естествознания. М.: МЦНМО, 2008.
49. Diaconis P. The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. 2009. V. 49. № 2. P. 179-205.
<http://www.ams.org/journals/bull/2009-46-02/S0273-0979-08-01238-X/S0273-0979-08-01238-X.pdf>
50. Joulin A., Ollivier Y. Curvature, concentration and error estimates for Markov chain Monte Carlo // Ann. Prob. 2010. V. 38. № 6. P. 2418-2442.
51. Красносельский М. А., Лишиц Е. А., Соболев А. В. Позитивные линейные системы. Метод положительных операторов. М.: Hayka, 1985.
52. Кингман Дж. Пуассоновские процессы / под ред. А. М. Вершика. М.: МЦНМО, 2007.

53. *Магарил-Ильяев Г. Г., Тихомиров В. М.* Выпуклый анализ и его приложения. М.: УРСС, 2003.
54. *Гасникова Е. В.* Двойственные мультиплекативные алгоритмы для задач энтропийно–линейного программирования // ЖВМ и МФ. 2009. Т. 49. № 3. С. 453–464.
55. *Нестеров Ю. Е.* Введение в выпуклую оптимизацию. М.: МЦНМО, 2010.
56. *Нурминский Е. А., Шамрай Н. Б.* Прогнозное моделирование автомобильного трафика Владивостока // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)
57. *Flajolet P., Sedgewick R.* Analytic combinatorics. Cambridge University Press, 2008. <http://algo.inria.fr/flajolet/Publications/book.pdf>
58. *Вершик А. М., Шмидт А. А.* Предельные меры, возникающие в асимптотической теории симметрических групп // ТВП. 1977. Т. 22. № 1. С. 72–88; 1978. Т. 23. № 1. С. 42–54.
59. *Синай Я. Г.* Вероятностный подход к анализу статистики выпуклых ломаных // Функц. анализ и его прил. 1994. Т. 28. № 2. С. 41–48.
60. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2004.
61. *Ledoux M.* Concentration of measure phenomenon. Providence, RI, Amer. Math. Soc., 2001 (Math. Surveys Monogr. V. 89).
62. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.
63. *Якымив А. Л.* Вероятностные приложения тауберовых теорем. М.: Наука, 2005.
64. *Ширяев А. Н.* Вероятность 1, 2. М.: МЦНМО, 2007.
65. *Malyshev V.A., Pirogov S.A., Rubco A.N.* Random walks and chemical networks // Mosc. Math. J. 2004. V. 4. № 2. P. 441–453.
66. *Батищева Я. Г., Веденягин В. В.* II-й закон термодинамики для химической кинетики // Мат. мод. 2005. Т. 17. № 8. С. 106–110.
67. *Веденягин В. В., Орлов Ю. Н.* О законах сохранения для полиномиальных гамильтонианов и для дискретных моделей уравнения Больцмана // ТМФ. 1999. Т. 121. № 2. С. 307–315.
68. *Рыбко А. Н.* Пуассоновская гипотеза для больших симметричных коммуникационных сетей // Глобус. Общематематический семинар / под ред. М. А. Цфасмана и В. В. Прасолова. № 4. М.: МЦНМО, 2009. С. 105–126.
69. *Федорюк М. В.* Метод перевала. М.: УРСС, 2010.

Введение в стохастические модели транспортных потоков

1. Введение

Математические модели автомобильного трафика могут быть весьма различны: от дифференциальных уравнений с частными производными, средств современной компьютерной физики до создания игровых моделей, где точки на видео движутся по сети улиц с перекрестками. Мы рассматриваем здесь некоторые строгие вероятностные подходы к транспортным сетям. Основная цель этого приложения — не столько представить технику решения задач, сколько представить методику (и искусство) составления адекватных моделей, которые отличаются наглядностью определений (основной объект там именно автомобиль, а не потоки) и основаны на простых интуитивных рассуждениях. Более того, все вводимые постулаты в этих моделях допускают статистическую проверку, широкие уточнения и обобщения, и не используют сомнительных физических аналогий. Вообще, вероятностные модели должны связываться с психикой водителей, если водители не роботы. Такой законченной и общепринятой теории пока не существует, здесь делаются, по-видимому, первые строгие попытки установить такую связь.

В текст включены упражнения для лучшего усвоения материала, а также задачи посложнее, в том числе и «публикальные».

Вероятностный подход к транспортным потокам существует уже более 50 лет, см. [1–3], однако здесь мы даем более современную трактовку и рассматриваем более сложные задачи. В то же время мы не говорим здесь о других вероятностных подходах к проблемам транспорта, например [12, 14], они отражены в других частях этой книги. Мы также ничего не говорим о гидродинамическом подходе, так как связь его со статистическим подходом пока математически плохо изучена.

Приложение состоит из трех частей. В первой дано построение случайных потоков и некоторые модели, отражающие качественные

явления на автомагистрали. В том числе новая модель, основанная на сравнительно недавней теории случайных грамматик. Во второй части показано, как можно получать явные формулы с помощью техники пуассоновских потоков. В третьей рассмотрены сложные сети дорог и вычисление критической нагрузки, выше которой начинаются пробки.

2. Потоки автомобилей

2.1. Маркованные точечные поля

Под словом поток в зависимости от контекста понимают либо среднее число J автомобилей в единицу времени, пересекающих сечение транспортного пути в данном направлении, либо статическую случайную конфигурацию

$$\dots < x_i < x_{i-1} < \dots$$

автомобилей в данный момент времени, но можно понимать его динамически как меру на множестве траекторий $\{x_i(t)\}$ автомобилей.

Что такое конфигурация автомобилей. Максимально детальное описание расположения автомобилей в данный момент времени таково. Автомобиль индивидуален и ему присваивается некий индекс α . Например, пусть есть автотрасса с k полосами $1, 2, \dots, k$, представляемая k прямыми, параллельными оси x . Тогда индекс $\alpha = (m, i)$ выделяет i -й автомобиль на полосе m . Индекс i нумерует автомобиль на полосе, так что автомобиль i следует за автомобилем $i - 1$. Пусть d_α — длина этого автомобиля, $x_\alpha(t)$ — его координата (например, переднего бампера). Автомобили движутся в положительном направлении оси x . Далее индекс полосы опускается — читатель может его добавлять где надо — и используем только индекс i .

Обозначим расстояние автомобиля i до предыдущего автомобиля в реальном потоке в момент t через

$$d_i^+(t) = x_{i-1}(t) - x_i(t) - d_{i-1}.$$

Обозначим (тоже важная величина для водителя)

$$d_i^-(t) = d_{i+1}^+(t)$$

— расстояние до следующего автомобиля.

Как вводятся вероятности на множестве конфигураций.

Формально точечный случайный поток на прямой задается вероятностной мерой на множестве всех счетных локально конечных (то есть конечных на каждом ограниченном интервале) подмножеств прямой. Иначе говоря, задается согласованной системой вероятностей

$$P(I_1, k_1; \dots; I_n, k_n)$$

того, что в интервалах $I_j, j = 1, \dots, n$ находится ровно k_j частиц.

Для более конкретного задания этих распределений существуют две большие науки: *теория восстановления* (см., например, [6]) и *теория гиббсовских точечных полей* [23, 24]. Первая теория существенно проще, но годится только в одномерном случае. Вторая глубоко связана с физикой, годится и для многомерных ситуаций, но довольно сложна, и мы не будем ее здесь касаться.

Самый простой случайный поток пуассоновский, см., например, [30]. Простейший способ его понять такой. Рассмотрим интервал $[-N, N]$ и бросим на него независимо и случайно (точнее равномерно) $M = [\rho N]$ точек, где $\rho > 0$ — некоторая константа, называемая плотностью. Легко вычислить биномиальную вероятность $P_{N,M}(k, I)$ того, что в конечный интервал I попадет ровно k точек. Последняя при $N \rightarrow \infty$ стремится к пуассоновскому выражению

$$P(k, I) = \frac{\{\rho|I|\}^k}{k!} e^{-\rho|I|}.$$

Более общие потоки легко строятся на полуправой $[0, \infty)$. Именно, случайные точки

$$x_0 = 0, x_1, \dots, x_n, \dots$$

определяются как суммы независимых одинаково распределенных случайных величин $\xi_i > 0, i = 1, 2, \dots$, с распределением $G(x)$:

$$x_1 = \xi_1, x_2 = \xi_1 + \xi_2, \dots$$

Для определения трансляционно-инвариантного потока на всей прямой остается одна проблема — где разместить начальную точку потока, от которой откладывать независимые величины налево и направо. Для этого надо воспользоваться следующим (одним из основных) утверждением теории восстановления. Пусть $P(t, t + \Delta t)$ — вероятность того, что в интервал $(t, t + \Delta t)$ попадет ровно одна

точка x_n . Тогда если ξ_i имеют плотность, то на полуправой прямой предел $\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t, t + \Delta t)$ существует и стремится при $t \rightarrow \infty$ к $\rho = (E\xi)^{-1}$. Предельная (при $t \rightarrow \infty$) плотность вероятности того, что расстояние от точки t до первой случайной точки $x_i > t$ больше s , равна произведению ρ на вероятность того, что $\xi_i = x_i - x_{i+1} > s$, то есть равна

$$\rho(1 - G(s)).$$

Поэтому первую после начала координат точку потока следует взять на случайном расстоянии с этой плотностью. Расстояния же между точками будут по-прежнему независимыми от функции распределения $G(x)$.

Альтернирующие потоки. Расстояния между соседними точками потока не обязательно одинаково распределены. Распределения могут чередоваться. Например, возьмем две последовательности случайных величин: ξ_1, ξ_2, \dots и η_1, η_2, \dots и положим

$$x_{2n} = \xi_1 + \dots + \xi_n + \eta_1 + \dots + \eta_n, x_{2n-1} = \xi_1 + \dots + \xi_n + \eta_1 + \dots + \eta_{n-1}.$$

Тогда построение потока на всей прямой делается как и выше. В нашем случае чередуются длины автомобилей d_i (независимые и одинаково распределенные) и функции d_i^+ (независимые и одинаково распределенные).

Маркированные потоки. Каждой точке x_i точечного процесса может быть сопоставлена величина σ_i , принимающая значения в некотором множестве S . Эту величину в разных случаях называют маркой или спином в точке i и говорят о случайном маркированном точечном множестве (потоке, процессе). Он определяется мерой на последовательностях пар (x_i, σ_i) . Проще всего, когда задана мера на счетных множествах, то есть задан поток без марок, а величины σ_i объявляются независимыми и одинаково распределенными. В разделе 1.5 мы построим маркированный процесс, где марками являются скорости, причем их распределение будет сложным образом коррелировать во времени с траекториями точек.

О марковских процессах. Эволюция во времени конфигураций автомобилей часто использует марковские процессы, и необходимо сказать о соответствующей терминологии. Часто определение самого

процесса и его свойств (например, эргодичности) отличаются в разных источниках. Поясним это. Для простоты ограничимся случаем дискретного времени.

Рассмотрим на некотором фазовом пространстве X систему мер (переходных вероятностей) $P(A|x)$, определяющих вероятности того, что в момент $t+1$ процесс попадет в множество $A \subset X$, если в момент t процесс находился в состоянии $x \in X$. Если все меры $P(A|x)$ одноточечные, то это эквивалентно заданию детерминированного отображения $T : X \rightarrow X$, точнее $P(\cdot|x) = \delta(T(x))$ — единичная мера в точке $T(x)$. Тогда говорят о детерминированном отображении, задающем динамическую систему.

Заметим, что система мер $P(A|x)$ определяет очевидным образом преобразование U множества вероятностных мер на X в себя:

$$U\mu = \int P(\cdot|x)d\mu(x).$$

Важным является понятие инвариантной (относительно U) меры. Обычно исследуется ее существование, единственность и другие свойства.

По системе переходных вероятностей можно построить разные последовательности случайных величин ξ_n со значениями в X или их распределений μ_n на X , где $P(\xi_n \in A) = \mu_n(A)$. Вероятностным пространством при этом служит множество траекторий $\{x_n\}$. Например, по вероятностной инвариантной мере строится стационарный марковский процесс как последовательность $\xi_n, n \in Z_+$ случайных величин со значениями в X . Или по заданной начальной мере μ_0 на X строится последовательность $\xi_n, n \in Z_+$.

Под марковским процессом может пониматься как одна из таких последовательностей случайных величин, так и все семейство таких последовательностей ξ_n . Соответственно разнится терминология, например, определение эргодичности. Динамическая система с заданной инвариантной мерой μ называется эргодической, если любое инвариантное множество имеет меру μ ноль или единицу. На любой стационарный марковский процесс можно смотреть как на динамическую систему — сдвиг в пространстве траекторий. Тогда понятие эргодичности совпадает с понятием эргодичности этой динамической системы.

Однако чаще, когда говорят о марковском процессе, имеют в виду не только стационарные процессы. Наиболее часто используемым

определенением эргодичности является следующее. Процесс называется эргодическим, если существует единственная инвариантная мера μ на X , и для любой начальной меры μ_0 при $n \rightarrow \infty$ имеет место слабая сходимость $\mu_n \rightarrow \mu$.

Отметим, что марковские процессы резко разделяются на два класса. Первый класс — для которых существует положительная мера на X (не обязательно вероятностная), относительно которой все меры $P(\cdot|x)$ абсолютно непрерывны. К ним относятся почти все классические марковские процессы — конечные и счетные цепи Маркова, диффузионные процессы и другие. Такие процессы называются эргодическими, если, во-первых, у преобразования нет нетривиальных инвариантных множеств, а во-вторых, существует единственная инвариантная вероятностная мера. Во многих случаях отсюда следует сходимость к этой инвариантной мере из любого начального распределения. Для счетных цепей эквивалентным условием является положительная возвратность, то есть конечность (для всех пар $x, y \in X$) среднего времени достижения y из x .

Второй класс характеризуется тем, что все меры $P(\cdot|x)$ взаимно сингулярны. К этому классу относятся почти все процессы с бесконечным числом частиц. Теория таких процессов существенно сложнее.

2.2. Связь скорости и плотности с пропускной способностью

Психика водителя в простейшем потоке. Полностью моделировать психику конечно невозможно, но многие закономерности очевидны. Так, водитель i видит несколько автомобилей (часто только один впереди себя) в потоке и выбирает оптимальное для себя расстояние до предыдущего автомобиля. Если скорость $v_{i-1}(t) = \frac{dx_{i-1}(t)}{dt}$ меняется медленно, то можно считать, что реакция водителя быстрее, и выбиаемое расстояние D_i^+ зависит только от этой скорости:

$$D_i^+ = D_i^+(v_{i-1})$$

(индекс i говорит, что функции $D_i^+(v)$ разные для разных водителей). Назовем поток алгоритмическим в момент t , если для всех i

$$d_i^+(t) = D_i^+(v_{i-1}(t)),$$

то есть все скорости последовательно определяются по скорости первого автомобиля. Конечно, нас интересуют не сами функции, а их статистические характеристики. При вероятностном подходе функции $D_i^+(v)$ становятся независимыми одинаково распределенными случайными функциями, зависящими от скорости v предыдущего водителя как от параметра. Распределение этих функций не может быть выведено из математических, статистических, физических и т. д. законов. Оно зависит от индивидуальной и коллективной психики водителя и должно находиться экспериментально, см. [22].

Детерминированная динамика без обгона. Если все автомобили, водители и скорости v одинаковы, то многие задачи решаются просто. Обозначим через d — длину автомобиля, $d^+ = D^+(v)$ — расстояние до впереди идущего автомобиля, которое водитель соблюдает. Уже такая динамика позволяет понять многие качественные эффекты.

Определим пропускную способность дороги как максимально возможный поток по ней:

$$J_{\max} = \max_v v \lambda(v),$$

где максимум берется по разрешенному интервалу скоростей, а

$$\lambda(v) = \frac{k}{d + D^+(v)}$$

— плотность автомобилей на k -полосной дороге при заданной скорости v . Отсюда видно, что пропускная способность может уменьшаться при увеличении скорости. Этот простой вывод говорит лишь о том, что многие водители увеличивают расстояние до впереди идущего автомобиля при увеличении его скорости.

Случайная динамика без обгона. То же самое получится, если скорости v одинаковы, функции d_i^+ случайны и независимы, а их средние равны (для заданного v) некоторому числу $d^+(v)$. Мы видим, что сам факт нетривиальной зависимости пропускной способности от скорости тривиален, и для него совершенно не нужны вероятностные модели. Однако для более тонких вопросов вероятностные модели необходимы. Сейчас мы введем довольно общую вероятностную модель с очень богатым спектром фаз. При этом процессы с

запретами (exclusion processes) появляются как вырожденный частный случай. Другие модели см. [10, 12, 22].

Случайная динамика с обгоном (случайные грамматики). Здесь естественно возникает связь с таким недавно открытым объектом, как случайные грамматики, см. [25]. Мы дадим краткое содержательное описание одной такой модели.

Пусть в момент $t = 0$ все автомобили находятся на левой полосе, движение однополосное. Мы разбиваем полосу движения на клетки определенной длины и считаем, что в каждой клетке не более одного автомобиля. Таким образом, конечная последовательность автомобилей изображается парой (S, r) , где $r \in Z$, а S — конечная последовательность (слово) из трех символов 0, 1, 2:

$$S = s_N \dots s_2 s_1.$$

При этом 0 соответствует пустой клетке, 1 — активному (быстрому) водителю в клетке, 2 — спокойному водителю в клетке. Длина слова $N = N(t)$ и все символы $s_k(t)$ могут меняться во времени, но так, что всегда $s_1(t) \neq 0$ для всех $t \geq 0$. В произвольный момент t каждый символ $s_k(t)$ имеет координату $x(s_k(t))$. Координаты однозначно определяются

$$x(s_k(t)) = x(s_1(t)) - k + 1 \quad (1)$$

координатой $x(s_1(t))$ первого символа, которую мы обозначим $r = r(t)$.

Динамика моделирует процесс ускорений и торможений отдельных водителей и определяется как цепь Маркова $(S(t), r(t))$ с непрерывным временем на множестве пар $\{(S, r)\}$. Интенсивности скачков определяются так. Изменения S и r независимы друг от друга. Изменение r моделирует движение всего потока с постоянной скоростью v . Именно, r увеличивается на единицу с вероятностью $v dt$ за время dt , и все координаты немедленно изменяются соответственно формуле (1). Динамика S , таким образом, будет описывать ситуацию относительно некоторого равномерного движения. Эта динамика задается случайной грамматикой, то есть списком возможных локальных замен подслов (всего будет 5 типов замен) S на другое под слово. Любые замены из приводимого ниже списка производятся независимо, случайно и имеют разные интенсивности (всего 4 параметра). Вот этот список.

- 1) $10 \rightarrow 01$ — быстрый водитель передвигается на 1-го вперед, освобождая свободное место за собой, с вероятностью $\lambda_0^+ dt$ за время dt ;
- 2) $120 \rightarrow 021$ — быстрый водитель обгоняет спокойного, с вероятностью $\lambda_1^+ dt$;
- 3) $22 \rightarrow 202, 21 \rightarrow 201$ — предусмотрительный водитель тормозит, увеличивая дистанцию перед собой, с вероятностью $\lambda_2^- dt$. Отметим, что здесь увеличивается длина S (возникает лишняя свободная ячейка), что ведет к сдвигу всех автомобилей сзади этого водителя на 1-го назад. Это нелокальный скачок, реально он растянут во времени, но это совместимо с правилом сложения относительных скоростей;
- 4) $200 \rightarrow 020$ — спокойный водитель ускоряется с вероятностью $\lambda_2^+ dt$ (если впереди с его точки зрения много свободного места).

Необходимо сказать, что для точной формулировки результатов, которые мы лишь обрисуем, надо делать разнообразные скейлинги параметров t, N, λ . Это будет сделано в отдельной статье. В зависимости от 4 параметров могут быть разнообразные типы (фазы) движения. Мы приведем только три из них.

Если λ_2^\pm малы по сравнению с остальными двумя параметрами, то автомобили типа 2 едут синхронно и с постоянной скоростью, а быстрые автомобили имеют дополнительную относительную скорость. Если быстрых автомобилей мало, то эта дополнительная скорость определяется движением одного автомобиля среди неподвижных препятствий и зависит от плотности ρ_2 автомобилей типа 2 и плотности дырок ρ_0 и примерно равна

$$v_{rel} = \lambda_0^+ \rho_0 + 2\lambda_1^+ \rho_2.$$

Если λ_2^- мала по сравнению с остальными двумя параметрами (нет нелокальных эффектов), а λ_2^+ имеет такой же порядок как λ_0^+, λ_1^+ , то разница между типами стирается. Мы имеем тогда процесс, близкий к так называемому полностью асимметричному процессу с запретами (TASEP — totally asymmetric exclusion process), а для значений

$$\lambda_0^+ = \lambda_1^+, \lambda_2^- = 0$$

полностью с ним совпадающий (о TASEP см. приложение М. Бланка и ссылки в нем).

Если λ_2^+ мала, а λ_2^- велика по сравнению с остальными двумя параметрами, то картина иная. Каждый обгон $120 \rightarrow 021$ вызывает немедленное торможение автомобиля 2 и, как следствие, все последующие автомобили замедляются. Для автомобилей ближе к концу слова замедление будет весьма существенным, если поток достаточнно плотный (мало ячеек с нулями), так как много автомобилей типа 2 будет тормозиться.

Можно усложнять введенную динамику, например, избежать дискретизации (см. конец этого раздела), вводя вместо нулей положительные вещественные числа — расстояния между последовательными автомобилями. Это потребует существенных (см. однако раздел 2.5) переформулировок, особенно для скачков типа 3, но сохранит грубые качественные эффекты.

2.3. Рост пробки

Если входной транспортный поток в некоторую фиксированную область равен J_{in} , а выходной $J_{out} < J_{in}$, то количество автомобилей в данной области за время t увеличится на

$$t(J_{in} - J_{out}).$$

Так будет, однако, только если рассматриваемая область не находится на самом транспортном пути. Например, если автомобили скапливаются в пробке на самой дороге, то ответ другой. Дело в том что область сама может расти за счет скапливающихся автомобилей. Чтобы уточнить эти утверждения надо уточнить модель.

Пусть автомобили одинаковой длины d едут в потоке (по одной полосе) со скоростью v один за другим на одинаковом расстоянии d^+ между ними. Пусть в течение времени t движение остановлено неким препятствием, например, красным светофором. При этом автомобили останавливаются на расстоянии $d_0^+ < d^+$ до предыдущего автомобиля.

Упражнение 1. Доказать, что за время $t \rightarrow \infty$ пробка (то есть максимальная длина $L(t)$ участка, где все автомобили стоят) не-

ред препятствием будет иметь длину, асимптотически равную

$$L(t) \sim_{t \rightarrow \infty} tv \frac{d + d_0^+}{d^+ - d_0^+}. \quad (2)$$

По-видимому, этот результат зависит в действительности лишь от средних величин и остается верным при возможности обгона. Это сделано в [16] для независимого движения автомобилей (то есть когда автомобили не мешают друг другу), причем скорости автомобилей имеют флуктуации, однако средние скорости всех автомобилей одинаковы и равны v . Однако доказательство там совсем не просто. Другие модели роста пробки см. в [17] и главе 2.

Локальные расширения и сужения трассы. Что происходит при переходе участка дороги с k полосами в участок с l полосами. Пусть этот переход происходит в точке с координатой $x = 0$.

Случай $k < l$. Пусть максимально разрешенная скорость равна v_{\max} и предполагается дисциплинированность водителей. Пусть автомобили движутся по k -полосной трассе со скоростью $v < v_{\max}$ и быстрее ехать невозможно по причине фундаментального соотношения между плотностью автомобилей ρ и их скоростью:

$$d + D^+(v) = \rho^{-1}.$$

Тогда по l полосной трассе длины L автомобили теоретически могут сохранить ρ и двигаться с такой же скоростью, но ρ может скорректироваться так, что автомобили смогут двигаться быстрее с некоторой большей скоростью v_1 . Выгода во времени будет

$$\frac{L}{v} - \frac{L}{v_1}.$$

Случай $k > l$. Тогда возможны три разных ситуации.

Свободный поток. Если поток очень редкий, то автомобили будут подъезжать к точке 0 в одиночку и не заметят перехода.

Растущая пробка. Обозначим J_k — текущий входящий поток и $J_{l,\max}$ — максимально возможный поток по l -полосной трассе. Если $J_k > J_{l,\max}$, то будет образовываться пробка, и число автомобилей в пробке в среднем будет расти как $t(J_k - J_{l,\max})$, а точнее как в формуле (2).

Задержка. В случае $J_k < J_{l,\max}$ практическое наблюдение таково: перед сужениями могут возникать пробки случайной длины, которые однако не растут слишком сильно. Соответствующих стохастических моделей пока нет, для этого прежде всего нужны нестационарные модели начала и остановки движения. Некоторые из этих моделей мы сейчас опишем.

2.4. Модели начала движения

В работе [9] автомобили задаются точками

$$\dots < x_i(t) < x_{i-1}(t) < \dots$$

на прямой. В начальный момент времени $t = 0$ автомобили стоят и образуют пуассоновское точечное поле с плотностью $\rho < 1$. Автомобили могут иметь две скорости: 0 или 1; обгоны запрещены. Каждый стоящий автомобиль, через независимое экспоненциально распределенное время со средним 1, начинает двигаться со скоростью 1. Может случиться так, что автомобиль с номером i доедет до автомобиля $i - 1$ пока тот еще не начал двигаться. Тогда он останавливается и начинает двигаться через экспоненциальное время после того как начнет двигаться автомобиль $i - 1$. Такое правило действует всегда. Этот процесс в некотором смысле описывает выезд автомобилей из пробки.

Основной результат состоит в том, что с вероятностью 1 каждый автомобиль будет останавливаться только конечное число раз (при условии $\rho < 1$). Пусть t_i — момент времени, начиная с которого автомобиль i все время движется. Тогда для любых i и k и любых $t_i, t_{i-1}, \dots, t_{i-k}$ случайные величины

$$x_{i-1}(t_{i-1}) - x_i(t_i), x_{i-2}(t_{i-2}) - x_{i-1}(t_{i-1}), \dots, x_{i-k}(t_{i-k}) - x_{i-k+1}(t_{i-k+1})$$

будут независимы и экспоненциально распределены. Иначе говоря, после выезда из пробки автомобили будут образовывать пуассоновскую конфигурацию той же самой интенсивностью ρ , что и в начале.

Пусть теперь в момент 0 пуассоновский точечный поток с плотностью ρ находится на левой полуоси. Каждая точка движется со скоростью $v > 0$, если расстояние до предыдущей точки не меньше некоторого $d_{eff} > 0$, и стоит в противном случае. Здесь очевидно, что каждая частица не останавливается, начиная с некоторого момента. Но здесь можно получить больше. Рассмотрим следующие

случайные величины: $\tau_k^{(1)}$ — случайное время начала движения k -й точки, $\tau_k^{(2)}$ — случайное время, начиная с которого эта точка больше не останавливается, x_k — расстояние до первой точки, начиная с момента $\tau_k^{(2)}$.

Задача 1. Найти асимптотику распределений этих случайных величин при $k \rightarrow \infty$.

Связь с задачей задержки очевидна. Пусть есть две полосы и на каждой полосе интенсивность потока ρ ; объединенный поток, таким образом, имеет плотность 2ρ . Автомобилям из первой полосы надо втиснуться во вторую. Алгоритмы втискивания могут быть разными. Например, любой автомобиль втискивается независимо от других, если его расстояние (по оси x) до предыдущего и последующего автомобиля из второй полосы было не менее некоторого числа d^+ .

2.5. Ближний и дальний порядок при меняющихся во времени скоростях автомобилей

Здесь автомобили представляются точками x_i . С автомобилем i связывается случайный процесс $w_i(t)$, определяющий его скорость в момент t «на свободной дороге» (то есть при отсутствии препятствия спереди). Величина этой скорости косвенно определяет активность водителя в данный момент времени. Будем говорить, что автомобиль i имеет впереди себя препятствие в момент t , если

$$x_i(t-0) = x_{i-1}(t).$$

Процессы $w_i(t)$ взаимно независимы и определяются лишь психикой индивидуального водителя. Предположим, что существуют константы $0 < C_1 < C_2 < \infty$ такие, что для всех t, i

$$C_1 < w_i(t) < C_2.$$

Поток задается начальным положением $x_i(0)$ автомобилей, а их движение определяется как

$$x_i(t) = x_i(0) + \int_0^t v_i(s) ds,$$

где $v_i(t)$ — определяемая ниже скорость автомобиля в потоке. При этом начальные положения таковы, что расстояния $d_i^+(0)$ независимы и, например, экспоненциальны с заданным параметром $\rho(0)$. Процесс будет полностью определен, если для всех $t_1, \dots, t_n, i_1, \dots, i_n$ мы зададим конечномерные распределения векторов:

$$(v_{i_1}(t_1), \dots, v_{i_n}(t_n)),$$

где среди индексов i_k могут быть одинаковые. Эти распределения полностью определяются следующими правилами:

- 1) (правило свободной дороги) если ни один из автомобилей i_1, \dots, i_k при $k \leq n$ не имеет впереди себя препятствия, то распределение вектора $v_{i_1}(t_1), \dots, v_{i_k}(t_k)$ совпадает с распределением вектора $w_{i_1}(t_1), \dots, w_{i_k}(t_k)$ и является независимым от распределения вектора $v_{i_{k+1}}(t_{k+1}), \dots, v_{i_n}(t_n)$;
- 2) (правило препятствия) если автомобиль i имеет впереди себя препятствие в момент t , то $v_i(t) = v_{i-1}(t)$;
- 3) (правила обгона) если автомобиль i имеет впереди себя препятствие в момент t , то он меняется местами с предыдущим автомобилем с некоторой интенсивностью λ в течение (случайного) интервала времени пока $w_i(t) > v_{i-1}(t)$. Смысл этого условия состоит в том, что водитель обгоняет, если его активность высока в течение некоторого промежутка времени.

Уже для этого простейшего определения транспортного потока с зависимыми от времени скоростями есть много задач. Некоторые из них мы сейчас сформулируем.

Назовем свободной фазой случай, когда автомобиль не задерживается при обгоне догоняемого автомобиля, то есть интенсивность обгона равна бесконечности. Тогда для любых автомобилей с индексами i, j их скорости независимы, и, значит, ковариации

$$\text{cov}_{ij}(t) = E v_i(t) v_j(t) - E v_i(t) E v_j(t) = 0.$$

Задача 2. Для заданных распределений процессов $w_i(t)$ существует константа $0 < \lambda_0 < \infty$ такая, что при $\lambda < \lambda_0$ существует предельный стационарный процесс (по i и по t), в котором ковариации $\text{cov}_{ij}(t)$ убывают экспоненциально по $|j - i|$.

