

ISSN 2072-6759



ТРУДЫ МФТИ

2010

Т. 2 № 4(8)

Научно-технический
журнал

ISSN 2072-6759-9



9 772072 675004

Специальный выпуск, посвященный математическому моделированию транспортных потоков
Редактор – вице-президент РАН академик Валерий Васильевич Козлов
Выпуск журнала подготовлен при поддержке
факультета управления и прикладной математики МФТИ

Высшая аттестационная комиссия Министерства образования и науки Российской Федерации
включила журнал «Труды МФТИ» в Перечень ведущих рецензируемых научных журналов и
изданий, в которых должны быть опубликованы основные научные результаты диссертаций на
соискание учёной степени доктора и кандидата наук.

Т Р У Д Ы
Московского
физико-технического
института
(государственного университета)

Научно-технический журнал
 Основан в 2008 году
 Том 2, № 4 (8) 2010 год
 Выходит 4 раза в год

Учредитель

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

Главный редактор

чл.-корр. РАН, ректор МФТИ Н.Н. Кудрявцев

Редакционный совет

<i>М.В. Алфимов — акад. РАН</i>	<i>А.С. Коротеев — акад. РАН</i>
<i>А.Ф. Андреев — акад. РАН</i>	<i>Н.А. Кузнецов — акад. РАН</i>
<i>О.М. Белоцерковский — акад. РАН</i>	<i>В.Л. Макаров — акад. РАН</i>
<i>С.Т. Беляев — акад. РАН</i>	<i>А.А. Петров — акад. РАН</i>
<i>Е.П. Велихов — акад. РАН</i>	<i>А.А. Солдатов — проф., д.ф.-м.н.</i>
<i>Ю.В. Гуляев — акад. РАН</i>	<i>Э.Е. Сон — чл.-корр. РАН</i>
<i>В.Г. Дмитриев — чл.-корр. РАН</i>	<i>В.Е. Фортвов — акад. РАН</i>
<i>В.П. Иванников — акад. РАН</i>	

Редакционная коллегия

<i>А.П. Алёхин — проф., д.т.н.</i>	<i>В.Е. Кривцов — доц., к.ф.-м.н.</i>
<i>Ю.М. Белоусов — проф., д.ф.-м.н.</i>	<i>А.Г. Леонов — проф., д.ф.-м.н.</i>
<i>А.С. Бугаев — акад. РАН</i>	<i>Д.С. Лукин — проф., д.ф.-м.н.</i>
<i>Ю.Н. Волков — доц., к.ф.-м.н.</i>	<i>С.С. Негодяев — к.т.н.</i>
<i>В.В. Вышинский — проф., д.т.н.</i>	<i>В.Г. Орлов — доц., к.ф.-м.н.</i>
<i>Э.М. Габидулин — проф., д.т.н.</i>	<i>И.Б. Петров — проф., д.ф.-м.н.</i>
<i>С.Н. Гаричев — с.н.с., д.т.н.</i>	<i>Е.С. Половинкин — проф., д.ф.-м.н.</i>
<i>А.Д. Гладун — проф., д.ф.-м.н.</i>	<i>Ю.А. Самарский — проф., к.ф.-м.н.</i>
<i>Е.В. Глухова — проф., д.т.н. —</i>	<i>Л.В. Стрыгин — к.ф.-м.н. —</i>
<i>заместитель главного редактора</i>	<i>ответственный секретарь редколлегии</i>
<i>И.Н. Грознов — доц., к.ф.-м.н.</i>	<i>П.А. Тодуа — проф., д.ф.-м.н.</i>
<i>А.Ю. Деревнина — доц., д.т.н.</i>	<i>М.Р. Трунин — д.ф.-м.н.</i>
<i>Г.Н. Дудин — проф., д.ф.-м.н.</i>	<i>Э.М. Трухан — проф., д.ф.-м.н.</i>
<i>Т.В. Кондранин — проф., д.ф.-м.н.</i>	<i>А.С. Холодов — чл.-корр. РАН</i>
<i>В.И. Кондауров — проф., д.ф.-м.н.</i>	<i>А.А. Шананин — проф., д.ф.-м.н.</i>

Редакция

Е.А. Болбот (начальник научно-исследовательской части МФТИ)
В.А. Дружинина (начальник редакционно-издательского отдела МФТИ)
А.В. Чудновский (технический редактор)

141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9
 Тел.: (495) 408-41-11, факс: (495) 408-58-22
<http://www.mipt.ru/nauka/trudy/>
zhurnaltrudy@mail.ru

**Proceedings
of Moscow Institute
of Physics and Technology
(State University)**

Journal of Science and Technology
Founded in 2008
Volume 2, N. 4 (8) 2010
The journal is published in four issues

Moscow Institute of Physics and Technology
(State University) (MIPT)

Editor-in-chief

rector of MIPT *N.N. Kudryavtsev*

Associative editors

M.V. Alfimov
A.F. Andreev
O.M. Belotserkovskii
S.T. Belyaev
E.P. Velichov
Yu.V. Gulyaev
V.G. Dmitriev
V.P. Ivannikov

A.S. Koroteev
N.A. Kuznetsov
V.L. Makarov
A.A. Petrov
A.A. Soldatov
E.E. Son
V.E. Fortov

Editorial Board

A.P. Alekhin
Y.M. Belousov
A.S. Bugaev
Y.N. Volkov
V.V. Vyshinsky
E.M. Gabidulin
S.N. Garichev
A.D. Gladun
E.V. Glukhova –
Deputy Editor-in-chief
I.N. Groznov
A.Y. Derevnina
G.N. Dudin
T.V. Kondranin
V.I. Kondaurov

V.E. Krivtsov
A.G. Leonov
D.S. Lukin
S.S. Negodyaev
V.G. Orlov
I.B. Petrov
E.S. Polovinkin
Y.A. Samarsky
L.V. Strygin –
Executive Secretary
P.A. Todua
M.R. Trunin
E.M. Trukhan
A.S. Kholodov
A.A. Shanarin

Editorial Staff

E.A. Bolbot, V.A. Druzhinina, A.V. Chudnovsky

Введение

Реально складывающаяся дорожно-транспортная ситуация в мегаполисах России требует комплексности решений не только в сфере развития транспортной инфраструктуры, но и в повседневном управлении транспортной системой для обеспечения надежности ее функционирования как в штатных, так и в кризисных ситуациях. Мероприятия по основным сферам обеспечения управления транспортным комплексом включают в себя следующие составляющие: организационно-управленческую, финансово-экономическую, архитектурно-градостроительную, научно-техническую, законодательную, информационную, социальную, учебную и воспитательно-просветительскую.

Фундаментальные проблемы и перспективы исследований в области организации движения автомобильного транспорта обсуждаются на ежемесячном семинаре «Научно-практические задачи развития автомобильно-дорожного комплекса России», проводимом с 2006 года под научно-методическим руководством РАН с участием представителей академической, вузовской и отраслевой науки.

Создание комплексной интеллектуальной транспортной системы мегаполиса может послужить той базой, платформой, на которой усилия фундаментальной и прикладной науки могут быть объединены для достижения единой цели — обеспечить устойчивое, комфортное и экологичное движение автотранспорта. В настоящем журнале представлены работы участников семинара и конференции РАН 2010 года по различным задачам и подходам, связанным с моделированием автотранспортных потоков.

Моделирование транспортных потоков начало формироваться в математическую дисциплину в пятидесятые годы прошлого века в связи с бурным развитием газовой динамики (обобщенные решения законов сохранения, устойчивые разностные схемы расчета решений). Тогда же появились первые макроскопические (гидродинамические) модели, в которых транспортный поток уподобляется потоку 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 (1955) транспортный поток уподобляется потоку сжимаемой жидкости и описывается законом сохранения количества автомобилей. При этом в модели постулируется существование функциональной зависимости (уравнения состояния) между интенсивностью потока автомобилей и плотностью. Эту зависимость часто называют фундаментальной диаграммой.

В последующие годы класс микро- и макромоделей был значительно расширен. В современном макроскопическом подходе (А. Эу и М. Раскль, 2000) транспортный поток 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 настоящем сборнике имеются две статьи (А.С. Холодов и др., Н.Н. Смирнов и др.), выполненные по этому направлению.

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). Если в таких моделях время течет дискретно и динамика движения автомобилей стохастическая (марковская), то их, как правило, называют моделями клеточных автоматов. В статье М.Л. Бланка продемонстрирован один из способов того, как с помощью простейших моделей клеточных автоматов можно математически строго получать правдоподобные макроскопические уравнения состояния транспортного потока (например, треугольную фундаментальную диаграмму).



Модели клеточных автоматов, к которым пришли западные исследователи К. Даганзо, К. Нагель и М. Шрекенберг в середине девяностых годов прошлого века при дискретизации непрерывных конструкций и создании компьютерных алгоритмов имитационного моделирования, еще ранее появились в работах специалистов по случайным блужданиям. Так, в 1980-х в работах Ю.К. Беляева и его учеников рассматривалась стохастическая модель движения поездов метро. Принципиальным вопросом в моделях клеточных автоматов для трафика является вопрос о том, что же такое клетка. Если клетка — это фиксированная часть полосы, которую за отведенное время преодолевает автомобиль, то получаем динамическую систему — дискретный аналог непрерывных моделей.

Если в качестве клетки рассматривать динамический габарит части автомобилей потока, например тех, которые по городу движутся с максимально разрешенной скоростью, то оставшаяся часть потока пытается с некоторой мерой просочиться сквозь законопослушных. Можно также рассматривать поведение участников движения как смесь коллективного, синхронизированного и индивидуального, случайного. Синхронизированное движение определяет динамический габарит, а индивидуальное поведение определяется частью дорожного полотна, которая остается после покрытия элементов потока их динамическими габаритами. Такой подход, в частности, означает, что явной зависимости между плотностью потока и скоростью, вообще говоря, нет. Эти потоки называются частично связными. Некоторые задачи в данном направлении рассматриваются в статье А.С. Бугаева и др.

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. Имеются модели промежуточные между кинетическими и гидродинамическими. Такой моделью, назовем ее газодинамической, пользуется коллектив, возглавляемый Б.Н. Четверушкиным. В работе Б.Н. Четверушкина и др. даны сравнительные результаты, полученные при моделировании потока с использованием газодинамической модели и модели клеточных автоматов.

Несмотря на то, что с момента появления первых фундаментальных работ прошло более полувека, по мнению ряда известных специалистов в области математического моделирования дорожного движения (Б. Кернер, К. Нагель, Х. Махмасани, М. Шрекенберг и др.), проблема образования предзаторных и заторных ситуаций еще до конца не изучена. Используя терминологию, предложенную Б. 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 различные модели: начиная от модели Лайтхилла–Уизема (А.А. Куржанский и др.), заканчивая моделями, в которых каждый водитель описывается своим вариационным принципом (И.А. Лубашевский и др.). Отметим также, что большинство исследователей ограничивается 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).

В статье С.Л. Кленова (коллеги Б. Кернера) приводится эмпирический базис 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, наблюдаемых в многочисленных экспериментах.

Математическая теория управления транспортными потоками сейчас активно развивается в работах школы, возглавляемой А.Б. Куржанским и П. Варайя (см. соответствующую статью в этом номере). Исходя из модели клеточных автоматов К. Даганзо (1994), которая состоит в применении схемы Годунова к модели Лайтхилла–Уизема с треугольной фундаментальной диаграммой, предлагается способ оптимального управления светофорами и въездами на магистралях в Калифорнии. Здесь стоит обратить внимание на соизмеримость грубости выбранной модели, качества имеющихся данных и простоты работы с этой моделью.

Одним из возможных подходов к моделированию является исследование усредненных показателей транспортной системы. Обгоны на многополосной дороге, очереди перед светофорами и многое другое можно описывать таким образом — о чем говорится в статьях Л.Г. Афанасьевой и Е.В. Булинской, а также А.А. Замятина и В.А. Малышева (в основе этих моделей 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 статье А.М. Райгородского 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).

Остановимся на проблеме целостного описания транспортного потока на полном графе транспортной сети. Будем считать, что есть информация о том, сколько людей живет и сколько рабочих мест есть в том или ином районе. В статьях П.П. Бобрика, А.В. Гасникова и Е.В. Гасниковой, Е.А. Нурминского и Н.Б. Шамрай приведены различные способы обоснования известной в приложениях энтропийной (гравитационной) модели А.Дж. Вильсона (1967) расчета матрицы корреспонденций. Матрица корреспонденций определяется как наиболее вероятное макросостояние, в окрестности которого и будет плотная концентрация стационарной меры «разумной» эргодической марковской динамики, порождающей изучаемую макросистему. Точнее говоря, эргодическая марковская динамика приводит на больших временах к стационарной пуассоновской 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 на множестве, заданном ограничениями в виде законов сохранения. Поскольку число ограничений, как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы базируются на решении двойственной задачи. Отметим, что описанная здесь задача энтропийно-линейного программирования имеет много общего с обычной транспортной задачей.

Далее, исходя из известных потребностей, водители начинают «нащупывать» некую равновесную конфигурацию потоков (конкурентное равновесие, равновесие Нэша–Вардроп (1952)). Корреспонденция не определяет, вообще говоря, однозначно путь следования. Однако 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 используемым путям, соответствующим данной корреспонденции, должны быть одинаковы. 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 статьях Е.В. Гасниковой и Ю.В. Дорна, Е.А. Нурминского и Н.Б. Шамрай, В.И. Швецова.

Рассмотрим в заключение задачу Монжа–Канторовича о перемещении масс (эквивалентную, 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 (Пуанкаре–Леви): площадь многомерной сферы (с выделенным северным и южным полюсами) практически полностью сосредоточена в маленькой полоске вокруг экватора. Этот принцип нашел широкие применения 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). В статье А.В. Колесникова много внимания уделяется изучению свойств отображения оптимального плана перевозок, в частности исследуется роль сжимаемости этого отображения. Это единственная статья в данном сборнике, в которой затрагивается очень старая (Г. Монж, 1781) и очень важная в транспортной проблематике задача о перемещении масс, сильно связанная с транспортной задачей, о которой написано в трех статьях настоящего журнала. Отметим также важность затронутых в этой статье вопросов для целостного восприятия статей всего номера. Речь идет о широких приложениях явления концентрации меры в других приведенных здесь статьях: концепция равновесия макросистем для расчета матрицы корреспонденций, исследования надежности графа транспортной сети по модели Эрдеша–Реньи, оценки скорости сходимости к равновесию (неравенства Пуанкаре, Чигера) и др.

В целом, несмотря на разнообразие задач по трафику и подходов к их решению со стороны авторов настоящего выпуска журнала, следует надеяться, что первый шаг к созданию коллектива специалистов для решения столь сложной задачи сделан.

*Козлов Валерий Васильевич,
вице-президент РАН,
академик*

Л.Г. Афанасьева, Е.В. Булинская

Московский государственный университет им. М.В. Ломоносова

Математические модели транспортных систем, основанные на теории очередей

Цель работы — получение условий эргодичности для ряда транспортных систем, учитывающих наличие светофоров. Поступление автомобилей описывается пуассоновскими процессами. Интервалы между последовательными переключениями светофора могут иметь произвольное распределение. Рассматривается также оптимизация функционирования светофора. Существование стационарного режима в системе с двумя последовательными светофорами установлено с учетом размера автомобилей. Изучено влияние светофоров и различия в скоростях автомобилей на плотность их потока. Методы теории очередей оказались полезными для исследования.

Ключевые слова: транспортные потоки, светофоры, нерегулируемый перекресток, эргодичность, оптимизация, плотность потока, высокая загрузка.

I. Введение

Изучение транспортных потоков началось достаточно давно, еще до второй мировой войны (см., например, [1–4] и приведенные там ссылки). Поэтому невозможно даже просто упомянуть всех исследователей, внесших вклад в разработку данной тематики. Отметим только, что для исследований используются различные методы такие, как клеточные автоматы [5], статистическая механика и математическая физика [6–10] или теория очередей [11].

В данной работе продолжаются исследования, начатые авторами в [12] и [13], и применяются методы теории очередей. Основная цель авторов получить условия эргодичности для стохастических транспортных моделей, учитывающих наличие светофоров (которые можно интерпретировать как обслуживающие устройства в системе массового обслуживания). Потоки автомобилей считаются пуассоновскими. Времена между переключениями света на светофорах предполагаются независимыми случайными величинами с произвольными распределениями. Первое из предположений вполне естественно ввиду хорошо известных результатов о суммировании случайных потоков (см., например, [14]). Второе условие позволяет, в частности, рассмотреть функционирование нерегулируемых перекрестков, введенных в [15]. В самом деле, безопасное пересечение основной трассы возможно в течение некоторого случайного интервала, который может быть интерпретирован как интервал, когда для пересекающей дороги горит зеленый свет. Кроме того, для приложений важно изучить влияние распределения интервалов между переключениями на такие характеристики модели, как среднее число автомобилей, ожидающих у светофора, или среднее время ожидания перед светофором. Согласно числен-

ным подсчетам эти средние почти одинаковы при коэффициенте загрузки, близком к половине. Напротив, среднее, соответствующее показательному распределению, гораздо больше среднего для постоянных интервалов при высокой загрузке. Наконец, для показательно распределенных интервалов можно получить явный вид для среднего числа ожидающих автомобилей. Это позволяет найти оптимальное правило функционирования светофора. Полученный результат можно использовать в качестве начального приближения для нахождения оптимальной политики в случае постоянных (или произвольно распределенных) интервалов.

II. Плотность потока автомобилей, движущихся с различными скоростями

II.1. Описание модели

Предположим, что автомобили двух типов (медленные и быстрые) движутся в одном направлении по трассе, которую далее считаем действительной прямой. Автомобили появляются на трассе независимо друг от друга, при этом автомобиль типа i ($i = 1, 2$) возникает в интервале $(x, x + dx)$ в промежутке времени $(t, t + dt)$ с вероятностью $\lambda_i dx dt$. Каждый автомобиль проходит по трассе некоторое случайное расстояние, после чего покидает ее. Пройденные расстояния являются независимыми показательно распределенными случайными величинами с параметром μ_i для автомобилей i -го типа. Пусть V_i — скорость автомобиля i -го типа, причем $V_1 < V_2$. Очевидно, что $\zeta_i = (\mu_i V_i)^{-1}$ — это среднее время, проводимое автомобилем i -го типа на трассе, если он не взаимодействует с другими автомобилями. Предполагается, что автомобиль второго типа (быстрый), догнавший автомобиль первого типа (медленный), начинает двигаться со скоростью V_1 до тех пор, пока

либо он сам покинет трассу, либо ее покинет задержавший его автомобиль первого типа. При таком поведении могут образовываться группы автомобилей произвольного размера. Когда трассу покидает блокирующий автомобиль первого типа, вся группа быстрых автомобилей начинает двигаться со скоростью V_2 . Размер автомобилей мы не принимаем во внимание. Следовательно, при фиксированном моменте времени t конфигурация автомобилей на трассе описывается последовательностью $\{x_s, n_s, e_s\}_{s=-\infty}^{+\infty}$, где $x_s \in R^1$ — положение s -й группы на прямой, n_s — число автомобилей в s -й группе, а $e_s = 1$, если в этой группе имеется автомобиль первого типа, и $e_s = 0$, если группа состоит только из автомобилей второго типа. Таким образом, мы изучаем случайный процесс $X(t)$, принимающий значения в пространстве конфигураций $X = \{(x_s, n_s, e_s)_{s=-\infty}^{+\infty}\}$, то есть *маркированный точечный процесс* (см., например, [16]). В сделанных предположениях это *однородный марковский процесс*. Строгое доказательство этого утверждения, а также построение производящего оператора достаточно сложно и опирается на теорию маркированных точечных процессов. Мы ограничимся здесь интуитивными соображениями, как это обычно делается в теории массового обслуживания. Итак, предположим, что известно состояние $X(t)$ в момент t . Поведение процесса после момента t определяется следующими событиями: появление на трассе новых автомобилей; уход с трассы некоторых из следующих по ней автомобилей, что приводит к исчезновению некоторых из точек x_s , когда $n_s = 1$, или уменьшению размера соответствующей группы на единицу, если $n_s > 1$; догон группой из автомобилей второго типа какого-нибудь автомобиля первого типа.

Поскольку поток автомобилей, появляющихся на трассе в интервале $(x, x + h)$, является пуассоновским с параметром λh , длина проходимого пути имеет показательное распределение, обладающее, как известно, свойством отсутствия памяти, а скорость автомобиля каждого типа постоянна, все указанные факторы не зависят от прошлого, если известно состояние $X(t)$.

II.2. Эргодическая теорема. Определение плотности потока

Мы начнем с теоремы, устанавливающей существование предельного распределения процесса $X(t)$.

Теорема 1. Если $V_i > 0$, $\mu_i > 0$ при $i = 1, 2$, то марковский процесс $X(t)$ эргодический.

Доказательство. Проводится в несколько этапов. Прежде всего строится мажорирующий процесс $\tilde{X}(t)$. Предполагается, что все автомобили движутся со скоростью $V = V_1$, а интенсивность их появления на участке $(x, x + dx)$ равна $\lambda = \lambda_1 + \lambda_2$.

Каждый автомобиль проходит по прямой случайное расстояние, имеющее показательное рас-

пределение с параметром $\mu = \min(\mu_1, \mu_2)$. Обозначим $N_A(t)$ (соотв. $\tilde{N}_A(t)$) число автомобилей в интервале $(-A, A)$ в момент t для процесса $X(\cdot)$ (соотв. $\tilde{X}(\cdot)$). Тогда $\tilde{N}_A(t)$ стохастически доминирует $N_A(t)$, это записывается $\tilde{N}_A(t) \geq_{st} N_A(t)$ и означает

$$P(\tilde{N}_A(t) > K) \geq P(N_A(t) > K) \text{ для любого } K.$$

В свою очередь процесс $\tilde{X}(t)$ эквивалентен процессу $\hat{X}(t)$, в котором автомобили не движутся, а появившись в некоторой точке, находятся в ней случайное время, имеющее показательное распределение с параметром μV , и потом исчезают. Процесс $\hat{X}(t)$ является пуассоновским точечным процессом и для него $\tilde{N}_A(t)$ имеет предельное (при $t \rightarrow \infty$) пуассоновское распределение с параметром $2\lambda A(\mu V)^{-1}$. Это позволяет проверить выполнение условий эргодической теоремы Боровкова (см. [17], теорема 1, с. 184) для процесса с дискретным временем $\{X(nh), n \geq 1\}$.

Наконец, устанавливается асимптотическая стохастическая непрерывность $X(t)$, то есть $\|P(x, t + h, \cdot) - P(x, t, \cdot)\| \rightarrow 0$, для любого x при $h \rightarrow 0, t \rightarrow \infty$. Здесь $P(x, t, \cdot)$ — переходная вероятность однородного марковского процесса $X(t)$, а $\|\cdot\|$ — расстояние по вариации. Детали опускаются ввиду громоздкости.

Процесс $X(t)$ в дальнейшем считается стационарным. Отметим еще одно свойство, которое в дальнейшем будем называть *свойством независимости и однородности*. Для любого конечного интервала I действительной прямой и момента времени t введем следующие обозначения: $\nu^I(t)$ — число точек, в которых есть автомобили (или группы автомобилей), случайные величины $x_s^I(t)$, $n_s^I(t)$, $e_s^I(t)$ имеют такой же смысл, как x_s , n_s , e_s , только относятся к автомобилям, находящимся в момент t в интервале I . Введем случайный элемент $X^I(t) = (\nu^I(t), x_s^I(t), n_s^I(t), e_s^I(t), s = 1, \dots, \nu^I(t))$.

Утверждение 1. (Свойство независимости и однородности). Пусть I_1 и I_2 — конечные непесекающиеся интервалы действительной прямой. Для фиксированного t случайные элементы $X^{I_1}(t)$ и $X^{I_2}(t)$ независимы, а если $|I_1| = |I_2|$, где $|I|$ — длина интервала I , то одинаково распределены.

Доказательство этого свойства опирается на теорию маркированных точечных процессов, хотя интуитивно оно достаточно очевидно. Дело в том, что появления автомобилей в малых окрестностях различных точек, а также в различные интервалы времени в окрестности одной точки независимы.

Под плотностью потока автомобилей i -го типа мы понимаем математическое ожидание числа автомобилей этого типа, находящихся в отрезке единичной длины. В силу стационарности процесса $X(t)$ и свойства однородности плотность потока постоянна, то есть не зависит от времени и положения отрезка на прямой. Дадим формальное определение.

Определение 1. Плотность потока автомобилей первого типа

$$\alpha_1 = \lim_{|I| \rightarrow \infty} \frac{1}{|I|} \sum_{s=1}^{\nu^I(t)} e_s^{|I|}(t), \quad (1)$$

а плотность потока автомобилей второго типа

$$\alpha_2 = \lim_{|I| \rightarrow \infty} \frac{1}{|I|} \sum_{s=1}^{\nu^I(t)} (n_s^I(t) - e_s^I(t)), \quad (2)$$

суммарная плотность потока

$$\Lambda = \alpha_1 + \alpha_2 = \lim_{|I| \rightarrow \infty} \frac{1}{|I|} \sum_{s=1}^{\nu^I(t)} n_s^{|I|}(t).$$

Существование пределов (1) и (2) с вероятностью единица следует из усиленного закона больших чисел и свойства независимости и однородности. Поскольку процесс $X(t)$ стационарный, эти пределы не зависят от времени.

II.3. Система уравнений для вычисления плотности потока автомобилей

Чтобы получить эту систему, введем функции $p_j(t, x)$ и $q_j(t, x)$, $j \geq 0$, такие, что $p_j(t, x)dx$ — вероятность того, что в момент t в интервале $(x, x+dx)$ содержится группа из j автомобилей второго типа и один автомобиль первого типа, а $q_j(t, x)dx$ — аналогичная вероятность, но нет автомобиля первого типа.

Используя свойства независимости и однородности, для $h \rightarrow 0$ можно записать

$$p_0(t+h, x) = (1 - \mu_1 V_1 h - \beta(V_2 - V_1)h)p_0(t, x - hV_1) + \lambda_1 h + \mu_2 V_1 h p_1(t, x - hV_1) + o(h),$$

а при $j > 0$

$$p_j(t+h, x) = (1 - \mu_1 V_1 h - \beta(V_2 - V_1)h - j\mu_2 V_1 h) \times \\ \times p_j(t, x - hV_1) + (j+1)\mu_2 V_1 h p_{j+1}(t, x - hV_1) + \\ + (V_2 - V_1)h \sum_{i=1}^{j-1} p_i(t, x - hV_1) q_{j-i}(t, x - hV_2) + o(h),$$

где $\beta = \sum_{j=1}^{\infty} q_j(t, x)$.

Аналогичные уравнения можно выписать для функций $q_j(t, x)$, $j \geq 1$. Используя тот факт, что $p_j(t, x)$ и $q_j(t, x)$ не зависят от t и x , следовательно, их производные по t и по x равны нулю, для $p_j = p_j(t, x)$ и $q_j = q_j(t, x)$ получаем систему уравнений

$$(\mu_1 V_1 + \beta(V_2 - V_1))p_0 = \mu_2 V_1 p_1 + \lambda_1, \quad (3)$$

$$(j\mu_2 V_1 + \mu_1 V_1 + \beta(V_2 - V_1))p_j = \mu_2 V_1 (j+1)p_{j+1} + \\ + (V_2 - V_1) \sum_{i=0}^{j-1} p_i q_{j-i}, \quad j > 0,$$

и

$$\left(\mu_2 V_2 + (V_2 - V_1) \sum_{j=0}^{\infty} p_j \right) q_1 = \\ = 2\mu_2 V_2 q_2 + \mu_1 V_1 p_1 + \lambda_2, \quad (4)$$

$$(j\mu_2 V_2 + (V_2 - V_1) \sum_{j=0}^{\infty} p_j) q_j =$$

$$= (j+1)\mu_2 V_2 q_{j+1} + \mu_1 V_1 p_j, \quad j > 1.$$

Мы замечаем, что $\sum_{j=0}^{\infty} p_j = \alpha_1$. Чтобы найти α_1 , введем функцию $\alpha_1(t, x)$ такую, что $\alpha_1(t, x)dx$ есть вероятность того, что в момент t в интервале $(x, x+dx)$ есть автомобиль первого типа. Тогда справедливо соотношение

$$\alpha_1(t+h, x) = (1 - \mu_1 V_1 h)\alpha_1(t, x - V_1 h) + \lambda_1 h + o(h).$$

Устремляя h к нулю, получаем

$$\frac{\partial \alpha_1}{\partial t} + V_1 \frac{\partial \alpha_1}{\partial x} = -\mu_1 V_1 \alpha_1 + \lambda_1.$$

Поскольку $\partial \alpha_1 / \partial t = 0$, $\partial \alpha_1 / \partial x = 0$, из выведенного соотношения имеем

$$\alpha_1 = \lambda_1 (\mu_1 V_1)^{-1} = \lambda_1 \zeta_1.$$

А для того, чтобы найти плотность потока автомобилей второго типа $\alpha_2 = \sum_{j=1}^{\infty} j(p_j + q_j)$, нужно решить бесконечную систему нелинейных уравнений (3)–(4).

II.4. Аппроксимационная процедура

Для получения решения системы (3)–(4) мы предлагаем следующий

Алгоритм. Обозначим $S^{(k)}$ систему первых $(k+1)$ уравнений соответственно из (3) и (4), где положим $p_j = 0$, $j > k$, и $q_j = 0$, $j > k+1$. Идея заключается в том, чтобы получать систему линейных уравнений, используя решение системы $S^{(k-1)}$. Более точно, процедура последовательных приближений состоит из следующих шагов.

Шаг 1. Пусть $k = 0$. Как было установлено, $\sum_{j=0}^{\infty} p_j = \alpha_1 = \lambda_1 \zeta_1$. Это значение будет подставлено во все уравнения (4). Тогда из второго уравнения системы $S^{(0)}$ получается $\beta^{(0)} = q_1^{(0)} = \lambda_2 (\mu_2 V_2 + \alpha_1 (V_2 - V_1))^{-1}$.

Шаг 2. Положим $\beta = \beta^{(0)}$ в первых двух уравнениях системы $S^{(1)}$, а вместо q_1 подставим $q_1^{(0)}$. В результате мы получим систему уравнений

$$(\mu_1 V_1 + \beta^{(0)}(V_2 - V_1))p_0 = \mu_2 V_1 p_1 + \lambda_1,$$

$$(\mu_2 V_1 + \mu_1 V_1 + \beta^{(0)}(V_2 - V_1))p_1 = (V_2 - V_1)p_0 q_1^{(0)},$$

$$(\mu_2 V_2 + \alpha_1 (V_2 - V_1))q_1 = \mu_1 V_1 p_1 + \lambda_2,$$

$$(2\mu_2 V_2 + \alpha_1 (V_2 - V_1))q_2 = \mu_1 V_1 p_1.$$

Эта система линейных уравнений имеет решение $(p_0^{(1)}, p_1^{(1)}, q_1^{(1)}, q_2^{(1)})$.

Шаг 3. Теперь рассматривается система $S^{(2)}$, в которой вместо β берется $\beta^{(1)} = q_1^{(1)} + q_2^{(1)}$, а вместо q_1 и q_2 , аналогично шагу 2, используем $q_1^{(1)}$ и $q_2^{(1)}$ в уравнениях для p_j , $j = 0, 1, 2$.

Следующие шаги с очевидными изменениями повторяют шаг 3.

Если известны (p_j, q_j) , то можно выписать в явном виде плотность автомобилей

$$\Lambda = \lambda_1 \zeta_1 + \sum_{j=1}^{\infty} j(p_j + q_j)$$

и вероятности различных конфигураций автомобилей на конечных интервалах R^1 .

II.5. Оценка сверху плотности автомобилей

С этой целью рассмотрим вспомогательную модель, в которой предполагается, что автомобиль второго типа, догнавший автомобиль первого типа, сам приобретает тип один. Это означает, что автомобили второго типа не образуют групп. Если $\zeta_2 \leq \zeta_1$, то плотность потока автомобилей в этой вспомогательной модели не меньше, чем в рассматриваемой основной. Поскольку теперь $q_j = 0$ при $j > 1$, вместо уравнений (3)–(4) получаем систему

$$\begin{aligned} (j\mu_1 V_1 + q_1(V_2 - V_1))\tilde{p}_j &= (j + 1)\mu_1 V_1 \tilde{p}_{j+1} + \\ &+ (V_2 - V_1)q_1 \tilde{p}_{j-1}, \end{aligned} \quad (5)$$

и

$$\begin{aligned} (\mu_1 V_1 + q_1(V_2 - V_1))\tilde{p}_1 &= 2\mu_1 V_1 \tilde{p}_2 + \lambda_1, \quad (6) \\ (\mu_2 V_2 + \tilde{\alpha}(V_2 - V_1))q_1 &= \lambda_2, \quad (7) \end{aligned}$$

где $\tilde{\alpha} = \sum_{j=1}^{\infty} \tilde{p}_j$, а $\tilde{p}_j dx$ означает вероятность того, что имеется группа из j медленных автомобилей в интервале $(x, x + dx)$ для любых $x \in R^1$ и $t \in R^1$.

Теорема 2. Решение уравнений (5)–(7) имеет вид

$$q_1 = \frac{\delta_0}{\zeta_1(V_2 - V_1)}, \quad \tilde{p}_j = \frac{\delta_0^{j-1}}{j!} \tilde{p}_1, \quad j \geq 1,$$

где $\tilde{p}_1 = \lambda_1 \zeta_1$ и δ_0 является единственным корнем уравнения

$$\lambda_1 \zeta_1 (e^{\delta} - 1) = \lambda_2 \zeta_1 - \delta((V_2 - V_1)\zeta_2)^{-1}, \quad (8)$$

кроме того,

$$\tilde{\Lambda} = \zeta_1(\lambda_1 + \lambda_2) + \frac{\zeta_2 - \zeta_1}{(V_2 - V_1)\zeta_1 \zeta_2} \delta_0. \quad (9)$$

Доказательство. Просуммировав уравнения (5) по j и прибавив уравнение (6), получаем

$$\tilde{p}_1 = \lambda_1(\mu_1 V_1)^{-1} = \lambda_1 \zeta_1. \quad (10)$$

Чтобы найти решение уравнений (5) и (6), положим

$$u_j = j\mu_1 V_1 \tilde{p}_j - q_1(V_2 - V_1)\tilde{p}_{j-1}.$$

Тогда (5) можно переписать в виде

$$u_j = u_{j+1} \text{ для } j > 1. \quad (11)$$

Подстановка (10) в (6) дает

$$\begin{aligned} 2\mu_1 V_1 \tilde{p}_2 &= (\mu_1 V_1 + q_1(V_2 - V_1))\tilde{p}_1 - \lambda_1 = \\ &= q_1(V_2 - V_1)\lambda_1 \zeta_1, \end{aligned}$$

поэтому

$$u_2 = 2\mu_1 V_1 \tilde{p}_2 - q_1(V_2 - V_1)\tilde{p}_1 = 0.$$

Отсюда, принимая во внимание (10), легко видеть, что $u_j = 0$ при $j > 1$ и

$$\tilde{p}_{j+1} = \frac{q_1(V_2 - V_1)}{\mu_1 V_1(j + 1)} \tilde{p}_j, \quad j \geq 1.$$

Обозначив $\delta = q_1 \zeta_1 (V_2 - V_1)$, мы получаем

$$\tilde{p}_j = \frac{\delta^{j-1}}{j!} \tilde{p}_1, \quad j \geq 1. \quad (12)$$

Используя (7), (10) и (12), устанавливаем справедливость следующих соотношений:

$$\tilde{\alpha} = \frac{\lambda_1(e^{\delta} - 1)}{\mu_1 V_1 \delta} = \frac{\lambda_2}{\mu_1 V_1 \delta} - \frac{\mu_2 V_2}{V_2 - V_1}, \quad (13)$$

а значит, и тот факт, что δ удовлетворяет (8). Это уравнение имеет единственное решение δ_0 и

$$\delta_0 < \delta_1 = \lambda_2(V_2 - V_1)\zeta_1 \zeta_2. \quad (14)$$

Наконец, мы имеем

$$\tilde{p}_j = \frac{\lambda_1 \zeta_1 \delta_0^{j-1}}{j!}, \quad j \geq 1,$$

в то время как среднее число медленных автомобилей в группе равно

$$\begin{aligned} m &= \lambda_1 \zeta_1 \sum_{j=1}^{\infty} j \frac{\delta_0^{j-1}}{j!} = \lambda_1 \zeta_1 e^{\delta_0} = (\lambda_1 + \lambda_2)\zeta_1 - \\ &- \frac{\delta_0}{\zeta_2(V_2 - V_1)}. \end{aligned}$$

Согласно (7) и (13) получаем $q_1 = \delta_0 \mu_1 V_1 / (V_2 - V_1)$. Следовательно, плотность потока автомобилей, равная $\tilde{\Lambda} = q_1 + m$, имеет вид (9). Тем самым доказательство закончено.

Если δ_1 в (14) мало, то используя разложение Тейлора для e^{δ} только с линейной частью, получим вместо (8) простое уравнение

$$\lambda_1 \zeta_1 \delta = \lambda_2 \zeta_1 - \delta(\zeta_2(V_2 - V_1))^{-1},$$

из которого вытекает следующее приближение для δ_0 :

$$\tilde{\delta} = \frac{\lambda_2}{\lambda_1 + (\zeta_1 \zeta_2 (V_2 - V_1))^{-1}}.$$

III. Плотность потока автомобилей на трассе со светофорами

В предыдущем разделе исследовано влияние на плотность потока автомобилей различия в их скоростях. Наличие светофоров способствует образованию скоплений автомобилей перед ними и увеличению плотности. Анализ этого явления посвящен данный раздел.

III.1. Описание модели

Предположим, что расстояния между соседними светофорами на основной трассе образуют последовательность $\{\eta_i\}_{i=-\infty}^{+\infty}$ независимых случайных величин с общей функцией распределения $F(x)$. Пересечение перекрестка по основной трассе возможно, если на светофоре горит зеленый свет. Красный свет, запрещающий движение по основной трассе, разрешает движение по перпендикулярному направлению (для него соответственно зажигается зеленый свет). Длительности промежутков между переключениями светофоров образуют последовательности независимых случайных величин, имеющих показательное распределение с параметрами соответственно γ_1 и γ_2 . Если в момент подхода автомобиля к светофору горит зеленый свет и нет очереди, он беспрепятственно следует далее. Если горит красный свет или есть очередь, он останавливается. В таком случае время прохождения светофора имеет показательное распределение с параметром ν . Все случайные величины, определяющие функционирование светофора, независимы и не зависят от входящего потока автомобилей. Пусть $\lambda dx dt$ — вероятность возникновения автомобиля в интервале $(x, x + dx)$ за время dt . Автомобили на трассе появляются как в точках расположения светофоров, так и между ними. В точку расположения светофора в перпендикулярном к основной трассе направлении поступают два независимых потока автомобилей. Первый из них, интенсивности a_2 , сворачивает на трассу после прохождения светофора, а второй, интенсивности a_3 , продолжает движение в перпендикулярном направлении. Все автомобили следуют по трассе с постоянной скоростью V и проходят по ней случайное расстояние, имеющее показательное распределение с параметром μ .

III.2. Интенсивность потока автомобилей и условие эргодичности

Выберем один из светофоров и введем случайные процессы: $Y_1(t)$ — количество автомобилей, подошедших к светофору за время t , $Z_1(t)$ — количество автомобилей, прошедших через перекресток по основной трассе за то же время, и пусть $X_1(t)$ — число автомобилей, стоящих перед светофором в момент t . Соответственно процессы $Y_2(t)$, $Z_2(t)$, $X_2(t)$ будут относиться к автомобилям, сворачивающим с перпендикулярного направления на трассу, а $Y_3(t)$, $Z_3(t)$, $X_3(t)$ к автомобилям, которые пересекают трассу, на нее не сворачивая.

Под интенсивностью процесса $Y_1(t)$ мы будем понимать (существующую в сделанных предположениях) величину

$$a_1 = \lim_{t \rightarrow \infty} \frac{Y_1(t)}{t}.$$

Первый результат касается этого предела.

Теорема 3. Если процессы $X_i(t)$, $i = 1, 2, 3$, стохастически ограничены, то

$$a_1 = \frac{a_2 f^*(\mu)}{1 - f^*(\mu)} + \frac{\lambda}{\mu}, \quad (15)$$

где $f^*(\mu) = \int_0^{\infty} e^{-\mu x} dF(x)$.

Доказательство. Поскольку $t^{-1}X_i(t) \rightarrow 0$ при $t \rightarrow \infty$ и $Z_i(t) = Y_i(t) - X_i(t)$, $i = 1, 2, 3$, то

$$\lim_{t \rightarrow \infty} \frac{Z_i(t)}{t} = \lim_{t \rightarrow \infty} \frac{Y_i(t)}{t} = a_i, \quad i = 1, 2, 3.$$

Интенсивность потока автомобилей, проходящих через перекресток и следующих по трассе, таким образом, будет $a_1 + a_2$. Некоторые из этих автомобилей покидают трассу до того, как они достигнут следующего светофора. Вероятность этого события равна $\int_0^{\infty} (1 - e^{-\mu x}) dF(x)$, так что интенсивность потока автомобилей на следующем светофоре, из числа прошедших предыдущий, составит

$$(a_1 + a_2) \int_0^{\infty} e^{-\mu x} dF(x) = (a_1 + a_2) f^*(\mu).$$

К следующему светофору также придут некоторые автомобили из числа появившихся на трассе между светофорами. Интенсивность этого потока будет равна

$$\lambda \int_0^{\infty} \int_0^y e^{-\mu x} dx dF(y) = \lambda \mu^{-1} (1 - f^*(\mu)).$$

Из соображений симметрии интенсивности потоков автомобилей, следующих по трассе, равны для всех светофоров, поэтому получаем уравнение

$$a_1 = (a_1 + a_2) f^*(\mu) + \lambda \mu^{-1} (1 - f^*(\mu)), \quad (16)$$

откуда немедленно вытекает (15).

Теорема 4. Если $V > 0$, $\mu > 0$ и выполнены условия

$$a_1 \theta < \nu \gamma_1^{-1}, \quad (a_2 + a_3) \theta < \nu \gamma_2^{-1}, \quad (17)$$

где $\theta = \gamma_1^{-1} + \gamma_2^{-1}$, а параметр a_1 задается соотношением (15), то процессы $X_i(t)$, $i = 1, 2, 3$, имеют собственное предельное распределение.

Доказательство. Совокупность светофоров представляет собой своеобразную сеть Джексона (см., например, [18]) с бесконечным числом узлов. Ее особенность состоит в том, что между переходами из узла i в узел $i + 1$ требования могут исчезать и появляться. Вероятность того, что требование, прошедшее узел с номером i , появится перед узлом с номером $i + 1$, равна $f^*(\mu)$, так что интенсивность потока таких требований равна $(a_1 + a_2) f^*(\mu)$. Интенсивность пуассоновского потока вновь поступающих требований составляет $(\lambda/\mu)(1 - f^*(\mu))$. Следовательно, соотношение

(16) представляет собой уравнение баланса для рассматриваемой сети. Поясним смысл соотношений (17). Циклом назовем временной интервал между последовательными включениями зеленого света в направлении основной трассы, так что θ — средняя продолжительность цикла. Первое из соотношений (17) означает, что среднее число автомобилей, приходящих к светофору за цикл по трассе, меньше математического ожидания количества автомобилей, которые могут пройти через светофор в этом направлении. Второе из соотношений (17) имеет тот же смысл для перпендикулярного к основной трассе направления. Строгое доказательство теоремы использует метод обновляющих событий, а также идеи и подходы, разработанные А.А. Боровковым [19–20] и В.В. Калашниковым и С.Г. Фоссом [21]. Детали опущены ввиду их громоздкости и прикладной направленности данной статьи.

III.3. Плотность потока автомобилей при наличии светофоров

Введем случайный процесс $X(t)$ со значениями в фазовом пространстве $X = \{x\}$, где $x = (y_i, l_i, n_i, k_i, z_1^{(i)}, \dots, z_{k_i}^{(i)}, e_i)_{i=-\infty}^{\infty}$, y_i — координата i -го светофора, l_i — число автомобилей на трассе перед i -м светофором, n_i — число автомобилей в перпендикулярном к трассе направлении перед i -м светофором, k_i — число автомобилей между i -м и $(i + 1)$ -м светофорами, $z_j^{(i)}$ — расстояние j -го из указанных автомобилей до i -го светофора, $j = 1, \dots, k_i$, наконец, $e_i = 1$, если у i -го светофора горит зеленый свет для трассы, и $e_i = 0$, если красный.

В сделанных предположениях процесс $X(t)$ — марковский и, если выполнены условия теоремы 4, у него есть стационарное распределение. Далее будем предполагать, что процесс $X(t)$ стационарен.

Под плотностью потока автомобилей будем понимать

$$\Lambda = \lim_{A \rightarrow \infty} \frac{N_A}{2A},$$

где, как и раньше, N_A — число автомобилей в промежутке $(-A, A)$.

Пусть αdx — вероятность наличия автомобилей в интервале $(x, x + dx)$, причем x лежит между светофорами. Если $\kappa = \int_0^\infty (1 - F(x)) dx$ — среднее расстояние между светофорами, то

$$\Lambda = \frac{m_1}{\kappa} + \alpha, \tag{18}$$

где m_1 — среднее число автомобилей перед светофором на трассе.

Чтобы найти α , введем функцию $\beta(x)$, представляющую собой условную плотность автомобилей при условии, что ближайший слева светофор

находится на расстоянии x . Обычным образом получаем уравнение

$$V \frac{d\beta}{dx} = -\mu V \beta(x) + \lambda \tag{19}$$

с начальным условием

$$\beta(0) = \frac{a_1 + a_2}{V}. \tag{20}$$

Решение (19) при условии (20) имеет вид

$$\beta(x) = \frac{\lambda}{V\mu} (1 - e^{-\mu x}) + \frac{a_1 + a_2}{V} e^{-\mu x}. \tag{21}$$

Теперь из теории восстановления (см., например, [22]) получаем

$$\alpha = (\kappa)^{-1} \int_0^\infty \beta(y) (1 - F(y)) dy.$$

Используя (21) и (15), после несложных выкладок находим

$$\alpha = \frac{a_2 + \lambda\kappa}{V\kappa\mu} = \frac{\lambda}{V\mu} + \frac{a_2}{V\kappa\mu},$$

откуда с учетом (18) получаем формулу

$$\Lambda = \frac{\lambda}{V\mu} + \frac{a_2}{V\kappa\mu} + \frac{m_1}{\kappa}.$$

Основная сложность состоит в вычислении m_1 . Этому посвящен следующий раздел.

IV. Функционирование регулируемого перекрестка

IV.1. Описание модели

Предположим, что автомобили движутся в двух перпендикулярных направлениях по однополосным трассам. Мы хотим изучить, как наличие светофора влияет на образование очередей автомобилей. Пусть $\{t_n^{(i)}\}_{n=1}^\infty$ — это последовательные моменты включения зеленого света в i -м направлении, $i = 1, 2$, и $0 = t_1^{(1)} < t_2^{(2)} < t_2^{(1)} < \dots$. Интервалы между переключениями

$$\tau_n^{(1)} = t_n^{(2)} - t_n^{(1)}, \quad \tau_n^{(2)} = t_{n+1}^{(1)} - t_n^{(2)}, \quad n = 1, 2, \dots,$$

образуют две независимые последовательности независимых одинаково распределенных (в каждой последовательности) случайных величин соответственно с функциями распределения $G_1(x)$ и $G_2(x)$ и средними γ_1^{-1} и γ_2^{-1} . Сумму $\tau_n^{(1)} + \tau_n^{(2)}$ назовем n -м циклом и положим $\theta = \gamma_1^{-1} + \gamma_2^{-1}$. Поток автомобилей, прибывающих к перекрестку в i -м направлении, $A_i(t)$ является пуассоновским с интенсивностью λ_i , $i = 1, 2$. Эти потоки предполагаются независимыми. Красный свет светофора запрещает движение, и автомобиль останавливается у светофора или в конце очереди ожидающих автомобилей. Когда загорается зеленый

свет, время прохождения перекрестка является по-казательно распределенной с параметром ν случайной величиной для каждого автомобиля. Зеленый свет в первом направлении соответствует красному во втором и наоборот. Случайные величины, определяющие функционирование светофора и прохождение перекрестка автомобилями, между собой независимы и не зависят от потоков прибывающих автомобилей.

IV.2. Эргодическая теорема

Рассмотрим случайный процесс $X(t) = (X_1(t), X_2(t))$, где $X_i(t)$ — число автомобилей, ожидающих в момент t перед перекрестком в i -м направлении, $i = 1, 2$. Согласно нашим предположениям процесс $X(t)$ не является марковским. Сначала мы установим, при каких условиях существует собственное предельное распределение этого процесса при $t \rightarrow \infty$.

Теорема 5. Пределы

$$\lim_{t \rightarrow \infty} P(X_i(t) = j) = p_j^{(i)}, \quad j = 0, 1, 2, \dots, \quad (22)$$

$$i = 1, 2,$$

с $p_j^{(i)} > 0$ и $\sum_{j=0}^{\infty} p_j^{(i)} = 1$ существуют тогда и только тогда, когда

$$\rho_i = \frac{\lambda_i \gamma_i \theta}{\nu} < 1, \quad i = 1, 2. \quad (23)$$

Доказательство. Поскольку процессы $A_1(t)$ и $A_2(t)$ — пуассоновские, процесс $X(t)$ является регенерирующим. Его точки регенерации — это такие моменты $t_j^{(1)}$, для которых $X_i(t_j^{(1)}) = 0, i = 1, 2$, то есть в обоих направлениях нет очереди автомобилей и для первого направления включился зеленый свет. Существование пределов (20) следует из теоремы Смита (см., например, [22]). Эти пределы $\{p_j^{(i)}\}_{j=0}^{\infty}$ образуют распределение тогда и только тогда, когда процесс $X_i(t)$ стохастически ограничен, а именно:

$$\lim_{k \rightarrow \infty} \limsup_{t \rightarrow \infty} P(X_i(t) > k) = 0. \quad (24)$$

Мы установим, что условие (23) эквивалентно (24). Доказательство этого факта основано на процедуре мажорирования. Введем вспомогательные процессы $Y_i(t), i = 1, 2$, числа автомобилей, которые могли бы пройти через перекресток, если все время имела очередь из ожидающих автомобилей. В нашей модели $Y_i(t), i = 1, 2$, — дважды стохастические пуассоновские процессы (ДСПП) (см., например, [23]) со случайными интенсивностями:

$$y_1(t, \omega) = \nu \sum_{k=1}^{\infty} I(t \in (t_k^{(1)}, t_k^{(2)}]),$$

$$y_2(t, \omega) = \nu \sum_{k=1}^{\infty} I(t \in (t_k^{(2)}, t_{k+1}^{(1)}]).$$

Здесь и далее $I(A)$ означает индикатор события A .

Рассмотрим первое направление и, без потери общности, положим $X_1(0) = 0$. Нас интересует поведение вложенного процесса $x_n = X_1(t_n^{(1)} - 0)$. Для его исследования введем два вспомогательных процесса с помощью следующих соотношений:

$$x_0^- = 0, \quad x_n^- = [x_{n-1}^- + \xi_n^{(1)} - \eta_n^{(1)}]^+,$$

$$x_0^+ = 0, \quad x_n^+ = [x_{n-1}^+ - \eta_n^{(1)}]^+ + \xi_n^{(1)}, \quad (25)$$

где

$$\xi_n^{(1)} = A_1(t_{n+1}^{(1)}) - A_1(t_n^{(1)}), \quad \eta_n^{(1)} = Y_1(t_n^{(2)}) - Y_1(t_n^{(1)}).$$

Таким образом, $\xi_n^{(1)}$ — это число автомобилей, прибывающих в первом направлении, а $\eta_n^{(1)}$ — число автомобилей, которые могут пройти перекресток в этом же направлении за n -й цикл.

Нетрудно проверить, что для любого $n \geq 1$

$$x_n^- \leq_{st} x_n \leq_{st} x_n^+. \quad (26)$$

В самом деле, пусть $\tilde{\eta}_n^{(1)}$ — это число автомобилей, прошедших через перекресток в первом направлении в течение n -го цикла, тогда

$$x_n = x_{n-1} + \xi_n^{(1)} - \tilde{\eta}_n^{(1)}.$$

Очевидно, что $\eta_n^{(1)} \geq_{st} \tilde{\eta}_n^{(1)}$. Вместе с (25) это неравенство дает первое из неравенств (26). Второе неравенство доказывается по индукции. Очевидно, $x_1 \leq_{st} x_1^+$, поскольку $x_0^+ = x_0 = 0, x_1^+ = \xi_1^{(1)}$ и $x_1 = \xi_1^{(1)} - \tilde{\eta}_1^{(1)}$. Теперь предположим, что $x_{n-1}^+ \geq_{st} x_{n-1}$. Отсюда немедленно вытекает, что

$$x_n = x_{n-1} + \xi_n^{(1)} - \tilde{\eta}_n^{(1)} \leq_{st} x_{n-1}^+ + \xi_n^{(1)} - \tilde{\eta}_n^{(1)}.$$

Если очередь перед светофором (в первом направлении) была не пуста в течение всего n -го цикла, то $\tilde{\eta}_n^{(1)} = \eta_n^{(1)}$ и $x_n \leq x_n^+ = [x_{n-1}^+ - \eta_n^{(1)}]^+ + \xi_n^{(1)}$. Если же были моменты, когда очередь была пустая, то $x_n \leq \xi_n^{(1)}$. Следовательно, и в этом случае $x_n \leq_{st} x_n^+$.

Процесс $\{x_n^-\}$ — это случайное блуждание с задерживающей границей в нуле. Если $\rho_1 \geq 1$, то процесс стохастически неограничен (см., например, [24]). А это означает, что $p_j^{(1)} = 0$ для любого j .

Теперь пусть $\rho_1 < 1$. Рассмотрим $\{x_n^+\}$, положив $S_k = \sum_{j=1}^k (\xi_j^{(1)} - \eta_{j+1}^{(1)})$, $S_0 = 0$ и $M_n = \xi_n^{(1)} + \max_{0 \leq j < n} S_j$. После некоторых преобразований из рекуррентных соотношений (23) нетрудно получить, что

$$P(x_n^+ \leq y) = P(M_n \leq y).$$

Поскольку $E(\xi_j^{(1)} - \eta_{j+1}^{(1)}) < 0$, случайное блуждание $\{S_n\}$ имеет отрицательный снос. Следовательно, процесс $\{M_n\}$ стохастически ограничен при $n \rightarrow \infty$.

Предложенная модель не учитывает зависимость времени прохода перекрестка от длины очереди. Одна из возможностей преодоления этого недостатка — предположение, что автомобиль, приходящий к светофору в то время, как для него горит зеленый свет, проходит его немедленно, если нет ожидающих автомобилей. Другими словами, время прохождения перекрестка равно нулю, а не является показательно распределенной случайной величиной. Такую ситуацию мы назовем *эффектом проскакивания*.

Обозначим $\hat{X}_i(t)$ аналог процесса $X_i(t)$, $i = 1, 2$, для модифицированной модели, то есть с проскакиванием. Нетрудно установить следующий результат.

Следствие 1. При выполнении условий (23) процессы $\hat{X}_i(t)$, $i = 1, 2$, являются эргодическими.