Назовем этот тип движения фазой с ближним порядком. Существование фазы дальнего порядка определяется следующей гипотезой.

Задача 3. Существует константа $0 < \lambda_{cr} < \infty$ такая, что при $\lambda > \lambda_{cr}$ существует предельный стационарный процесс, в котором ковариации $\text{cov}_{ij}(t)$ не стремятся к нулю при $|j - i| \rightarrow \infty$.

Будет ли $\lambda_{cr} = \lambda_0$ предыдущей задачи?

Эти три фазы могут иметь отношение к фазам, определенным Б. С. Кернером [15].

Задача 4. Определить подобный процесс с длинами d_i, d_i^+ , а также с дополнительными индексами, соответствующими полосам движения, и с поведением водителя, зависящим не только от следующей, но и от предыдущего автомобиля. Какие дополнительные качественные эффекты могут ловить эти модели?

3. Расчет средней скорости на автотрасse

Мы приводим здесь простейшую постановку задачи о снижении средней скорости движения автомобиля по автотрасse из-за случайных неподвижных (аварии и ремонтные работы) и движущихся (медленные автомобили) препятствий. Цель — показать (полностью решив модельные задачи), что во многих случаях можно получить простые красивые формулы, позволяющие понять основные причины замедления. Мы четко формулируем технические предположения для получения таких формул. Основное предположение касается однородности трассы, именно въезда, выезда автомобилей, специфики обгона.

3.1. Дорога как одномерная сеть массового обслуживания

Следующая модель заимствована из [8], с. 117. Пусть есть бесконечная дорога и два типа автомобилей, задаваемые точками на бесконечной прямой, которые движутся в одном направлении. Автомобили первого типа (быстрые) двигаются с постоянной скоростью v_1 , автомобили второго типа (медленные) имеют постоянную скорость v_2 , где $v_1 > v_2$.

Предположим, что быстрые автомобили в начальный момент времени образуют пуассоновскую случайную конфигурацию (пуассоновский точечный поток) на всей прямой с плотностью λ_1 . Медленные автомобили расположены в момент $t = 0$ в точках

$$x_0 = 0 < x_1 < \dots < x_n < \dots,$$

причем расстояния $x_k - x_{k-1}$ одинаково распределены со средним λ_2^{-1} (не обязательно экспоненциально). Медленные автомобили едут независимо, не замечая других автомобилей. Быстрые же «взаимодействуют» с каждым автомобилем, с которым их координаты совпадают. Именно, быстрым автомобилям разрешено обгонять медленные. Когда быстрый автомобиль догоняет медленный, то есть их координаты совпадают, то он сколько-то времени едет вместе с медленным, то есть со скоростью v_2 . Через экспоненциально распределенное время с параметром μ он обгоняет медленного, то есть начинает ехать со скоростью v_1 . Если быстрый автомобиль догоняет группу быстрых автомобилей, следующих за медленным, то обгон происходит в порядке очереди, точнее, в том порядке, в котором быстрые автомобили догоняли данный медленный автомобиль. Без ограничения общности, скорости медленных автомобилей можно считать равными нулю $v_2 = 0$, а скорости быстрых соответственно равными $v = v_1 - v_2$. Поэтому каждый медленный автомобиль можно представлять узлом обслуживания, на который приходят клиенты (быстрые автомобили) и в порядке очереди (то есть прибытия) ждут обслуживания (обгона), и обслуживаются с интенсивностью обслуживания μ .

Теперь эта задача может быть сведена к линейной сети массового обслуживания, которую мы сейчас опишем. Имеется бесконечная последовательность

$$S_0 \rightarrow \dots \rightarrow S_k \rightarrow S_{k+1} \rightarrow \dots$$

узлов обслуживания двух типов. Каждый узел S_k представляет собой систему типа $M/M/1$ с дисциплиной обслуживания FIFO (first-in-first-out), то есть обслуживание в порядке естественной очереди. Эти узлы соответствуют медленным автомобилям, а требования — быстрым. Например, узел S_0 соответствует крайнему левому медленному автомобилю. Вторая буква M означает экспоненциальность времени обслуживания. Это вместе с дисциплиной FIFO отвечает

формулировке нашей модели. Первая буква M означает пуассоновость входящего потока прибывающих требований. Так, на узел S_0 поступление требований образует стационарный пуассоновский поток с интенсивностью $\lambda_1 v$. Из элементарной теории очередей известно, во первых, что если $\lambda_1 v < \mu$, то устанавливается стационарный режим с вероятностями P_n того, что длина очереди равна n

$$P_n = (1 - r)r^n, r = \frac{\lambda_1 v}{\mu}.$$

Во вторых, известно (теорема Бюрке (Burke)), что в стационарном режиме выходящий поток из системы типа $M/M/1$ будет пуассоновским с интенсивностью равной интенсивности входящего потока, то есть в нашем случае это $\lambda_1 v$.

После первого узла, со случайным, но одинаковым для всех автомобилей временным сдвигом $\frac{x_1 - x_0}{v}$, поток требований поступает на узел S_1 , где также устанавливается стационарный режим.

Найдем среднюю скорость быстрого автомобиля на интервале (x_0, x_N) , $N \rightarrow \infty$. При этом мы будем предполагать, что стационарный режим уже установленся. Время проезда этого участка складывается из N обгонов и N путей между медленными автомобилями.

Среднее время, затрачиваемое быстрым автомобилем на обгон медленного, составит

$$\sum_{n=0}^{\infty} (1 - r)r^n \frac{(n + 1)}{\mu} = \frac{1}{(1 - r)\mu} = \frac{1}{\mu - \lambda_1 v},$$

в то время как среднее время движения до следующего медленного автомобиля есть

$$\frac{1}{\lambda_2 v}.$$

Поэтому расстояние между соседними медленными автомобилями (в среднем равное λ_2^{-1}) быстрый автомобиль в среднем проходит за время $(\mu - \lambda_1 v)^{-1} + (\lambda_2 v)^{-1}$.

Таким образом, средняя скорость быстрого автомобиля составит

$$v_{mean} = \frac{\lambda_2^{-1}}{(\mu - \lambda_1 v)^{-1} + (\lambda_2 v)^{-1}}.$$

В следующих разделах мы рассмотрим более сложную ситуацию с более общими распределениями.

3.2. Снижение средней скорости из-за ремонтных работ

По длинной автотрассе едут автомобили с постоянной скоростью v , встречая препятствия. Препятствия обычно имеют малый размер в сравнении с расстояниями между ними, поэтому можно представлять их точками. Они возникают на произвольном участке дороге $(x, x + dx) \subset R$ за время $(t, t + dt) \subset R$ с вероятностью $\lambda dx dt$. Точнее говоря, пары (место и момент возникновения препятствия) $(x_j, t_j) \in R \times R_+$ образуют пуассоновское точечное поле Π на $R \times R_+$ с интенсивностью λ . Другое эквивалентное определение состоит в том, что для любого интервала $I \subset R$ есть пуассоновский поток прибывающих препятствий интенсивности $\lambda|I|$, причем в момент прибытия препятствие выбирает точку равномерно на интервале I .

Предположим, что j -е препятствия находится на дороге некоторое случайное время τ_j , после чего оно убирается с дороги. Будем считать, что τ_j – независимые одинаково распределенные случайные величины с функцией распределения $Q(t)$, не зависящие от пуассоновского точечного поля Π . Предположим, что первые два момента с.в. τ_j конечны. Обозначим $m_Q = E\tau_j$ и $m_Q^{(2)} = E\tau_j^2$.

Далее мы будем рассматривать два случая. В первом случае обезд запрещен и автомобиль вынужден стоять до тех пор, пока не оберут препятствие, после чего автомобиль мгновенно набирает свою скорость v . Во втором случае обезд разрешен. Более точно, автомобилю требуется некоторое случайное время для того, чтобы обехать препятствие или группу автомобилей, стоящих перед препятствием, причем время обгона не зависит от размера этой группы. Обозначим через $\eta_{m,i}$ случайное время обезда i -м автомобилем m -го препятствия. Мы предполагаем, что $\eta_{m,i}$ независимы и одинаково распределены с функцией распределения $F(u)$. Эти предположения естественны для слабой нагрузки дороги, тогда перед препятствием не будет много автомобилей. Случай большой нагрузки рассматривается ниже.

Нашей первой задачей будет вычисление средней скорости автомобиля. При сделанных предположениях автомобили не мешают друг другу, поэтому достаточно рассмотреть кого-нибудь одного из них. Обозначим через $T(x)$ случайное время, затрачиваемое автомобилем на прохождение расстояния x . Мы хотим найти предел отношения $\frac{x}{T(x)}$ при $x \rightarrow \infty$.

Пусть $b = \lambda m_Q$, ζ – с.в. с плотностью распределения

$$h(t) = m_Q^{-1}(1 - Q(t)). \quad (3)$$

Упражнение 2. Показать, что h – плотность.

Отметим, что

$$\begin{aligned} E\zeta &= \frac{1}{m_Q} \int_0^\infty t(1 - Q(t))dt = \frac{1}{m_Q} \int_0^\infty (1 - Q(t))d\left(\frac{t^2}{2}\right) = \\ &= \frac{1}{2m_Q} \int_0^\infty t^2 dQ(t) = \frac{m_Q^{(2)}}{2m_Q}, \end{aligned}$$

где $m_Q^{(2)}$ – второй момент распределения $Q(t)$.

Определим с.в. $\alpha = \min(\eta, \zeta)$, где равенство по распределению, при этом с.в. η, ζ считаются независимыми и с.в. η имеет функцию распределения $F(u)$. Положим

$$a = E\alpha.$$

Теорема 1. С вероятностью 1 при $x \rightarrow \infty$

$$\frac{x}{T(x)} \rightarrow \frac{v}{1 + abv}. \quad (4)$$

Доказательство. Без ограничения общности можно считать, что автомобиль выезжает в точке $x = 0$ в момент времени $t = 0$. Пусть $T_0(x)$ – время простоя автомобиля. Тогда, очевидно, $T(x) - T_0(x) = v^{-1}x$ и

$$\frac{x}{T(x)} = \frac{x}{T(x) - T_0(x) + T_0(x)} = \frac{1}{v^{-1} + x^{-1}T_0(x)}.$$

Поэтому достаточно найти предел отношения $\frac{T_0(x)}{x}$ при $x \rightarrow \infty$. Мы хотим показать, что

$$T_0(x) = \sum_{i=1}^{\pi(x)} \alpha_i, \quad (5)$$

где α_i – н.о.р. с.в., распределенные как α , $\pi(x)$ – с.в. с пуассоновским распределением с параметром bx , причем α_i и $\pi(x)$ независимы. Смысл этой формулы в том, что автомобиль при прохождении расстояния x встретит $\pi(x)$ препятствий и потеряет случайное время α_i на i -м препятствии.

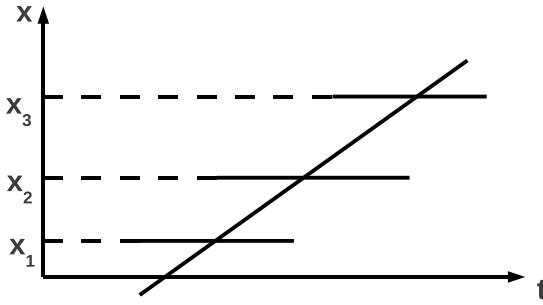


Рис. 1

Из (5) и усиленного закона больших чисел легко следует, что $\frac{T_0(x)}{x} \rightarrow ab$ п.н. при $x \rightarrow \infty$.

Докажем (5). Введем маркированное пуассоновское точечное поле Π_1 на $R \times R_+$ с конфигурацией (x_j, t_j, τ_j) , то есть τ_j — марка в точке (x_j, t_j) . Следующее утверждение можно найти в [5]:

Лемма 1. *Маркированное точечное поле Π_1 эквивалентно по распределению пуассоновскому полю на $R \times R_+^2$ с интенсивностью*

$$\lambda dx dt dQ(t).$$

Препятствия, возникающие на дороге, удобно представлять в виде горизонтальных отрезков, изображенных на рис. 1. Координаты начальной точки определяют место и время возникновения препятствия (пара (x_j, t_j)). Длина отрезка — время пребывания препятствия на дороге (марка τ_j).

Возьмем произвольную прямую $c_1 t + c_2$ и рассмотрим точки пересечения этой прямой с горизонтальными отрезками. Обозначим через $\{x_i\}$ — пространственные координаты этих точек, как показано на рис. 1. Следующая лемма доказана в [7].

Лемма 2. *Конфигурация $\{x_i\}$ образует пуассоновский процесс интенсивности $b = \lambda m_Q$.*

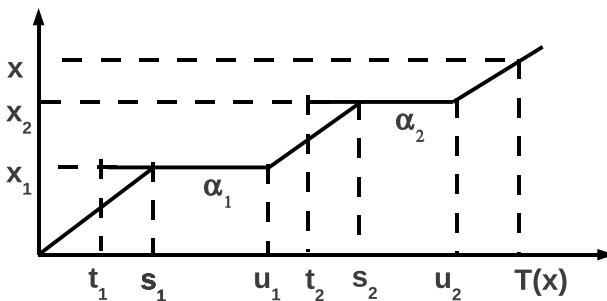


Рис. 2

На рис. 2 изображена траектория движения автомобиля, который стартует в точке $x = 0$ в момент времени $t = 0$. Обозначим через x_i пространственные координаты препятствий, которые возникают при движении автомобиля, t_i – моменты их возникновения, s_i – моменты времени, когда автомобиль встречает препятствие, u_i – моменты времени, когда автомобиль избавляется от препятствия, либо в результате объезда препятствия, либо в результате исчезновения препятствия; $\alpha_i = u_i - s_i$ – задержка автомобиля на i -м препятствии.

Из леммы 2 и пространственно-временной однородности пуассоновского точечного поля Π следует, что точки x_i образуют пуассоновский процесс интенсивности b .

Под временем жизни препятствия будем понимать время его пребывания на дороге. Назовем остаточным временем жизни препятствия – время его нахождения на дороге после того как его догнал автомобиль. Другими словами, это задержка автомобиля, если объезд невозможен.

Лемма 3. *Остаточное время жизни препятствия имеет распределение с плотностью $h(s)$, где $h(s)$ определяется формулой (3).*

В самом деле, из свойств пуассоновского точечного поля следует, что условное распределение остаточного времени жизни препятствия при условии, что полное время жизни равно t , совпадает с равномерным распределением на отрезке $[0, t]$. В силу леммы 2 ве-

роятность возникновения препятствия в интервале длины dx равна $\lambda m_Q dx + o(dx)$, а вероятность возникновения препятствия с фиксированным временем жизни t в интервале длины dx есть $\lambda t dQ(t)dx + o(dx)$, что вытекает из леммы 1. Поскольку

$$\frac{\lambda t dQ(t)dx + o(dx)}{\lambda m_Q dx + o(dx)} = \frac{t dQ(t)}{m_Q}$$

есть условная вероятность возникновения препятствия с фиксированным временем жизни t , то плотность распределения остаточного времени жизни препятствия имеет вид

$$\int_s^{\infty} \frac{t dQ(t)}{m_Q} \frac{ds}{t} = m_Q^{-1} (1 - Q(s)) ds = h(s) ds.$$

Лемма доказана.

В том случае, когда объезд возможен, автомобиль потеряет время, которое есть минимум из времени обгона и остаточного времени жизни препятствия, т.е. $\alpha_i = \min(\eta, \zeta)$. Теорема доказана.

Обсудим результат. Смысл константы a мы уже пояснили, а константа b имеет смысл стационарной плотности препятствий в пространстве.

Этот результат довольно точен при малой плотности автомобилей, так как около препятствий будет по одному автомобилю. При высокой плотности автомобилей время обгоза будет пересчитываться (увеличиваться) в зависимости от средней длины очереди перед препятствием.

3.3. Снижение средней скорости из-за медленных автомобилей

Автотрасса описывается действительной осью R . Потоки считаются не очень плотными, поэтому длина автомобиля роли не играет, и в данный момент времени положение автомобиля задается точкой $x_i(t) \in R$, где i — индекс, нумерующий автомобили. Каждый автомобиль имеет фиксированный маршрут: место и время въезда $x_{i,in}, t_{i,in}$, а также предписанное ему место выезда $x_{i,out}$. Но время выезда $t_{i,out}$ зависит от степени загруженности дороги. Мы опреде-

ляем среднюю скорость автомобиля i как

$$V_i = \frac{x_{i,out} - x_{i,in}}{t_{i,out} - t_{i,in}}.$$

Есть два типа автомобилей: быстрые и медленные, каждый движется с постоянной скоростью слева направо. У быстрых автомобилей скорость v_1 , у медленных — v_2 , где $v_1 > v_2 > 0$. Пусть $v = v_1 - v_2$. Заметим, что случай неподвижных препятствий соответствует нулевой скорости v_2 . Медленные автомобили движутся до пункта назначения нигде не останавливаясь, а быстрые до тех пор, пока не догонят впереди идущий медленный автомобиль. После этого быстрый автомобиль i движется вместе с этим медленным автомобилем j некоторое случайное время $\tau_{i,j}$ и затем обгоняет его, сразу набирая скорость v_1 . Основное предположение состоит в том, что эти случайные величины независимы и одинаково распределены с функцией распределения $F(s)$.

Эта функция распределения может быть найдена статистически двумя способами как путем прямой выборки (оценки времени ожидания обгона), так и по статистике препятствий к обгону — плотности встречного потока.

Прибытие медленных автомобилей задается тем же самым пуассоновским точечным полем Π интенсивности λ , которое было определено в предыдущем разделе. Нам потребуются также новые обозначения. С каждым медленным автомобилем мы свяжем случайное расстояние, которое ему необходимо проехать. Будем предполагать, что j -му медленному автомобилю необходимо проехать случайное расстояние ρ_j , после чего он съедет с дороги. С.в. ρ_i независимы и одинаково распределены с общей функцией распределения $G(r)$. С.в. ρ_j не зависят также от пуассоновского точечного поля Π . Будем предполагать существование первых двух моментов с.в. ρ_1 . Обозначим $m_G = E\rho_1$, $m_G^{(2)} = E\rho_1^2$.

Медленный автомобиль не встречает на своем пути препятствий и проходит свой путь со скоростью v_2 . Быстрым автомобилям могут мешать медленные. Мы рассмотрим два случая. В первом случае обгон запрещен и быстрый автомобиль вынужден следовать за медленным до тех пор пока медленный автомобиль не доедет до нужного места, после чего быстрый автомобиль мгновенно набирает свою скорость v_1 . Во втором случае обгон разрешен. Более точно, когда i -й быстрый автомобиль догоняет j -й медленный или группу быстрых

автомобилей (следующих за j -м медленным), ему требуется случайное время $\tau_{i,j}$ для того, чтобы обогнать j -й медленный автомобиль или всю эту группу автомобилей. При этом время обгона не зависит от размера группы. С.в. $\tau_{i,j}$ предполагаются независимыми и одинаково распределенными с функцией распределения $F(u)$.

Пусть $d = \lambda m_G (v_2^{-1} - v_1^{-1})$. Введем с.в. β с плотностью распределения $g(x) = m_G^{-1}(1 - G(x))$ и с.в. $\gamma = \min(v_2\tau_{1,1}, \beta)$, где равенство по распределению и с.в. $\tau_{1,1}$, β считаются независимыми. Отметим, что

$$E\beta = \frac{m_G^{(2)}}{2m_G}.$$

Положим $c = E\gamma$.

Теорема 2. С вероятностью 1 при $x \rightarrow \infty$

$$\frac{x}{T(x)} \rightarrow \bar{v}_1 = \frac{1 + dc}{1 + dc v_1 v_2^{-1}} v_1. \quad (6)$$

Доказательство. Покажем, что этот случай сводится к рассмотренному случаю $v_2 = 0$. Введем систему координат, которая движется со скоростью v_2 относительно исходной. Найдем среднюю скорость быстрого автомобиля относительно новой системы координат по формуле (4), подставляя $v = v_1 - v_2$, $b = \frac{\lambda m_G}{v_2}$, $a = \frac{c}{v_2}$:

$$\frac{1}{(v_1 - v_2)^{-1} + \frac{\lambda m_G}{v_2^2}} = \frac{v_1 - v_2}{1 + dc v_1 v_2^{-1}}.$$

Тогда средняя скорость быстрого автомобиля относительно исходной системы координат составит

$$\bar{v}_1 = \frac{v_1 - v_2}{1 + dc v_1 v_2^{-1}} + v_2 = \frac{1 + dc}{1 + dc v_1 v_2^{-1}} v_1.$$

4. Критерии образования пробок в сложных транспортных сетях

Обязательным атрибутом транспортной сети (например, городских улиц) является граф, где множество V вершин представляет перекрестки (узлы или пункты обслуживания), а множество ребер $L =$

$= \{(i, j)\}$ — отрезки путей без перекрестков. Пусть число перекрестков равно N . Мы предполагаем, что между двумя перекрестками существует не более одного пути без перекрестков.

Наиболее разработанными являются два класса сетей. С одной стороны, это (по имени авторов и в порядке увеличения общности) сети Джексона, BCMP-сети, DB-сети (см., например, [8, 26]). В них требование (сообщение, автомобиль, работа) обслуживается в каждом проходимом ими узле и затем выбирает случайно следующий узел. С другой стороны, — сети Келли (см. [8]), где каждое требование имеет заранее фиксированный маршрут. Эти два класса сетей связаны как общей техникой, так и близостью результатов. Именно они обладают замечательным свойством мультиплексивности — стационарные распределения в них имеют вид так называемой, продукт-формы. Мы рассматриваем только первый класс сетей.

4.1. Замкнутые сети

Если предполагается, что автомобили не прибывают извне и не убывают вовне, то такая сеть называется **замкнутой**. Таким образом, число автомобилей в сети сохраняется и далее обозначается через M . Движение отдельного автомобиля определяется так. Автомобиль ждет некоторое время на перекрестке i и затем направляется на перекресток j . Выбор j определяется стохастической матрицей маршрутизации: $P = \{p_{ij}\}_{i,j=1,\dots,N}$, где p_{ij} — вероятность того, что с перекрестка i автомобиль поедет (после ожидания) на перекресток j (например, прямо, налево, направо), то есть по улице (i, j) .

Стохастическая матрица P определяет конечную цепь Маркова с дискретным временем, множеством состояний $V = \{1, \dots, N\}$. Эта цепь Маркова предполагается неразложимой. В этом случае система линейных уравнений

$$\rho P = \rho, \quad \rho = (\rho_1, \dots, \rho_N) \iff \sum_{i=1}^N \rho_i p_{ij} = \rho_j, \quad j = 1, \dots, N \quad (7)$$

имеет единственное решение (с точностью до произвольного множителя). Нормированное решение имеет вид

$$\pi_i = \frac{\rho_i}{\sum_{i=1}^N \rho_i}, \quad i = 1, \dots, N.$$

С каждым узлом $i \in V$ свяжем функцию $\mu_i(n_i)$ от числа автомобилей n_i в i -м узле, где $\mu_i(0) = 0$ и $\mu_i(n_i) > 0$ при $n_i > 0$. Эта функция характеризует пропускную способность данного узла и определяет интенсивность выходящего из узла потока автомобилей. Именно вероятность того, что за малый промежуток времени dt из узла i выйдет автомобиль, равна $\mu_i(n_i)dt + o(dt)$ при условии, что в узле находится n_i автомобилей. Используя терминологию теории очередей, будем называть $\mu_i(n_i)$ интенсивностью обслуживания в узле i .

Порядок, в котором пропускаются (обслуживаются) прибывающие в узел автомобили, определяется дисциплиной обслуживания. Простейший вариант дисциплины обслуживания – это обслуживание в порядке поступления. В узле прибывающие автомобили становятся в очередь друг за другом в том порядке, в котором они приехали, и узел пропускает автомобили согласно этой очереди. Если в узле i находится n_i автомобилей, то первый автомобиль в очереди обслуживается с интенсивностью $\mu_i(n_i)$.

Более общая дисциплина обслуживания – это дисциплина разделения общего ресурса, где под ресурсом в данном случае понимается пропускная способность узла. Согласно этой дисциплине ресурс делится в некоторой пропорции между всеми автомобилями, находящимися в данный момент в узле. В общем случае предположим, что k -й автомобиль в i -м узле обслуживается с интенсивностью $\mu_{i,k}(n_i) \leq \mu_i(n_i)$. При этом потребуем, чтобы

$$\sum_{k=1}^{n_i} \mu_{i,k}(n_i) = \mu_i(n_i).$$

Например, общий ресурс может быть разделен в равной степени между всеми автомобилями в очереди:

$$\mu_{i,k}(n_i) = \frac{\mu_i(n_i)}{n_i}.$$

В этом случае каждый из n_i автомобилей потратит экспоненциальное время со средним $n_i \mu_i^{-1}(n_i)$ на прохождение этого узла, при условии, что число автомобилей будет сохраняться равным n_i . Если взять $\mu_{i,1}(n_i) = \mu_i(n_i)$, то получим дисциплину обслуживания в порядке поступления. Таким образом, интенсивности $\mu_{i,k}(n_i)$ полностью определяют дисциплину обслуживания в узлах.

Динамика сети описывается с помощью N -мерной марковской цепи с непрерывным временем $\xi(t) = (\xi_i(t), i = 1, \dots, N)$, где $\xi_i(t)$

— число автомобилей, скопившихся в i -м узле в момент времени t . Случайный процесс $\xi(t)$ принимает значение в пространстве S_M , где S_M — множество всех таких векторов с неотрицательными целочисленными координатами $\bar{n} = (n_1, \dots, n_N)$, что $n_1 + \dots + n_N = M$.

Пусть e_i — базисный вектор, в котором i -я координата равна 1, а остальные координаты равны 0. Из состояния \bar{n} марковская цепь $\xi(t)$ может перейти в одно из состояний $T_{i,j}\bar{n} = \bar{n} - e_i + e_j$, $i \neq j$, с интенсивностью

$$\alpha(\bar{n}, T_{i,j}\bar{n}) = \mu_i(n_i)p_{i,j}, \quad (8)$$

при условии, что $n_i \neq 0$. Переход $\bar{n} \rightarrow T_{i,j}\bar{n}$ соответствует тому, что, выехав из узла i , автомобиль поступает в узел j .

Отметим, что марковская цепь $\xi(t)$ однозначно определяется матрицей маршрутизации P и набором интенсивностей обслуживания в узлах $(\mu_i(n_i), i = 1, \dots, N)$.

Пусть $\rho = (\rho_1, \dots, \rho_N)$ — решение уравнения (7), которое рассматривается как формальное уравнение для интенсивностей ρ_i входящих потоков в узлы (в стационарном режиме они равны выходящим). Решая эти уравнения, находим ρ_i , и тогда стационарное распределение $\nu(n_1, \dots, n_N)$ марковской цепи $\xi(t)$ будет иметь вид

$$\nu(n_1, \dots, n_N) = \frac{1}{Z_{N,M}} \prod_{i=1}^N \frac{\rho_i^{n_i}}{\mu_i(1)\mu_i(2)\dots\mu_i(n_i)}, \quad (9)$$

где нормирующий множитель (малая статсумма)

$$Z_{N,M} = \sum_{n_1+\dots+n_N=M} \prod_{i=1}^N \frac{\rho_i^{n_i}}{\mu_i(1)\mu_i(2)\dots\mu_i(n_i)},$$

что проверяется подстановкой ответа (9) в уравнения Колмогорова для стационарных вероятностей, см., например, [8, 29].

4.2. Открытые сети

Рассмотрим сеть, состоящую из N узлов. В отличие от замкнутой сети, общее число автомобилей в сети теперь не фиксировано. Предположим, что извне сети в узел i поступает пуассоновский поток автомобилей интенсивности λ_i , $i \in \{1, \dots, N\}$.

Зададим матрицу маршрутизации $P = \{p_{ij}\}_{i,j=1,\dots,N}$, где матрица P неразложима и

$$\forall i : \sum_{j=1}^N p_{ij} \leq 1, \quad \exists i_0 : \sum_{j=1}^N p_{i_0 j} < 1. \quad (10)$$

Как и в случае замкнутой сети, $p_{i,j}$ — это вероятность того, что из узла i автомобиль едет в узел j . В отличие от замкнутой сети, добавляется вероятность того, что, выйдя из узла i , автомобиль покидает сеть. Эта вероятность по определению равна

$$p_{i0} = 1 - \sum_{j=1}^N p_{ij}.$$

Как и в случае замкнутой сети, пусть $\mu_i(n_i)$ — интенсивность обслуживания в i -м узле. Тогда с интенсивностью $\mu_i(n_i)p_{i,0}$ автомобиль покидает сеть после выхода из узла i .

Мы будем описывать динамику сети с помощью N -мерного случайного процесса с непрерывным временем $\eta(t) = (\eta_i(t), i = 1, \dots, N)$, где $\eta_i(t)$ — число автомобилей в i -м узле в момент времени t . Случайный процесс $\eta(t)$ является марковской цепью с непрерывным временем и с пространством состояний S , где S — множество N -мерных векторов с неотрицательными целочисленными координатами $\bar{n} = (n_1, \dots, n_N)$.

Из состояния \bar{n} марковская цепь $\xi(t)$ может перейти в одно из состояний $T_{i,j}\bar{n} = \bar{n} - e_i + e_j$, $T_{i,0}\bar{n} = \bar{n} - e_i$, $T_i\bar{n} = \bar{n} + e_i$ с интенсивностями

$$\begin{aligned} \alpha(\bar{n}, T_{i,j}\bar{n}) &= \mu_i(n_i)p_{i,j}, \\ \alpha(\bar{n}, T_{i,0}\bar{n}) &= \mu_i(n_i)p_{i,0}, \\ \alpha(\bar{n}, T_i\bar{n}) &= \lambda_i, \end{aligned} \quad (11)$$

при условии, что $T_{i,j}\bar{n}, T_{i,0}\bar{n}, T_i\bar{n} \in S$.

Таким образом, марковская цепь $\eta(t)$ однозначно определяется триплетом (λ, μ, P) , где $\lambda = (\lambda_1, \dots, \lambda_N)$ — вектор интенсивностей внешних потоков, $\mu = (\mu_i(n_i), i = 1, \dots, N)$ — набор интенсивностей обслуживания в узлах и P — матрица маршрутизации.

Рассмотрим формальное уравнение для интенсивностей входящих потоков в узлы (в стационарном режиме они равны выходящим):

$$\rho = \lambda + \rho P \iff \rho_i = \lambda_i + \sum_{k=1}^N \rho_k p_{ki}, \forall i. \quad (12)$$

В силу условия (10) и неразложимости матрицы P это уравнение имеет единственное решение, которое можно представить в виде

$$\rho = \lambda + \sum_{n=1}^{\infty} \lambda P^n.$$

Далее рассмотрим случай, когда интенсивности обслуживания $\mu_i(n_i) \equiv \mu_i$ не зависят от числа автомобилей в узлах. Введем нагрузки в узлах по формуле

$$r_i = \frac{\rho_i}{\mu_i}, i = 1, \dots, N.$$

Следующую теорему можно найти, например, в [8, 20], она называется иногда теоремой Гордона—Ньюэлла.

Теорема 3. *Марковская цепь $\eta(t)$ является эргодической тогда и только тогда, когда для всех $i = 1, \dots, N$ будет $r_i < 1$. При этом стационарное распределение цепи имеет вид*

$$\sigma(n_1, \dots, n_N) = \prod_{i=1}^N (1 - r_i) r_i^{n_i}.$$

Из этой теоремы легко следует, что средние длины очередей в стационарном режиме равны

$$m_i = \frac{r_i}{1 - r_i}.$$

Если в некоторых узлах i_1, \dots, i_k нагрузка строго больше 1, то марковская цепь $\eta(t)$ является транзиентной (см., например, [33]). Это свидетельствует о том, что средние очереди в узлах i_1, \dots, i_k стремятся к бесконечности с течением времени. Подробный анализ открытых сетей дан в работе [19]. В частности показано, что в узлах, где нагрузка больше 1, средние очереди увеличиваются линейно с ростом времени. При этом найдены скорости роста средних очередей.

4.3. Алгоритм вычисления критической нагрузки в замкнутых сетях

Этот раздел основан на работе [18]. Мы рассмотрим последовательность замкнутых сетей J_N , $N = 1, 2, \dots$. Сеть J_N состоит из N узлов и $M = M(N)$ автомобилей. Интенсивности обслуживания в узлах сети J_N не зависят от длины очереди: $\mu_{i,N}(n_i) \equiv \mu_{i,N}$. Пусть $P_N = \{p_{i,j,N}\}$ – матрица маршрутизации в N -й сети; P_N предполагается неразложимой.

Пусть $\rho_N = (\rho_{1,N}, \dots, \rho_{N,N})$ – вектор с положительными компонентами, удовлетворяющий уравнению

$$\rho_N = \rho_N P_N. \quad (13)$$

Относительные нагрузки в узлах определяются как

$$r_{i,N} = C_N^{-1} \rho_{i,N} \tau_{i,N},$$

где $\tau_{i,N} = \mu_{i,N}^{-1}$ и $C_N = \max_{i=1, \dots, N} \rho_{i,N} \tau_{i,N}$. Очевидно, что $r_{i,N} \in [0, 1]$.

В соответствии с (9) стационарное распределение числа автомобилей $\xi_{i,N,M}$ в узлах сети J_N равно

$$P_{N,M}(\xi_{i,N,M} = n_i, i = 1, \dots, N) = \frac{1}{Z_{N,M}} \prod_{i=1}^N r_{i,N}^{n_i},$$

где нормирующий множитель (малая статсумма)

$$Z_{N,M} = \sum_{n_1 + \dots + n_N = M} \prod_{i=1}^N r_{i,N}^{n_i}. \quad (14)$$

Многие важные характеристики сети выражаются через статсумму.

Упражнение 3. Показать, что среднее число автомобилей в i -м узле в стационарном режиме равно

$$m_{i,N,M} = E\xi_{i,N,M} = \frac{r_{i,N}}{Z_{N,M}} \frac{\partial Z_{N,M}}{\partial r_{i,N}}. \quad (15)$$

Ниже мы будем требовать слабую сходимость относительных нагрузок $r_{i,N}$. Точнее, определим выборочную меру на отрезке $[0, 1]$:

$$I_N(A) = \frac{1}{N} \sum_{i:r_{i,N} \in A} 1,$$

где A – произвольное борелевское множество из отрезка $[0, 1]$. Предположим, что при $N \rightarrow \infty$ меры I_N слабо сходятся к некоторой вероятностной мере I заданной на отрезке $[0, 1]$.

Нас будет интересовать случай больших N, M , точнее $N, M \rightarrow \infty$, причем так, что $\frac{M}{N} \rightarrow \lambda = \text{const}$, то есть удельное число автомобилей на один узел постоянно. Именно это число определяет существование пробок.

Замечание 1. *Интересно найти конкретные последовательности растущих графов, для которых предельная мера I явно описывается. Некоторые примеры, где мера I одноточечна см. в ссылках к работе [18], см. также с. 157–160 в [29].*

В терминах предельной меры I мы найдем критическое значение плотности λ_{cr} , так что при $\lambda < \lambda_{cr}$ средние длины очередей равномерно ограничены. Если $\lambda \geq \lambda_{cr}$, то в узле с максимальной относительной нагрузкой средняя длина очереди стремится к бесконечности, что означает возникновение пробки.

Положим

$$h(z) = \int_0^1 \frac{r}{1 - zr} dI(r),$$

где $z \in \mathbb{C} \setminus [1, +\infty)$. Функция $h(z)$ строго возрастает на $[0, 1]$. Обозначим

$$\lambda_{cr} = \lim_{z \rightarrow 1^-} h(z).$$

Будем предполагать, что $\lambda_{cr} > 0$.

Теорема 4. • Если $\lambda < \lambda_{cr}$, то средние очереди равномерно ограничены: существует такая константа B , что $m_{i,N} < B$, равномерно по $N \geq 1$ и $1 \leq i \leq N$.

• Если $\lambda \geq \lambda_{cr}$ и $i(N)$ удовлетворяет условию $r_{i(N),N} = 1$, то $m_{i(N),N} \rightarrow \infty$, при $N \rightarrow \infty$ т.е. пробки будут в тех узлах, где нагрузка максимальна.

При $z \in \mathbb{C} \setminus [1, +\infty)$ положим

$$S_N(z) = -\lambda(1 + \varepsilon_N) \ln z - \frac{1}{N} \sum_{i=1}^N \ln(1 - zr_{i,N}), \quad (16)$$

$$S(z) = -\lambda \ln z - \int_0^1 \ln(1 - zr) dI(r),$$

где $\lambda(1 + \varepsilon_N) = \frac{M}{N}$.

Введем производящую функцию (большую статсумму):

$$\Xi_N(z) = \sum_{M=0}^{\infty} z^M Z_{N,M} = \prod_{i=1}^N \frac{1}{1 - zr_i}, |z| < 1.$$

По формуле Коши и формуле (16) имеем следующее выражение для статсуммы (14):

$$Z_{N,M} = \frac{1}{2\pi i} \int_{\gamma} \frac{\Xi_N(z)}{z^{M+1}} dz = \frac{1}{2\pi i} \int_{\gamma} \frac{\exp(NS_N(z))}{z} dz, \quad (17)$$

где $\gamma = \{z \in \mathbb{C} : |z| = \sigma < 1\}$. Для средних, согласно (15), имеем

$$m_{i,N} = \frac{1}{2\pi i Z_N} \int_{\gamma} \frac{r_{i,N}}{1 - zr_{i,N}} \exp(NS_N(z)) dz. \quad (18)$$

Можно показать, что для стационарного распределения длин очередей справедлива формула

$$P_{N,M}(\xi_{1,N,M} = n_1, \dots, \xi_{K,N,M} = n_K) = \\ = \frac{1}{2\pi i Z_N} \int_{\gamma} z^{-1} \prod_{i=1}^K (1 - zr_{i,N}) (zr_{i,N})^{n_i} \exp(NS_N(z)) dz. \quad (19)$$

В доказательстве теорем этого раздела существенную роль играет *метод перевала* (см. [27]), точнее его обобщение, поскольку функция в показателе экспоненты зависит от N . Из уравнения

$$\frac{\partial S_N(z)}{\partial z} = 0 \quad (20)$$

находятся точки перевала. Пусть $z_{0,N}$ — корень этого уравнения, лежащий в интервале $(0, 1)$.

Упражнение 4. Показать, что все корни уравнения (20) действительны и положительны. Всегда существует единственный корень, лежащий в интервале $(0, 1)$.

Пусть z_0 – корень уравнения

$$h(z) = \frac{\lambda}{z} \Leftrightarrow \frac{\partial S(z)}{\partial z} = 0, \quad (21)$$

лежащий в интервале $(0, 1)$.

Упражнение 5. • Доказать, что при всех λ существует предел $\lim_{N \rightarrow \infty} z_{0,N} = z_0 = z_0(\lambda) > 0$.

- Если $\lambda < \lambda_{cr}$, то $z_0(\lambda)$ – корень уравнения (21); $z_0(\lambda)$ строго возрастает по λ , $z_0(\lambda) \in (0, 1)$, $\lim_{\lambda \rightarrow \lambda_{cr}^-} z_0(\lambda) = 1$.
- Если $\lambda \geq \lambda_{cr}$, то $z_0 = 1$.

В следующей теореме мы находим асимптотику статсуммы и предельное распределение для последовательности замкнутых сетей J_N .

Теорема 5. Пусть $\lambda < \lambda_{cr}$.

- При $N \rightarrow \infty$ статсумма Z_N и свободная энергия $F_N = \frac{1}{N} \ln Z_N$ имеют следующие асимптотики:

$$Z_N \sim \frac{\exp(NS_N(z_{0,N}))}{z_0 \sqrt{2\pi N S''(z_0)}}, \quad F_N = \frac{1}{N} \ln Z_N \sim S(z_0).$$

- Если при $i = 1, \dots, K$ существуют пределы $r_i = \lim_{N \rightarrow \infty} r_{i,N}$, то

$$\lim_{N \rightarrow \infty} m_{i,N} = \frac{z_0 r_i}{1 - z_0 r_i},$$

$$\lim_{N \rightarrow \infty} P_{N,M}(\xi_{1,N,M} = n_1, \dots, \xi_{K,N,M} = n_K) = \prod_{i=1}^K (1 - z_0 r_i)(z_0 r_i)^{n_i}.$$

Таким образом, в пределе мы получаем открытую сеть, состоящую из независимых очередей.

Доказательство теорем 4 и 5. Мы приведем более общий результат, из которого будут следовать теоремы 4 и 5. Пусть $U_d(v) = \{z \in \mathbb{C} : |z - v| < d\}$. Рассмотрим контур $\gamma = \{z \in \mathbb{C} : |z| = z_0(\lambda)\}$.

Теорема 6. Пусть $\lambda < \lambda_{cr}$ и $f(\theta, z)$, $\theta \in \Theta$, – семейство функций голоморфных в кольце $\{z \in C : z_0(\lambda) - \delta_0 < |z| < z_0(\lambda) + \delta_0\}$ при некотором $\delta_0 > 0$, равномерно ограниченных в этом кольце и таких, что для заданного достаточно малого $\epsilon > 0$ существует такое $\delta_u > 0$ и такая ненулевая действительная константа f_u , что $|f(\theta, z)/f_u - 1| < \epsilon$ при $z \in U_{2\delta_u}(z_0)$, $\theta \in \Theta$.

Тогда при достаточно больших N , равномерно по $\theta \in \Theta$

$$\frac{1}{2\pi i} \int_{\gamma} f(\theta, z) \exp(NS_N(z)) dz = \frac{f_u \exp(NS_N(z_0, N))}{\sqrt{2\pi N S''(z_0)}} (1 + \zeta_N),$$

где $|\zeta_N| < 25\epsilon$.

Доказательство этой теоремы основано на применении метода перевала (saddle-point method)(см. [27]). Отличие от стандартной ситуации состоит в том, что функция в показателе экспоненты зависит от N . Подробное доказательство можно найти в оригинальной статье [18].

Доказательство теоремы 5. Используя теорему 6, докажем первый пункт теоремы 5. Согласно (17) имеем

$$Z_{N,M} = \frac{1}{2\pi i} \int_{\gamma} \frac{\exp(NS_N(z))}{z} dz,$$

где $\gamma = \{z \in \mathbb{C} : |z| = z_0(\lambda)\}$. Положив $f(\theta, z) = z^{-1}$, $f_u = z_0^{-1}$ и применив теорему 6, получим, что для любого достаточно малого $\epsilon > 0$ при достаточно больших N

$$Z_N = \frac{\exp(NS_N(z_0, N))}{z_0 \sqrt{2\pi N S''(z_0)}} (1 + \zeta_N), \quad |\zeta_N| < 25\epsilon. \quad (22)$$

Второй пункт теоремы 5 доказывается аналогично с использованием формулы (18) для средней очереди и формулы (19) для совместного распределения длин очередей.

Упражнение 6. Доказать третье утверждение теоремы 5, используя теорему 6 и формулы (18), (19).

Доказательство теоремы 4. Чтобы доказать первый пункт теоремы 4, рассмотрим семейство функций

$$f(\theta, z) = \frac{A}{z} + \frac{\theta}{1 - z\theta}, \quad \theta \in \Theta = [0, 1], \quad A > 0, \quad f_u = \frac{A}{z_0}.$$

Зафиксируем малое $\epsilon > 0$ и выберем $\sigma_u = \frac{\epsilon}{8}$, $A = \frac{16z_0}{(1-z_0)\epsilon}$. По теореме 6 имеем для достаточно больших N и всех $\theta \in \Theta$

$$\begin{aligned} & \frac{1}{2\pi i} \int_{\gamma} \left(\frac{A}{z} + \frac{\theta}{1 - z\theta} \right) \exp(NS_N(z)) dz = \\ & = \frac{A \exp(NS_N(z_{0,N}))}{z_0 \sqrt{2\pi N S''(z_0)}} (1 + \zeta_N), \quad |\zeta_N| < 25\epsilon. \end{aligned}$$

Разделив на Z_N и применив (22) к правой части получившегося равенства, получим при достаточно больших N

$$A + \frac{1}{Z_N} \frac{1}{2\pi i} \int_{\gamma} \frac{\theta}{1 - z\theta} \exp(NS_N(z)) dz = A (1 + \zeta'_N), \quad |\zeta'_N| < 30\epsilon.$$

Из последнего равенства и формулы (18) следует равномерная ограниченность $m_{i,N}$.

Докажем второе утверждение теоремы 4. Для этого потребуется следующее свойство монотонности: при любых $M_2 \geq M_1 > 0$ и любом $N \geq 1$ выполнено $m_{i,M_2,N} \geq m_{i,M_1,N}$.

Поскольку $z_0(\lambda)$ строго возрастает по λ , $z_0(\lambda) \in (0, 1)$ и

$$\lim_{\lambda \rightarrow \lambda_{cr}^-} z_0(\lambda) = 1,$$

то функция

$$\frac{z_0(\lambda)}{1 - z_0(\lambda)}$$

монотонно возрастает и стремится к ∞ , когда $\lambda \nearrow \lambda_{cr}$. Поэтому для любого $m > 0$ существует такое $\lambda' = \lambda'(m) < \lambda_{cr}$, что

$$\frac{z_0(\lambda')}{1 - z_0(\lambda')} = m + 1.$$

Без ограничения общности можно считать, что $i(N) \equiv 1$ и $r_{1,N} = 1$. Если взять $M'(N) = [\lambda'N]$, то по теореме 5

$$\lim_{N \rightarrow \infty} m_{1,M'(N),N} = \frac{z_0(\lambda')}{1 - z_0(\lambda')}.$$

Следовательно, при достаточно больших N

$$m_{1,M'(N),N} > \frac{z_0(\lambda')}{1 - z_0(\lambda')} - 1 = m.$$

Но $M/N \rightarrow \lambda \geq \lambda_{cr} > \lambda'$, поэтому при достаточно больших N имеем $M(N) \geq M'(N)$. По свойству монотонности $m_{1,N} = m_{1,M(N),N} \geq m_{1,M',N} > m$ для достаточно больших N . Это доказывает, что $m_{1,N} \rightarrow \infty$.

Технические обобщения и математические проблемы. Мы предполагали мгновенное перемещение между перекрестками. При этом не учитываются времена движения по улицам. Это допущение однако легко устраняется усложнением графа. Именно введением дополнительных вершин u_{ij} , соответствующих улицам, и средних времен $\tau_{ij} = \mu_{ij}^{-1}$ пребывания на улицах. В терминах теории очередей это значит, что улицы рассматриваются как узлы обслуживания с бесконечным числом обслуживающих устройств и время обслуживания экспоненциально распределено со средним $\tau_{ij} = \mu_{ij}^{-1}$.

Отметим, что результаты раздела 4.1 могут быть обобщены на случай, когда сеть содержит узлы с бесконечным числом обслуживающих устройств. Пусть, например, сеть содержит один узел такого типа ($i = 0$) и $\mu_{0,N}(n) = n\nu_N$ – интенсивность обслуживания в этом узле. Пусть $\rho_N = (\rho_{0,N}, \dots, \rho_{N,N})$ – решение уравнения (13). Тогда относительные нагрузки определим по формуле

$$r_{i,N} = \frac{\mu_{0,N}}{\rho_{0,N}} \frac{\rho_{i,N}}{\mu_{i,N}},$$

так что $r_{0,N} = 1$. Согласно (9), стационарное распределение длин очередей имеет вид

$$P_{N,M}(\xi_{i,N,M} = n_i, i = 1, \dots, N) = \frac{1}{\widehat{Z}_{N,M}} \frac{1}{(M - \sum_{i=1}^M n_i)!} \prod_{i=1}^N r_{i,N}^{n_i},$$

где

$$\widehat{Z}_{N,M} = \sum_{n_1 + \dots + n_N \leq M} \frac{1}{(M - \sum_{i=1}^M n_i)!} \prod_{i=1}^N r_{i,N}^{n_i},$$

и большая статсумма сети равна

$$\widehat{\Xi}_N(z) = e^z \prod_{i=1}^N \frac{1}{1 - zr_{i,N}}.$$

Положим

$$q_{i,N} = \frac{r_{i,N}}{p_N}, \quad w = zp_N, \quad p_N = \max_{1 \leq i \leq N} r_{i,N}.$$

Тогда

$$\widehat{\Xi}_N(w) = e^{w/p_N} \prod_{i=1}^N \frac{1}{1 - wq_{i,N}}.$$

В предположении, что $p_N N \rightarrow \alpha > 0$ при $N \rightarrow \infty$ мы можем найти критическое значение плотности λ по формуле

$$\lambda_{cr} = \alpha^{-1} + \lim_{w \rightarrow 1^-} \int_0^1 \frac{q}{1 - wq} dI(q),$$

где, как и раньше, мера I есть слабый предел при $N \rightarrow \infty$ выборочных мер

$$I_N(A) = \frac{1}{N} \sum_{i: q_{i,N} \in A} 1,$$

где A – произвольное борелевское множество из отрезка $[0, 1]$.

В работе [21] для замкнутых сетей аналогичные результаты получаются для случая более общей зависимости интенсивностей от длин очередей в узлах.

Прием усложнения графа позволяет устраниить также другое ограничение, что для данного перекрестка средняя длительность красного света одна для всех направлений. Необходимо вместо вершины i , соответствующей перекрестку, ввести несколько вершин (i, d) , где d перечисляет возможные направления движения на перекрестке i . Это конечно налагает ограничение на соответствующие времена обслуживания $\tau_{i,d}$ в новых вершинах типа

$$\sum_d \tau_{i,d} = \tau_i.$$

Мы ограничились задачей, когда в системе возникала хотя бы одна пробка. Интереснее рассмотреть ситуацию, когда в разных местах графа одновременно возникает много пробок.

Связь с практикой. Эта модель удобна тем, что все ее параметры можно оценить. Именно на практике статистические оценки параметров p_{ij}, μ_i имеют вид (например, для постоянных μ_i)

$$p_{ij} \sim \frac{N_{ij}(T)}{\sum_j N_{ij}(T)}, \mu_i = \frac{1}{T} \sum_j N_{ij}(T),$$

где $N_{ij}(T)$ — число автомобилей, повернувших за время T на перекрестке i в направлении j .

Практически интересна прежде всего задача оптимизации светофоров, которая может достигаться выбором $\tau_{i,d}$, и за счет изменения матрицы P подсказками о выборе маршрута. Более того, в экспоненциальной модели многие случайные величины независимы, и, значит, полностью игнорируется проблема синхронизации светофоров.

Следует сказать, что проблема светофоров в стохастическом контексте только начинает изучаться, и постановки задач там должны быть более тонкими. Ранее данная проблема изучалась в жидкостных моделях. Однако пока нет понимания (а тем более вывода) связи жидкостных транспортных моделей со стохастическими (как в статистической физике).

Литература

1. Хейт Ф. Математическая теория транспортных потоков. М.: Мир, 1966.
2. Renyi A. On two mathematical models of the traffic on a divided highway // Journal of Applied Probability. 1964. V. 1. P. 311–320.
3. Solomon H., Wang P. Nonhomogeneous Poisson fields of random lines with applications to traffic flow // Proc. Sixth Berkeley Symp. on Math. Statist. and Prob. 1972. V. 3. P. 383–400.
4. Solomon H. Geometric Probability. Philadelphia: SIAM, 1978.
5. Daley D., Vere-Jones D. An Introduction to the Theory of Point Processes V. 1. Springer, 2003.
6. Кокс Д., Смит Б. Теория восстановления. М.: Мир, 1967.
7. Cox D. R., Isham V. Point processes. Chapman and Hall, 1980.
8. Kelly F. Reversibility and stochastic networks. N.Y.: Wiley, 1979.
9. Caceres F., Ferrari P., Pechersky E. A slow-to-start traffic model related to a M/M/1 queue // Journal of Statistical Mechanics: Theory and Experiment. 2007. [arXiv:0703709 cond-mat].
10. Иносэ Х., Хамада Т. Управление дорожным движением. М.: Транспорт, 1983.
11. Traffic flow theory: A state-of-the-art report. Editors Gartner N. H., Messer C. J., Rathi A. K. Washington DC: Transportation Research Board, 2001.
12. Blank M. Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111. P. 903–930.
13. Jost D., Nagel K. Probabilistic Traffic flow breakdown in stochastic car following models // Traffic and Granular Flow. 2005 V. 03. Part 2. P. 87–103.
14. Lotito P., Mancinelli E., Quadrat J.-P. A min-plus derivation of the fundamental car-traffic law // Automatic Control IEEE Transactions. May 2005. V. 50. N 5. P. 699–705.

15. Kerner B. S. Introduction to modern traffic flow theory and control. Berlin: Springer, 2009.
16. Замятин А. А., Малышев В. А. Накопление на границе для одномерной стохастической системы частиц // Проблемы передачи информации. 2007. Т. 43. № 4. С. 68–82.
17. Lighthill M. J., Whitham G. B. On kinematic waves. II. Theory of traffic flow on long crowded roads // Proc. R. Soc. London, Se. A. 1955. V. 229. P. 281–345.
18. Malyshev V., Yakovlev A. Condensation in large closed Jackson networks // Ann. Appl. Probab. 1996. V. 6. N. 1. P. 92–115.
19. Botvich D. D., Zamyatin A. A. On fluid approximations for conservative networks // Markov Processes and Related Fields. 1995. V. 1. N. 1. P. 113–141.
20. Fayolle G., Malyshev V., Menshikov M. Topics in the constructive theory of countable Markov chains. Cambridge University Press, 1995.
21. Fayolle G., Lasgouttes J.-M. Asymptotics and Scalings for Large Product-Form Networks via the Central Limit Theorem // Markov Processes and Related Fields. 1996. V. 2. N. 2. P. 317–349.
22. Revised Traffic Flow Theory. A state-of-art report. Editors Gartner J.-M., Messer C. J., Rathi A. K. Washington DC: Transportation Research Board, 2001.
23. Рюэль Д. Статистическая механика. Строгие результаты. М.: Мир, 1971.
24. Малышев В. А., Минлос Р. А. Гиббсовские случайные поля. М.: Наука, 1985.
25. Малышев В. А. Случайные грамматики // Успехи мат. наук. 1998. Т. 53. № 2. С. 107–134.
26. Serfozo R. Introduction to stochastic networks. Springer, 1999.
27. Федорюк М. В. Метод перевала. М.: Наука, 1977.

28. *Буслаев А. П., Новиков А. В., Приходько В. М., Таташев А. Г., Яшина М. В.* Вероятностные и имитационные подходы к оптимизации автодорожного движения. М.: Мир, 2003.
29. *Афанасьев Л. Г.* Очерк Исследования Операций. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2007.
30. *Кингман Дж.* Пуассоновские процессы. М.: МЦНМО, 2007.
31. *Blythe R. A., Evans M. R.* Nonequilibrium steady states of matrix-product form: a solver's guide // Journal of Physics A: Mathematical and Theoretical. 2007. V. 40. N. 46.
32. *Derrida B.* Non-equilibrium steady states: fluctuations and large deviations of the density and of the current // Journal of Statistical Mechanics: Theory and Experiment. July 2007.
33. *Малышев В. А.* Кратчайшее ведение в современные вероятностные модели. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2009.
<http://mech.math.msu.su/~malyshev/Malyshev/Lectures/course.pdf>

Транспортная задача и концентрация

Первые результаты о концентрации были получены П. Леви в его книге по функциональному анализу [10]. Само название «концентрация мер» было предложено В. Мильманом. Благодаря ему же явление концентрации приобрело большую популярность в математическом сообществе и нашло многочисленные приложения в функциональном анализе, геометрии, вероятности, комбинаторике и технических науках.

Среди сугубо математических приложений упомянем: 1) новое доказательство теоремы Дворецкого о «почти круглых» сечениях выпуклых тел, 2) изопериметрические теоремы сравнения М. Громова для многообразий положительной кривизны Риччи, 3) приложения в теории гауссовых случайных процессов (например, оценки статистического супремума), 4) применения к другим функциональным и вероятностным неравенствам (неравенства типа Соболева, неравенства типа Брунна–Минковского для выпуклых тел и т.п.). Подробнее об этом можно узнать в книгах [2; 7–9]. О вероятностных приложениях см. статью [11]. Для ознакомления с недавними результатами о концентрации и функциональных неравенствах для логарифмически вогнутых распределений также рекомендуем статью [13]. Теоремы о концентрации также позволяют оценить скорость сходимости системы к равновесному состоянию (см. комментарий ниже и приложение Е. В. Гасниковой настоящего пособия).

Классический пример свойства концентрации дает стандартное нормальное (гауссовское) d -мерное распределение γ . Как известно, плотность такого распределения задается формулой

$$\rho_\gamma(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|x|^2}{2}\right).$$

Для произвольного множества A со свойством $\gamma(A) > \frac{1}{2}$ рассмотрим его r -окрестность:

$$A_r = \{x : \exists y \in A : |x - y| \leq r\}.$$

Тогда выполнено следующее неравенство концентрации:

$$\gamma(A_r) \geq 1 - \frac{1}{2}e^{-\frac{r^2}{2}}. \quad (1)$$

Как мы видим, $P(A_r)$ очень быстро (квадратично экспоненциально) стремится к единице.

Равномерное распределение σ на единичной сфере $S^{d-1} \subset R^d$ также обладает аналогичным свойством:

$$\sigma(A_r) \geq 1 - \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-d\frac{r^2}{2}}. \quad (2)$$

Одно из важных следствий неравенств такого типа — неравенства для колебаний липшицевых функций. Пусть f — 1-липшицева функция, т.е. удовлетворяющая соотношению $|f(x) - f(y)| \leq |x - y|$. Используя формулу коплощади

$$\int g(x) d\gamma = \int_{-\infty}^{\infty} \gamma(\{g > t\}) dt,$$

из (1) можно получить неравенство вида

$$\gamma(\{x : f(x) - \int f d\gamma > t\}) \leq 2e^{-ct^2}$$

для некоторой универсальной константы c . Полученное свойство обычно формулируется в виде «липшицевы функции с большой дрейфом вероятности мало отличаются от своего среднего значения».

Несмотря на простой вид, доказательство (1) нетривиально. Классический способ основан на описании так называемых «изопериметрических множеств» — множеств, имеющих наименьшую границу среди множеств такой же меры (вероятности). В евклидовом пространстве таким, как известно, является шар. На сфере их роль выполняют сферические «шары», а в пространстве, наделенном гауссовым распределением, — полупространства $\{x : \langle x, a \rangle < c\}$, $a \in \mathbb{R}^d$, $c \in \mathbb{R}$. Это — хорошо известный в теории вероятностей результат, доказанный В. Судаковым и Б. Цирельсоном (и независимо от них К. Бореллем, см. [1]). Из того факта, что r -окрестности полупространств являются полупространствами (т.е. опять изопериметрическими множествами), несложно извлечь следствие, что функция

$r \mapsto F(A, r) = \gamma(A_r)$ среди всех множеств фиксированной вероятности r растет медленнее всего для изопериметрического A . Для этого множества функция $F(A, r)$ явно вычисляется, и мы получаем (1).

Указанный способ плох тем, что явно найти изопериметрические множества в более общем случае невозможно. Существует несколько подходов к доказательству неравенства концентрации. В настоящем пособии мы опишем связь явления концентрации с транспортной задачей, возникшей и развившейся в совершенно другой области математики. Связь эта была найдена в работе К. Мартон [12].

Транспортная задача ведет свою историю от классической работы Г. Монжа [14], написанной в 1781 году. В этой работе задача была сформулирована следующим образом: имеется куча песка и яма одинаковых объемов. Как засыпать песком яму, потратив наименьшие усилия на перевозку? Конечно, это не единственная возможная «экономическая» формулировка транспортной задачи. Речь, например, может идти о перевозке грузов со складов по заданным адресам.

В дискретной постановке мы имеем набор точек $\{x_i\}$, $1 \leq i \leq N$. Задано N других точек $\{y_i\}$ и функция стоимости $c(x, y)$ (например, расстояние или квадрат расстояния). Как построить взаимно-однозначное отображение T , сопоставляющее каждой точке из первого набора точку из второго набора, так, чтобы суммарная стоимость $\sum_{i=1}^N c(x_i, y_i)$ была наименьшей?