Доказательство. Рассмотрим только первое направление, для второго рассуждения аналогичны. Пусть $\rho_1 < 1$ и процесс $\hat{X}_1(t)$ стохастически неограничен, то есть $P(\hat{X}_1(t) = j) \rightarrow 0$ при $t \rightarrow \infty$ для любого j . Введем следующее событие:

$$C_n = \{\hat{X}_1(t_n^{(1)} + t) > 0 \quad \forall t \in [0, \tau_n^{(1)}]\}.$$

Из стохастической неограниченности $\hat{X}_1(t)$ вытекает

$$P(C_n) \rightarrow 1 \text{ при } n \rightarrow \infty. \quad (27)$$

Можно получить следующую оценку сверху:

$$\begin{aligned} E\hat{x}_{n+1} &= E\hat{X}_1(t_{n+1}^{(1)} - 0) \leq E(\hat{x}_n + \xi_n^{(1)} - \eta_n^{(1)})I(C_n) + \\ &+ E(\hat{x}_n + \xi_n^{(1)})I(\bar{C}_n) = \\ &= E\hat{x}_n + E\xi_n^{(1)} - E\eta_n^{(1)} + E\eta_n^{(1)}I(\bar{C}_n). \end{aligned}$$

В силу существования $E\eta_n^{(1)}$ и условия (27) последнее слагаемое $E\eta_n^{(1)}I(\bar{C}_n) \rightarrow 0$ при $n \rightarrow \infty$. Поэтому $E\hat{x}_{n+1} \leq E\hat{x}_n$ в предположении $\rho_1 < 1$, а это противоречит стохастической неограниченности $\hat{X}_1(t)$.

IV.3. Стационарный режим функционирования светофора при произвольном распределении интервалов между переключениями

Начнем с рассмотрения вложенной цепи Маркова $X_n^{(i)} = X_i(t_n^{(i)} - 0)$, где $t_n^{(i)}$ — момент n -го включения зеленого света в i -м направлении, $n \geq 1, i = 1, 2$. Предполагается, что выполнено условие (23).

Наша цель — предложить алгоритм для подсчета стационарного распределения процесса $\{X_n^{(i)}\}$, $i = 1, 2$, а также среднего числа ожидающих автомобилей в стационарном режиме. Можно рассмотреть только $\{X_n^{(1)}\}$, поскольку результаты для $\{X_n^{(2)}\}$ получаются заменой λ_1 на λ_2 и G_1 на G_2 .

Введем вспомогательный процесс рождения и гибели $Z(t)$ с поглощающим состоянием $\{0\}$, с интенсивностью рождения λ_1 и интенсивностью гибели ν . Положим

$$\varphi_{kj}(t) = P(Z(t) = j | Z(0) = k), \quad k > 0, \quad j \geq 0.$$

Эти функции удовлетворяют следующей системе уравнений:

$$\begin{aligned} \varphi'_{k0}(t) &= \nu\varphi_{k1}(t), \\ \varphi'_{k1}(t) &= -(\lambda_1 + \nu)\varphi_{k1}(t) + \nu\varphi_{k2}(t), \\ \varphi'_{kj}(t) &= -(\lambda_1 + \nu)\varphi_{kj}(t) + \nu\varphi_{k,j+1}(t) + \lambda_1\varphi_{k,j-1}(t), \\ & \quad j > 1, \end{aligned} \quad (28)$$

с начальными условиями $\varphi_{kk}(0) = 1$ и $\varphi_{kj}(0) = 0$ при $j \neq k$. Решение уравнений (28) может быть получено с помощью применения преобразования Лапласа и его последующего обращения. Таким образом, мы установим явный вид $\varphi_{kj}(t)$ в терминах обобщенных функций Бесселя первого рода.

С этой целью введем производящие функции

$$\Phi_k(z, t) = \sum_{n=0}^{\infty} \varphi_{kn}(t)z^n, \quad |z| < 1.$$

Из (28) немедленно вытекает, что

$$z \frac{\partial \Phi_k(z, t)}{\partial t} = (1 - z)(\nu - \lambda_1 z)(\Phi_k(z, t) - \varphi_{k0}(t)),$$

с $\Phi_k(z, 0) = z^k$.

Используя преобразование Лапласа по t ,

$$f^*(s) = \int_0^{\infty} e^{-st} f(t) dt, \quad \text{Res} \geq 0,$$

получаем

$$\Phi_k^*(z, s) = \frac{z^{k+1} - \varphi_{k0}^*(s)(1 - z)(\nu - \lambda_1 z)}{sz - (1 - z)(\nu - \lambda_1 z)}.$$

Поскольку $\Phi_k^*(z, s)$ аналитична в единичном круге, нули знаменателя $\alpha_l(s)$ с $|\alpha_l(s)| < 1$ должны совпадать с нулями числителя. Это дает следующее выражение:

$$\varphi_{k0}(s) = \frac{\alpha_2(s)^{k+1}}{(1 - \alpha_2(s))(\nu - \lambda_1 \alpha_2(s))},$$

где

$$\alpha_2(s) = \frac{\lambda_1 + \nu + s - \sqrt{(\lambda_1 + \nu + s)^2 - 4\lambda_1\nu}}{2\lambda_1}.$$

Обращая преобразование Лапласа $\Phi_k^*(z, s)$ и разлагая $\Phi_k(z, t)$ в ряд по степеням z , приходим к следующему виду для коэффициентов:

$$\begin{aligned} \varphi_{kj}(t) &= e^{-(\lambda_1 + \nu)t} (\lambda_1 / \nu)^{(j-2)/2} (J_{|j-k|}(2\sqrt{\lambda_1\nu}t) - \\ & - J_{j+k}(2\sqrt{\lambda_1\nu}t)), \quad j > 0. \end{aligned}$$

Здесь

$$J_l(u) = \sum_{m=0}^{\infty} \frac{(u/2)^{l+2m}}{m! \Gamma(l+m+1)}$$

— это обобщенная функция Бесселя первого рода, а

$$\varphi_{k0}(t) = 1 - \sum_{j=1}^{\infty} \varphi_{kj}(t).$$

Итак, мы получили в явном виде решение уравнений (28).

Теперь подсчитаем переходные вероятности для вложенной цепи Маркова $\{X_n^{(1)}\}$. Пусть d_j — вероятность того, что j автомобилей прибудут к перекрестку в первом направлении, пока для них горит красный свет, следовательно,

$$d_j = \int_0^{\infty} e^{-\lambda_1 y} (\lambda_1 y)^j (j!)^{-1} dG_2(y), \quad j = 0, 1, \dots$$

Введем также

$$b_{00} = 1, \quad b_{kj} = \int_0^{\infty} \varphi_{kj}(y) dG_1(y), \quad k, j > 0,$$

и положим $b_{k0} = 1 - \sum_{j=1}^{\infty} b_{kj}$.

Переходные вероятности вложенной цепи Маркова задаются формулами

$$P_{kj}^{(1)} = P(X_{n+1}^{(1)} = j | X_n^{(1)} = k) = \sum_{m=0}^j b_{km} d_{j-m},$$

$$k, j = 0, 1, \dots$$

Стационарное распределение $\{\pi_k^{(0)}\}$ вложенной цепи Маркова получается как решение следующей системы уравнений:

$$\pi_j^{(0)} = \sum_{k=0}^{\infty} \pi_k^{(0)} P_{kj}^{(1)}, \quad \sum_{j=0}^{\infty} \pi_j^{(0)} = 1.$$

Далее, стационарное распределение $\{\pi_j\}$ процесса $X_1(t)$ имеет вид

$$\pi_j = \theta^{-1} \sum_{i=0}^j \sum_{k=0}^{\infty} \pi_k^{(0)} b_{ki} \int_0^{\infty} e^{-\lambda_1 y} (\lambda_1 y)^{j-i} [(j-i)!]^{-1} \times$$

$$\times (1 - G_2(y)) dy +$$

$$+ \theta^{-1} \sum_{k=0}^{\infty} \pi_k^{(0)} \int_0^{\infty} \varphi_{kj}(y) (1 - G_1(y)) dy. \quad (29)$$

Используя эти результаты, нетрудно построить алгоритм для вычисления среднего числа автомобилей в стационарном режиме.

IV.4. Нерегулируемый перекресток как перекресток со случайными интервалами между переключениями светофора

Модель, изученная в разделе IV.3, может быть использована для анализа нерегулируемого перекрестка, рассмотренного в [15]. Мы имеем дело с пересечением двух перпендикулярных дорог, причем первое направление считается главным. Следовательно, автомобили, движущиеся в первом направлении, не останавливаются перед перекрестком. С другой стороны, автомобиль, прибывающий к перекрестку во втором направлении, может проехать только в том случае, когда на главной дороге нет автомобилей в интервале длины I перед перекрестком. Предположим, что скорость автомобилей в первом направлении равна V , следовательно, время прохождения интервала безопасности I будет I/V .

Остальные предположения такие же, как в предыдущих параграфах. А именно, потоки прибывающих автомобилей пуассоновские. Время прохождения перекрестка для первого направления равно нулю (автомобили движутся с постоянной скоростью, и можно считать, что для первого направления все время включен зеленый свет). Для второго направления можно считать, что зеленый свет загорается в тот момент, когда промежуток $[-I, 0]$ на главной дороге становится пустым. (Точка 0 означает перекресток.) Красный свет загорается в тот момент, когда в указанном интервале на главной дороге появляется автомобиль. Таким образом, для второго направления у нас будет светофор со случайными интервалами между переключениями.

Пусть $X_2(t)$ — это число автомобилей, ожидающих в момент t у перекрестка, чтобы проехать во втором направлении. Мы рассмотрим систему $M|D|\infty$, то есть бесконечно-канальную систему с входящим потоком интенсивности λ_1 и постоянным временем обслуживания I/V . Число клиентов в этой системе в момент t обозначим $q_1(t)$. Таким образом, «зеленый интервал» во втором направлении — это свободный период в рассматриваемой системе обслуживания, его функция распределения равна $G_2(x)$. А «красный интервал» — это период занятости в системе $M|D|\infty$, и его функция распределения равна $G_1(x)$. Так как входящий поток пуассоновский с интенсивностью λ_1 , мы получаем $G_2(x) = 1 - e^{-\lambda_1 x}$.

Следствие 2. Процесс $X_2(t)$ имеет собственное предельное распределение тогда и только тогда, когда

$$\lambda_2 e^{\lambda_1 I/V} < \nu. \quad (30)$$

Доказательство. Чтобы получить из теоремы 5 условия эргодичности для числа автомобилей $X_2(t)$, ожидающих во втором направлении, нам необходимо знать соответствующие средние

$\gamma_i^{-1} = \int_0^\infty x dG_i(x)$, $i = 1, 2$. Очевидно, $\gamma_2 = \lambda_1$ и в стационарном режиме

$$P(q_1(t) = 0) = \frac{\gamma_1}{\gamma_1 + \gamma_2} = e^{-\lambda_1 I/V},$$

поэтому

$$\gamma_1 = \lambda_1^{-1}(e^{\lambda_1 I/V} - 1) \text{ и } \theta = \gamma_1^{-1} + \gamma_2^{-1} = \lambda_1^{-1} e^{\lambda_1 I/V}.$$

Следовательно, согласно теореме 5 условия эргодичности для процесса $X_2(t)$ имеют вид (30).

Чтобы использовать результаты раздела IV.3 для построения алгоритма подсчета стационарного распределения $X_2(t)$, нам необходимо знать $G_1(x)$. Для этого используем следующий результат из [15]. Пусть $W(t)$ — время ожидания автомобиля, прибывающего к перекрестку во втором направлении, если других ждущих автомобилей нет. Доказано, что

$$F(x) = \lim_{t \rightarrow \infty} P(W(t) < x) = \sum_{j=0}^{k_x} (-1)^j e^{-\lambda_1(j+1)I/V} \left(\frac{\lambda_1^{j+1}(x - jI/V)^{j+1}}{(j+1)!} - \frac{\lambda_1^{j+1}(x - jI/V)^j}{j!} \right),$$

где $k_x = [xV/I]$ и $[\cdot]$ означает целую часть.

Как хорошо известно из теории восстановления (см., например, [22]):

$$F(x) = \frac{1}{\gamma_2 \theta} + \frac{1}{\gamma_1 \theta} \gamma_1 \int_0^x (1 - G_1(y)) dy, \quad (31)$$

откуда мы немедленно получаем

$$G_1(x) = 1 - \theta F'(x).$$

Для преобразования Лапласа $g_1^*(s) = \int e^{-sx} dG_1(x)$ результат записывается в более компактном виде. Без потери общности положим $I/V = 1$. Тогда, как показано в [15],

$$\int_{0-}^\infty e^{-sx} dF(x) = (s + \lambda_1) e^{-s\lambda_1} (s e^s + \lambda_1 e^{-\lambda_1})^{-1},$$

что дает вместе с (31) следующее выражение:

$$g_1^*(s) = (\lambda_1 + s)(\lambda_1 + s e^{s+\lambda_1})^{-1}.$$

Таким образом, нерегулируемый перекресток может быть исследован с помощью модели со светофором, имеющим случайные промежутки между переключениями света.

IV.5. Показательно распределенные интервалы между переключениями

Теперь предположим, что $G_i(x) = 1 - e^{-\gamma_i x}$ для $x > 0$, $i = 1, 2$, и автомобиль, прибывающий в то время, когда для него горит зеленый свет и

нет ожидающих автомобилей, немедленно проходит перекресток.

Рассмотрим цепь Маркова $\{X_1(t), e(t)\}$, где $e(t) = 1$, если для первого направления горит зеленый свет, и $e(t) = 0$, если горит красный свет. Далее предполагается, что условия (23) выполнены и процесс $\{X_1(t), e(t)\}$ стационарный. Обозначим

$$p_j = P(X_1(t) = j, e(t) = 1),$$

$$r_j = P(X_1(t) = j, e(t) = 0),$$

$$P(z) = \sum_{j=0}^\infty p_j z^j, \quad R(z) = \sum_{j=0}^\infty r_j z^j, \quad |z| \leq 1.$$

Используя уравнения для стационарного распределения процесса $\{X_1(t), e(t)\}$:

$$p_0 \gamma_1 = p_1 \nu + r_0 \gamma_2,$$

$$p_1(\gamma_1 + \nu + \lambda_1) = p_2 \nu + r_1 \gamma_2,$$

$$p_j(\gamma_1 + \nu + \lambda_1) = p_{j+1} \nu + p_{j-1} \lambda_1 + r_j \gamma_2, \quad j \geq 2,$$

$$r_0(\lambda_1 + \gamma_2) = p_0 \gamma_1,$$

$$r_j(\lambda_1 + \gamma_2) = r_{j-1} \lambda_1 + p_j \gamma_1, \quad j \geq 1,$$

мы получаем следующие соотношения для производящих функций $P(z)$ и $R(z)$:

$$P(z) = p_0 \frac{(\lambda_1 z - \nu)(\gamma_2 + \lambda_1(1 - z))}{(\lambda_1 z - \nu)(\gamma_2 + \lambda_1(1 - z)) + \lambda_1 \gamma_1 z},$$

$$R(z) = \frac{\gamma_1 P(z)}{\gamma_2 + \lambda_1(1 - z)}. \quad (32)$$

Поскольку $P(1) = P(e(t) = 1) = \gamma_2(\gamma_1 + \gamma_2)^{-1}$, легко видеть, что

$$p_0 = \frac{\nu \gamma_2 - \lambda_1(\gamma_1 + \gamma_2)}{(\gamma_1 + \gamma_2)(\nu - \lambda_1)}. \quad (33)$$

Вводя новые параметры $c_i = \lambda_i \nu^{-1}$, $x = (\gamma_1 \theta)^{-1}$, $d = \nu \theta$, можно переписать (23) в следующем виде:

$$c_i \geq 0, \quad i = 1, 2, \quad c_1 + c_2 < 1, \quad x \in (c_1, 1 - c_2),$$

здесь x — это доля времени, когда для первого направления горит зеленый свет.

Средняя длина очереди в первом направлении (то есть среднее число автомобилей, ждущих перед светофором), подсчитанная на основе (32) и (33), имеет вид

$$m_1(x) = P'(1) + R'(1) = \frac{c_1(1-x)(1+d(1-c_1)x(1-x))}{(1-c_1)(x-c_1)}.$$

Аналогичное выражение справедливо для второго направления

$$m_2(x) = \frac{c_2 x(1+d(1-c_2)x(1-x))}{(1-c_2)(1-x-c_2)}.$$

Пусть θ фиксировано, мы выбираем $x \in (c_1, 1 - c_2)$ таким образом, чтобы оно обеспечивало минимум $f(x) = m_1(x) + m_2(x)$. Иными словами, мы ищем

единственную точку, удовлетворяющую уравнению $m'_1(x) = -m'_2(x)$. Очевидно, что в симметричном случае $c_1 = c_2 = c$ будет $x_{\min} = 0,5$, а

$$f(0,5) = 2c[1 + 0,25d(1 - c)][(1 - c)(1 - 2c)]^{-1}.$$

Условия высокой загрузки возникают, если либо $x \downarrow c_1$, либо $x \uparrow 1 - c_2$, соответственно возникает длинная очередь ожидающих автомобилей в первом или во втором направлении. Кроме того, если $x \downarrow c_1$, то

$$m_1(x) \approx \frac{c_1(1 + c_1d(1 - c_1)^2)}{x - c_1}.$$

Замечание 1. Соотношения (29) позволяют найти средние числа автомобилей m_i , $i = 1, 2$, ожидающих у светофора в i -м направлении. Численные расчеты показывают, что средние для показательных промежутков между переключениями светофора много больше, чем для постоянных интервалов, если ρ_i близки к единице, в то время как для $\rho_i \approx 0,5$ они мало отличаются.

Некоторые результаты приведены в табл. 1, 2 и 3, где индекс d у средних m_i^d указывает, что речь идет о постоянных интервалах между переключениями, а индекс e соответствует показательному распределению, $m_{total} = m_1 + m_2$, $\theta = \gamma_1^{-1} + \gamma_2^{-1}$.

Т а б л и ц а 1

Средние длины очередей в первом направлении, $\nu = 15, \theta = 5$

λ_1	5	5	5	5	5	7	7	7	7	9	9	9	10	10
γ_1^{-1}	4,5	4,0	3,0	2,5	2,0	4,5	4,0	3,0	2,8	4,5	4,0	3,7	4,5	4,0
γ_2^{-1}	0,5	1,0	2,0	2,5	3,0	0,5	1,0	2,0	2,2	0,5	1,0	1,3	0,5	1,0
ρ_1	0,37	0,42	0,58	0,67	0,83	0,52	0,58	0,78	0,83	0,67	0,75	0,81	0,74	0,83
m_1^d	2,5	5,0	10,0	12,7	16,8	3,5	7,0	14,7	16,9	4,5	9,3	12,8	5,1	11,1
m_1^e	0,5	1,9	9,8	20,3	58,5	0,9	3,9	27,8	44,8	1,9	8,7	18,9	2,8	15,0

Т а б л и ц а 2

Средние длины очередей в двух направлениях,
 $\lambda_1 = \lambda_2, \nu = 20, \theta = 5$

λ_1	5	5	5	5	5	5	5	5	5	5
γ_1^{-1}	3,6	3,5	3,2	3,0	2,5	2,0	1,8	1,5	1,4	
γ_2^{-1}	1,4	1,5	1,8	2,0	2,5	3,0	3,2	3,5	3,6	
ρ	0,35	0,36	0,39	0,42	0,50	0,63	0,69	0,83	0,89	
m_1^d	7,0	7,5	9,0	10,0	12,7	16,8	20,0	26,5	28,4	
m_1^e	3,2	3,7	5,6	7,2	13,2	25,3	35,5	78,2	129,0	
m_2^d	28,4	26,5	20,0	16,8	12,7	10,0	9,0	7,5	7,0	
m_2^e	129,0	78,2	35,5	25,3	13,2	7,2	5,6	3,7	3,2	
m_{total}^d	35,4	34,1	29,0	26,8	25,4	26,8	29,0	34,1	35,4	
m_{total}^e	132,2	81,9	41,1	32,6	26,3	32,6	41,1	81,9	132,2	

Т а б л и ц а 3

Средние длины очередей в двух направлениях, $\lambda_1 \neq \lambda_2$,
 $\nu = 20, \theta = 5$

λ_1	5	5	5	5	5	5	5	5	5
λ_2	3	3	3	3	3	3	3	3	3
γ_1^{-1}	3,6	3,5	3,2	3,0	2,5	2,0	1,8	1,5	1,4
γ_2^{-1}	1,4	1,5	1,8	2,9	2,5	3,0	3,2	3,5	3,6
ρ	0,35	0,36	0,39	0,42	0,50	0,63	0,69	0,83	0,89
m_1^d	7,0	7,5	9,0	10,0	12,7	16,8	20,0	26,5	28,4
m_1^e	3,2	3,7	5,6	7,2	13,2	25,3	35,5	78,2	129,0
m_2^d	11,5	10,9	9,7	9,0	7,5	6,0	5,4	4,5	4,2
m_2^e	17,7	15,5	11,1	9,1	5,6	3,4	2,7	1,8	1,6
m_{total}^d	18,5	18,4	18,7	19,0	20,2	22,8	25,4	31,0	32,6
m_{total}^e	20,9	19,2	16,7	16,3	18,8	28,7	38,1	80,0	130,5

В частности, мы видим, что в симметричном случае, табл. 2, суммарное среднее числа ожидающих автомобилей достигает минимума при коэффициенте загрузки 0,5. А из табл. 3 ясно, что в несимметричном случае минимум достигается в

окрестности точки 0,42 для показательного распределения интервалов и 0,36 для постоянных.

IV.6. Функционирование светофора в условиях высокой загрузки

Мы будем рассматривать только поведение $X_1(t)$ при $\rho_1 \uparrow 1$, изучение $X_2(t)$ производится аналогично. Ситуация высокой загрузки может возникать различными способами. Предположим, что интенсивность поступления автомобилей зависит от малого параметра ε следующим образом: $\lambda_1^\varepsilon = (1 - \varepsilon)\lambda_1/\rho_1$. Это значит, что $\rho_1^\varepsilon = (1 - \varepsilon) \rightarrow 1$ при $\varepsilon \rightarrow 0$. Пусть $\{X_n^{(1)}(\varepsilon)\}_{n \geq 1}$ — стационарная вложенная цепь Маркова, соответствующая потоку автомобилей с интенсивностью λ_1^ε . Для удобства записи обозначим τ_i , $i = 1, 2$, случайную величину, имеющую такое же распределение, как $\tau_n^{(i)}$ при любом n .

Теорема 6. Если $E\tau_i^{2+\delta} < \infty$, $i = 1, 2$, для некоторого $\delta > 0$, то

$$P(\varepsilon X_n^{(1)}(\varepsilon) > y) \rightarrow \exp\{-2\nu\mu_1 y/\sigma^2\}, \text{ при } \varepsilon \rightarrow 0,$$

где $\sigma^2 = \lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1 + \lambda_1^2 D\tau_2 + (\lambda_1 - \nu)^2 D\tau_1$, а $\mu_i = \gamma_i^{-1}$, $\lambda_1 = \nu\mu_1\theta^{-1}$.

Схема доказательства. Пусть $\{\xi_n^\varepsilon\}_{n \geq 1}$ и $\{\eta_n\}_{n \geq 1}$ — две независимые последовательности случайных величин независимых и одинаково распределенных для каждой последовательности. Здесь ξ_n^ε — число клиентов (автомобилей), поступающих в систему в соответствии с пуассоновским потоком интенсивности λ_1^ε в течение интервала $\tau_n^{(1)} + \tau_n^{(2)}$, в то время как η_n — это число клиентов, поступающих в соответствии с пуассоновским потоком интенсивности ν в течение интервала $\tau_n^{(1)}$. Положим $S_0^\varepsilon = 0$, $S_k^\varepsilon = \sum_{j=1}^k (\xi_j^\varepsilon - \eta_j)$ и $S^\varepsilon = \sup_{k \geq 0} S_k^\varepsilon$. Мы устанавливаем, что для любого фиксированного $y \geq 0$:

$$P(\varepsilon X_n^{(1)}(\varepsilon) > y) / P(\varepsilon S^\varepsilon > y) \rightarrow 1 \text{ при } \varepsilon \rightarrow 0.$$

Как показано в книге [25], глава 4, теорема 18, распределение $\varepsilon S^\varepsilon$ при $\varepsilon \rightarrow 0$ сходится к показательному с параметром $2\nu\mu_1/\sigma^2$, где σ^2 приведено в условиях теоремы 6. Это заканчивает доказательство.

Следствие 3. В условиях теоремы 6

$$\varepsilon E X_n^{(1)}(\varepsilon) \rightarrow \sigma^2 / 2\nu\mu_1 \text{ при } \varepsilon \rightarrow 0.$$

Легко подсчитать, что в случае показательного распределения интервалов между переключениями $D\tau_j = \mu_j^2$, $j = 1, 2$, следовательно,

$$\begin{aligned} \sigma^2 &= \lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1 + \lambda_1^2\mu_2^2 + (\lambda_1 - \nu)^2\mu_1^2 = \\ &= \sigma_d^2 + \lambda_1^2\mu_2^2 + (\lambda_1 - \nu)^2\mu_1^2, \end{aligned}$$

где $\sigma_d^2 = \lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1$ — это коэффициент σ^2 для случая постоянных интервалов.

Нетрудно установить асимптотическое поведение средних чисел ожидающих автомобилей на главной трассе в условиях высокой загрузки.

Следствие 4. В случае постоянных интервалов между переключениями

$$m_1^d = E X_n^{(1)}(\varepsilon) \approx \frac{\lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1}{2\nu\mu_1\varepsilon}, \quad \varepsilon \rightarrow 0.$$

Для показательного распределенных интервалов

$$m_1^\varepsilon \approx \frac{\lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1}{2\nu\mu_1\varepsilon} + \frac{\lambda_1^2\mu_2^2 + (\lambda_1 - \nu)^2\mu_1^2}{2\nu\mu_1\varepsilon},$$

$$\varepsilon \rightarrow 0.$$

Очевидно, что в условиях высокой загрузки средняя длина очереди автомобилей существенно зависит от распределения интервалов между переключениями светофора. Так, для показательного случая

$$\frac{m_1^\varepsilon}{m_1^d} \approx 1 + \frac{\lambda_1^2\mu_2^2 + (\lambda_1 - \nu)^2\mu_1^2}{\lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1},$$

а для произвольного распределения интервалов τ_1 и τ_2 :

$$\frac{m_1^d}{m_1^d} \approx 1 + \frac{\lambda_1^2 D\tau_2 + (\lambda_1 - \nu)^2 D\tau_1}{\lambda_1\mu_2 + (\lambda_1 + \nu)\mu_1}.$$

Более того, отношение возрастает с ростом дисперсий.

Итак, можно сделать вывод, что среднее минимально для случая постоянных интервалов. Наша модель позволяет оптимальным образом выбрать соотношение между τ_1 и τ_2 .

V. Модель двух последовательных светофоров, учитывающая размер автомобилей

Пусть расстояние между двумя светофорами на главной дороге достаточно мало. В этом случае мы не можем, как делалось ранее, не принимать во внимание размер автомобилей.

V.1. Описание модели

Мы предполагаем, что пуассоновский поток клиентов (автомобилей) интенсивности λ поступает на первый обслуживающий прибор (светофор). Прибор может быть в рабочем состоянии (горит зеленый свет) или недоступен (красный свет). Соответствующие интервалы чередуются и образуют две независимые последовательности, каждая из которых состоит из независимых показательных распределенных случайных величин с параметрами соответственно γ_1 и γ_2 . Второй прибор (следующий светофор) функционирует аналогично, но соответствующие параметры равны γ_3 и γ_4 . Предполагается также, что оба прибора работают независимо. В отличие от предыдущих разделов мы теперь учитываем размеры автомобилей. Это означает, что перед вторым светофором найдется место только для k автомобилей.

Автомобиль, прибывающий к светофору в то время, когда горит зеленый свет и нет ожидающих автомобилей (прибор в рабочем состоянии и свободен), немедленно проезжает перекресток (время обслуживания равно нулю). В противном

случае автомобиль останавливается перед светофором или в конце очереди ожидающих проезда автомобилей. В этом случае, когда прибор снова начнет работать (зажжется зеленый свет), время обслуживания (проезд перекрестка) имеет показательное распределение с параметром ν . Более того, первый прибор блокируется, если заняты все места ожидания перед вторым прибором. Когда появляется свободное место, обслуживание заблокированного клиента (автомобиля) повторяется.

Функционирование такой двухфазной системы описывается цепью Маркова $X(t) = \{X_1(t), X_2(t), e_1(t), e_2(t)\}$, где $X_i(t)$ — число автомобилей перед i -м светофором (прибором), $e_i(t) = 1$, если i -й прибор в рабочем состоянии, и $e_i(t) = 0$ в противном случае, $i = 1, 2$.

В.2. Условия эргодичности

Для того чтобы получить необходимые и достаточные условия эргодичности рассматриваемой системы, мы введем вместо второго прибора вспомогательную одноканальную систему, которая будет представлять собой случайную среду для первого прибора.

Предположим, что на второй прибор поступает дважды стохастический пуассоновский поток со случайной интенсивностью $\lambda(t, \omega) = \nu e_1(t)$. Число мест для ожидания перед рассматриваемым прибором равно k . Другими словами, поведение вспомогательной системы совпадает с работой второго прибора, если очередь перед первым все время не пустая.

Пусть $u_2(t)$ — число клиентов во вспомогательной системе, а $e_i(t)$, $i = 1, 2$, были определены в предыдущем параграфе. Цепь Маркова $U(t) = \{u_2(t), e_1(t), e_2(t)\}$ эргодическая, так как число состояний конечно. Нам понадобятся следующие стационарные вероятности:

$$p_j = P(u_2 = j, e_1 = 1, e_2 = 1),$$

$$q_j = P(u_2 = j, e_1 = 0, e_2 = 1),$$

$$r_j = P(u_2 = j, e_1 = 1, e_2 = 0),$$

$$s_j = P(u_2 = j, e_1 = 0, e_2 = 0).$$

Используя прямую систему уравнений Колмогорова, мы получим систему уравнений для соответствующих производящих функций $P(z)$, $Q(z)$, $R(z)$ и $S(z)$:

$$\begin{aligned} & ((\gamma_1 + \gamma_2)z - \nu(1-z)^2)P(z) - \gamma_3 z R(z) - \gamma_2 z Q(z) = \\ & = -\nu p_0(1-z)^2 + \nu z^{k+1}(1-z)p_k, \\ & -\gamma_1 z P(z) + ((\gamma_2 + \gamma_3)z - \nu(1-z))Q(z) - \gamma_4 z S(z) = \\ & = -\nu q_0(1-z), \\ & -\gamma_3 P(z) + ((\gamma_1 + \gamma_4) + \nu(1-z))R(z) - \gamma_2 S(z) = \\ & = \nu r_k z^k(1-z), \\ & -\gamma_1 R(z) - \gamma_3 Q(z) + (\gamma_2 + \gamma_4)S(z) = 0. \end{aligned}$$

Введем дополнительные обозначения

$$a_1(z) = (\gamma_1 + \gamma_3)z - \nu(1-z)^2,$$

$$a_2(z) = \gamma_2(\gamma_2 + \gamma_3 + \gamma_4)(\gamma_2 + \gamma_4)^{-1}z - \nu(1-z),$$

$$a_3(z) = \gamma_4(\gamma_1 + \gamma_2 + \gamma_4)(\gamma_2 + \gamma_4)^{-1} + \nu(1-z),$$

$$b_1(z) = -\nu(1-z)p_0 + \nu z^{k+1}p_k,$$

$$b_2 = \nu q_0, \quad b_3(z) = \nu z^k r_k.$$

Положим также

$$g_1(z) = \gamma_1 z(\gamma_3 \gamma_4 + (\gamma_2 + \gamma_4)a_3(z)),$$

$$g_2(z) = \gamma_3(a_2(z)(\gamma_2 + \gamma_4) + \gamma_1 \gamma_2 z),$$

$$d_1(z) = (\gamma_2 + \gamma_4)(\gamma_1 z b_3(z) - \gamma_3 b_2)g_2(z)^{-1},$$

$$d_2(z) = \gamma_3^{-1}b_3(z) - \gamma_2(\gamma_2 + \gamma_4)^{-1}d_1(z),$$

наконец,

$$f(z) = [(a_1(z)a_3(z)\gamma_3^{-1} - \gamma_4 z)g_2(z) - (\gamma_2(\gamma_2 + \gamma_4)^{-1} \times \\ \times a_1(z) + \gamma_2 z)g_1(z)](1-z)^{-1}.$$

Решая уравнения для производящих функций, мы получим

$$R(z) = \frac{b_1(z) + \gamma_2 z d_1(z) - a_1(z)d_2(z)}{f(z)} g_2(z).$$

Нетрудно показать, что $f(z)$ — это полином третьей степени, имеющий три действительных корня z_i , $i = 1, 2, 3$. Поскольку функция $R(z)$ аналитическая, мы имеем три уравнения для определения неизвестных коэффициентов p_0 , q_0 , p_k , q_k :

$$b_1(z_i) + \gamma_2 z d_1(z_i) - a_1(z_i)d_2(z_i) = 0,$$

$$i = 1, 2, 3. \quad (34)$$

Четвертое уравнение — это условие нормировки

$$R(1) = \frac{\gamma_2 \gamma_3}{(\gamma_1 + \gamma_2)(\gamma_3 + \gamma_4)}. \quad (35)$$

Теорема 7. Цепь Маркова $X(t)$ эргодична тогда и только тогда, когда

$$\lambda < \nu(\gamma_2(\gamma_1 + \gamma_2)^{-1} - p_k - r_k), \quad (36)$$

где постоянные p_k , r_k удовлетворяют уравнениям (34) и (35).

Доказательство. Пусть $A(t)$ — это число автомобилей, которые прибыли к первому светофору до момента t , а $Y(t)$ — это число автомобилей, которые могли пройти перекресток (быть обслужены), если бы очередь была все время не пустая. Тогда $A(t)$ и $Y(t)$ независимы, $A(t)$ — пуассоновский процесс с параметром λ , в то время как $Y(t)$ — дважды стохастический пуассоновский процесс. Введенный выше процесс $U(t)$ может рассматриваться как случайная среда для первого прибора (светофора) согласно определению, данному в [26], кроме того

$$E(A(t)|U(s), 0 \leq s \leq t) = \int_0^t g(U(s)) ds = \lambda t,$$

$$E(Y(t)|U(s), 0 \leq s \leq t) = \int_0^t \varphi(U(s))ds = \\ = \nu \int_0^t I(u_2(s) < k, e_1(s) = 1)ds.$$

Поскольку у процесса $U(t)$ существует стационарное распределение π , нагрузка первого прибора, функционирующего в случайной среде, в соответствии с [26] имеет вид

$$\rho = \frac{E_\pi g(U)}{E_\pi \varphi(U)} = \frac{\lambda}{\nu(P(e_1 = 1) - P(e_1 = 1, u_2 = k))}.$$

В силу доказанной в [26] теоремы число автомобилей перед первым светофором стохастически ограничено тогда и только тогда, когда $\rho < 1$. Поскольку цепь Маркова $X(t)$ неприводимая и непериодическая, ее стохастическая ограниченность эквивалентна эргодичности.

Замечание 2. Найти явный вид решения уравнений (34)–(35) не удастся. Мы проанализировали влияние параметра k на условия эргодичности в симметрическом случае $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0,5$. Долгие и утомительные вычисления показали, что условие (36) при больших k может быть переписано в следующем виде:

$$\frac{\lambda}{\nu} < \frac{1}{2} - \frac{\nu + 2}{2k} + o(k^{-1}),$$

откуда при $k = \infty$ мы получаем $\lambda < \nu/2$, что согласуется с теоремой 5.

Интересно также рассмотреть тот случай, когда часть автомобилей, прибывающих к первому светофору во втором направлении (перпендикулярном к основной трассе), может завернуть, то есть после перекрестка следовать в первом направлении. При этом возникает следующая задача.

Предположим, что первый светофор расположен на пересечении двух перпендикулярных дорог. Пусть $\lambda_i, i = 1, 2$, — интенсивность пуассоновского потока, прибывающего к первому светофору по i -й дороге. Автомобили, проезжающие в первом направлении перекресток (в то время, как для них горит зеленый свет), продолжают движение ко второму светофору в том же направлении (вдоль главной дороги). С другой стороны, каждый из автомобилей, прибывающих к перекрестку по второй (второстепенной) дороге, с вероятностью β , независимо от других автомобилей, может повернуть на перекрестке на главную дорогу (в то время как для главного направления горит красный свет), а с вероятностью $1 - \beta$ продолжить движение в первоначальном направлении. Предположим, что светофоры работают синхронно, иначе говоря, зеленый (и красный) свет на них загорается одновременно, то есть $\gamma_3 = \gamma_1$ и $\gamma_4 = \gamma_2$. Такая система описывается цепью

Маркова $X(t) = (X_1(t), X_2(t), X_3(t), e(t))$. Здесь $X_1(t)$ и $X_2(t)$ имеют тот же смысл, что и ранее, а $X_3(t)$ — это число автомобилей перед первым светофором на второстепенном направлении, кроме того, $e(t) = 1$, если на обоих светофорах горит зеленый свет для главной дороги, в противном случае $e(t) = 0$.

Чтобы сформулировать условия эргодичности, введем вспомогательную одноканальную систему со входящим дважды стохастическим пуассоновским потоком, имеющим интенсивность $\lambda(t, \omega) = \nu I(e(t) = 1) + \beta \nu I(e(t) = 0)$. Остальные предположения сохраняются. В частности, $u_2(t)$ — это число клиентов во вспомогательной системе в момент t , и мы изучаем стационарный режим. Положим теперь $p_j = P(u_2(t) = j, e(t) = 1)$ и $q_j = P(u_2(t) = j, e(t) = 0)$. Производящая функция $P(z) = \sum_{j=0}^{\infty} p_j z^j$, как нетрудно показать, имеет следующий вид:

$$P(z) = \frac{p_0 v(z) + p_k g(z) + q_k w(z)}{-\nu \mu z^2 + (\gamma_1 \mu + \gamma_2 \nu + 2\mu \nu)z - \gamma_2 \nu - \nu \mu} \quad (37)$$

с $\mu = \beta \nu$ и $v(z) = -\nu(1 - z)(\gamma_2 + \mu(1 - z))$, $g(z) = \nu(\gamma_2 + \mu(1 - z)z^{k+1})$, $w(z) = \gamma_2 \mu z^{k+1} q_k$.

Знаменатель в выражении (37) имеет два действительных корня z_1 и z_2 . Значит, мы получаем два уравнения для определения коэффициентов p_0, p_k и q_k :

$$v(z_i)p_0 + g(z_i)p_k + w(z_i)q_k = 0, \quad i = 1, 2. \quad (38)$$

Третье уравнение получается из условия нормировки

$$P(1) = \frac{\gamma_2(p_k + \beta q_k)}{\gamma_1 \beta} = \frac{\gamma_2}{\gamma_1 + \gamma_2},$$

откуда немедленно вытекает

$$p_k + \beta q_k = \pi_2 \beta. \quad (39)$$

Вспомним, что $\pi_2 = P(e(t) = 0) = \gamma_1 / (\gamma_1 + \gamma_2)$. Решая уравнения (38)–(39), мы получаем

$$q_k = \frac{\beta \pi_2 (g(z_1)v(z_2) - g(z_2)v(z_1))}{w(z_2)v(z_1) - w(z_1)v(z_2) + \beta (g(z_1)v(z_2) - g(z_2)v(z_1))}, \quad (40)$$

$$p_k = \beta(\pi_2 - q_k). \quad (41)$$

Теорема 8. Система эргодична, если выполнены следующие неравенства

$$\lambda_1 < \nu(\gamma_2(\gamma_1 + \gamma_2)^{-1} - p_k), \\ \lambda_2 < \nu(\gamma_1(\gamma_1 + \gamma_2)^{-1} - \beta q_k), \quad (42)$$

где p_k и q_k задаются с помощью (40) и (41).

Для доказательства эргодичности достаточно установить стохастическую ограниченность $X_1(t)$ и $X_3(t)$. Достаточность первого из условий (42) для стохастической ограниченности $X_1(t)$ доказывается также, как в теореме 7, только

$r_k = 0$. Обратимся к процессу $X_3(t)$. Очевидно, что $X_3(t)$ — это длина очереди в одноканальной системе, функционирующей в случайной среде $U_2(t) = (u_2(t), e(t), j(t))$, где $j(t) = 1$, если автомобиль, приходящий к перекрестку во втором направлении, поворачивает на главную трассу. Вероятность такого события равна β . В противном случае полагаем $j(t) = 0$.

Пусть $V_2(t)$ — это число автомобилей, которые могут пройти до момента t через перекресток в перпендикулярном направлении (быть обслужены) в предположении, что прибор все время занят (имеется очередь автомобилей). Тогда

$$E(V_2(t)|U_2(s), 0 \leq s \leq t) = \\ = \nu \int_0^t (I(e(s) = 0, j(s) = 0) + I(e(s) = 0, \\ j(s) = 1, u_2(s) < k)) ds.$$

Нетрудно установить, что процесс $U_2(t)$ имеет собственное предельное распределение при $t \rightarrow \infty$ и $P(u_2(t) = k, e(t) = 0) \rightarrow \tilde{q}_k$. Таким образом, загрузка равна $\rho_2 = \lambda_2 \nu^{-1} (\gamma_1 (\gamma_1 + \gamma_2)^{-1} - \beta \tilde{q}_k)^{-1}$.

Это означает, что необходимое и достаточное условие стохастической ограниченности $X_3(t)$ задается вторым неравенством в (42), где вместо q_k поставлено \tilde{q}_k . Очевидно, что вероятность \tilde{q}_k зависит от поведения процесса $X_1(t)$. Например, если $X_1(t)$ стохастически неограничен, то число автомобилей $N_1(t)$, проходящих перекресток по главной трассе, является дважды стохастическим пуассоновским процессом с интенсивностью $\lambda_1(t, \omega) = \nu I(e(t) = 1)$. Более того, в любом случае такой дважды стохастический процесс доминирует $N_1(t)$, следовательно, $q_k \geq \tilde{q}_k$. А это значит, что второе условие (42) достаточно для стохастической ограниченности $X_3(t)$.

Замечание 3. Для $\beta = 0$ мы получаем предыдущую модель с синхронизацией светофоров. Устремив k к бесконечности в первом соотношении (42), мы приходим к условию эргодичности из теоремы 5.

VI. Выводы и направления будущих исследований

Мы рассмотрели ряд стохастических моделей транспортных потоков, для анализа которых полезными оказались методы теории очередей.

Для случая отдельного перекрестка, когда потоки прибывающих машин пуассоновские, установлены условия эргодичности в предположении, что интервалы между переключениями светофора имеют произвольное распределение (в частности, постоянны). Предложена процедура оптимизации функционирования светофора, что позволяет минимизировать суммарное среднее число ожидающих на перекрестке автомобилей. Численные

расчеты показывают, что при малой загрузке показательно распределенные интервалы дают лучший результат, чем постоянные, а при высокой загрузке ситуация обратная. Более того, если коэффициент загрузки близок к 0,5, то характеристики системы мало зависят от типа распределения. Именно поэтому здесь можно использовать явный вид средних $m_i(x)$, полученный для случая показательного распределения. Изучение поведения системы в условиях высокой загрузки показало, что в такой ситуации оптимально иметь постоянные интервалы между переключениями.

Исследована также плотность потока автомобилей, движущихся с двумя разными скоростями по трассе без светофоров и предложен алгоритм ее подсчета с заданной степенью точности. При наличии светофоров плотность рассмотрена в предположении, что все автомобили движутся с одной и той же скоростью.

Модель с двумя последовательными светофорами, учитывающая размеры автомобилей, выглядит достаточно сложной. Однако полученные условия эргодичности позволяют определять интенсивности потоков и/или средние интервалы между переключениями, предотвращающие образование пробок.

Для применения полученных в работе результатов необходимо также решить целый ряд статистических задач. Оценка параметров входящих потоков может проводиться традиционными методами. С другой стороны, для того чтобы оценить среднее число автомобилей, проходящих через перекресток, можно применить нетрадиционный подход, например, воспользоваться теорией клеточных автоматов.

Возникают и другие проблемы для будущих исследований. Так, для оптимизации работы светофоров можно использовать другие критерии. Например, зеленый свет может включаться, если число ожидающих автомобилей превысит определенный порог. Полезно также рассмотреть другие «точки задержки» на трассе, кроме светофоров. Это могут быть пункты оплаты за проезд, дорожно-транспортные происшествия или знаки дорожного движения, требующие уменьшения скорости. Интересно изучить неоднородные потоки автомобилей, влияние обгона и более сложную конфигурацию транспортной сети.

Работа выполнена при поддержке РФФИ, грант 10-01-00266.

Литература

1. *Greenshields B.D.* A Study of Highway Capacity // Proc. Highway Res. — 1935. — V. 14. — P. 448–477.
2. *Greenberg H.* An Analysis of Traffic Flows // Oper. Res. — 1959. — V. 7. — P. 79–85.
3. *Иносе Х., Хамада Т.* Управление дорожным транспортом. — М.: Транспорт, 1983.

4. *May A.D.* Traffic Flow Fundamentals. — Englewood Cliffs, New Jersey: Prentice Hall, 1990.
5. *Maerivoet S. and De Moor B.* Cellular automata models of road traffic // Phys. Rep. — 2005. — V. 419. — P. 1–64.
6. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. — 2003. — V. 11. — P. 903–930.
7. *Chowdhury D.* Vehicular Traffic: A System of Interacting Particles Driven Far from Equilibrium // arXiv: cond-mat /9910173 v1 [cond-mat. stat-mech] 12 Oct. 1999.
8. *Fuks H. and Boccara N.* Convergence to equilibrium in a class of interacting particle system evolving in discrete time // Phys. Rev. E. — 2001. — V. 64, 016117.
9. *Helbing D.* Traffic and related self-driven many-particle systems // Rev. Modern Phys. — 2001. — V. 73. — P. 1067–1141.
10. *Schadschneider A.* Statistical Physics of Traffic Flow. — arXiv: cond-mat /0007418 v1 [cond-mat. stat-mech] 26 Jul 2000.
11. *Cáceres F.C., Ferrari P.A. and Pechersky E.* A slow to start traffic model related to a M / M / 1 queue. — arXiv: cond-mat /0703709 v2 [cond-mat. stat-mech] 31 May 2007.
12. *Афанасьева Л.Г., Булинская Е.В.* Некоторые задачи для потоков взаимодействующих частиц // Сборник, посвященный 70-летию ректора МГУ В.А. Садовниченко. «Современные проблемы математики и механики». — 2009. — Т. 2. — С. 55–68.
13. *Bulinskaya E., Afanas'eva L.* Estimation of interacting particles density // Proceedings of IWAP. Compiègne. — 2008. — P. 1–6.
14. *Григелионис Б.И.* О сходимости сумм ступенчатых случайных процессов к пуассоновскому // Теория вероятностей и ее применения. — 1963. — Т. 8, в. 2. — С. 189–194.
15. *Tanner J.C.* The Delay to Pedestrians Crossing a Road // Biometrika. — 1951. — V. 38. — P. 383–392.
16. *Franken P., König D., Arndt M., Schmidt V.* Queues and Point Processes. — Berlin: Akademie-Verlag, 1981.
17. *Боровков А.А.* Эргодичность и устойчивость случайных процессов. — М.: Эдиториал УРСС, 1999.
18. *Kelly F.P. and Williams J. R.* Stochastic Networks // IMA Volumes in Mathematics and Applications. — V. 71. — New York: Springer, 1995.
19. *Боровков А.А.* Предельные теоремы для сетей обслуживания, I // Теория вероятностей и ее применения. — 1986. — Т. 31, В. 3. — С. 472–490.
20. *Боровков А.А.* Предельные теоремы для сетей обслуживания, II // Теория вероятностей и ее применения. — 1987. — Т. 32, В. 2. — С. 280–290.
21. *Foss S.G., Kalashnikov V.V.* Regeneration and renovation in queues // Queueing Systems. — 1991. — V. 8. — P. 211–223.
22. *Афанасьева Л.Г., Булинская Е.В.* Случайные процессы в теории массового обслуживания и управления запасами. — М.: Изд-во МГУ, 1980.
23. *Grandell J.* Doubly Stochastic Poisson Processes. // Lecture Notes in Mathematics. — V. 529. — New York: Springer, 1976.
24. *Феллер В.* Введение в теорию вероятностей и ее приложения. — Т. 2. — М.: Мир, 1967.
25. *Боровков А.А.* Вероятностные процессы в теории массового обслуживания. — М.: Наука, 1972.
26. *Афанасьева Л.Г.* Системы массового обслуживания с циклическими управляющими процессами // Кибернетика и системный анализ. — 2005. — № 1. — С. 54–68.

Поступила в редакцию 15.10.2010.

М.Л. Бланк

Институт проблем передачи информации им. А.А. Харкевича РАН

Синхронно обновляемые процессы с запретами в моделях транспортных потоков

Изучено несколько простых моделей транспортных потоков в виде процессов с запретами (как решеточных, так и с непрерывным пространством) и получены явные формулы для некоторых связанных с ними статистик. В частности, получена так называемая фундаментальная диаграмма, выражающая зависимость средней скорости движения от плотности частиц.

Ключевые слова: процессы с запретами, транспортные потоки, системы взаимодействующих частиц, каплинг.

I. Введение

Одним из естественных способов математического моделирования транспортных потоков представляется 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 имеются 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.

В работе будет изучено несколько простых моделей транспортных потоков этого типа (как решеточных, так и с непрерывным пространством) и получены явные формулы для некоторых связанных с ними статистик. В частности, будет 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 (напри-

мер, 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 этой 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 только как модели транспортных потоков, но и с чисто математической точки зрения. Первые результаты на эту тему были получены недавно в [5], где была разработана оригинальная техника, позволяющая изучать эргодиче-

ские (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.

II. 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 этих 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 – занятым 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 имеются 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 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, при $\forall n$ 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 соотношением $x_i^* = 1 - x_i$ 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 предыду-

щей тем, что на каждом шагу частица сдвигается вперед до следующей занятой позиции.

Теорема 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 близких 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 многополосного 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 этой 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 части-

цы (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 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).

III. 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 $r \geq 0$. 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 ча-

стицы, 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 осмысленна в задачах транспортного моделирования. В общем случае при анализе локальных скоростей обоих знаков определения становятся существенно сложнее, но как результаты, так и доказательства почти не изменяются (см. [5]).

Для неотрицательных локальных скоростей рассматриваемые нами 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:

$$\mathcal{N}(v_i^t, x^t) := \begin{cases} v_i^t & \text{если } v_i^t \leq \Delta_i^t, \\ \Delta_i^t & \text{в противном случае.} \end{cases}$$

¹Заметим, что $\{v_0^t\}$ может быть как траекторией детерминированного хаотического отображения $f : [0,1] \rightarrow [0,1]$, т.е. $v_0^{t+1} := v f^t(v_0^t/v)$, так и (несмотря на название) реализацией настоящей стохастической цепи Маркова.

Здесь важно отметить, что между любыми двумя конфигурациями частиц $x(r), \hat{x}(\hat{r})$ с общей последовательностью зазоров $\Delta := \{\Delta_i\}$ имеется взаимно однозначное соответствие φ :

$$\hat{x}_i(\hat{r}) = \varphi(x_i(r)) := x_i(r) - 2i(r - \hat{r}) \quad \forall i \in \mathbb{Z}.$$

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 $\exists \bar{v}(\gamma) := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \min(v_0^s, \gamma) \quad \forall \gamma > 0$ (п.н.);
- (b) $\{v_i^t\}$ являются н.о.р. (как по i , так и по t) случайными величинами.

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,¹ а о постановке (b) как о стохастической.

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 6).

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.

IV. 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 $|I_n| \xrightarrow{n \rightarrow \infty} \infty$ 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 1.

- (а) Если $\rho(x) < \infty$, то $\|x_n - x_m\| \xrightarrow{|n-m| \rightarrow \infty} 1/\rho(x)$.
- (б) 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 1. Верхняя/нижняя плотности $\rho_{\pm}(x^t)$ инвариантны относительно динамики, т.е. $\rho_{\pm}(x^t) = \rho_{\pm}(x^{t+1}) \forall t$.

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 2. Для любой конфигурации $x \in X$ выполнено $|V(x, j, t) - V(x, i, t)| \xrightarrow{t \rightarrow \infty} 0$ п.н. $\forall i, j \in \mathbb{Z}$.

Следствие 1. 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 3. Пусть $x \in X$ и рассматривается только чисто детерминированная постановка (т.е. $v_i^t \equiv v$). Предположим, что $\forall i, t \exists j > t : \Delta_j(x^t) < v$. Тогда $\forall i \exists t_i < \infty : \Delta_i(x^t) < 2v \forall t \geq t_i$.

V. Эргодические свойства

Сформулируем теперь основные результаты для процессов с запретами в непрерывном пространстве.

Теорема 4. 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. Фундаментальная диаграмма в детерминированной постановке

$$V(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \min(1/\rho, v_0^s) = \begin{cases} v, & \text{если } \rho(x) \leq 1/v, \\ 1/\rho(x) & \text{в противном случае,} \end{cases} \quad \text{если } v_0^t \equiv v.$$

Следствие 2. 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 2. Последний результат совпадает с соответствующим утверждением о процессе на решетке, описанном в теореме 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 различно: 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.² 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 (обозначение $h_{\text{top}}(T_v)$) как супремум по всем метрическим энтропиям этого отображения относительно его вероятностных инвариантных мер.

Теорема 6. 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:

$$(\sigma_v x)_i := x_i + v \quad i \in \mathbb{Z}, x \in X.$$

Лемма 4. 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.

VI. 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 па-

ры 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:

$$P((x^t, y^t) \in A \times X) = P(x^t \in A)$$

и

$$P((x^t, y^t) \in X \times A) = P(y^t \in A),$$

т.е. проекции нового процесса пар ведут себя точно так же, как исходные процессы.

Обсудим теперь конструкцию динамического каплинга между двумя копиями x^t, \hat{x}^t рассматриваемого нами марковского процесса. Обычно при анализе систем взаимодействующих частиц на решетке с асинхронными взаимодействиями используется так называемый *равный* каплинг (см., напр., [14]). В этом случае каплинг состоит в спаривании частиц процессов x^t, \hat{x}^t , занимающих одинаковое 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, \hat{x}^t , обгоняют друг друга, но при этом ни в какой момент времени не занимают одинаковые позиции. Более 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: $\bullet^\circ \longrightarrow \circ \circ$. 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 \hat{x} -частицам.

Чтобы обойти это препятствие, мы и вводим *динамический*³ каплинг, описанный в [5, 6]. Отметим для сравнения идейно близкую конструкцию каплинга в [1, 11], предложенную для случая решеточных систем с асинхронными взаимодействиями. Важным преимуществом 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, \hat{x}^t понимается последовательное спаривание достаточно близко расположенных частиц в разных

²Обычно говорят, что отображение хаотическое, если его топологическая энтропия положительна, поэтому бесконечное значение энтропии говорит об очень высоком уровне хаотичности.

³Слово *динамический* используется для того, чтобы подчеркнуть то, что взаимное положение частиц в одной паре меняется со временем, в отличие от равного каплинга.

процессах, удовлетворяющее следующим условиям:

- (A1) В момент $t = 0$ все частицы предполагаются неспаренными. Локальные скорости взаимно спаренных частиц всегда одинаковы.
- (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, двигаются 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. 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 \acute{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 ($\circ \bullet$ или $\bullet \circ$) в процессе пар (x^t, \acute{x}^t) понимаются две взаимно спаренные частицы и x -дефект, находящийся между ними, индекс которого отличается на единицу от индекса спаренной x -частицы. \acute{x} -тройка ($\bullet \circ$ или $\circ \bullet$) определяется аналогично.

Говорят, что две пары частиц *пересекают* друг друга, если прямые линии, соединяющие позиции частиц из одной пары, пересекаются, т.е. $\bullet \star \bullet$ (здесь взаимно спаренные частицы обозначены одинаковыми символами).

x -дефект в x_i^t вместе с ближайшим к нему⁴ \acute{x} -дефектом в \acute{x}_j^t (\circ или \bullet) назовем d -парой, если $|x_i^t - \acute{x}_j^t| < v$, эта пара дефектов не пересекается с другими взаимно спаренными частицами, и интервал (x_i^t, \acute{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 $\circ \bullet \bullet$ две первых x -частицы вместе с первой \acute{x} -частицей образуют x -тройку, несмотря на наличие дополнительной спаренной частицы в интервале между ними. С другой стороны, набор $\bullet \circ \circ$ не содержит ни троек, ни d -пар.

Пара конфигураций (x^t, \acute{x}^t) называется *правильной*, если она не содержит x - или \acute{x} -троек, d -пар и пересекающихся взаимно спаренных частиц.