В дальнейшем транспортная задача переживала как периоды забвения, так и бурного развития. На языке современной математики транспортная задача была переформулирована и решена Л. Канторовичем в 40-х годах XX-го века (см. [3]) и получила в дальнейшем название задачи Монжа—Канторовича. Важным шагом в работах Канторовича было применение развитого им в теории линейного программирования метода двойственности и формулировка транспортной задачи на языке теории меры и функционального анализа. О приложениях двойственности и задач типа транспортной в технических науках см., например, главу 2 и приложение Е. В. Гасниковой настоящего пособия.

Пусть задана пара вероятностных распределений μ и ν на пространствах $X = Y = \mathbb{R}^d$. Решением задачи Монжа—Канторовича называется распределение m на \mathbb{R}^{2d} , удовлетворяющее следующим условиям:

1. Проекции m на X и Y равны соответственно μ и ν :

$$\text{pr}_X m = \mu, \quad \text{pr}_Y m = \nu. \quad (3)$$

2. Распределение m реализует минимум следующего функционала:

$$\mathcal{F}(m) : m \rightarrow \int_{X \times Y} c(x, y) dm,$$

где $c : X \times Y \rightarrow \mathbb{R}$ — некоторая функция, называемая функцией стоимости (cost function).

При весьма общих предположениях задача Монжа—Канторовича имеет решение.

В дальнейшем мы будем интересоваться только случаем $c(x, y) = |x - y|^2$.

Обратим теперь внимание на важное отличие задачи Монжа—Канторовича от исходной задачи Монжа. В задаче Монжа речь идет о *перевозке груза*, что на математическом языке соответствует задаче существования *отображения* $T : X \rightarrow Y$, преобразующего распределение μ в распределение ν (последнее означает, что $\nu(A) = \mu(\{x : T(x) \in A\})$) и реализующего минимум функционала

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}^d} |x - T(x)|^2 d\mu.$$

Оказывается, что при весьма общих условиях (например, распределения μ и ν непрерывны) эти задачи эквивалентны. Если m — решение задачи Монжа—Канторовича, то m сосредоточено на графике некоторого отображения T : $m\{(x, y) : y = T(x)\} = 1$. Мы будем называть T оптимальным отображением. Существование T было доказано Я. Бренье в [5]. Более того, имеет место следующий удивительный факт.

Теорема 1. T имеет вид

$$T(x) = \nabla \varphi(x),$$

где φ — некоторая выпуклая функция.

Величина

$$W_2(\mu, \nu) = \sqrt{\int_{\mathbb{R}^d} |x - T(x)|^2 d\mu}$$

называется расстоянием Канторовича (также можно встретить названия «расстояние Канторовича–Рубинштейна» и «расстояние Варштейна»). Действительно, можно проверить, что $W_2(\mu, \nu)$ является расстоянием на пространстве вероятностных распределений.

Пусть теперь распределения μ и ν заданы плотностями $\mu = \rho_1 dx$, $\nu = \rho_2 dx$. Свойство T отображать μ в ν аналитически записывается с помощью формулы замены переменной:

$$\rho_2(\nabla\varphi) \det D^2\varphi = \rho_1.$$

Если рассматривать φ как неизвестную функцию, то мы получаем *уравнение Монжа–Ампера*. Под $D^2\varphi = D(\nabla\varphi)$ подразумевается матрица вторых производных (гессиан) функции φ .

Сделаем важное техническое замечание. Выполнено очевидное тождество

$$\int_{\mathbb{R}^d} |x - y|^2 dm = \int_{\mathbb{R}^d} |x|^2 d\mu - 2 \int_{\mathbb{R}^d} \langle x, y \rangle dm + \int_{\mathbb{R}^d} |y|^2 d\nu.$$

Поэтому поиск минимума функционала $\int_{\mathbb{R}^d} |x - y|^2 dm$ эквивалентен поиску максимума функционала $\int_{\mathbb{R}^d} \langle x, y \rangle dm$.

Существование φ может быть доказано разными способами. Стандартный подход состоит в применении метода двойственности Канторовича и работе с так называемыми циклически монотонными множествами. При этом выпуклость φ получается автоматически. Двойственная задача Канторовича принимает вид

$$\int \Phi(x) d\mu + \int \Psi(y) d\nu \rightarrow \max,$$

где функционал максимизируется среди функций, удовлетворяющих условию $\Phi(x) + \Psi(y) \leq |x - y|^2$. Отображение T связано с Φ следующим образом: $T(x) = x - \nabla\Phi(x)$ (см. подробнее гл. 1, [16]).

Формальное, но поучительное доказательство того факта, что T является градиентом, можно получить путем вывода уравнения Эйлера–Лагранжа (см. [6]). Пусть T — произвольное отображение из μ в ν . Решение задачи Монжа–Канторовича можно искать, как условный экстремум функционала

$$\int_{\mathbb{R}^d} \langle T(x), x \rangle d\mu$$

при условии $\rho_\nu(T) \det DT = \rho_\mu$. Составим функционал Лагранжа:

$$\int_{\mathbb{R}^d} \left(\langle T(x), x \rangle \rho_\mu + \lambda(x)(\rho_\nu(T) \det DT - \rho_\mu) \right) dx.$$

Функция λ играет роль множителя Лагранжа. Чтобы найти первую вариацию функционала Лагранжа, рассмотрим инфинитезимальную вариацию

$$T_\varepsilon(x) = T(x) + \varepsilon \omega(x)$$

отображения T . Здесь ω — гладкое векторное поле с компактным носителем. Очевидно,

$$\rho_\nu(T_\varepsilon) \approx \rho_\nu(T) + \varepsilon \langle \omega, \nabla \rho_\nu(T) \rangle.$$

Можно проверить, что

$$\begin{aligned} \det(DT + \varepsilon D\omega) &= \det DT \cdot \det(I + \varepsilon(DT)^{-1}D\omega) \approx \\ &\approx \det DT \left(1 + \varepsilon \text{Tr}[DT^{-1} \cdot D\omega] \right). \end{aligned}$$

Таким образом, первая вариация функционала Лагранжа равна

$$\begin{aligned} \int_{\mathbb{R}^d} & \left(\langle \omega(x), x \rangle \rho_\mu + \lambda(x) \cdot \rho_\nu(T) \text{Tr}[DT^{-1}D\omega] + \right. \\ & \left. + \lambda(x) \langle \nabla \rho_\nu(T), \omega \rangle \frac{\rho_\mu}{\rho_\nu(T)} \right) dx. \end{aligned}$$

Заметим, что $\text{div}(\omega(T^{-1})) = \text{Tr}D[\omega(T^{-1})] = \text{Tr}[DT^{-1}D\omega](T^{-1})$.

Интегрируя по частям и применяя замену переменных, несложно убедиться в том, что

$$\begin{aligned} \int_{\mathbb{R}^d} \lambda(x) \cdot \text{Tr}[DT^{-1} \cdot D\omega] \rho_\mu dx &= \int_{\mathbb{R}^d} \lambda(T^{-1}) \text{div}(\omega(T^{-1})) \rho_\nu dx = \\ &= - \int_{\mathbb{R}^d} \langle \nabla [\lambda(T^{-1})], \omega(T^{-1}) \rangle \rho_\nu dx - \int_{\mathbb{R}^d} \lambda(T^{-1}) \langle \omega(T^{-1}), \frac{\nabla \rho_\nu}{\rho_\nu} \rangle \rho_\nu dx. \end{aligned}$$

Следовательно, вариация равна

$$\int_{\mathbb{R}^d} \left(\langle \omega(x), x \rangle \rho_\mu - \int_{\mathbb{R}^d} \langle \nabla [\lambda(T^{-1})], \omega(T^{-1}) \rangle \rho_\nu dx \right) = 0.$$

Пусть $\lambda = u(T)$, где u — некоторая функция. Применяя опять формулу замены переменных, получаем, что для любого гладкого поля ω выполнено

$$\int_{\mathbb{R}^d} \left(\langle \omega(x), x \rangle - \langle \nabla u(T), \omega(x) \rangle \right) \rho_\mu dx = 0.$$

Следовательно:

$$\nabla u(T) = x \implies T^{-1} = \nabla u.$$

Таким образом, $T^{-1} = \nabla u$. В силу симметричности задачи относительно μ и ν то же утверждение можно сделать для самого отображения T .

В качестве иллюстрации эффективного использования оптимальной транспортировке в анализе приведем доказательство М. Громова классического изопериметрического неравенства.

Пример 1. Среди множеств фиксированной меры Лебега шары имеют наименьшую поверхностную меру.

Доказательство. Пусть $A \subset \mathbb{R}^d$ — борелевское множество, $B_r = \{x : |x| \leq r\}$ — шар, удовлетворяющий условию $\lambda(A) = \lambda(B_r)$, где λ — мера Лебега на \mathbb{R}^d . Пусть $T = \nabla \varphi$ — оптимальная транспортировка, отображающая $\lambda|_A$ в $\lambda|_{B_r}$. По формуле замены переменных $\det D^2 \varphi = 1$ на A (для простоты изложения считаем, что φ — гладкая функция, хотя аргументы ниже легко обобщаются на негладкий случай). Матрица $D^2 \varphi$ симметрична и неотрицательна, поэтому $1 = \sqrt[d]{\det D^2 \varphi} \leq \frac{\Delta \varphi}{d}$ в силу неравенства между средним арифметическим и средним геометрическим. Проинтегрируем это неравенство по A и применим теорему Остроградского-Гаусса:

$$d\lambda(A) \leq \int_A \Delta \varphi \, dx = \int_{\partial A} \langle \nabla \varphi, n_A \rangle d\mathcal{H}^{d-1} \leq r \mathcal{H}^{d-1}(\partial A).$$

Здесь n_A — единичная нормаль к ∂A , \mathcal{H}^{d-1} — $(d-1)$ -мерная мера Хаусдорфа. Из соотношения $\lambda(A) = \lambda(B_r) = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+\frac{d}{2})} r^d$ получаем классическое изопериметрическое неравенство

$$\lambda^{1-\frac{1}{d}}(A) \leq \kappa_d \mathcal{H}^{d-1}(\partial A),$$

где $\kappa_d = \frac{\left[\Gamma(1+\frac{d}{2}) \right]^{\frac{1}{d}}}{d\sqrt{\pi}}$. Из доказательства следует, что неравенства становятся равенствами в случае $A = \{\|x - x_0\| \leq r\}$. Таким образом,

шары имеют наименьшую поверхностную меру среди множеств фиксированной меры Лебега. \square

Пусть ν — некоторое вероятностное распределение. Энтропией вероятностного распределения $g \cdot \nu$ относительно ν называется величина $\text{Ent}_\nu(g) = \int g \log g \, d\nu$ (мы считаем, что функция $x \log x$ равна нулю в точке 0).

Следующий результат, доказанный М. Талаграном [15], связывает теорию оптимальной транспортировки с функциональными неравенствами.

Теорема 2. *Пусть $\mu = \gamma$ — стандартное гауссовское распределение. Предположим, что $\nu = g \cdot \gamma$ — другое вероятностное распределение. Тогда квадрат расстояния Канторовича между этими распределениями оценивается относительной энтропией g :*

$$\frac{1}{2}W^2(\gamma, g \cdot \gamma) \leq \int_{\mathbb{R}^d} g \cdot \log g \, d\gamma := \text{Ent}_\gamma(g).$$

Доказательство. Для простоты изложения предположим, что g и φ — гладкие функции (это бывает не всегда, но общий случай можно свести к этому). Рассмотрим формулу замены переменной:

$$e^{-\frac{x^2}{2}} = g(\nabla \varphi) e^{-\frac{|\nabla \varphi|^2}{2}} \det D^2 \varphi.$$

Прологарифмируем это соотношение:

$$-\frac{x^2}{2} = \log g(\nabla \varphi) - \frac{|\nabla \varphi|^2}{2} + \log \det D^2 \varphi.$$

Перепишем его в виде

$$\frac{1}{2}|x - \nabla \varphi|^2 = \langle x, x - \nabla \varphi \rangle + \log g(\nabla \varphi) + \log \det D^2 \varphi.$$

Проинтегрируем полученное неравенство по γ . Заметим, что $\nabla e^{-\frac{x^2}{2}} = -xe^{-\frac{x^2}{2}}$. Из формулы интегрирования по частям следует:

$$\int_{\mathbb{R}^d} \langle x, x - \nabla \varphi \rangle \, d\gamma = \int_{\mathbb{R}^d} (d - \text{Tr} D^2 \varphi) \, d\gamma$$

(напомним, что d — размерность). Следовательно,

$$\frac{1}{2} \int_{\mathbb{R}^d} |x - \nabla \varphi|^2 \, d\gamma + \int_{\mathbb{R}^d} (\text{Tr} D^2 \varphi - d - \log \det D^2 \varphi) \, d\gamma \leq \int_{\mathbb{R}^d} \log g(\nabla \varphi) \, d\gamma.$$

Заметим теперь, что

$$\mathrm{Tr} D^2\varphi - d - \log \det D^2\varphi \geq 0.$$

Действительно, если λ_i — собственные значения матрицы $D^2\varphi$, то

$$\mathrm{Tr} D^2\varphi - d - \log \det D^2\varphi = \sum_{i=1}^d \lambda_i - 1 - \log \det \lambda_i \geq 0$$

(в силу неотрицательности функции $x - 1 - \ln x$). Таким образом, получаем

$$\frac{1}{2} \int_{\mathbb{R}^d} |x - \nabla\varphi|^2 d\gamma \leq \int_{\mathbb{R}^d} \log g(\nabla\varphi) d\gamma = \int_{\mathbb{R}^d} g \log g d\gamma.$$

□

Теорема 3 (К. Мартон). *Если вероятностное распределение μ удовлетворяет неравенству Талаграна:*

$$W^2(\mu, g \cdot \mu) \leq C \int_{\mathbb{R}^d} g \cdot \log g d\mu,$$

то μ удовлетворяет неравенству гауссовой концентрации:

$$\mu(A_r) \geq 1 - 2e^{-\frac{r^2}{4C}}, \quad \mu(A) \geq \frac{1}{2}. \quad (4)$$

В частности, неравенству гауссовой концентрации удовлетворяет гауссовское распределение.

Доказательство. Положим $(A^r)^c = \mathbb{R}^d \setminus A^r$. Рассмотрим оптимальную транспортировку $\nabla\varphi$ вероятностного распределения $\mu_1 = \frac{1}{\mu(A)} \times I_A \cdot \mu$ в вероятностное распределение $\mu_2 = \frac{1}{\mu((A^r)^c)} I_{(A^r)^c} \cdot \mu$. В силу того, что расстояние между носителями μ_1, μ_2 превосходит r , имеем $W_2(\mu_1, \mu_2) \geq r$. В силу неравенства треугольника (напомним, что W_2 — расстояние):

$$r \leq W_2(\mu_1, \mu_2) \leq W_2(\mu_1, \mu) + W_2(\mu_2, \mu).$$

По неравенству Талаграна

$$r \leq \sqrt{2C\mathrm{Ent}_\mu \mu_1} + \sqrt{2C\mathrm{Ent}_\mu \mu_2}.$$

Так как $\text{Ent}_\mu \mu_1 = \log \frac{1}{\mu(A)}$, $\text{Ent}_\mu \mu_2 = \log \frac{1}{\mu((A^r)^c)}$, немедленно получаем

$$\frac{r^2}{4C} \leq \log \frac{1}{\mu(A)} + \log \frac{1}{\mu((A^r)^c)}.$$

Следовательно,

$$\mu((A^r)^c) \leq 2e^{-\frac{r^2}{4C}}.$$

Остается заметить, что $\mu((A^r)^c) = 1 - \mu(A^r)$. Теорема доказана. \square

Несложно проверить, что приведенные выше аргументы применимы к случаю распределения с плотностью e^{-V} , где $D^2V \geq K \cdot \text{Id}$, $K > 0$ (неравенство понимается в матричном смысле, эквивалентная формулировка: $\langle D^2V \cdot v, v \rangle \geq K$ для любого вектора $v \in \mathbb{R}^d$ единичной длины). В этом случае также получаем гауссовскую концентрацию. Те же самые аргументы работают для сферы или, более общим образом, для многообразия с ограниченным снизу тензором Риччи.

Напоследок кратко обсудим еще одно приложение транспортной задачи — оценку скорости сходимости к равновесному состоянию. Пусть V — равномерно выпуклый потенциал:

$$D^2V \geq K \cdot \text{Id}, K > 0.$$

Рассмотрим решение уравнения Фоккера—Планка:

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \text{div}(\nabla V \cdot \mu_t),$$

$\mu_t = \rho_t dx$ — вероятностное распределение. Оказывается, для двух решений этого уравнения выполнено неравенство

$$\frac{d}{dt} W_2^2(\mu_t, \nu_t) \leq -2K \cdot W_2^2(\mu_t, \nu_t)$$

(см. [16], пример 9.10). Это можно проверить, непосредственно про-дифференцировав расстояние Канторовича по параметру t . Очевидно, эта оценка дает экспоненциальную скорость сходимости μ_t к равновесному распределению:

$$W_2(\mu_t, \nu_t) \leq W_2(\mu_0, \nu_0) e^{-Kt}.$$

Приложения такого рода включают в себя широкий класс уравнений, являющихся градиентными потоками относительно метрики Канторовича. Подробнее об этом см. в [4].

Литература

1. Богачев В. И. Гауссовские меры. М.: Наука, 1997.
2. Зорич В. А. Математический анализ задач естествознания. М.: МЦНМО, 2008.
3. Канторович Л. В. О перемещении масс // ДАН СССР. 1942. Т. 37. С. 227–229.
4. Ambrosio L., Gigli N., Savaré G. Gradient flows in metric spaces and in the Wasserstein spaces of probability measures. Lectures in Math. ETH Zurich, 2008.
5. Brenier Y. Polar factorization and monotone rearrangement of vector valued functions // Comm. Pure Appl. Math. 1991. V. 44. P. 375–417.
6. Evans L. C. Partial differential equations and Monge-Kantorovich mass transfer. In «Current developments in mathematics». Cambridge, 1997; Boston: Int. Press, 1999. P. 65–126.
<http://math.berkeley.edu/~evans/>
7. Gromov M. Metric structure for Riemannian and non-Riemannian spaces. V. 152. Boston: Birkhäuser, 1998.
8. Milman V., Schechtman G. Asymptotic theory of finite dimensional normed vector spaces. Lect Notes in Math. V. 1200. Springer, 1986.
9. Ledoux M. The concentration of measure phenomenon. Mathematical Surveys and Monographs 89. Amer. Math. Soc., 2001.
10. Levy P. Problème concretes d'analyse fonctionnelle. Paris: Gauthier-Villars, 1951.
11. Lugosi G. Concentration of measures inequalities. Barcelona, 2009.
<http://www.econ.upf.edu/~lugosi/anu.pdf>
12. Marton K. A measure concentration inequality for contractive Markov chains // Geom. Func. Anal. 1997. V. 6. P. 556–571.
13. Milman E. On the role of Convexity in Isoperimetry, Spectral-Gap and Concentration // Invent. Math. 2009. V. 177. N. 1. P. 1–43.

14. *Monge G.* Mémoire sur la théorie des déblais et de remblais. Histoire de l'Acad'emic Royale des science, 1781.
15. *Talagrand M.* Transportation cost for Gaussian and other product measures // Geom. Funct. Anal. 1996. V. 6. P. 587–600.
16. *Villani C.* Topics in Optimal Transportation. Amer. Math. Soc. Providence. Rhode Island, 2003.
<http://math.univ-lyon1.fr/~villani/>

Модели случайных графов и их применения¹

В этом приложении дается обзор основных современных направлений в теории случайных графов. Делается акцент на связь моделей случайного графа с транспортной проблематикой.

1. Введение

Теория графов играет огромную роль в фундаментальной и прикладной математике. Ей посвящены сотни монографий и тысячи – если не десятки тысяч – статей. Разумеется, мы не можем ставить перед собою цель дать на этих страницах сколь-нибудь подробное изложение теории графов. Нас будет интересовать лишь одно направление, которое с каждым годом становится все более актуальным. В рамках этого направления графы изучаются с вероятностной точки зрения. Типичная постановка вопроса (говоря не совсем строго) такова: *велика ли вероятность того, что граф обладает данным свойством?* Вопрос исключительно важный, и мы в этом не раз убедимся ниже. Правда, в нем ни слова не сказано о том, как именно мы понимаем термин *вероятность*. Всякий человек, имеющий представление об аксиоматике Колмогорова, хорошо знает, что можно вложить множество разных смыслов в этот термин. И его можно действительно определять по-разному. В зависимости от определения получится та или иная модель *случайного графа*. С чисто математических позиций любая такая модель имеет право на существование. Однако для приложений – в том числе приложений к транспортной проблематике – некоторые из этих моделей более интересны, некоторые – менее. Соответственно ниже мы расскажем о двух классах моделей, каждый из которых за десятилетия, прошедшие с момента

¹Работа выполнена при финансовой поддержке гранта РФФИ 09-01-00294, гранта Президента РФ МД-8390.2010.1, гранта поддержки ведущих научных школ НШ-8784.2010.1.

своего появления, зарекомендовал себя плодотворным как в рамках «чистой» математики, так и в рамках ее разнообразных приложений, среди которых надежность транспортной сети, рост Интернета и других социальных и биологических сетей, теория алгоритмов и пр.

Не претендуя на полноту изложения (это было бы нелепо, т.к. и здесь наука разрослась безгранично), мы постараемся выделить лишь самые основные и принципиальные моменты теории случайных графов.

2. Модель Эрдеша—Ренъи

Этот раздел мы посвятим описанию модели случайного графа, которая возникла исторически первой. На рубеже 50-х и 60-х годов XX века эту модель предложили классики современной комбинаторики и теории вероятностей П. Эрдеш и А. Ренни (см. [1–3]). Отметим, что Эрдеш – это, пожалуй, одна из самых ярких фигур в математике XX века. Ему принадлежат сотни статей и задач, которые оказали огромное влияние на развитие многих математических дисциплин. Ренни также сыграл значительную роль в формировании венгерской вероятностной школы, и его именем назван математический институт в Будапеште.

2.1. Формальное описание модели

Пусть дано множество $V_n = \{1, \dots, n\}$, элементы которого мы назовем *вершинами*. Именно на этом множестве мы будем «строить» случайный граф. Понятно, что случайным будет множество ребер графа. Мы не хотим сейчас рассматривать графы с кратными ребрами (мультиграфы), графы с петлями (псевдографы) и ориентированные графы (орграфы). Поэтому мы считаем, что потенциальных ребер у графа не больше, чем C_n^2 штук. Будем соединять любые две вершины i и j ребром с некоторой вероятностью $p \in [0, 1]$ независимо от всех остальных $C_n^2 - 1$ пар вершин. Иными словами, ребра появляются в соответствии со стандартной схемой Бернулли, в которой C_n^2 испытаний и «вероятность успеха» p . Обозначим через E случайное множество ребер, которое возникает в результате реализации такой схемы. Положим $G = (V_n, E)$. Это и есть случайный граф в модели

Эрдеша—Ренни. Напомним, что E также обозначает математическое ожидание (из контекста будет ясно, какое E имеется в виду в том или ином случае).

Если записывать приведенное только что определение в формате аксиоматики Колмогорова, то мы имеем вероятностное пространство

$$G(n, p) = (\Omega_n, \mathcal{F}_n, P_{n,p}),$$

в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n,p}(G) = p^{|E|}(1-p)^{C_n^2 - |E|}. \quad (1)$$

Здесь через $|A|$ обозначена мощность множества A (количество элементов), а 2^A — это совокупность всех подмножеств множества A .

Элемент сигма-алгебры \mathcal{F}_n — это набор графов. Если нам хочется найти вероятность, с которой граф на n вершинах обладает данным свойством A , то мы просто берем множество $\mathcal{A} \in \mathcal{F}_n$, состоящее из всех графов, для которых выполнено свойство A , и вычисляем

$$P_{n,p}(\mathcal{A}) = \sum_{G \in \mathcal{A}} P_{n,p}(G).$$

Таким образом, вероятность того, например, что случайный граф связан, — это величина, равная сумме вероятностей всех связных графов (на фиксированном множестве вершин). Казалось бы, все совсем просто и мы вряд ли имеем шансы наткнуться здесь на нечто особенно интересное. Однако дело обстоит прямо противоположным образом: специфика вероятностных методов, которые эффективно работают в задачах о случайных графах, позволит нам пронаблюдать весьма нетривиальные и, главное, неожиданные явления, которые возникают даже в этой простой модели и которые влекут приятные следствия для приложений.

Прежде чем двигаться дальше, сделаем еще ряд полезных замечаний. Во-первых, если $p = \frac{1}{2}$, то, как видно из формулы (1), вероятность любого графа равна $2^{-C_n^2}$. Иными словами, в этом специальном случае все графы равновероятны и всякое утверждение о вероятности какого-либо свойства — это, по сути, утверждение о доле графов данным свойством обладающих.

В действительности, мы не только не обязаны предполагать, что $p = \frac{1}{2}$ (хотя и этот случай очень важен), мы даже можем считать,

что с ростом величины n (числа вершин) вероятность p возникновения ребра изменяется. Иначе говоря, $p = p(n)$ – любая функция, значения которой заключены между нулем и единицей. Как правило, в науке о случайных графах важны даже не сами вероятности событий, но их предельные значения. Почему это так, мы скоро увидим.

Скажем, наконец, что свойство выполнено *почти всегда*, если его вероятность стремится к единице при $n \rightarrow \infty$.

2.2. Транспортная интерпретация модели

Представим себе, что в некоторой стране есть 10 городов, которые *попарно* соединены дорогами. Это довольно сильное предположение, но пока сохраним его. Допустим, каждая из дорог за определенный срок изнашивается (т.е. становится непроезжей) с известной вероятностью q . При этом износ данной дороги никак не зависит от совокупного износа остальных дорог. Спрашивается: какова максимальная вероятность q , при которой с вероятностью больше $1/2$ не исчезнет возможность перемещения между любыми двумя городами? По существу, это вопрос о *надежности* транспортной сети: чем выше искомая вероятность q , тем, разумеется, сеть надежнее.

Нетрудно видеть, что вопрос о надежности сети – это в свою очередь вопрос о *связности* случайного графа. В самом деле, составим каждому городу вершину $i \in V_{10}$. Тогда «дорога» между «городами» i и j – это ребро. Износ дороги – это исчезновение ребра. Значит, утверждение «дорога изнашивается с вероятностью q » равносильно утверждению «ребро появляется с вероятностью $p = 1 - q$ ». Таким образом, нас интересует, какова *минимальная* вероятность p , при которой в модели Эрдеша–Ренни $G(n, p)$ вероятность связности графа больше половины (граф, скорее, связан, чем несвязан).

Понятно, что если мы заменим число 10 другим числом, то соответствующее минимальное p может измениться. Этим и обусловлено наше желание рассматривать не только постоянные p , но и нетривиальные функции $p = p(n)$.

В разделе 2.4 мы обсудим ряд строгих утверждений, касающихся сформулированного выше вопроса. Однако есть у нас и более срочное дело: все-таки предположение о том, что города связаны дорогами *попарно*, черезсчур сильное, и в следующем разделе мы приведем модификацию модели Эрдеша–Ренни, в рамках которой это пред-

положение можно будет адекватно ослабить.

2.3. Обобщения модели Эрдеша—Ренни

Пусть по-прежнему $V_n = \{1, \dots, n\}$. Однако теперь вероятность ребра между вершинами i и j мы обозначим через p_{ij} . Иными словами, мы, как и раньше, проводим ребра независимо друг от друга, но с *разными* вероятностями. В формате аксиоматики Колмогорова мы получаем вероятностное пространство

$$G(n, p_{ij}) = (\Omega_n, \mathcal{F}_n, P_{n, p_{ij}}),$$

в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n, p_{ij}}(G) = \prod_{(i,j) \in E} p_{ij} \cdot \prod_{(i,j) \notin E} (1-p_{ij}).$$

Важный частный случай описанного пространства получается, коль скоро мы фиксируем некоторый граф $H_n = (V_n, E_n)$ и полагаем

$$p_{ij} = \begin{cases} p, & (i, j) \in E_n, \\ 0, & (i, j) \notin E_n. \end{cases}$$

Иначе говоря, ребра графа H_n возникают в случайному графе независимо друг от друга с одной и той же вероятностью $p = p(n) \in [0, 1]$, а ребра, которых в графе H_n нет, не возникают в случайному графе вовсе. Этот вариант модели принято обозначать $G(H_n, p)$. В ней

$$P_{n, p_{ij}}(G) = p^{|E|} (1-p)^{|E_n| - |E|}.$$

Ясно, что модель $G(H_n, p)$ есть та самая модель, которая вполне адекватна вопросу о надежности транспортной сети. На сей раз мы не обязаны предполагать, что города попарно соединены дорогами, мы можем с самого начала зафиксировать граф дорог H_n и следить за износом его ребер.

2.4. Некоторые математические результаты о надежности сети

Прежде всего справедлива следующая теорема.

Теорема 1. Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c \ln n}{n}$. Если $c > 1$, то почти всегда случайный граф связен. Если $c < 1$, то почти всегда случайный граф не является связным.

Этот довольно простой с точки зрения доказательства факт мы обоснуем в разделе 2.5. Однако при всей своей формальной простоте теорема 1 несет весьма содержательную и в каком-то смысле неожиданную информацию. Действительно, вернемся к вопросу о надежности сети. Пусть число n городов, попарно соединенных дорогами, растет. Тогда, разумеется, величина $p = \frac{c \ln n}{n}$ довольно быстро стремится к нулю. Тем не менее теорема 1 утверждает, что вероятность сохранения связности графа при уничтожении его ребер с вероятностью $q = 1 - p$ стремится к единице. Грубо говоря, если городов 1000, то мы можем позволить дорогам разрушаться с вероятностью ≈ 0.993 , так что в результате с вероятностью, близкой к единице, перемещение между любыми двумя городами останется возможным. Поначалу это кажется противоречащим интуиции, но, по здравому размышлению, становится понятно, в чем здесь смысл. Дорог у нас $C_n^2 = \Theta(n^2)$, вероятность износа дороги равна $1 - \Theta(\ln n/n)$. Значит, ожидаемое количество неизношенных дорог имеет порядок $n \ln n$. Этого хватает для сохранения связности.

При определенной аккуратности в выкладках, которые мы частично проведем в разделе 2.5, можно доказать, например, такой факт.

Теорема 1'. Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c \ln n}{n}$. Если $c > 3$, то при $n > 100$

$$P_{n,p}(G \text{ связен}) \geq 1 - \frac{1}{n}.$$

Этот результат совсем замечателен своей конкретностью. Получается, что при той же тысяче городов и вероятности износа дороги $1 - \frac{3 \ln 1000}{1000} \approx 0.98$ вероятность сохранения связности не меньше, чем 0.999!

Теорема 1 любопытна еще и тем, что в ней наблюдается резкий скачок от «почти всегда связности» к «почти всегда несвязности». Функция $p(n) = \frac{\ln n}{n}$ служит своего рода рубежом, преодоление которого означает переход от ненадежности к надежности. Такой переход принято называть *фазовым*, а соответствующую функцию $p(n)$ — *пороговой*.

Следующая теорема содержит в себе еще более глубокую информацию о природе связности-надежности. Она была доказана самими Эрдешем и Ренни (см. [1–3]).

Теорема 2. *Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c}{n}$. Если $c < 1$, то найдется такая константа $\beta = \beta(c)$, что почти всегда размер каждой связной компоненты случайного графа не превосходит $\beta \ln n$. Если же $c > 1$, то найдется такая константа $\gamma = \gamma(c)$, что почти всегда в случайном графе есть ровно одна компонента размера $\geq \gamma n$.*

И снова мы имеем фазовый переход – резкое изменение свойств случайного графа при преодолении некоторого порога. В данном случае порогом служит функция $p = \frac{1}{n}$. Оказывается, что если вероятность ребра в $c > 1$ раз «ниже» порога, то все связные компоненты графа, скорее всего, крошечные – имеющие логарифмический от общего числа вершин размер; если же вероятность ребра в $c > 1$ раз «выше» порога, то, скорее всего, найдется компонента с числом вершин порядка n . Такая компонента называется *гигантской*.

Теорема 2 допускает различные уточнения. Например, можно утверждать, что при $c > 1$, помимо единственной гигантской компоненты, в случайном графе ничего сколь-нибудь крупного почти никогда не возникает: все остальные компоненты снова логарифмические. Можно еще аккуратнее описывать размер гигантской компоненты. В действительности, верно не только неравенство $\geq \gamma n$, но и асимптотика $\sim \gamma n$. В. Е. Степанов доказал, что при $p = \frac{c}{n}$, $c > 1$, размер гигантской компоненты асимптотически нормален (см. [4–6]).

В целом изменение свойств случайного графа при изменении вероятности ребра p принято трактовать как эволюцию графа. Нам кажется, что несколько правильнее говорить о своего рода истории мира. Сначала (при $p \ll \frac{1}{n}$) имеет место «феодализм» – весь граф поделен на несвязанные между собой логарифмические кусочки. Затем (при $p \gg \frac{1}{n}$) возникает «империя» – гигантская компонента. Наконец, при $p \gg \frac{\ln n}{n}$ империя уничтожает «окраины» и добивается мирового господства – связности.

В терминах надежности смысл теоремы 2 также очевиден: можно еще в $\ln n$ раз уменьшить вероятность сохранности отдельной дороги, и если не вся страна, то значительная ее часть окажется консолидированной, т.е. не лишенной инфраструктуры – возможности сообщения между любыми двумя городами.

Глубокий интерес представляет, конечно, устройство мира «внутри» фазовых переходов, т.е. при $p \sim \frac{1}{n}$ и при $p \sim \frac{\ln n}{n}$. В первом случае все совсем сложно, и мы отсылаем читателя к книгам [7–9]. Во втором случае можно сформулировать, например, следующий понятный результат.

Теорема 3. Пусть $p = \frac{\ln n + c + o(1)}{n}$. Тогда $P_{n,p}(G \text{ связен}) \rightarrow e^{-e^{-c}}$. В частности, при $p = \frac{\ln n}{n}$ вероятность стремится к e^{-1} .

Здесь уже речь не идет о «почти всегда связности» или «почти всегда несвязности». Здесь асимптотическая вероятность связности есть, но она лежит в строгих пределах от нуля до единицы.

Все, о чем мы говорили до сих пор, касалось модели $G(n, p)$. Естественно, модель $G(H_n, p)$, будучи более адекватной реальности, является и более сложной для изучения. Главный результат относительно этой модели принадлежит Г. А. Маргулису (см. [11]).

Теорема 4. Пусть $\{H_n\}$ – последовательность графов, реберная связность которых стремится к бесконечности при $n \rightarrow \infty$. Тогда существует пороговая функция p для свойства связности случайного графа в модели $G(H_n, p)$.

Теорема 4 нетривиальна, и ее доказательство (а также массу ссылок на близкие результаты) можно найти в книге [8]. Разумеется, поиск пороговой функции, существование которой доказывается в теореме 4, – это всякий раз сложная задача, связанная на специфику графов из последовательности $\{H_n\}$.

Практический смысл теоремы 4 банален: надо строить дороги так, чтобы связность получающегося графа неуклонно росла. К сожалению, в России положение, как правило, противоположное. Даже в Москве есть много улиц, перекрытие которых означает фактическую потерю связности. Например, таковы улицы, проходящие под железнными дорогами: они крайне редки и служат единственными лазейками с одной стороны полотна на другую.

В следующем разделе мы докажем теорему 1, а в разделе 2.6 мы обсудим основные идеи доказательства теоремы 2. Отметим, что дополнительную информацию о поведении случайных графов в модели Эрдеша–Рены можно найти в [7–10].

2.5. Доказательство теоремы 1

Сперва обсудим случай $c > 1$.

Введем случайную величину на пространстве $G(n, p)$:

$$X_n = X_n(G) = \begin{cases} 0, & \text{если } G \text{ связен,} \\ k, & \text{если у } G \text{ ровно } k \text{ компонент.} \end{cases}$$

Таким образом, X_n принимает неотрицательные целые значения, причем $X_n \neq 1$. Нам нужно показать, что $P_{n,p}(X_n = 0) \rightarrow 1$ при $n \rightarrow \infty$. Это равносильно асимптотике $P_{n,p}(X_n \geq 1) \rightarrow 0$. По неравенству Чебышёва: $P_{n,p}(X_n \geq 1) \leq EX_n$, и нам остается обосновать стремление к нулю математического ожидания.

Представим X_n в виде суммы

$$X_n = X_{n,1} + \dots + X_{n,n-1},$$

где $X_{n,k} = X_{n,k}(G)$ – число k -вершинных компонент графа G . Занумеруем все k -элементные подмножества множества вершин V_n случайного графа в некотором (произвольном) порядке: $K_1, \dots, K_{C_n^k}$. Тогда в свою очередь

$$X_{n,k} = X_{n,k,1} + \dots + X_{n,k,C_n^k},$$

коль скоро

$$X_{n,k,i} = X_{n,k,i}(G) = \begin{cases} 1, & \text{если } K_i \text{ образует компоненту в } G, \\ 0, & \text{иначе.} \end{cases}$$

В итоге

$$EX_n = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} EX_{n,k,i}.$$

Очевидно,

$$\begin{aligned} EX_{n,k,i} &= P_{n,p}(K_i \text{ образует компоненту в } G) \leq \\ &\leq P_{n,p}(\text{из } K_i \text{ в } V_n \setminus K_i \text{ нет ребер в } G). \end{aligned}$$

Получая последнее неравенство, мы просто пренебрегли условием связности той части графа G , которая «сидит» на множестве вершин

K_i (такую часть принято называть *индуцированным подграфом* и обозначать $G|_{K_i}$). Далее,

$$P_{n,p}(\text{из } K_i \text{ в } V \setminus K_i \text{ нет ребер в } G) = (1-p)^{k(n-k)},$$

и, значит,

$$EX_n \leq \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} (1-p)^{k(n-k)} = \sum_{k=1}^{n-1} C_n^k (1-p)^{k(n-k)}.$$

Последняя сумма симметрична в том смысле, что ее слагаемые при k и $n - k$ равны. Рассмотрим $k = 1$:

$$n(1-p)^{n-1} \leq ne^{-p(n-1)} = ne^{-\frac{c(\ln n)(n-1)}{n}} = n \left(\frac{1}{n}\right)^{c(1+o(1))} = o(1),$$

поскольку $c > 1$.

Оставшаяся часть рассуждения состоит в доказательстве того, что слагаемые с $k > 1$ и $k < n - 1$ пренебрежимо малы по сравнению с первым слагаемым. Соответствующую выкладку мы пропустим. Если же поверить в ее справедливость, то получится, что вся сумма доминируется первым и последним слагаемыми, а стало быть, и она стремится к нулю.

Теорема 1 для случая $c > 1$ доказана.

Теперь рассмотрим случай $c < 1$. Обозначим через X_n количество изолированных вершин в случайному графе. Запишем

$$X_n = X_{n,1} + \dots + X_{n,n},$$

где

$$X_{n,k} = X_{n,k}(G) = \begin{cases} 1, & \text{если вершина } k \in V_n \text{ изолированная в } G, \\ 0, & \text{иначе.} \end{cases}$$

Тогда

$$EX_n = EX_{n,1} + \dots + EX_{n,n}.$$

В свою очередь

$$EX_{n,k} = P_{n,p}(k \text{ изолированная в } G) = (1-p)^{n-1}.$$

Таким образом,

$$\begin{aligned} EX_n &= n(1-p)^{n-1} = n(1-p)^n(1+o(1)) = \\ &= (1+o(1))ne^{-c \ln n} = (1+o(1))n^{1-c}. \end{aligned}$$

Заметим, что ввиду неравенства $c < 1$ выполнено $EX_n \rightarrow \infty$.
Посчитаем дисперсию случайной величины X_n :

$$\begin{aligned} DX_n &= EX_n^2 - (EX_n)^2 = E(X_{n,1} + \dots + X_{n,n})^2 - (EX_n)^2 = \\ &= EX_{n,1}^2 + \dots + EX_{n,n}^2 + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2 = \\ &= EX_{n,1} + \dots + EX_{n,n} + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2 = \\ &= EX_n + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2. \end{aligned}$$

Далее,

$$\begin{aligned} EX_{n,i}X_{n,j} &= P_{n,p}(i \text{ и } j \text{ изолированы в } G) = \\ &= (1-p)^{2n-1} = (1+o(1))(1-p)^{2n}, \end{aligned}$$

т.е.

$$\begin{aligned} \sum_{i \neq j} EX_{n,i}X_{n,j} &= n(n-1)(1+o(1))(1-p)^{2n} = \\ &= (1+o(1))n^{2-2c} = (1+o(1))(EX_n)^2. \end{aligned}$$

В итоге

$$DX_n = EX_n + (1+o(1))(EX_n)^2 - (EX_n)^2 = o((EX_n)^2).$$

По неравенству Чебышёва

$$\begin{aligned} P_{n,p}(G \text{ связен}) &\leq P_{n,p}(X_n = 0) = P_{n,p}(X_n \leq 0) = P_{n,p}(-X_n \geq 0) = \\ &= P_{n,p}(EX_n - X_n \geq EX_n) \leq \frac{DX_n}{(EX_n)^2} = o(1), \end{aligned}$$

и вторая часть теоремы доказана.

2.6. Идеи доказательства теоремы 2

Метод, о котором мы будем здесь говорить, восходит к Р. Карпу (см. [12]), и в таком виде он описан в книге [9]. Мы лишь перечислим ниже основные шаги рассуждения.

2.6.1. Простейший ветвящийся процесс

Пусть Z_1, \dots, Z_t, \dots – независимые пуассоновские величины с одним и тем же средним λ . Положим

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Представлять себе описанный только что процесс можно так. В начальный момент времени есть одна частица. Затем она приносит Z_1 потомков и умирает. Заметим, что она может умереть, даже не принеся потомства, т.к. величина Z_1 равна нулю с положительной вероятностью. На следующем шаге все повторяется: какая-то частица (порядок роли не играет) порождает Z_2 новых частиц, а сама гибнет. И так далее. Популяция может выродиться, а может и жить вечно. Хорошо известно, что имеют место следующие результаты.

Теорема 5. *Пусть $\lambda \leq 1$. Тогда с вероятностью 1 процесс Y_t вырождается, т.е. $P(\exists t : Y_t \leq 0) = 1$.*

Теорема 6. *Пусть $\lambda > 1$. Пусть $\gamma \in (0, 1)$ – единственное решение уравнения $1 - \gamma = e^{-\lambda\gamma}$. Тогда процесс Y_t вырождается с вероятностью $1 - \gamma$, т.е. $P(\exists t : Y_t \leq 0) = 1 - \gamma$.*

Доказательства теорем 5 и 6 можно найти, например, в [9]. Впрочем, это стандартные факты теории ветвящихся процессов (см. [13]). Забегая вперед, скажем, что величина γ в теореме 6 и одноименная величина в теореме 2 суть одно и то же. Вероятность вырождения процесса Y_t и размер гигантской компоненты напрямую связаны.

2.6.2. Ветвящийся процесс на случайном графе

Пусть дан граф $G = (V, E)$: конкретный граф, не случайный. Зафиксируем какую-нибудь его вершину $v_1 \in V$. Назовем ее «живой», а все остальные вершины – «нейтральными». Выберем среди нейтральных вершин всех соседей вершины v_1 . После этого объявим

вершину v_1 «мертвой», ее соседей – живыми, а все остальные вершины – нейтральными. Снова зафиксируем какую-нибудь живую вершину v_2 . Выберем всех ее соседей среди *нейтральных*. Вершину v_2 отправим в царство мертвых, а в живых останутся все, кто был жив, кроме v_2 , и новорожденные «потомки» вершины v_2 . Продолжая этот ветвящийся процесс, мы в конце концов получим кладбище вершин и ни одной живой вершины. Кладбище будет компонентой, содержащей v_1 .

Обозначим число живых вершин в момент времени t через Y_t , число нейтральных вершин – через N_t , а число потомков очередной живой вершины, отправляющейся в последний путь, – через Z_t . Тогда, очевидно,

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Разумеется, все введенные величины зависят от графа G и от последовательности выбираемых вершин v_1, \dots . Если граф G посчитать случайным, то при любом выборе вершин v_1, \dots получатся случайные величины Y_t, N_t, Z_t на пространстве $G(n, p)$. На самом деле, ясно, конечно, что распределения этих величин не зависят от v_1, \dots ; поэтому мы нигде зависимость от вершин и не указываем.

Сразу понятно, что Z_t не является пуассоновской. Тем не менее она похожа на пуассоновскую, а именно, имеет биномиальное распределение с «числом испытаний» N_t и вероятностью «успеха» p . Правда, число испытаний само случайно. По счастью, это не проблема. Удается доказать, что Y_t имеет вид

$$Y_t = \xi_t + 1 - t, \quad \xi_t \sim \text{Binom}(n - 1, 1 - (1 - p)^t).$$

Подробности можно найти в [9], и мы их здесь не касаемся.

Как известно, биномиальное распределение сходится к пуассоновскому, коль скоро вероятность успеха обратно пропорциональна числу испытаний. Нечто подобное имеет место и у нас ($p = c/n$), и ровно на этом мы сыграем в итоге.

2.6.3. Случай $c < 1$

Положим $t_0 = [\beta \ln n]$, где $\beta = \beta(c)$ – константа, которую мы подберем позднее. Нам хочется доказать, что с большой вероятностью каждая из компонент случайного графа имеет размер не больше t_0 . Но размер компоненты – это момент вырождения процесса Y_t

на случайном графе. Значит, интересующее нас утверждение можно записать в следующем виде:

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \rightarrow 0, \quad n \rightarrow \infty.$$

Поскольку

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \leq n P_{n,p}(Y_{t_0} > 0),$$

достаточно найти такое β , при котором

$$P_{n,p}(Y_{t_0} > 0) = o\left(\frac{1}{n}\right).$$

Дальнейшие выкладки будут слегка некурутными, но при желании их можно сделать безукоризненно строгими. Мы же хотим максимально прояснить основную суть подхода. Итак,

$$\begin{aligned} P_{n,p}(Y_{t_0} > 0) &= P_{n,p}(\xi_{t_0} \geq t_0) \approx P_{n,p}(\text{Binom } (n, 1 - (1-p)^{t_0}) \geq t_0) \approx \\ &\quad (\text{с учетом асимптотики } 1 - (1-p)^{t_0} \sim pt_0) \\ &\approx P_{n,p}(\text{Binom } (n, pt_0) \geq t_0) \approx \\ &\quad (\text{с учетом центральной предельной теоремы}) \end{aligned}$$

$$\approx \int_{\frac{t_0 - np t_0}{\sqrt{np t_0 (1 - p t_0)}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Поскольку $c < 1$, нижний предел интегрирования имеет порядок $\sqrt{t_0}$. Значит, весь интеграл не превосходит величины $e^{-\delta t_0}$. Выберем β таким, чтобы $e^{-\delta t_0}$ оказалось меньше, чем $e^{-2 \ln n} = \frac{1}{n^2}$, и в случае $c < 1$ теорема доказана.

2.6.4. Случай $c > 1$

Этот случай гораздо сложнее предыдущего. Здесь ветвящийся процесс на графе нужно «запускать» не один раз, а несколько. Только так удается доказать, что почти наверняка хотя бы в одном запуске возникнет гигантская компонента. Подробности можно найти

в [9], мы же лишь поясним, откуда в текущей ситуации появляется константа γ из формулировки теоремы и почему она совпадает с одноименной константой из теоремы 6.

Грубо говоря, идея следующая. Нам хочется доказать, что есть гигантская компонента. Тогда, как следствие, нам нужно, чтобы ветвящийся процесс на графе не вырождался даже при $t \approx \gamma n$. Иными словами, необходимо, чтобы

$$P_{n,p}(Y_t \leq 0) \rightarrow 0, \quad t \approx \gamma n, \quad n \rightarrow \infty.$$

У нас $p = \frac{c}{n}$. Значит, при $t \sim \alpha n$ выполнено

$$1 - (1 - p)^t \sim 1 - e^{-pt} \sim 1 - e^{-c\alpha}.$$

Применим центральную предельную теорему к

$$P_{n,p}(Y_t \leq 0) \approx P_{n,p}(\text{Binom}(n, 1 - e^{-c\alpha}) \leq \alpha n).$$

Интегрирование пойдет от минус бесконечности до

$$\frac{\alpha n - n(1 - e^{-c\alpha})}{\sqrt{n(1 - e^{-c\alpha})e^{-c\alpha}}}.$$

Если $\alpha < 1 - e^{-c\alpha}$, то мы получим искомое стремление вероятности к нулю. Если $\alpha > 1 - e^{-c\alpha}$, то вероятность, напротив, будет стремиться к единице. Таким образом, критическое значение α , вплоть до которого есть именно стремление к нулю, – это решение уравнения $\alpha = 1 - e^{-c\alpha}$ или, что равносильно, $1 - \alpha = e^{-c\alpha}$. А это и есть уравнение из теоремы 6, коль скоро λ мы заменяем на c .

3. Модели Барабаши–Альберт

В этом разделе мы поговорим о самых современных моделях случайных графов, которые призваны описывать рост различных сетей – социальных, биологических, транспортных. Но в первую очередь речь пойдет об интернете. В 90-е годы XX века, когда Интернет только зарождался, исследователи уже задались вопросом, каким законам подчиняется рост интернета и какова наиболее адекватная модель для описания свойств этой сети. Одними из первых здесь были А.-Л. Барабаши и Р. Альберт. Они нашли ряд важных эмпирических закономерностей в поведении Интернета и на их основе

придумали модель, которую впоследствии по-разному формализовывали многие авторы. Соответственно мы построим наше изложение следующим образом. В разделе 3.1 мы расскажем о результатах Барабаши—Альберт. В разделе 3.2 мы опишем модель Б. Боллобаша и О. Риордана, которая весьма неплохо ложится на статистики Барабаши—Альберт. В разделе 3.3 мы обсудим возможные уточнения модели Боллобаша—Риордана.

3.1. Наблюдения Барабаши—Альберт

В своих работах [14–16] Барабаши и Альберт, а также Х. Джонг описали те статистики Интернета, которые легли в основу науки о росте этой сети – науки, имеющей глубокие приложения как в собственно интернетской проблематике, так и в многочисленных близких дисциплинах. В действительности, большинство реальных сетей (социальные, биологические, транспортные и пр.) имеют похожую «топологию».

Итак, сперва договоримся о том, что мы понимаем под сетью Интернет. Это так называемый *веб-граф*, вершины которого суть какие-либо конкретные структурные единицы в Интернете: речь может идти о страницах, сайтах, хостах, владельцах и пр. Для определенности будем считать, что вершинами веб-графа служат именно сайты. Ребрами же мы будем соединять те вершины, между которыми имеются ссылки. При этом разумно проводить столько ребер между двумя вершинами, сколько есть ссылок между соответствующими сайтами. Более того, ребра естественно считать направленными. Таким образом, веб-граф ориентирован и он может иметь кратные ребра, петли и даже кратные петли (ссылки вполне могут идти с одной страницы данного сайта на другую его страницу). Это такой «псевдомультиорграф». Сразу понятно, что для подобного «зверя» модель Эрдеша—Рены вряд ли подходит.

Теперь мы готовы перечислить самые основные моменты исследования Барабаши—Альберт. По существу, этих моментов всего три. Во-первых, веб-граф – это весьма разреженный граф. У него на t вершинах примерно kt ребер, где $k \geq 1$ – некоторая константа. Для сравнения, у полного графа на t вершинах $C_t^2 = \Theta(t^2)$ ребер. Однако – и это, во-вторых, – диаметр веб-графа исключительно скромен. В 1999 году он имел величину 5–7 (см. [16]). Это хорошо всем известное свойство любой социальной сети, которое принято в обыденной

речи характеризовать выражением «мир тесен». Например, говорят о том, что любые два человека в мире «знакомы через 5–6 рукопожатий». Точно так же и сайты: «кликая» по ссылкам, можно с любого сайта на любой другой перейти за 5–7 нажатий клавиши компьютерной мыши. Конечно, тут есть важная оговорка. Некоторые из двух появившихся сайтов могут не быть связаны с внешним по отношению к ним миром. Несколько правильнее сказать, что в веб-графе есть гигантская компонента, и уже ее диаметр невелик. Таким образом, веб-граф очень специфичен: будучи разреженным, он, тем не менее, в известном смысле тесен.

В-третьих, у веб-графа весьма характерное распределение степеней вершин. Эмпирическая вероятность того, что вершина веб-графа имеет степень d , оценивается как c/d^λ , где $\lambda \approx 2.1$, а c – нормирующий множитель, вычисляемый из условия «сумма вероятностей равна 1». Этот любопытный факт роднит Интернет с очень многими реальными сетями – биологическими, социальными, транспортными. Все они подчиняются степенному закону, только у каждой из них свой показатель λ .

Ввиду перечисленных наблюдений не остается никаких сомнений в том, что модель Эрдеша–Ренъи не применима для описания роста Интернета и подобных сетей. Если подбором вероятности p еще можно добиться разреженности и тесноты (хотя и не с теми параметрами), то степенной закон совсем уж не имеет отношения к схеме Бернулли, в рамках которой появляются ребра обычного случайного графа. В модели $G(n, p)$ степень каждой вершины случайного графа биномиальна с параметрами $n - 1$ и p , и при тех p , которые маломальски гарантируют разреженность (т.е. при $p = \Theta(1/n)$), указанное биномиальное распределение аппроксимируется пуассоновским, а вовсе не степенным.

Барабаши и Альберт предложили очень разумный взгляд на процесс формирования Интернета. Давайте считать, сказали они, что в каждый момент времени появляется новый сайт, и этот сайт ставит фиксированное количество ссылок на своих предшественников. На кого он предпочтет сослаться? Наверное, на тех, кто и так уже популярен. Можно допустить, что вероятность, с которой новый сайт поставит ссылку на один из прежних сайтов, пропорциональна числу уже имевшихся на тот сайт ссылок.

Модели случайных графов, основанные на описанной идеи, называются моделями *предпочтительного присоединения*. В своих рабо-

таких Барабаши и Альберт никак не конкретизировали, какую именно из этих моделей они предлагают рассматривать. А эти модели исключительно разнородны по своим свойствам. Ведь можно ставить ссылки независимо друг от друга, а можно еще и зависимости между разными ссылками с одного сайта учитывать. В итоге удается доказать даже такой забавный факт.

Теорема 7. *Пусть $f(n)$, $n \geq 2$, — произвольная целочисленная функция, такая, что $f(2) = 0$, $f(n) \leq f(n+1) \leq f(n) + 1$ для всех $n \geq 2$ и $f(n) \rightarrow \infty$ при $n \rightarrow \infty$. Тогда существует такая модель типа Барабаши—Альберт, что в ней с вероятностью, стремящейся к единице при $n \rightarrow \infty$, случайный граф содержит в точности $f(n)$ треугольников.*

Одну из наиболее правильных спецификаций модели Барабаши—Альберт предложили в начале двухтысячных годов Б. Боллобаш и О. Риордан. В следующем разделе мы ее обсудим.

3.2. Модель Боллобаша—Риордана

Наиболее полно эта модель описана в книге [8] и обзоре [17]. Также имеется малодоступная книга [18]. Мы представим здесь две основных и, по сути, совпадающих модификации этой модели. В одной дается динамическое, а в другой статическое описание случайности. Интуитивно более понятна динамическая модификация, с нее и начнем.

3.2.1. Динамическая модификация

Построим последовательность (случайных) графов $\{G_1^n\}$, в которой у графа с номером n число вершин и ребер равно n . Затем сделаем из нее последовательность $\{G_k^n\}$, в которой у графа с номером n число вершин равно n , а число ребер равно kn , $k \in \mathbb{N}$.

Итак, пусть $G_1^1 = (\{1\}, \{(1, 1)\})$, т.е. в начальный момент времени есть одна вершина и одна петля. Пусть теперь граф G_1^{n-1} уже построен. У него вершины образуют множество $\{1, \dots, n-1\}$, а ребер у него тоже $n-1$ штука. Добавим вершину n и ребро (n, i) , у которого $i \in \{1, \dots, n\}$. Ребро (n, n) будет появляться с вероятностью $\frac{1}{2n-1}$; ребро (n, i) возникнет с вероятностью $\frac{\deg i}{2n-1}$, где $\deg i$ — степень вершины i в графе G_1^{n-1} . Очевидно, что распределение вероятностей

задано корректно, поскольку

$$\sum_{i=1}^{n-1} \frac{\deg i}{2n-1} + \frac{1}{2n-1} = \frac{2n-2}{2n-1} + \frac{1}{2n-1} = 1.$$

Случайный граф G_1^n построен, и он удовлетворяет принципу предпочтительного присоединения.

Осталось перейти к G_k^n . Берем G_1^{kn} . Это граф с kn вершинами и kn ребрами. Делим множество его вершин на последовательные куски размера k :

$$\{1, \dots, k\}, \quad \{k+1, \dots, 2k\}, \quad \dots, \quad \{k(n-1)+1, \dots, kn\}.$$

Объявляем каждый кусок «вершиной», а ребра сохраняем, т.е. если были ребра внутри куска, то будут кратные петли, а были ребра между двумя различными кусками – будут кратные ребра. Внешне – вполне себе Интернет, как мы его и представляли. Вершин стало n , а ребер – по-прежнему kn . Цель реализована.

3.2.2. Статическая модификация, или LCD-модель

Введем такой объект, который называется *линейной хордовой диаграммой*. Вообще-то, он возник в топологии и теории узлов (см., например, [19]), но его комбинаторика оказывается напрямую связана с формированием веб-графа.

Итак, зафиксируем на оси абсцисс на плоскости $2n$ точек: $1, 2, 3, \dots, 2n$. Разобьем эти точки на пары, и элементы каждой пары соединим дугой, лежащей в верхней полуплоскости. Полученный объект назовем *линейной хордовой диаграммой* (*linearized chord diagram*, или LCD). Дуги в нем могут пересекаться, лежать друг под дружкой, но не могут иметь общих вершин. Количество различных LCD легко считается, оно равно

$$l_n = \frac{(2n)!}{2^n n!}.$$

По каждой LCD построим граф с n вершинами и n ребрами. Действуем так. Идем слева направо по оси абсцисс, пока не встретим впервые правый конец какой-либо дуги. Пусть этот конец имеет номер i_1 . Объявляем набор $\{1, \dots, i_1\}$ первой вершиной будущего графа. Снова идем от $i_1 + 1$ направо до первого правого конца i_2 какой-либо дуги. Объявляем второй вершиной графа набор $\{i_1 + 1, \dots, i_2\}$.

И так далее. Поскольку правых концов у дуг в данной диаграмме n штук, получаем всего n вершин. А ребра порождаем дугами. Иными словами, две вершины соединяем ребром, коль скоро между соответствующими наборами есть дуга. Ребра ориентируем справа налево. Аналогично возникают петли. Дуг n , и ребер n .

Теперь считаем LCD случайной, т.е. полагаем вероятность каждой диаграммы равной $1/l_n$. Возникают случайные графы. Можно показать, что такие графы по своим вероятностным характеристикам практически неотличимы от графов G_1^n .

Графы с n вершинами и kn ребрами получаем тем же способом, что и в предыдущем пункте.

3.2.3. Некоторые результаты

Замечательна модель Боллобаша–Риордана не только тем, что с ее помощью наводится порядок в «каше», которую «заварили» Барабаш и Альберт, но еще и тем, что она полностью адекватна эмпирическим наблюдениям. Прежде всего справедлива

Теорема 8. Для любого $k \geq 2$ и любого $\varepsilon > 0$

$$P\left((1 - \varepsilon)\frac{\ln n}{\ln \ln n} \leq \text{diam } G_k^n \leq (1 + \varepsilon)\frac{\ln n}{\ln \ln n}\right) \rightarrow 1, \quad n \rightarrow \infty.$$

На первый взгляд утверждение кажется непонятным. Ну, хорошо: диаметр плотно сконцентрирован (по вероятности) около величины $\ln n / \ln \ln n$. А у нас ведь какие-то 5–7 были? Так ничего странного. Вершин в интернете образца 1999 года около 10^7 . Значит,

$$\frac{\ln 10^7}{\ln \ln 10^7} = \frac{7 \ln 10}{\ln 7 + \ln \ln 10} \approx 6.$$

Фантастическое попадание. Отметим, что при недавней проверке с другими цифрами эмпирика снова подтвердилась.

Теорема 8 доказана в работе [20] авторами модели. А в работе [21] была внесена ясность и в вопрос о распределении степеней вершин.