Правильность пары конфигураций (x^t, \acute{x}^t) в момент t в общем случае не препятствует тому, что под действием динамики в момент $(t + 1)$ пара $(x^{t+1}, \acute{x}^{t+1})$ может оказаться неправильной. В частности, могут возникнуть тройки обоих типов и d -пары, например $\bullet \circ \bullet \rightarrow \bullet \circ \bullet$ или $\circ \circ \circ \rightarrow \circ \circ \circ$. Здесь важно, что ввиду сохранения порядка частиц пересекающиеся взаимно спаренные частицы не могут появиться.

Лемма 5. Пусть пара конфигураций (x^t, \acute{x}^t) не имеет пересекающихся взаимно спаренных частиц. Тогда тройки одного типа не могут иметь общих элементов.

Поэтому все тройки одного типа могут быть устранены одновременно. Под *устранением* x - или \acute{x} -тройки будем понимать следующее: бывший дефект спаривается с частицей из другого процесса, а спаренная ранее частица становится неспаренной: $\circ \bullet \bullet \rightarrow \bullet \circ \bullet$.

Устранение d -пары еще проще: дефекты «аннигилируют» друг друга, образуя спаренную пару частиц: $\circ \circ \rightarrow \bullet \bullet$. 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: $\circ \bullet \bullet \rightarrow \bullet \circ \bullet$.
- (2) Каждая \acute{x} -тройка рекурсивно устраняется: $\bullet \circ \circ \rightarrow \bullet \circ \circ$.
- (3) Наименьшая⁵ d -пара рекурсивно устраняется: $\circ \circ \rightarrow \bullet \bullet$.

Лемма 6. 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 \acute{x} -тройки может привести к созданию новой тройки того же типа:

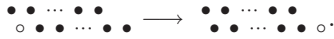
$$\circ \bullet \bullet \rightarrow \bullet \circ \bullet \rightarrow \bullet \bullet \circ$$

Заметим теперь, что при рекурсиях в процедуре каплинга дефект может сдвинуться на произ-

⁴Если имеется несколько ближайших \acute{x} -дефекта, то выбирается тот, который имеет минимальный индекс.

⁵Порядок d -пар может меняться после каждой процедуры рекурсии.

вольно большое расстояние от его начальной позиции:



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, \acute{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., напр., [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 7. Пусть $x, \acute{x} \in X$ с $\rho(x) = \rho(\acute{x}) > 0$, и предположим, что имеет место почти удачный каплинг (x^t, \acute{x}^t) , удовлетворяющий дополнительному условию о том, что расстояния между взаимно спаренными частицами равномерно ограничены сверху величиной $\gamma(t) = o(t)$. Тогда

$$|V(x, 0, t) - V(\acute{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0.$$

VII. Схема доказательства основных эргодических результатов

Начнем с двух технических результатов.

Лемма 8. Супремум $|W_{ij}^t| := x_i^t - \acute{x}_j^t$ по всем взаимно спаренным частицам при динамическом каплинге (см. раздел 6) процессов (x^t, \acute{x}^t) равномерно ограничен величиной v для любого $t \in \mathbb{Z}_0$.

Лемма 9. Пусть $\rho(x) = \rho(\acute{x})$ и пусть при каплинге $\forall i, j$ найдется такой (случайный) момент времени $t_{ij} < \infty$, что $x_i^t > \acute{x}_j^t$ для любого $t \geq t_{ij}$. Тогда каплинг почти успешен.

При наших предположениях (стандартный удачный каплинг⁶ может не существовать (например, в случае чисто детерминированной постановки с двумя равно распределенными начальными конфигурациями, сдвинутыми друг относительно друга). Поэтому лемма 8 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 4 состоит в следующем. Рассмотрим две произвольные допустимые конфигурации x, \acute{x} одинаковой плотности $\rho > 0$. Если предположить, что имеется почти удачный динамический каплинг процесса пар с

начальными условиями x, \acute{x} , то по лемме 8 выполняются условия леммы 7, откуда следует, что $|V(x, 0, t) - V(\acute{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0$. Применение леммы 2 доказывает наше утверждение.

Предположим теперь, что нет почти удачного динамического каплинга. Определим новые случайные величины

$$W_{ij}^t := x_i^t - \acute{x}_j^t, i, j \in \mathbb{Z}, t \in \mathbb{Z}_0.$$

Тогда

$$V(x, i, t) - V(\acute{x}, j, t) = W_{ij}^t/t - W_{ij}^0/t.$$

Согласно лемме 2 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 следующие ситуации:

- (a) $\lim_{t \rightarrow \infty} W_{00}^t/t = 0$. Тогда $|V(x, 0, t) - V(\acute{x}, 0, t)| \leq |W_{00}^t|/t + |W_{00}^0|/t \xrightarrow{t \rightarrow \infty} 0$, что по следствию 1 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 \acute{x} -процесса, исходно расположенную правее точки x_i^0 . Это наблюдение вместе с условием равенства плотностей позволяет применить лемму 9, согласно которой каплинг почти успешен. С другой стороны, по лемме 8 расстояния между взаимно спаренными частицами не могут превышать величины v . Поэтому по лемме 7 имеем $|V(x, 0, t) - V(\acute{x}, 0, t)| \xrightarrow{t \rightarrow \infty} 0$, что противоречит предположению (b).
- (c) $\limsup_{t \rightarrow \infty} W_{00}^t/t < 0$. Меняя роли процессов x^t, \acute{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 4 о том, что конфигурации одинаковой плотности имеют одинаковые средние скорости, получаем требуемое утверждение. □

Работа была частично поддержана грантами РФФИ.

⁶Когда почти все частицы со временем оказываются спаренными.

Литература

1. *Angel O.* The Stationary Measure of a 2-type Totally Asymmetric Exclusion Process // J. Combin. Theory Ser. A — 2006 — V. 4. — P. 625–635.
2. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. — 2003. — V. 3–4. — P. 903–930.
3. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. — 2005. — V. 3–4. — P. 627–658.
4. *Blank M.* Travelling with/against the Flow. Deterministic Diffusive Driven Systems // J. Stat. Phys. — 2008. — V. 4. — P. 773–796.
5. *Blank M.* Metric properties of discrete time exclusion type processes in continuum // J. Stat. Phys. — 2010. — V. 1. — P. 170–197.
6. *Blank M., Pirogov S.* On quasi successful couplings of Markov processes // Problemy Peredachi Informacii — 2007. — V. 4. — P. 316–330.
7. *Borodin A., Ferrari P.L., Sasamoto T.* Large time asymptotics of growth models on space-like paths II: PNG and parallel TASEP — 2007.
8. *D. Chowdhury, L. Santen, A. Schadschneider.* Statistical physics of vehicular traffic and some related systems // Physics Reports — 2000. — V. 329. — P. 199–329.
9. *Comtet A., Majumdar S.N., Ouvry S. and Sabhapandit S.* Integer partitions and exclusion statistics: limit shapes and the largest parts of Young diagrams // J. Stat. Mech. — 2007.
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 — 2008.
12. *Gray L., Griffeth D.* The ergodic theory of traffic jams // J. Stat. Phys. — 2001. — V. 3/4. — P. 413–452.
13. *Kornfeld E.P., Sinai Ya.G., Fomin S.V.* Ergodic theory — M.: Nauka, 1980.
14. *Liggett T.M.* Interacting particle systems // Springer-Verlag — NY, 1985.
15. *Maerivoet S., De Moor B.* Cellular Automata Models of Road Traffic // Physics Reports — 2005. — V. 1. — P. 1–64.
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.

Поступила в редакцию 15.10.2010.

П.П. Бобрж

Институт проблем транспорта им. Н.С. Соломенко РАН

Обоснование гравитационной модели транспортных корреспонденций при помощи закона убывающей предельной полезности

В статье исследуется связь между поведенческими реакциями экономического типа и часто возникающими на практике примерами пространственного распределения транспортных корреспонденций.

Ключевые слова: транспорт, гравитационная модель, закон убывающей предельной полезности, полезность поездки.

I. Транспортная активность

Рассмотрим модель, где из некоторой точки отправления v совершается $A = \text{const}$ поездок в единицу времени в некоторое конечное число пунктов u_i , $i = 1, \dots, n$. Каждая поездка происходит по маршруту (v, u_i, v) без попутного посещения других пунктов. При этом будем предполагать, что каждая точка u_i посещается равномерно через некоторый период времени T_i .

Если обозначить переменной ν_i долю потока из точки v в точку u_i , то

$$T_i = \frac{1}{A\nu_i}.$$

Величина A является внешним параметром модели, которая определяется не из каких-то вычислений, а исходя исключительно из специфики реальной исследуемой транспортной системы. Она является характеристикой транспортной активности в точке v .

В общем случае движение из точки v полностью описывается последовательностью времен отправлений, времен прибытий с указанием точки прибытия и времен возврата с точку v . Далее в статье будет предполагаться, что время в пути до точки u_i пренебрежимо мало по сравнению с интервалом T_i между поездками. Это позволяет в первом приближении рассматривать лишь такие расписания движения, которые задаются только временами отправления, то есть движение до точки и возврат обратно в модели происходит мгновенно.

Хотя A предполагается постоянной величиной, но из этого, вообще говоря, не следует, что транспортные средства должны выходить из пункта v равномерно через одинаковые промежутки времени с интервалом $T = 1/A$. Например, первая машина вышла через треть периода, а вторая — через две трети. Тогда всего за период выйдет две машины. Т. е. величину A скорее надо понимать как некоторую усредненную по времени интенсивность исходящего потока.

II. Гравитационная модель

Для того чтобы полностью определить динамику рассматриваемой модели, требуется определить множество периодов $\{T_i\}$, $i = 1, \dots, n$ посещения каждой точки u_i . Или, что равносильно, частоты ν_i ее посещения. В данной модели предлагается это сделать при помощи широко распространенной как в теории, так и при практических исследованиях гравитационной модели. См., например, [3, 4].

В разных источниках под ней понимаются хотя и похожие, но все-таки порой отличающиеся друг от друга формулы в зависимости от решаемой задачи. Поэтому приведем точные формулировки, которые далее будут использоваться в данной статье.

Обозначим символом n_i численность населения пункта u_i , и пусть время пути из пункта v в пункт u_i равно $t_i = t(v, u_i)$. Присвоим каждому пункту u_i вес по формуле

$$m_i = \frac{n_i}{(t_i)^2},$$

и пусть $M_v = \sum_i m_i$ — сумма всех весов пунктов корреспонденций из точки v .

Будем предполагать (и это и является сутью гравитационной модели), что поездки из пункта v распределяются по остальным пунктам таким образом, чтобы доля ν_i поездок в точку u_i была бы равна

$$\nu_i = \frac{m_i}{M}.$$

Отметим, что в таком виде (есть и другие формулировки [5, 6]) гравитационная модель указывает, как делится исходящий поток по пунктам назначения, но ничего не говорит о том, какой должна быть сама интенсивность A .

Гравитационная модель является эмпирическим наблюдением над реальными потоками. Но ее точность невелика. Порой ошибки, особенно на малодетальных направлениях, могут достигать сотен процентов.

Но тем не менее пока ничего лучше не существует. При этом модель была апробирована в

большом числе исследований, и для большинства из них она оказалась лучшим первым приближением. К тому же ошибки, как правило, ложатся в разные стороны от теоретических формул, то есть в среднем по многим экспериментам гравитационная модель является намного более точной. Поэтому, несмотря на ее значительные погрешности, она де-факто является стандартом.

Более того, она сама начала выступать как эмпирическое наблюдение, результат непосредственного опыта. Поэтому те теоретические рассуждения, которые будут противоречить ей, скорее всего надо признать неверными, как не соответствующие объективной реальности. И наоборот, те, которые хорошо согласуются с ней, получают дополнительное подтверждение своей адекватности.

В данной статье гравитационная модель используется для определения интенсивностей потоков к пунктам посещения для рассматриваемой идеальной модели. Но целью статьи является получение тех же формул, исходя из более общих законов человеческого поведения. Например, исходя из экономической целесообразности.

III. Закон убывания предельной полезности

В настоящее время не существует общепризнанного теоретического обоснования гравитационной модели, исходя из каких-либо поведенческих закономерностей. Отчасти такая ситуация сложилась по причине трудности и неоднозначности оценки экономических результатов большинства поездок, которые могут быть совершены, например, в личных целях. Но то, что распределение корреспонденций должно порождаться (по крайней мере, в среднем по всему множеству поездок) более простыми поведенческими стереотипами, прежде всего экономическими — весьма вероятно.

В статье предлагается объяснить распределение транспортных потоков исходя из хорошо известного экономического закона убывающей предельной (marginal) полезности продуктов. Будет строго определено понятие полезности поездки, и в соответствии с ним потоки будут выбираться так, чтобы полезность от совокупности всех поездок за период была бы наибольшей. Другими словами, гравитационная модель будет являться экстремальным решением исходя из некоторого общего принципа.

Напомним, что в соответствии с законом убывания предельной полезности предельная полезность от каждой новой порции продукции или услуг снижается. Самая распространенная популярная интерпретация этого закона заключается в том, что полезность каждого следующего бутерброда меньше предыдущего.

В такой формулировке закон является неким принципом, а не строгим детерминированным со-

отношением. Требуется строго определить факторы, от которых зависит полезность поездки, а также сам вид функции полезности, чтобы его можно было бы использовать в вычислениях. При этом функция полезности должна подбираться так, чтобы в результате выполнялась бы классическая гравитационная модель транспортных корреспонденций.

IV. Полезность поездки

Качество жизни как набор удовлетворяемых потребностей является во многом гуманитарным понятием, не допускающим однозначного формализованного определения [1, 2]. Оно также является субъективным, то есть различным для разных личностей при одних и тех же внешних условиях. Понимая ограниченность любого подхода к этому определению, предложим собственную трактовку, которая имеет своей целью оценку зависимости качества жизни от пространственного местоположения.

Прежде всего определим, зачем совершаются поездки.

Будем считать, что удовлетворение тех или иных потребностей человека происходит только в результате контактов с другими людьми. При этом чем больший доход имеет человек, с которым контактируют, тем больше услуг и (или) более качественные услуги он может оказать. Поэтому ценность контактов при посещении точки u_i тем больше, чем выше доходы d_i населения на территории, а также чем больше число n_i проживающих или посещающих данную территорию людей.

Отметим, что при таком подходе многие необитаемые территории в виде лесов, парков, полей, дорог и даже озер также будут являться полезными для поездок, если там систематически появляются люди.

Далее. В соответствии с законом убывания предельной полезности, чем чаще посещается то или иное место, тем меньше полезности эта поездка приносит. Другим словами, полезность поездки тем выше, чем реже посещается место u_i или чем больше период его посещения T_i .

Оставляя за скобками вопросы, как была получена требуемая зависимость, будем предполагать в статье, что предельная полезность поездки r_i в точку u_i (то есть полезность только от одной последней поездки, без учета полезности предыдущих поездок) прямо пропорциональна корню квадратному от произведения численности населения, дохода населения и периода посещения:

$$r_i = C\sqrt{n_i d_i T_i}.$$

Здесь C — некоторый коэффициент пропорциональности, постоянный для всех точек посещения u_i , который выбором величин размерностей положим равным единице.

Чем выше предельная полезность r_i в точке, тем более желательна поездка к ней и тем более продуктивной будет сама поездка. Специально отметим, что предельная полезность поездки r_i при таком подходе не является величиной с денежной размерностью, а это, скорее, некоторый интегральный усредненный показатель.

Поскольку предельная полезность посещения каждого пункта растет со временем, то рано или поздно наступит такой момент, когда посещение данного пункта станет самым выгодным вариантом. Это значит, что, выбирая в качестве поведенческой реакции критерий максимальной полезности поездки, все пункты будут посещаться.

И наконец, отметим, что при таком определении полезности закон уменьшающейся предельной полезности выполняется также и для параметров численности населения n_i , и дохода населения d_i .

Далее от благ, которые позволяет получить поездка, перейдем к рассмотрению затрат, которые требуется для этого понести.

V. Выбор поездок

Отличительной особенностью данного подхода является тот факт, что для повышения уровня жизни недостаточно увеличивать предложение тех или иных товаров или (и) услуг. Надо еще обеспечить подъезд до мест их доступности.

Ранее были сделаны предположения, что полезность от поездки пропорциональна трем факторам: численности населения n_i , доходу населения d_i и периоду посещения T_i . Однако чтобы получить эту полезность, следует доехать до пункта назначения, что требует затратить некоторое время $t_i = t(v, u_i)$.

Рассматривая при таком подходе поездку как обычное экономическое действие, мы получаем, что удельная на единицу времени поездки предельная полезность e_i этой операции должна быть равна отношению выгоды от поездки к затратам на ее приобретение:

$$e_i = \frac{r_i}{t_i} = \frac{\sqrt{n_i d_i T_i}}{t_i}.$$

Далее величину e_i будем называть просто удельной полезностью.

В соответствии с экономическим правилом максимизации полезности всего множества поездок они должны совершаться с такими периодами T_i , что удельная полезность от посещения каждого пункта была бы равной:

$$e_i = K = \text{const}, \quad i = 1, \dots, N.$$

Тем самым показатель $K = K(v)$ является характеристикой пункта v . Главными отличиями этого показателя от других, известных в литературе показателей транспортной доступности, являются учет фактора доходов населения, а также

зависимость от периода посещения. Это позволяет объяснить ряд наблюдаемых на практике явлений. Например, замыкание значительной доли транспортных коммуникаций в богатых областях.

Покажем, что подобные предположения приводят к гравитационной модели корреспонденций.

VI. Обоснование гравитационной модели

Делая необходимые преобразования в выражении для удельной полезности e_i , получаем следующую формулу для периода посещения:

$$\sqrt{n_i d_i T_i} = K t_i$$

и

$$T_i = \frac{(K t_i)^2}{n_i d_i}.$$

Поскольку период посещения обратно пропорционален доли поездок до пункта u_i из пункта v , то

$$A \nu_i = \frac{n_i d_i}{(K t_i)^2}.$$

Или, по-другому, доля поездок до пункта u_i с точностью до постоянного множителя обратно пропорциональна квадрату времени поездки t_i и прямо пропорциональна общему доходу пункта:

$$\nu_i = \frac{1}{AK^2} \frac{n_i d_i}{t_i^2} = \frac{1}{AK^2} \frac{n_i d_i}{t_i^2}.$$

Если мы положим $d_i \equiv D$ для всех пунктов одинаковым, что означает, что все люди равны по своей коммуникационной полезности (в гравитационной модели люди вообще не различаются по доходам), то доли поездок получаются такими же, как если бы они рассчитывались в соответствии с гравитационной моделью:

$$\nu_i = \frac{D}{AK^2} \frac{n_i}{t_i^2} = \frac{1}{M} \frac{n_i}{t_i^2},$$

где $M = \frac{AK^2}{D}$ — константа, равная сумме всех весов точек корреспонденций в гравитационной модели.

Другими словами, гравитационная модель является частным случаем предложенной в данной работе поведенческой модели распределения потоков.

VII. Пример

Рассмотрим простейшую транспортную систему, состоящую из трех точек, как показано на рис. 1. Левый пункт будем считать первым, правый — третьим.

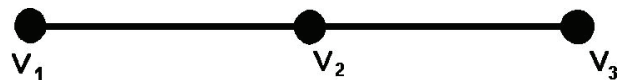


Рис. 1. Простейшая транспортная сеть из трех узлов

Будем предполагать, что времена движения в смежную вершину равны единице, численности

населенных пунктов равны единице, доходы населения также во всех узлах равны единице. Исходящий поток из всех вершин равен единице.

1. Рассчитаем вначале потоки из среднего узла.

Весы обоих вершин корреспонденций равны

$$m_1 = m_3 = \frac{1}{1} = 1, M = 2.$$

Соответственно равны доли поездок в левый и правый узел:

$$\nu_1 = \nu_3 = 0,5.$$

Это означает, что в первый период времени одна машина идет в левый пункт, а во второй период — в правый. Далее поездки повторяются.

В этом случае периоды посещения также равны между собой и равны 2:

$$T_1 = T_3 = \frac{1}{\nu_i} = 2.$$

Удельные полезности поездок снова равны в обоих случаях

$$e_1 = e_3 = \frac{\sqrt{1 \cdot 1 \cdot 2}}{1} = \sqrt{2} \approx 1,4142.$$

Таким образом, коэффициент K для центральной точки приблизительно равен 1,41.

2. Рассмотрим теперь потоки из левого узла. Для правого узла ввиду симметрии системы формулы будут аналогичны.

Весы узлов равны соответственно

$$m_2 = \frac{1}{1} = 1, \quad m_3 = \frac{1}{2} = 0,25,$$

$$M = m_1 + m_3 = \frac{5}{4} = 1,25.$$

Доли потоков равны

$$\nu_2 = \frac{4}{5}, \quad \nu_3 = \frac{1}{5}.$$

Периоды посещения

$$T_2 = \frac{5}{4}, \quad T_3 = 5.$$

Таким образом, второй узел будет посещаться в четыре раза чаще третьего, хотя времена подъезда отличаются всего в два раза.

Полезности поездок будут равны

$$e_2 = \frac{\sqrt{1 \cdot 1 \cdot 1,25}}{1} = \frac{\sqrt{5}}{2} \approx \frac{2,236}{2} = 1,118,$$

$$e_3 = \frac{\sqrt{1 \cdot 1 \cdot 5}}{2} = \frac{\sqrt{5}}{2} \approx \frac{2,236}{2} = 1,118.$$

Таким образом, коэффициент окружения для граничных точек сети равен приблизительно 1,118.

Сравнивая коэффициенты окружения между собой можно сделать вывод, что при одних и тех же предположениях только за счет своего местоположения общественная эффективность транспортного сообщения в центре будет на 0,296, то есть почти на треть выше, чем на окраинах.

Это позволяет с качественной точки зрения объяснить, почему жилье в центре городов, как правило, дороже, чем на периферии.

Отметим, что для количественного предсказания должного разброса цен на различные объекты, например на недвижимость, надо дополнительно учитывать другие факторы, в том числе и не транспортного характера.

VIII. Выводы

В предложенной работе изучались поведенческие реакции выбора направлений поездок, основанных на экономическом законе убывания предельной полезности. Показано, что один из частных случаев функционирования подобных систем приводит к широко распространенной на практике гравитационной модели корреспонденций.

Литература

1. *Байсеркеев О.Н., Бугроменко В.Н.* Региональная пространственно-временная среда. — Алматы: Рауан. — 1993. — 243 с.
2. *Тархов С.А.* Эволюционная морфология транспортных сетей. — Смоленск: Универсум, 2005. — 384 с.
3. *Бобрик П.П.* О преимуществе треугольной топологии сети над квадратной // Транспорт: наука, техника, управление. — 2005. — № 3. — С. 32–34.
4. *Agassiant A. A, Strelnikov A.I.* Rational Development of urban transportation systems, with due consideration given to environmental protection. — М., 1989. — 97 p.
5. *Стенбринк П.А.* Оптимизация транспортных сетей. — М.: Транспорт, 1981.
6. *Вильсон А. Дж.* Энтропийные методы моделирования сложных систем. — М.: Наука, 1978.

Поступила в редакцию 15.10.2010.

А.С. Бугаев¹, А.П. Буслаев^{2,3}, А.Г. Таташев^{3,2}, М.В. Яшина^{3,2}

¹ Московский физико-технический институт (государственный университет)

² Московский автомобильно-дорожный институт (государственный технический университет)

³ Московский технический университет связи и информатики

Оптимизация частично-связных потоков в детерминированно-стохастической модели

Рассматривается модель многополосного частично связанного движения, в которой скорость движения частиц представляет собой сумму детерминированной и стохастической составляющей. Решается задача максимизации средней скорости и интенсивности движения по параметрам, определяющим модель.

Ключевые слова: стохастические модели, случайное блуждание, автотранспортные потоки, многополосное движение, оптимизация.

I. Введение

В [1] рассмотрена модель движения частиц (автотранспортных средств) по фрагменту дороги с несколькими полосами. В этой модели скорость движения является суммой детерминированной и стохастической составляющих, причем детерминированная составляющая движения соответствует фоновому движению на полосе, а стохастическая составляющая — индивидуальным перемещениям частиц. Каждая полоса состоит из последовательности клеток, размер которых определяется *динамическим габаритом, то есть длиной участка дороги, на котором размещается одна частица с учетом расстояния между соседними, определяемым требованиями безопасности и зависящим от величины скорости движения*. Стохастическое движение описывается монотонным случайным блужданием частиц по клеткам полосы, а регулярное движение описывается равномерным перемещением всех клеток полосы [1, 2].

В [3] рассматривалась модель движения частиц по неподвижной полосе и получена формула для средней скорости частиц. Эта формула позволяет вычислить стохастическую составляющую скорости движения в модели с движущимися полосами в предположении, что частицы не переходят с одной полосы на другую. Предполагается, что зависимость динамического габарита от скорости регулярного движения известна [4].

В случае, когда вероятность перемещения частицы в свободном состоянии стремится к единице, результаты [3] соответствуют результатам работы [5], где рассматривалась сходная модель, в которой случайность проявлялась лишь в начальном размещении частиц на клеточном поле. Исследованию модели данного класса посвящена также работа [6].

В [7] рассматривалась двухполосная стохастическая модель, в которой разрешались переходы частиц на соседнюю полосу. Разработан способ

приближенного вычисления средней скорости движения частиц. Точность приближения этого способа проверялась имитационным моделированием.

В [8] результаты работы [7] обобщены на случай, когда частицы различаются по типу.

В одной из пионерских работ, посвященных клеточным моделям потоков [9], получены содержательные результаты методом имитационного моделирования.

При изучении одноканального движения частиц широко применялись также различные модели следования за лидером, представляющие собой дифференциальные уравнения с запаздывающим аргументом [10], приводящие к гладким решениям.

В [11, 12] модели случайного блуждания по дискретной решетке, аналогичные введенной в [1], были использованы при решении некоторых проблем оптимизации дорожного движения.

В настоящей работе рассматриваются различные постановки задачи максимизации средней скорости и интенсивности по скоростям регулярного движения, определяющим динамический габарит. Исследуются случаи, когда переходы частиц на соседние полосы запрещены (*канализированное движение*) и когда такие переходы разрешены. При канализированном движении оптимизация осуществляется также по распределению частиц по полосам. Задача для канализированного движения решается с помощью формулы, полученной в [3]. Для решения задачи оптимизации в случае, когда переходы на соседнюю полосу разрешены, частично используется метод имитационного моделирования.

Дадим описание рассматриваемой в настоящей работе модели. Предположим, что имеются k последовательностей ячеек (k полос) (рис. 1). Движение на каждой полосе представляет собой сумму двух составляющих [1, 2]. Первой из этих составляющих является регулярное движение, которое определяет разбиение полосы на клетки. Бу-

дем обозначать через $v = v_{det,i}$ скорость регулярного движения на i -й полосе, $i = 1, \dots, k$. Зависимость размера динамического габарита $d(v)$ (м) от $v = v_{det}$ (м/с) будем описывать, например, формулой [4]:

$$d(v) = 5,7 + 0,504v + 0,0285v^2. \quad (1)$$

Как отмечается в [4], зависимость (1) получена аппроксимацией связи между скоростью и средним интервалом между автомобилями посредством параболы, при этом коэффициенты определены методом наименьших квадратов и имеют прозрачный физический смысл.

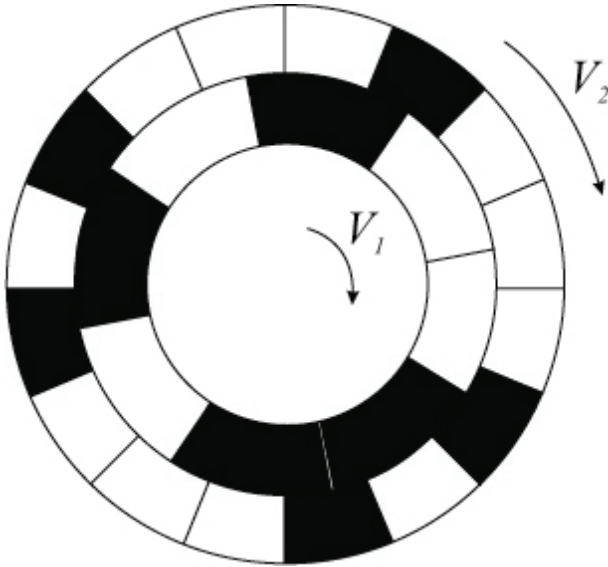


Рис. 1. Модель двухполосного движения

Вторая составляющая скорости движения частицы обусловлена индивидуальными перемещениями частиц в соседние ячейки по ходу движения. За единицу времени с некоторой вероятностью p частица имеет намерение переместиться вперед и осуществляет это намерение, если позволяет конфигурация.

Будем рассматривать следующие задачи.

Задача 1. Оптимизация распределения частиц при канализованном движении. Требуется при фиксированной плотности ρ и регулярной скорости v максимизировать интенсивность I потока частиц, то есть среднее число частиц, пересекающих сечение дороги в единицу времени:

$$I(v, r_1, \dots, r_k) \rightarrow_{\vec{r}} \max, \\ \frac{r_1 + \dots + r_k}{d(v)} = \frac{m}{nd(v)} = \rho,$$

где r_i ($i = 1, 2, \dots, k$) — отношение числа m_i частиц на i -й полосе к числу ячеек n на этой полосе, $i = 1, 2, \dots, k$; m — число всех частиц, $m_1 + m_2 + \dots + m_k = m$; ρ — плотность потока, то есть отношение числа m частиц к длине пути $L = nd(v)$.

Задача 2. В условиях задачи 1 полагаем, что регулярная скорость движения может зависеть от

полосы: $\vec{v} = (v_1, \dots, v_k)$:

$$I(\vec{v}, \vec{r}, p) \rightarrow_{\vec{v}, \vec{r}} \max,$$

$$\frac{r_1}{d(v_1)} + \frac{r_2}{d(v_2)} + \dots + \frac{r_k}{d(v_k)} = \rho.$$

Задача 3. В условиях задачи 1 разрешаются переходы частиц с одной полосы на другую. Требуется исследовать зависимость средней скорости частицы от плотности потока.

В случае, когда полоса единственна, задача 3 и эквивалентная задача 1 сводятся к задаче, рассматривавшейся в [3]. Как и для случая одной полосы, рассматриваемая *общая задача 3* может быть описана цепью Маркова. Однако решение соответствующей системы уравнений не выражается в виде удобной формулы, как это имеет место для случая одной полосы [3]. Зависимость интенсивности потока от плотности может быть представлена приближенно.

Назовем *средней стохастической скоростью частиц на j -й полосе* $u(r_j)$ отношение числа перемещений в единицу времени частиц, находящихся на полосе, к числу всех частиц на этой полосе.

В соответствии с формулой, полученной в [3], где рассматривается задача вычисления средней скорости стохастического движения по неподвижной полосе, имеем ($m = rn$):

$$u(r) = u\left(\frac{m}{n}\right) = u(r, p) = \frac{1}{m} Q = \\ = \frac{n}{m} \sum_{j=1}^{\min(m, n-m)} \frac{C p}{(1-p)^{j-1}} C_{m-1}^{j-1} C_{n-m-1}^{j-1}, \quad (2)$$

где C определено следующим соотношением:

$$C = \left(\sum_{j=1}^{\min(m, n-m)} \frac{n}{j} \cdot C_{m-1}^{j-1} C_{n-m-1}^{j-1} \frac{1}{(1-p)^{j-1}} \right)^{-1}.$$

График функции $u(r)$, $0 < r < 1$, при $n = 20$, $m = 10$, $p = 0,5$ изображен на рис. 2. Эта функция не является выпуклой по r , так как вторая разность $\Delta^2 u(r) = u(r+2h) - 2u(r+h) + u(r)$, $h = n^{-1}$ не всегда отрицательна. На рис. 3, 4 изображены зависимости соответственно первой $\Delta u(r) = u(r+h) - u(r)$ и второй разности $\Delta^2 u(r)$ от r при $n = 20$, $m = 10$, $p = 0,5$.

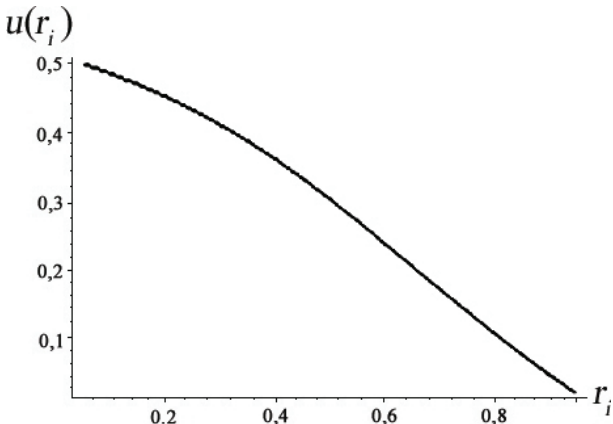


Рис. 2. Зависимость $u(r_i)$ от r_i при $n = 20$, $m = 10$, $p = 0,5$.

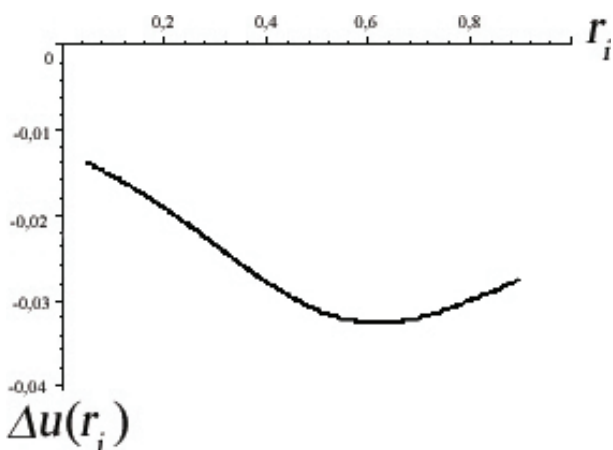


Рис. 3. Зависимость $\Delta u(r_i)$ от r_i при $n = 20$, $m = 10$, $p = 0,5$.

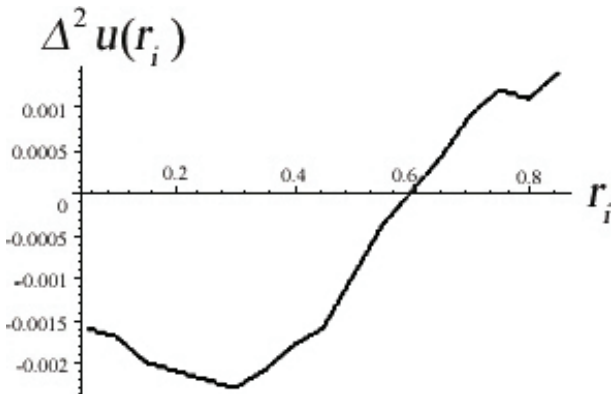


Рис. 4. Зависимость $\Delta^2 u(r_i)$ от r_i при $n = 20$, $m = 10$, $p = 0,5$.

II. Оптимальное распределение потоков при одинаковых скоростях регулярного движения

II.1. Двухполосное канализированное движение

Будем рассматривать двухполосное канализированное движение с одинаковыми детерминированными скоростями движения, равными v для

каждой полосы (рис. 1). Каждая полоса представляет собой замкнутую последовательность ячеек, причем число ячеек на каждой полосе равно n .

На j -й полосе находится m_j частиц, $j = 1, 2$. Обозначим через (i, j) i -ю ячейку j -й полосы, $i = 1, 2, \dots, n$, $j = 1, 2$. Ячейка $(i + 1, j)$ следует за ячейкой (i, j) , $i = 1, 2, \dots, n - 1$, $j = 1, 2$. За ячейкой (n, j) следует ячейка $(1, j)$, $j = 1, 2$. Перемещения частиц могут происходить в целочисленные моменты времени $1, 2, \dots$. Если в некоторый целочисленный момент времени t в ячейке (i, j) находится частица, а следующая ячейка j -й полосы свободна, то с вероятностью $0 < p < 1$ в промежутке времени $(t, t + 1)$ независимо от других осуществляется переход частицы в следующую ячейку.

Пусть $r = m/2n$, $r_j = m_j/n$, $j = 1, 2$,

$$r_1 + r_2 = 2r,$$

$$u(r_1, r_2) = (r_1 u(r_1) + r_2 u(r_2)) / (2r), \quad (3)$$

где $u(r_1, r_2)$ — средняя стохастическая скорость частиц по двум полосам. Тогда *средняя физическая скорость* частиц определяется как

$$v(r_1, r_2) = v + u(r_1, r_2) d(v).$$

Если $\rho = m/L$ — плотность потока частиц, отношение числа частиц к длине L пути и соответственно по полосам $\rho_i = m_i/L$, $i = 1, 2$, то интенсивность $I(r_1, r_2)$ потока определяется формулой

$$I(r_1, r_2) = \rho v(r_1, r_2).$$

Интенсивность потока представляет собой *среднее число частиц, пересекающих в единицу времени сечение пути, неподвижное в системе отсчета, относительно которой каждая из полос движется со скоростью v* .

Для простоты считаем, что отношение L к $d(v)$ — целое число. Тогда число ячеек на каждой полосе определяется по формуле

$$n = L/d(v).$$

Будем решать задачу максимизации интенсивности потока путем распределения m частиц по полосам (m задано), то есть выбора возможных значений r_1 и r_2 , максимизирующих $I(r_1, r_2)$. Эта задача редуцируется к максимизации $v(r_1, r_2)$, которая в свою очередь сводится к максимизации $u(r_1, r_2)$ при $0 \leq r_1, r_2 \leq 1$, $r_1 + r_2 = 2r$, $r = m/2n$.

На рис. 5 представлена зависимость функции $u(r_1, 2r - r_1)$ от r_1 при $n = 20$, $m = 10$ и p , равным соответственно 0,1, 0,5 и 0,9.

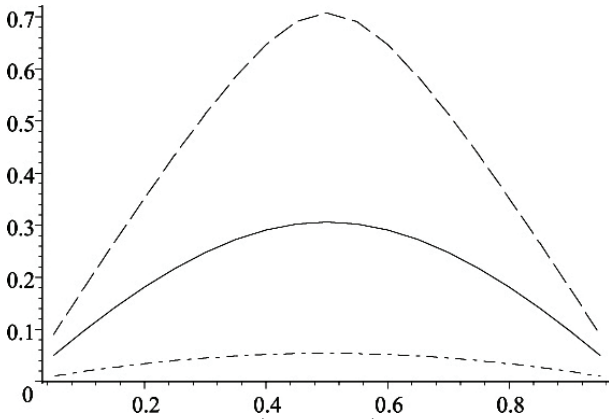


Рис. 5. Зависимость $u(r_1, 2r - r_1)$ при $n = 20$, $m = 20$, $p = 0,1$, $p = 0,5$ (сплошная линия), $p = 0,9$ (пунктирная линия)

Лемма 1. Для любого фиксированного $r = n^{-1}, 2n^{-1}, \dots, 1 - n^{-1}$ последовательность $u(r_1, 2r - r_1)$ монотонна по переменной r_1 на множестве $r_1 = n^{-1}, 2n^{-1}, \dots, r$.

Доказательство. Функция

$$U(x) = u(x, 2r - x) = xu(x) + (2r - x)u(2r - x),$$

$0 < x < r$, имеет производную $(u(x) - u(2r - x)) + (xu'(x) - (2r - x)u'(2r - x))$.

Первое слагаемое очевидно положительно при $x < r$, а второе — из-за того, что $|u'(x)| \leq |u'(2r - x)|$, то есть $|u(x + h) - u(x)| \leq |u(2r - x) - u(2r - x - h)|$, при $x < r$ ($h = n^{-1}$).

Учитывая лемму 1 и симметрию функции $u(r_1, r_2)$, можно сформулировать следующую теорему.

Теорема 1. Решение задачи 1 при $k = 2$ достигается на равном распределении потока по полосам.

Лемма 2. Если функция $f(x)$ на отрезке $[0, 1]$ монотонно убывает и имеет отрицательную вторую производную, то функция $F(x) = xf(x) + (1 - x)f(1 - x)$ на этом отрезке также имеет отрицательную вторую производную, то есть выпукла.

Доказательство. Имеем

$$F''(x) = 2f'(x) + xf''(x) + 2f'(1 - x) + (1 - x)f''(1 - x), x \in [0, 1].$$

Так как $f'(x) < 0$, $f'(1 - x) < 0$, $f''(x) < 0$, $f''(1 - x) < 0$, то и $F''(x) < 0$, $x \in [0, 1]$, что и требовалось доказать.

Лемма 3. Пусть функция $f(x_1, \dots, x_k)$ не зависит от порядка аргументов, а функция $F(x_1, \dots, x_{k-1}) = f(x_1, \dots, x_{k-1}, 1 - x_1 - \dots - x_{k-1})$, определенная на множестве $x_i \geq 0$ ($i = 1, \dots, k - 1$), $x_1 + \dots + x_{k-1} \leq 1$, строго выпукла по каждой переменной при фиксированных значениях остальных переменных, то эта функция достигает наибольшего значения при $x_1 = \dots = x_{k-1} = 1 - x_1 - \dots - x_{k-1} = x_k$.

Доказательство. Функция $F(x_1, \dots, x_{k-1})$ достигает в некоторой точке своего наибольшего значения. Пусть $x_{10}, \dots, x_{k-1,0}$ — набор значений аргументов функции, для которого $x_{i0} \neq 1 - x_{10} - \dots - x_{i0} - \dots - x_{k-1,0}$ при некотором $1 \leq i \leq k - 1$. Обозначим $1 - x_{10} - \dots - x_{i-1,0} - x_{i+1,0} - \dots - x_{k-1,0}$ через a .

В силу выпуклости функции $F(x_1, \dots, x_{k-1})$ по переменной x_i

$$F(x_{10}, \dots, x_{i-1,0}, x_{i0}, x_{i+1,0}, \dots, x_{k-1,0}) + F(x_{10}, \dots, x_{i-1,0}, 1 - x_{10} - \dots - x_{k-1,0}, x_{i+1,0}, \dots, x_{k-1,0}) < 2F(x_{10}, \dots, x_{i-1,0}, \frac{a}{2}, x_{i+1,0}, \dots, x_{k-1,0}).$$

Отсюда, учитывая, что в силу симметрии функции $f(x_1, \dots, x_k)$ относительно ее аргументов

$$F(x_{10}, \dots, x_{i-1,0}, x_{i0}, x_{i+1,0}, \dots, x_{k-1,0}) = F(x_{10}, \dots, x_{i-1,0}, 1 - x_{10} - \dots - x_{k-1,0}, x_{i+1,0}, \dots, x_{k-1,0}),$$

получаем, что на рассматриваемом наборе значений аргументов функция F не может достигать наибольшего значения, откуда следует доказываемое утверждение.

Лемма 4. Пусть функция $g(x)$ на отрезке $[0, 1]$ монотонно убывает и имеет отрицательную вторую производную, тогда функция

$$H(x_1, \dots, x_{k-1}) = x_1g(x_1) + \dots + x_{k-1}g(x_{k-1}) + (1 - x_1 - \dots - x_{k-1})g(1 - x_1 - \dots - x_{k-1})$$

удовлетворяет условиям леммы 2 для функции $F(x_1, \dots, x_{k-1})$.

Доказательство.

Имеем

$$H(x_1, \dots, x_{k-1}) = h(x_1, \dots, 1 - x_1 - \dots - x_{k-1}),$$

где

$$h(x_1, \dots, x_k) = x_1g(x_1) + \dots + x_kg(x_k).$$

Значение функции $h(x_1, \dots, x_k)$ не зависит от порядка аргументов.

Покажем, что функция $H(x_1, \dots, x_{k-1})$ имеет отрицательную вторую частную производную по переменной x_i для любого $i = 1, \dots, k - 1$. Действительно,

$$H''_{x_i} = 2g'_{x_i}(x_i) + x_i g''_{x_i x_i}(x_i) + 2g'_{x_i}(1 - x_1 - \dots - x_{k-1}) + (1 - x_1 - \dots - x_{k-1}) g''_{x_i x_i}(1 - x_1 - \dots - x_{k-1}).$$

Так как $g'_{x_i}(x_i) < 0$, $g'_{x_i}(1 - x_1 - \dots - x_{k-1}) < 0$, $g''_{x_i x_i}(x_i) < 0$, $g''_{x_i x_i}(1 - x_1 - \dots - x_{k-1}) < 0$, $x \in [0, 1]$. Таким образом, утверждение леммы 4 верно.

При произвольном k из леммы 3 получаем следующее утверждение.

Теорема 2. Если на интервале $n^{-1}, \dots, kr - n^{-1}$, $r = \frac{m}{kn}$ функция $u(r_1)$ выпукла, то решение задачи 1 при произвольном k достигается на равном распределении потока по полосам.

II.2. Исследование случая $p \sim 0$

Рассмотрим случай, когда p мало, а n велико. В соответствии с (3) имеем

$$u_j(r_j) = p \cdot \frac{n - m_j}{n - 1} + o(p), p \sim 0. \quad (4)$$

Пренебрегая в (3) слагаемым порядка малости выше p , перепишем (2) в виде

$$u_j(r_j) = p \left(1 - \frac{m_j - 1}{n - 1} \right). \quad (5)$$

Формулу (5) можно записать в виде

$$u_j(r_j) = p(1 - r_j) + O\left(\frac{1}{n}\right), \frac{1}{n} \sim 0. \quad (6)$$

Пренебрегая бесконечно малым слагаемым, перепишем (6) в виде

$$u_j(r_j) = p(1 - r_j). \quad (7)$$

Из (2) и (7), учитывая, что $r_2 = 2r - r_1$, сводим задачу к максимизации:

$$u(r_1, 2r - r_1) = p(r_1(1 - r_1) + (2r - r_1)(1 - (2r - r_1)))/(2r) \quad (8)$$

при ограничении

$$\max(0, 2r - 1) < r_1 < \min(2r, 1). \quad (9)$$

Нетрудно установить, что максимум (8), (9) достигается при $r_1 = r_2$.

II.3. Исследование случая $p \sim 1$

Перейдем к рассмотрению случая, когда значение p близко к 1. При $p \sim 1$ имеем [3]:

$$u_j(r_j) = \begin{cases} 1, & r_j \leq \frac{1}{2}, \\ \frac{1}{r_j} - 1 + O(1 - p), & r_j > \frac{1}{2}. \end{cases} \quad (10)$$

Пренебрегая в (10) бесконечно малым слагаемым, получим ($j = 1, 2$):

$$u(r_1, r_2) = (r_1 \min(1, \frac{1}{r_1} - 1) + r_2 \min(1, \frac{1}{r_2} - 1))/(2r). \quad (11)$$

Максимум (11) с ограничениями (9) достигается при $r_1 = r_2 = r$.

На рис. 6 приводятся зависимости $u(r_1, 2r - r_1)$ от r_1 при $p \sim 1$, $r = 0,5$, $0 \leq r_1 \leq 1$, $r = 0,4$, $0 \leq r_1 \leq 0,8$, $r = 0,6$, $0,2 \leq r_1 \leq 1$. При расчете полагаем $m = 10$. Величина r_1 принимает значения в интервале $\max(0, 2r - 1) < r_1 < \min(2r, 1)$.

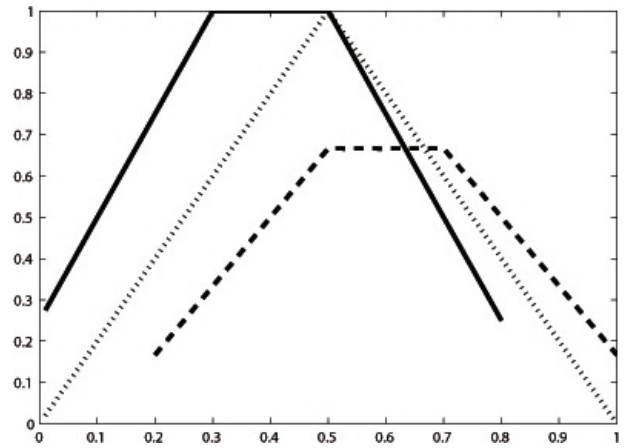


Рис. 6. Зависимость $u(r_1, 2r - r_1)$ от r_1 при $p \sim 1$, $r = 0,5$ — пунктирная линия, $r = 0,4$ — жирная линия, $r = 0,6$ — пунктирная и жирная

Как видно из приведенных результатов, зависимость $u(r_1, r_2)$ от r_1 выпукла по r_1 , симметрична и достигает максимума при $r_1 = r_2$.

III. Оптимизация детерминированной и стохастической компонент при канализированном движении

В силу вышесказанного в условиях задачи 1 при фиксированном значении v для обеих полос получаем равномерное распределение частиц по полосам. Следовательно, задачу оптимизации потоковой скорости v можно рассматривать для одной полосы. Считаем, что на полосе имеется n ячеек и на ней находится m частиц. Естественно предположить, что значение параметра p , определяющего индивидуальное поведение в потоке, зависит от v , то есть определяется заданной функцией $p = p(v)$, $0 \leq p \leq 1$. Рассмотрим следующую модель поведения:

$$v + p(v)d(v) = v_0, \quad (12)$$

которая означает, что существует равновесная средняя скорость для частиц рассматриваемого потока, то есть ожидаемая скорость движения в случае незначительного уровня помех.

Реальная средняя скорость $v(m, n)$ движения частицы по полосе представляет собой сумму детерминированной составляющей v и стохастической составляющей $u(m, n)$:

$$v(m, n) = v + d(v)u(m, n), \quad (13)$$

$$u(m, n) = \frac{n}{m} \sum_{k=1}^{\min(m, n-m)} \frac{Cp(v)}{(1 - p(v))^{k-1}} C_{m-1}^{k-1} C_{n-m-1}^{k-1},$$

$$C = \left(\sum_{k=1}^{\min(m, n-m)} \frac{n}{k} \cdot C_{m-1}^{k-1} C_{n-m-1}^{k-1} \frac{1}{(1 - p(v))^{k-1}} \right)^{-1},$$

$$m \leq n(v) = L/d(v).$$

Число ячеек n определяется величиной динамического габарита, который в свою очередь зависит от детерминированной составляющей скорости. При росте значения v возрастает динамический габарит, а следовательно, уменьшается значение n , что в свою очередь приводит к увеличению $r = m/n$ и уменьшению значения $u(m,n)$. При достижении n значения m величина стохастической составляющей скорости уменьшается до нуля. Дальнейшее увеличение детерминированной составляющей недопустимо, так как при больших ее значениях расстояние между частицами не может удовлетворять требованиям динамического габарита.

Минимальное значение $v = v_{\min}$, при котором условие (12) может быть выполнено, определяется условием

$$v_{\min} + d(v_{\min}) = v_0. \quad (14)$$

Считаем, что выполняется условие

$$m < L/d(v_{\min}).$$

При невыполнении этого условия требования на динамический габарит не удовлетворялись бы даже при минимальном допустимом значении v .

Максимальное допустимое значение v^* , при котором еще могут выполняться требования динами-

ческого габарита, определяется равенством

$$d(v^*) = \frac{L}{m}. \quad (15)$$

Так как из (12) следует, что должно выполняться неравенство $v \leq v_0$, то, учитывая (15), получаем для максимального допустимого значения $v = v_{\max}$:

$$v_{\max} = \min(d^{-1}(L/m), v_0) = \min(v^*, v_0). \quad (16)$$

При фиксированных значениях m , L и v_0 задача максимизации интенсивности эквивалентна в рассматриваемом случае максимизации средней скорости одноканального потока (13) на интервале допустимых значений v :

$$v_{\min} \leq v \leq v_{\max}.$$

Пример 1. Пусть $L = 100$ м, $m = 5$, $v_0 = 10$ (м/с). Для этого случая из (14), (16) найдем $v_{\min} = 2,6$ м/с, $v_{\max} = 10$ м/с.

В табл. 1 и на рис. 7 приводятся результаты расчета значения средней скорости движения при допустимых значениях детерминированной составляющей.

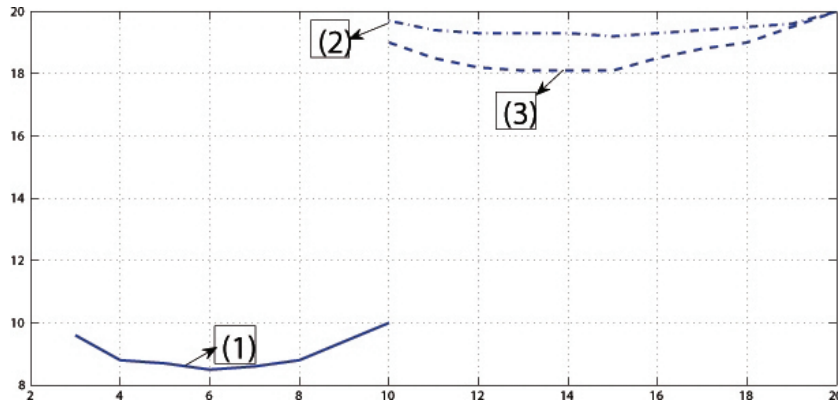


Рис. 7. Зависимость $v(m,n)$ от v_0 при $L = 100$ (м), $m = 5$, $v_0 = 10$ (м/с) — пунктирная линия, $L = 1000$ (м), $m = 10$, $v_0 = 20$ (м/с) — жирная линия, $L = 1000$ (м), $m = 20$, $v_0 = 20$ (м/с) — пунктирная и жирная

Т а б л и ц а 1

Зависимость средней скорости $v(m,n)$ (м) от величины детерминированной составляющей v (м) при $L = 100$ (м), $m = 5$, $v_0 = 10$ (м/с)

v (м/с)	3	4	5	6	7	8	9	10
$d(v)$ (м)	7.5	8.2	8.9	9.8	10.6	11.8	12.5	13.6
n	13	12	11	10	9	8	7	7
r	0.38	0.42	0.45	0.50	0.56	0.62	0.71	0.71
$p(v)$	0.94	0.73	0.56	0.41	0.28	0.17	0.07	0
$u(v)$	0.89	0.58	0.42	0.25	0.15	0.07	0.03	0
$u(v)d(v)$ (м/с)	6.6	4.8	3.7	2.5	1.6	0.8	0.4	0
$v(m,n)$ (м/с)	9.6	8.8	8.7	8.5	8.6	8.8	9.4	10.0

Т а б л и ц а 2

Зависимость средней скорости $v(m,n)$ (м) от величины детерминированной составляющей v (м) при $L = 1000$ (м), $m = 10$, $v_0 = 20$ (м/с)

v	10	11	12	13	14	15	16	17	18	19	20
$d(v)$ (м)	13.6	14.7	15.8	17	18.3	19.6	21	22.5	24	25.6	27.1
n	73	68	63	58	54	50	47	44	41	39	36
r	0.14	0.15	0.16	0.17	0.19	0.2	0.21	0.23	0.24	0.26	0.28
$p(v)$	0.74	0.61	0.5	0.41	0.33	0.25	0.19	0.13	0.08	0.04	0
$u(v)$ (м)	0.71	0.57	0.46	0.37	0.29	0.21	0.16	0.1	0.06	0.03	0
$u(v)d(v)$ (м/с)	9.66	8.40	7.26	6.25	5.28	4.24	3.32	2.36	1.51	0.59	0
$v(m,n)$ (м/с)	19.7	19.4	19.3	19.3	19.3	19.2	19.3	19.4	19.5	19.6	20

Т а б л и ц а 3

Зависимость средней скорости $v(m,n)$ (м) от величины детерминированной составляющей v (м) при $L = 1000$ (м), $m = 20$, $v_0 = 20$ (м/с)

v (м/с)	10	11	12	13	14	15	16	17	18	19	20
$d(v)$ (м)	13.6	14.7	15.8	17	18.3	19.6	21	22.5	24	25.6	27.1
n	73	68	63	58	54	50	47	44	41	39	36
r	0.27	0.29	0.32	0.34	0.37	0.4	0.43	0.45	0.49	0.51	0.56
$p(v)$	0.74	2.11	2.02	1.93	1.85	1.78	1.71	1.64	1.58	1.53	1.47
$u(v)$	0.66	0.51	0.39	0.30	0.23	0.16	0.12	0.08	0.04	0.02	0
$u(v)d(v)$ (м/с)	9.0	7.5	6.2	5.1	4.1	3.1	2.5	1.8	1.0	0.5	0
$v(m,n)$ (м/с)	19.0	18.5	18.2	18.1	18.1	18.1	18.5	18.8	19.0	19.5	20

Значение $v(m,n)$ достигает максимума, равного v_0 , при $v = v_{\min}$ ($p(v_{\min}) = 1$) и при $v = v_0$, ($p(v_{\min}) = 0$). При $v_{\min} < v < v_{\max}$ имеем $u(v_{\min}) < p(v_{\min})$ и $v(m,n) < v_0$. Как видно из результатов расчетов, при росте v значение $v(m,n)$ сначала убывает, а затем возрастает.

Пример 2. Пусть $L = 1000$, $m = 10$, $v_0 = 20$ (м/с). Для данного случая $v_{\min} = 8,2$ м/с, $v_{\max} = 20$ м/с.

Результаты расчетов приводятся в табл. 2 и на рис. 7.

Как и в предыдущем примере, значение $v(m,n)$ сначала убывает при возрастании v , а затем возрастает.

Пример 3. Пусть $L = 1000$ м, $m = 20$, $v_0 = 20$ м/с. Для данного случая $v_{\min} = 8,2$ м/с, $v_{\max} = 20$ м/с.

Результаты расчета значения средней скорости движения при допустимых значениях детерминированной составляющей приводятся в табл. 3 и на рис. 7. Значения $d(v)$, n , $p(v)$ и v_{\min} не зависят от m и при тех же допустимых значениях v остаются теми же, что и в примере 2. Так как в данном примере $v_{\max} = v_0$, то значение v_{\max} также остается таким же, как и в примере 2.

Теорема 3. При выполнении условий

$$r(v_{\min}) = \frac{md(v_{\min})}{2L} \leq \frac{1}{2} \tag{17}$$

и

$$d^{-1}\left(\frac{L}{m}\right) \geq v_0 \tag{18}$$

наибольшее возможное значение $v(m,n)$, равное v_0 , достигается как при $v = v_{\min}$, так и при $v = v_{\max}$.

Если выполняется (17) и не выполняется (18), то наибольшее значение $v(m,n)$ достигается только при $v = v_{\min}$.

Если же выполняется (18) и не выполняется (17), то наибольшее значение $v = v(m,n)$ достигается только при $v = v_{\max}$.

Доказательство. Рассмотрим случай, когда $v = v_{\min}$. Тогда $p(v_{\min}) = 1$. Поведение функции $u(m,n)$ при значениях p , близких к 1, рассматривалось в п. 1.3. Учитывая формулу (10), при выполнении (17) имеем

$$v(m,n) = v + d(m,n) = (v_0 - d(m,n)) + d(m,n) = v_0,$$

а если (17) не выполняется, то получаем

$$v(m,n) = v_0 - d(v_{\min}) + d(v_{\min}) \left(\frac{1}{r(v_{\min})} - 1 \right) < v_0.$$

Предположим, что выполняется (18). Тогда $v_{\max} = v_0$, $p(v_{\max}) = 0$, $v(m,n) = v_0$. Если же (18) не выполняется, то $0 < p(v_{\max}) < 1$, $u(m,n) < p(v_{\max})$ и при $v = v_{\max}$

$$v(m,n) = v_{\max} + u(m,n)d(v_{\max}) < < v_{\max} + p(v_{\max})d(v_{\max}) = v_0.$$

Аналогично, при $v_{\min} < v < v_{\max}$ имеем $0 < p(v) < 1$, $u(m,n) < p(v)$,

$$v(m,n) = v + u(m,n)d(v) < v + p(v)d(v) = v_0.$$

Из приведенных соотношений для значений функции $v(m,n)$ на концах и во внутренних точках допустимых значений v следует утверждение теоремы 3.

IV. Оптимизация канализированных потоков по распределению и скоростям

Рассмотрим задачу 2. Пусть m — число частиц, ρ — плотность потока, m_1 — число частиц на первой полосе. Тогда $(m_2 = m - m_1)$ — число частиц на второй полосе. Пусть L — длина каждой полосы. Число ячеек $n_i = n_i(v_i)$ на i -й полосе определяется формулой

$$n_i = n_i(v_i) = \frac{L}{d(v_i)}. \quad (19)$$

Будем использовать следующие обозначения: v_i — скорость регулярного движения по i -й полосе; $v(m_i, n_i, v_i)$ — средняя скорость движения частицы на i -й полосе в предположении, что на ней находится m_i частиц, $i = 1, 2$; $v(\vec{m}, \vec{n}, \vec{v}) = v(m_1, m_2, n_1, n_2, v_1, v_2)$ — средняя скорость частицы:

$$v(m_i, n_i, v_i) = v_i + d(v_i)u(m_i, n_i), \quad (20)$$

$$v(\vec{m}, \vec{n}, \vec{v}) = (m_1 v(m_1, n_1, v_1) + m_2 v(m_2, n_2, v_2)) / m, \quad (21)$$

где $u(m_i, n_i)$ вычисляется по формуле (3), $p(v)$ определяется из (12), а число $n = n_i$ ячеек на i -й полосе вычисляется по формуле $n_i = L/d(v_i)$, $i = 1, 2$.

Интенсивность потока

$$\begin{aligned} I &= I_1 + I_2 = \rho_1 v(m_1, n_1, v_1) + \rho_2 v(m_2, n_2, v_2) = \\ &= \rho \left(\frac{\rho_1}{\rho_1 + \rho_2} v(m_1, n_1, v_1) + \frac{\rho_2}{\rho_1 + \rho_2} v(m_2, n_2, v_2) \right) = \\ &= \rho v(\vec{m}, \vec{n}, \vec{v}), \end{aligned} \quad (22)$$

где ρ определяется из соотношения

$$\rho = \frac{m}{L}, \quad (23)$$

по-прежнему представляет собой среднее число частиц, пересекающих неподвижное сечение дороги. Относительно этого сечения детерминированная составляющая скорости первой полосы равна v_1 , а второй — v_2 .

Задача максимизации (22) по m_1 при фиксированных m , v_1 и v_2 и задача максимизации (22) по m_1 , v_1 и v_2 при фиксированных m сводятся к соответствующим задачам максимизации (21), то есть максимизации средней скорости частицы.

Задачи максимизации (21) по m_1 решаются с учетом того, что на полосе не может находиться частиц больше, чем необходимо при соблюдении динамического габарита:

$$m_i \leq n_i(v_i), \quad i = 1, 2. \quad (24)$$

На рис. 8 приводятся зависимости $v(m_1, m - m_1)$ от m_1 при фиксированном значении

m . Наибольшее значение $v(m_1, m - m_1)$ достигается при $m_1 = 6$, $m_2 = 4$. Если значение m_1 увеличить до 7, то условия на динамический габарит для первой полосы не будут выполняться.

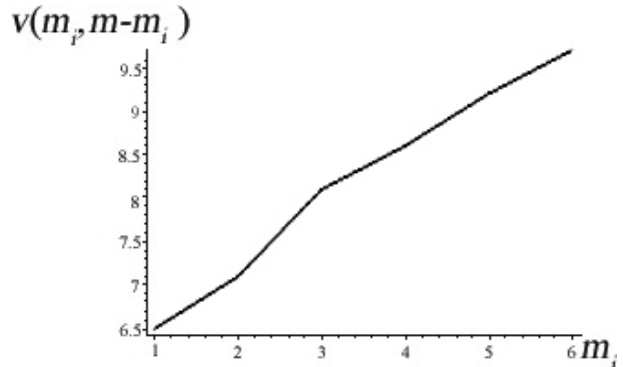


Рис. 8. Зависимость средней скорости $v(m_1, m - m_1)$ от m_1 ; $m = 10$, $p = 0,5$, $v_1 = 10$ (м/с), $v_2 = 5$ (м/сек), $L = 100$ м

Формулы (1), (3), (20), (21) позволяют также решать задачу в предположении, что скорости v_1, v_2 регулярного движения не фиксированы и осуществляется оптимизация по возможным значениям этих скоростей (задача 2).

V. Многополосное движение частиц, при котором разрешаются переходы на другую полосу

Опишем модель, в которой осуществляется движение частиц по двум полосам по следующим правилам.

Имеются две замкнутые последовательности (полосы) ячеек. В каждой последовательности число ячеек равно n . Каждой ячейке ставится в соответствие пара чисел (i, j) , где j — номер полосы ($j = 1, 2$), а i — порядковый номер ячейки на полосе, $i = 1, \dots, n$. Перемещение частицы осуществляется по следующим правилам.

Если ячейка, следующая за той, в которой находится частица, свободна, то в текущий момент времени с вероятностью p осуществляется перемещение частицы на ячейку вперед.

Если ячейка, следующая за ячейкой, в которой находится частица, занята и занята также клетка, находящаяся позади на соседней полосе, то рассматриваемая частица остается на месте (рис. 9).

Если ячейка, следующая за ячейкой, в которой находится частица, занята, а три прилегающие клетки на соседней полосе свободны, то осуществляется переход рассматриваемой частицы на соседнюю полосу с продвижением вперед (рис. 10).

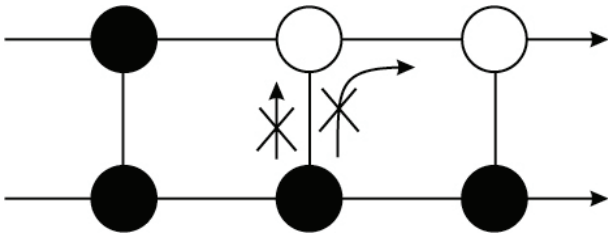


Рис. 9. Переход на другую полосу запрещен

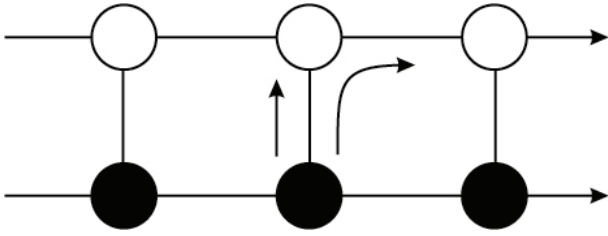


Рис. 10. Переход на другую полосу с продвижением вперед

Если ячейка, следующая за ячейкой, в которой находится частица, и ячейка на другой полосе впереди заняты, а две ячейки позади нее свободны, то осуществляется переход рассматриваемой частицы на соседнюю полосу без продвижения вперед (рис. 11).

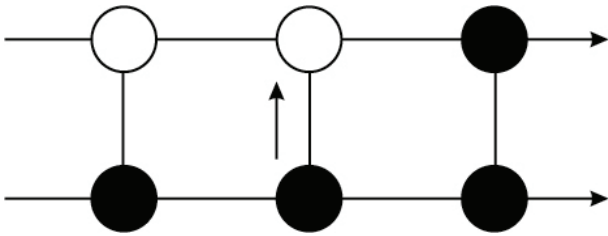


Рис. 11. Переход на другую полосу без продвижения вперед

Возникает задача исследования этой модели, в частности, задача вычисления средней скорости частицы для данной модели.

В [5] рассматривалась аналогичная модель с той разницей, что переход частиц на соседнюю полосу без перемещения вперед не происходит, а переход на другую полосу при наличии двух свободных ячеек на соседней полосе осуществляется независимо от того, занята или свободна ячейка, находящаяся на соседней полосе позади, и разработан приближенный способ вычисления характеристик рассматриваемой системы. Аналогичный способ может быть применен и для исследования характеристик модели, описанной в настоящем параграфе.

Поставленную задачу можно интерпретировать как задачу исследования случайного блуждания частиц по многограннику (рис. 12, 13).

Аналогичным образом модель может быть обобщена на случай многополосного движения с числом полос больше двух.

В табл. 4 приводятся результаты имитационного моделирования для движения частиц по двум полосам при $m = 10$, $p = 0,5$. Интервал моделирования полагается равным 10 000 тактам. Полученные значения сравниваются с соответствующими

значениями средней скорости частиц, вычисленными по формуле (3) для движения частиц по полосе при том же значении отношения числа частиц к числу ячеек, то есть со значениями средней скорости частиц для двухполосного движения с одинаковыми полосами, при котором переходов частиц на другую полосу не происходит.

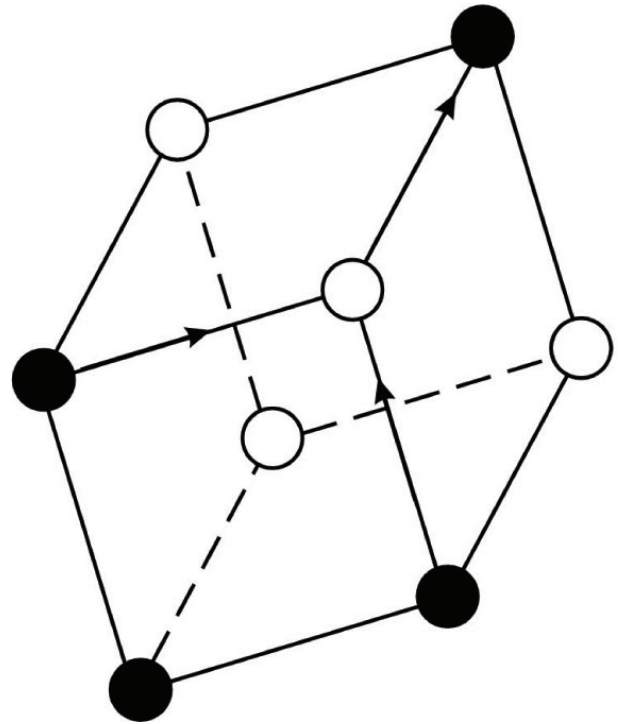


Рис. 12. Двухканальный четырехугольник

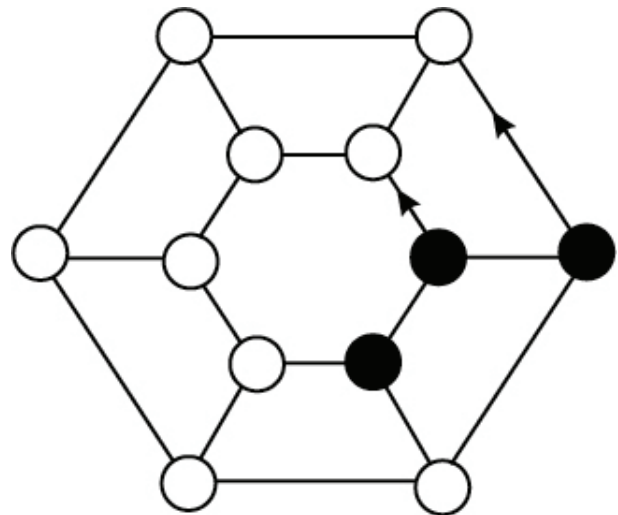


Рис. 13. Двухканальный шестиугольник

Значение $u^*(r)$ получено с помощью имитационного моделирования (разрешены переходы на соседнюю полосу). Значение $u^{**}(r)$ вычислено по формуле (3) для канализованного движения; Δ — относительная величина (в процентах) превышения $u^{**}(r)$ над $u^*(r)$:

$$\Delta(r) = \left| \frac{u^{**}(r) - u^*(r)}{u^*(r)} \right|.$$

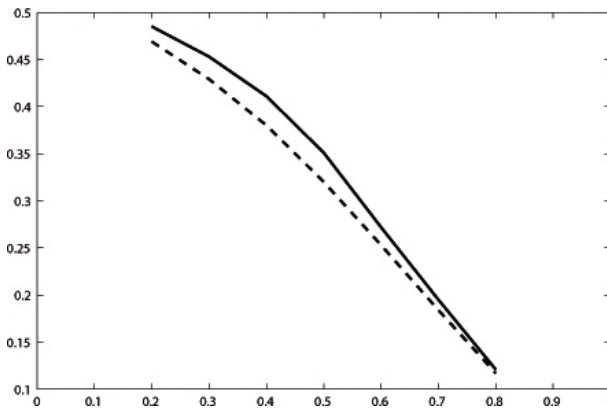
Т а б л и ц а 4

Зависимость средней скорости $u(r)$ от отношения r числа частиц к числу ячеек $n = 10, p = 0,5$

r	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$u^*(r)$	0,485	0.453	0.411	0.351	0.272	0.195	0.121
$u^{**}(r)$	0.469	0.429	0.380	0.320	0.253	0.184	0.117
Δ (%)	4.9	8.4	8.0	8.3	8.7	7.1	3.4

Из табл. 4 видно, что значения $u^*(r)$ превышают значения $u^{**}(r)$, что, очевидно, объясняется тем, что имитационной модели соответствовал случай, когда частицы при наличии препятствия движения на полосе имеют возможность перейти на другую полосу. Графики функций $u^*(r)$ и $u^{**}(r)$ изображены на рис. 14.

В табл. 5–6 приведены аналогичные результаты сравнения средних скоростей в двух моделях при некоторых других значениях параметров. В таблице 5 приводятся результаты для $m = 20, p = 0,5$.

Рис. 14. Зависимости $u^*(r)$ и $u^{**}(r)$

Т а б л и ц а 5

Зависимость средней скорости $u(r)$ от отношения r числа частиц к числу ячеек $n = 20, p = 0,5$

r	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$u^*(r)$	0,476	0.444	0.400	0.339	0.267	0.191	0.120
$u^{**}(r)$	0.463	0.412	0.363	0.302	0.241	0.177	0.112
Δ (%)	5.1	8.0	10.2	12.2	10.8	5.8	7.1

В табл. 6 приводятся результаты для $n = 30, p = 0,5$.

Т а б л и ц а 6

Зависимость средней скорости $u(r)$ от отношения r числа частиц к числу ячеек $n = 30, p = 0,5$

r	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$u^*(r)$	0,471	0.445	0.397	0.338	0.263	0.189	0.119
$u^{**}(r)$	0.448	0.408	0.359	0.300	0.240	0.172	0.112
Δ (%)	6.5	9.1	5.1	12.7	9.6	9.9	6.2

В рассмотренных примерах величина $\Delta(r)$ принимает значения от 1.0 до 12.7 процентов.

При фиксированных значениях n и p наибольшее значение $\Delta(r)$ достигается при $r = 0,5$. Это можно объяснить тем, что при малых значениях плотности потока редко возникает ситуация, когда требуется переход на другую полосу, а при больших нагрузках переходы на другую полосу затруднены и поэтому также редко происходят. При $p = 0,8$ значения $\Delta(r)$ меньше соответствующих значений для $p = 0,2$ и $p = 0,5$.

Работа поддержана Грантом РФФИ 08-01-0939а и проектом АВЦП 2.1.2/369.

Авторы выражают благодарность студенту 5-го курса, первого за восьмидесятилетнюю историю существования МАДИ выпуска «Прикладной математики» Ярошенко А. за активное участие в приведенных в работе численных результатах.

Литература

1. Buslaev A.P., Prichodko V.M., Tatashev A.G., Yashina M.V. The deterministic-stochastic flow model // arXiv: physics /0504139 v1[physics / soc.-ph] 20 Apr. 2005, P. 1–21.
2. Буслев А.П., Новиков А.Н., Приходько В.М., Таташев А.Г., Яшина М.В. Вероятностные и имитационные подходы к оптимизации автомобильного движения. — М.: Мир, 2003. — 286 с.
3. Buslaev A.P., Tatashev A.G. Particles Flow on the Regular Polygon [accepted to Journal of Concrete and Applicable Mathematics (JCAMM) 10.10.2010].
4. Inose H., Hamada T. Road Traffic Control. — University of Tokio Press, 1975.
5. Бланк М.Л. Точный анализ динамических систем, возникающих в моделировании транспортных потоков // Успехи математических наук. — 2000. — Т. 55. — N 3 (333). С. 167–168.
6. Blank M. L Dynamics of traffic jams: order and chaos // Mosc. Math. J. — V. 1, N. 1. — 2001. — P. 1–26.
7. Беляев Ю.К., Буслев А.П., Селезнев О.В., Таташев А.Г., Яшина М.В. Марковская аппроксимация стохастической модели движения по двум полосам // Московский автомобильно-дорожный институт (ГТУ). — М., 2002. — 32 с. Деп. ВИНТИ 03.07.2002 Г. N 1234-B — 2002.
8. Buslaev A.P., Tatashev A.G., Yashina M.V. Traffic flow stochastic model (2×2) with discrete set of states and continuous time. — arXiv: [cond-mat. other] 21 May 2004.
9. Schreckenberg M., Schadschneider A., Nagel K., Ito N. Discrete stochastic models for traffic flow // Phys. Rev. E. 1995. — V. 51, P. 2939–2949.
10. Daganzo C.F. A finite difference approximation for the kinetic wave model // Institute of transportation studies. — Trans. Res. — 1993. — V. 29. В (4).P. 261–276.
11. Бугаев А.С., Буслев А.П., Таташев А.Г. Монотонное случайное блуждание частиц по целочисленной полосе и LYuMen проблема // Мат. моделирование. Т. 18, № 12. — 2006. С. 19–34.
12. Бугаев А.С., Буслев А.П., Таташев А.Г. О моделировании сегрегации двухполосного потока частиц // Мат. моделирование. — 2008. — Т. 20, № 9. — С. 111–119.

Поступила в редакцию 15.10.2010.

А.В. Гасников, Е.В. Гасникова

Московский физико-технический институт (государственный университет)

О возможной динамике в модели расчета матрицы корреспонденций (А.Дж. Вильсона)

Приводится основной аппарат, необходимый для исследования динамики *макросистем* при больших значениях времени. В основе динамики лежит *эргодическая марковская цепь* с огромным числом состояний. При больших значениях времени распределение макросистемы по *макросостояниям* будет близко к *стационарному*. С ростом размерности макросистемы (количества состояний марковской цепи) стационарное распределение будет *концентрироваться* в окрестности *наиболее вероятного макросостояния*, которое и принимается за *равновесное* для данной макросистемы. В качестве примера применения описанного формализма приводится вывод статической *гравитационной модели* расчета матрицы корреспонденций (одной из наиболее популярных в приложениях), исходя из «разумной» (индивидуально выгодной) динамики обменов местами жительства.

Ключевые слова: эргодическая теорема, функция Ляпунова, энтропия, гравитационная модель расчета матрицы корреспонденций, концентрация инвариантной (стационарной) меры, канонический скейлинг, динамика квазисредних, условие динамического равновесия, принцип детального равновесия.

I. Введение

После работ ряда крупных биологов первой половины прошлого века, а также работ Е.Т. Дрейфа (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 достаточно прочно укрепились 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 стацио-

нарному) распределением эргодического (например, марковского) случайного процесса, порождающего рассматриваемую макросистему [35–40]. Если размерность макросистемы увеличивается, то, как правило, распределение сосредотачивается в окрестности наиболее вероятного макросостояния⁷. 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 зависимости 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].⁸

⁷Заметим, что отмеченное обстоятельство (концентрация) может быть по-разному обосновано (как правило, достаточно элементарных комбинаторных соображений и формулы Стирлинга [1, 2, 30, 33]).

⁸Укажем 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); 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]).

**II. 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, 46]). 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 $s_i(0)$ (монетками, достоинством в один рубль). Со временем пронумерованные жители (количество которых не изменяется, так же как и суммарное количество рублей) случайно разыгрывают свое имущество. Пусть в момент времени $t \geq 0$ r -й житель имеет k рублей, а l -й житель — m рублей. Тогда $p_{k;m}(t)\Delta t + o(\Delta t)$ есть вероятность того, что жители с номерами r и l ($1 \leq r < l \leq N$) 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

$$S = \sum_{i=1}^N s_i(0), \quad \bar{s} = \frac{S}{N}.$$

⁹Эргодическая 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 [47].

¹⁰Необходимо (для простоты формулировок, в рамках этой сноски считаем время дискретным) асимптотически (по размеру макросистемы) оценить второе по величине собственное значение матрицы переходных вероятностей — инфинитезимальной матрицы (первое собственное значение, которое для неотрицательных матриц часто называют числом Фробениуса–Перрона, равно единице, поскольку матрица стохастическая — все элементы неотрицательны и сумма всех элементов в любой строке равна единице), определяющее основание геометрической прогрессии, мажорируемой последовательностью 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 [48], 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. [48, 49] и цитированную там литературу, а также статью А.В. Колесникова).

Тогда по эргодической теореме для конечных однородных марковских цепей (см. [18, 19; 35–38] и упражнение ниже):⁹

$$\forall q = 0, \dots, S \exists \lambda_q > 0, T_q = O(N) : \forall t \geq T_q \rightarrow \\ \rightarrow P \left(\left| \frac{c_s(t)}{C e^{-s/\bar{s}}} - 1 \right| \leq \frac{\lambda_q}{\sqrt{N}}, s = 0, \dots, q \right) \geq 0,999,$$

где C определяется из условия нормировки:

$$\sum_{s=0}^S C e^{-s/\bar{s}} = 1, \quad \text{то есть } C \approx \frac{1}{\bar{s}}.$$

Скорость сходимости оценивается сверху, исходя из оценок в доказательстве эргодической теоремы для однородных марковских цепей с конечным числом состояний.¹⁰ 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 трюке Коровьева-Фагота. Помните булгаковского героя, который разбрасывал в варьете 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 имеющая однозначного решения. И тем не менее ответ

есть: примерно квадратный корень из миллиона, то есть тысяча зрителей.

Точнее говоря, как следует из вышенаписанного, 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 $N/\bar{s} \sim 1$.

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 меняется¹¹ 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, то есть

$$x_{ij}(t) \geq 0, \sum_{j=1}^n x_{ij}(t) \equiv L_i, \\ \sum_{i=1}^n x_{ij}(t) \equiv W_j, i, j = 1, \dots, n. \quad (A)$$

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

$$p_{k,m;p,q}^L(t) \equiv \\ \equiv p^L N^{-1} \exp(\gamma \cdot (t_{km} + t_{pq} - (t_{pm} + t_{kq}))/2) > 0,$$

где $p^L > 0$, $\gamma > 0$. Тогда по эргодической теореме для конечных однородных марковских цепей (см. [18, 19; 35–38]):

$$\forall \{x_{ij}\}_{i=1,j=1}^{n,n} \in (A) \rightarrow \\ \rightarrow \lim_{t \rightarrow \infty} P(x_{ij}(t) = x_{ij}, i, j = 1, \dots, n) = \\ = Z^{-1} \prod_{i=1,j=1}^{n,n} \exp(-\gamma t_{ij} x_{ij}) \cdot (x_{ij}!)^{-1} \stackrel{def}{=} \\ \stackrel{def}{=} p(\{x_{ij}\}_{i=1,j=1}^{n,n}),$$

где статсумма Z находится из условия нормировки получившейся пуассоновской вероятностной меры [50]. Распределение $p(\{x_{ij}\}_{i=1,j=1}^{n,n})$ на множестве (A) сконцентрировано при $N \gg 1$ (см. ниже) в окрестности наиболее вероятного значения

¹¹Поскольку 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).

¹²Нетрудно заметить, что задача поиска наиболее вероятного состояния асимптотически (по n) эквивалентна задаче максимизации энтропийного функционала (воспользовались формулой Стирлинга):

$$\ln p(\{x_{ij}\}_{i=1,j=1}^{n,n}) \sim - \sum_{i=1,j=1}^{n,n} x_{ij} \ln(x_{ij}/e) - \gamma \sum_{i=1,j=1}^{n,n} t_{ij} x_{ij} + \text{const}n$$

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), то есть по принципу Лагранжа [51] (балансовые ограничения (A) перенесли в функционал),

$$L(\{x_{ij}\}_{i=1,j=1}^{n,n}; \bar{\lambda}^L, \bar{\lambda}^W) = - \sum_{i=1,j=1}^{n,n} x_{ij} \ln(x_{ij}/e) - \gamma \sum_{i=1,j=1}^{n,n} t_{ij} x_{ij} + \\ + \sum_{i=1}^n \lambda_i^L \cdot \left(\sum_{j=1}^n x_{ij} - L_i \right) + \sum_{j=1}^n \lambda_j^W \cdot \left(\sum_{i=1}^n x_{ij} - W_j \right) \rightarrow \sup_{x_{ij} \geq 0, i,j=1,\dots,n}$$

имеет, и притом единственное решение $\{x_{ij}^*\}_{i=1,j=1}^{n,n}$, которое определяется из системы:

$$\partial L(\{x_{ij}\}_{i=1,j=1}^{n,n}; \bar{\lambda}^L, \bar{\lambda}^W) / \partial x_{ij} = - \ln x_{ij} + \lambda_i^L + \lambda_j^W - \gamma t_{ij} = 0, i, j = 1, \dots, n.$$

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, намного 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, 52]. Можно показать, см. [52], что многие популярные в литературе [30, 45] численные методы решения этой двойственной задачи являются барьерно-мультипликативными аналогами квазиградиентных итерационных методов. В частности, к ним относится один из первых методов: «метод балансировки», заключающийся в подстановке прямых переменных как функций двойственных, в

$\{x_{ij}^*\}_{i=1,j=1}^{n,n}$, которое находится, как решение задачи энтропийно линейного программирования: ¹²

$$\sum_{i=1,j=1}^{n,n} x_{ij} \ln(x_{ij}/e) + \gamma \sum_{i=1,j=1}^{n,n} t_{ij} x_{ij} \rightarrow \min_{\{x_{ij}\}_{i=1,j=1}^{n,n} \in (A)} . \quad (1)$$

Решение задачи (1) можно представить как

$$x_{ij} = \exp(-1 - \lambda_i^L - \lambda_j^W - \gamma t_{ij}),$$

где множители Лагранжа (двойственные переменные) $\{\lambda_i^L\}_{i=1}^n$ и $\{\lambda_j^W\}_{j=1}^n$ определяются из равенств системы (А). На практике мы имеем информацию о $\{L_i, W_i\}_{i=1}^n$ и $\{t_{ij}\}_{i=1,j=1}^{n,n}$. Решив задачу (А), мы найдем

$$\{x_{km}(\{L_i, W_i\}_{i=1}^n; \{t_{ij}\}_{i=1,j=1}^{n,n})\}_{k=1,m=1}^{n,n}.$$

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]:

$$A_i = \left(\sum_{j=1}^n B_j W_j \exp(-\gamma t_{ij}) \right)^{-1},$$

$$B_j = \left(\sum_{i=1}^n A_i L_i \exp(-\gamma t_{ij}) \right)^{-1}.$$

Описанная модель (точнее говоря, рассчитанная по этой модели матрица корреспонденций) активно используется в теоретико-игровых моделях (например, базирующихся на принципах Дж.Г. Вардропа) равновесного распределения потоков [32] (см. также главу 1 в пособии [54]). Подробнее об этих моделях речь пойдет в статьях Е.А. Нурминского и Н.Б. Шамрай, В.И. Швецова, Е.В. Гасниковой и Ю.В. Дорна из этого же сборника. Один из возможных способов определения времени в пути, в зависимости от загрузки ребра, предложен в статье М.Л. Бланка. Иной подход к выводу гравитационной модели имеется в статье П.П. Бобрика. Заметим также, что задача (1) по своим свойствам очень похожа на транспортную задачу (см. статью А.В. Колесникова).

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 сепарабельных (функционал декомпозируется в аддитивную сумму функций одного аргумента) алгоритмов типа Нестерова–Немировского для задач энтропийного программирования, возникающих при нахождении равновесий макросистем [53].

¹³Отметим, 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]).

Перейдем теперь к исследованию полученного стационарного распределения вероятностей $p(\{x_{ij}\}_{i=1,j=1}^{n,n})$ на макросостояниях $\{x_{ij}\}_{i=1,j=1}^{n,n} \in (A)$. Если¹³

$$N \sim n^{\alpha+1}, L_i, W_j \sim n^\alpha \quad (\alpha > 1), n \gg 1,$$

$$\forall i, j = 1, \dots, n \rightarrow t_{ij} = \tau > 0,$$

то распределение вероятностей $p(\{x_{ij}\}_{i=1,j=1}^{n,n})$ на множестве (А) 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, 51]. Формулу (2) можно интерпретировать как наличие у районов «потенциалов притяжения» [2]:

$$L_i / \sqrt{N} \sim \exp(\lambda_i^L), \quad i = 1, \dots, n$$

и

$$W_j / \sqrt{N} \sim \exp(\lambda_j^W), \quad j = 1, \dots, n,$$

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 $p(\{x_{ij}\}_{i=1,j=1}^{n,n})$ 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 сноску 12) следует

$$\forall \{x_{ij}\}_{i=1,j=1}^{n,n} \in (A) \rightarrow \sum_{i=1,j=1}^{n,n} \frac{\partial \ln p(\{x_{ij}^*\}_{i=1,j=1}^{n,n})}{\partial x_{ij}} \times (x_{ij} - x_{ij}^*) \leq 0.$$

Поэтому

$$\forall \{x_{ij}\}_{i=1,j=1}^{n,n} \in (A) \quad \exists \theta \in [0, 1] : \\ \ln p(\{x_{ij}\}_{i=1,j=1}^{n,n}) \leq \ln p(\{x_{ij}^*\}_{i=1,j=1}^{n,n}) + \\ + \sum_{i=1,j=1}^{n,n} \frac{\partial^2 \ln p(\{x_{ij}^* \theta + x_{ij} \cdot (1 - \theta)\}_{i=1,j=1}^{n,n})}{\partial x_{ij}^2} \times \\ \times \frac{(x_{ij} - x_{ij}^*)^2}{2}.$$

Но

$$\frac{\partial^2}{\partial x_{ij}^2} (\ln p(\{x_{ij}\}_{i=1,j=1}^{n,n})) =$$

$$= \frac{\partial^2}{\partial x_{ij}^2} \left(- \sum_{i=1, j=1}^{n, n} x_{ij} \ln x_{ij} \right) = - \frac{1}{x_{ij}}.$$

Следовательно, приходим к «неравенству о концентрации меры»:

$$\forall M > 0, \quad \forall \{x_{ij}\}_{i=1, j=1}^{n, n} \in (A) :$$

$$\sum_{i=1, j=1}^{n, n} \frac{(x_{ij} - x_{ij}^*)^2}{2 \max\{x_{ij}, x_{ij}^*\}} \geq M \rightarrow$$

$$\rightarrow p(\{x_{ij}\}_{i=1, j=1}^{n, n}) \leq e^{-M} p(\{x_{ij}^*\}_{i=1, j=1}^{n, n}).$$

Из этого неравенства имеем результат о концентрации распределения $p(\{x_{ij}\}_{i=1, j=1}^{n, n})$ на множестве (A) в $O(n^{(\alpha-1)/2})$ окрестности наиболее вероятного значения $\{x_{ij}^*\}_{i=1, j=1}^{n, n}$:

$$\exists \lambda > 0 : \lim_{t \rightarrow \infty} P(\forall i, j = 1, \dots, n \rightarrow |x_{ij}(t)/x_{ij}^* - 1| \leq \leq \lambda/n^{(\alpha-1)/2}) \geq 0,999.$$

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 [55], см. также статью А.А. Замятина и В.А. Малышева в этом сборнике и конец следующего пункта). Исследование концентрации в окрестности математического ожидания можно также приводить, например, используя предельные меры [56], метод канонических ансамблей [57] или обобщённую схему размещения [58], нашедшие применения к задачам асимптотической перечислительной комбинаторики,¹⁴ 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 [58], метод пуассоновской и гауссовской аппроксимации (метод локальной предельной теоремы) [7, 58]. Другой способ восходит к принципу концентрации А. Пуанкаре и П. Леви, получившему дальнейшее развитие в работах В.Д. Мильмана и др. [59] — в этом случае концентрация наблюдается в окрестности медианы.¹⁵ В заключение краткого обзора

методов исследования концентрации меры упомянем теоремы тауберова типа [61] и мартингалльные неравенства [60]. 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 [59].

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, 62] и цитированную там литературу)), в том числе встречающихся в экономике, биологии, социальной сфере [3, 4; 15-24].

III. 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 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,

¹⁴В частности, к теории случайных графов (компьютерные, транспортные сети — вопросы надёжности и т.п. [24, 39], см. также статью А.М. Райгородского в этом сборнике).

¹⁵Этот 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 [60] (см. также статьи А.М. Райгородского и А.В. Колесникова в этом сборнике).

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 (Мальшев–Пирогов–Рыбко, 2004 [19, 63]) 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):

$$c_i(t) \stackrel{\text{п.н.}}{=} \lim_{M \rightarrow \infty} n_i(t)/M.$$

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 $n_i(t)$.¹⁶ Пусть существует хотя бы одно положительное решение $\vec{\xi}$ системы (условие динамического равновесия, В.В. Веденяпин (2001) [18], Мальшев–Пирогов–Рыбко (2004) [19, 63], Батищева–Веденяпин [64]):

$$\sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = \sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j}, \quad (4)$$

где $K_{\vec{\beta}}^{\vec{\alpha}}(\vec{c}) \equiv K_{\vec{\beta}}^{\vec{\alpha}}$, а $\vec{\alpha}$ — произвольный, но такой, что $\exists \vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J$.

Частным, но часто встречающимся в приложениях, случаем условия (4) является условие детального равновесия:

$$K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j}.$$

¹⁶Заметим, 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».

¹⁷Стоит заметить, что аттрактор системы (3) с постоянными коэффициентами реакции, по-видимому, в общем случае может быть сколь угодно сложным множеством [19, 65].

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»)¹⁷. 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):

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

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

$$\begin{aligned} \sum_{(\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} &= \sum_{(\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j} y_j^{\alpha_j} = \\ &= \sum_{(\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\beta_j}. \end{aligned}$$

Таким образом,

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

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

Оказывается (Мальшев–Пирогов–Рыбко [19, 63]), что условие (4) можно проинтерпретировать

как условие инвариантности пуассоновской меры [50]:¹⁸

$$\mu(\vec{n}) = \prod_i \lambda_i^{n_i} e^{-\lambda_i} / n_i!,$$

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 $A = \|A_{kl}\|$ — матрица максимального ранга ($k = 1, \dots, m$). Обозначим через A множество неотрицательных целочисленных векторов \vec{n} , удовлетворяющих $A\vec{n} = \vec{d}$. Тогда равновесие \vec{n}^* находится как решение задачи $E(\vec{n}) \rightarrow \max_{\vec{n} \in A}$ (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 уравнений $A\vec{n}(\vec{y}^*) = \vec{d}$.

Приведем, во многом следуя [2], другой путь (восходящий к Дарвину–Фаулеру), по которому можно прийти к аналогичным формулам.

Для этого введем производящую функцию:

$$\begin{aligned} F(\vec{z}; A) &= \sum_{\vec{n} \geq \vec{0}} \mu(\vec{n}) \prod_k z_k^{\sum_l A_{kl} n_l} = \\ &= \prod_i \sum_{n_i \geq 0} \frac{(\lambda_i \prod_k z_k^{A_{ki}})^{n_i} e^{-\lambda_i}}{n_i!} = \\ &= \prod_i \exp \left(\lambda_i \cdot \left(\left(\prod_k z_k^{A_{ki}} \right) - 1 \right) \right). \end{aligned}$$

Тогда по формуле Коши:

$$\begin{aligned} E[n_{r_1}^{p_1} \dots n_{r_Q}^{p_Q}] &= \frac{1}{Z} \frac{1}{(2\pi i)^m} \oint dz_1 \dots dz_m \times \\ &\times \left[\left\{ \prod_k z_k^{-d_k-1} \right\} \cdot \left\{ \prod_q \left(\frac{1}{\ln z_1} \frac{\partial}{\partial A_{1r_q}} \right)^{p_q} F(\vec{z}; A) \right\} \right], \\ Z &= \frac{1}{(2\pi i)^m} \oint dz_1 \dots dz_m \cdot \left[\left\{ \prod_k z_k^{-d_k-1} \right\} \cdot F(\vec{z}; A) \right]. \end{aligned}$$

Здесь математическое ожидание $E[n_{r_1}^{p_1} \dots n_{r_Q}^{p_Q}]$ 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 [67], асимптотически оценим математическое ожидание:

$$\begin{aligned} E[n_{r_1}^{p_1} \dots n_{r_Q}^{p_Q}] &\approx \frac{1}{F(\vec{z}^*; A)} (\ln z_1^*)^{-\sum_q p_q} \times \\ &\times \left\{ \prod_q \left(\frac{\partial}{\partial A_{1r_q}} \right)^{p_q} F(\vec{z}^*; A) \right\}, \end{aligned}$$

где «точка перевала» \vec{z}^* определяется как решение системы:

$$z_k \cdot (\partial F(\vec{z}; A) / \partial z_k) \approx d_k \cdot F(\vec{z}; A), \quad k = 1, \dots, m.$$

В частности,¹⁹

$$E[n_i] \approx \lambda_i \prod_k (z_k^*)^{A_{ki}}, \quad D[n_i] \approx \lambda_i \prod_k (z_k^*)^{A_{ki}},$$

где \vec{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» \vec{z}^* с двойственными переменными \vec{y}^* : $z_k^* = \exp(y_k^*)$.

IV. Заключение

В статье 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 статье схему.

¹⁸Пуассоновская 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) [66].

¹⁹Обратим 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 $E[n_i]$.

1. Макросистема состоит из огромного числа пронумерованных агентов, каждый из которых может находиться в одном из возможных состояний. Число состояний, как минимум, на несколько порядков 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 макрозакон сохранения. Соотношения (как правило, линейные) между макровеличинами, 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. Приятной особенностью такого класса задач является явная (легко выписываемая) зависимость решения прямой задачи через двой-

ственные переменные. Поскольку число ограничений, как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы базируются на решении двойственной задачи: задачи минимизации выпуклой функции.

Авторы выражают благодарность за ряд ценных замечаний В.В. Веденяпину, И.А. Лубашевскому, В.А. Малышеву, Е.А. Нурминскому, С.А. Пирогову, Ю.С. Попкову, И.Г. Поспелову, В.Н. Разжевайкину, С.В. Чуканову, Н.Б. Шамрай, А.А. Шананину, В.И. Швецову.

Работа поддержана грантами РФФИ № 08-01-00959-а, 08-07-00501-а, 08-07-00158-а, 10-01-00321-а, 10-07-00620-а, 11-01-00494-а, РГНФ № 08-02-00347, ПФИ ОМН РАН № 3, ПФИ Президиум РАН П-2. Работа 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).

Настоящая статья представляет собой запись нескольких лекций, прочитанных первым автором студентам МФТИ в весеннем семестре 2009/2010 учебного года в рамках курса по выбору «Математическое моделирование транспортных потоков».

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. Свирежжев 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
22. Занг В.-Б. Синергетическая экономика: время и перемены в нелинейной экономической теории. М.: Мир, 1999.
23. Dragulescu A., Yakovenko V.M. Statistical mechanics of money // The European Physical Journal B. 2000. V. 17. P. 723–729. arXiv: cond-mat/0001432 v4
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 системостатика. М.: Наука, 1986.
28. Малышевский 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. Квантовая 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: Марковские процессы как отправная точка теории случайных процессов и их приложения. М.: МЦНМО, 2010.
39. Вишневецкий В.М. Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003.
40. Ивницкий 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. Богданов К.Ю. Прогулки с физикой. Библиотечка «Квант» В. 98. М.: Бюро Квантум, 2006 (глава 18).
47. Зорич В.А. Математический анализ задач естествознания. М.: МЦНМО, 2008.
48. 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>
49. Joulin A., Ollivier Y. Curvature, concentration and error estimates for Markov chain Monte Carlo // Ann. Prob. 2010. V. 38, № 6. — P. 2418–2442.
50. Кингман Дж. Пуассоновские процессы / под ред. А.М. Вершика. М.: МЦНМО, 2007.
51. Магарил-Ильяев Г.Г., Тихомиров В.М. Выпуклый анализ и его приложения. М.: УРСС, 2003.
52. Гасникова Е.В. Двойственные мультипликативные алгоритмы для задач энтропийно-линейного программирования // ЖВМ и МФ. 2009. Т. 49, № 3. — С. 453–464.
53. Нестеров Ю.Е. Введение в выпуклую оптимизацию. М.: МЦНМО, 2010.

- 54.** 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.
- 55.** Flajolet P., Sedgewick R. Analytic combinatorics. Cambridge University Press, 2008. <http://algo.inria.fr/flajolet/Publications/book.pdf>
- 56.** 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.
- 57.** Синай Я.Г. Вероятностный подход к анализу статистики выпуклых ломаных // Функц. анализ и его прил. 1994. Т. 28. № 2. С. 41–48.
- 58.** Колчин В.Ф. Случайные графы. М.: Физматлит, 2004.
- 59.** Ledoux M. Concentration of measure phenomenon. Providence, RI, Amer. Math. Soc., 2001 (Math. Surveys Monogr. V. 89).
- 60.** Алон Н., Спенсер Дж. Вероятностный метод. М.: Бином, 2007.
- 61.** Якимив А.Л. Вероятностные приложения тауберовых теорем. М.: Наука, 2005.
- 62.** Ширяев А.Н. Вероятность 1, 2. М.: МЦНМО, 2007.
- 63.** Malyshev V.A., Pirogov S.A., Rubco A.N. Random walks and chemical networks // Mosc. Math. J. 2004. V. 4, № 2. — P. 441–453.
- 64.** Батищева Я.Г., Веденяпин В.В. II-й закон термодинамики для химической кинетики // Матем. моделирование. 2005. Т. 17, № 8. — С. 106–110.
- 65.** Веденяпин 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.
- 66.** Рыбко А.Н. Пуассоновская гипотеза для больших симметричных коммуникационных сетей // Глобус. Общественно-математический семинар / под ред. М.А. Цфасмана и В.В. Прасолова. № 4. М.: МЦНМО, 2009. — С. 105–126.
- 67.** Федоряк М.В. Метод перевала. М.: УРСС, 2010.

Поступила в редакцию 15.10.2010.

Е.В. Гасникова, Ю.В. Дорн

Московский физико-технический институт (государственный университет)

О стохастической марковской динамике, приводящей к равновесию Нэша–Вардропа в модели распределения потоков

Описывается возможная динамика, приводящая к равновесию Нэша–Вардропа в модели распределения потоков. Следует заметить, что сделанные выводы базируются в основном на результатах численных экспериментов, которые показали в ряде игр очень быструю сходимость предложенной стохастической (гиббсовской) марковской динамики наилучших ответов в соответствующей эволюционной игре к равновесию Нэша.

Ключевые слова: эволюционные игры, равновесие Нэша, парадокс Брайеса, принципы Дж. Г. Вардропа, эффективность по Парето, алгоритм Григориадиса–Хачияна.

Ориентированный граф $\Gamma = (V, E)$ представляет собой транспортную сеть города (V – узлы сети (вершины), $E \subset V \times V$ – дуги сети (рёбра графа)). Пусть $W = \{w = (i, j) : i, j \in V\}$ – множество пар источник–сток; $p = \{v_1, v_2, \dots, v_m\}$ – путь из v_1 в v_m , если $(v_k, v_{k+1}) \in E, k = 1, \dots, m - 1$; P_w – множество путей, отвечающих корреспонденции $w \in W$; $P = \bigcup_{w \in W} P_w$ – совокупность всех путей в сети Γ ; x_p – величина потока по пути p , $x = \{x_p : p \in P\}$; $G_p(x)$ – удельные затраты на проезд по пути p , $G(x) = \{G_p(x) : p \in P\}$; y_e – величина потока по дуге e :

$$y_e = \sum_{p \in P} x_p \delta_{pe}, \quad \text{где} \quad \delta_{pe} = \begin{cases} 1, e \in p, \\ 0, e \notin p, \end{cases}$$

$\tau_e(y_e)$ – удельные затраты на проезд по дуге e (как правило, возрастающие, выпуклые, гладкие функции), при этом естественно считать, что $G_p(x) = \sum_{e \in E} \tau_e(y_e) \delta_{pe}$. Если $\tau_e(y_e)$ – возрастающие функции, то отображение $G(x)$ – строго монотонное. Заметим, что в приложениях часто требуется учитывать и затраты на прохождения вершин графа (в свою очередь эти затраты могут зависеть, вообще говоря, от величин всех потоков, проходящих через каждую рассматриваемую вершину). Пусть также известны потоки корреспонденций $d_w, w \in W$. Тогда вектор x , характеризующий распределение потоков, должен лежать в допустимом множестве:

$$X = \left\{ x \geq 0 : \sum_{p \in P_w} x_p = d_w, w \in W \right\}.$$

Это множество может иметь и другой вид, если дополнительно учитывать, например, конечность пропускных способностей рёбер (ограничения сверху на y_e).

²⁰Например, 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 $G_p(x(n))$ 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.

Рассмотрим игру, в которой каждому элементу $w \in W$ соответствует свой, достаточно большой ($d_w \gg 1$), набор однотипных игроков (сидящих на корреспонденции w). Множеством чистых стратегий каждого такого игрока является P_w , а выигрыш (потери со знаком минус) определяются формулой $-G_p(x)$ (игрок выбирает путь следования $p \in P_w$, при этом он пренебрегает тем, что от его выбора также немного зависят $|P_w|$ компонент вектора x и, следовательно, сам выигрыш $-G_p(x)$). Тогда, считая отображение $G(x)$ непрерывным и строго монотонным (этого достаточно), можно показать, что отыскание (единственного!) равновесия Нэша (1951) $x^* \in X$ (макроописание равновесия) равносильно решению задачи дополнителности (принцип Дж. Г. Вардропа (1952)), что в свою очередь равносильно решению вариационного неравенства:

$$\forall w \in W, p \in P_w \rightarrow x_p^* \cdot (G_p(x^*) - \min_{q \in P_w} G_q(x^*)) = 0 \quad \Leftrightarrow \quad \forall x \in X \rightarrow \langle G(x^*), (x - x^*) \rangle \geq 0.$$

Вариационное неравенство можно переписать как проекционное уравнение

$$x^* = \Pi_X(x^* - \lambda G(x^*)), \quad \lambda > 0,$$

где $\Pi_X(x^* - \lambda G(x^*))$ – такая «точка» множества X , которая доставляет минимум функционалу расстояния от точки $x \in X$ до фиксированной точки $x^* - \lambda G(x^*)$. Выписанное проекционное уравнение можно далее численно решать, например, с помощью метода простой итерации $x^{n+1} = \Pi_X(x^n - \lambda G(x^n))$. Более того, в рассматриваемом случае задача отыскания равновесия Нэша–Вардропа сводится к решению следующей задачи выпуклого программирования [1]:

$$\sum_{p \in P} x_p \delta_{pe} \int_0^{\tau_e(z)} \tau_e(z) dz \rightarrow \min_{x \in X}.$$

В данной заметке предлагается возможная динамика в этой игре, приводящая к равновесию Нэша–Вардропа. Свой путь на $(n + 1)$ -м шаге²⁰ игрок, сидящий на корреспонденции w , выбирает согласно смешанной стратегии (в независимости от всех остальных): с вероятностью

$$Prob_p^w(n + 1) = \gamma_n \cdot (x_p(n) + 1) \times \\ \times \exp(-G_p(x(n))/T_n) / Z_n^w, \quad w \in W,$$

выбрать путь $p \in P_w$ ($0 < \gamma_n \leq 1$), а с вероятностью $1 - \gamma_n$ действовать согласно стратегии, использованной на предыдущем n -м шаге. Здесь $x_p(n)$ — количество игроков, сидящих на корреспонденции w и выбравших на n -м шаге стратегию $p \in P_w$, а

$$Z_n^w = \sum_{p \in P_w} \gamma_n \cdot (x_p(n) + 1) \exp(-G_p(x(n))/T_n).$$

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 статье 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 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 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.

Введение в динамику стохастичности сближает предложенный подход с поиском так называемых «стохастических равновесий в транспортных сетях» [2], с другой стороны, подход данной статьи принципиально от них 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 [3]–[5], при этом имеется много общего с концепциями quantal response equilibria [6] (используется похожая рандомизация) и minority games [7] (наблюдаются похожие колебания около равновесия). Также близким к предложенному итерационному процессу является концепция генетических алгоритмов [8] и предложенный на их основе эффективный вероятностный (с гиббсовским распределением) алгоритм Григориадиса–Хачияна (1995) [9] поиска ε приближенного равновесия Нэша в матричной игре $n \times n$ за $O(n \cdot (\log^2 n) / \varepsilon^2)$ операций с плавающей точкой. Стоит заметить, что в классе детерминированных алгоритмов необходимо осуществлять не менее $\sim n^2$ таких операций.

В заключение рассмотрим пример, демонстрирующий, что в результате строительства новой дороги новое равновесие Нэша–Вардропа окажется не эффективным по Парето и будет строго хуже, чем то, которое было до строительства. Тем не менее предложенная выше марковская динамика наилучших ответов приводит именно к такому, не оптимальному по Парето, равновесию.

Пример (парадокс Брайеса, 1968).

Пусть корреспонденция $x_{14} = 6$ (тысяч автомобилей/час). 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/час). Например, в случае 2: $y_{24} = x_{124} + x_{1324}$ (рис. 1). Естественно считать, что время движения — возрастающая функция потока.

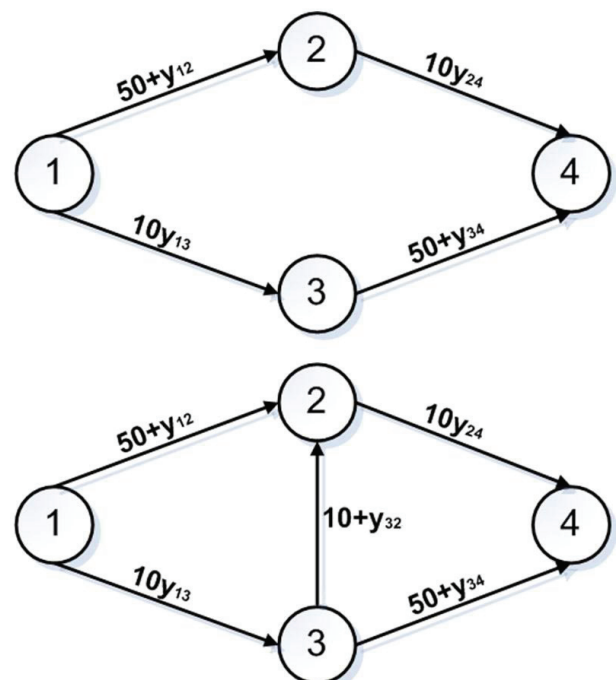


Рис. 1. Случай 1: $x_{124} = x_{134} = 3$ (полное время в пути $T = 83$ мин). Случай 2: $x_{124} = x_{1324} = x_{134} = 2$ (полное время в пути $T = 92$ мин)

Оба равновесия Нэша–Вардропа (в случае 1 и 2) являются притягивающими положениями равновесия описанной выше динамики (положили $\gamma \sim 1$, $T \sim 15 - 35$), рис. 2, 3 (для случая 2).