Теорема 9. Для любого $k \geq 1$ и любого $d \leq n^{1/15}$

$$\begin{aligned} E\left(\frac{\left|\{i = 1, \dots, n : \deg_{G_k^n} i = d\}\right|}{n}\right) &\sim \\ &\sim \frac{2k(k+1)}{(d+k+1)(d+k+2)(d+k+3)}. \end{aligned} \quad (2)$$

Поскольку k – константа, выражение в правой части (2) имеет вид const/d^3 . Это в точности степенной закон. Правда, в формулировке теоремы написано математическое ожидание, а не вероятность, но одно из другого получается за счет мартингальных неравенств и соответствующих теорем о плотной концентрации меры около среднего (см. [21]).

У теоремы 9 есть все же два неприятных момента. Первый состоит в том, что степень d в степенном законе, который в ней устанавливается, равна не 2.1, а 3. Второй – это ограничение $d \leq n^{1/15}$, которое ставит крест на практической применимости теоремы. Даже при $n \approx 10^{12}$, чего в природе (пока) не бывает, мы имеем лишь $d \leq 10^{4/5}$, и это нелепо.

Последний недостаток недавно устранил Е. А. Гречников – исследователь-разработчик в Яндексе, который получил более точный результат практически без ограничений на d . Статья Гречникова еще не опубликована.

Первым же недостатком занимались много и, в частности, предлагали различные альтернативные модели. Одну из таких моделей мы обсудим в разделе 3.3. Но прежде скажем еще несколько слов о свойствах LCD-модели.

Пусть H – фиксированный граф. Обозначим через $\sharp(H, G_k^n)$ случайную величину, равную количеству подграфов графа G_k^n , изоморфных графу H . Как распределена эта величина? Изучали ее математическое ожидание в разных специальных случаях. Например, в работе [17] приводится громоздкая общая формула и пара ее симплических следствий, которые мы выпишем и здесь.

Теорема 10. Пусть $k \geq 2$. Пусть также K_3 – полный граф на трех вершинах. Тогда при $n \rightarrow \infty$

$$E(\sharp(K_3, G_k^n)) = (1 + o(1)) \cdot \frac{k \cdot (k-1) \cdot (k+1)}{48} \cdot (\ln n)^3.$$

Теорема 11. Пусть фиксированы $k \geq 2$ и $l \geq 3$. Пусть также C_l – цикл на l вершинах. Тогда

$$E(\#(C_l, G_k^n)) = (1 + o(1)) \cdot c_{k,l} \cdot (\ln n)^l$$

при $n \rightarrow \infty$, где $c_{k,l}$ – это положительная константа. Более того, при $k \rightarrow \infty$ имеем $c_{k,l} = \Theta(k^l)$.

Студенты МФТИ А. Рябченко и Е. Самосват недавно (в несколько иной, но очень близкой модели) установили следующий общий факт.

Теорема 12. Пусть задан граф H , степени вершин которого равны d_1, \dots, d_s . Обозначим через $\#(d_i = m)$ число вершин в H , степень каждой из которых равна m . Тогда

$$E(\#(H, G_k^n)) = \Theta\left(n^{\#\{d_i=0\}} \cdot (\sqrt{n})^{\#\{d_i=1\}} \cdot (\ln n)^{\#\{d_i=2\}}\right).$$

Зависимость от k занесена в константу Θ .

Надо полагать, что нечто подобное было известно и авторам статьи [17], но мы ничего похожего в литературе не встречали. А такая запись результата очень удобна. Скажем, в теореме 10 речь идет про K_3 . Ясно, что для K_3 выполнено

$$\#(d_i = 0) = \#(d_i = 1) = 0, \quad \#(d_i = 2) = 3.$$

По теореме 12

$$E(\#(K_3, G_k^n)) = \Theta((\ln n)^3),$$

и это прекрасно согласуется с теоремой 10. Аналогично можно разобраться и с циклами (теорема 11). А если взять K_4 – полный граф на четырех вершинах, – то теорема 12 скажет, что средняя его встречаемость в веб-графе постоянна. Иными словами, «тетраэдров» в веб-графах почти не бывает.

Отметим, что в реальном вебе случаются не только тетраэдры, но и клики куда большей мощности. Это связано с деятельностью спамеров, которые искусственно расставляют ссылки, желая повысить рейтинги сайтов, заплативших за раскрутку. Спам в модели Боллобаша–Риордана не учтен, и это тоже минус.

3.3. Модель копирования

Здесь мы опишем еще одну очень интересную модель, которая также призвана объяснить феномен степенного закона в реальных сетях. Эта модель возникла практически в одно время с моделью Барабаши—Альберт. Она принадлежит Р. Кумару, П. Рагхавану, С. Раджагопалану, Д. Сивакумару, А. Томкинсу и Э. Упфалу (см. [22]).

Фиксируем $\alpha \in (0, 1)$ и $d \geq 1$, $d \in \mathbb{N}$. Случайный граф будет расти, и это будет похоже на процесс из пункта 3.2.1. Однако здесь процесс будет устроен совсем по-другому.

В качестве начального графа возьмем любой d -регулярный граф (граф, у которого степень каждой вершины равна d). Пусть построен граф $G_t = (V_t, E_t)$ с номером t . Здесь $V_t = \{u_1, \dots, u_s\}$, где s отличается от t на число вершин начального графа, т.е. на некоторую константу, выражаемую через d . Добавим к G_t одну новую вершину u_{s+1} и d ребер, выходящих из u_{s+1} . Для этого сперва выберем случайную вершину $p \in V_t$ (все вершины в V_t равновероятны). Одно за другим строим ребра из u_{s+1} в V_t . На шаге с номером i , $i \in \{1, \dots, d\}$ разыгрываем случайную величину, которая с вероятностью α принимает значение 1 («монетка падает решкой кверху») и с вероятностью $1 - \alpha$ принимает значение 0 («монетка падает орлом кверху»). Если выпала единица, то выпускаем ребро из u_{s+1} в случайную вершину из V_t (все вершины в V_t равновероятны). Если вышел ноль, то берем i -го по номеру соседа вершины p . Последнее действие всегда возможно, т.к. по построению у каждой вершины не менее d соседей.

Интуиция за всем этим примерно такая. Появляется новый сайт. Проставляя очередную ссылку, его владелец с некоторой вероятностью будет ориентироваться на кого-то из своих предшественников. Скажем, сайт посвящен автомобилям. Вероятно, владелец возьмет один из уже существовавших сайтов про автомобили и скопирует оттуда ссылку (с точки зрения стороннего наблюдателя вполне случайную). Это ситуация, когда монетка выпала орлом кверху (p – это сайт, с которого копируются ссылки). Однако при простановке ссылки владелец может и никого не копировать, а случайно (по нашему мнению) цитировать кого-то из предшественников. Это случай выпадения решки. Таким образом, $1 - \alpha$ – это вероятность копирования или, если угодно, вероятность выбора, мотивированного тематикой сайта.

Основной результат из [22] – это теорема 13.

Теорема 13. Пусть $N_{t,r}$ – это математическое ожидание числа вершин степени r в графе G_t . Тогда

$$\lim_{t \rightarrow \infty} \frac{N_{t,r}}{t} = \Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right).$$

Пафос теоремы в том, что в ней мы снова приходим к степенному закону. Более того, если вероятность копирования близка к 1 (а величина α – к нулю), то показатель степени может равняться ожидаемой величине 2.1, чего до сих пор у нас не было.

В целом распределение степеней вершин в модели копирования очень похоже на распределение степеней вершин в модели Боллобаша–Риордана. В остальном модели сильно разнятся. Например, в модели Боллобаша–Риордана практически отсутствуют плотные двудольные подграфы (см. теорему 12); в модели копирования таких подграфов очень много. Это особенно важно ввиду того, что спамерские структуры, о которых мы вскользь говорили в конце пункта 3.2.3, зачастую образуют именно двудольные графы с плотной перелинковкой.

Литература

1. Erdős P., Rényi A. On random graphs I // Publ. Math. Debrecen. 1959. V. 6. P. 290–297.
2. Erdős P., Rényi A. On the evolution of random graphs // Publ. Math. Inst. Hungar. Acad. Sci. 1960. V. 5. P. 17–61.
3. Erdős P., Rényi A. On the evolution of random graphs // Bull. Inst. Int. Statist. Tokyo. 1961. V. 38. P. 343–347.
4. Степанов В. Е. О вероятности связности случайного графа $g_m(t)$ // Теория вероятностей и ее применения. 1970. Т. 15. № 1. С. 55–67.
5. Степанов В. Е. Фазовый переход в случайных графах // Теория вероятностей и ее применения. 1970. Т. 15. № 2. С. 187–203.
6. Степанов В. Е. Структура случайных графов $g_n(x|h)$ // Теория вероятностей и ее применения. 1972. Т. 17. № 3. С. 227–242.
7. Колчин В. Ф. Случайные графы. М.: Физматлит, 2004.
8. Bollobás B. Random Graphs. Cambridge: Cambridge Univ. Press, 2001.
9. Алон Н., Спенсер Дж. Вероятностный метод. М: Бином. Лаборатория знаний, 2007.
10. Janson S., Luczak T., Rucinski A. Random graphs. N.Y.: Wiley, 2000.
11. Маргулис Г. А. Вероятностные характеристики графов с большой связностью // Проблемы передачи информации. 1974. Т. 10. С. 101–108.
12. Karp R. The transitive closure of a random digraph // Random structures and algorithms. 1990. V. 1. P. 73–94.
13. Карлин С. Основы теории случайных процессов. М: Мир, 1971.
14. Barabási L.-A., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.

15. *Barabási L.-A., Albert R., Jeong H.* Scale-free characteristics of random networks: the topology of the world-wide web // *Physica A*. 2000. V. 281. P. 69–77.
16. *Albert R., Jeong H., Barabási L. A.* Diameter of the world-wide web // *Nature*. 1999. V. 401. P. 130–131.
17. *Bollobás B., Riordan O.* Mathematical results on scale-free random graphs. *Handbook of graphs and networks*. Weinheim: Wiley-VCH. 2003. P. 1–34.
18. *Райгородский А. М.* Экстремальные задачи теории графов и анализ данных. М.–Ижевск: НИЦ «РХД», 2009.
19. *Stoimenow A.* Enumeration of chord diagrams and an upper bound for Vassiliev invariants // *J. Knot Theory Ramifications*. 1998. V. 7. N. 1. P. 93–114.
20. *Bollobás B., Riordan O.* The diameter of a scale-free random graph // *Combinatorica*. 2004. V. 24. N. 1. P. 5–34.
21. *Bollobás B., Riordan O., Spencer J., Tusnády G.* The degree sequence of a scale-free random graph process // *Random Structures Algorithms*. 2001. V. 18. N. 3. P. 279–290.
22. *Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E.* Stochastic models for the web graph // *Proc. 41st Symposium on Foundations of Computer Science*. 2000.

Задачи

Задачи к главам пособия и приложениям

Задача к главе 1 (сублинейный приближенный вероятностный алгоритм для матричных игр; Григориадис–Хачиян, 1995). Рассматривается симметричная антагонистическая игра двух лиц X и Y. Смешанные стратегии X и Y будем обозначать соответственно \vec{x} и \vec{y} . При этом x_k – вероятность того, что игрок X выберет стратегию с номером k , аналогично определяется y_k . Таким образом,

$$\vec{x}, \vec{y} \in S = \left\{ \vec{x} \in \mathbb{R}^n : \vec{e}^T \vec{x} = 1, \vec{x} \geq \vec{0} \right\}, \text{ где } \vec{e} = (1, \dots, 1)^T.$$

Выигрыш игрока X:

$$V_X(\vec{x}, \vec{y}) = \vec{y}^T A \vec{x},$$

а выигрыш игрока Y:

$$V_Y(\vec{x}, \vec{y}) = -\vec{y}^T A \vec{x} \text{ (игра антагонистическая).}$$

Каждый игрок стремится максимизировать свой выигрыш при заданном ходе оппонента. Равновесием Нэша (в смешанных стратегиях) называется такая пара стратегий (\vec{x}^*, \vec{y}^*) , что

$$\vec{x}^* \in \operatorname{Arg} \max_{\vec{x} \in S} \vec{y}^{*T} A \vec{x}, \quad \vec{y}^* \in \operatorname{Arg} \min_{\vec{y} \in S} \vec{y}^T A \vec{x}^*.$$

Ценой игры называют

$$\max_{\vec{x} \in S} \min_{\vec{y} \in S} \vec{y}^T A \vec{x} = \min_{\vec{y} \in S} \max_{\vec{x} \in S} \vec{y}^T A \vec{x} = \vec{y}^{*T} A \vec{x}^*.$$

Поскольку по условию игра также симметричная, то $A = -A^T$ – матрица $n \times n$. С помощью стандартной редукции можно свести к этому случаю общий случай произвольной несимметричной матричной игры. В рассматриваемом же случае цена игры (выигрыш игроков в положении равновесия Нэша) есть 0, а множества оптимальных стратегий игроков совпадают. Требуется найти с точностью $\varepsilon > 0$ положение равновесия Нэша (оптимальную стратегию), т.е. требуется найти такой вектор $\vec{x} \in S$, что $A \vec{x} \leq \varepsilon \vec{e}$. Покажите, считая элементы матрицы A равномерно ограниченными, скажем, единицей, что приводимый ниже алгоритм находит с вероятностью не меньшей $1/2$ (вместо $1/2$ можно взять любое положительное число, меньшее единицы) такой \vec{x} за время

$$O(\varepsilon^{-2} n \log^2 n),$$

т.е. в определенном смысле даже не вся матрица (из n^2 элементов) просматривается. Отметим также, что в классе детерминированных алгоритмов время работы растет с ростом n не медленнее, чем $\sim n^2$ (эта нижняя оценка получается из информационных соображений [1, 2]). Другими словами, никакой детерминированный алгоритм не может также асимптотически быстро находить приближенно равновесие Нэша. Точнее говоря, описанный ниже вероятностный алгоритм дает почти квадратичное ускорение по сравнению с детерминированными.

Алгоритм

1. **Инициализация:** $\vec{X} = \vec{U} = \vec{0}$, $\vec{p} = \vec{e}/n$, $t = 0$.
2. **Повторить:**
3. **Счетчик итераций:** $t := t + 1$.
4. **Датчик случайных чисел:** выбираем $k \in \{1, \dots, n\}$ с вероятностью p_k .
5. **Модификация \vec{X} :** $X_k := X_k + 1$.
6. **Модификация \vec{U} :** $U_i := U_i + a_{ik}$, $i = 1, \dots, n$.
7. **Модификация \vec{p} :** $p_i := p_i \exp(\varepsilon a_{ik}/2) / \left(\sum_{j=1}^n p_j \exp(\varepsilon a_{jk}/2) \right)$,
 $i = 1, \dots, n$.
8. **Критерий останова:** если $\vec{U}/t \leq \varepsilon \vec{e}$, то останавливаемся и печатаем $\vec{x} = \vec{X}/t$.

Указание. Покажите, что с вероятностью не меньшей, чем $1/2$, алгоритм остановится через $t^* = 4\varepsilon^{-2} \ln n$ итераций. Для этого введите

$$P_i(t) = \exp(\varepsilon a_{ik}(t)/2) \text{ и } \Phi(t) = \sum_{i=1}^n P_i(t).$$

Покажите, что

$$\begin{aligned} E[\Phi(t+1)|\vec{P}(t)] &= \Phi(t) \sum_{i,k=1}^n p_i(t) p_k(t) \exp(\varepsilon a_{ik}/2) \text{ и} \\ &\quad \exp(\varepsilon a_{ik}/2) \leq 1 + \varepsilon a_{ik}/2 + \varepsilon^2/6. \end{aligned}$$

Используя это и кососимметричность матрицы A , покажите, что

$$E[\Phi(t+1)] \leq n E[\Phi(t)] (1 + \varepsilon^2/6).$$

Следовательно,

$$E[\Phi(t)] \leq n \exp(t\varepsilon^2/6) \text{ и } E[\Phi(t^*)] \leq n^{5/3}.$$

Отсюда по неравенству Маркова имеем, что ($n \geq 8$):

$$P(\Phi(t^*) \leq n^2) \geq P(\Phi(t^*) \leq 2n^{5/3}) \geq 1/2.$$

Тогда

$$P(\varepsilon U_i(t^*)/2 \leq 2 \ln n, i=1, \dots, n) \geq 1/2.$$

Откуда уже следует, что

$$P(\bar{x}(t^*) \leq \varepsilon \bar{e}) \geq 1/2.$$

Приведенные выше идеи случайного квазиградиентного спуска сейчас активно используются в современных численных методах решения задач выпуклой оптимизации в пространствах огромной размерности [3].

Литература

- Хачян Л. Г. Избранные труды / сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.
- <ftp://ftp.mccme.ru/users/shen/kolmbook.pdf>
- <http://www2.isye.gatech.edu/~nemirovs/>
<http://www.core.ucl.ac.be/staff/biosketchNesterov.html>
<http://elis.dvo.ru/~nurmi/>

Задача к главе 1 (стохастическая марковская динамика, приводящая к равновесию Нэша–Вардропа в модели распределения потоков; Гасникова–Дорн, 2010). Свой путь на $(n+1)$ -м шаге¹ игрок, сидящий на корреспонденции w , выбирает согласно смешанной стратегии (в независимости от всех остальных: с вероятностью

$$\text{Prob}_p^w(n+1) = \gamma_p \cdot (x_p(n)+1) \exp(-G_p(\bar{x}(n))/T) / Z_n^w, \quad w \in W,$$

выбрать путь $p \in P_w$ ($0 < \gamma_p \leq 1$), а с вероятностью $1 - \gamma_p$ – действовать согласно стратегии, использованной на предыдущем n -м шаге. Здесь $x_p(n)$ – количество игроков, сидящих на корреспонденции w и вы-

¹ Например, шаг с периодом в день можно проинтерпретировать как выбор утром маршрута следования (пути) из дома на работу, исходя из «опыта» вчерашнего дня. Заметим, что информацию о $G_p(\bar{x}(n))$ водители (игроки) черпают из открытых источников типа Яндекс-пробки, а множитель $(x_p(n)+1)$ определяется исходя из случайного опроса соседей, знакомых, коллег и т.п.

бравших на n -м шаге стратегию $p \in P_w$, а

$$Z_n^w = \sum_{p \in P_w} \gamma_n \cdot (x_p(n) + 1) \exp(-G_p(\bar{x}(n))/T).$$

Множитель $(x_p(n) + 1)$ характеризует желание имитировать, а также надежность использования этой стратегии. Именно этот множитель подмечает специфику рассматриваемой задачи (без него сходимость будет в общем случае не к равновесию Нэша–Вардропа) и отличает предложенную в задаче динамику от многих других (см. ниже краткий обзор). Параметр γ характеризует «консерватизм» («ленивость»), чем меньше γ , тем более консервативный игрок; «температура» T характеризует отношение к риску («горячность»), чем больше температура, тем более «горячий игрок», склонный к более рискованным действиям.

Как показали разнообразные численные эксперименты, часто вполне разумно выбирать $\gamma_n \sim 1/n$. При таком выборе γ_n наблюдается сходимость к равновесию Нэша–Вардропа при наиболее общих условиях относительно T (вне зависимости от точки старта). Строго говоря, наблюдается сходимость не к равновесию, а к некоторой его окрестности, уменьшающейся с уменьшением T . Стоит также обратить внимание на высокую эффективность предложенной процедуры «нащупывания равновесия» с точки зрения количества итераций. Иначе говоря, на предложенный итерационный процесс можно смотреть просто как на эффективный способ численного нахождения равновесия Нэша–Вардропа. В экспериментах с людьми (студентами 5-го курса ФУПМ) также наблюдалась сходимость к равновесию и колебания около него.

Введение в динамику стохастичности сближает предложенный подход с поиском так называемых «стохастических равновесий в транспортных сетях» [1], с другой стороны, он принципиально отличается тем, что предполагает знание транспортных расходов по маршрутам (используется достоверная информация вчерашнего дня), на основе которых производится рандомизированный выбор. В стохастическом же равновесии водитель узнает лишь случайную оценку времени проезда по каждому из маршрутов и затем выбирает маршрут с минимальным временем.

Предложенную схему можно трактовать скорее как стохастическую динамику наилучших ответов в эволюционной (популяционной) игре [2–6], при этом имеется много общего с концепциями «quantal response equilibria» [7] (используется похожая рандомизация) и «minority games» [8] (наблюдаются похожие колебания около положения равно-

весия). Близкой к предложенному итерационному процессу является концепция генетических алгоритмов [9] и эффективный приближенный вероятностный (с гиббсовским («logit») распределением вероятностей) алгоритм Григориадиса–Хачияна [10].

Исходя из результатов главы 1, предложите ситуацию, когда равновесие Нэша–Вардропа не единственно (считайте, например, что за проезд по дороге с водителями взимается некоторая сумма; см. также пример В. И. Швецова ниже). Для описанной выше динамики, исследуйте, к какому положению равновесия (в зависимости от точки старта) она будет сходиться.

Более подробно о моделях распределения потоков и связанных с ними задачах можно прочитать, например, в главе 1 и [1, 11, 12].

В заключение рассмотрим два примера. Первый демонстрирует, что в результате строительства новой дороги новое равновесие Нэша–Вардропа окажется неэффективным по Парето и будет строго хуже, чем то, которое было до строительства. Тем не менее предложенная выше марковская динамика наилучших ответов (также как и эксперимент) приводит именно к такому, не оптимальному по Парето, равновесию.

Пример (парадокс Брайеса, 1968). Пусть корреспонденция $x_{14} = 6$ (тысяч автомобилей/ч). Вес ребра (удельные затраты на проезд по этому ребру) есть время движения по ребру (в минутах), если поток через ребро есть y_{ij} (тысяч автомобилей/ч). Например, в случае 2: $y_{24} = x_{124} + x_{1324}$ (см. рис. 1). Естественно считать, что время движения – возрастающая функция потока.

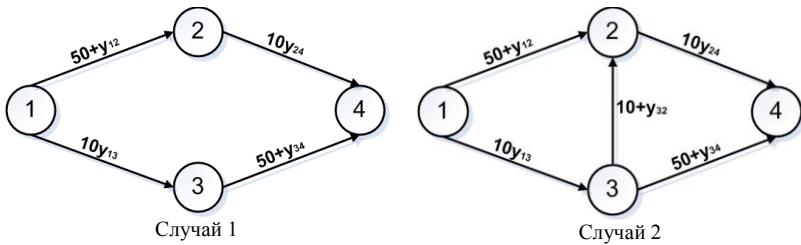


Рис. 1

Случай 1: $x_{124} = x_{134} = 3$.

Полное время в пути $T = 83$ мин

Случай 2: $x_{124} = x_{1324} = x_{134} = 2$.

Полное время в пути $T = 92$ мин

Оба равновесия Нэша–Вардропа (в случаях 1 и 2) являются «притягивающими» положениями равновесия описанной выше динамики (положили $\gamma \sim 1$, $T \sim 15 - 35$), см. рис. 2–4 (для случая 2).

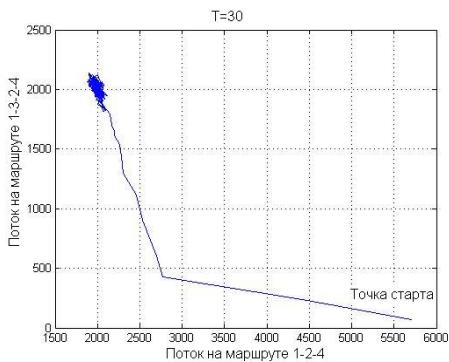


Рис. 2

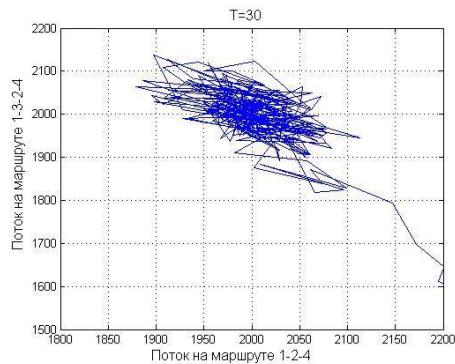


Рис. 3

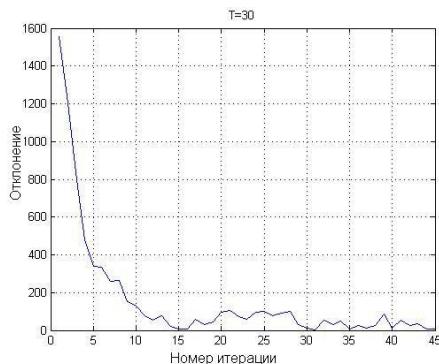


Рис. 4

Второй пример демонстрирует, что при весьма естественных условиях вектор-функция удельных затрат пользователей на проезд $\vec{G}(\vec{x})$ может не быть строго монотонной (см. п. 1.1.3.1):

$$\exists \vec{x}, \vec{y} \in X (\vec{x} \neq \vec{y}): \vec{G}(\vec{x}) = \vec{G}(\vec{y}) \Rightarrow \langle \vec{G}(\vec{x}) - \vec{G}(\vec{y}), \vec{x} - \vec{y} \rangle = 0.$$

Связанно это может быть, например, с тем, что (см. п. 1.1.4)

$$\vec{G}(\vec{x}) = \Theta^T \vec{\tau}(\vec{y}), \quad \vec{y} = \Theta \vec{x},$$

где вектор $\vec{y} = \{y_e\}_{e \in E}$ описывает загрузку ребер (дуг) графа транспортной сети, $\vec{\tau}(\vec{y}) = \{\tau_e(y_e)\}_{e \in E}$ – вектор-функция затрат на проезд по ребрам графа транспортной сети, Θ – матрица инцидентности ребер и путей, и разные векторы распределения потоков \vec{x} могут соответствовать одному и тому же вектору $\vec{y} = \Theta \vec{x}$.

Пример (неединственность равновесия; В. И. Швецов, 2010). На рис. 5 показано равновесное распределение потоков для любого значения параметра $x \in [0, 0.5]$. Подробности имеются в статье [14].

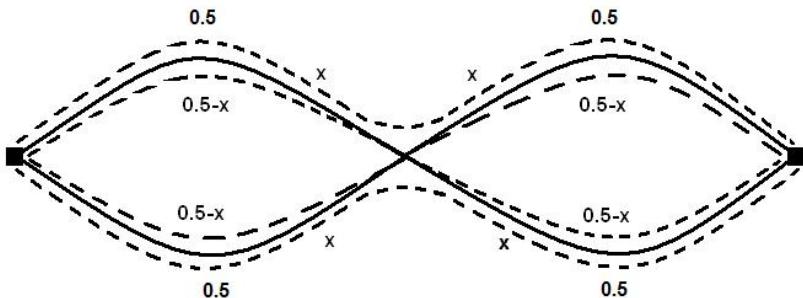


Рис. 5

Литература

1. Sheffi Y. Urban transportation networks: Equilibrium analysis with mathematical programming methods. N.J.: Prentice-Hall Inc., Englewood Cliffs, 1985.
2. Foster D., Young P. Stochastic evolutionary game dynamics // Theoretical population biology. 1990. V. 38. № 2.
3. Cressman R. Evolutionary game theory and extensive form games. Cambridge, Mass.: MIT Press, 2003.

4. Hofbauer J., Sigmund K. Evolutionary game dynamics // Bulletin of the AMS. 2003. V. 40. № 4. P. 479–519.
5. Васин А. А., Краснощеков П. С., Морозов В. В. Исследование операций. М.: Издательский центр «Академия», 2008.
6. <http://www.cs.cornell.edu/home/kleinber/networks-book/>
7. McKelvey R. D., Palfrey T. R. Quantal response equilibria for extensive form games // Experimental economics. 1998. V. 1. P. 9–41.
8. Marsili M. Toy models of markets with heterogeneous interacting agents // e-print www.unifr.ch/econophysics/, 2001.
9. Fogel D. B. Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. New York: IEEE Press, 2000.
10. Хачян Л. Г. Избранные труды / сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.
11. Стенбринк П. А. Оптимизация транспортных сетей. М.: Транспорт, 1981.
12. Гасникова Е. В., Дорн Ю. В. О стохастической марковской динамике, приводящей к равновесию Нэша–Вардропа в модели распределения потоков // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)
13. Como G., Salva K., Acemoglu D., Dahleh M.A., Frazzoli E. Stability analysis of transportation networks with multiscale driver decisions // e-print [arXiv:1101.2220v1](https://arxiv.org/abs/1101.2220v1), 2011.
14. Швецов В. И. Проблемы моделирования передвижений в транспортных сетях // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)

Задача (сходимость к равновесию Нэша; Малишевский–Опойцев, 1972). Рассмотрим систему обыкновенных дифференциальных уравнений (СОДУ):²

² Такие СОДУ возникают, например, при исследовании коллективного поведения, в частности «при нащупывании равновесия Нэша». Действительно, пусть $D_i(x_1, \dots, x_i, \dots, x_n)$ – функция выигрыша игрока i , если игроки придерживаются набора стратегий $\bar{x} = (x_1, \dots, x_i, \dots, x_n)$. Пусть «функция цели» $f_i(\bar{x})$ игрока i однозначным образом определяется из условия

$$D_i(x_1, \dots, f_i(\bar{x}), \dots, x_n) = \max_{x_i \in X_i} D_i(x_1, \dots, x_i, \dots, x_n),$$

где X_i – множество возможных стратегий игрока i . Тогда СОДУ (D), очевидным образом, выражает стремление игроков двигаться по направлению своей цели (по мере движе-

$$\dot{x}_i = \gamma_i(t)(f_i(x_1, \dots, x_n) - x_i), \quad (\text{D})$$

- $\gamma_i(t) > 0$ при $t \geq 0$ и $\int_0^\infty \gamma_i(t) dt = \infty$, $i = 1, \dots, n$;
- $f_i(\cdot)$, $i = 1, \dots, n$ – непрерывные функции;
- $\exists K \subset \mathbb{R}^n$ (компакт): $\vec{f}(K) \subseteq K$.

Докажите, что приводимое ниже условие обеспечивает существование равновесия СОДУ (единственность в K) и его асимптотическую устойчивость:

$$\forall \vec{x} \in K, i = 1, \dots, n \rightarrow \sum_{j=1}^n \left| \frac{\partial f_i}{\partial x_j}(\vec{x}) \right| < 1.$$

Указание. В кубической норме ($\|\vec{x}\|_\square = \max_{i=1, \dots, n} |x_i|$) $\vec{f}(\vec{x})$ – сжимающее отображение компакта K в себя. Устойчивость показывается с помощью второго метода Ляпунова. В качестве функции Ляпунова следует взять $V(\vec{x}) = \|\vec{x} - \vec{x}^*\|_\square$, где \vec{x}^* – единственное положение равновесия в K .