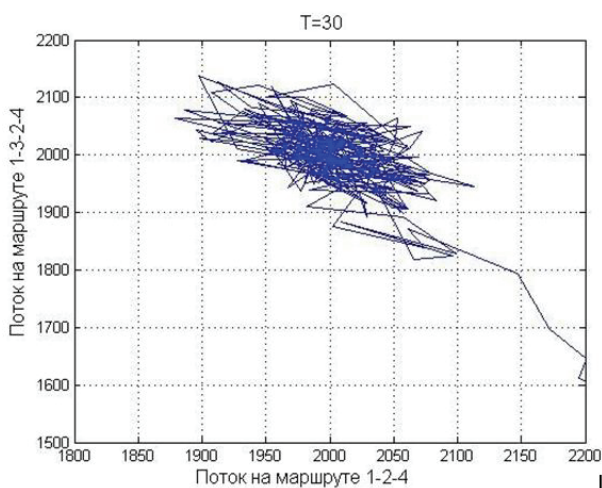
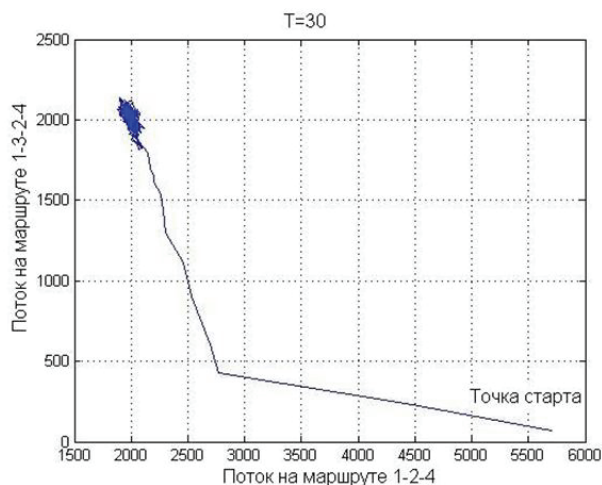


Рис. 2

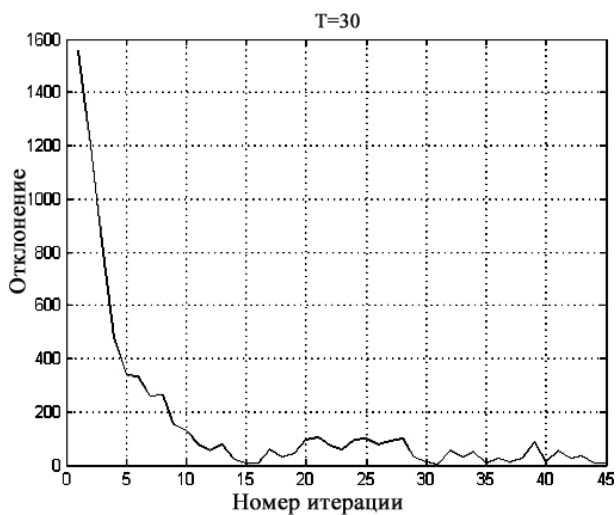


Рис. 3

Более подробно о моделях распределения потоков и связанных с ними задачах можно прочи-

тать, например, в книгах [1, 2, 10]. Значительно более подробные материалы, посвященные эффективным численным методам отыскания равновесий Нэша–Вардропа и решениям задач выпуклого программирования, содержатся в следующих электронных ресурсах [11].

Авторы выражают благодарность за ряд ценных замечаний А.В. Гасникову, И.С. Меньшикову, Е.А. Нурминскому, С.П. Тарасову, А.А. Шананину, В.И. Швецову.

Работа поддержана грантами РФФИ № 10-07-00620-а, 10-01-00321-а, 11-01-00494-а. Работа проведена в рамках реализации ФЦП «Кадры инновационной России» на 2009–2013 годы (меропр. 1.3.1, НК-215П, П1490).

Литература

1. Гасников А.В., Кленов С.Л., Нурминский Е.А., Холодов Я.А., Шамрай Н.Б. Введение в математическое моделирование транспортных потоков; учебное пособие / под ред. А.В. Гасникова с приложениями М.Л. Бланка, Е.В. Гасниковой, А.А. Замятина и В.А. Малышева, А.В. Колесникова, А.М. Райгородского. М.: МФТИ, 2010.
2. Sheffi Y. Urban transportation networks: Equilibrium analysis with mathematical programming methods. N.J.: Prentice-Hall Inc., Englewood Cliffs, 1985.
3. Foster D., Young P. Stochastic evolutionary game dynamics // Theoretical population biology. 1990. V. 38. № 2.
4. Cressman R. Evolutionary game theory and extensive form games. Cambridge: Mass. MIT Press, 2003.
5. Hofbauer J., Sigmund K. Evolutionary game dynamics // Bulletin of the AMS. 2003. V. 40, № 4. P. 479–519.
6. McKelvey R.D., Palfrey T.R. Quantal response equilibria for extensive form games // Experimental economics, 1998. V. 1. P. 9–41.
7. Marsili M. Toy models of markets with heterogeneous interacting agents // e-print. — www.unifr.ch/econophysics/
8. Fogel D.B. Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. — New York: IEEE Press, 2000.
9. Хачиян Л.Г. Избранные труды. [сост. С.П. Тарасов] М.: МЦНМО, 2009.
10. Стенбринк П.А. Оптимизация транспортных сетей. М.: Транспорт, 1981.
11. <http://www2.isye.gatech.edu/~nemirovs/> — <http://www.core.ucl.ac.be/staff/biosketchNestorov.html> — <http://elis.dvo.ru/~nurmi/>

Поступила в редакцию 15.10.2010.

А.А. Замятин, В.А. Малышев

Московский государственный университет им. М.В. Ломоносова

Транспортные автомобильные потоки — введение в вероятностный подход

Вероятностные модели транспортных потоков, отражающие качественные явления на автомагистрали, весьма разнообразны: от массового обслуживания середины прошлого века до статистической физики и процессов с локальным взаимодействием. В статье перечислены возможные постановки задач и подробно разобраны некоторые примеры.

Ключевые слова: точечные случайные процессы, сети массового обслуживания, транспортные потоки, случайные грамматики.

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 обсуждаем здесь строгие вероятностные подходы к транспортным потокам и сетям. Вероятностные модели транспортных потоков, отражающие качественные явления на автомагистрали, весьма разнообразны: от массового обслуживания середины прошлого века до статистической физики и процессов с локальным взаимодействием. Основная цель этой статьи — не столько представить технику решения задач, сколько представить методику (и искусство) составления адекватных моделей, которые отличаются наглядностью определений (основной объект там именно автомобили, 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. В статье перечислены возможные постановки задач и подробно разобраны некоторые примеры. Вообще, вероятностные модели должны связываться с психикой водителей, если водители — не роботы. Хотя такая математическая теория пока не вполне понята (в этом и есть как раз существенное отличие от статистической физики), здесь делаются, по-видимому, первые попытки построить строгие модели.

Вероятностный подход к транспортным потокам существует уже более 50 лет, см. [20, 24, 36], однако здесь мы даем более современную трактовку и рассматриваем более сложные задачи. Несмотря на то, что рассмотренные модели демонстрируют богатство возможностей, наш обзор, конечно, далеко не полон, см., например, [1, 4, 17, 27].

Статья состоит из трех частей. В первой дано построение случайных потоков и некоторые модели, в том числе и ряд новых, отражающих разнообразные качественные явления на автомагистрали. Во второй части показано, как можно полу-

чать явные формулы с помощью техники пуассоновских потоков. В третьей рассмотрены сложные сети дорог и вычисление критической нагрузки, выше которой начинаются пробки.

I. Потоки автомобилей

I.1. Маркированные 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 поток динамически — как меру на множестве траекторий $\{x_i(t)\}$ автомобилей.

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, ..., k , представляемая k прямыми, параллельными оси x . Тогда индекс $\alpha = (m, i)$ выделяет i -й автомобиль на полосе m . Индекс i нумерует автомобили 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 мы везде опускаем — читатель может его добавлять где надо — и используем только индекс i .

Обозначим расстояние от 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 за ним 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, [31]) и теория гиббсовских точечных полей [32, 34]. 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 (не пуассоновские) легко строятся на полупрямой $[0, \infty)$. 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 полупрямой предел

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t, t + \Delta t)$$

существует и стремится при $t \rightarrow \infty$ к $\rho = (E\xi)^{-1}$. Тогда 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 с функцией 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 так же, как и выше. В нашем случае чередуются длины автомобилей d_i (независимые и одинаково распределенные) и расстояния d_i^+ между ними (также независимые и одинаково распределенные).

Маркированные потоки. Каждой точке x_i точечного процесса может быть сопоставлена величина σ_i , принимающая значения в некотором множестве \mathcal{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 счетных множествах 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 I.5 мы построим маркированный процесс, где марками будут скорости, причем их распределение будет сложным образом коррелировать во времени с траекториями точек.

I.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 выбираемое расстояние D_i^+ зависит только от этой скорости:

$$D_i^+ = D_i^+(v_{i-1})$$

(индекс i свидетельствует о том, 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 водителей и должно находиться экспериментально, см. [21].

Детерминированная динамика без обгона. Если все автомобили, водители и скорости v одинаковы, то многие задачи решаются просто. Обозначим d — длину автомобиля и через $d^+ = D^+(v)$ 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 и где

$$\lambda(v) = \frac{k}{d + D^+(v)}$$

есть плотность автомобилей на k -полосной дороге при заданной скорости v в простейшем алгоритмическом потоке. Отсюда видно, что пропускная способность может уменьшаться при увеличении скорости. Этот простой вывод говорит лишь о том, что многие водители увеличивают расстояние до предыдущего автомобиля при увеличении им скорости.

Случайная динамика без обгона. То же самое получится, если скорости v одинаковы, а 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 необязательны вероятностные или физические 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. [1, 21, 29].

Случайная динамика с обгоном (случайные грамматики). Здесь естественно возникает связь с таким недавно открытым объектом, как случайные грамматики, см. [33]. Мы дадим краткое содержательное описание одной такой модели.

Пусть в момент $t = 0$ все автомобили находятся на левой полуси, 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 $s_1(t), s_{N(t)}(t) \neq 0$ 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 подслово. 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 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 N слова S (возникает лишняя свободная ячейка), что ведет к сдвигу всех автомобилей сзади данного водителя на 1 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).

5. Некоторые скачки в конце слова S могут привести к тому, что последний символ слова станет нулевым. В этом случае он немедленно уничтожается, то есть длина слова уменьшается на 1.

Необходимо сказать, что для точной формулировки результатов, которые мы лишь обрисовем, надо делать разнообразные скейлинги параметров t, N и $\lambda_0^+, \lambda_1^+, \lambda_2^-, \lambda_2^+$. Это будет сделано в отдельной статье. В зависимости от 4 параметров $\lambda_0^+, \lambda_1^+, \lambda_2^-, \lambda_2^+$ 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, 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. [1]).

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 достаточно плотный (мало ячеек с нулями), так как много автомобилей типа 2 будет тормозиться. В результате перед каждым автомобилем типа 2 возникнет много свободного места и поток разрядится.

Можно усложнять введенную динамику, например избежать дискретизации, вводя вместо нулей положительные вещественные числа — расстояния между последовательными автомобилями. Это потребует существенных переформулировок, особенно для скачков типа 3, но сохранит грубые качественные эффекты.

1.3. 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, 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 одна за другой 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.

Нетрудно доказать, что за время $t \rightarrow \infty$ пробка (то есть максимальная длина $L(t)$ участка, где все автомобили стоят) перед 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 [28] 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 . Но доказательство там совсем не просто. Другие модели роста пробки см. в [16].

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 $v < v_{max}$. Причем быстрее 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:

$$\frac{L}{v} - \frac{L}{v_1}.$$

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.

1.4. Модели начала движения

В работе [5] 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 движется без остановки. Тогда для любых i и k и любого $t > t_i, t_{i-1}, \dots, t_{i-k}$ случайные величины

$$x_{i-1}(t) - x_i(t), x_{i-2}(t) - x_{i-1}(t), \dots, x_{i-k}(t) - x_{i-k+1}(t)$$

будут независимы и экспоненциально распределены. Иначе говоря, после выезда из пробки автомобили будут образовывать пуассоновскую конфигурацию с той же самой интенсивностью ρ , что и в начале.

Рассмотрим более простую ситуацию, когда в момент 0 пуассоновский точечный поток

$$\dots < x_2 < x_1 < 0$$

с плотностью ρ находится на левой полуоси. Каждая точка движется со скоростью $v > 0$, если расстояние до предыдущей точки не меньше некоторого $d > 0$, 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 $\tau_k^{(2)}$. Интересно было бы 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 . Автомобилиям 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 автомобилей из второй полосы было не менее некоторого числа $d^+(v)$.

1.5. 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. 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 препятствия) если автомобиль i имеет впереди себя препятствие в момент t , то $v_i(t) = v_{i-1}(t)$. При этом мы говорим, что автомобиль i имеет впереди себя препятствие в момент t , если

$$x_i(t - 0) = x_{i-1}(t);$$

- 2) (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 $v_{i_{k+1}}(t_{k+1}), \dots, v_{i_n}(t_n)$;
- 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. Несколько из них мы сейчас сформулируем в виде гипотез.

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

$$\text{cov}_{ij}(t) = Ev_i(t)v_j(t) - Ev_i(t)Ev_j(t) = 0.$$

Гипотеза 1. 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.

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 $|j - i| \rightarrow \infty$. Неизвестно, будет ли $\lambda_{cr} = \lambda_0$.

Эти три фазы могут иметь отношение к фазам, определенным Б.С. Кернером [15]. Можно определить подобный процесс в более общем случае. Именно с длинами d_i, d_i^+ , а также с дополнительными индексами, соответствующими полосам движения, и с поведением водителя, зависящим не только от следующего, но и от предыдущего автомобиля. При этом могут возникать и другие качественные эффекты.

II. Расчет средней скорости на автотрассе

Мы приводим здесь простейшую постановку задачи о снижении средней скорости движения автомобилей по автотрассе 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.

II.1. Дорога как одномерная сеть массового обслуживания

Следующая модель заимствована из [14, 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 медленный, 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. 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 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 установился. 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 равно 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.

II.2. О связи вероятностного подхода с кинетическими уравнениями

Каждой частице в точке x_i на прямой сопоставляется марка — скорость $v_i > 0$. Тогда поток машин определяется счетным подмножеством $\{(x_i, v_i)\}$ точек фазового пространства $R \times R_+$. А случайный поток определяется мерой μ на множестве таких конфигураций. Обозначим $n(A) = n_\mu(A)$ — среднее число частиц в ограниченном объеме A фазового пространства. Если существует такая функция $f(x, v)$, что для всех A

$$n(A) = \int_A f(x, v) dx dv,$$

то f называется одночастичной (корреляционной) функцией распределения μ .

Хотя кинетический подход к транспортным задачам известен давно [19] и много разрабатывался в физических работах [11, 12, 37], следует сразу сказать, что строгий математический вывод кинетических уравнений (то есть уравнений для $f(t; x, v)$) для транспортных потоков остается открытым. Мы даем лишь краткий комментарий.

Пусть в момент $t = 0$ задано распределение μ с одночастичной корреляционной функцией $f(0, x, v)$. Если движение частиц свободное, то есть каждая частица движется со своей, но строго постоянной, скоростью, то

$$f(t + \delta; x, v) = f(t; x - v\delta, v).$$

Вычитая $f(t; x, v)$ из обеих частей этого равенства, деля на δ и переходя к пределу $\delta \rightarrow 0$, получаем

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} = 0. \quad (3)$$

Это простейшее кинетическое уравнение для системы частиц без столкновений. Решением будет, конечно,

$$f(t; x, v) = f(0; x - vt, v). \quad (4)$$

В общем случае в кинетическом уравнении должен быть также столкновительный член. Столкновением частиц-автомобилей могло бы служить их сближение, так что один мешает другому совершить определенные маневры, например, обгон. Как бы происходит образование новой составной частицы (кластера из нескольких автомобилей), которая через некоторое время распадается на отдельные частицы. Это не так легко формализовать.

Другой подход — рассматривать движение быстрых автомобилей в среде медленных. Предположим сначала, что автомобили и водители одинаковы, по-прежнему движутся независимо, но скорость $v(x, t)$ каждой зависит от x (например, от состояния дороги в данном месте) и от t (например, от погодных условий). Иначе говоря, движение каждой частицы определяется уравнением

$$\frac{dx}{dt} = v(x, t), x(0) = x_0.$$

При этом траектории частиц не пересекаются. Тогда вместо $f(t; x, v)$ достаточно плотности $\rho(t, x)$ частиц в точке x в момент t . Если $x(t, x_0)$ — решение этого уравнения, то плотность меняется по закону

$$\rho(t, x) = \rho(0, x_0),$$

если $x = x(t, x_0)$. Пусть теперь имеются два типа автомобилей — со скоростями v и $v_1 > v$ соответственно. Медленные автомобили едут независимо и для них одночастичная функция имеет вид (4). Перейдем в систему координат, движущуюся со скоростью v . Если задать правила обгона для быстрых автомобилей и выразить через них измененную скорость $v_1(t, x)$ «около» точек, соответствующих медленным автомобилям, то получится плотность $\rho_1(t, x)$ быстрых автомобилей в движущейся системе координат. Плотность же в исходной системе координат имеет вид $\rho(t, x - vt)$.

С гидродинамическими уравнениями (типа Бюргерса) такие же проблемы для трафика, хотя существуют строгие работы по выводу гидродинамических уравнений для exclusion processes, см. обзоры [2, 8]. Заметим, что, безусловно, интересно было увидеть некоторые аналогии между потоком автомобилей и движением жидкости. Однако, что касается связи между стохастическим и гидродинамическим подходами, есть две возможности — либо выводить гидродинамические уравнения для автомобильных потоков из микромоделей (по аналогии со статистической физикой), либо получать качественные закономерности транспортных потоков напрямую из детерминированного или случайного движения частиц, что кажется авторам этого обзора более естественным.

II.3. 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 дороги $(x, x + dx) \subset R$ 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 препятствие 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 объезд запрещен и автомобиль вынужден стоять до тех пор, пока не уберут препятствие, после чего автомобиль мгновенно набирает свою скорость v . Во втором случае объезд разрешен. Более точно, автомобилю требуется некоторое случайное время для того, чтобы объехать препятствие или группу автомобилей, стоящих перед препятствием, причем время обгона не зависит от размера этой группы. Обозначим через $\eta_{m,i}$ случайное время объезда 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 какой-то один 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

$$h(t) = m_Q^{-1}(1 - Q(t)). \quad (5)$$

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 $x \rightarrow \infty$

$$\frac{x}{T(x)} \rightarrow \frac{v}{1 + av}. \quad (6)$$

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

$$T_0(x) = \sum_{i=1}^{\pi(x)} \alpha_i, \quad (7)$$

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.

Из (7) и усилленного закона больших чисел легко следует, что $\frac{T_0(x)}{x} \rightarrow ab$ п.н. при $x \rightarrow \infty$.

Докажем (7). 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 [7]:

Лемма 1. Маркированное точечное поле Π_1 эквивалентно по распределению пуассоновскому полю на $R \times R_+^2$ с интенсивностью $\lambda dx dt dQ(t)$.

Фактически мы предполагаем, что в момент $t = 0$ на дороге нет препятствий. При $t \rightarrow \infty$ устанавливается стационарный режим со скоростью, зависящей от хвоста времени жизни препятствия. Впрочем можно рассмотреть поле на $R \times R$, это означало бы стационарный режим на дороге.

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).

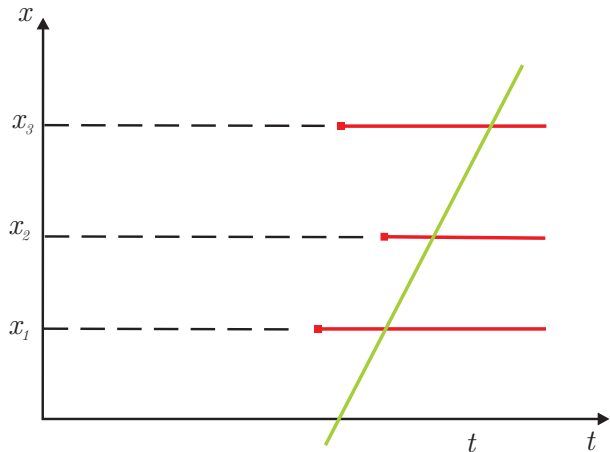


Рис. 1

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 [6].

Лемма 2. Конфигурация $\{x_i\}$ образует пуассоновский процесс интенсивности $b = \lambda m_Q$.

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.

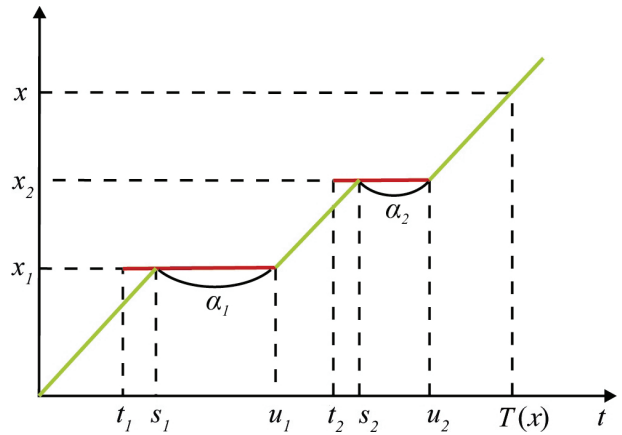


Рис. 2

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 (5).

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 вероятность 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 времени 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, то есть $\alpha_i = \min(\eta, \zeta)$. 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 объезда будет пересчитываться (увеличиваться) в зависимости от средней длины очереди перед препятствием.

II.4. 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 ей место выезда $x_{i,out}$. Но время выезда $t_{i,out}$ зависит от степени загруженности дороги. Мы определяем 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 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 $x \rightarrow \infty$

$$\frac{x}{T(x)} \rightarrow \bar{v}_1 = \frac{1 + dc}{1 + dc v_1 v_2^{-1}} v_1 \quad (8).$$

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 (6), подставляя $v = v_1 - v_2$, $b = \frac{\lambda m_G}{v_2}$, $a = \frac{c}{v_2}$:

$$\frac{1}{(v_1 - v_2)^{-1} + \frac{\lambda m c}{v_2^2}} = \frac{v_1 - v_2}{1 + dc v_1 v_2^{-1}}.$$

Тогда средняя скорость быстрого автомобиля относительно исходной системы координат составит

$$\bar{v}_1 = \frac{v_1 - v_2}{1 + dc v_1 v_2^{-1}} + v_2 = \frac{1 + dc}{1 + dc v_1 v_2^{-1}} v_1.$$

III. 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 сети Джексона, ВСМР-сети (см., например, [14, 22]). В них требование (сообщение, автомобиль, работа) обслуживается в каждом проходимом ими узле и затем выбирает случайно следующий узел. С другой стороны, сети Келли (см. [14]), где каждое требование имеет заранее фиксированный маршрут. Эти два класса сетей связаны как общей техникой, так и близостью результатов. Именно они обладают замечательным свойством мультипликативности — стационарные распределения в них имеют вид так называемой проакт-формы. Мы рассматриваем только первый класс сетей.

III.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

$$\rho P = \rho, \rho = (\rho_1, \dots, \rho_N) \iff \sum_{i=1}^N \rho_i p_{ij} = \rho_j, \quad j = 1, \dots, N \quad (9)$$

имеет единственное решение (с точностью до произвольного множителя). Нормированное решение имеет вид

$$\pi_i = \frac{\rho_i}{\sum_{i=1}^N \rho_i}, i = 1, \dots, N.$$

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 выедет 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

$$\alpha(\bar{n}, T_{i,j}\bar{n}) = \mu_i(n_i)p_{i,j} \quad (10)$$

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 (9), 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

$$\nu(n_1, \dots, n_N) = \frac{1}{Z_{N,M}} \prod_{i=1}^N \frac{\rho_i^{n_i}}{\mu_i(1)\mu_i(2)\dots\mu_i(n_i)}, \quad (11)$$

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 в уравнения Колмогорова для стационарных вероятностей, см., например, [14, 26].

III.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

$$\forall i : \sum_{j=1}^N p_{ij} \leq 1, \quad \exists i_0 : \sum_{j=1}^N p_{i_0j} < 1. \quad (12)$$

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 $S = Z_+^N$ — 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

$$\begin{aligned} \alpha(\bar{n}, T_{i,j}\bar{n}) &= \mu_i(n_i)p_{i,j}, \\ \alpha(\bar{n}, T_{i,0}\bar{n}) &= \mu_i(n_i)p_{i,0}, \\ \alpha(\bar{n}, T_i\bar{n}) &= \lambda_i \end{aligned} \quad (13)$$

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):

$$\begin{aligned} \rho = \lambda + \rho P &\iff \rho_i = \lambda_i + \sum_{k=1}^N \rho_k p_{ki}, \\ \forall i &= 1, \dots, N. \end{aligned} \quad (14)$$

В силу условия (12) и неразложимости матрицы P это уравнение имеет единственное решение, которое можно представить в виде сходящегося (см. [9]) ряда:

$$\rho = \lambda + \sum_{n=1}^{\infty} \lambda P^n.$$

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

$$r_i = \frac{\rho_i}{\mu_i}, \quad i = 1, \dots, N.$$

Следующую теорему можно найти, например, в [9, 14], она иногда называется теоремой Гордона–Ньюэлла.

Теорема 3. Марковская цепь $\eta(t)$ эргодическая тогда и только тогда, когда для всей $i = 1, \dots, N$ 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 транзитной. 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 [3]. В частности, показано, что в узлах, где нагрузка больше 1, средние очереди увеличиваются линейно с ростом времени. При этом найдены скорости роста средних очередей.

III.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

$$\rho_N = \rho_N P_N. \quad (15)$$

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 (11) 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)

$$Z_{N,M} = \sum_{n_1 + \dots + n_N = M} \prod_{i=1}^N r_{i,N}^{n_i}. \quad (16)$$

Многие важные характеристики сети выражаются через статсумму.

Нетрудно проверить, что среднее число автомобилей в i -м узле в стационарном режиме равно

$$m_{i,N,M} = E \xi_{i,N,M} = \frac{r_{i,N}}{Z_{N,M}} \frac{\partial Z_{N,M}}{\partial r_{i,N}}. \quad (17)$$

Ниже мы будем требовать слабую сходимост 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. 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

$$S_N(z) = -\lambda(1 + \varepsilon_N) \ln z - \frac{1}{N} \sum_{i=1}^N \ln(1 - z r_{i,N}), \quad (18)$$

$$S(z) = -\lambda \ln z - \int_0^1 \ln(1 - zr) dI(r),$$

где $\lambda(1 + \varepsilon_N) = \frac{M}{N}$.

Введем производящую функцию (большую статсумму):

$$\Xi_N(z) = \sum_{M=0}^{\infty} z^M Z_{N,M} = \prod_{i=1}^N \frac{1}{1 - z r_i}, \quad |z| < 1.$$

По формуле Коши и формуле (18) имеем следующее выражение для статсуммы (16):

$$Z_{N,M} = \frac{1}{2\pi i} \int_{\gamma} \frac{\Xi_N(z)}{z^{M+1}} dz = \frac{1}{2\pi i} \int_{\gamma} \frac{\exp(NS_N(z))}{z} dz, \quad (19)$$

где $\gamma = \{z \in \mathbb{C} : |z| = \sigma < 1\}$. Для средних, согласно (17), имеем

$$m_{i,N} = \frac{1}{2\pi i Z_N} \int_{\gamma} \frac{r_{i,N}}{1 - z r_{i,N}} \exp(NS_N(z)) dz. \quad (20)$$

Можно показать, что для стационарного распределения длин очередей справедлива формула

$$P_{N,M}(\xi_{1,N,M} = n_1, \dots, \xi_{K,N,M} = n_K) = \frac{1}{2\pi i Z_N} \int_{\gamma} z^{-1} \prod_{i=1}^K (1 - z r_{i,N}) (z r_{i,N})^{n_i} \exp(NS_N(z)) dz. \quad (21)$$

В доказательстве теорем этого раздела существенную роль играет метод перевала (см. [35]), точнее, его обобщение, поскольку функция в показателе экспоненты зависит от N . Из уравнения

$$\frac{\partial S_N(z)}{\partial z} = 0 \quad (22)$$

находятся точки перевала. Пусть $z_{0,N}$ — корень этого уравнения, лежащий в интервале $(0,1)$. Можно показать, что все корни уравнения (22) действительны и положительны, и всегда существует единственный корень, лежащий в интервале $(0, 1)$.

Пусть z_0 — корень уравнения

$$h(z) = \frac{\lambda}{z} \Leftrightarrow \frac{\partial S(z)}{\partial z} = 0, \quad (23)$$

лежащий в интервале $(0,1)$. Имеют место следующие факты:

- 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 (23); $z_0(\lambda)$ 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:

$$Z_N \sim \frac{\exp(NS_N(z_{0,N}))}{z_0 \sqrt{2\pi NS''(z_0)}}, F_N = \frac{1}{N} \ln Z_N \sim S(z_0).$$

- 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 $\{z \in \mathbb{C} : z_0(\lambda) - \delta_0 < |z| < z_0(\lambda) + \delta_0\}$ при некотором $\delta_0 > 0$, равномерно ограниченных в этом кольце и таких, что для заданного достаточно малого $\varepsilon > 0$ существует такое $\delta_u > 0$ и такая ненулевая действительная константа f_u , что $|f(\theta, z)/f_u - 1| < \varepsilon$ при $z \in U_{2\delta_u}(z_0)$, $\theta \in \Theta$.

Тогда при достаточно больших N равномерно по $\theta \in \Theta$:

$$\frac{1}{2\pi i} \int_{\gamma} f(\theta, z) \exp(NS_N(z)) dz = \frac{f_u \exp(NS_N(z_{0,N}))}{\sqrt{2\pi NS''(z_0)}} (1 + \zeta_N),$$

где $|\zeta_N| < 25\varepsilon$.

Доказательство этой теоремы основано на применении метода перевала (saddle-point method) (см. [35]). 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 (19) 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 применим теорему 6, получим, что для любого достаточно малого $\varepsilon > 0$ при достаточно больших N

$$Z_N = \frac{\exp(NS_N(z_{0,N}))}{z_0 \sqrt{2\pi NS''(z_0)}} (1 + \zeta_N), |\zeta_N| < 25\varepsilon. \quad (24)$$

Второй пункт теоремы 5 доказывается аналогично с использованием формулы (20) для средней очереди и формулы (21) для совместного распределения длин очередей. Третье утверждение теоремы 5 доказывается, используя теорему 6 и формулы (20), (21).

Доказательство теоремы 4. Чтобы доказать первый пункт теоремы 4, рассмотрим семейство функций

$$f(\theta, z) = \frac{A}{z} + \frac{\theta}{1 - z\theta}, \theta \in \Theta = [0,1], A > 0, f_u = \frac{A}{z_0}.$$

Зафиксируем малое $\varepsilon > 0$ и выберем $\sigma_u = \frac{\varepsilon}{8}$, $A = \frac{16z_0}{(1-z_0)\varepsilon}$. По теореме 6 имеем для достаточно

больших N и всех $\theta \in \Theta$

$$\begin{aligned} & \frac{1}{2\pi i} \int_{\gamma} \left(\frac{A}{z} + \frac{\theta}{1-z\theta} \right) \exp(NS_N(z)) dz = \\ & = \frac{A \exp(NS_N(z_{0,N}))}{z_0 \sqrt{2\pi NS''(z_0)}} (1 + \zeta_N), |\zeta_N| < 25\varepsilon. \end{aligned}$$

Разделив на Z_N и применив (24) к правой части получившегося равенства, получим при достаточно больших N :

$$\begin{aligned} A + \frac{1}{Z_N} \frac{1}{2\pi i} \int_{\gamma} \frac{\theta}{1-z\theta} \exp(NS_N(z)) dz = \\ = A(1 + \zeta'_N), |\zeta'_N| < 30\varepsilon. \end{aligned}$$

Из последнего равенства и формулы (20) 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 $m_{1,N} = m_{1,M(N),N} \geq m_{1,M'(N),N} > m$ 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 III 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 (15). Тогда относительные нагрузки определим по формуле

$$r_{i,N} = \frac{\mu_{0,N} \rho_{i,N}}{\rho_{0,N} \mu_{i,N}},$$

так что $r_{0,N} = 1$. Согласно (11), стационарное распределение длин очередей имеет вид

$$P_{N,M}(\xi_{i,N,M} = n_i, i = 1, \dots, N) =$$

$$= \frac{1}{\hat{Z}_{N,M}} \frac{1}{(M - \sum_{i=1}^M n_i)!} \prod_{i=1}^N r_{i,N}^{n_i},$$

где

$$\hat{Z}_{N,M} = \sum_{n_1 + \dots + n_N \leq M} \frac{1}{(M - \sum_{i=1}^M n_i)!} \prod_{i=1}^N r_{i,N}^{n_i},$$

и большая статсумма сети равна

$$\hat{\Xi}_N(z) = e^z \prod_{i=1}^N \frac{1}{1 - zr_{i,N}}.$$

Положим

$$q_{i,N} = \frac{r_{i,N}}{p_N}, w = zp_N, \quad p_N = \max_{1 \leq i \leq N} r_{i,N}.$$

Тогда

$$\hat{\Xi}_N(w) = e^{w/p_N} \prod_{i=1}^N \frac{1}{1 - wq_{i,N}}.$$

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 [10] для замкнутых сетей аналогичные результаты получаются для случая более общей зависимости интенсивностей от длин очередей в узлах.

Прием усложнения графа позволяет устранить 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 возникла хотя бы одна пробка. Интереснее рассмотреть ситуацию, когда в разных местах графа одновременно возникает много пробок.

Рассмотренные в этом разделе модели удобны также тем, что все их 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 она изучалась в жидкостных моделях, см. [21].

Литература

1. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // *J. Stat. Phys* — 2003. — V. 111. — P. 903–930.
2. *Blythe R.A., Evans M.R.* Nonequilibrium steady states of matrix-product // *Journal of Physics A: Mathematical and Theoretical*. — 2007. — V. 40, N. 46.
3. *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.
4. *Buslaev A.P., Prikhodko V.M., Tatashev A.G., Yashina M.V.* The deterministic – stochastic flow model. — e-print arXiv: physics /0504139 v1, 2005.
5. *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. — V. 2007.
6. *Cox D.R., Isham V.* Point processes. — Chapman and Hall, 1980.
7. *Daley D., Vere-Jones D.* An Introduction to the Theory of Point Processes. — Springer, 2003. — V. 1.
8. *Derrida B.* Non-equilibrium steady states: fluctuations and large deviations // *Journal of Statistical Mechanics: Theory and Experiment*. — 2007.
9. *Fayolle G., Malyshev V., Menshikov M.* Topics in the constructive theory of countable Markov chains. — Cambridge University Press, 2008. — 2nd edition.
10. *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.
11. *Helbing D.* Verkehrsdynamik. — Berlin: Springer, 1997.
12. *Helbing D.* Traffic and related self-driven many particle systems // *Rev. Mod. Phys.* 73. — 2001. — P. 1067–1141.
13. *Jost D., Nagel K.* Probabilistic Traffic flow breakdown in stochastic car following models // *Traffic and Granular Flow* — 2005. — Part 2. — P. 87–103.
14. *Kelly F.* Reversibility and stochastic networks. — N.-Y.: Wiley, 1979.
15. *Kerner B.S.* Introduction to modern traffic flow theory and control. — Springer, 2009.
16. *Lighthill M.J., Whitham G.B.* On kinematic waves: II. Theory of traffic flow on long crowded roads // *Proc. R. Soc.* — London, 1955. — V. 229. — P. 281–345.
17. *Lotito P., Mancinelli E., Quadrat J.-P.* A min-plus derivation of the fundamental car-traffic law // *Automatic Control IEEE Transactions*. — 2005. — V. 50. — P. 699–705.
18. *Malyshev V., Yakovlev A.* Condensation in large closed Jackson networks // *Ann. Appl. Probab.* — 1996. — V. 6, N. 1. — P. 92–115.
19. *Prigogine I., Herman R.* Kinetic theory of vehicular traffic. — N.-Y.: Elsevier, 1971.
20. *Renyi A.* On two mathematical models of the traffic on a divided highway // *Journal of Applied Probability*. — 1964. — V. 1. — P. 311–320.
21. *Gartner N.H., Messer C.J., Rathi A.K.* Revised Traffic Flow Theory. A state-of-art report. — Washington DC: Transportation Research Board, 2001.
22. *Serfozo R.* Introduction to stochastic networks. — Springer, 1999.
23. *Solomon H.* Geometric Probability. — SIAM, Philadelphia, 1978.
24. *Solomon H., Wang P.* Nonhomogeneous Poisson fields of random lines with applications to traffic flow // *Proc. Sixth Berkeley Symp. On Math. Statist. and Prob.* — Univ. of Calif. Press, 1972. — V. 3. — P. 383–400.
25. *Gartner N.H., Messer C.J., Rathi A.K.* Traffic flow theory: A state-of-the-art report. — Washington DC: Transportation Research Board, 2001.
26. *Афанасьева Л.Г.* Очерк исследования операций. — М.: Изд-во МГУ, 2007.

- 27.** Буслаев А.П. Вероятностные и имитационные подходы к оптимизации автодорожного движения. — М.: Мир, 2003.
- 28.** Замятин А.А., Малышев В.А. Накопление на границе для одномерной стохастической системы частиц // Проблемы передачи информации. — 2007. — Т. 43, вып. 4. — С. 68–82.
- 29.** Иносэ Х., Хамада Т. Управление дорожным движением. — М.: Транспорт, 1983.
- 30.** Кингман Д. Пуассоновские процессы. — М.: МЦНМО, 2007.
- 31.** Кокс Д., Смит В. Теория восстановления. — М., 1967.
- 32.** Малышев В.А., Минлос Р.А. Гиббсовские случайные поля. — М.: Наука, 1985.
- 33.** Малышев В.А. Случайные грамматики. — Успехи Мат. Наук, 1998. — Т. 53, вып. 2. — С. 107–134.
- 34.** Рюэль Д. Статистическая механика. Строгие результаты. — М.: Мир, 1971.
- 35.** Федорюк М.В. Метод перевала. — М.: Наука, 1977.
- 36.** Хейт Ф. Математическая теория транспортных потоков. — М., 1966.
- 37.** Четверушкин Б.Н., Трапезникова М.А., Фурманов И.Р., Чубарова Н.Г. Макро- и микроскопические модели для описания движения автотранспорта на многополосных магистралях // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков, под ред. акад. В.В. Козлова). — 2010. — Т. 2, № 4(8).

Поступила в редакцию 15.10.2010.

С.Л. Кленов

Московский физико-технический институт (государственный университет)

Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий

Кратко рассматриваются основные положения и результаты предложенной Кернером теории трех фаз в транспортном потоке. Эта теория позволяет объяснить и предсказать эмпирические пространственно-временные свойства перехода к плотному потоку и результирующих структур в плотном транспортном потоке. Приводятся результаты измерения пространственно-временных структур в транспортном потоке на скоростных автомагистралях в Германии, Англии и США. Эти результаты являются эмпирическим базисом как для определения фаз транспортного потока, так и для остальных положений теории Кернера. Показывается, что в теории Кернера возникновение плотного транспортного потока (traffic breakdown) объясняется как фазовый переход первого рода от фазы свободного потока к фазе синхронизованного потока. Такой переход с некоторой вероятностью может происходить в широком диапазоне значений транспортного потока, что отвечает бесконечному числу значений пропускной способности автодороги. Анализируются характерные свойства процесса образования фазы движущихся кластеров машин (движущихся заторов) в синхронизованном потоке, дается классификация пространственно-временных структур транспортного потока в теории Кернера. Приводятся примеры новых интеллектуальных транспортных технологий, основанные на теории трех фаз Кернера.

Ключевые слова: транспортные потоки, теория трех фаз Кернера в транспортном потоке, интеллектуальные транспортные технологии, переход от свободного к плотному транспортному потоку, бесконечное число значений пропускной способности скоростных автомагистралей.

I. Введение. Фундаментальные эмпирические свойства фазовых переходов в транспортном потоке

В 1996–2002 годах 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 Кернера [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.

Поэтому 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 Кернера [1]. В на-

стоящей статье кратко излагаются основные положения теории трех фаз Кернера в соответствии с книгами [1, 2].

Цель данной статьи состоит в том, чтобы дать представление об эмпирическом базисе и основных идеях теории трех фаз Кернера. Это означает, что статья не может заменить книг Кернера, чтение которых необходимо тем, кто хочет разобраться в этой теории. По этой причине основное внимание уделяется не математическому обоснованию тех или иных положений теории трех фаз Кернера, а качественному описанию. Единственное исключение составляет раздел XII, 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 подчеркнуть фундаментальные достижения предыдущих моделей транспортного потока по математической формулировке многочисленных эффектов взаимодействия между водителями [3–25]. К ним относятся, в первую очередь, введенная в моделях Дженера Моторс задержка водителей [10–13], приводящая к переторможению как реакции на замедление машины впереди [13, 19, 20], формулировка задержки водителей через модельные флуктуации, введенные в моделях Нагеля, Шрекенбер-

га и Шашнайдера (см. обзоры [19–22]), микроскопическое описание *slow-to-start rule* в работах группы Шрекенберга (см. обзоры [19–22]). Эти и другие математические формулировки в поведении водителя являются также важными элементами математических моделей трех фаз (см. главу 11 книги Кернера [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, наблюдаемых в эмпирических данных. Детальное объяснение этого «парадокса» дано в главе 10 книги Кернера [1]. Этот парадокс объясняется очень просто: анализ эмпирических данных, который позволил выявить фундаментальные эмпирические свойства перехода от свободного к плотному потоку 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, обнаруженные Кернером, детально описаны в разделе 2.4 и части II первой книги Кернера [2]. В рамках данной статьи нет возможности останавливаться на рассмотрении всех эмпирических свойств транспортного потока. Тем не менее сформулируем сначала фундаментальные эмпирические свойства перехода к плотному транспортному потоку и эмпирические свойства фазовых переходов в транспортном потоке.

Фундаментальные эмпирические свойства перехода к плотному транспортному потоку следующие:

1. Переход к плотному транспортному потоку (*traffic breakdown*) является $F \rightarrow S$ переходом (буква F соответствует «free flow», то есть 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 машин.

3. Может быть как спонтанный, так и индуцированный $F \rightarrow S$ переход около одного и того же узкого места на дороге (*bottleneck*).

В главе 10 своей последней книги [1] Кернер показывает, что теория Лайтхилла–Уизема–Ричард-

са (ЛВР) кинематических и ударных волн в транспортном потоке [3–5], дискретной версией которой является *sell transmission* модель Даганцо [9], не может описывать пункты 2 и 3, а модели, относящиеся к классу Джeneralал Моторс (см. главу 10 в [1] и обзоры [19–22]), не могут описывать пункты 1–3.

Фундаментальные эмпирические свойства фазовых переходов в транспортном потоке следующие:

а) в соответствии с пунктами 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 машин (*wide moving jams*, обозначается ниже буквой J) возникают спонтанно только в синхронизованном потоке, то есть 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.

ЛВР теория кинематических волн не может описывать пункты б) и в), а теория, основанная на моделях класса Джeneralал Моторс, не может описывать пункты а)–в).

II. Три фазы транспортного потока

Теория трех фаз, основанная Кернером, фокусируется главным образом на физике плотного транспортного потока на скоростных автомагистралях. Кернер 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). Согласно Кернеру в плотном потоке выделяются две фазы, соответственно существуют три фазы транспортного потока:

- 1) свободный поток — фаза F ,
- 2) синхронизованный 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 Кернера разделение на свободный и плотный поток точно такое же, как и в классических теориях Лайтхилла–Уизема и Джeneralал Моторс. Фундаментальное отличие теории 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 разделах II–XI, основаны исключительно на эмпирических данных.

III. 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 машин в единицу времени (проходящих через данную точку дороги), и плотностью ρ , измеряемой в количестве машин 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 $\rho = \rho_{\text{crit}}$ (рис. 1), которые могут быть достигнуты в свободном потоке.

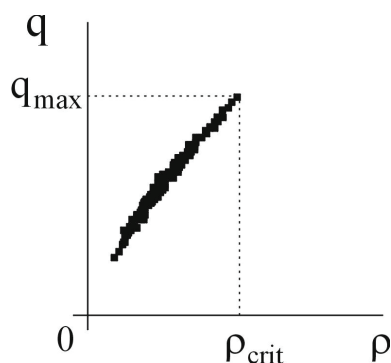


Рис. 1. Величина потока q в зависимости от плотности машин ρ в свободном потоке [14, 20]

IV. Плотный транспортный поток

В плотном транспортном потоке, который определяется так же, как и в классических теориях Лайтхилла–Уизема и Джеренал Моторс, скорость машин меньше, чем минимально возможная скорость машин в свободном потоке. Это означает, что прямая с наклоном, равным минимальной скорости в свободном 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.

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 транс-

портному потоку, в дальнейшем будет называться узким местом, или «бутылочным горлом» [14, 20].

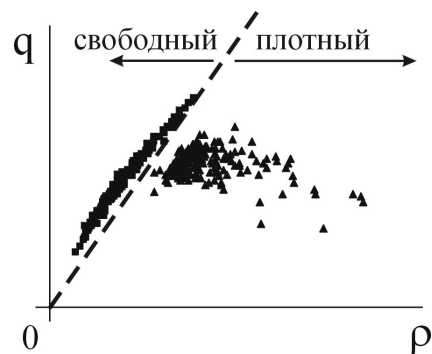


Рис. 2. Величина потока q в зависимости от плотности машин ρ в свободном и плотном потоке [14, 20]

V. Определение фаз J и S в плотном транспортном потоке

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 затор) — «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. Кернер определил фазы J и S следующим образом.

Определение [J] фазы широкого движущегося кластера:

Задний по направлению движения фронт широкого движущегося кластера (локального движущегося затора), где машины, выезжающие из кластера, ускоряются вплоть до свободного или до синхронизованного 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 [S] фазы синхронизованного потока:

Задний по направлению движения фронт области синхронизованного потока, где машины ускоряются вплоть до свободного потока, НЕ обладает характеристическим свойством широкого движущегося кластера. В частности, задний фронт синхронизованного 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, то есть исходно не имеет никакого отношения к какой-либо математике. Теоретический смысл этих определений можно понять, прочтя раздел 6.1 книги Кернера [1].

Данные измерений средней скорости машин (рис. 3а) иллюстрируют определения фаз J и S . На рис. 3а имеются две пространственно-временные структуры плотного потока с низкой скоростью машин. 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 машин с автомагистрали. Согласно определению [S], эта область плотного потока относится к фазе «синхронизованного потока» (рис. 3а и 3б).

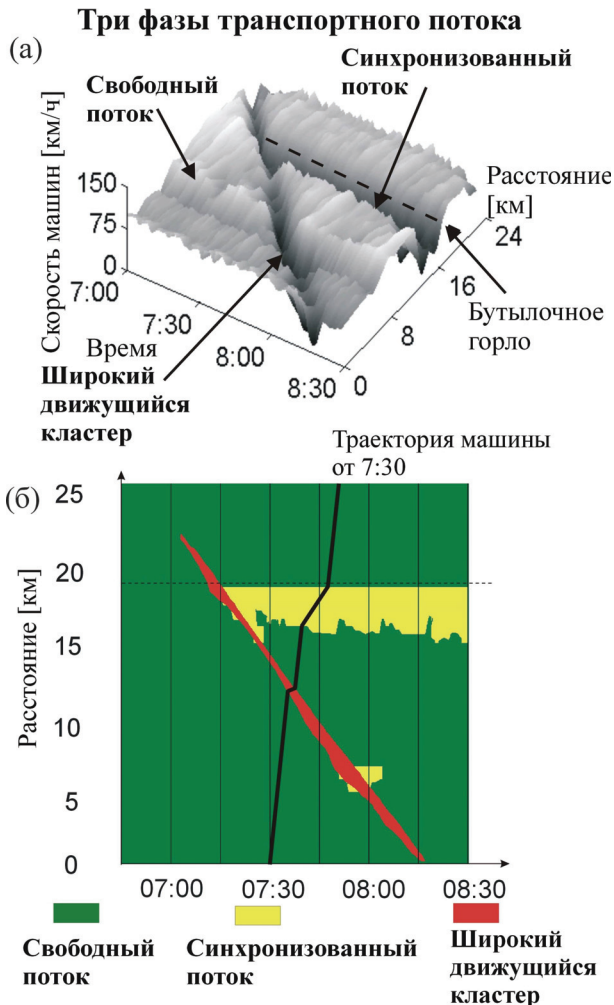


Рис. 3. Данные измерений скорости машин в пространстве и времени (а) и их представление на координатно-временной плоскости (б). Взято из книги Кернера [1]

VI. Возникновение плотного потока (traffic breakdown) — $F \rightarrow S$ фазовый переход

Переход от свободного к плотному потоку в англоязычной литературе известен как traffic breakdown. В теории трех фаз Кернера такой переход объясняется возникновением фазы синхронизованного потока, то есть $F \rightarrow S$ 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 синхронизованного потока.

Возникновение плотного потока ($F \rightarrow S$ переход)

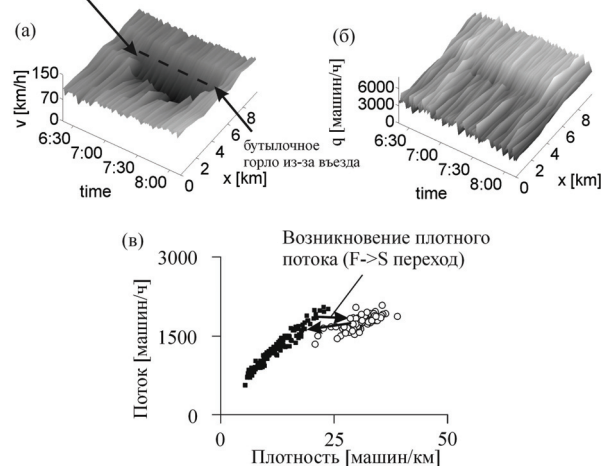


Рис. 4. Эмпирический пример возникновения плотного потока и эффект гистерезиса у бутылочного горла из-за въезда на автодорогу. (а, б) Средняя скорость (а) и поток (б) на автодороге в пространстве и времени (увеличение потока после въезда на (б) связано с потоком въезжающих на дорогу машин). (в) Эффект гистерезиса в плоскости поток–плотность, обозначенный двумя стрелками, показывающими переход к плотному потоку и обратный переход к свободному потоку. 1-минутные данные. Взято из книги Кернера [1]

В самом деле, типичный пример перехода из свободного в синхронизованный поток вблизи въезда машин показан на рис. 4. Из рисунка видно, что в то время как скорость машин резко уменьшается в процессе перехода (рис. 4а), поток меняется мало (рис. 4б). Скачок скорости при мало меняющемся потоке особенно наглядно виден на рис. 4в. В течение всего времени после перехода задний фронт между плотным и свободным потоками фиксирован вблизи въезда на дорогу. По этой причине плотный поток соответствует определению фазы синхронизованного потока, поэтому весь плотный поток относится к фазе синхронизованного потока. Образование плотного потока примерно в 6:30 мин (показанное стрелкой

слева направо на рис. 4в) и его исчезновение примерно в 7:45 (показанное стрелкой справа налево на рис. 4в) сопровождается гистерезисом, хорошо известным в теории фазовых переходов I 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 транспортном потоке образуются широкие движущиеся кластеры (рис. 5б, координата дороги $x = 0$ км). Видно, что в результате перехода к плотному потоку на узком месте, связанным с въездом на скоростную магистраль, сначала образуется фаза S синхронизованного потока. Действительно, в течение всего времени существования плотного потока на этом узком месте задний фронт плотного потока, на котором машины ускоряются из плотного потока до свободного, фиксирован на данном месте. Поэтому по определению фаз в теории 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 $F \rightarrow S$ фазовый переход будет рассмотрен ниже, в разделе X.

Таким образом, переход от свободного к плотному потоку в эмпирических данных является $F \rightarrow S$ переход первого рода. Это эмпирическое свойство есть общее свойство реальных транспортных потоков на скоростных магистралях. Напротив, в моделях класса Джеренал Моторс, как объяснено в книге Кернера [1], переход от свободного к плотному потоку связан с возникновением широких движущихся кластеров.

Используя данные измерений, Кернер сделал 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 пото-

ка 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.

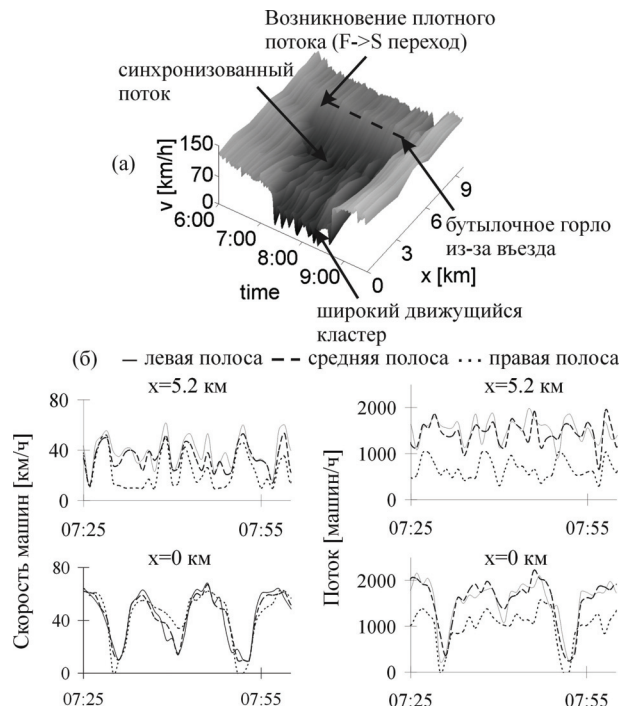


Рис. 5. Эмпирический пример возникновения широких движущихся кластеров в синхронизованном потоке: (а) Скорость машин в пространстве и времени; (б) Скорость (слева) и поток (справа) на трех полосах дороги в области синхронизованного потока ($x = 5.2$ км) и в области широких движущихся кластеров ($x = 0$ км). Взято из книги Кернера [1]

Кернер объяснил природу $F \rightarrow S$ фазового перехода с помощью «соревнования» во времени и пространстве двух противоположных процессов: ускорения машины при обгоне более медленной машины впереди, названном «переускорением», и в случае, когда обгон невозможен, торможения машины до скорости более медленной машины, названном «адаптация скорости». «Переускорение» поддерживает дальнейшее существование свободного потока. Напротив, «адаптация скорости» ведет к синхронизованному потоку. Кернер постулировал, что вероятность обгона, которая совпадает с вероятностью «переускорения», является разрывной функцией плотности (рис. 6): при данной плотности машин вероятность обгона в свободном потоке много больше, чем в синхронизованном потоке.

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. $F \rightarrow S$ пере-

ход 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).

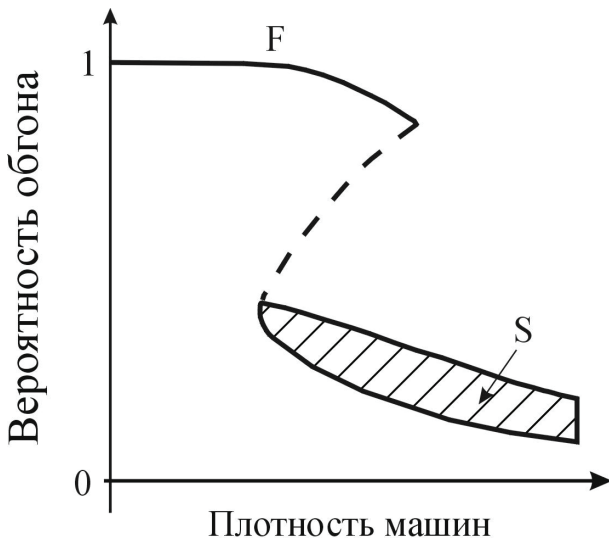


Рис. 6. Объяснение фазового перехода к плотному потоку (traffic breakdown) на основе Z-образной нелинейной функции вероятности обгона (вероятности «переускорения») в теории Кернера. Пунктирная линия описывает критическое значение вероятности обгона как функцию плотности машин. Взято из книги Кернера [1]

Отметим, что $F \rightarrow S$ фазовый переход и обратный $S \rightarrow F$ фазовый переход сопровождаются гистерезисом. Однако данный гистерезис не имеет никакого отношения к хорошо известному гистерезису в математических моделях класса Джеренал-Моторс, впервые найденному в теории Кернера-Конхойзера [26, 27]. Этот известный гистерезис, описанный в огромном количестве математических работ (см. работы [19–22, 28, 29] 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.

VII. Бесконечное число пропускных способностей скоростной автомагистрали

Спонтанное образование плотного потока, то есть спонтанный $F \rightarrow S$ фазовый переход, может произойти в широком диапазоне значений величины потока q в свободном транспортном потоке. Основываясь на эмпирических данных измерений, Кернер сделал 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 (рис. 7).

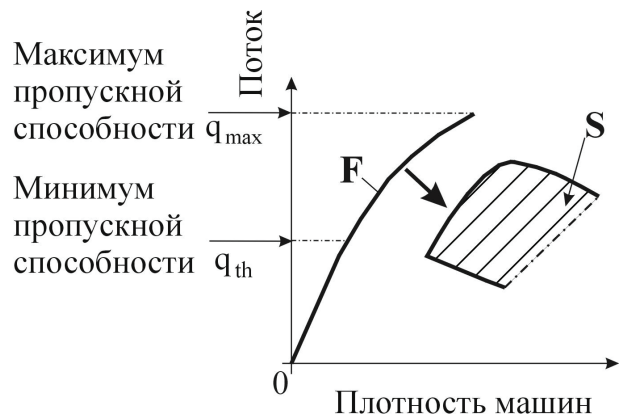


Рис. 7. Максимум и минимум пропускной способности скоростной автомагистрали в теории трех фаз Кернера. Взято из книги Кернера [1]

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.