Литература

1. Опойцев В. И. Равновесие и устойчивость в моделях коллективного поведения. М.: Наука, 1977.
2. Мулен Э. Теория игр с примерами из математической экономики. М.: Мир, 1985.
3. Малишевский А. В. Качественные модели в теории сложных систем. М.: Наука, 1998.

Задача (устойчивые системы большой размерности; В. И. Опойцев, 1985). Из курсов функционального анализа и вычислительной математики хорошо известно, что если спектральный радиус матрицы $A = \|a_{ij}\|_{i,j=1}^n$ меньше единицы: $\rho(A) < 1$, то итерационный процесс $\vec{x}^{n+1} = A\vec{x}^n + \vec{b}$ (СОДУ $\dot{\vec{x}} = A\vec{x} + \vec{b} - \vec{x}$, см. СОДУ (D) из предыдущей задачи с $\vec{f}(\vec{x}) = A\vec{x} + \vec{b}$) в независимости от точки старта \vec{x}^0 сходится к

ния цель меняется), т.е. определенную рациональность игроков. Отметим здесь нечувствительность результатов к тому, насколько сильно (это характеризуется функциями $\gamma_i(t) > 0$, $i = 1, \dots, n$) каждый из игроков стремится к своей цели.

единственному решению уравнения $\vec{x}^* = A\vec{x}^* + \vec{b}$. Скажем, если $\|A\|_{\square} = \max_i \sum_{j=1}^n |a_{ij}| < 1$, то и $\rho(A) < 1$ (обратное, конечно, не верно).

Предположим, что существует такое малое $\varepsilon > 0$, что

$$\frac{1}{n} \sum_{i,j=1}^n |a_{ij}| < 1 - \varepsilon. \quad (\text{S})$$

Очевидно, что отсюда тем более не следует: $\rho(A) < 1$. Тем не менее, введя на множестве матриц, удовлетворяющих условию (S) равномерную меру, покажите, что относительная мера тех матриц (удовлетворяющих условию (S)), для которых спектральный радиус не меньше единицы, стремится к нулю с ростом n (ε – фиксировано и от n не зависит).

Указание. 1. Покажите, что при доказательстве можно ограничиться матрицами с неотрицательными элементами.

2. Покажите, что, не ограничивая общности, можно также считать, что в определении множества S стоит не неравенство, а равенство. Так определенное множество матриц будем называть SE.

3. Положите, например,³ $a_{ij} \in \text{Exp}(n/(1-\varepsilon))$ – н.о.р. и покажите, что при $n \rightarrow \infty$ распределение элементов случайной матрицы $A = \|a_{ij}\|_{i,j=1}^n$ будет сходиться (уточните в каком смысле) к равномерному распределению на SE.

4. Покажите, введя обозначение $P_n = P(\|A\|_{\square} \geq 1) \geq P(\rho(A) \geq 1)$ и используя неравенство Чебышева, что

$$P_n \leq nP\left(\sum_{j=1}^n a_{1j} \geq 1\right) \leq \frac{n}{\varepsilon^4} E\left[\left(\sum_{j=1}^n a_{1j} - (1-\varepsilon)\right)^4\right] = O\left(\frac{1}{n}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Литература

1. Опойцев В. И. Устойчивые системы большой размерности // Автоматика и телемеханика. 1986. № 6. С. 43–49.
2. Опойцев В. И. Нелинейная системостатистика. М.: Наука, 1986.

³ Независимые, одинаково распределенные с.в. (случайные величины) по показательному закону с параметром $n/(1-\varepsilon)$, т.е. $P(a_{ij} > x) = \exp(-(n/(1-\varepsilon))x)$.

Задача к главам 2, 3 и приложению М. Л. Бланка. Предложите модель клеточных автоматов многополосного транспортного потока, в которой бы учитывались перестраивания АТС из полосы в полосу (например, для осуществления необходимых маневров в окрестности вершин графа транспортной сети), и которая бы описывала три фазы Кернера. Для проверки последнего можно прибегнуть к численным экспериментам с предложенной моделью.

Задача к приложению А. А. Замятину и В. А. Малышева (пуассоновский процесс как процесс восстановления). В приложениях (в теории надежности, теории массового обслуживания) широко используются *процессы восстановления (потоки Пальма)*. Основным (и наиболее удобным для анализа) представителем таких процессов является *пуассоновский процесс*, который можно определить следующим образом:⁴

$$K(t) = \max \left\{ k : \sum_{i=1}^k T_i < t \right\},$$

где⁵ н.о.р. с.в. $T_i \in \text{Exp}(\lambda)$, т.е.⁶ $P(T_i > t) = e^{-\lambda t}$, $\lambda > 0$. Покажите, что пуассоновский процесс может быть определен таким образом. Будет ли пуассоновский процесс марковским?

Задача к приложению А. А. Замятину и В. А. Малышева (парадокс времени ожидания). Автобусы прибывают на остановку в соответствии с пуассоновским процессом с параметром $\lambda > 0$. Вы приходите на остановку в фиксированный момент времени (скажем, в полночь). Каково математическое ожидание времени, в течение которого Вы ждёте автобуса?

Литература

1. Секей Г. Парадоксы в теории вероятностей и математической статистике. М.–Ижевск: ИКИ, 2002.

⁴ $K(t)$ – число отказов приборов к моменту времени $t \geq 0$ (отказавший прибор сразу же заменяется исправным). Все приборы идентичны, т.е. имеют одинаковое распределение времени безотказной работы. Кроме того, приборы работают независимо друг от друга.

⁵ Параметр $\lambda > 0$ принято называть интенсивностью пуассоновского процесса.

⁶ Напомним важную особенность показательного распределения $\text{Exp}(\lambda)$ «отсутствие последействия»: $P(T_i > t + \tau | T_i > t) = P(T_i > \tau)$. Заметим также, что общие процессы восстановления задаются аналогичной формулой с той лишь разницей, что н.о.р. $T_i \geq 0$ п.н. уже не обязательно распределены по показательному закону.

2. Фёллер В. Введение в теорию вероятностей и ее приложения Т. 2. М.: УРСС, 2010.
3. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. 2. Марковские процессы как отправная точка теории случайных процессов и их приложения. М.: МЦНМО, 2010.

Задача к приложению А. А. Замятиной и В. А. Малышева (замкнутая сеть, теорема Гордона–Ньюэлла, 1967; Л. Г. Афанасьева, 2007). Рассматривается транспортная сеть, в которой между N станциями курсируют M такси. Клиенты пребывают в i -й узел в соответствии с пуассоновским потоком с параметром $\lambda_i > 0$ ($i = 1, \dots, N$). Если в момент прибытия в i -й узел там есть такси, клиент забирает его и с вероятностью $p_{ij} \geq 0$ направляется в j -й узел, по прибытии в который покидает сеть. Такси остается ждать в узле прибытия нового клиента. Времена перемещений из узла в узел – независимые случайные величины, имеющие показательное распределение с параметром $v_{ij} > 0$ для пары узлов (i, j) . Если в момент прихода клиента в узел там нет такси, клиент сразу покидает узел. Считая $p_{ij} = N^{-1}$, $\lambda_i = \lambda$, $v_{ij} = v$, покажите, что вероятность того, что клиент, поступивший в узел (в установившемся (стационарном) режиме работы сети), получит отказ, равна

$$p_{\text{отказа}}(N, M) = \sum_{k=0}^M \frac{C_{N-2+k}^k \rho^{M-k}}{(M-k)!} \left/ \sum_{k=0}^M \frac{C_{N-1+k}^k \rho^{M-k}}{(M-k)!} \right., \quad \rho = N\lambda/v.$$

Методом перевала покажите справедливость следующей асимптотики при $N \rightarrow \infty$:

$$p_{\text{отказа}}(N, rN) = 1 - \frac{2r}{\lambda/v + r + 1 + \sqrt{(\lambda/v + r + 1)^2 - 4\lambda r/v}} + O\left(\frac{1}{N}\right).$$

Литература

1. Афанасьева Л. Г. Очерки исследования операций. М.: Изд-во механико-математического факультета МГУ, 2007.
2. Федорюк М. В. Метод перевала. М.: УРСС, 2010.
3. Афанасьева Л. Г., Булинская Е. В. Математические модели транспортных систем, основанные на теории очередей // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков / под ред. акад. В. В. Козлова). 2010. Т. 2. № 4(8). (в печати)

Задача к приложению А. М. Райгородского (обобщенная схема размещений; В. Ф. Колчин, 1984). а) Пусть для неотрицательных целочисленных с.в. η_1, \dots, η_N существуют независимые одинаково распределенные с.в. ξ_1, \dots, ξ_N , такие, что

$$P(\eta_1 = k_1, \dots, \eta_N = k_N) = P(\xi_1 = k_1, \dots, \xi_N = k_N | \xi_1 + \dots + \xi_N = n). \quad (*)$$

Введем независимые одинаково распределенные с.в. $\xi_1^{(r)}, \dots, \xi_N^{(r)}$, где r – целое неотрицательное число и

$$P(\xi_1^{(r)} = k) = P(\xi_1 = k | \xi_1 \neq r), \quad k = 0, 1, \dots$$

Пусть $p_r = P(\xi_1 = r)$ и $S_N = \xi_1 + \dots + \xi_N$, $S_N^{(r)} = \xi_1^{(r)} + \dots + \xi_N^{(r)}$. Пусть $\mu_r(n, N)$ – число тех с.в. η_1, \dots, η_N , которые приняли значение r . Покажите, что с.в. типа $\mu_r(n, N)$ можно изучать с помощью *обобщенной схемы размещений*: для любого $k = 0, 1, \dots, N$:

$$P(\mu_r(n, N) = k) = C_n^k p_r^k (1 - p_r)^{N-k} \frac{P(S_{N-k}^{(r)} = n - kr)}{P(S_N = n)}.$$

Напомним, что в классической схеме размещений n различных частиц по N различным ячейкам было доказано, что распределение заполнений ячеек η_1, \dots, η_N имеет вид

$$P(\eta_1 = k_1, \dots, \eta_n = k_n) = \frac{n!}{k_1! \dots k_n! N^n},$$

где k_1, \dots, k_n – неотрицательные целые числа, такие, что $k_1 + \dots + k_n = n$. Если положить $\xi_1, \dots, \xi_N \in Po(\lambda)$ – н.о.п. ($\lambda > 0$ произвольно), то получим (*).

б)* Дан случайный граф⁷ (модель Эрдёша–Ренни) $G(n, p)$. Пусть

⁷ Даны n вершин, любые две вершины соединены ребром с вероятностью p независимо от того, какие еще пары вершин соединены ребрами. Таким образом, $q = 1 - p$ есть вероятность отказа ребра. По сути, в задаче приведен некий порог \bar{q} для q (по аналогии со статистической механикой иногда говорят, что этот порог характеризует «фазовый переход случайного графа»). Если в полном графе на n вершинах ребра отказывают с вероятностью, «большей» \bar{q} , то транспортная система почти наверное разрушится, если же ребра отказывают с вероятностью, «меньшей» \bar{q} , то транспортная система (несмотря на то, что может потерять много ребер) почти наверное сохранит свое основное свойство – возможность добраться по ребрам из любой вершины в любую другую.

$p = (c \ln n)/n$. Покажите, что при $c > 1$ граф $G(n, p)$ почти наверное связен⁸, а при $c < 1$ – почти наверное не связен.

в)** Пусть $p \geq \sqrt{(2 \ln n)/n}$, причем длина (вес) r_{ij} каждого появившегося ребра есть независимая от того, какие еще пары вершин соединены ребрами и какие длины у этих ребер – случайная величина, имеющая равномерное распределение на отрезке $[0, 2r]$. Покажите, что тогда почти наверное граф $G(n, p)$ имеет гамильтонов цикл, причём длина почти всех гамильтоновых циклов стабилизируется (имеет место плотная концентрация) около nr .

Литература

1. *Перепелица В. А.* Асимптотический подход к решению некоторых экстремальных задач на графах // Проблемы кибернетики. 1973. Т. 26. С. 291–314.
2. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2004.
3. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.

Задача к приложениям А. В. Колесникова, Е. В. Гасниковой (принцип концентрации площади сферы; А. Пуанкаре, 1911). Покажите, что если в многомерном шаре задано равномерное распределение вероятностей и согласно этому распределению вероятностей сгенерировано два случайных вектора, то с вероятностью, близкой к единице, концы этих векторов будут лежать почти на границе шара и эти два случайных вектора будут почти ортогональны.

Указание. Нетривиально второе утверждение (про ортогональность). Для того чтобы его установить, покажите, что доля от площади всей сферы S_r^n (радиуса r) в \mathbb{R}^n ($n \geq 3$), которую занимает площадь сегмента, проектирующегося в отрезок $[a, b]$, скажем, оси x_1 , равна

$$P[a, b] = \frac{\int\limits_a^b \left(1 - (x/r)^2\right)^{\frac{n-3}{2}} dx}{\int\limits_{-r}^r \left(1 - (x/r)^2\right)^{\frac{n-3}{2}} dx}.$$

Фиксируя $r = 1$ и устремляя n к бесконечности, получите

⁸ Из любой вершины можно добраться в любую другую по ребрам. Словосочетание «почти наверное» означает, что вероятностная мера тех графов, для которых это не так, стремится к нулю с ростом n .

$$P[-\delta, \delta] \sim 1 - \sqrt{\pi/2} \exp(-\delta^2 n/2).$$

В статистической физике

$$\sum_{i=1}^n V_i^2 = \frac{2E_n}{m} \sim n.$$

Поэтому если известно, что вектор скоростей молекул газа равномерно распределен по поверхности постоянной энергии,⁹ то для того чтобы найти (следуя Максвеллу) распределение компонент вектора скорости, скажем, V_1 , нужно осуществить термодинамический скейлинг (термодинамический предельный переход) $n \rightarrow \infty$, $r = \sigma n^{1/2} \rightarrow \infty$:

$$P[a, b] = \frac{\int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx}{\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

Таким образом, получаем нормальный закон распределения Максвелла в статистической физике.

Литература

1. Зорич А. В. Математический анализ задач естествознания. М.: МЦНМО, 2008. С. 48–56.

Задача к приложениям Е. В. Гасниковой, А. В. Колесникова, А. М. Райгородского (изопериметрическое неравенство и принцип концентрации меры; П. Леви, 1919). Число μ_f называют медианой функции f , если

$$\mu(\vec{x} \in S_1^n : f(\vec{x}) \geq \mu_f) \geq 1/2 \text{ и } \mu(\vec{x} \in S_1^n : f(\vec{x}) \leq \mu_f) \geq 1/2,$$

где $\mu(d\vec{x})$ – равномерное мера на единичной сфере S_1^n в \mathbb{R}^n . Пусть A – измеримое (борелевское) множество на сфере S_1^n . Через A_δ будем обозначать δ -окрестность множества A на сфере S_1^n . Предположим теперь, что в некотором царстве, расположенному на S_1^3 , царь предло-

⁹ Равномерное распределение на поверхности постоянной энергии возникло из-за того, что инвариантной (и предельной по эргодической гипотезе) мерой для гамильтоновой системы будет как раз равномерная мера Лиувилля (фазовый объем сохраняется). Поскольку выполняется закон сохранения энергии, то система «живет» на поверхности постоянной энергии. Следовательно, носитель инвариантной меры сосредоточен именно на этой поверхности.

жил царице Дионе построить «огород с заданной длиной забора». Царица хочет, чтобы её огород при заданном периметре имел наибольшую площадь. Таким образом, царице надо решить изопериметрическую задачу (такие задачи обычно рассматриваются в курсах вариационного исчисления). Решение этой задачи хорошо известно – «круглый огород». Для нас же полезно рассмотрение двойственной задачи, имеющей такое же решение: при заданной площади огорода спроектировать его так, чтобы он имел наименьшую длину забора, его ограждающего. Используя решение обобщения двойственной задачи на случай $n \geq 3$, покажите, что если $\mu(A) \geq 1/2$, то

$$\mu(A_\delta) \geq 1 - \sqrt{\pi/2} \exp(-\delta^2 n/2).$$

Пусть теперь на S_1^n задана функция с модулем непрерывности

$$\omega_f(\delta) = \sup \{ |f(\bar{x}) - f(\bar{y})| : \rho(\bar{x}, \bar{y}) \leq \delta, \bar{x}, \bar{y} \in S_1^n \}.$$

Тогда

$$\mu(\bar{x} \in S_1^n : |f(\bar{x}) - \mu_f| \geq \omega_f(\delta)) \leq \sqrt{\pi/2} \exp(-\delta^2 n/2).$$

Можно показать, что при весьма естественных условиях медиана асимптотически близка к среднему значению (математическому ожиданию). Аналогичное неравенство можно получить (М. Талагран, 1996), например, для модели случайных графов (Эрдёша–Рényи). И исследовать плотную концентрацию около среднего значения различных функций на случайных графах: число независимости, хроматическое число и т.п.

Литература

1. Ledoux M. Concentration of measure phenomenon. Providence, RI, Amer. Math. Soc., 2001 (Math. Surveys Monogr. V. 89).
2. Алон Н., Спенсер Дж. Вероятностный метод. М.: Бином, 2007.

Задача к приложениям Е. В. Гасниковой, А. В. Колесникова, А. М. Райгородского (изопериметрическое неравенство Талаграна и его приложения; М. Талагран, 1996). а)* Пусть заданы множества Ω_i , $i = 1, \dots, n$ элементарных исходов. На этих множествах заданы вероятностные меры P_i , $i = 1, \dots, n$. Положим

$$\Omega = \prod_{i=1}^n \Omega_i, P = \prod_{i=1}^n P_i.$$

Введём взвешенную метрику Хэмминга:

$$d_\alpha(\vec{x}, \vec{y}) = \sum_{x_i \neq y_i} \alpha_i \sqrt{\sum_{i=1}^n \alpha_i^2}$$

и определим $d_\alpha(\vec{x}, A) = \min_{\vec{y} \in A} d_\alpha(\vec{x}, \vec{y})$, $\rho(\vec{x}, A) = \sup_{\vec{\alpha} \in \mathbb{R}^n} d_\alpha(\vec{x}, A)$. Пусть $A \subset \Omega$. Определим t -окрестность ($t \geq 0$) множества A по формуле

$$A_t = \left\{ \vec{x} \in \Omega : \rho(\vec{x}, A) \leq t \right\}.$$

Покажите, что тогда справедливо следующее неравенство:

$$P(A)(1 - P(A_t)) \leq \exp(-t^2/4).$$

б)** Пусть в сельском районе, имеющем форму квадрата со стороной 1, находится n домов ($n \gg 1$), размерами которых можно пренебречь по сравнению с линейным размером района. Будем считать, что при строительстве домов застройщик случайно (согласно равномерному распределению $R[0, 1]^2$) и независимо выбирал их местоположения. Почтальону необходимо обойти все n домов ровно по одному разу (от любого дома почтальон может направиться к любому другому по прямой). Обозначим через TSP длину наикратчайшего из таких путей (кратчайший гамильтонов путь). Используя п. а), покажите, что найдутся такие постоянные $c > 0$ и $\beta > 0$, не зависящие от n , что

$$P(|TSP - E[TSP]| \geq t) \leq \exp(-t^2/(4c)), \text{ где } E[TSP] \sim \beta\sqrt{n}.$$

в)*** Пусть в условиях п. б) требуется построить систему дорог минимальной суммарной длины *SteinerTree*, по которой можно было бы добраться из любого дома в любой другой (дерево Штейнера с минимальной суммарной длиной ребер). Получите неравенство о плотной концентрации с.в. *SteinerTree* в окрестности своего математического ожидания, аналогичное неравенству п. б). Как себя асимптотически ведёт $E[\text{SteinerTree}]$ при $n \rightarrow \infty$?

Литература

1. Ledoux M. Concentration of measure phenomenon. Providence, RI, Amer. Math. Soc., 2001 (Math. Surveys Monogr. V. 89).
2. Алон Н., Спенсер Дж. Вероятностный метод. М.: Бином, 2007.
3. Dubhashi D. P., Panconesi A. Concentration of measure for the analysis of randomized algorithms. Cambridge University Press, 2009.
4. Gromov M. Metric Structures for Riemannian and Non-Riemannian Spaces. With Appendices by M. Katz, P. Pansu, and S. Semmes. Modern Birkhäuser Classics, 2001; Chapter 3₂₊.

Задача Штейнера и задачи на графах транспортных сетей

Е. Г. Молчанов

(при участии Р. А. Гимадеева, Ю. В. Дорна,
С. П. Тарасова, П. А. Шишкина)

Приведенные задачи разделяются на два типа: задачи, связанные с задачей Штейнера на плоскости, и задачи на транспортных графах. Последние обычно являются (возможно, нетривиальными) случаями, разновидностями и усложнениями нескольких стандартных задач: задачи поиска минимального остовного дерева, кратчайшего пути, максимального потока, а также транспортной задачи. Этот задачный раздел будет состоять из нескольких частей, во главе каждой из которых будет одна из таких «стандартных» задач, далее – задачи, которые связаны с первоначальной «стандартной» задачей.

Если в задачах требуется оценить время работы программы, будем считать, что она выполняется на компьютере, совершающем порядка 10^{10} операций с плавающей точкой в секунду. Остальные параметры этого компьютера (например, оперативную память) будем считать неограниченными.

Часть 1. Минимальная сумма расстояний

Задача 1. Согласно плану в районе строятся несколько домов, в каждом из которых будет проживать определенное количество человек. Нужно построить парковку, чтобы сумма расстояний, пройденных каждым человеком до парковки, была минимальной.

- а)** Где нужно построить парковку, если есть три одинаковых дома?
- б)** Докажите, что если все дома не лежат на одной прямой, то у задачи ровно одно решение.

Задача 2. Для решения задачи 1 о минимальной сумме расстояний можно воспользоваться следующей гамильтоновой системой: на плоский фанерный лист кладется карта района и в центрах домов про сверливаются маленькие дырочки. Через дырочки пропущены нерастяжимые нити, к которым прикреплены грузики с весом, пропорциональным количеству жителей в соответствующем доме. Начала всех нитей завязаны в один узел. Будем считать узел достаточно большим: он не будет пролезать в дырочки.

а) Докажите, что (при условии, что все точки не лежат на одной прямой), если «отпустить» все грузики, узел застрянет в точке с минимальной суммой.

б) Может ли узел застрять в вершине?

в) На основе описанной гамильтоновой системы предложите итерационный процесс решения задачи о минимальной сумме расстояний и оцените его скорость сходимости.

Задача 3. Предположим теперь в задаче 1, что от домов до парковки можно строить только «вертикальные» и «горизонтальные» дорожки (т.е. дорожки направления юг–север и запад–восток).

а) Где теперь нужно строить парковку, чтобы сумма расстояний до домов по дорожкам была минимальная.

б) Может ли у пункта а) быть не единственное решение? Как в общем случае выглядит область с минимальной суммой расстояний?

Указание. Сведите эту задачу к задаче о поиске минимального расстояния в одномерном случае (все дома лежат на одной прямой)

Часть 2. Минимальное остовное дерево

Задача 4. N точек-пунктов (вершин графа) соединены между собой дорогами. Каждая дорога соединяет две точки и имеет некоторую длину. Известно, что из любой такой точки до любой другой можно добраться по дорогам. Нужно реконструировать некоторое количество дорог, так чтобы от любой точки можно было добраться до любой другой только по реконструированным дорогам, и чтобы сумма длин дорог была минимальной.

Иными словами, в графе со взвешенными ребрами нужно найти минимальное остовное дерево.

а) Покажите, что при решении этой задачи можно считать, что произвольные две точки соединены дорогой (т.е. *граф полный*).

б) Какую сложность имеют стандартные алгоритмы Прима и Краскала; как эта сложность зависит от способов хранения информации, нужной для алгоритмов (например, матрицы расстояний).

в) Оцените время работы стандартных алгоритмов, если количество точек-вершин – 1000.

Задача 5.

а) В задаче 4 проектировщик решил искать нужные дороги среди гамильтоновых циклов, т.е. среди маршрутов, выходящих из какой-то

вершины и последовательно проходящих через все остальные вершины. Покажите, что найденное проектировщиком решение может в любое количество раз по длине превосходить минимальное (уточнить, что вкладывается в слово «минимальное»).

6) Оцените время поиска гамильтонова цикла. Примите количество вершин равным 1000 (100, 10). Сравните его с временем поиска минимального оствовного дерева.

Задача 6. Дорожная сеть города была построена следующим образом: каждый год строился новый микрорайон в последовательности, предусмотренной генпланом, и он соединялся дорогой с тем микрорайоном, до которого расстояние было наименьшим. После постройки города микрорайоны были связаны между собой системой дорог, образующей дерево. Будем считать, что расстояния между каждыми двумя микрорайонами известно до постройки и никаких географических ограничений на эти расстояния нет.

Сумма длин всех построенных дорог зависит от того, в каком порядке были построены микрорайоны. Докажите, что отношение максимально возможной суммы к минимально возможной может быть сколь угодно большим числом.

Задача 7. В некоторых местах построены несколько микрорайонов и несколько станций метро. Будем считать, что от любой станции можно на метро добраться до любой другой. Длины возможных дорог от микрорайонов до станций метро и между микрорайонами известны. Нужно построить минимальную по суммарной длине систему дорог так, чтобы из любого города можно было добраться до любого другого, возможно, воспользовавшись метро.

Как свести эту задачу к задаче 4 о минимальном оствовном дереве?

Указание. Будем считать, что от любой станции метро до любой другой можно добраться за нулевое время. Тогда, считая все станции метро одной вершиной, определите расстояния от метро до микрорайонов нужным образом.

Часть 3. Задача Штейнера

Задача 8. Несколько городов на плоскости нужно соединить дорогами так, чтобы из любого города можно было бы доехать до любого другого по построенными дорогам, и суммарная длина построенных до-

рог была бы минимальна. Расстояния считаются по обычной, Евклидовой метрике, при построении дорог можно строить дополнительные узлы – пересечения нескольких дорог.

а) Решите задачу для четырёх городов, расположенных в вершинах квадрата. Объясните, почему найденное решение не обладает всеми теми симметриями, которыми обладает исходная постановка задачи.

б) Докажите, что в оптимальной системе дорог в любом «вспомогательном» узле должны сходиться ровно три дороги под углами в 120° друг к другу.

в) Решите задачу для десяти городов, расположенных в вершинах правильного 10-угольника.

Задача 9. Сетью Штейнера называется систем прямых дорог, такая, что любая дорога соединяет напрямую какие-то две точки, где точками могут быть города, а также вспомогательные узлы, обладающая следующими свойствами:

1. В каждом вспомогательном узле сходятся ровно три дороги под углами в 120° друг к другу.
2. Из каждого города должно выходить не более трех дорог; если из города выходят две дороги – они должны образовывать угол не менее 120° , если три – все образованные ими углы должны быть по 120° .

а) Обязана ли такая сеть обладать минимальной суммой расстояний среди всех возможных сетей дорог (таких, что из любого города можно доехать до любого другого)?

б) Покажите, что сеть с минимальной суммой расстояний нужно исключить среди возможных сетей Штейнера.

Задача 10. Проектировщик соединил некоторые пары городов на плоскости прямыми дорогами так, чтобы из любого города в любой другой можно было бы доехать по построенным дорогам, и суммарная длина построенных дорог при этом была бы минимальной. Будем считать, что все дороги пересекаются вне города на разных уровнях – съезжать с одной дороги на другую можно только в городах.

а) Докажите, что в оптимальном решении задачи 9 не будет дорог, пересекающихся вне городов.

б) Докажите, что решение задачи 9 по суммарной длине построенных дорог не более чем в два раза превосходит решение задачи 8.

Замечание. Отношение минимального оствового дерева (веса ребер – расстояния на плоскости) к минимальной сети Штейнера назы-

вается числом Джильберта–Поллака. В п. б) предстоит доказать, что это отношение не превосходит числа 2. Это грубая оценка сверху, точная оценка — $\sqrt{3}/2$.¹⁰

в) Приведите пример, когда оценка $\sqrt{3}/2$ достигается.

Замечание. Построение минимальной сети Штейнера является *NP*-трудной задачей, поэтому нахождение минимального остовного дерева среди графа городов (вершины — города, расстояния между парами вершин считаются по обычной, Евклидовой метрике) является полиномиальным *приближенным* решением задачи Штейнера с коэффициентом приближения $\sqrt{3}/2$ (найденная длина дорог минимального остовного дерева не более чем в $\sqrt{3}/2$ превосходит оптимальное решение).

Часть 4. Кратчайший путь

Задача 11. Пусть есть некоторая система дорог, каждая из которых соединяет 2 пункта-вершины, причем для каждой дороги известно время движения по этой дороге. Требуется доехать из пункта А в пункт Б за минимальное время. *Иными словами в графе со взвешенными ребрами требуется найти кратчайший путь из одной вершины в другую.*

а) Оцените сложность алгоритма Дейкстры и время его работы, если количество вершин-пунктов равно 1000.

б) Будем считать, что время движения из одного пункта до другого зависит от направления (т.е. *граф — ориентированный*). Как изменится время решения пункта а)?

Замечание. Разумеется, при поиске кратчайшего расстояния на реальной сети дорог всегда рассматривают ориентированные графы. В последующих задачах этого параграфа, если неуказано обратное, графы будут ориентированными.

Задача 12. В задаче 11 будем считать, что время проезда по определенному участку пути зависит от того, в какое время вы въехали на этот участок. Для упрощения время проезда каждого участка мы будем считать целым неотрицательным числом $T_{ij}(t_0)$: мы движемся от i -го перекрестка к j -му, начиная движение в момент времени t_0 (очевидно, t_0 также целое неотрицательное).

¹⁰ Хотя и принято считать, что «точность» этой оценки доказана около 20 лет назад, в известном нам оригинальном доказательстве имеются «лакуны».

Как свести эту задачу к задаче 11 о кратчайшем пути в графе, вебса ребер которого не изменяются со временем?

Указание. Постройте новый граф, сопоставив одной вершине исходного графа несколько вершин, каждая из которых отвечает своему времени.

Задача 13. *M* автомобилям, первоначально находящимся в каких-то пунктах (вершинах графа), необходимо встретиться в одном месте (может быть, на дороге, соединяющей две вершины) и передать друг другу важную информацию. К сожалению, машинам негде припарковаться, поэтому останавливаться не разрешено, но разрешено поворачивать или разворачиваться в любом направлении в каждой вершине. Будем считать, что все машины движутся с одинаковыми скоростями, а все дороги, соединяющие некоторые пары перекрестков, требуют единицу времени для их прохождения.

Найдите минимальное время, необходимое для встречи.

Задача 14. Из пункта А в пункт Б (вершин транспортного графа) нужно добраться с помощью общественного транспорта, возможно, с пересадками. Каждый вид общественного транспорта ходит от какого-то одного пункта-вершины до какого-то другого строго по расписанию; расписание представляет собой таблицу времен отправлений из начального пункта и соответствующих им времен прибытий в конечный. Пересадка на другой вид транспорта в промежуточном пункте возможна, если время отправления превосходит время прибытия в этот пункт на предыдущем виде транспорта. Будем считать, что ровно в 12:00 мы оказались в пункте А. Требуется найти минимальное время, за которое мы сможем добраться до пункта Б, пользуясь общественным транспортом.

Приведите алгоритм решений этой задачи (например, метод ветвей и границ) и оцените сложность его работы.

Задача 15. В задаче 11 будем считать, что на некоторых вершинах могут быть установлены ограничения, запрещающие поворот с одной дороги на другую в этой вершине.

Как свести эту задачу к задаче 11 о минимальном графе, в котором не существует запрещенных поворотов.

Указание. Как и в задаче 12, нужно превратить одну исходную вершину в несколько.

Замечание. См. ниже задачу «Преобразования запрещенных маневров», являющуюся обобщением этой задачи.

Задача 16. Пусть в условиях задачи 11 кратчайший путь ровно один. Требуется найти путь, который будет являться вторым по величине длины (времени прохождения).

Задача 17. В городе были построены дороги, соединяющие некоторые перекрестки (вершины графа транспортной сети). Длина каждой дороги известна. Строителям для отчета требуется найти кратчайшие пути между каждой парой перекрестков.

а) Оцените сложность алгоритма и время работы соответствующей программы (1000 перекрестков), если разрешается пользоваться только алгоритмом Дейкстры поиска кратчайшего расстояния (правда, много раз).

б) Предложите более быстрый алгоритм (например, алгоритм Флойда–Уоршала) и также оцените его время работы при тех же условиях.

в) В каком смысле поиск кратчайших путей (алгоритмом Флойда) эквивалентен вычислению в идемпотентном полуполе (см. п. 2.1.3 главы 2) некоторой степени (укажите какой степени и предложите эффективный способ возведения в эту степень) матрицы, заполненной длинами всевозможных ребер–дорог? Как действовать, если какие-то вершины не соединены ребром?

Задача 18. Сопоставим каждому перекрестку (вершине графа транспортной сети) некоторое число – потенциал P_i и определим новое расстояние (время прохождения) между перекрестками i и j :

$$New_{ij} = Old_{ij} + P_j - P_i,$$

где Old_{ij} – «старое» расстояние между вершинами i и j .

а) Докажите, что кратчайший путь в графе с «новыми» расстояниями совпадает с кратчайшим путем в графе со «старыми» расстояниями.

б) Какие потенциалы нужно расставить в вершины, чтобы алгоритм Дейкстры нашел нужный кратчайший путь, не добавляя в множество рассмотренных вершин «лишние» вершины, т.е. не лежащие на кратчайшем пути?

Задача 19. В городе были выделены несколько перекрестков (вершин графа транспортной сети) и были подсчитаны все расстояния от выделенных перекрестков до всех остальных.

а) Как, используя эту информацию, оценить снизу кратчайшее расстояние между любыми двумя перекрестками (вершинами)?

б) Какие перекрестки (вершины) мы должны выделить (их количество заранее оговорено), чтобы оценка из п. а) была более точной?

Часть 5. Максимальный поток

Задача 20 (О максимальном потоке). Будем считать, что из пункта А в пункта Б едет (возможно, через другие пункты) непрерывный поток автомобилей величиной X автомобилей в единицу времени. Также будем считать, что у каждой дороги, соединяющей два пункта, пропускная способность ограничена числом c_{ij} машин в единицу времени. Необходимо найти, какой максимальный поток можно запустить из пункта А в пункт Б с учетом пропускных способностей дорог.

а) Запишите эту задачу, как задачу линейного программирования, напишите двойственную к ней задачу.

Указание. Считайте граф ориентированным. Используйте матрицу инциденций графа.

б) Пусть c_{ij} – неотрицательные целые числа. Покажите, что в оптимальном решении задачи поток, проходящий через каждую дорогу, будет также целым.

Указание. Используйте вид матрицы инциденций графа (в каждом столбце только два числа не равны нулю; эти числа: 1 и минус 1).

в) Как при доказательстве целочисленности решения (см. пункт б)) можно использовать теорему Гофмана–Краскала [6] о том, что полиэдр, задаваемый ограничениями вида $A\vec{x} \leq \vec{b}$, $\vec{x} \geq 0$ с абсолютно унимодулярной матрицей A и произвольным вектором \vec{b} (таким, что система $A\vec{x} \leq \vec{b}$, $\vec{x} \geq 0$ совместна), имеет все вершины с целочисленными компонентами?

Замечание. Отметим, что по теореме Пуанкаре [5], матрица инциденций произвольного ориентированного графа абсолютно унимодулярна, т.е. определитель любой её квадратной подматрицы равняется одному из трех чисел: минус 1, 0, 1.

г) Оцените сложность алгоритма Форда–Фалкерсона (c_{ij} – неотрицательные целые числа) и время работы соответствующей программы для поиска максимального потока (количество пунктов – 1000).

д) Сойдется ли за конечное время к оптимальному решению алгоритм Форда–Фалкерсона, если c_{ij} – неотрицательные вещественные числа?

е) Модифицируйте алгоритм Форда–Фалкерсона для решения задачи

максимального потока при неотрицательных вещественных c_{ij} (например, алгоритмы Эдмонса–Карпа и Диница), оцените их сложность и время работы (количество вершин – 1000).

ж) Оцените пространственную и временную сложность алгоритма Карзанова нахождения максимального потока.¹¹ В чем заключаются преимущества алгоритма Карзанова, например, над алгоритмом Форда–Фалкерсона?

Задача 21. В контексте предыдущей задачи будем считать, что суммарный поток машин через вершину также ограничен.

Как эту задачу свести к задаче 20 поиска максимального потока?

Задача 22. Во вражеской стране некоторые города соединены прямыми дорогами. Партизаны хотят сорвать перевозку оружия от города А до города Б: для этого им нужно взорвать минимальное количество дорог так, чтобы от города А стало бы нельзя доехать до города Б.

а) Предложите алгоритм нахождения дорог, которые предполагается взорвать (*минимальный разрез*) и оцените его время работы (количество вершин – 1000).

б) Докажите, что количество взорванных дорог равняется максимальному потоку из вершины А в вершину Б (веса всех ребер – 1).

Задача 23. Чиновник хочет проинспектировать все дороги города. Для этого он составил маршрут, проходящий по каждой дороге хотя бы один раз. Будем считать, что чиновнику нужна единица времени, чтобы проинспектировать одну дорогу (проезжая дорогу несколько раз, он её соответствующее число раз и инспектирует). Необходимо найти минимальное время, нужное чиновнику для инспекции всех дорог (как минимум по разу).

Как свести эту задачу к задаче 20 о максимальном потоке?

Указание. Когда в графе нельзя найти Эйлеров цикл? Какие ребра нужно сделать двойными, чтобы Эйлеров цикл нашелся?

Задача 24. Некоторые объекты (подмножество вершин) в городе назовем стратегически-важными. Найдите минимальное по суммарной длине подмножество дорог, таких, что из каждого стратегически-

¹¹ Описание этого алгоритма имеется, например, в лекции М. Г. Фуругяна, прочитанной студентам второго курса ФУПМ МФТИ весной 2009 г.,

<http://www.intuit.ru/department/algorithms/algomodex/2/>.

важного объекта до каждого стратегически важного можно добраться по дорогам из выбранного множества.

а) Частным случаем каких вышеупомянутых задач является данная задача? Что известно про сложность решения частных случаев?

б) Докажите *NP*-полноту этой задачи путем сведения к ней задачи 3-SAT о выполнимости булевых формул в 3-конъюнктивной нормальной форме.

Замечание. Условие задачи 3-SAT см. в книге [16].

Задача 25 (Транспортная задача). Имеется m производителей и n потребителей некоторого товара, расположенных в узлах транспортной сети. Пусть $x_{ij} \geq 0$ обозначает количество продукта, перевозимого из i -го узла в j -й, $c_{ij} \geq 0$ – стоимость перевозки, $a_i \geq 0$ – объём производства в i -м узле, $b_j \geq 0$ – суммарную потребность – в j -м (при этом $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$). Поставим задачу о составлении плана перевозок минимизирующего общую стоимость перевозок от производителей к потребителям:

$$\sum_{i=1, j=1}^{m, n} c_{ij} x_{ij} \rightarrow \min, \quad \sum_{j=1}^n x_{ij} = a_i, \quad \sum_{i=1}^m x_{ij} = b_j.$$

а) Напишите двойственную задачу к вышеприведенной задаче.

б) Дополнительно введя сток и исток и пользуясь матрицей инцидентий, запишите задачу нахождения максимального потока минимальной стоимости и двойственную к ней.

в) Рассматривая эту задачу как задачу линейного программирования, получите оценку для времени работы известных вам алгоритмов в наихудшем случае и в среднем. Примите $m = n = 500$.

Задача 26. Рассматривается та же самая транспортная задача с целыми неотрицательными параметрами

$$a_i \in \mathbb{Z}, \quad i = 1, \dots, n; \quad b_j \in \mathbb{Z}, \quad j = 1, \dots, m.$$

а) Покажите, что решения x_{ij} задачи в классе неотрицательных действительных чисел будут целыми.

Указание. Запишите задачу нахождения максимального потока минимальной стоимости (см. задачу 25 б)), затем воспользуйтесь теоремой Гофмана–Краскала (см. задачу 20 в)).

б) Как нужно модифицировать алгоритм Форда–Фалкерсона нахож-

дения максимального потока для решения транспортной задачи? Оцените время работы получившегося алгоритма при $m = n = 500$.

в) Сойдется ли за конечное время модифицированный алгоритм из пункта а), если параметры задачи могут быть не целыми?

Замечание. Вообще говоря, нахождение целочисленных решений, как правило, значительно усложняет алгоритм, иногда делая задачу, решаемую за полиномиальное время, *NP*-трудной (например, в задаче размещения производства, см. [15]). Однако приведенная задача как раз является исключением из этого правила.

Задача 27. Пусть есть граф со взвешенными ребрами, причем веса ребер могут быть отрицательными числами.

а) Как в данном графе найти цикл, сумма длин ребер которого отрицательна?

б) Как нахождение таких циклов использовать при решении транспортной задачи 24?

в) Оцените время работы полученной программы при $m = n = 500$.

Часть 6. Построение оптимальной транспортной сети

Задача 28 (П. П. Бобрик, 2000). Распределение мест проживания населения по городу, имеющему форму квадрата со стороной A км и не имеющего дорог, равномерное. Условное распределение мест работы жителей (при условии, что зафиксировано место жительства) также равномерное (и, как следствие, не зависит от места жительства). Руководство города решило построить сеть дорог в виде квадратной решётки со стороной квадратной ячейки, равной $a = A/n$. Известно, что затратами на строительство сети дорог можно пренебречь по сравнению с последующими суммарными затратами на её поддержание. Пусть C_L – стоимость поддержания 1 км дороги (в одну сторону) в течение 1 года, C_T – стоимость 1 часа, потраченного одним человеком в дороге. Число жителей в городе N . Каждый человек $Q = 300$ раз в год направляется из дома на работу и обратно с работы домой. Скорость движения человека не по дороге v км/ч, скорость движения человека по дороге V км/ч ($V \gg v$), время, потраченное на прохождение одного перекрёстка, есть $t \geq 0$ мин.

а) Постройте целевую функцию, отвечающую за эффективность построенной транспортной сети.

б) Какое значение n «оптимально» выбрать руководству города?

Задача 29. В условиях предыдущей задачи было решено построить радиально-кольцевую систему дорог, состоящую из k колец и m радиусов.

а) Постройте целевую функцию, отвечающую за эффективность построенной транспортной сети.

б) При каких условиях эта сеть окажется эффективней оптимальной квадратной сети из предыдущей задачи?

Задача 30. В условиях задачи 28 был предложен альтернативный план транспортной сети – регулярная сеть дорог, с элементарной «ячейкой» в виде правильного (равностороннего) треугольника со стороной b (из расчета, что длина сети совпадает с «квадратной»). При этом оказалось, что время, потраченное на прохождение перекрестка в вершине графа транспортной сети, $t = m_i \tau$, пропорционально степени m_i этой вершины. Определить, при каком значении τ эффективность сетей совпадает (если такое τ существует).

Задача 31 (А. П. Буслاءв, 2001). В литературе имеется довольно большое количество различных характеристик графов транспортных сетей. Интересной задачей является исследование этих макрохарактеристик на различных типах случайных графов. Например, во второй части монографии [23] вводятся такие показатели «хорошести» графа транспортной сети (отметим, что терминология пока не устоялась, поэтому в других местах аналогичные термины могут иметь иное содержание):

$$r_G(p) = \frac{N - E_p \xi_G}{N - 1} \text{ – надежность,}$$

где N – число вершин в графе G , ξ_G – число связных компонент в графе \tilde{G}_p , полученном из исходного графа G путем случайного и независимого разыгрывания его ребер (с вероятностью $p > 0$ ребро оставляют и с вероятностью $1 - p$ ребро стирают), $E_p \xi_G$ – математическое ожидание с.в. ξ_G ;

$$s_G(p) = \frac{E_p \eta_G - 1}{N - 1} \text{ – устойчивость,}$$

где η_G – размер (число вершин) максимальной связной компоненты в графе \tilde{G}_p ;

$$rt_G(p) = \frac{1}{(N-1)N} \sum_{i,j \neq i} E_p \rho_{ij}^G - \text{доступность},$$

где ρ_{ij}^G – длина кратчайшего пути в графе \tilde{G}_p , ведущего из вершины i в вершину j (доопределим $\rho_{ij} = 1$ (i – мнимая единица), если из вершины i нельзя добраться по графу \tilde{G}_p в вершину j).

а) Исследуйте качественно зависимость приведенных показателей от p для графов транспортной сети со структурой задачи 28 и 30 (считайте, что $N \gg 1$).

б) Сравните эффективность транспортных сетей задач 28 и 30 с помощью приведенных показателей.

в) (гигантская компонента; см. приложение А. М. Райгородского) Покажите, что если взять в качестве G полный граф на N вершинах, то $\tilde{G}_p \equiv G(N, p)$ (модель Эрдёша–Ренъи) и для любого $c > 1$ найдётся такое $0 < \gamma_c < 1$, что

$$P(s_G(c/N) \geq \gamma_c) \xrightarrow{n \rightarrow \infty} 1.$$

Задача 32 (показатель качества графа транспортной сети; Уотс–Строгатц, 1998)*. Рассматривается такой ориентированный взвешенный граф транспортной сети, что из любой его вершины можно добраться по ребрам (с учетом ориентации ребер) в любую другую. Обозначим через d_{ij} – длину кратчайшего пути по рассматриваемому графу транспортной сети из вершины i в вершину j . Показателем качества исследуемого графа назовем величину:

$$e = (n-1)^{-1} n^{-1} \sum_{i, j \in V: i \neq j} d_{ij}^{-1},$$

где $V = \{1, \dots, n\}$ – множество вершин графа. Оцените значение этого показателя для графа транспортной сети из задачи 28, считая $n \gg 1$.

Задача 33 (нахождение минимума функционала среднего расстояния; Е. О. Степанов, 2005).** На множестве всех замкнутых связных множеств $\Sigma \subset \mathbb{R}^2$ хаусдорфовой размерности 1 ($H^1(\Sigma) < \infty$), удовлетворяющих ограничению на длину, $H^1(\Sigma) \leq l$, требуется найти решение Σ^* следующей задачи:

$$F_\varphi(\Sigma) = \int_{\mathbb{R}^2} A(\text{dist}(\vec{x}, \Sigma)) d\varphi(\vec{x}) \rightarrow \min_{\Sigma: H^1(\Sigma) \leq l},$$

где $\varphi(\vec{x})$ – мера, отвечающая распределению населения в окрестности точки с координатой \vec{x} , число $A(d)$ задает затраты жителя на проезд на расстояние d . Транспортная сеть Σ , таким образом, проектируется исходя из соображений минимизации полных затрат населения на достижение сети из мест проживания. Покажите, что при весьма общих условиях решение Σ^* этой задачи существует, не содержит петель, имеет конечное число концевых точек и точек ветвления, а также обладает некоторыми свойствами регулярности (см. раздел, посвященный задаче Штейнера). Предложите эффективный способ численного поиска Σ^* .

Литература

К частям 1, 3:

1. *Hwang F. K., Richards D., Winter P.* The Steiner tree problem. Elsevier Science Publishers, 1992.
2. *Гордеев Э. Н., Тарасцов О. Г.* Задача Штейнера. Обзор // Дискретная математика. 1993. Т. 5. № 2. С. 3–28. <http://www.mathnet.ru/>¹²
3. *Иванов А. О., Тужилин А. А.* Теория экстремальных сетей. Москва–Ижевск: Институт компьютерных исследований, 2003.
4. *Протасов В. Ю.* Максимумы и минимумы в геометрии. № 31. М.: МЦНМО, Библиотечка «Математическое просвещение», 2005. http://www.mccme.ru/mnmf-lectures/books/books/book_31.pdf

К частям 2, 4, 5:

5. *Берж К.* Теория графов и ее приложения. М.: ИЛ, 1962.
6. *Гвишiani А. Д., Гурвич В. А.* Динамические задачи классификации и выпуклое программирование в приложениях. М.: Наука, 1992.
7. *Stoer Mechthild, Wagner Frank* A Simple Min-Cut Algorithm // Journal of the ACM. 1997. V. 44. № 4. P. 585–591.
8. *Разборов А. А.* О сложности вычислений // Матем. просв. 1999. № 3. <http://www.mccme.ru/free-books/matpros4.html>

¹² Обратим внимание, что на этом электронном ресурсе открыт свободный доступ к полным текстам статей ряда ведущих российских научных журналов (Успехи математических наук, Математический сборник, Математические заметки, Функциональный анализ, Известия РАН, Дискретная математика, Математическое моделирование, Журнал вычислительной математики и математической физики и др.).

9. Смайл С. О проблемах вычислительной сложности // Матем. просв. 2000. № 4. <http://www.mccme.ru/free-books/matpros5.html>
10. Schrijver Alexander On the history of the transportation and maximum flow problems // Math. Program., Ser B. 2002. V. 91. P. 437–445. <http://oai.cwi.nl/oai/asset/10084/10084A.pdf>
11. Вяльи М. Н. Линейные неравенства и комбинаторика. М.: МЦНМО, Летняя школа «Современная математика», 2003. <http://www.mccme.ru/free-books/dubna/vyalyi.pdf>
12. Зыков А. А. Основы теории графов. М.: Вуз. книга, 2004.
13. Diestel R. Graph Theory. Electronic Edition, NY: Springer, 2005.
14. Goldberg A. V., Harrelson C. SODA '05 Proceedings of the sixteenth annual ACM-SIAM symposium on discrete algorithms. USA Society for Industrial and Applied Mathematics Philadelphia. 2005. P. 156–165.
15. Кормен Т. Х., Лейзерсон Ч. И., Ривест Р. Л., Штайн К. Алгоритмы: Построение и анализ. М.: Вильямс, 2005.
16. Кузюрин Н. Н., Фомин С. А. Эффективные алгоритмы и сложность вычислений. М.: МФТИ, 2007. <http://discopal.ispras.ru/ru.lectures-mipt.htm>
17. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. Лекции по теории графов. М.: УРСС, 2009.
18. Annual IEEE Symposium on Foundations of Computer Science, 1 – 51. <http://theory.stanford.edu/focs2010/>
19. <http://e-maxx.ru/algo/>; <http://www.machinelearning.ru/>
20. Шень А. Х. Программирование: теоремы и задачи. М.: МЦНМО, 2011.

К части 6:

21. Стенбринк П. А. Оптимизация транспортных сетей. М.: Транспорт, 1981.
22. Бобрик П. П. Сравнение эффективностей треугольной и квадратичной регулярных транспортных сетей // Транспорт: наука, техника, управление. М.: Изд-во ВИНИТИ, 2000. № 7.
23. Луканин В. Н., Буслаев А. П., Трофимов Ю. В., Яшина М. В. Авто-транспортные потоки и окружающая среда. Ч. 1, 2. М.: ИНФРА-М, 1998, 2001.
24. Latora V., Marchiori M. Efficient behavior of small-world networks // Phys. Rev. Letters. 2001. V. 87. № 19. http://www.w3.org/People/Massimo/papers/2001/efficiency_prl_01.pdf
25. Степанов Е. О. Математические модели оптимизации транспортных сетей. Санкт-Петербург: СПбГУ ИТМО, 2005.

Преобразования запрещенных маневров (от Яндекс-пробки)

А. И. Верещагин, В. Б. Гольдштейн, И. И. Колесниченко,
М. В. Левин, Андрей А. Петров

Постановка задачи

При нахождении маршрутов проезда по городу важно учитывать не только взаимное расположение дорог, но и правила дорожного движения. Картографические компании собирают информацию о расположении дорог и дорожных знаках. Эта информация предоставляется отдельно, однако для маршрутизации обязана учитываться вместе. Граф дорог состоит из перекрестков и улиц их соединяющих. Встречаются улицы с односторонним и двусторонним движением, а также улицы, по которым проезд машин запрещен вовсе. Однако из-за запрещающих знаков на некоторые улицы можно выехать не всегда. Простейшим примером служит запрещенный левый поворот. Несмотря на то, что по улицам АМВ и СМД разрешен проезд в обоих направлениях, нельзя проехать по ним последовательно. То есть проезд по пути АМС и ВМД запрещен. Обобщая этот пример, назовем запрещенным маневром последовательность дорог, проезд по которым от начала и до конца запрещен. Запрещенные маневры могут состоять как из двух ребер (запрет левого поворота), так и из 4–5 дорог (сложный выезд на трассу).

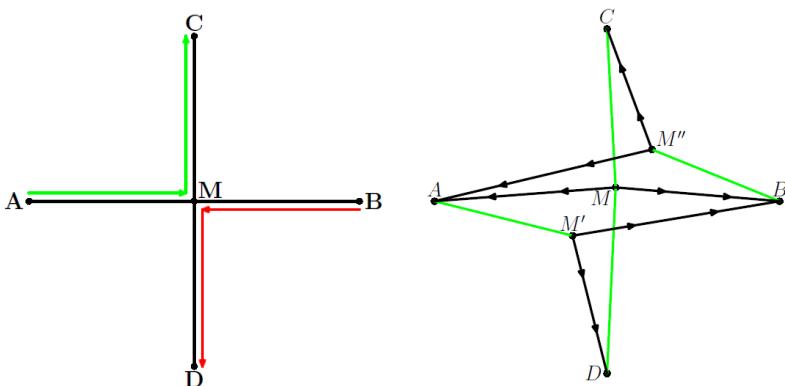


Рис. 1

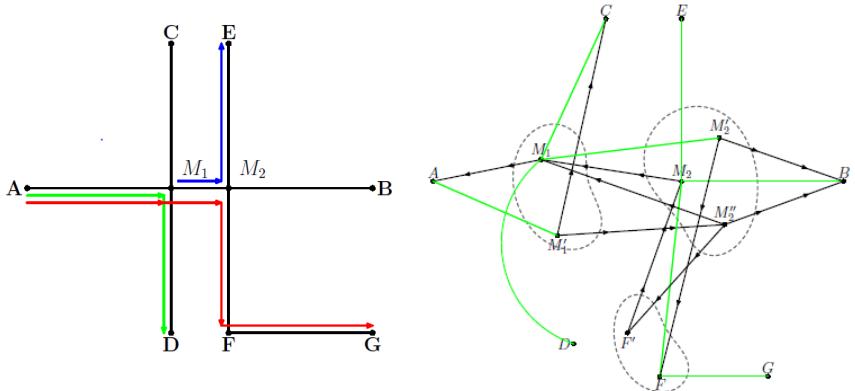


Рис. 2

Так на рис. 1 два запрещенных маневра:

$$P_1 = (AM, MC) \text{ и } P_2 = (BM, MD).$$

На рис. 2 три запрещенных маневра:

$$P_1 = (AM_1, M_1D) \text{ и } P_2 = (AM_1, M_1M_2, M_2F, FG) \text{ и } P_3 = (M_1M_2, M_2E).$$

Формальная постановка задачи

Задан ориентированный (для простоты в примерах будем рассматривать неориентированный граф, каждому ребру без стрелок на рис. 1, 2 соответствуют два ребра в разных направлениях) граф **G** без петель и набор запрещенных маршрутов **P**. Маршрут P_i – последовательность ребер графа. Запрет проезда по маршруту P_i означает запрет проезда по всему маршруту начиная с первого ребра и заканчивая последним. Проезд по любой части маршрута не запрещен. Требуется построить такой граф **H**, в котором не будет запрещенных маневров. Каждой вершине (ребру) графа **G** соответствует одна или несколько вершин (ребер) графа **H**. Каждому разрешенному пути (чередующаяся последовательность вершин и ребер) в графе **G** соответствует путь в графе **H**. Запрещенному пути в графе не должно соответствовать ни одного пути в графе **H**. Разрешенный путь в графе **G** – путь, не содержащий запрещенных маневров. Запрещенный путь в графе **G** – путь, содержащий хотя бы один запрещенный маневр.

Все вершины графа **G** имеют вполне конкретный географический смысл и находятся в некоторых различных точках земли. Различные

вершины графа **H** могут находиться в одной точке Земли, однако отражать разное состояние.

На рис. 1 знание, что машина находится в точке M , не достаточно для понимания, куда может ехать машина. В графе **H** вершине M будет соответствовать три вершины:

- 1) если мы приехали из вершины A ;
- 2) из вершины B ;
- 3) из другой вершины.

Рассмотрим граф **H** для графа, изображенного на рис. 2. Светлым обозначены двусторонние дороги. Вершина M'_1 означает, что машина находится в вершине M_1 и запрещены следующие пути: $\{M_1M_2E, M_1D, M_1M_2FG\}$, вершина M'_2 находится в M_2 и запрещен путь M_2E , M''_2 находится в M_2 и запрещены пути $\{M_2E, M_2FG\}$, F' находится в F и запрещен путь FG .

Используемые сокращения

m – метр, *c* – секунда, *km* – километр, *ч* – час, *мин* – минута
т.е. – то есть, *т.д.* – так далее, *т.п.* – тому подобное
др. – другие, *см.* – смотри, *п.* – пункт, *пп.* – пункты, *г.* – год
УДС – улично-дорожная сеть
ЗОС – загрязнение окружающей среды
LWR – Лайтхилл–Уизем–Ричардс
ATC – автотранспортное средство (автомобиль)
RRH – Риман–Ранкин–Гюгонио
E-условие – энтропийное условие О. А. Олейник
Г–Я – Гамильтон–Якоби, *Г–Я–Б* – Гамильтон–Якоби–Беллман
T.V. – total variation (полная вариация)
CTM – Cell Transmission Model (модель клеточных автоматов Даганзо)
IDM – Intelligent Driver Model (модель разумного водителя)
ПДД – правила дорожного движения
CA – Cellular Automata (клеточные автоматы)
СОДУ – система обыкновенных дифференциальных уравнений
КdФБ – Кортевег–де Фриз–Бюргерс
ДМ – Джениерал Моторс
F – free flow, *S* – synchronized flow, *J* – wide moving jam
СП – синхронизованный поток
ЛСП – локализованный синхронизованный поток
РСП – расширяющийся синхронизованный поток
МСП – мигрирующий синхронизованный поток
ОП – общей структурой плотного потока
ЕП – единой структурой плотного потока
ЕСП – единая структура синхронизованного потока
ЕОП – единая общая структура плотного потока
РОП – рассасывающаяся общая структура плотного потока
п.н. – почти наверное, *н.о.р.* – независимые одинаково распределенные
с.в. – случайная(ые) величина(ы), *д.м.ц.* – дискретная марковская цепь
E – Expected value (математическое ожидание), *D* – Dispersion (дисперсия)
Н.О.Д. – наибольший общий делитель
СЛАУ – система линейных алгебраических уравнений
TASEP – totally asymmetric exclusion process (полностью асимметричные процессы с запретами)
LCD – linearized chord diagram (линейная хордовая диаграмма)
 \mathbb{R} (**R**) – множество действительных чисел
 \mathbb{N} (*N*) – множество натуральных чисел (1, 2, 3, …)
 \mathbb{Z} (**Z**) – множество целых чисел

Учебное издание

**ВВЕДЕНИЕ
В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
ТРАНСПОРТНЫХ ПОТОКОВ**

**Гасников Александр Владимирович
Кленов Сергей Львович
Нурминский Евгений Алексеевич
Холодов Ярослав Александрович
Шамрай Наталья Борисовна**

Приложения:

**Бланк Михаил Львович
Гасникова Евгения Владимировна
Замятин Андрей Андреевич и Малышев Вадим Александрович
Колесников Александр Викторович
Райгородский Андрей Михайлович**

Редакторы *B. A. Дружинина, И. А. Волкова*
Корректор *О.П. Котова, Л.В. Себова*

Подписано в печать 30.11.2010. Формат 60 × 84 ¹/₁₆.
Печать офсетная. Усл. печ л. 22,5. Уч.-изд. л. 22,0. Тираж 250 экз.
Заказ № ф-003.

Государственное образовательное учреждение
высшего профессионального образования
«Московский физико-технический институт (государственный университет)»
141700, Московская обл., г. Долгопрудный, Институтский пер., 9

Отдел автоматизированных издательских систем «ФИЗТЕХ-ПОЛИГРАФ»
141700, Московская обл., г. Долгопрудный, Институтский пер., 9

**Учебное пособие посвящено математическому моделированию транспортных потоков.
Излагается математический аппарат и некоторые физические концепции, которые могут пригодиться при создании (модернизации) комплексной интеллектуальной транспортной системы.**

С глубокой благодарностью дорогому А.А. Петрову



**Академик РАН Александр Александрович Петров
(03.02.1934 – 23.02.2011)**

ISBN 5-7417-0334-8

A standard linear barcode representing the ISBN 5-7417-0334-8. Below the barcode, the numbers 9785741703342 are printed vertically.