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 $F \rightarrow S$ фазового перехода.

Если величина потока q меньше чем минимальная пропускная способность q_{th} , то возникновение плотного потока ($F \rightarrow S$ 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, то есть является устойчивым относительно малых возмущений. Однако для больших возмущений свободный поток оказывается неустойчивым, и происходит $F \rightarrow S$ фазовый переход к синхронизованному потоку.

Как уже упоминалось в разделе VI, Кернер объяснил природу $F \rightarrow S$ фазового перехода с помощью «соревнования» во времени и пространстве двух противоположных процессов: ускорения машины при обгоне более медленной машины впереди, названном «переускорением», и в слу-

чае, когда обгон невозможен, торможения машины до скорости более медленной машины, названном «адаптация скорости». На рис. 8 поясняется такое соревнование более детально. На рис. 8а, который соответствует рис. 6, стрелочка вниз означает, что если в свободном потоке вблизи узкого места возникает локальное уменьшение скорости машин, то вероятность обгона внутри этого возмущения падает. Если это уменьшение вероятности обгона становится меньше, чем критическая величина вероятности обгона, показанная пунктирной линией на рис. 8а, то $F \rightarrow S$ фазовый переход происходит внутри возмущения; в противоположном случае возмущение затухает и свободный поток остается на узком месте. Стрелочка вверх на рис. 8а означает, что если исходное состояние отвечает фазе синхронизованного потока и в этом состоянии возникает случайное локальное увеличение скорости машин, то внутри возмущения вероятность обгона возрастает. Если возрастание вероятности обгона превышает критическое значение (как отмечалось выше, эта критическая вероятность обгона отвечает пунктирной кривой на рис. 8а), то в области возмущения происходит $S \rightarrow F$ переход; в противоположном случае возмущение затухает, и синхронизованный поток остается.

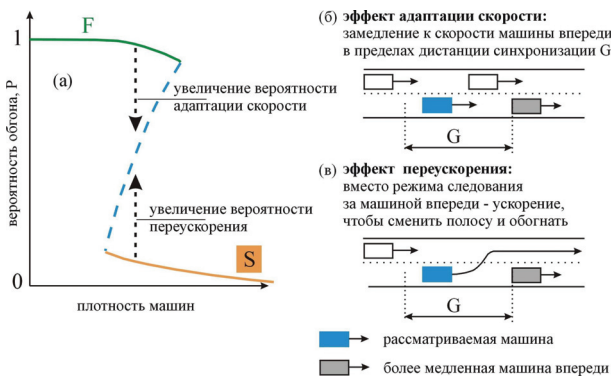


Рис. 8. Пояснение соревнования между переускорением и адаптацией скорости, которое объясняет бесконечное число значений пропускной способности автомагистрали. Взято из книги Кернера [1]

Бесконечное число значений пропускной способности автомагистрали вблизи узкого места в теории трех фаз Кернера фундаментально противоречит классическим теориям транспортного потока и методам управления и автоматического регулирования транспортных потоков, которые предполагают существование в любой момент времени некоторой (фиксированной или случайной) пропускной способности.

VIII. 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 машин внутри широкого движущегося кластера много меньше, чем скорость машин в свободном потоке. На заднем (в направлении потока) фронте кластера машины могут ускоряться вплоть до свободного потока. На переднем фронте кластера машины, подъезжающие к фронту, должны сильно уменьшать свою скорость. Согласно определению фазы 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.

Эмпирические результаты показывают (см. часть II в книге [2] и ссылки в ней), что характеристические параметры широких движущихся кластеров не зависят от величины потока на дороге и особенностей узкого места (где и когда кластер возник). Однако характеристические параметры зависят от погоды, дорожных условий, конструктивных характеристик машин, 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). Это означает, что разные широкие движущиеся кластеры имеют одинаковые параметры при одинаковых условиях. В результате данные параметры могут быть предсказаны. Движение заднего фронта широкого движущегося кластера может быть показано на плоскости поток — плотность с помощью прямой, называемой линией 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 машин ρ_{max} в широком движущемся кластере [1].

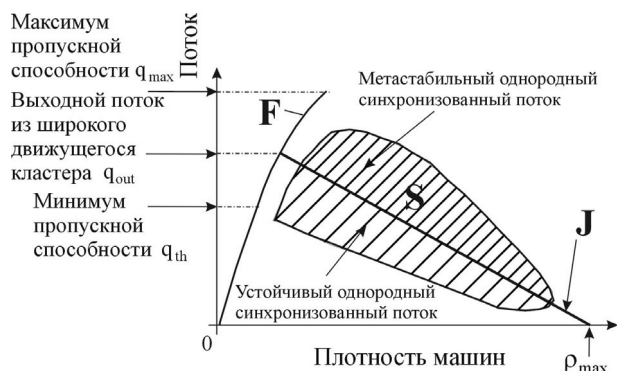


Рис. 9. Три фазы транспортного потока в плоскости поток–плотность в теории трех фаз Кернера. Взято из книги Кернера [1]

Кернер подчеркивает, 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 $F \rightarrow S$ фазовый переход вблизи узкого места, то есть возникновение плотного потока (traffic breakdown). В свою очередь величина выходного потока из широкого движущегося кластера q_{out} характеризует условия существования таких кластеров, то есть 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 возможен $F \rightarrow S$ 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 максимально возможный поток q_{max} в свободном потоке перед кластером. Это означает, что в свободном потоке водители могут выбирать более короткую временную дистанцию до машины впереди, чем та дистанция, которую они принимают, ускоряясь на заднем фронте широкого движущегося кластера.

IX. Синхронизированный транспортный поток — фаза S

В отличие от широких движущихся кластеров, в синхронизированном потоке как величина потока q , так и скорость машин могут меняться заметным образом. Задний по направлению потока фронт синхронизированного потока часто фиксирован в пространстве (см. определение [S]), обычно вблизи расположения узкого места. Величина потока q в фазе синхронизированного потока может оставаться почти такой же, как и в свободном потоке, даже если скорость машин сильно уменьшается.

Синхронизированный поток не имеет характеристического свойства [J] фазы широкого движущегося кластера J . В теории трех фаз Кернера предполагается, что гипотетические однородные состояния синхронизированного потока покрывают двумерную область в плоскости поток–плотность (см. заштрихованную область на рис. 9).

X. $S \rightarrow J$ фазовый переход

Широкие движущиеся кластеры не возникают в свободном потоке, но они могут возникать в области синхронизированного потока. Это называется $S \rightarrow J$ фазовым переходом. Эмпирический пример $S \rightarrow J$ перехода показан на рис. 5. Таким образом, образование широких движущихся кластеров в свободном потоке наблюдается в результате каскада $F \rightarrow S \rightarrow J$ фазовых переходов: сначала область синхронизированного потока возникает внутри свободного потока. Как было объяснено выше, такой $F \rightarrow S$ фазовый переход происходит в большинстве случаев вблизи узкого места. Далее внутри синхронизированного потока происходит

«сжатие» потока, то есть плотность машин возрастает, в то время как их скорость падает. Такое сжатие называется «пинч»-эффект. В области синхронизированного потока, где происходит пинч-эффект, возникают узкие движущиеся кластеры. Кернер показал, что частота возникновения узких движущихся кластеров тем выше, чем выше плотность в синхронизированном 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 Кернера линия J делит все однородные состояния синхронизированного потока на две области (рис. 9). Состояния выше линии J являются метастабильными относительно образования широких движущихся кластеров, в то время как состояния ниже линии J 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.

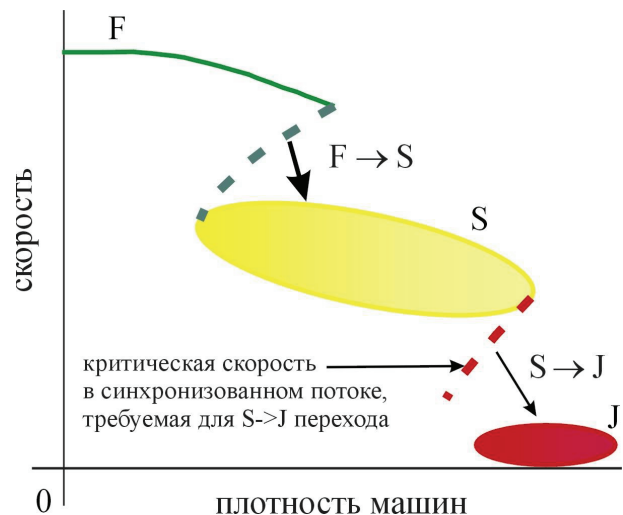


Рис. 10. Двойная Z -характеристика в теории трех фаз Кернера, поясняющая каскад $F \rightarrow S \rightarrow J$ фазовых переходов. Взято из книги Кернера [1]

Каскад $F \rightarrow S \rightarrow J$ фазовых переходов можно пояснить на основе двойной Z характеристики в теории трех фаз Кернера (рис. 10). Пунктирная линия между фазами F и 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 фазами F и S на рис. 10).

Пунктирная линия между фазами S и J качественно соответствует критической скорости внутри локального возмущения синхронизованного потока, при которой происходит $S \rightarrow J$ фазовый переход. Другими словами, $S \rightarrow J$ фазовый переход происходит внутри локального возмущения синхронизованного потока, в котором скорость меньше чем критическая скорость (символически переход показан стрелкой между фазами S и J на рис. 10).

XI. 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 синхронизованного потока (СП). Когда задний фронт СП фиксирован вблизи узкого места на дороге, а передний фронт не распространяется против потока, такая СП называется локализованной структурой синхронизованного потока (ЛСП). Однако во многих случаях передний фронт структуры синхронизованного потока распространяется в направлении против потока. Если при этом задний фронт по-прежнему остается фиксированным вблизи узкого места, то ширина области синхронизованного потока увеличивается. Такая структура называется расширяющейся структурой синхронизованного потока (РСП). Возможна также ситуация, когда задний фронт синхронизованного потока уже не фиксирован вблизи узкого места, а оба фронта синхронизованного потока движутся в направлении против потока. Такая структура называется бегущей, или мигрирующей структурой синхронизованного потока (МСП).

Разница между пространственно-временными структурами, состоящими из только синхронизованного потока, и широкими движущимися кластерами становится особенно ясной, когда РСП или МСП достигают следующего узкого места, расположенного вверх по течению транспортного потока. В таком случае структура синхронизованного 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, то есть пробегая через узкое место на дороге. Кроме того, в отличие от широкого движущегося кластера, структура

синхронизованного 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 может быть разной у разных МСП. Данные особенности структур синхронизованного 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 синхронизованный поток охватывает два и более узких места, называется единой структурой плотного потока (ЕП). ЕП может состоять только из синхронизованного 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) (рис. 11).

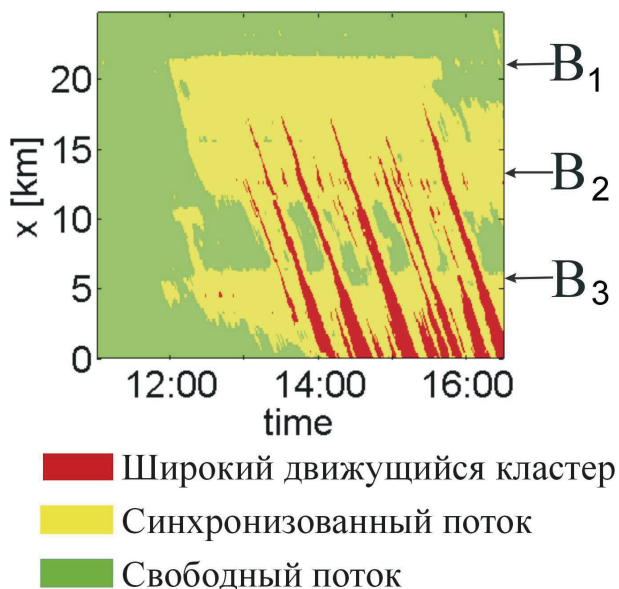


Рис. 11. ЕОП, измеренная на магистрали с тремя узкими местами B_1 , B_2 и B_3 . Взято из книги Кернера [1]

В данном разделе были рассмотрены основные качественные положения теории трех фаз Кернера, которые 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 в книге [1]). В следующем разделе дается краткий обзор стохастической микроскопической модели в рамках теории трех фаз Кернера и приведены некоторые результаты численных расчетов.

ХИ. 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 Кернера. Впервые микроскопическая трехфазная модель, которая может воспроизводить эмпирические свойства перехода к плотному потоку (traffic breakdown) и результирующих пространственно-временных структур, была разработана Кернером и Кленовым в 2002 году [30]. Несколькими месяцами позже Кернер, Кленов и Вольф предложили трехфазную модель на основе клеточных автоматов (ККВ модель) [31]. Позднее были также разработаны другие модели транспортного потока в рамках теории трех фаз Кернера: Дэвис [32], а также Кернер и Кленов [33] предложили детерминистические микроскопические модели трех фаз; Ли и Шрекенберг с соавторами [34], Янг и Ву [35], и Гао с соавторами [36] разработали различные трехфазные модели клеточных автоматов (КА); Лаваль [37] и Хугендорн с соавторами [38] разработали макроскопические модели трех фаз. Последние результаты моделирования транспортного потока, проведенные различными научными группами в США, Германии, Голландии, Китае, Южной Кореи и Японии в рамках теории трех фаз Кернера, можно найти в [39–62].

В данном разделе кратко рассматривается стохастическая микроскопическая трехфазная модель транспортного потока, предложенная Кернером и Кленовым в [30, 63]. В этой модели [63] использовалось дискретное время с шагом τ , в то время как перемещение в пространстве предполагалось непрерывным. Ниже будет рассмотрена дискретная версия модели [63], сформулированная в [60], в которой используется достаточно мелкая дискретизация пространства с шагом δx (см. также формулировку дискретной версии модели в [61]). При этом в приведенных ниже формулах координата измеряется в единицах δx , в то время как скорость машины и ее ускорение измеряются соответственно в единицах $\delta v = \delta x/\tau$ и $\delta a = \delta v/\tau$, где временной шаг $\tau = 1$ сек.

Уравнения движения машины в дискретной версии [60, 61] стохастической трехфазной модели [63] транспортного потока на скоростной 2-полосной автодороге в приближении идентичных машин задаются следующими формулами:

$$v_{n+1} = \max(0, \min(v_{free}, \tilde{v}_{n+1} + \xi_n, v_n + a\tau, v_{s,n})),$$

$$x_{n+1} = x_n + v_{n+1}\tau, \quad (1)$$

$$\tilde{v}_{n+1} = \max(0, \min(v_{free}, v_{s,n}, v_{c,n})), \quad (2)$$

$$v_{c,n} = \begin{cases} v_n + \Delta_n, & \text{если } g_n \leq G_n, \\ v_n + a_n\tau, & \text{если } g_n > G_n, \end{cases} \quad (3)$$

$$\Delta_n = \max(-b_n\tau, \min(a_n\tau, v_{\ell,n} - v_n)), \quad (4)$$

где индекс n отвечает дискретному времени $t = n\tau$, $n = 0, 1, 2, \dots$; x_n и v_n — координата и скорость машины на временном шаге n ; v_{free} — максимальная скорость машины в свободном потоке; $g_n = x_{\ell,n} - x_n - d$ — расстояние до машины впереди, индекс ℓ относится ко всем переменным и функциям, описывающим машину впереди, d — длина машины, которая предполагается одинаковой и включает в себя также среднее расстояние между машинами, когда они стоят внутри широкого движущегося кластера; \tilde{v}_n — величина скорости машины без шумовой компоненты ξ_n ; $v_{s,n}$ — безопасная скорость; $a_n \geq 0$ и $b_n \geq 0$; a — максимальное ускорение; G_n — максимальное расстояние, на котором водитель синхронизирует свою скорость со скоростью машины впереди (так называемая дистанция «синхронизации скорости»):

$$G_n = G(v_n, v_{\ell,n}), \quad (5)$$

где

$$G(u, w) = \max(0, [k\tau u + a^{-1}u(u - w)]), \quad (6)$$

где $k > 1$ — константа, $[z]$ означает целую часть действительного числа z .

Решения модели (1)–(6) при стационарном и однородном движении машин (которое отвечает потоку, где машины находятся на одинаковом расстоянии друг от друга и движутся с постоянной скоростью) показаны на рис. 12а и 12б. В соответствии с фундаментальной гипотезой теории трех фаз Кернера (рис. 9) фаза синхронизованного потока (буква S) как на плоскости расстояние–скорость (рис. 12а), так и на плоскости поток–плотность (рис. 12б) покрывает двумерную область (заштрихованная область на рис. 12). Для того чтобы пояснить смысл этой двумерной области для стационарных состояний синхронизованного потока, рассмотрим более подробно рисунок 12а. Видно, что при некоторой заданной величине скорости $v < v_{free}$ в стационарном состоянии существует бесконечное количество расстояний между машинами в диапазоне

$$g_{safe} \leq g \leq G,$$

где g — расстояние между машинами, $G = G(v, v)$ — дистанция синхронизации скорости (5), взятая при одинаковых скоростях машин, g_{safe} — безопасное расстояние между машинами, которое является решением уравнения $v = v_s(g_{safe})$, где безопасная скорость v_s в однородном стационарном состоянии определяется как $v_s(g) = g/\tau$. Таким образом, стационарные однородные состояния стохастической трехфазной модели (рис. 12) принципиально отличаются от таких же состояний всех предшествующих моделей, в которых заданной скорости отвечает одно единственное расстояние между машинами, соответствующее точке на фундаментальной диаграмме.

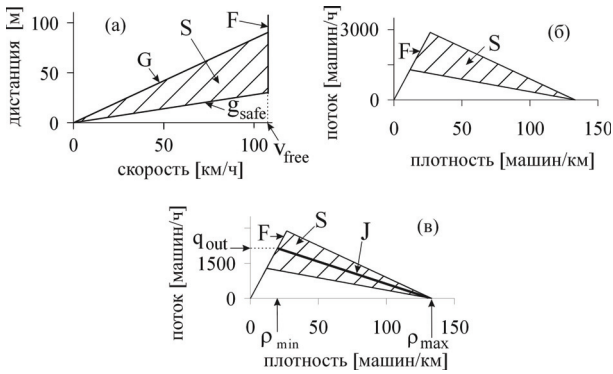


Рис. 12. Стационарные однородные состояния и линия J Кернера в стохастической трехфазной модели: (а, б) свободный поток (F) и синхронизованный поток (затрихованная двумерная область, обозначенная буквой S) на плоскости расстояние между машинами — скорость (а) и на плоскости поток-плотность (б). На (в) показано соответствие между линией J и однородными стационарными состояниями модели на плоскости поток-плотность

Из рис. 12в видно, 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, разделяет двумерную область стационарных однородных состояний синхронизованного потока на две части. Оказывается, что в соответствии с теорией трех фаз Кернера стационарные однородные состояния выше линии J являются метастабильными к образованию широких движущихся кластеров, а состояния ниже линии J — 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 (3) и (4) задаются как случайные функции:

$$a_n = a\theta(P_0 - r_1),$$

$$b_n = a\theta(P_1 - r_1),$$

$$P_0 = \begin{cases} p_0, & \text{если } S_n \neq 1, \\ 1, & \text{если } S_n = 1, \end{cases}$$

$$P_1 = \begin{cases} p_1, & \text{если } S_n \neq -1, \\ p_2, & \text{если } S_n = -1, \end{cases}$$

где величины $1 - P_0$ и $1 - P_1$ представляют собой вероятности для времени задержки водителя соответственно при ускорении и при замедлении машины; $p_0(v)$ и $p_2(v)$ являются функциями скорости машины, p_1 — константа; $r_1 = rand(0,1)$ случайная величина, равномерно распределенная 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

машина 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 — состояние движения машины 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 машины:

$$\xi_a = a\tau\theta(p_a - r),$$

$$\xi_b = a\tau\theta(p_b - r),$$

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 машины, $p^{(0)}$ и $a^{(0)}$ — константы, $r = rand(0,1)$, ступенчатая функция $\theta(z)$ определяется как $\theta(z) = 0$ при $z < 0$ и $\theta(z) = 1$ при $z \geq 0$.

Безопасная скорость $v_{s,n}$ вычисляется следующим образом:

$$v_{s,n} = \min\left(v_n^{(safe)}, v_\ell^{(a)} + \frac{g_n}{\tau}\right),$$

где

$$v_\ell^{(a)} = \max(0, \min(v_{\ell,n}^{(safe)}, v_{\ell,n}, g_{\ell,n}/\tau) - a\tau)$$

— так называемая «ожидаемая» (прогнозируемая) скорость машины 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 Краусса с соавторами [64] в 1997 году и которая в свою очередь является решением уравнения Гипса [65]:

$$v^{(safe)}\tau + X_d(v^{(safe)}) = g_n + X_d(v_{\ell,n}),$$

где $X_d(u)$ — тормозной путь, проходимый машиной, движущейся с первоначальной скоростью u и тормозящей с постоянным ускорением b вплоть до полной остановки. В модели с дискретным временем этот путь дается формулой $X_d(u) = b\tau^2(\alpha\beta + 0,5\alpha(\alpha - 1))$, α и β — соответственно целая и дробная части величины $u/b\tau$.

В рассматриваемой модели автодороги с двумя полосами смена полосы машинами происходит независимо от того, находятся ли эти машины вдали или вблизи неоднородностей дороги, связанной с въездом/выездом. Машина меняет полосу, если 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

$$R \rightarrow L: v_n^+ \geq v_{\ell,n} + \delta_1 \quad \text{и} \quad v_n \geq v_{\ell,n},$$

$$L \rightarrow R: v_n^+ > v_{\ell,n} + \delta_1 \quad \text{или} \quad v_n^+ > v_n + \delta_1.$$

Условия безопасности при смене полосы имеют вид

$$g_n^+ > \min(v_n \tau, G_n^+), \quad g_n^- > \min(v_n^- \tau, G_n^-), \quad (7)$$

где

$$G_n^+ = G(v_n, v_n^+), \quad G_n^- = G(v_n^-, v_n),$$

верхние индексы «+» и «-» относятся соответственно к машинам впереди и позади на соседней полосе.

Если условия (7) не выполняются, то используются более жесткие условия для «вдавливания» машины на соседнюю полосу:

$$x_n^+ - x_n^- - d > g_{t \arg et}^{(\min)}, \quad \text{где} \quad g_{t \arg et}^{(\min)} = \lfloor \lambda v_n^+ + d \rfloor. \quad (8)$$

В дополнение к (8) используется условие, что машина проходит среднюю точку $x_n^{(m)} = \lfloor (x_n^+ + x_n^-)/2 \rfloor$ между двумя соседними машинами на соседней полосе, то есть выполняются следующие условия:

$$x_{n-1} < x_{n-1}^{(m)} \quad \text{и} \quad x_n \geq x_n^{(m)} \quad \text{или} \quad x_{n-1} \geq x_{n-1}^{(m)} \quad \text{и} \quad x_n < x_n^{(m)}.$$

Если условия для смены полосы выполняются, машина 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 машины не меняется, если выполнены условия (7), и она устанавливается равной $x_n = x_n^{(m)}$, если выполнено условие (8). Величины $\Delta v^{(1)}$, p_c , δ_1 и λ являются константами.

Более подробно об использованных параметрах модели, условиях смены полосы и модели поведения водителя на въезде/выезде со скоростной автодороги можно прочитать в главе 11 книги Кернера [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} .

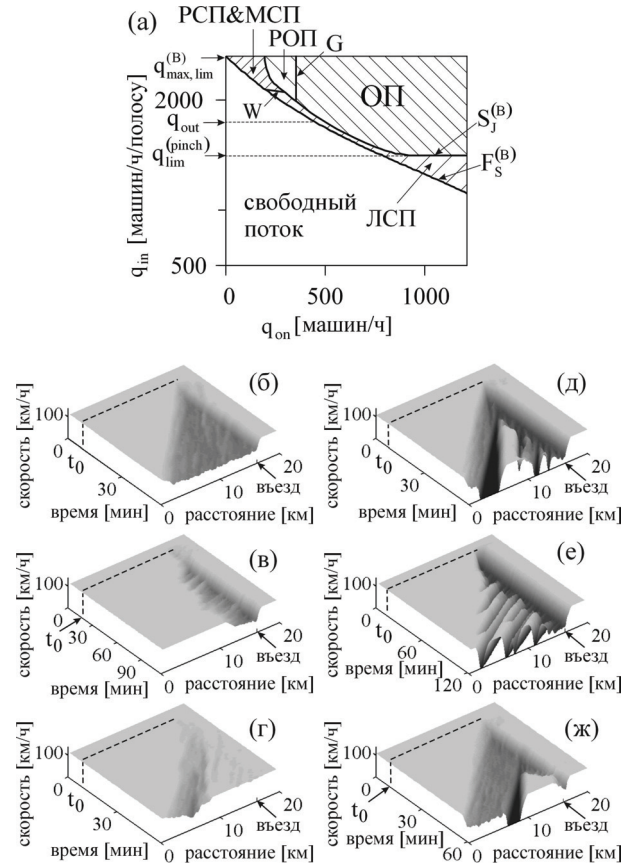


Рис. 13. Диаграмма пространственно-временных структур транспортного потока вблизи изолированного въезда (а) и соответствующие типы пространственно-временных структур (б)–(ж), относящиеся к диаграмме (а): СП (б)–(г) и ОП (д)–(ж) (из книги Кернера [2])

На рис. 13а показана диаграмма этих пространственно-временных структур. По оси абсцисс данной диаграммы откладывается поток машин q_{on} , по оси ординат указан поток машин по основной дороге q_{in} . Граница $F_S^{(B)}$ на диаграмме разделяет свободный поток влево от границы от структур плотного потока, возникающих вблизи въезда машин. 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 плотного потока — МСП (рис. 13г). При увеличении q_{on} МСП превращается в расширяющуюся структуру плотного потока — РСП (рис. 13б). Если уменьшать q_{in} и увеличивать q_{on} , то на границе, обозначенной буквой W, РСП превращаются в локализованную структуру плотного потока — ЛСП (рис. 13в).

Укажем еще некоторые особенности общей структуры плотного потока — ОП, которая возникает выше границы $S_J^{(B)}$ на диаграмме (область ОП на рис. 13а). Существует граница, обозначенная буквой G внутри области возникновения

структур ОП. Слева от этой границы, после того как широкий кластер возникает в синхронизованном потоке, новые широкие кластеры больше не формируются (рис. 13ж), и возникает рассасывающаяся общая структура плотного потока (РОП). Правее от границы G возникает общая структура, в которой непрерывно рождаются новые кластеры внутри синхронизованного потока (рис. 13д). Однако если уменьшать поток по дороге q_{in} , оставаясь в области ОП, и перейти к значениям q_{in} меньше чем q_{out} (поток q_{out} соответствует величине свободного потока, образованного машинами, выезжающими из широкого движущегося кластера), то широкие движущиеся кластеры, спонтанно возникающие в синхронизованном потоке ОП, постепенно рассасываются, распространяясь против течения. В результате возникает ОП, показанная на рис. 13е. Другая особенность общей структуры состоит в том, что если поток из въезда q_{in} дальше увеличивать, то возникает некоторый эффект насыщения, связанный с возникновением плотного потока на дороге, ведущей к въезду машин. Этот эффект насыщения связан с тем, что поток внутри области синхронизованного потока в ОП достигает своего предельного минимального значения, обозначенного на рис. 13а как $q_{lim}^{(pin ch)}$.

Распределения скорости и потока внутри ОП структуры на рис. 13д как функции времени при разных координатах вдоль дороги показаны на рис. 14. Можно видеть, что сначала на 16,5 км возникает фазовый $F \rightarrow S$ переход из свободного в синхронизованный поток. Через 2 км против течения ($x = 14,5$) видно развивающиеся узкие кластеры внутри синхронизованного 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).

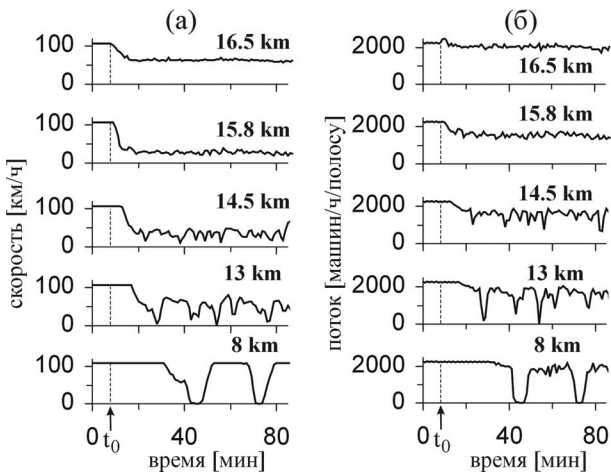


Рис. 14. Одноминутные данные виртуальных детекторов, отвечающие ОП на рис. 13д. Въезд на скоростную дорогу отвечает координате 16 км (из книги Кернера [2])

Более подробно свойства пространственно-временных структур, полученные в результате моде-

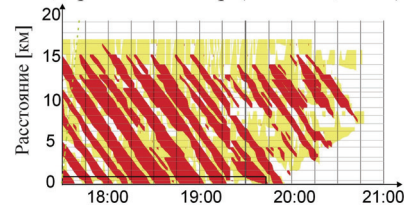
лирования, рассматриваются в книгах Кернера [1] и [2].

ХIII. Применение теории трех фаз Кернера для интеллектуальных транспортных технологий

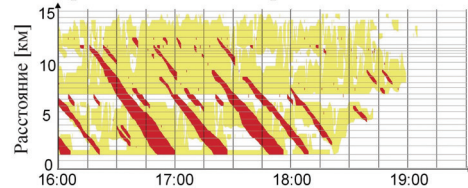
Кернер с сотрудниками предложил и частично внедрил в эксплуатацию целый ряд новых методов интеллектуальных транспортных технологий. Одним из внедренных и уже установленных на скоростных автодорогах применений теории трех фаз Кернера является метод ASDA / FOTO. Метод ASDA / FOTO функционирует в работающей 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 Кернера. 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).

Единая общая структура плотного потока (ЕОП)

(а) Автодорога А5-Север (14 июня, 2006) в Германии



(б) Автодорога М-42 (11 января, 2008) в Великобритании



(в) Автодорога I405-Юг (04 марта, 2003) в США

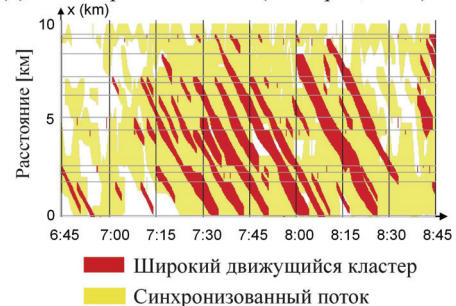


Рис. 15. Пространственно-временная структура транспортного потока, полученная методом ASDA / FOTO в трех странах (из книги Кернера [1])

Дальнейшее развитие приложений теории трех фаз Кернера связано с разработкой и усовершенствованием моделей для транспортных симуляторов, методов регулирования въездного по-

тока на автомагистраль (ANCONA), методов коллективного регулирования транспортных потоков, системы автоматического ассистента водителя и методов детектирования состояния транспортного потока, описанных в книгах Кернера [1, 2].

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. *Узлем 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, N 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. — New York: Englewood Cliffs, 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.* Moving bottlenecks: a theory grounded on experimental observations // Traffic and Transportation Theory. Ed. by Taylor M.A.P. — Oxford: Pergamon, 2002. — P. 441–462.
18. Traffic Flow Theory / Ed. by Gartner N.H., Messer C.J., Rathi A. — 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., Woessler R.* Still flowing: Approaches to traffic flow and traffic jam modelling // Oper. Res. — 2003. — V. 51. — P. 681–716.
23. *Mahnke R., Kaupuys 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. *Kerner B.S., Konhäuser P.* Cluster effect in initially homogeneous traffic flow // Phys. Rev. E. — 1993. — V. 48. — P. 2335–2338.
27. *Kerner B.S., Konhäuser P.* Structure and parameters of clusters in traffic flow // Phys. Rev. E. — 1994. — V. 50. — P. 54–83.
28. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. — 2005. — V. 120, N 3–4. — P. 627–658.
29. *Maerivoet S., De Moor B.* Cellular automata models of road traffic // Phys. Rep. — 2005. — V. 419, N 1. — P. 1–64.
30. *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.
31. *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.
32. *Davis L.C.* Multilane simulations of traffic phases // Phys. Rev. E. — 2004. — V. 69. — P. 016108.
33. *Kerner B.S., Klenov S.L.* Deterministic microscopic three-phase traffic flow models // J. Phys. A: Math. Gen. — 2006. — V. 39. — P. 1775–1809.
34. *Lee H.K., Barlovic R., Schreckenberg M., Kim D.* Mechanical Restriction versus Human Overreaction Triggering Congested Traffic States // Phys. Rev. Lett. — 2004. — V. 92. — P. 238702.
35. *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.
36. *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. — P. 026105.

- 37.** *Laval J.A.* Linking synchronized flow and kinematic waves // *Traffic and Granular Flow'05*. Ed. by Schadschneider A., Pöschel T., Kühne R., Schreckenberg M., Wolf D.E. — 2007. — P. 521–526.
- 38.** *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.
- 39.** *Davis L.C.* Controlling traffic flow near the transition to the synchronous flow phase // *Physica. A.* — 2006. — V. 368. — P. 541–550.
- 40.** *Davis L.C.* Effect of cooperative merging on the synchronous flow phase of traffic // *Physica. A.* — 2006. — V. 361. — P. 606–618.
- 41.** *Davis L.C.* Effect of adaptive cruise control systems on mixed traffic flow near an on-ramp // *Physica. A.* — 2007. — V. 379. — P. 274–290.
- 42.** *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.
- 43.** *Jiang R., Wu Q.-S.* Toward an improvement over Kerner–Klenov–Wolf three-phase cellular automaton model // *Phys. Rev. E.* — 2005. — V. 72. — 067103.
- 44.** *Jiang R., Wu Q.-S.* Dangerous situations in a synchronized flow model // *Physica. A.* — 2007. — V. 377. — P. 633–640.
- 45.** *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.
- 46.** *Pottmeier A., Thiemann C., Schadschneider A., Schreckenberg M.* Mechanical Restriction Versus Human Overreaction: Accident Avoidance and Two-Lane Traffic Simulations // *Traffic and Granular Flow'05*. Ed. by Schadschneider A., Pöschel T., Kühne R., Schreckenberg M., Wolf D.E. — Berlin: Springer, 2007. — P. 503–508.
- 47.** *Siebel F., Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // *Phys. Rev. E.* — 2006. — V. 73. — 066108.
- 48.** *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.
- 49.** *Kerner B.S., Klenov S.L., Hiller A., Rehborn H.* Microscopic features of moving traffic jams // *Phys. Rev. E.* — 2007. — V. 73. — 046107.
- 50.** *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.
- 51.** *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.
- 52.** *Kerner B.S.* A theory of traffic congestion at heavy bottlenecks // *J. Phys. A: Math. Theor.* — 2008. — V. 41. — 215101–369801.
- 53.** *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.* — 2008. — V. 387. — P. 6395–6410.
- 54.** *Davis L.C.* Realizing Wardrop equilibria with real-time traffic information // *Physica. A.* — 2009. — V. 388. — P. 4459–4474.
- 55.** *Davis L.C.* Predicting travel time to limit congestion at a highway bottleneck // *Physica. A.* — 2010. — V. 389. — P. 3588–3599.
- 56.** *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.
- 57.** *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.
- 58.** *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. — P. 1944–1987. — DOI: 10.1080/18128600903312789.
- 59.** *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.
- 60.** *Kerner B.S., Klenov S.L.* Phase transitions in traffic flow on multi-lane roads // *Phys. Rev. E.* — 2009. — V. 80. — 056101.
- 61.** *Kerner B.S., Klenov S.L.* A theory of traffic congestion on moving bottlenecks // *J. Phys. A: Math. Theor.* — 2010. — V. 43. — 42510.
- 62.** *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.
- 63.** *Kerner B.S., Klenov S.L.* Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks // *Phys. Rev. E.* — 2003. — V. 68. — 036130.
- 64.** *Krauß S., Wagner P., Gawron C.* Metastable states in a microscopic model of traffic flow // *Phys. Rev. E.* — 1997. — V. 55. — P. 5597–5602.
- 65.** *Gipps P.G.* A behavioural car-following model for computer simulation // *Transportation Research B.* — 1981. — V. 15. — P. 105–111.

А.В. Колесников

Московский государственный университет печати
Государственный университет — Высшая школа экономики

Транспортировка масс и сжимающие отображения

Согласно известному результату Л. Каффарелли, оптимальная транспортировка стандартной гауссовской меры в логарифмически вогнутую меру $e^{-W} dx$, удовлетворяющую условию $D^2W \geq Id$, является 1-липшицевым отображением. Настоящая работа представляет собой краткий обзор различных результатов и приложений, полученных в этом направлении.

Ключевые слова: оптимальная транспортировка, уравнение Монжа–Ампера, логарифмически вогнутые меры, гауссовы меры, изопериметрические неравенства, неравенства Соболева.

I. Введение

Пусть α — неотрицательное число. Будем называть отображение $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ α -липшицевым, если

$$|T(x) - T(y)| \leq \alpha|x - y|.$$

Для гладкого T это эквивалентно условию

$$\sup_{x \in \mathbb{R}^d} \|DT(x)\| \leq \alpha,$$

где $\|\cdot\|$ — операторная норма. В случае $\alpha = 1$ мы будем просто писать «сжатие».

Аналогично, если $T : X \rightarrow Y$ — отображение между метрическими пространствами, будем говорить, что T — сжатие, если $\rho_Y(T(x_1), T(x_2)) \leq \rho_X(x_1, x_2)$.

Пусть μ — вероятностная мера в метрическом пространстве (M, ρ) . Для произвольного множества $A \subset M$ определим поверхностную меру μ^+ границы ∂A :

$$\mu^+(\partial A) = \lim_{h \rightarrow 0} \frac{\mu(A_h) - \mu(A)}{h},$$

где $A_h = \{x : \rho(x, A) \leq h\}$.

Множество A называется изопериметрическим, если оно имеет минимальную поверхностную меру среди множеств такой же меры $\mu(A)$. Изопериметрическая функция \mathcal{I}_μ определяется следующим образом:

$$\mathcal{I}_\mu(t) = \inf\{\mu^+(\partial A) : \mu(A) = t\}.$$

Изопериметрические множества в большинстве случаев невозможно найти. В то же время для изопериметрических функций существуют различные оценки, имеющие обширные приложения в анализе, геометрии и теории вероятностей. Например, хорошо известно, что изопериметрические неравенства влекут неравенства типа Соболева. Подробнее см. в [9, 18, 22, 24, 27].

Многочисленные приложения сжимающих отображений основаны на следующем элементарном факте:

Пусть X, Y — два метрических пространства, и X наделено мерой μ . Предположим, что существует сжатие $T : X \rightarrow Y$ между метрическими пространствами X и Y . Тогда мера-образ $\nu = \mu \circ T^{-1}$ удовлетворяет неравенству

$$\mathcal{I}_\nu \geq \mathcal{I}_\mu.$$

В настоящей работе в основном изучается случай оптимальной транспортировки мер. Пусть нам даны две борелевские вероятностные меры μ и ν на \mathbb{R}^d . Рассмотрим оптимальную транспортировку $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$, минимизирующую «стоимость транспортировки»

$$W_2^2(\mu, \nu) = \int |x - T(x)|^2 d\mu,$$

среди отображений, отображающих μ в ν , $\nu = \mu \circ T^{-1}$. Последнее означает, что $\mu \circ T^{-1}(A) = \nu(A)$ для любого борелевского множества A . Если $\mu = \rho_0 dx$ и $\nu = \rho_1 dx$ абсолютно непрерывны, то такое T существует и может быть получено из решения транспортной задачи Монжа–Канторовича. Более того, это отображение μ -единственно и имеет вид $T = \nabla \Phi$, где Φ — выпуклая функция (см. [27]). Предполагая гладкость Φ , легко проверить, что Φ является решением следующего нелинейного уравнения (уравнения Монжа–Ампера):

$$\rho_1(\nabla \Phi) \det D^2 \Phi = \rho_0.$$

Настоящая статья содержит обзор работ о сжимаемости оптимальных транспортных отображений. Первый результат в этом направлении был доказан Л. Каффарелли (см. [6]). Согласно этому результату, если μ — стандартная гауссовская мера $\mu = \frac{1}{(2\pi)^{d/2}} e^{-\frac{x^2}{2}} dx$ и $\nu = e^{-W} dx$ с $D^2W \geq Id$, то соответствующее отображение T является сжатием. Из этого наблюдения немедленно следует теорема сравнения Бакри–Леду [2] и различные функциональные неравенства, включающие логарифмическое неравенство Соболева для равномерно логарифмически вогнутых мер. Среди других приложений отметим гауссовское

корреляционное неравенство и неравенство Браскампа–Либа. Мы также обсудим некоторые обобщения теоремы Каффарелли и некоторые открытые проблемы.

II. Теорема Каффарелли о сжатии

Замечание 1. Теоремы 1 и 2 будут цитироваться в работе как «теорема Каффарелли о сжатии». Оригинальная формулировка была дана в теореме 2.

Теорема 1 (L. Caffarelli). Пусть $T = \nabla\Phi$ оптимальная транспортировка, отображающая вероятностную меру $\mu = e^{-V}dx$ в вероятностную меру $\nu = e^{-W}dx$. Пусть V и W дважды непрерывно дифференцируемы и $D^2W \geq K$. Тогда для любого единичного вектора e :

$$\sup_{x \in \mathbb{R}^d} \Phi_{ee}^2 \leq \frac{1}{K} \sup_{x \in \mathbb{R}^d} V_{ee}.$$

В частности, если μ — стандартная гауссовская мера и $K \geq 1$, то T — сжатие.

Идея доказательства:

1) Принцип максимума

Приведем идею доказательства, основанного на принципе максимума. Функции V, W и Φ предполагаются достаточно регулярными. Впрочем, гладкость Φ можно вывести из гладкости V, W (плюс некоторые ограничения на рост, см. теорему 4.14 [27]). Запишем формулу замены переменных

$$e^{-V} = e^{-W(\nabla\Phi)} \det D^2\Phi.$$

Возьмем логарифм от обеих частей

$$V = W(\nabla\Phi) - \log \det D^2\Phi.$$

Зафиксируем единичный вектор e и продифференцируем эту формулу вдоль e . Для этого применим фундаментальное соотношение

$$\partial_e \ln \det D^2\Phi = \frac{\partial_e \det D^2\Phi}{\det D^2\Phi} = \text{Tr}(D^2\Phi)^{-1} D^2\Phi_e.$$

Продифференцировав эту формулу вдоль другого направления v и пользуясь тем, что

$$D^2\Phi_v(D^2\Phi)^{-1} + D^2\Phi[(D^2\Phi)^{-1}]_v = 0,$$

получаем

$$\begin{aligned} \partial_{ev} \ln \det D^2\Phi &= \text{Tr}(D^2\Phi)^{-1} D^2\Phi_{ev} - \\ &- \text{Tr}[(D^2\Phi)^{-1} D^2\Phi_e (D^2\Phi)^{-1} D^2\Phi_v]. \end{aligned}$$

Используем опять формулу замены переменных

$$V_e = \langle \nabla W(\nabla\Phi), D^2\Phi \cdot e \rangle - \text{Tr}(D^2\Phi)^{-1} D^2\Phi_e$$

и

$$V_{ee} = \langle D^2W(\nabla\Phi) D^2\Phi \cdot e, D^2\Phi \cdot e \rangle + \langle \nabla W(\nabla\Phi), \nabla\Phi_{ee} \rangle -$$

$$- \text{Tr}(D^2\Phi)^{-1} D^2\Phi_{ee} + \text{Tr}[(D^2\Phi)^{-1} D^2\Phi_e]^2.$$

Предположим, что Φ_{ee} достигает максимума в точке x_0 . Тогда

$$\nabla\Phi_{ee}(x_0) = 0, \quad D^2\Phi_{ee} \leq 0.$$

Заметим, что $\text{Tr}[(D^2\Phi)^{-1} D^2\Phi_e]^2 > 0$, потому что это равно $\text{Tr}C^2$, где

$$C = (D^2\Phi)^{-1/2} D^2\Phi_e (D^2\Phi)^{-1/2}$$

— симметричная матрица.

Очевидно, $\text{Tr}(D^2\Phi(x_0))^{-1} D^2\Phi_{ee}(x_0) \leq 0$, следовательно

$$V_{ee}(x_0) \geq K \|D^2\Phi(x_0) \cdot e\| \geq K \Phi_{ee}^2(x_0).$$

Таким образом,

$$\sup_{x \in \mathbb{R}^d} \Phi_{ee}^2 \leq \frac{1}{K} \sup_{x \in \mathbb{R}^d} V_{ee}(x_0).$$

2) Разностные приращения

Вместо того чтобы дифференцировать уравнение Монжа–Ампера, можно рассмотреть разностные приращения

$$\delta_2\Phi(x) = \Phi(x + th) + \Phi(x - th) - 2\Phi(x) \geq 0$$

для некоторого вектора $h \in \mathbb{R}^d, |h| = 1$. Используя приближения, можно свести ситуацию к случаю, когда $\text{supp}(\nu)$ — ограниченная выпуклая область и V, W — локально гильдеровы. Из результатов Каффарелли о гильдеровой регулярности следует $\Phi \in C_{loc}^{2,\alpha}(\mathbb{R}^d)$.

Кроме того, используя приближения, можно полагать, что μ убывает не быстрее гауссовской меры, то есть $V(x) \leq C_1 + C_2|x|^2$ для некоторых $C_1, C_2 \geq 0$. Тогда имеет место следующая лемма (см. лемму 4 в [6]).

Лемма 1. $\lim_{x \rightarrow \infty} \delta_2\Phi(x) = 0$.

Следовательно, существует точка максимума x_0 функции $\delta_2\Phi(x)$. Дифференцирование по x_0 влечет

$$\nabla\Phi(x_0 + th) + \nabla\Phi(x_0 - th) = 2\nabla\Phi(x_0), \quad (1)$$

$$D^2\Phi(x_0 + th) + D^2\Phi(x_0 - th) \leq 2D^2\Phi(x_0).$$

Из вогнутости определителя следует

$$\begin{aligned} \det D^2\Phi(x_0) &\geq \det \left(\frac{D^2\Phi(x_0 + th) + D^2\Phi(x_0 - th)}{2} \right) \geq \\ &\geq \left(\det D^2\Phi(x_0 + th) \det D^2\Phi(x_0 - th) \right)^{\frac{1}{2}}. \end{aligned}$$

Применив формулу замены переменных $\det D^2\Phi = e^{W(\nabla\Phi) - V}$, получаем

$$\begin{aligned} V(x_0 + th) + V(x_0 - th) - 2V(x_0) &\geq \\ &\geq W(\nabla\Phi(x_0 + th)) + W(\nabla\Phi(x_0 - th)) - 2W(\nabla\Phi(x_0)). \end{aligned} \quad (2)$$

Из (1) вытекает, что

$$v := \nabla\Phi(x_0 + th) - \nabla\Phi(x_0) = \nabla\Phi(x_0) - \nabla\Phi(x_0 - th).$$

Таким образом, получаем из (2), что

$$\begin{aligned} \sup V_{hh} \cdot t^2 &\geq K |\nabla\Phi(x_0 + th) - \nabla\Phi(x_0)|^2 = \\ &= K |\nabla\Phi(x_0 - th) - \nabla\Phi(x_0)|^2 = K |v|^2. \end{aligned}$$

В силу выпуклости Φ :

$$\begin{aligned} \Phi(x_0 + th) + \Phi(x_0 - th) - 2\Phi(x_0) &\leq \\ &\leq t \langle \nabla\Phi(x_0 + th) - \nabla\Phi(x_0 - th), h \rangle = \\ &= 2t \langle v, h \rangle \leq 2t |v|. \end{aligned}$$

Таким образом,

$$\frac{\sup_{x \in \mathbb{R}^d} V_{hh}}{K} \geq \left(\frac{\delta_2 \Phi}{2t^2} \right)^2.$$

Отсюда вытекает, что

$$\Phi_{hh} \leq 2C$$

при $C = \sqrt{\frac{\sup_{x \in \mathbb{R}^d} V_{hh}}{K}}$. Но эта оценка хуже желаемой. Чтобы получить нужную оценку, используем дополнительную информацию, что $\Phi_{hh} \leq a_0 C$, где $a_0 = 2$. Применим соотношение

$$\begin{aligned} \Phi(x_0 + th) + \Phi(x_0 - th) - 2\Phi(x_0) &= \\ &= \int_0^t \langle \nabla\Phi(x_0 + sh) - \nabla\Phi(x_0 - sh), h \rangle ds. \end{aligned}$$

В силу выпуклости Φ $\langle \nabla\Phi(x_0 + sh) - \nabla\Phi(x_0 - sh), h \rangle \leq \langle \nabla\Phi(x_0 + th) - \nabla\Phi(x_0 - th), h \rangle$, выполнена оценка

$$\Phi(x_0 + th) + \Phi(x_0 - th) - 2\Phi(x_0) \leq \int_0^t \min(2a_0 C s, 2|v|) ds.$$

Вычисляя правую часть и принимая во внимание, что $|v| \leq Ct$, получаем

$$\Phi(x_0 + th) + \Phi(x_0 - th) - 2\Phi(x_0) \leq a_1 C t^2,$$

где $a_1 = \frac{3}{2}$. Таким образом, $\Phi_{hh} \leq a_1 C$. Повторяя эти аргументы бесконечное число раз, получим $\Phi_{hh} \leq a_n C$ и $\lim_n a_n = 1$. Теорема доказана.

3) L^p -оценки

См. раздел 6.

Замечание 2. Заметим, что теорема из [6] несколько отличается от результата выше. Ниже приведен оригинальный результат Каффарелли.

Теорема 2 (L. Caffarelli). Пусть $\mu = e^{-Q} dx$ — произвольная гауссовская мера. Тогда для любой меры $\nu = e^{-Q-P} dx$, где P — выпуклая функция, соответствующая оптимальная транспортировка T является сжатием.

Идея доказательства: Применим принцип максимума. Мы ищем максимум $\Phi_{ee}(x)$ среди единичных e и $x \in \mathbb{R}^d$. Применим соотношение, полученное выше

$$Q_{ee} = \langle D^2(Q + P)(\nabla\Phi)D^2\Phi \cdot e, D^2\Phi \cdot e \rangle +$$

$$\begin{aligned} &+ \langle \nabla(Q + P)(\nabla\Phi), \nabla\Phi_{ee} \rangle - \text{Tr}(D^2\Phi)^{-1} D^2\Phi_{ee} + \\ &+ \text{Tr} \left[(D^2\Phi)^{-1} D^2\Phi_e \right]^2. \end{aligned}$$

Из тех же самых аргументов, что и выше, следует

$$Q_{ee} \geq \langle D^2(Q + P)(\nabla\Phi)D^2\Phi \cdot e, D^2\Phi \cdot e \rangle.$$

Учтем, что P — выпуклая функция и e — собственный вектор $D^2\Phi$. Получаем

$$Q_{ee} \geq \Phi_{ee}^2 \cdot Q_{ee}(\nabla\Phi).$$

Из того, что Q_{ee} постоянно, следует искомое утверждение.

III. Равномерно выпуклые меры общего вида

Доказательство, основанное на изучении дифференциальных разностей, может быть легко обобщено на случай равномерно логарифмически вогнутых мер (в обобщенном смысле). Последнее означает, что потенциал W удовлетворяет соотношению

$$W(x + y) + W(x - y) - W(x) \geq \delta(|y|)$$

для некоторой возрастающей неотрицательной функции δ . Следующий результат был доказан в [15].

Теорема 3. Предположим, что V и W удовлетворяют соотношению

$$V(x + y) + V(x - y) - 2V(x) \leq A_p |y|^{p+1},$$

$$W(x + y) + W(x - y) - 2W(x) \geq A_q |y|^{q+1}$$

для некоторых $0 \leq p \leq 1$, $1 \leq q$, $A_p > 0$, $A_q > 0$.

Тогда Φ удовлетворяет неравенству

$$\Phi(x + th) + \Phi(x - th) - 2\Phi(x) \leq 2 \left(\frac{A_p}{A_q} \right)^{\frac{1}{q+1}} t^{1+\alpha} \quad (3)$$

для любого единичного вектора $h \in \mathbb{R}^d$ с $\alpha = \frac{p+1}{q+1}$.

Замечание 3. Константа в (3) в общем случае не оптимальна.

Из (3) следует, что $\nabla\Phi$ — глобально гёльдерово отображение. Это верно и без предположения выпуклости Φ , но выпуклый случай проще и следует из леммы, сообщенной автору Сашей Содиным.

Лемма 2. Для любой выпуклой функции f и единичного вектора h выполнено

$$\begin{aligned} &|\nabla f(x + th) - f(x)| \leq \\ &\leq \frac{2}{t} \sup_{v: |v|=1} \left(f(x + 2tv) + f(x - 2tv) - 2f(x) \right). \end{aligned}$$

Используя эту лемму, можно усилить результат о гёльдерово транспортировке.

Теорема 4. Предположим, что

$$V(x + y) + V(x - y) - 2V(x) \leq |y|^2,$$

и

$$W(x + y) + W(x - y) - W(x) \geq \delta(|y|)$$

для некоторой неотрицательной возрастающей функции δ . Тогда

$$|\nabla\Phi(x) - \nabla\Phi(y)| \leq 8\delta^{-1}(4|x - y|^2).$$

Применяя эту оценку, можно перенести знаменитое гауссовское изопериметрическое неравенство Судакова-Цирельсона на случай (обобщенной) равномерно выпуклой меры. Напомним (см. [1]), что стандартная гауссова мера γ удовлетворяет гауссовому изопериметрическому неравенству

$$\gamma(A^r) \geq \Phi(\Phi^{-1}(\gamma(A)) + r),$$

где $A^r = \{x \in \mathbb{R}^d : \exists a \in A : |a - x| < r\}$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Следовательно, применяя теорему 4 к мерам $\mu = \gamma$ и $\nu = e^{-W} dx$ с потенциалом W , удовлетворяющим

$$W(x + y) + W(x - y) - W(x) \geq \delta(|y|),$$

мы получаем

$$\nu(A^r) \geq \Phi\left(\Phi^{-1}(\nu(A)) + \frac{1}{2}\sqrt{\delta(r/8)}\right).$$

В частности, ν обладает безразмерным свойством концентрации

$$\nu(A^r) \geq 1 - \frac{1}{2} \exp\left(-\frac{1}{8}\delta(r/8)\right),$$

если $\nu(A) \geq 1/2$. Близкий результат был получен Е. Мильманом и С. Содиным в работе [23] с помощью локализационных аргументов. Из результатов Е. Мильмана [21] следует, что изопериметрические неравенства и неравенства концентрации эквивалентны для логарифмически вогнутых мер. Подробнее о неравенствах концентрации см. [18, 22].

IV. Мера Лебега на выпуклом множестве

В этом разделе мы обсудим следующую проблему.

Проблема 1. Рассмотрим «хорошую» μ (например, произведение гауссовских или экспоненциальных мер). Требуется эффективно оценить липшицеву константу оптимального отображения, отображающего μ в нормированную меру Лебега на выпуклом множестве K .

Замечание 4. Прямые произведения мер являются «хорошими», потому что константы для неравенств типа Соболева (константы Чигера, Пуанкаре и т.д.) легко оцениваются для таких мер. К другим хорошим мерам можно отнести логарифмически вогнутые с равномерно выпуклым или радиально-симметричным потенциалом.

Эта задача была мотивирована известной гипотезой Каннана-Ловаша-Симоновица (КЛС-гипотеза). Напомним, что константой Чигера $C_{\text{ch}}(K)$

выпуклого тела K называется наименьшая константа, для которой выполнено неравенство

$$\int_K \left| f - \frac{1}{\lambda(K)} \int_K f dx \right| dx \leq C_{\text{ch}}(K) \int_K |\nabla f| dx$$

для любой гладкой функции f .

КЛС-гипотеза. Существует такая универсальная константа c , что

$$C_{\text{ch}}(K) \leq c$$

для любого выпуклого $K \subset \mathbb{R}^d$, удовлетворяющего соотношениям

$$\int_K x_i dx = 0, \frac{1}{\lambda(K)} \int_K x_i x_j dx = \delta_i^j.$$

Подробнее о КЛС-гипотезе см. [5, 12, 21].

Некоторые результаты такого типа были получены в [15]. Аргументы ниже обобщают доказательство, основанное на принципе максимума. Рассмотрим оптимальную транспортировку, отображающую $e^{-V} dx$ в $\frac{1}{\lambda(K)} \lambda|_K$. Зафиксируем единичный вектор h . Найдем такую функцию ψ , что функция

$$\psi(\Phi_h) + \log \Phi_{hh}$$

будет ограничена сверху. Предположим, что x_0 — точка максимума этой функции. В этой точке

$$\psi'(\Phi_h) \nabla \Phi_h + \frac{1}{\Phi_{hh}} \nabla \Phi_{hh} = 0, \tag{4}$$

$$\begin{aligned} \psi''(\Phi_h) \nabla \Phi_h \oplus \nabla \Phi_h + \psi'(\Phi_h) D^2 \Phi_h + \frac{1}{\Phi_{hh}} D^2 \Phi_{hh} - \\ - \frac{1}{\Phi_{hh}^2} \nabla \Phi_{hh} \oplus \nabla \Phi_{hh} \leq 0. \end{aligned} \tag{5}$$

Дифференцирование формулы замены переменных дает соотношения (см. раздел 1):

$$V_h = -\text{Tr}(D^2 \Phi)^{-1} D^2 \Phi_h,$$

$$V_{hh} = -\text{Tr}(D^2 \Phi)^{-1} D^2 \Phi_{hh} + \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_h]^2.$$

Умножим (5) на $(D^2 \Phi)^{-1}$, возьмем след и подставим выражение для V_{hh} в формулу. Получим

$$\begin{aligned} V_{hh} \geq -\frac{1}{\Phi_{hh}} \text{Tr}[(D^2 \Phi)^{-1} \cdot \nabla \Phi_{hh} \oplus \nabla \Phi_{hh}] + \\ + \Phi_{hh} \cdot \psi''(\Phi_h) \text{Tr}[(D^2 \Phi)^{-1} \cdot \nabla \Phi_h \oplus \nabla \Phi_h] + \\ + \Phi_{hh} \cdot \psi'(\Phi_h) \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_h] + \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_h]^2. \end{aligned}$$

Заметим, что $\text{Tr}(D^2 \Phi)^{-1} \cdot \nabla \Phi_h \oplus \nabla \Phi_h = \Phi_{hh}$. Из (4) вытекает, что $\nabla \Phi_{hh} = -\Phi_{hh} \psi'(\Phi_h) \nabla \Phi_h$. Подставляя это в неравенство для V_{hh} , получаем

$$\begin{aligned} V_{hh} \geq \Phi_{hh}^2 [\psi'' - (\psi')^2] \circ \Phi_h + \\ + \Phi_{hh} \cdot \psi'(\Phi_h) \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_h] + \\ + \text{Tr}[(D^2 \Phi)^{-1} D^2 \Phi_h]^2. \end{aligned}$$

Заметим, что

$$\mathrm{Tr}[(D^2\Phi)^{-1}D^2\Phi_h] = \mathrm{Tr}C,$$

$$\mathrm{Tr}[(D^2\Phi)^{-1}D^2\Phi_h]^2 = \mathrm{Tr}C^2,$$

где

$$C = (D^2\Phi)^{-1/2}(D^2\Phi_h)(D^2\Phi)^{-1/2}$$

— симметричная матрица. По неравенству Коши

$$V_{hh} \geq \Phi_{hh}^2 \left[\psi'' - \left(1 + \frac{d}{4}\right) (\psi')^2 \right] \circ \Phi_h.$$

Предположим, что V_{hh} ограничено константой C . Пусть ψ — функция, удовлетворяющая соотношению $\psi'' - \left(1 + \frac{d}{4}\right) (\psi')^2 \geq e^{2\psi}$. Получаем

$$C \geq \Phi_{hh}^2(x_0) e^{2\psi(\Phi_h(x_0))} = \sup_{x \in \mathbb{R}^d} \Phi_{hh}^2 e^{2\psi(\Phi_h)}.$$

В частности, выбирая подходящую функцию ψ , получим следующее утверждение (см. детали в [15]).

Теорема 5. 1) Оптимальная транспортировка T стандартной гауссовой меры γ в $\frac{1}{\lambda(K)}\lambda|_K$, где K — выпуклое тело, удовлетворяет оценке

$$\|DT\| \leq c\sqrt{d}\mathrm{diam}(K),$$

где c — универсальная константа, а $\mathrm{diam}(K)$ — диаметр K .

2) Оптимальная транспортировка T меры $\mu = e^{-V}dx$ в $\frac{1}{\lambda(K)}\lambda|_K$, где $V_{hh} \leq C$, $|V_h| \leq C$ для некоторого C и всех h , $|h| = 1$, удовлетворяет оценке

$$\|DT\| \leq c\mathrm{diam}(K),$$

где c зависит только от C .

К сожалению, оценки теоремы 5 недостаточно сильны, чтобы дать новое доказательство даже известных результатов о константе Чигера для выпуклых тел. Возникает следующая естественная проблема.

Проблема 2. Существует ли не зависящая от размерности оценка для $\|DT\|$, где $\mu = \gamma$ и $\nu = \frac{1}{\lambda(K)}\lambda|_K$? Тот же самый вопрос для произведения экспоненциальных распределений.

Заметим также, что было бы достаточно получить оценки для $\int \|DT\|^p d\mu$. Это следует из результата Е. Мильмана об эквивалентности норм для логарифмически вогнутых мер [21].

V. Сжатие для транспортировки мер, порожденной полугруппами

Результат о сжатии для другого типа транспортировки был недавно получен в [13].

Рассмотрим полугруппу $P_t = e^{tL}$, порожденную

$$L = \Delta - \langle \nabla V, \nabla \rangle = e^V \mathrm{div}(e^{-V} \cdot \nabla),$$

и поток вероятностных мер

$$\nu_t = P_t(e^{-W+V}) \cdot \mu.$$

Очевидно, μ — инвариантная мера для P_t , $\nu_0 = \nu$, и $\nu_\infty = \mu$.

Запишем уравнение для ν_t :

$$\begin{aligned} \frac{d}{dt}\nu_t &= LP_t(e^{-W+V}) \cdot \mu = \mathrm{div}[\nabla P_t(e^{-W+V}) \cdot e^{-V}] = \\ &= \mathrm{div}[\nabla \log P_t(e^{-W+V}) \cdot \nu_t]. \end{aligned}$$

Соответствующий поток диффеоморфизмов определяется уравнением

$$\frac{d}{dt}S_t = -\nabla \log P_t(e^{-W+V}) \circ S_t, S_0 = Id, \quad (6)$$

где ν_t и S_t связаны соотношением

$$\nu_t = \nu \circ S_t^{-1}.$$

В частности, предельное отображение $S_\infty = \lim_{t \rightarrow \infty} S_t$ отображает ν в μ . Обозначим обратное отображение через T_t :

$$T_t \circ S_t = Id, T = \lim_{t \rightarrow \infty} T_t.$$

Свойство сжатия для $T = S^{-1}$ эквивалентно свойству «расширения» для S . Так как T и S — диффеоморфизмы, то достаточно доказать, что $(DS_t)^* DS_t \geq Id$. Используя (6), получаем

$$\frac{d}{dt}DS_t(x) = -DW_t(S_t) \cdot DS_t, W_t = \nabla \log P_t(e^{-W+V}).$$

Таким образом,

$$\frac{d}{dt}(DS_t)^* DS_t = 2(DS_t)^* \cdot DW_t(S_t) \cdot DS_t.$$

Если

$$DW_t(S_t) = -D^2 \log P_t(e^{-W+V}) \geq 0,$$

то S_t обладает нужным свойством.

Предположим, что функция U , определенная по формуле

$$\nu = e^{-U} \cdot \mu, U = W - V,$$

выпукла. Тогда свойство $-D^2 \log P_t(e^{-W+V}) = -D^2 \log P_t e^{-U} \geq 0$ означает, что P_t сохраняет логарифмически вогнутые функции. Так мы получили следующую теорему.

Теорема 6. Предположим, что U — выпуклая функция. Если $U_t = -\log P_t e^{-U}$ — выпуклая функция для любого $t \geq 0$, то каждое отображение T_t является 1-сжатием.

Заметим, что согласно результату из [14], свойством сохранять все логарифмически вогнутые функции обладают только диффузионные гауссовские полугруппы. Тем не менее Ким и Мильман показали, что при наличии некоторой симметрии логарифмическая вогнутость может сохраняться. Доказательство основано на применении принципа максимума.

В частности, ими было получен следующий результат (см. более общую формулировку в [13]).

Теорема 7. Предположим, что μ — продакт-мера, V и U — выпуклые функции, причем U удовлетворяет условию $U(x_1, \dots, x_n) = U(\pm x_1, \dots, \pm x_n)$, и V имеет вид $V(x) = \sum_{i=1}^d \rho_i(|x_i|)$, где $\rho_i''' \leq 0$.

Тогда T — сжатие.

Кратко обсудим идею доказательства. Пусть t_0 — первый момент, когда U_t теряет выпуклость. Предположим, что минимум $\partial_{ee}U_{t_0}$ достигается в некоторой точке x_0 для некоторого направления e . Тогда $(d/dt - \Delta)\partial_{ee}U_t|_{t_0, x_0} \leq 0$. Кроме этого, $\nabla\partial_e U_t = 0$ и $\nabla\partial_{ee}U_t = 0$. Используя это, можно показать, что

$$(d/dt - \Delta)\partial_{ee}U_t|_{t_0, x_0} = -\langle \nabla U_t, \nabla V_{ee} \rangle|_{t_0, x_0}.$$

В момент t_0 функция U_t еще выпукла и легко показать, что правая часть равенства неотрицательна. Это ведет к противоречию.

VI. L^p -сжатие

В этом разделе мы обсудим L^p -обобщения теоремы Каффарелли (см. [16]). Результаты доказаны с помощью так называемой леммы о касательной (см. [16]). Огромное преимущество этого подхода состоит в том, что заранее не требуется никакой регулярности функции Φ . Детали и обсуждение связи с транспортными неравенствами см. в [16].

Замечание 5. Оценки, полученные в этом разделе, не зависят от размерности и являются априорными соболевскими глобальными оценками для оптимальной транспортировки. В частности, они могут быть обобщены на случай бесконечномерных мер.

Теорема 8. Предположим, что $D^2W \geq K \cdot \text{Id}$. Для любого единичного вектора e , $p \geq 1$, выполнены оценки

$$K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \|(V_{ee})_+\|_{L^p(\mu)},$$

$$K \|\Phi_{ee}^2\|_{L^p(\mu)} \leq \frac{p+1}{2} \|V_e^2\|_{L^p(\mu)}.$$

Доказательство. Зафиксируем единичный вектор e . Согласно результатам МакКэна [19]: формула замены переменных

$$V(x) = W(\nabla\Phi(x)) - \log \det D_a^2\Phi$$

выполнена для всех точек, за исключением множества A меры нуль: $\mu(A) = 0$. Здесь $D_a^2\Phi$ — обсолютно непрерывная часть второй производной $D^2\Phi$, понимаемой в смысле обобщенных функций (производная Александрова). Имеем

$$V(x+te) - V(x) = W(\nabla\Phi(x+te)) - W(\nabla\Phi(x)) - \log \left[(\det_a D^2\Phi(x))^{-1} \cdot \det_a D^2\Phi(x+te) \right].$$

В силу равномерной выпуклости W :

$$\begin{aligned} & V(x+te) - V(x) \geq \\ & \geq \langle \nabla\Phi(x+te) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle + \\ & \quad + \frac{K}{2} |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 - \\ & \quad - \log \left[(\det_a D^2\Phi(x))^{-1} \cdot \det_a D^2\Phi(x+te) \right]. \end{aligned}$$

Умножим это соотношение на $(\delta_{te}\Phi)^p$, где $p \geq 0$,

$$\delta_{te}\Phi = \Phi(x+te) + \Phi(x-te) - 2\Phi(x),$$

и проинтегрируем по μ . Применим следующую простую лемму

Лемма 3. Пусть $\varphi : A \rightarrow \mathbb{R}$, $\psi : B \rightarrow \mathbb{R}$ — выпуклые функции на выпуклых множествах A , B . Предположим, что $\nabla\psi(B) \subset A$. Тогда

$$\text{div}(\nabla\varphi \circ \nabla\psi) \geq \text{Tr} [D_a^2\varphi(\nabla\psi) \cdot D_a^2\psi] dx \geq 0,$$

где div — дивергенция в смысле обобщенных функций.

Интегрируя по частям и применяя лемму, получаем

$$\begin{aligned} & \int \langle \nabla\Phi(x+te) - \nabla\Phi(x), \nabla W(\nabla\Phi(x)) \rangle (\delta_{te}\Phi)^p d\mu = \\ & = \int \langle \nabla\Phi(x+te) \circ (\nabla\Psi) - x, \nabla W(x) \rangle (\delta_{te}\Phi)^p \circ (\nabla\Psi) dv \geq \\ & \geq \int \left(\text{Tr} [D_a^2\Phi(x+te) \cdot (D_a^2\Phi)^{-1}] \circ (\nabla\Psi) - d \right) (\delta_{te}\Phi)^p \circ \\ & \quad \circ (\nabla\Psi) dv + \\ & + p \int \left\langle \nabla\Phi(x+te) \circ (\nabla\Psi) - x, (D^2\Psi) \nabla\delta_{te}\Phi \circ (\nabla\Psi) \right\rangle \\ & \quad (\delta_{te}\Phi)^{p-1} \circ (\nabla\Psi) dv. \end{aligned}$$

Заметим, что

$$\text{Tr} A - d - \log \det A \geq 0$$

для любого A вида $A = BC$, где B и C симметричны и положительны. Действительно,

$$\begin{aligned} \text{Tr} A - d - \log \det A &= \text{Tr} C^{1/2} B C^{1/2} - d - \\ & - \log \det C^{1/2} B C^{1/2} = \sum_i \lambda_i - 1 - \log \lambda_i, \end{aligned}$$

где λ_i — собственные значения $C^{1/2} B C^{1/2}$.

Таким образом,

$$\begin{aligned} & \int (V(x+te) - V(x)) (\delta_{te}\Phi)^p d\mu \geq \\ & \geq \frac{K}{2} \int |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu + \\ & + p \int \left\langle \nabla\Phi(x+te) - \nabla\Phi(x), (D^2\Psi) \circ \nabla\Phi(x) \nabla\delta_{te}\Phi \right\rangle \\ & \quad (\delta_{te}\Phi)^{p-1} d\mu. \end{aligned}$$

Применим то же неравенство к $-te$ и рассмотрим сумму

$$\begin{aligned} & \int (V(x+te) + V(x-te) - 2V(x))(\delta_{te}\Phi)^p d\mu \geq \\ & \geq \frac{K}{2} \int |\nabla\Phi(x+te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu + \\ & + \frac{K}{2} \int |\nabla\Phi(x-te) - \nabla\Phi(x)|^2 (\delta_{te}\Phi)^p d\mu + \\ & + p \int \langle \nabla\delta_{te}\Phi, (D_a^2\Phi)^{-1}\nabla\delta_{te}\Phi \rangle (\delta_{te}\Phi)^{p-1} d\mu. \end{aligned}$$

Заметим, что последнее слагаемое неотрицательно. Разделив на t^{2p} и перейдя к пределу, получаем

$$\begin{aligned} & \int V_{ee}\Phi_{ee}^p d\mu \geq K \int \|D^2\Phi \cdot e\|^2 \Phi_{ee}^p d\mu + \\ & + p \int \langle (D^2\Phi)^{-1}\nabla\Phi_{ee}, \nabla\Phi_{ee} \rangle \Phi_{ee}^{p-1} d\mu. \quad (7) \end{aligned}$$

Для доказательства первой части заметим, что

$$\int V_{ee}\Phi_{ee}^p d\mu \geq K \int \Phi_{ee}^{p+2} d\mu.$$

Из неравенств Гельдера следует

$$\|(V_{ee})_+\|_{L^{(p+2)/2}(\mu)} \|\Phi_{ee}^p\|_{L^{(p+2)/p}(\mu)} \geq \int V_{ee}\Phi_{ee}^p d\mu.$$

Отсюда вытекает нужный результат.

Для доказательства второй части утверждения применим интегрирование по частям

$$\begin{aligned} & \int V_{ee}\Phi_{ee}^p d\mu = -p \int V_e\Phi_{eee}\Phi_{ee}^{p-1} d\mu + \int V_e^2\Phi_{ee}^p d\mu = \\ & = -p \int \langle \nabla\Phi_{ee}, V_e \cdot e \rangle \Phi_{ee}^{p-1} d\mu + \int V_e^2\Phi_{ee}^p d\mu. \end{aligned}$$

В силу неравенства Коши эта величина не превосходит

$$\begin{aligned} & p \int \langle (D^2\Phi)^{-1}\nabla\Phi_{ee}, \nabla\Phi_{ee} \rangle \Phi_{ee}^{p-1} d\mu + \\ & + \frac{p}{4} \int V_e^2 \langle D^2\Phi e, e \rangle \Phi_{ee}^{p-1} d\mu + \int V_e^2\Phi_{ee}^p d\mu. \end{aligned}$$

Неравенство (7) влечет

$$\frac{p+4}{4} \int V_e^2\Phi_{ee}^p d\mu \geq K \int |\nabla\Phi_e|^2 \Phi_{ee}^p d\mu \geq K \int \Phi_{ee}^{p+2} d\mu.$$

Конец доказательства такой же, как и в первой части.

Следствие 1. В пределе $p \rightarrow \infty$ мы снова получаем теорему Каффарелли:

$$K \|\Phi_{ee}\|_{L^\infty(\mu)}^2 \leq \|(V_{ee})_+\|_{L^\infty(\mu)}.$$

Более сложная оценка для операторной нормы $\|\cdot\|$ также была получена в [16].

Теорема 9. Предположим, что $D^2W \geq K \cdot \text{Id}$. Тогда для любого $r \geq 1$ выполнено

$$K \left(\int \|D^2\Phi\|^{2r} d\mu \right)^{\frac{1}{r}} \leq \left(\int \|(D^2V)_+\|^r d\mu \right)^{\frac{1}{r}}.$$

VII. Сжатие бесконечных мер

В этом разделе мы обсудим результаты о сжатии бесконечных мер. Заметим, что, в отличие от вероятностного случая, здесь нет естественной нормировки мер.

Начнем с одномерного примера.

Пример 1. Пусть $d = 1$, $\mu = \lambda|_{\mathbb{R}^+}$, $\nu = \int_{[0,+\infty)} \rho dx$ и $\rho \geq 1$. Стандартная монотонная транспортировка T является сжатием.

Доказательство. Действительно, это следует из явного представления T :

$$\int_0^T \rho dx = x.$$

Посмотрим, что произойдет в случае $d = 2$ и сферически инвариантной меры-образа.

Пример 2 (F. Morgan). Пусть $d = 2$ и $\mu = \lambda$, $\nu = \Psi(r)dx$. Естественная транспортировка имеет вид

$$T(x) = \varphi(r) \cdot n, n = \frac{x}{r}.$$

Очевидно,

$$\nu(T(B_r)) = 2\pi \int_0^{\varphi(r)} s\Psi(s) dr = \pi r^2 = \mu(B_r).$$

Вычислим DT в базисе (n, v) , где $v = \frac{(-x_2, x_1)}{r}$. Получаем

$$\partial_n T = \varphi' \cdot n, \partial_v T = \frac{\varphi}{r} \cdot v.$$

Очевидно, необходимым и достаточным условием того, чтобы T было сжатием, является

$$\varphi' \leq 1$$

или $\psi' \geq 1$ для $\psi = \varphi^{-1}$. Из формулы замены переменной мы получаем

$$\psi(r) = \sqrt{2 \int_0^r s\Psi(s) ds}.$$

Условие $\psi' \geq 1$ эквивалентно $\int_0^r s\Psi(s) ds \leq \frac{(r\Psi(r))^2}{2}$.

Последнее выполнено, например, если

$$(s\Psi(s))' \geq 1.$$

Действительно, в этом случае

$$\int_0^r s\Psi(s) ds \leq \int_0^r s\Psi(s)(s\Psi(s))' ds = \frac{(r\Psi(r))^2}{2}.$$

Пример 3. Аналогично, если размерность равна d , достаточным условием для того, чтобы

транспортировка $T = \varphi(r) \frac{x}{r}$ меры λ в меру $\Psi(r)dx$ была сжатием, является

$$(r\Psi^{\frac{1}{d-1}}(r))' \geq 1.$$

Следствие 2. В d -мерном евклидовом пространстве с плотностью $\Psi(r)$, удовлетворяющей

$$(r\Psi^{\frac{1}{d-1}}(r))' \geq 1,$$

выполнено евклидово изопериметрическое неравенство.

Некоторые примеры сжимающих отображений естественным образом возникают в геометрии (см. [20], предложения 1.1 и 1.2).

Предложение 1. Пусть M — плоскость, наделенная метрикой

$$dr^2 + g^2(r)r^2d\theta^2$$

(поверхность революции), $g \geq 1$. Тогда тождественное отображение M в евклидову плоскость с мерой gdx является сжатием, сохраняющим объем.

В частности, $\cosh^2(r)dx$ является липшицевым образом H^2 (с метрикой $dr^2 + \cosh^2(r)d\theta^2$).

Следующая теорема сравнения была получена в [17]. Оказывается, что естественная модель логарифмически вогнутого распределения на прямой имеет следующий вид:

$$\nu_A = \frac{dx}{\cos Ax}, \quad -\frac{\pi}{2A} < x < \frac{\pi}{2A}.$$

Его потенциал V удовлетворяет условию $V''e^{-2V} = A^2$. Используя результат [25] о симметрии изопериметрических множеств, несложно вычислить изопериметрическую функцию ν_A :

$$\mathcal{I}_{\nu_A}(t) = e^{At/2} + e^{-At/2}.$$

Предложение 2. Пусть $\mu = e^W dx$ — мера на \mathbb{R}^1 с четным выпуклым потенциалом W . Предположим, что

$$W''e^{-2W} \geq A^2$$

и $W(0) = 0$. Тогда μ является образом ν_A при 1-липшицевом возрастающем отображении.

Доказательство. Без потери общности можно предположить, что W — гладкая функция и $W''e^{-2W} > A^2$. Пусть φ — выпуклый потенциал, т.ч. $T = \varphi'$ отображает μ в ν_A . Кроме этого, мы требуем, чтобы отображение T было нечетным. Очевидно, φ' удовлетворяет уравнению

$$e^W = \frac{\varphi''}{\cos A\varphi'}.$$

Пусть x_0 — точка локального максимума для φ'' . Тогда в этой точке

$$\varphi^{(3)}(x_0) = 0, \quad \varphi^{(4)}(x_0) \geq 0.$$

Дифференцируя формулу замены переменной в x_0 дважды, мы получаем

$$W'' = \frac{\varphi^{(4)}}{\varphi''} - \left(\frac{\varphi^{(3)}}{\varphi''}\right)^2 + \frac{A^2}{\cos^2 A\varphi'}(\varphi'')^2 + A \frac{\sin A\varphi'}{\cos A\varphi'} \varphi''''.$$

Следовательно, в точке x_0

$$W'' \leq \frac{A^2}{\cos^2 A\varphi'}(\varphi'')^2 = A^2 e^{2W}.$$

Но это противоречит основному предположению.

Таким образом, φ'' не имеет локального максимума. Заметим, что φ — выпуклая функция. Из этого следует, что 0 является точкой глобального минимума φ'' . Тогда $\varphi'' \geq \varphi''(0) = 1$. Очевидно, T^{-1} является искомым отображением.

VIII. Другие результаты и приложения

Немедленным следствием теоремы о сжатии является теорема сравнения Багри–Леду, вероятностный аналог теоремы сравнения Леви–Громова для многообразий положительной кривизны Риччи.

Теорема 10. Предположим, что $\mu = e^{-V} dx$, где $D^2V \geq Id$ — вероятностная мера на \mathbb{R}^d . Тогда

$$\mathcal{I}_\mu \geq \mathcal{I}_\gamma,$$

где γ — стандартная гауссовская мера.

Таким же образом теорема о сжатии применима к различным функциональным неравенствам и неравенствам концентрации для логарифмически вогнутых мер (логарифмическое неравенство Соболева, неравенство Пуанкаре и т.д.).

Следующая нерешенная проблема известна под названием *гауссова корреляционная гипотеза*.

Гауссова корреляционная гипотеза. Пусть A и B — симметрические выпуклые множества, а γ — стандартная гауссовская мера. Тогда

$$\gamma(A \cap B) \geq \gamma(A)\gamma(B). \tag{8}$$

Гауссова корреляционная гипотеза возникла в 70-е годы. Основные положительные результаты — неравенство верно в двумерном случае и в случае, когда одно из множеств — эллипсоид. Случай эллипсоида был доказан Ж. Арже ([10]), а транспортное решение получено Д. Кордеро–Ераскином [8].

Предложение 3. Пусть B — эллипсоид. Тогда (8) выполнено.

Доказательство. Применив линейное преобразование мер, можно свести неравенство к случаю, когда B — шар, а γ некоторая гауссова мера. Рассмотрим оптимальную транспортировку T между γ и $\gamma_A = \frac{1}{\gamma(A)}\gamma|_A$. По теореме 2 T — сжатие. В силу симметрии $T(0) = 0$. Таким образом, $T(B) \subset B$ и

$$\frac{\gamma(A \cap B)}{\gamma(A)} = \gamma_A(B) = \gamma(T^{-1}(B)) \geq \gamma(B).$$

Теорема доказана.

Следующее красивое наблюдение [11] следует из теоремы о сжатии и свойств полугруппы Орнштейна–Уленбека.

Теорема 11. Если γ — стандартная гауссовская мера, g — симметричная выпуклая функция, f — симметричная логарифмически вогнутая функция, то

$$\int f g d\gamma \leq \int f d\gamma \cdot \int g d\gamma.$$

Доказательство. Пусть $T(x) = x + \nabla\varphi(x)$ — оптимальная транспортировка γ в $\frac{f \cdot \gamma}{\int f d\gamma}$. Достаточно доказать, что

$$\int g(x + \nabla\varphi(x)) d\gamma \leq \int g d\gamma.$$

Положим

$$\psi(t) = \int g(x + P_t(\nabla\varphi(x))) d\gamma,$$

где $P_t = e^{tL}$ — полугруппа Орнштейна–Уленбека с генератором $L = \Delta - \langle x, \nabla \rangle$. Заметим, что

$$\begin{aligned} \frac{\partial}{\partial t} \psi(t) &= \frac{\partial}{\partial t} \int g(x + P_t(\nabla\varphi(x))) d\gamma = \\ &= \int \langle \nabla g(x + P_t(\nabla\varphi(x))), L P_t(\nabla\varphi(x)) \rangle d\gamma. \end{aligned}$$

Интегрируя по частям, получаем

$$\frac{\partial}{\partial t} \psi(t) = - \int \text{Tr} \left[D^2 g(x + P_t(\nabla\varphi(x))) \cdot (I + M) M \right] d\gamma,$$

где

$$M = D P_t(\nabla\varphi(x)) = e^{-t/2} P_t(D^2\varphi).$$

В силу теоремы о сжатии $I + M \geq 0$ и $M \leq 0$. Тогда $\text{Tr} \left[D^2 g \cdot (I + M) M \right] \leq 0$ и функция $\psi(t)$ возрастает. Заметим, что $P_{+\infty}(\nabla\varphi) = \int \nabla\varphi d\gamma = \frac{\int x f d\gamma}{\int f d\gamma} = 0$. Таким образом, $\int g(x + \nabla\varphi(x)) d\gamma \leq \psi(+\infty) = \int g d\gamma$. Теорема доказана.

Некоторые другие приложения к корреляционным неравенствам были получены в [8, 13]. Обобщение предложения 3 на негауссовы меры было получено в [13] (см. следствие 4.1).

Другие приложения, полученные в работах [6, 8, 10, 13], касаются неравенств вида

$$\int \Gamma(x) d\mu \leq \int \Gamma(x) d\nu,$$

где $\Gamma(x)$ — выпуклая функция (неравенства моментов и т.д.).

Следующая теорема была получена в [7] с помощью теоремы о сжатии. В частности, она дает положительное решение для так называемой (В)-гипотезы для гауссовских мер.

Теорема 12. Пусть K — симметричное выпуклое множество и γ — стандартная гауссовская мера. Тогда функция

$$t \rightarrow \gamma(e^t K)$$

— логарифмически вогнутая. В частности, $\gamma(\sqrt{ab}K)^2 \geq \gamma(aK)\gamma(bK)$ для любых $a > 0, b > 0$.

Заметим, что кроме результатов из предыдущего раздела ничего не известно о сжимающих отображениях многообразий.

Следующий результат был получен С.И. Вальдимарссоном (см. [26]). Пусть M — неотрицательная симметричная матрица. Обозначим через γ_M гауссову меру с плотностью

$$\sqrt{\det M} e^{-\pi \langle Mx, x \rangle}.$$

Теорема 13. Пусть A, G и B — положительно определенные линейные преобразования, $A < G$, $GB = BG$, H — выпуклая функция, а μ_0 — вероятностная мера. Оптимальная транспортировка $T = \nabla\Phi$ вероятностных мер

$\mu = \gamma_{B^{-1/2}GB^{-1/2}} * \mu_0$ и $\nu = \text{Ce}^{-H} \cdot \gamma_{B^{-1/2}A^{-1}B^{-1/2}}$ удовлетворяет отношению

$$D^2\Phi \leq G.$$

Специальный вид меры μ позволил Вальдимарссону (см. также работу Ф. Барта [3]) получить с помощью транспортных аргументов новую форму неравенства Браскампа–Либа. См. детали в [26].

Мы закончим раздел следующим наблюдением из [4].

Предложение 4. Пусть $\mu = I_{[0,+\infty)} e^{-x} dx$ — односторонняя экспоненциальная мера $\nu = e^g \cdot \mu$, удовлетворяющая $|g'| \leq c$ для некоторого $c < 1$. Монотонная транспортировка T , отображающая ν в μ , удовлетворяет

$$T'(x) \in [1 - c, 1 + c]$$

для всех $x \in [0, \infty)$. Обратное отображение $S = T^{-1}$ является $\frac{1}{1-c}$ -сжатием.

Результат вытекает из точного представления T , но может быть эвристически доказан с помощью принципа максимума, примененного к S . Действительно,

$$g(S) - S + \log S' = -x.$$

Если x_0 — точка максимума для S' , имеем $S''(x_0) = 0$. Кроме этого,

$$g'(S(x_0))S'(x_0) - S'(x_0) + \frac{S''(x_0)}{S'(x_0)} = -1.$$

Очевидно, $S'(x_0) = \frac{1}{1-g'(S(x_0))} \leq \frac{1}{1-c}$.

Используя это свойство, можно транспортными методами доказать 1-мерное неравенство Талагранна для показательных распределений (см. [4], предложение 6.6).

Работа написана при поддержке грантов РФФИ 07-01-00536, РФФИ 08-01-90431-Укр и Гранта

ДААД А1008062. Автор благодарен Фрэнку Моргану и Эмануэлю Мильману за поддержку и ценные замечания.

Литература

1. *Богачев В.И.* Гауссовские меры. — М.: Наука, 1997.
2. *Bakry D., Ledoux M.* Lévy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator // *Invent. Math.* — 2005. — V. 123(1). — P. 259–281.
3. *Barthe F.* On a reverse form of the Brascamp–Lieb inequality // *Invent. Math.* — 1998. — V. 2. — P. 335–361.
4. *Barthe F., Kolesnikov A.V.* Mass transport and variants of the logarithmic Sobolev inequality // *Journal. Geom. Analysis.* — 2008. — V. 18(4) — P. 921–979.
5. *Bobkov S.* On isoperimetric constants for log-concave probability distributions. In *Geometric aspects of functional analysis // Israel Seminar 2004-2005, volume 1910 of Lecture Notes in Math.* — P. 81–88.
6. *Caffarelli L.A.* Monotonicity properties of optimal transportation and the FKG and related inequalities // *Comm. Math. Phys.* — 2000. — V. 214(3) — P. 547–563.
7. *Cordero-Erausquin D., Fradelizi M., Maurey B.* The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems // *J. Funct. Anal.* — 2004. — V. 214. — P. 410–427.
8. *Cordero-Erausquin D.* Some applications of mass transport to Gaussian type inequalities // *Arch. Rational Mech. Anal.* — 2002. — V. 161. — P. 257–269.
9. *Gromov M.* Metric structure for Riemannian and non-Riemannian spaces // *Birkhäuser — Boston*, 1998. — V. 152.
10. *Hargé G.* A particular case of correlation inequality for the Gaussian measure // *Ann. Probab.* — 1999. — V. 27. — P. 1939–1951.
11. *Hargé G.* A convex / log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces // *Probab. Theory Related Fields.* — 2004. — V. 13, N. 3 — P. 415–440.
12. *Kannan R., Lovász L., Simonovits S.* Isoperimetric problems for convex bodies and a localization lemma // *Discrete Comput. Geom.* — 1995. — V. 13(3-4) — P. 541–559.
13. *Kim Y.-H., Milman E.* A Generalization of Caffarelli's Contraction Theorem via (reverse) Heat Flow. — arXiv:1002.0373.
14. *Kolesnikov A.V.* On diffusion semigroups preserving the log-concavity // *J. Funct. Anal.* — 2001. — V. 186.
15. *Kolesnikov A.V.* On global Hölder estimates of optimal transportation. — arXiv: 0810.5043.
16. *Kolesnikov A.V.* On Sobolev regularity of mass transport and transportation inequalities. — arXiv:1007.1103.
17. *Kolesnikov A.V., Zhdanov R.I.* On isoperimetric sets of radially symmetric measures. — arXiv:1002.1829.
18. *Ledoux M.* The concentration of measure phenomenon. *Mathematical Surveys and Monographs 89 // Amer. Math. Soc.* — 2001.
19. *McCann R.J.* A convexity principle for interacting gases // *Adv. Math.* — 1997. — V. 128. — P. 153–179.
20. *Maurmann Q., Morgan F.* Isoperimetric comparison theorems for manifolds with density // *Calculus of Variations and Partial Differential Equations* — 2009. — V. 36, N. 1. — P. 1–5.
21. *Milman E.* On the role of Convexity in Functional and Isoperimetric Inequalities // *Proc. London Math. Soc.* — 2009. — V. 99 (3) — P. 32–66.
22. *Milman V., Schechtman G.* Asymptotic theory of finite dimensional normed vector spaces. *Lect Notes in Math. // Springer.* — 1986.
23. *Milman E., Sodin S.* An isoperimetric inequality for uniformly log-concave measures and uniformly convex bodies // *Jour. Funct. Anal.* — 2008. — V. 254(5). — P. 1235–1268.
24. *Ros A.* The isoperimetric problem. Lecture at Clay Mathematical Institute on the Global Theory of Minimal Surfaces. — 2001.
25. *Rosales C., Cañete A., Bayle V., Morgan F.* On the isoperimetric problem in Euclidean space with density // *Calc.* — 2007. — V. 31. — P. 27–46.
26. *Valdimarsson S.I.* On the Hessian of optimal transport potential // *Ann. Sc. Norm. Sup. Pisa Cl Sci.* — 2007. — V. 6(3) — P. 441–456.
27. *Villani C.* *Topics in Optimal Transportation // Amer. Math. Soc. Providence.* — Rhode Island, 2003.

Поступила в редакцию 15.10.2010.

А.Б. Куржанский^{1,2}, А.А. Куржанский², П. Варайя²

¹ Московский государственный университет им. М.В. Ломоносова

² University of California at Berkeley, USA

Роль макро моделирования в активном управлении транспортной сетью

Система активного управления транспортной сетью (АУТС) нацелена на динамическое регулирование автомобильных потоков на дорогах в режимах регулярно повторяющихся и внезапно возникающих перегрузок. Её работа основана на измерении и оценивании текущей ситуации на дороге и направлена на повышение эффективности использования сети дорог в условиях перегрузок. Она представляет собой непрерывный процесс выполнения следующих операций: (1) получения и анализа данных измерений потоков транспорта; (2) оперативного планирования и компьютерной симуляции различных сценариев и управляющих стратегий; (3) внедрения наиболее подходящих управленческих решений; (4) обеспечения в реальном времени работы системы поддержки принятия решений, отфильтровывающей текущие измерения; это делается с целью предсказания ожидаемого состояния транспортной сети на ближайшее время и подготовки к внедрению соответствующих наилучших управленческих решений к моменту возможного возникновения ожидаемой критической ситуации. АУТС опирается на работу быстрого и надёжного симулятора дорожного движения, который способен в течение нескольких минут протестировать множество сценариев и управленческих стратегий для наблюдаемой дорожной сети и выбрать адекватные для использования. Для реализации подобной схемы действий возможно использовать бесплатный программный пакет для макросимуляций — систему Aurora Road Network Modeler. В настоящей статье изложено описание следующих элементов АУТС: динамической модели транспортных потоков, лежащей в её основе; приспособления и подготовки модели к процессу симуляции; критериев эффективности транспортной сети и оценки сценариев как части оперативного планирования; указания рамок, в пределах которых оцениваются управленческие решения, и их моделирование; предлагаемых алгоритмов оценивания состояния транспортной сети и предсказания её будущего поведения на ближайших коротких отрезках времени.

Ключевые слова: активное управление, транспортная сеть, макроскопическая транспортная модель, калибровка, автоматизированная система, геоинформационная система, фундаментальная диаграмма, коммуникационная сеть, быстрый и надёжный симулятор, оперативное планирование, динамическая фильтрация, оперативная стратегия, оценивание состояния, гарантированное прогнозирование.

1. Введение

Автомобильные заторы в городах являются причиной длительных задержек, пониженной производительности труда, неэкономного расхода бензина и чрезмерного загрязнения окружающей среды. Стремительный рост числа автомобилей на дорогах, стоимость земли и проблемы экологии делают задачу расширения существующих дорог и строительство новых в густонаселённых городах всё более трудно выполнимой. Активное управление транспортной сетью (АУТС) — это эффективное средство борьбы с повторяющимися изо дня в день и внезапно возникающими заторами, учитывающая конкретные для данного времени года и суток условия и не требующая физического увеличения пропускной способности дорог. АУТС использует автоматизированные системы управления дорожным движением (АСУД), а также вмешательство диспетчера для по возможности оптимальной организации работы транспортной сети. В данной статье предлагается определённая

структура АУТС и подробно рассматриваются его отдельные составляющие. В основе статьи лежит опыт, полученный авторами в результате участия в исследовательском проекте TOPL [1], проводимом в Берклейском отделении Калифорнийского университета в США, начиная с 2006 года. Цель проекта TOPL состоит в том, чтобы помочь Департаменту Транспорта штата Калифорния оптимизировать эксплуатацию особенно загруженных магистралей штата.

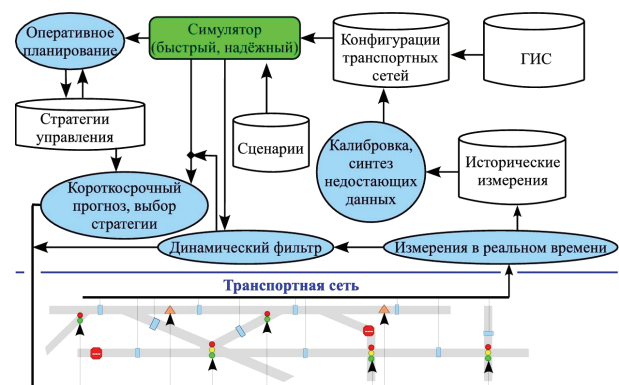


Рис. 1. Схема работы АУТС

АУТС представляет собой циклический процесс, состоящий из следующих факторов: (1) измерения транспортного потока и анализа полученных данных, без которых управление транспортной сетью будет беспомощным и практически невозможным; (2) оперативного планирования, которое включает изучение поведения транспортной сети при возможных сценариях (таких, например, как плохая погода, ДТП, дорожные работы, специальные мероприятия, увеличение спроса и т.п.), а также разработку стратегий управления, призванных увеличить продуктивность транспортной сети и оценку пригодности разработанных стратегий с точки зрения их стоимости и потенциальной эффективности; (3) внедрения наиболее обещающих стратегий непосредственно на магистралях путём установки необходимого оборудования и программного обеспечения; (4) работающей системы поддержки принятия решений, которая осуществляет фильтрацию поступающих в реальном времени измерений, прогнозирует поведение транспортной системы на ближайшие несколько часов и помогает диспетчеру выбрать стратегию управления, наиболее соответствующую ситуации.

Схема работы АУТС представлена на рис. 1. В основе её лежит коммуникационная сеть, по которой в центр сбора информации и управления пересылаются данные измерений с дорожных датчиков, и в свою очередь команды АСУД или диспетчеров передаются на управляющие устройства, установленные на перекрёстках городских улиц и автостадах. Центральным элементом схемы АУТС является «быстрый и надёжный симулятор». Надёжность симулятора обуславливается теоретически обоснованной транспортной моделью, адекватно отображающей ситуацию на дороге, для идентификации параметров которой достаточно лишь существующих измерений. В качестве кандидата в такие симуляторы нами предлагается система Aurora Road Network Modeler (RNM) [2], бесплатный программный пакет для моделирования и симуляций автомобильного движения на автостадах и городских улицах со светофорами на перекрёстках. Используемая в этом симуляторе макроскопическая динамическая транспортная модель, а также метод её построения при помощи данных геоинформационных систем (ГИС) и калибровки на основании данных измерений описываются в параграфе 2.

Симулятор может работать в трёх режимах. Первый из них, режим оперативного планирования, используется для прогона большого числа симуляций, тестирующих различные сценарии и потенциальные меры по улучшению дорожного движения. Сценарии могут включать перераспределение транспортных потоков из-за возможных дорожных работ, особых мероприятий или ДТП, уменьшение максимальной пропускной способности дорог из-за плохой погоды и т.п. Конфигу-

рации дорожных сетей и структуры возможных систем управления сохраняются в соответствующих базах данных, которые используются в механизме поддержки принятия решений. Параграф 3 затрагивает тему оперативного планирования, а параграф 4 посвящён моделированию управления транспортным потоком. Второй режим работы — это динамическая фильтрация: симуляция системы с неопределённостью в параметрах модели и входных данных. Она используется для корректировки измерений, поступающих в реальном времени с дорожных датчиков. Отфильтрованные через динамическую модель измерения используются для организации обратной связи в соответствующих алгоритмах управления транспортным потоком. В третьем режиме прогнозирования ближайшего будущего и выбора оперативной стратегии, исходя из начальных условий, полученных фильтрацией текущих измерений, вычисляются множества достижимости состояния транспортной сети при вероятных сценариях и возможных стратегиях управления для ближайших нескольких часов. Диспетчер выбирает наиболее подходящую из предлагаемых оперативных стратегий и вводит её в действие посылкой соответствующих команд на управляющие устройства. Параграф 5 описывает алгоритмы гарантированного прогнозирования и оценивания состояния транспортной сети. В параграфе 6 приведены отдельные элементы АУТС на примере автострады I-80 в Северной Калифорнии.

II. Динамическая модель транспортной сети

II.1. Описание динамической модели

Начнём с описания динамической модели транспорта — вариации макроскопической модели Даганзо [3, 4], на базе которой основан предлагаемый надёжный симулятор.

Транспортная сеть состоит из направленных соединений и узлов, где соединения представляют сегменты дорог, а узлы служат коммутаторами соединений. Обозначим множество соединений через \mathcal{L} , а множество узлов — через \mathcal{N} . Узел обязан иметь минимум одно входное и минимум одно выходное соединение. Соединение называется обычным, если у него присутствуют начальный и конечный узлы. Соединение без начального узла именуется далее как въезд, соединение без конечного узла — соответственно как выезд.

В табл. 1 приведены параметры модели с указанием возможных диапазонов их значений, а также используемые переменные. Каждое соединение $l \in \mathcal{L}$ характеризуется длиной Δx_l , количеством полос k_l , своей фундаментальной диаграммой (отражающей максимальную пропускную способность F_l , свободную скорость v_l и ско-

рость распространения затора w_l (см. рис. 2)²¹, и учитывающей число полос. Въезды — это соединения, через которые автомобили попадают в систему. С ними ассоциируются функции требуемых потоков $r_l(\cdot)$. Каждый узел $\nu \in \mathcal{N}$ характеризуется распределительной матрицей $\mathcal{B}_\nu(\cdot)$, устанавли-

вающей, каким образом входящие в узел потоки распределяются между выходящими из узла соединениями. Узлы могут использоваться не только для моделирования перекрёстков, точек слияния и разветвления дорог, но и просто для разбиения длинных соединений на более короткие.

Т а б л и ц а 1

Параметры модели и переменные

Символ	Описание	Возможное значение	Размерность
Δx_l	длина соединения $l \in \mathcal{L}$	$\in [0,05,0,4]$	км
k_l	число полос в соединении $l \in \mathcal{L}$		полосы
Δt	размер временного шага	$\in [\frac{1}{3600}, \frac{1}{360}]$	часы
F_l	максимальная пропускная способность $l \in \mathcal{L}$	для 1 полосы $\in [1500, 2000]$	авто/ч
v_l	свободная скорость в $l \in \mathcal{L}$	для автострад $\in [90, 130]$; для улиц $\in [30, 70]$	км/ч
w_l	скорость распространения затора в $l \in \mathcal{L}$	$\in [10, 40]$	км/ч
$\bar{\rho}_l$	максимальная плотность автомобилей в $l \in \mathcal{L}$	для 1 полосы $\in [95, 125]$	авто/км
ρ^c_l	критическая плотность автомобилей в $l \in \mathcal{L}$	для 1 полосы $\in [12, 30]$	авто/км
t	время	$\in [0, \infty)$	часы
$\mathcal{B}_\nu(t)$	распределительная матрица в узле $\nu \in \mathcal{N}$	$\beta_{ij}(t) \in [0, 1]$, $\sum_j \beta_{ij}(t) = 1$	безразмерна
$r_l(t)$	требуемый поток на въезде $l \in \mathcal{L}$	переменная	авто/ч
$f^u_l(t)$	фактический поток, входящий в $l \in \mathcal{L}$	переменная	авто/ч
$f^d_l(t)$	фактический поток, исходящий из $l \in \mathcal{L}$	переменная	авто/ч
$V_l(t)$	средняя скорость потока в $l \in \mathcal{L}$	переменная	км/ч
$\rho_l(t)$	плотность автомобилей в $l \in \mathcal{L}$	переменная	авто/км
$\rho(t)$	вектор плотностей из всех соединений в \mathcal{L}	переменная	авто/ч

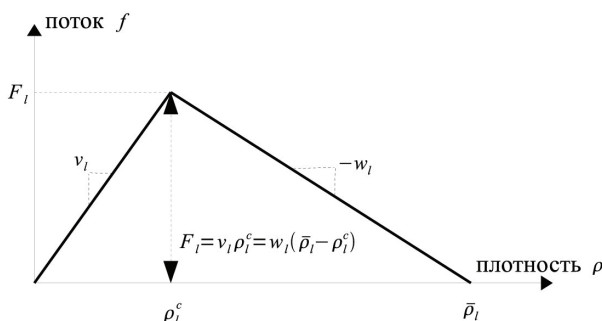


Рис. 2. Фундаментальная диаграмма для соединения l

В простом примере автострады с одним направлением, изображённом на рис. 3, узлы нахо-

дятся либо в точках, где въезды соединяются с автострадой, а выезды ответвляются от неё, либо в точках, где меняются свойства автострады, определяемые фундаментальной диаграммой. В данном примере обычные соединения — это отрезки автострады, идущие от одного узла до следующего, въезды соответствуют попаданию на автостраду, выезды — съездам с неё.

Состояние транспортной сети во время t определяется плотностью движущихся объектов (то есть автомобилей) в каждом соединении: $\rho_l(t)$, $l \in \mathcal{L}$. При начальных данных $\rho_l(t_0) = \rho_l^0$, обычно определяемых путём измерений, система развива-

²¹Фундаментальная диаграмма — это функция потока от плотности, обычно вогнутая. В нашем случае фундаментальная диаграмма имеет треугольную форму, как на рис. 2, определяемую тремя значениями: максимальной пропускной способностью F_l , свободной скоростью (средней скоростью потока, измеряемой при малой плотности автомобилей) v_l и скоростью распространения затора (скорость, с которой волна затора распространяется в направлении обратном движению автомобильного потока) w_l . Альтернативно, треугольная фундаментальная диаграмма задаётся тройкой значений: максимальной пропускной способностью F_l , критической плотностью (количеством автомобилей на километр в соединении, при котором поток достигает максимальной пропускной способности) ρ^c_l и максимальной плотностью (количеством автомобилей на километр в соединении, при котором движение полностью останавливается) $\bar{\rho}_l$. Хотя в данной статье мы имеем дело с треугольной формой, возможны и другие формы фундаментальной диаграммы, например, парабола [5], трапеция [6], или кусочно-линейная, так называемая «развёрнутая лямбда» [7, 8].

ется во времени соответственно уравнению

$$\rho_l(t + \Delta t) = \rho_l(t) + \frac{\Delta t}{\Delta x_l} (f_l^u(t) - f_l^d(t))$$

для всех $l \in \mathcal{L}$, (1)

где Δt должно удовлетворять условию

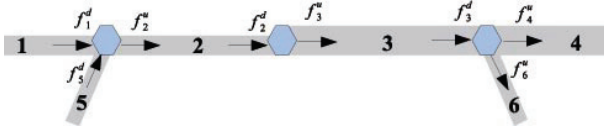


Рис. 3. Простая транспортная сеть: соединения пронумерованы от 1 до 6, узлы обозначены шестиугольниками

$\Delta t \leq \min_l \left\{ \frac{\Delta x_l}{v_l} \right\}$ и $\Delta t \leq \min_l \left\{ \frac{\Delta x_l}{w_l} \right\}$.²² Для въездов $f_l^u(t) = r_l(t)$, где $r_l(t)$ — значение требуемого потока во время t . Для выездов $f_l^d(t) = v_l \rho_l(t) \min \left\{ 1, \frac{F_l}{v_l \rho_l(t)} \right\}$. Для обычного соединения 1 поток $f_l^u(t)$ определяется в соответствующем ему начальном узле, а поток $f_l^d(t)$ — в конечном.

Узел с m входными и n выходными соединениями имеет распределительную матрицу размерности $m \times n$: $B_\nu(t) = \{\beta_{ij}(t)\}_{i=1, \dots, m}^{j=1, \dots, n}$. Эта матрица имеет неотрицательные элементы, лежащие в интервале $[0, 1]$, причём сумма элементов каждого ряда равна 1. Элемент $\beta_{ij}(t)$ определяет долю потока, исходящего из входного соединения i и ответвляющегося в направлении выходного соединения j в момент t . Входные потоки $f_i^d(t)$ и выходные потоки $f_j^u(t)$, $i = 1, \dots, m$, $j = 1, \dots, n$, вычисляются согласно приведённому ниже алгоритму.

Процедура вычисления входных и выходных потоков узла

1. Вычислим фактический допускаемый приём для каждого выходного соединения (какой поток готово принять данное выходное соединение):

$$s_j(t) = \min \{ F_j, w_j (\bar{\rho}_j - \rho_j(t)) \}, j = 1, \dots, n. \tag{2}$$

Здесь индекс j обозначает номер выходного соединения.

2. Установим итерацию $p = 0$.
3. Вычислим поступающий поток для каждого входного соединения:

$$d_i^{[0]}(t) = v_i \rho_i(t) \min \left\{ 1, \frac{F_i}{v_i \rho_i(t)} \right\}, i = 1, \dots, m, \tag{3}$$

где индекс i обозначает номер входного соединения. Переменная $d_i^{[0]}(t)$ обозначает поступающий поток, требующий выхода из входного соединения i .

4. Вычислим требующий вливания поток для каждого выхода:

$$d_j^{*[p]}(t) = \sum_{i=1}^m \beta_{ij}(t) d_i^{[p]}(t), j = 1, \dots, n. \tag{4}$$

Переменная $d_j^{*[0]}(t)$ обозначает полный поток, требующий вливания в выходное соединение j .

5. Для $p = 1, \dots, n$, и для каждого входного соединения $i = 1, \dots, m$, повторяем

- (a) при необходимости сократим требуемые входные потоки так, чтобы они не превышали фактический допускаемый приём для соответствующего выходного соединения:

$$d_i^{[p]}(t) = \begin{cases} d_i^{[p-1]}(t), & \text{если } \beta_{ip}(t) = 0, \\ d_i^{[p-1]}(t) \min \left\{ 1, \frac{s_p(t)}{d_p^{*[p-1]}(t)} \right\}, & \\ \text{иначе;} & \end{cases} \tag{5}$$

- (b) пересчитаем требуемые выходные потоки $d_j^{*[p]}(t)$, $j = 1, \dots, n$, по формуле (4).

На этом шаге использовано правило «пропорциональных приоритетов» для сходящихся соединений и правило очереди (первый вошёл, первый вышел) для расходящихся соединений так, как они сформулированы в [4]²³.

6. Поток, исходящий из входного соединения i :

$$f_i^d(t) = d_i^{[n]}(t), i = 1, \dots, m. \tag{6}$$

7. Поток, входящий в выходное соединение j :

$$f_j^u(t) = \sum_{i=1}^m \beta_{ij}(t) d_i^{[n]}(t), j = 1, \dots, n. \tag{7}$$

При начальной позиции $t = t_0$, $\rho(t_0) = \rho_0$ на каждом временном шаге Δt эволюционирующей системы вычисляются входные и выходные потоки для всех узлов сети; затем состояние системы обновляется по закону сохранения (1).

II.2. Построение модели

²²Это необходимое условие сходимости при численном решении гиперболических уравнений в частных производных, известное как условие Куранта–Фридрихса–Леви (КФЛ) [9].

²³Правило пропорциональных приоритетов означает, что каждое выходное соединение принимает автомобили из входных потоков пропорционально интенсивности этих потоков. Правило очереди означает, что отношение доли входящего потока из соединения i , распределённой в выходное соединение j , к входящему потоку $f_i^d(t)$ должно равняться соответствующему элементу распределительной матрицы узла $\beta_{ij}(t)$.

Построение динамической модели, готовой для симуляций, требует следующего: (1) составления дорожной сети с верным расположением узлов, длиной соединений и числом полос в соединениях; (2) калибровки системы, то есть идентификации фундаментальных диаграмм для всех соединений; (3) определения требуемых потоков для въездов и распределительных матриц для узлов. Эта работа является, как правило, весьма трудоёмкой, так как практически не имеется универсальных информационных систем, собирающих в едином месте те данные, что необходимы для выполнения всех трёх задач.

Для начала необходимо получить картографическую информацию об интересующих нас дорогах через доступные коммерческие [10, 11] или бесплатные [12] геоинформационные системы (ГИС). Данные ГИС содержатся в виде специальных файлов с более или менее детальным описанием дорожных сегментов, которые могут быть использованы для составления сети из узлов и соединений. В пакет *Auroga RNM* [2] входит программа *GIS Importer*, конвертирующая данные ГИС в форму соединения-узлы. В особых случаях, например, для автострад Калифорнии или автострад города Портленд, штат Орегон, существуют общедоступные информационные системы *PeMS* [13] и *PORTAL* [14], содержащие как картографические данные, так и данные измерений с дорожных датчиков.

Данные измерений с автострадных датчиков, предлагаемые такими информационными системами, как *PeMS* или *PORTAL*, позволяют идентифицировать фундаментальные диаграммы для соответствующих соединений автострады. Полное описание алгоритма калибровки дано в [15]. Здесь мы предлагаем сжатое описание.

1. Для каждого надёжного²⁴ датчика на выбранном участке автострады извлечём данные измерений плотности и потока. Обычно расстояние между датчиками бывает больше, чем длина соединений в нашей модели. Поэтому извлечённые данные измерений могут относиться (и чаще всего относятся) к более чем одному соединению.
2. Среди измерений найдём максимальный поток. Обычно этот измеренный максимальный поток и берётся за значение параметра F_l .
3. Используя метод наименьших квадратов для точек, определённых парой плотность–поток, на плоскости найдём свободную скорость v_l . Установим критическую плотность $\rho_l^c = \frac{F_l}{v_l}$. На практике обычно прямая линия хорошо аппроксимирует множество точек плотность–поток, когда плотность ниже критической.

4. Используя метод наименьших квадратов с ограничениями, найдём скорость распространения затора w_l .

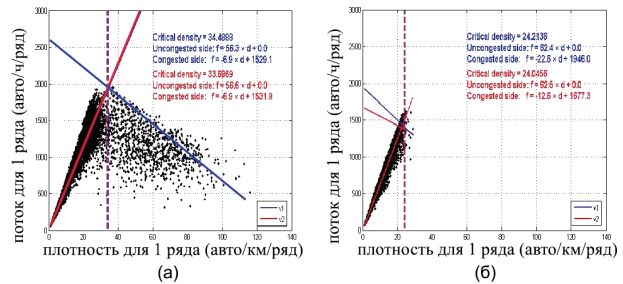


Рис. 4. Идентификация фундаментальной диаграммы: (а) данные высокого качества; (б) данные низкого качества

Если данных измерений достаточно и они поступают с исправных датчиков, то калибровочные шаги 1–3 дают хороший результат (рис. 4а). Однако при низком качестве измерений (например, ввиду поломки датчика либо просто потому, что на данном участке дороги поток не достигает максимальной пропускной способности (рис. 4б)) необходимо использовать фундаментальные диаграммы соседних участков либо синтезировать недостающие данные (например, при помощи способа, предложенного в [16]) и повторить шаги 1–3. Изображение на рис. 4 иллюстрирует тот факт, что измерения в зоне свободного движения фундаментальной диаграммы хорошо аппроксимируются прямой линией, в то время как измерения в зоне затора (правая часть диаграммы) имеют больший разброс, что в свою очередь может служить обоснованием альтернативных форм фундаментальной диаграммы (например, в виде трапеции или «развёрнутой лямбды»).

Наиболее трудная часть калибровки соединений, моделирующих городские улицы, состоит в оценивании их максимальной пропускной способности. Если максимальная пропускная способность уличного соединения определена (из измерений или соображений здравого смысла), то свободную скорость можно установить равной максимальной пропускной способности, а максимальную плотность вычислить путём оценки максимального числа автомобилей, которые могут разместиться в данном соединении с последующим делением этого числа на длину соединения.

Наконец, остаётся определить функции требуемого потока для въездов и распределительные матрицы для узлов. Для автострад это осуществляется при помощи данных информационных систем типа *PeMS* и *PORTAL*, предоставляющих измерения потоков на въездах, которые можно брать за основу при построении функций требуемого потока. Элементы распределительных матриц могут быть вычислены, исходя из измерений потоков на главной магистрали и съездах: в примере, показанном на рис. 3, это будут измерения

²⁴*PeMS* [13], например, предоставляет ежедневный бюллетень состояния для всех зарегистрированных датчиков.

потоков f^u_4 и f^u_6 . На практике, однако, измерений потоков на въездах и съездах может не быть, что, к примеру, имеет место на ряде калифорнийских автострад. В таких случаях требуемые потоки на въездах и распределительные матрицы должны быть синтезированы таким образом, чтобы потоки на основной магистрали, полученные в результате работы динамической модели, соответствовали измерениям потоков на основной магистрали. Это требует решения обратной задачи, являющейся некорректной по Адамару. Решение данной задачи неединственно. Для регуляризации её решения необходимо вводить дополнительные ограничения на максимальные требуемые потоки, их вариации и значения элементов распределительных матриц. Задача синтеза недостающих данных на автострадных въездах и съездах освещена в [17].

III. Оперативное планирование

Цель оперативного планирования для транспорта заключается в разработке стратегий управления, призванных улучшить продуктивность перегруженных транспортных сетей. В основе оперативного планирования лежат измерения и базирующийся на них анализ этой продуктивности. Подобный анализ выделяет узкие места, являющиеся причинами заторов, обозначает участки дорог с повышенным риском ДТП и позволяет наметить главные элементы управляющих стратегий, которые должны облегчить нагрузку на транспортную сеть.

Выработанные стратегии могут включать в себя увеличение пропускной способности узких мест путём строительства дополнительных полос или таких методов оперативного управления, как регулирование спроса, целью которого является сокращение перенасыщенных потоков на въездах в транспортную сеть в пиковые часы; профилактика и контроль ДТП, которые концентрируют ресурсы служб дорожного патруля и первой помощи поблизости от дорожных участков с повышенным риском ДТП; предоставление информации для водителей — ожидаемое время путешествия до определённых пунктов, предупреждение о ДТП впереди, возможные пути объезда пробки; управление транспортным потоком, включающее регулирование въездных потоков на загруженные магистрали, динамическое ограничение скорости для более равномерного движения автомобилей, оптимизация режимов и координация светофоров в группах перекрёстков на городских улицах. Для оперативного планирования нужно быстрое оценивание эффективности управляющих стратегий, которое в комбинации со стоимостью внедрения этих стратегий позволяет выбрать оптимальную из них с точки зрения отношения эффективности к стоимости.

Приведём общие параметры оценки эффективности соединения.

- Средняя скорость транспортного потока:

$$V_l(t) = \frac{f^d_l(t)}{\rho_l(t)}. \quad (8)$$

- Предполагаемое время в пути — время, необходимое автомобилю, чтобы проехать соединение от начала до конца в предположении, что скорость потока, зафиксированная в момент времени t , $V(t)$, останется постоянной:

$$\text{ПВП}_l(t) = \frac{\Delta x_l}{V_l(t)}. \quad (9)$$

- Фактическое время в пути — время, нужное автомобилю, чтобы проехать соединение от начала до конца, учитывая изменения в скорости потока после момента времени t :

$$\text{ФВП}_l(t) = T_l(t)\Delta t, \quad (10)$$

где

$$T_l(t) = \arg \max_{\tau} \left\{ \sum_{\tau'=0}^{\tau-1} V_l(t + \tau')\Delta t \leq \Delta x_l \right\}. \quad (11)$$

- Суммарный поток, измеряемый в автомобиле-километрах, оценивает фактическую пропускную способность соединения:

$$\text{СП}_l(t) = \rho_l(t)V_l(t)\Delta x_l\Delta t. \quad (12)$$

- Суммарная плотность, измеряемая в автомобиле-часах, отражает количество автомобилей, которые провели данный отрезок времени в данном соединении:

$$\text{СПЛ}_l(t) = \rho_l(t)\Delta x_l\Delta t. \quad (13)$$

- Задержка, измеряемая в автомобиле-часах:

$$\text{З}_l(t) = \text{СПЛ}_l(t) - \frac{\text{СП}_l(t)}{v_l}. \quad (14)$$

- Потеря продуктивности, измеряемая произведением километро-часов на число полос — степень недоиспользования пропускной способности полос соединения из-за затора:

$$\text{ПП}_l(t) = \begin{cases} 0, & \text{если } V_l(t) = v_l, \\ \left(1 - \frac{f^d_l(t)}{F_l}\right) k_l \Delta x_l \Delta t, & \text{иначе.} \end{cases} \quad (15)$$

Здесь надо обратить внимание на то, что потеря продуктивности зависит от числа полос. Например, при одном и том же исходящем из соединения потоке $f^d_l(t)$ и той же самой максимальной пропускной способности соединения F_l , потеря продуктивности двухполосного соединения вдвойне превышает потерю продуктивности соединения с одной полосой.

За исключением фактического времени в пути все перечисленные оценки эффективности вычисляются по ходу реализации траектории системы. Вычисление же фактического времени проводится апостериорно, требуя знания всей траектории системы. Если известны значения времени в пути, суммарного потока, суммарной плотности, задержки и потери продуктивности для каждого соединения сети, то можно элементарно вычислить эти же оценки для конкретных маршрутов внутри сети.

Соединения, представляющие городские улицы, чьи конечные узлы являются перекрёстками со светофорами, имеют дополнительные параметры эффективности, а именно следующие: задержка за цикл — число автомобиле-часов проведённых перед светофором; размер очереди — число ожидающих автомобилей в соединении; мера использования фазы — доля зелёной фазы цикла, при которой фактический поток, исходящий из соединения, равен требуемому потоку; нехватка цикла — процентная доля автомобилей, которым необходимо переждать красный свет более одного раза; отношение потока к максимальной пропускной способности, которое характеризует степень использования существующей пропускной способности; качество проезда — процентная доля автомобилей, прибывающих к перекрёстку во время зелёной фазы. Если известны значения этих параметров для каждого входящего соединения узла-перекрёстка, то нетрудно оценить эффективность перекрёстка в целом.

Для изучения поведения транспортной сети при различных сценариях, таких как плохая погода, ДТП, ремонт дороги, особые мероприятия, колебания спроса на дорогу и т.п. с помощью макроэкономической модели (1)–(7) используем переключения параметров модели (фундаментальных диаграмм в соединениях и распределительных матриц в узлах) либо функций требуемых потоков на въездах в определённые моменты времени.

Переключения в фундаментальных диаграммах моделируют уменьшение свободной скорости и максимальной пропускной способности из-за плохой погоды или блокирование полос в силу ДТП или дорожных работ. Например, допустим для какого-то соединения l : $F_l = 6000$ авто/ч, $v_l = 100$ км/ч, $w_l = 20$ км/ч, в 10:00 происходит авария, блокирующая половину полос, и длится до 10:30, когда движение полностью восстановлено. Мы моделируем это ДТП, дважды переключая фундаментальную диаграмму: $F_l(10) = 3000$ означает начало происшествия, и $F_l(10,5) = 6000$ означает возвращение к обычному режиму работы. Размер заторов, образующихся в результате ДТП, существенно зависит от времени реагирования на ДТП дорожно-патрульных служб. Изучение этой зависимости принадлежит разделу планирования по профилактике и контролю ДТП. Переключения в фундаментальных диаграммах мо-

гут быть применены также для выяснения целесообразности открытия в часы пик вспомогательных и специальных полос для общего пользования.

Особые мероприятия, влияние информации для водителей и прочие меры по перенаправлению автомобильных потоков моделируются переключениями в распределительных матрицах. Перенаправление потока с одного выходного соединения на другое может вызвать появление узкого места и, как следствие, затора, если максимальная пропускная способность этого другого соединения не будет удовлетворять требуемому потоку. Колебания в спросе моделируются путём умножения требуемых потоков на въздах на неотрицательные коэффициенты.

В симуляторе Aurora RNM эти переключения воплощены в качестве событий, которые пользователь может создавать и активизировать в заданные моменты времени. Влияние любого сценария на продуктивность транспортной сети возможно быстро проверить, поскольку этот симулятор вычисляет параметры эффективности для соединений и заданных маршрутов.

Ещё один важный аспект оперативного планирования, так же, как и системы поддержки реального времени, — управление потоком и оценка его эффективности рассматриваются в следующем параграфе.

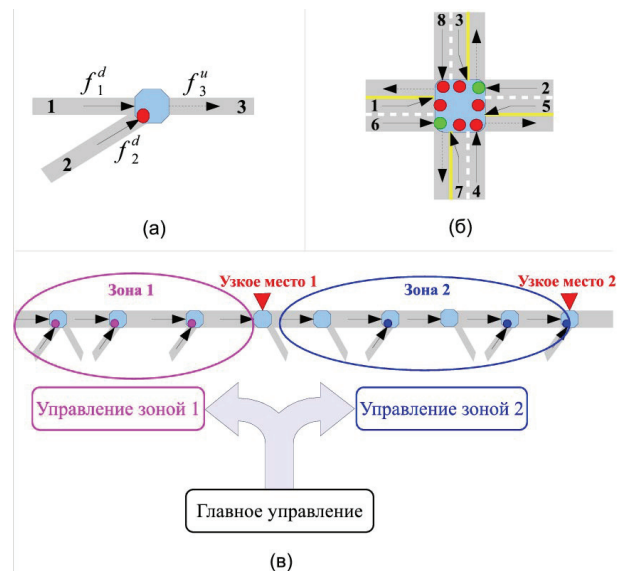


Рис. 5. Иерархия управляющих элементов: примеры (а) локального управления (регулятор въездного потока); (б) узлового управления (перекрёсток со светофором); (в) комплексное управление (координированное регулирование въездных потоков)

IV. Управление транспортным потоком

Среди упомянутых в предыдущем параграфе элементов оперативных стратегий минимизации заторов, влияние таких, как регулирование спроса, профилактика/контроль ДТП или информа-

ция для водителей, на поведение транспортной системы является лишь косвенным. А именно оценить, как изменилось поведение потока в результате упомянутых мер, можно лишь приблизительно, но невозможно напрямую измерить. Прямое же и доступное измерению влияние на транспортную систему достигается путём непосредственно-го управления автомобильным потоком.

В макроскопической модели (1)–(7) состояние соединения (плотность) эволюционирует во времени согласно закону сохранения. Для всякого соединения l , не являющегося выездом (то есть имеющего конечный узел), поток f_l^d , исходящий из этого соединения, может быть ограничен структурой узла.²⁵ Таким образом, управление потоками сосредоточено в узлах.

Далее предлагается следующая иерархия управляющих элементов, иллюстрированная рис. 5.

- Локальное управление привязано к конкретному входному соединению узла и регулирует только единственный поток, исходящий из этого соединения. Пример локального управления — это регулятор потока при въезде на автостраду, показанный на рис. 5а: узел имеет два входных соединения, соединение 1 без управления (автострада) и управляемое соединение 2 (въезд на автостраду).
- Узловое управление привязано к узлу и управляет потоками, исходящими из всех входных соединений узла. Пример узлового управления — перекрёсток со светофором, изображённый на рис. 5б. В этом примере в каждый момент времени управление пропускает только не конфликтующие между собой потоки: одновременно могут быть пропущены потоки 1 и 5, 2 и 6, 3 и 7 или 4 и 8, при этом остальные потоки блокируются. Каждое входное соединение индивидуально по-прежнему находится под контролем локального управления, но все локальные управляющие элементы подчинены узловому управлению.
- Комплексное управление оперирует несколькими локальными, узловыми или даже другими комплексными управляющими элементами, координируя их действия ради общей цели. Пример комплексного управления — координированное регулирование въездных потоков на автостраду, схема которого приведена на рис. 5в. В этом примере участок автострады имеет два выявленных узких места, делящих его на две зоны, 1 и 2, каждая из которых заканчивается соответствующим узким местом. Местоположения узких мест не фиксированы, и ожидается, что в течение

суток они меняются (такие узкие места называются плавающими). Зональные управления отвечают за координацию локальных регуляторов въездных потоков в своих зонах, в то время как главное управление отслеживает местоположение узких мест и соответственно модифицирует конфигурацию зональных управляющих элементов. И главное, и зональные управления являются комплексными.

Управление вводится в динамическую модель (1)–(7) заменой формулы (3) выражением

$$\tilde{d}_i^{[0]}(t) = v_i \rho_i(t) \min \left\{ 1, \frac{F_i}{v_i \rho_i(t)}, \frac{C_i(t, \boldsymbol{\rho}(t))}{v_i \rho_i(t)} \right\},$$

$$i = 1, \dots, m, \quad (16)$$

где m — число входных соединений узла, а $C_i(t, \boldsymbol{\rho}(t))$ — управляющая функция для потока, поступающего в узел из соединения i .

Управление может быть программным: $C_i(t, \boldsymbol{\rho}(t)) = C_i(t)$ или с обратной связью. Простейшим программным управлением является светофор, регулирующий въездной поток на автостраду, с фиксированной для данного времени суток длительностью красного и зелёного сигналов. Система светофоров с фиксированным циклом на перекрёстке представляет собой программное узловое управление.

Управление с обратной связью должно реагировать на состояние системы и желательно адаптироваться к специальным ситуациям типа ДТП. Примером управления с обратной связью является алгоритм ALINEA [18], локальное управление, регулирующее въездной поток на автостраду, которое для схемы рис. 5а определяется формулой

$$A_2(t, \rho_3(t)) = A_2(t - \Delta t, \rho_3(t - \Delta t)) + v_3(\rho_3^c - \rho_3(t)), \quad (17)$$

где индексы «2» и «3» относятся соответственно к соединениям 2 и 3, а функция A_2 — рекомендуемое ALINEA ограничение на поток, исходящий из соединения 2. Имеет смысл использовать алгоритмы управления типа ALINEA в комбинации с алгоритмами управления очередью, которые должны предотвращать формирование слишком больших заторов на въездах, которые могут распространиться на более узкие городские улицы. Наиболее известный алгоритм управления очередью, так называемое преодоление очереди (queue override) выражается следующим образом:

$$Q_2(t, \rho_2(t)) = f_2^u(t) + v_2(\rho_2(t) - \rho_2^c), \quad (18)$$

где функция Q_2 — ограничение на поток, исходящий из соединения 2, рекомендуемое алгоритмом преодоления очереди. При включении обоих алгоритмов одновременно они начинают работать

²⁵ По определению, поток, исходящий из выезда, ограничен лишь максимальной пропускной способностью выезда.

друг против друга: в то время как ALINEA создаёт очередь, алгоритм преодоления очереди пытается её сократить. Таким образом, комбинированный регулятор потока на въезде 2 имеет форму

$$C_2(t, \rho(t)) = \max\{A_2(t, \rho_3(t)), Q_2(t, \rho_2(t))\}. \quad (19)$$

Координированный алгоритм ALINEA, известный под названием HERO [19], идея которого заключается в том, чтобы сначала применять управление ALINEA (17) на ближайшем перед узким местом въезде на автостраду до тех пор, пока очередь не вырастет до рекомендуемого предела, затем надо включить ALINEA на следующем по удалённости въезде перед узким местом, пока очередь не заполнится там, и так, по мере необходимости, переключаться от въезда к въезду, удаляясь от узкого места, представляет собой комплексное управление. Примером комплексного управления с обратной связью, координирующего узловые управления на перекрёстках со светофорами, является алгоритм TUC [20].

Ограничение скорости, фиксированное и динамическое, также воздействует на автомобильный поток через узлы. Допустим, какой-либо регулятор скорости вычисляет желаемое ограничение скорости v_l^* для соединения l , тогда соответствующее ограничение на поток, устанавливаемое соответствующим локальным управлением, имеет форму

$$C_l(t, \rho_l(t)) = v_l^* \rho_l(t). \quad (20)$$

Нередко случается, что при моделировании управления с обратной связью с точки зрения уменьшения задержки и потери продуктивности эффективность алгоритмов очень высока. Однако как только дело доходит до внедрения этих алгоритмов в реальные транспортные системы, результаты разочаровывают. Причина здесь лежит в низком качестве обратной связи — из-за больших помех в измерениях либо из-за поломок в датчиках, либо из-за сбоев в коммуникационном оборудовании. Например, бюллетени состояния датчиков PeMS [13] свидетельствуют о том, что до 30% датчиков на калифорнийских автострадах не полностью функциональны в течение года.

Поэтому для симуляции алгоритмов управления с обратной связью и тестирования их эффективности в условиях, приближённых к реальным, мы предлагаем задействовать устройства измерения в транспортную модель. Такие элементы мы называем виртуальными датчиками. Виртуальные датчики моделируют работу реальных датчиков, сообщая данные счётчиков автомобилей и/или скорости автомобилей с определённых участков транспортной сети, основываясь на состоянии соединений сети, вычисленном посредством динамической модели. В каждом таком виртуальном датчике можно настроить свой уровень помех в диапазоне от идеального (полного отсутствия помех) до неудовлетворительного (все дан-

ные — одни сплошные помехи). Данные, сообщаемые виртуальными сенсорами, используются для формирования обратных связей при управлении потоком. Изображение на рис. 6 иллюстрирует идею виртуальных сенсоров на абстрактном уровне. В идеальных условиях (рис. 6а) управление получает информацию о состоянии системы напрямую из уравнения состояния (1). Вводя виртуальные датчики, мы подменяем в модели обратную связь по состоянию обратной связью по измерению.

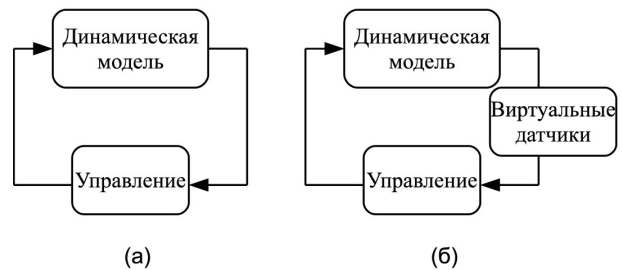


Рис. 6. Система с обратной связью: (а) обратная связь по состоянию; (б) обратная связь по измерению

Существующие устройства измерения автомобильных потоков можно условно подразделить на точечные датчики, такие, как индуктивные или магнитные детекторы, мобильные датчики, как, например, транспортные средства оборудованные приборами GPS, и пространственные устройства сбора данных, среди которых можно назвать камеры наблюдения, спутниковую и аэрофотосъёмку.

1. Точечные датчики имеют фиксированное местоположение на дороге и работают как счётчики проезжающих поверх них автомобилей, измеряя в определённых конфигурациях и скорость этих автомобилей.
2. Мобильные датчики передвигаются вместе с транспортным потоком в пространстве и времени и периодически сообщают свои координаты и скорость, снабжённые временной меткой.
3. Пространственные устройства сбора данных фотографирует ситуацию на данном участке дороги в данный момент времени, и затем из полученного снимка можно извлечь информацию о плотности автомобилей.

Мы здесь моделируем точечные и мобильные датчики. Каждый точечный датчик приписан к своему соединению и его местоположение внутри соединения строго определено. Соединения могут иметь по несколько точечных датчиков, приписанных к ним. Конкретные модели воплощают поведение индуктивных или магнитных детекторов. Измерения индуктивных детекторов содержат данные счётчиков автомобилей и, возможно, скорости. Измерения магнитных детекторов, таких, как, например, беспроводные сенсоры Sensys [21] могут быть использованы для вычисления

плотности автомобилей в соединении, как описано в [22]. Каждый же мобильный датчик приписан к конкретному маршруту между двумя пунктами в сети. В определённые моменты времени он сообщает соединение, в котором он находится, своё местоположение в этом соединении и скорость. Мобильные датчики — это «автомобили-фантомы» в том смысле, что они не оказывают влияния на состояние системы, диктуемое моделью. Применение мобильных датчиков возможно для вычисления фактического времени в пути для конкретных маршрутов.

Ещё одно применение виртуальных датчиков — это интерфейс к реальным измерительным устройствам, данные с которых поступают в центр управления транспортом в сыром, необработанном виде. В то же время симулятор с динамической моделью (1)–(7) исполняет функцию фильтрации этих данных прежде, чем использовать их для обратной связи с адаптивными управляющими элементами. Гарантированное прогнозирование и основанный на нём динамический фильтр описываются в следующем параграфе.

V. Прогноз и оценивание состояния транспортной системы

Для прогнозирования и оценивания состояния транспортной сети требуется знать начальные условия, входные данные (требуемые потоки на въездах) и параметры системы. Начальные условия формируются на основе поступающих измерений. Будущие значения требуемых потоков заранее неизвестны, но могут быть более или менее достоверно предугаданы, базируясь на статистическом анализе данных прошлых измерений, как предлагается в [23–25]. Параметры системы, то есть фундаментальные диаграммы и распределительные матрицы также определяются на основе данных прошлых измерений.²⁶ Задача измерения максимальной пропускной способности и калибровки фундаментальных диаграмм исследована в [15]. Как уже было упомянуто в подпараграфе II.2, данные измерений в левой части фундаментальной диаграммы (в зоне свободного движения) хорошо аппроксимируются прямой линией, в то время как данные в правой части фундаментальной диаграммы (в зоне затора) имеют значительный разброс. Этот разброс в комбинации с неточностью в определении максимальной пропускной способности создаёт неопределённость в фундаментальной диаграмме.

Таким образом, мы предполагаем, что значения требуемых потоков на въездах в сеть, $r_l(t)$, известны не точно, а лежат в известных интервалах, $r_l^-(t) \leq r_l(t) \leq r_l^+(t)$ ($r_l^-(t)$ и $r_l^+(t)$ известны); значения максимальной пропускной способности F_l и максимальной плотности $\bar{\rho}_l$, $l \in \mathcal{L}$, также ле-

жат в известных интервалах, $F_l^- \leq F_l \leq F_l^+$ (F_l^- и F_l^+ известны) и $\bar{\rho}_l^- \leq \bar{\rho}_l \leq \bar{\rho}_l^+$ ($\bar{\rho}_l^-$ и $\bar{\rho}_l^+$ известны), и соответственно

$$w_l^- = \frac{F_l^-}{\bar{\rho}_l^- - F_l^-/v_l} \text{ и } w_l^+ = \frac{F_l^+}{\bar{\rho}_l^+ - F_l^+/v_l},$$

а свободные скорости v_l известны точно. Изображение на рис. 7 иллюстрирует неопределённость в фундаментальной диаграмме. Ещё одно предположение заключается в том, что распределительные матрицы в узлах, $\mathcal{B}_\nu(t)$, $\nu \in \mathcal{N}$, известны точно.

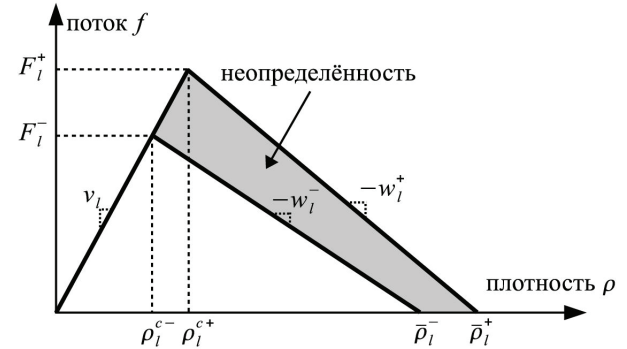


Рис. 7. Неопределённость в фундаментальной диаграмме соединения $l \in \mathcal{L}$

Для каждого соединения $l \in \mathcal{L}$ имеются неточные измерения исходящего потока:

$$y_l^{(f)}(t) = f_l^d(t) + \omega_l^{(f)}(t), \quad (21)$$

и скорости:

$$y_l^{(V)}(t) = V_l(t) + \omega_l^{(V)}(t). \quad (22)$$

В последних двух выражениях $\omega_l^{(f)}(t) \in [-\omega_l^{0,(f)}(t), \omega_l^{0,(f)}(t)]$ обозначает помеху при измерении потока, $\omega_l^{(V)}(t) \in [-\omega_l^{0,(V)}(t), \omega_l^{0,(V)}(t)]$ — помеху при изменении скорости, а ограничения $\omega_l^{0,(f)}(t)$ и $\omega_l^{0,(V)}(t)$ известны. Таким образом, для каждого соединения имеем оценку плотности по измерениям:

$$\hat{\rho}_l^-(t) \leq \hat{\rho}_l(t) \leq \hat{\rho}_l^+(t), \quad (23)$$

где

$$\hat{\rho}_l^-(t) = \frac{y_l^{(f)}(t) - \omega_l^{0,(f)}(t)}{y_l^{(V)}(t) + \omega_l^{0,(V)}(t)} \text{ и } \hat{\rho}_l^+(t) = \frac{y_l^{(f)}(t) + \omega_l^{0,(f)}(t)}{y_l^{(V)}(t) - \omega_l^{0,(V)}(t)}, \quad V_l(t) > 0. \quad (24)$$

Теперь мы можем сформулировать задачи, требующие решения.

1. **Гарантированное прогнозирование состояния транспортной сети.** Найти кривые $\rho^-(\cdot)$ (лучший случай) и $\rho^+(\cdot)$ (худший случай) такие, что любая возможная траектория системы (1)–(7) с описанной выше

²⁶Параметры Δx_l известны точно, а размер временного шага Δt выбирается, исходя из значений Δx_l и v_l , $l \in \mathcal{L}$ так, чтобы удовлетворялось условие КФЛ.

множественной неопределённостью в требуемых потоках и фундаментальных диаграммах, $\rho(\cdot)$, ограничена:

$$\rho_l^-(t) \leq \rho_l(t) \leq \rho_l^+(t), t \geq t_0, \quad (25)$$

для всех $l \in \mathcal{L}$.

2. Гарантированное оценивание состояния транспортной сети. В определённые моменты времени $t_0 < \tau_1 < \tau_2 < \dots < \tau_q < \dots$ для каждого соединения $l \in \mathcal{L}$ использовать оценку по измерениям $[\hat{\rho}_l^-(\tau), \hat{\rho}_l^+(\tau)]$ (если таковые имеются) для корректировки прогнозируемого интервала $[\rho_l^-(\tau), \rho_l^+(\tau)]$, $\tau = \tau_1, \tau_2, \dots, \tau_q, \dots$

V.1. Гарантированное прогнозирование состояния системы

Основываясь на законе сохранения (1), определим уравнения для нижних и верхних ограничений на плотность:

$$\rho_l^\pm(t + \Delta t) = \rho_l^\pm(t) + \frac{\Delta t}{\Delta x_l} (f_l^{u\pm}(t) - f_l^{d\pm}(t))$$

для каждого $l \in \mathcal{L}$. (26)

Здесь и далее символ « \pm » означает либо « $-$ » для вычисления нижней (лучший случай), либо « $+$ » для вычисления верхней (худший случай) границы плотности. Как и раньше, Δt удовлетворяет условию КФЛ $\Delta t \leq \min_l \left\{ \frac{\Delta x_l}{v_l} \right\}$ и $\Delta t \leq \min_l \left\{ \frac{\Delta x_l}{w_l} \right\}$. Для въездов $f_l^{u\pm}(t) = r_l^\pm(t)$. Для выездов $f_l^{d-}(t) = v_l \rho_l^-(t) \min \left\{ 1, \frac{F_l^+}{v_l \rho_l^+(t)} \right\}$ и $f_l^{d+}(t) = v_l \rho_l^+(t) \min \left\{ 1, \frac{F_l^-}{v_l \rho_l^-(t)} \right\}$. Для обычных соединений $f_l^{u\pm}(t)$ определяется своим начальным, а $f_l^{d\pm}(t)$ — своим конечным узлами.

В общем случае для узла $\nu \in \mathcal{N}$ с m входными, n выходными соединениями и распределительной матрицей: $\mathcal{B}_\nu(t) = \{\beta_{ij}(t)\}_{i=1, \dots, m, j=1, \dots, n}$, значения $f_i^{d\pm}$, $i = 1, \dots, m$, и $f_j^{u\pm}$, $j = 1, \dots, n$, вычисляются согласно следующему алгоритму.

Алгоритм вычисления входных и выходных потоков в лучшем и худшем случаях

1. Для каждого выходного соединения $j = 1, \dots, n$ вычислим нижние границы фактической приёмной способности:

$$\underline{s}_j^\pm(t) = \min\{F_j^-, \max\{0, w_j^-(\bar{\rho}_j^- - \rho_j^\pm(t))\}\} \quad (27)$$

и её верхние границы:

$$\bar{s}_j^\pm(t) = \min\{F_j^+, w_j^+(\bar{\rho}_j^+ - \rho_j^\pm(t))\}. \quad (28)$$

Из формул (27) и (28) следует, что $\underline{s}_j^\pm(t) \leq \bar{s}_j^\pm(t)$, $\underline{s}_j^+(t) \leq \underline{s}_j^-(t)$ и $\bar{s}_j^+(t) \leq \bar{s}_j^-(t)$.

На рис. 8 показано, что значат выражения $\underline{s}_j^\pm(t)$ и $\bar{s}_j^\pm(t)$. Так как возможно $\rho_j^+(t) > \bar{\rho}_j^-$, то необходимо следить, чтобы всегда выполнялось условие $\underline{s}_j^+(t) \geq 0$.

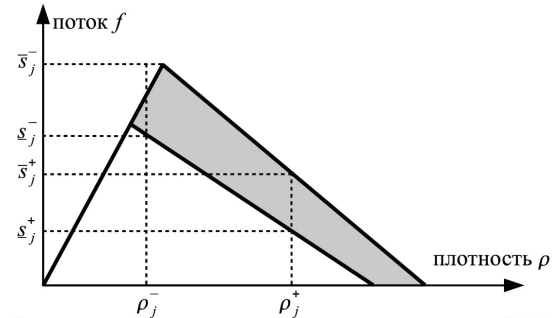


Рис. 8. Границы фактически допустимого приёма для $f_j^{u-}(t)$ и $f_j^{u+}(t)$

2. Установим итерацию $p = 0$.
3. Для каждого входного соединения $i = 1, \dots, m$ вычислим нижние границы требуемого потока:

$$\underline{d}_i^{\pm[0]}(t) = v_i \rho_i^\pm(t) \min \left\{ 1, \frac{F_i^-}{v_i \rho_i^\pm(t)}, \frac{C_i^-(t, \rho^-, \rho^+)}{v_i \rho_i^\pm(t)} \right\} \quad (29)$$

и верхние границы требуемого потока:

$$\bar{d}_i^{\pm[0]}(t) = v_i \rho_i^\pm(t) \min \left\{ 1, \frac{F_i^+}{v_i \rho_i^\pm(t)}, \frac{C_i^+(t, \rho^-, \rho^+)}{v_i \rho_i^\pm(t)} \right\}, \quad (30)$$

где

$$C_i^-(t, \rho^-, \rho^+) = \min\{C_i(t, \bar{\rho}) \mid \rho_l^-(t) \leq \bar{\rho}_l \leq \rho_l^+(t), \forall l \in \mathcal{L}\}$$

и

$$C_i^+(t, \rho^-, \rho^+) = \max\{C_i(t, \bar{\rho}) \mid \rho_l^-(t) \leq \bar{\rho}_l \leq \rho_l^+(t), \forall l \in \mathcal{L}\}.$$

Из формул (29), (30) следует, что $\underline{d}_i^{\pm[0]}(t) \leq \bar{d}_i^{\pm[0]}(t)$, $\underline{d}_i^{-[0]}(t) \leq \underline{d}_i^{+[0]}(t)$ и $\bar{d}_i^{-[0]}(t) \leq \bar{d}_i^{+[0]}(t)$. На рис. 9 показано, что значат выражения $\underline{d}_i^{\pm[0]}(t)$ и $\bar{d}_i^{\pm[0]}(t)$.

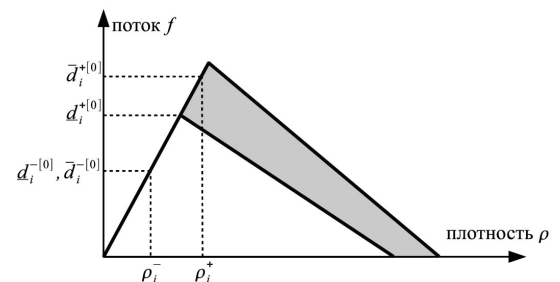


Рис. 9. Границы требуемого потока для $f_i^{d-}(t)$ и $f_i^{d+}(t)$

4. Для каждого выходного соединения $j = 1, \dots, n$ вычислим нижние границы требуемого потока:

$$\underline{d}_j^{*\pm[p]}(t) = \sum_{i=1}^m \beta_{ij}(t) \underline{d}_i^{\pm[p]}(t) \quad (31)$$

и его верхние границы:

$$\overline{d}_j^{*\pm[p]}(t) = \sum_{i=1}^m \beta_{ij}(t) \overline{d}_i^{\pm[p]}(t). \quad (32)$$

Очевидно, что $\underline{d}_i^{*\pm[0]}(t) \leq \overline{d}_i^{*\pm[0]}(t)$, $\underline{d}_i^{*-[0]}(t) \leq \underline{d}_i^{*+[0]}(t)$ и $\overline{d}_i^{*-[0]}(t) \leq \overline{d}_i^{*+[0]}(t)$.

5. Для $p = 1, \dots, n$ и для каждого входного соединения $i = 1, \dots, m$ повторяем:

(а) сократим, если необходимо, верхнюю границу требуемого потока для лучшего случая, $\overline{d}_i^{-[p]}(t)$, чтобы она удовлетворяла верхней границе фактической приёмной способности для лучшего случая, $\overline{s}_p^{-}(t)$:

$$\overline{d}_i^{-[p]}(t) = \begin{cases} \overline{d}_i^{-[p-1]}(t), & \text{если } \beta_{ip}(t) = 0, \\ \overline{d}_i^{-[p-1]}(t) \times \\ \times \min \left\{ 1, \frac{\overline{s}_p^{-}(t)}{\overline{d}_p^{*-[p-1]}(t)} \right\}, & \text{иначе;} \end{cases} \quad (33)$$

(б) сократим, если необходимо, нижнюю границу требуемого потока для худшего случая, $\underline{d}_i^{-[p]}(t)$, чтобы она удовлетворяла нижней границе фактической приёмной способности для худшего случая, $\underline{s}_p^{+}(t)$:

$$\underline{d}_i^{-[p]}(t) = \begin{cases} \underline{d}_i^{-[p-1]}(t), & \text{если } \beta_{ip}(t) = 0, \\ \underline{d}_i^{-[p-1]}(t) \times \\ \times \min \left\{ 1, \frac{\underline{s}_p^{+}(t)}{\underline{d}_p^{*+[p-1]}(t)} \right\}, & \text{иначе;} \end{cases} \quad (34)$$

(с) пересчитаем верхние границы требуемых выходных потоков для лучшего случая, $\overline{d}_q^{*-[p]}(t)$, по формуле (32) и нижние границы требуемых выходных потоков для худшего случая, $\underline{d}_q^{*+[p]}(t)$, по формуле (31), $q = p + 1, \dots, n$.

6. Для каждого входного соединения $i = 1, \dots, m$ поток, исходящий из этого соединения, в лучшем случае:

$$f_i^{d-}(t) = \overline{d}_i^{-[n]}(t), \quad (35)$$

а в худшем случае:

$$f_i^{d+}(t) = \underline{d}_i^{+[n]}(t). \quad (36)$$

7. Для каждого выходного соединения $j = 1, \dots, n$ повторяем

(а) для $p = 1, \dots, n$ повторяем

i. сократим, если необходимо, нижнюю границу требуемого потока для лучшего случая, $\underline{d}_i^{-[p]}(t)$, чтобы она удовлетворяла нижней границе фактической приёмной способности для худшего случая, $\underline{s}_p^{+}(t)$, при $p \neq j$ и нижней границе фактической приёмной способности для лучшего случая, $\underline{s}_j^{-}(t)$, при $p = j$:

$$\underline{d}_i^{-[p]}(t) = \begin{cases} \underline{d}_i^{-[p-1]}(t), & \\ \text{если } \beta_{ip}(t) = 0, & \\ \underline{d}_i^{-[p-1]}(t) \times & \\ \times \min \left\{ 1, \frac{\underline{s}_p^{+}(t)}{\underline{d}_p^{*-[p-1]}(t)} \right\}, & \\ \text{если } \beta_{ip} > 0 \text{ и } p \neq j, & \\ \underline{d}_i^{-[p-1]}(t) \times & \\ \times \min \left\{ 1, \frac{\underline{s}_j^{-}(t)}{\underline{d}_p^{*+[p-1]}(t)} \right\}, & \\ \text{если } \beta_{ip} > 0 \text{ и } p = j; & \end{cases} \quad (37)$$

ii. сократим, если необходимо, верхнюю границу требуемого потока для худшего случая, $\overline{d}_i^{+[p]}(t)$, чтобы она удовлетворяла верхней границе фактической приёмной способности для лучшего случая, $\overline{s}_p^{-}(t)$, при $p \neq j$ и верхней границе фактической приёмной способности для худшего случая, $\overline{s}_p^{+}(t)$, при $p = j$:

$$\overline{d}_i^{+[p]}(t) = \begin{cases} \overline{d}_i^{+[p-1]}(t), & \\ \text{если } \beta_{ip}(t) = 0, & \\ \overline{d}_i^{+[p-1]}(t) \times & \\ \times \min \left\{ 1, \frac{\overline{s}_p^{-}(t)}{\overline{d}_p^{*+[p-1]}(t)} \right\}, & \\ \text{если } \beta_{ip} > 0 \text{ и } p \neq j, & \\ \overline{d}_i^{+[p-1]}(t) \times & \\ \times \min \left\{ 1, \frac{\overline{s}_j^{+}(t)}{\overline{d}_p^{*-[p-1]}(t)} \right\}, & \\ \text{если } \beta_{ip} > 0 \text{ и } p = j; & \end{cases} \quad (38)$$

iii. пересчитаем нижние границы требуемых выходных потоков для лучшего случая, $\underline{d}_q^{*-[p]}(t)$, по формуле (31) и верхние границы требуемых выходных потоков для худшего случая, $\overline{d}_q^{*+[p]}(t)$, по формуле (32), $q = p + 1, \dots, n$;

(б) поток, входящий в выходное соединение j , в лучшем случае:

$$f_j^{u-}(t) = \sum_{i=1}^m \beta_{ij}(t) \underline{d}_i^{-[n]}(t), \quad (39)$$

а в худшем случае:

$$f_j^{u+}(t) = \min \left\{ \bar{s}_j^+(t), \sum_{i=1}^m \beta_{ij}(t) \bar{d}_i^{+[n]}(t) \right\}. \quad (40)$$

При начальной позиции $t = t_0$, $\rho^-(t_0) = \rho_0^-$ и $\rho^+(t_0) = \rho_0^+$, таких, что $\rho_l^-(t_0) \leq \rho_l^+(t_0)$ для всех $l \in \mathcal{L}$,²⁷ прогнозирующая система развивается во времени, во-первых, вычисляя $f_l^{u\pm}$ и $f_l^{d\pm}$ по шагам 1–7 для всех $l \in \mathcal{L}$, и, во-вторых, вычисляя границы плотности согласно (26) для всех $l \in \mathcal{L}$.

V.2. Гарантированное оценивание состояния системы

Динамическая фильтрация измерений, множественное оценивание состояния транспортной сети — это комбинированный процесс, состоящий из вычисления множества достижимости (прогноза состояния) системы (26)–(40) и его корректировки при помощи оценок измерений (23)–(24), поступающих в моменты времени $t_0 < \tau_1 < \tau_2 < \dots < \tau_q < \dots$. Далее применяем методы теории гарантированного оценивания, следуя публикациям [26–28].

Алгоритм гарантированного оценивания

1. В момент времени $\tau_0 = t_0$ вычисляем оценку состояния по измерениям $\hat{\rho}_l^-(\tau_0)$, $\hat{\rho}_l^+(\tau_0)$ по формуле (24) для всех $l \in \mathcal{L}$.

2. Присваиваем значения

$$\rho_l^-(\tau_0) = \hat{\rho}_l^-(\tau_0) \text{ и } \rho_l^+(\tau_0) = \hat{\rho}_l^+(\tau_0) \quad (41)$$

для всех $l \in \mathcal{L}$.

3. Для $q = 1, 2, \dots$ повторяем:

(а) вычисляем $\rho_l^-(\tau_q)$ и $\rho_l^+(\tau_q)$ с помощью уравнений (26)–(40) с начальными условиями $\rho_l^-(\tau_{q-1})$, $\rho_l^+(\tau_{q-1})$ для всех $l \in \mathcal{L}$;

(б) получаем оценку состояния по измерениям $\hat{\rho}_l^-(\tau_q)$, $\hat{\rho}_l^+(\tau_q)$ согласно (24) для всех $l \in \mathcal{L}$;

(с) в предположении, что

$$\begin{aligned} [\rho_l^-(\tau_q), \rho_l^+(\tau_q)] \cap [\hat{\rho}_l^-(\tau_q), \hat{\rho}_l^+(\tau_q)] \neq \\ \neq \emptyset, \end{aligned} \quad (42)$$

корректируем прогноз плотности:

$$\begin{aligned} \rho_l^-(\tau_q) \leftarrow \max\{\rho_l^-(\tau_q), \hat{\rho}_l^-(\tau_q)\} \text{ и} \\ \rho_l^+(\tau_q) \leftarrow \min\{\rho_l^+(\tau_q), \hat{\rho}_l^+(\tau_q)\} \end{aligned} \quad (43)$$

для всех $l \in \mathcal{L}$.

В любое время $t \geq t_0$ оценка плотности для соединения $l \in \mathcal{L}$ — это интервал $[\rho_l^-(t), \rho_l^+(t)]$, а если $t = \tau_q$, то значения $\rho_l^-(\tau_q)$, $\rho_l^+(\tau_q)$ берутся после корректировки (43).

На рис. 10 схематично представлен алгоритм динамической фильтрации: для каждого $q = 1, 2, \dots$ множественная траектория системы (26)–(40) вычисляется от τ_{q-1} до τ_q , после чего пересекается с измерениями, а результат пересечения берётся за новое начальное условие.

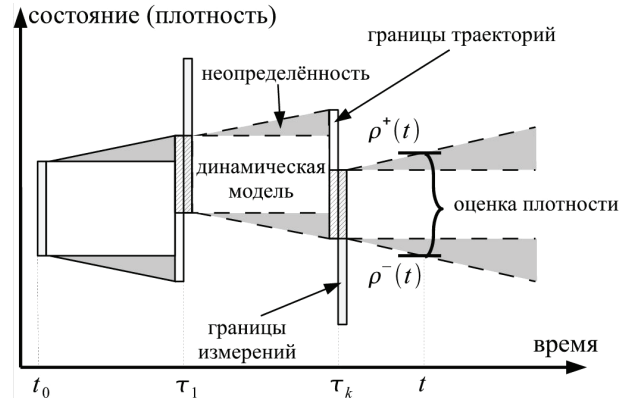


Рис. 10. Схема работы динамического фильтра

Согласно теории пересечение интервалов (42) должно быть непустым, иначе получается, что сделаны неверные предположения о диапазоне помех при измерениях или при неопределённости в динамической системе. Однако в реальной жизни пустые пересечения имеют место. Что делать, если условие (42) не выполнено, зависит от конкретной ситуации. Если мы доверяем модели больше, чем измерениям, то просто отказываемся от корректировки (43). Более того, динамическую модель (26)–(40) можем использовать для выявления неисправных датчиков. С другой стороны, если полагаем, что данные измерений надёжны, то корректировка (43) превращается в

$$\rho_l^-(\tau_q) \leftarrow \hat{\rho}_l^-(\tau_q), \text{ и } \rho_l^+(\tau_q) \leftarrow \hat{\rho}_l^+(\tau_q). \quad (44)$$

Возможно, что из каких-то соединений в \mathcal{L} измерения не поступают. Для таких соединений корректировка (43) не делается. Задача же наблюдаемости (когда, например, поток измерить можно, а скорость нельзя или наоборот) рассмотрена в [29].

VI. Пример: автострада I-80 в Калифорнии

В качестве примера транспортной сети рассмотрим одну из наиболее загруженных магистралей США — автостраду I-80 в районе залива Сан-Франциско. Интересующий нас участок западного направления — это 30 км, ведущие от моста Карквинес Бридж на севере восточного берега залива через города Сан Пабло, Ричмонд, Беркли и Окленд к мосту Бэй Бридж — основной магистрали, соединяющей восточный берег залива с

²⁷Начальные границы плотности могут быть получены из измерений по формуле (24).

городом Сан-Франциско (см. карту на рис. 11а, на которой начало и конец участка обозначены, соответственно, буквами **А** и **Б**). На данном участке автострады имеется 29 въездов, 24 съезда и 57 датчиков — индуктивных детекторов, функционирующих с переменным успехом. Кроме того, при построении модели этой транспортной сети нами дополнительно предусмотрено возможное регулирование потоков на отдельных въездах. Места возможного регулирования также указаны на рис. 11а.

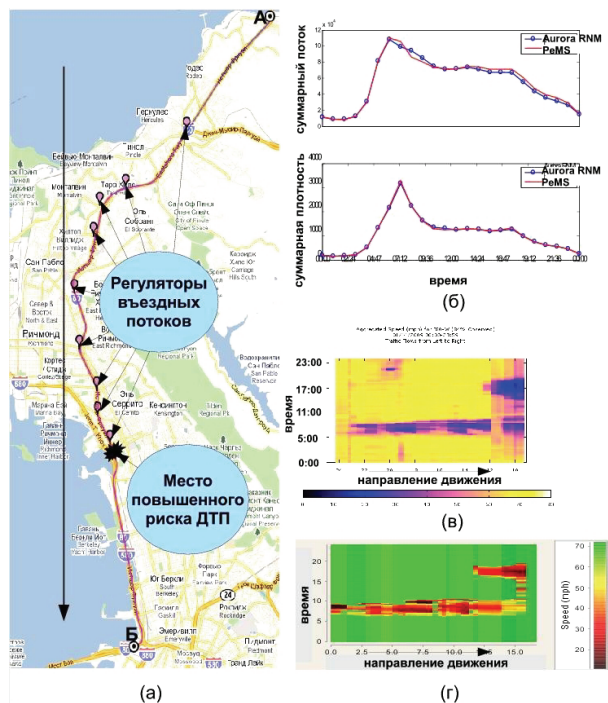


Рис. 11. Модель западного направления автострады I-80 с данными автомобильных въездных потоков на 14 января 2009 года, где (а) карта автострады — поток движется от пункта А к пункту Б; (б) суммарный поток и суммарная плотность — сравнение измерений (PeMS) с результатом симуляции (Aurora RNM) (значения даны из расчёта за 1 час, а суммарный поток указан в автомобиле-милях); контурные диаграммы скорости (скорость указана в милях в час) (в) по измерениям, (г) по результатам симуляции

Построение сети, состоящей из соединений и узлов, и калибровка фундаментальных диаграмм соединений есть первый этап в исследовании интересующей нас дороги. Далее, чтобы проверить правильность полученной модели, нами выбраны дни, когда по данным системы PeMS [13] количество неисправных датчиков было минимальным, а качество измерений соответственно наилучшим. Примером такого дня может служить среда, 14 января 2009 года.²⁸ В том, что предлагаемая нами модель хорошо воспроизводит реальную ситу-

ацию на дороге, имевшую место 14 января 2009 года, можно убедиться, глядя на сравнительные графики суммарного потока²⁹ и суммарной плотности для измерений и симуляции, где значения берутся с интервалом в 1 час, приведённые на рис. 11б, и контурные диаграммы скорости, полученные в результате измерений (рис. 11в) и симуляции (рис. 11г). Контурные диаграммы следует читать так: горизонтальная ось обозначает направление движения автомобильного потока (слева направо) при длине маршрута 30 км, вертикальная ось — это ось времени от 0 до 24 часов (снизу вверх), а цвет в данной точке плоскости указывает скорость в данной точке маршрута в данное время соответственно приведённой рядом шкале³⁰.

Одним из элементов оперативного планирования является тестирование потенциальных стратегий управления на предмет того, насколько выгодно их внедрение. В данном примере мы прогнали по 50 стохастических симуляций для каждой из трёх конфигураций: (1) с регулированием въездных потоков по методу ALINEA [18] в местах, указанных на рис. 11а; (2) с комбинированным регулированием ALINEA и преодоления очереди в этих же местах; (3) без всякого управления, то есть, как это происходит в реальной жизни на данный момент. Стохастика симуляций состоит в том, что для каждой из них максимальные пропускные способности соединений выбирается случайным образом из диапазона $\pm 5\%$ от значений, полученных в результате калибровки, а разброс значений требуемых потоков на въездах составляет $\pm 25\%$ от данных на 14 января 2009 года. На рис. 12 приведены результаты этих стохастических симуляций, показанные в виде зависимости задержки от суммарного потока за сутки для транспортной сети. Значения суммарного потока характеризуют спрос. Очевидно, что при низком спросе (малых значениях суммарного потока) задержка отсутствует, и нет различия между конфигурациями с управлением и без. Зато при высоком спросе (больших значениях суммарного потока) наблюдается большой разброс в значениях задержки: в конфигурации без управления разброс значений задержки для одного и того же суммарного потока составляет около 8000 автомобиле-часов в сутки, что, из консервативного расчёта: 1 автомобиль = 1 человек и стоимости 1 часа водителя равной 50 долларам, составляет 200 тысяч долларов в сутки. При регулировании въездных потоков разброс значений задержки незначителен, а сами значения существенно ниже, чем в конфигурации без управления.

²⁸На самом деле, мы не ограничиваемся одним днём, но стараемся собрать, по меньшей мере, по одной конфигурации на каждый день недели, так как наблюдения показывают, что хотя автомобильные потоки на одной и той же дороге отличаются друг от друга день ото дня, их поведение качественно повторяется в одни и те же дни недели. В нашем же случае 14 января 2009 года является прототипом среднестатистической среды на западном направлении автострады I-80.

²⁹Значения суммарного потока указаны в автомобиле-милях.

³⁰Скорость на контурных диаграммах указана в милях в час: учитывая, что 1 миля примерно равна 1,6 км, 70 миль/ч соответствует, примерно, 112 км/ч.

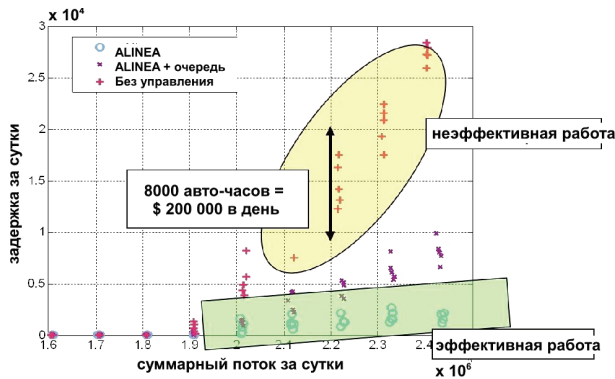


Рис. 12. Зависимость задержки за сутки, исчисляемой в автомобиле-часах, от суммарного потока за сутки, указанного в автомобиле-милях, для транспортной сети по результатам 50 стохастических симуляций с каждой из трёх конфигураций, (1) с регуляторами типа ALINEA на въездах, (2) с регуляторами ALINEA в комбинации с преодолением очереди и (3) без управления

Рассмотрим пример краткосрочного прогнозирования. Исходя из данных контурной диаграммы скорости на рис. 11г, затор на дороге начинается после 6 часов утра. Представим себя на месте диспетчера, контролирующего автостраду I-80 в 6:00, в среду. Оценив неопределённость в максимальных пропускных способностях в $\pm 1,5\%$ и неопределённость в предполагаемых требуемых потоках на въездах — в $\pm 2\%$ на ближайшие два часа (с 6 до 8 утра), далее можем вычислить границы плотности для всех соединений сети, а следовательно и оценки эффективности для лучшего и худшего случаев: суммарного потока, суммарной плотности, задержки и потери продуктивности. Графики этих оценок изображены на рис. 13а: период с 5:00 до 6:00 уже в прошлом, и для него имеются измерения, определяющие наши начальные условия в 6:00, а период с 6:00 до 8:00 — это ближайшее будущее, для которого рассчитывается прогноз. Заметим, что для суммарной плотности, задержки и потери продуктивности кривые лучшего и худшего случаев определяют границы, между которыми гарантированно будут лежать соответствующие параметры реальной системы (если не произойдёт ничего неожиданного, например, аварии). Для суммарного потока кривые лучшего и худшего случаев не являются границами, они могут пересекаться, так как более высокое значение потока не всегда соответствует более высокому (или более низкому) значению плотности и наоборот. На рис. 13б изображены графики тех же оценок, но при включенных регуляторах ALINEA в комбинации с преодолением очереди на въездах, указанных на рис. 11а. Здесь наблюдается сужение прогнозируемых интервалов для суммарной плотности, задержки и потери продуктивности, то есть система стала более предсказуемой. Следует обратить внимание на то, что хотя значения этих

трёх параметров сократились по сравнению с конфигурацией без управления, сама потеря продуктивности сократилась значительно сильнее, чем задержка. Это произошло потому, что задержка частично переместилась из заторов на автостраде в очереди на въездах.

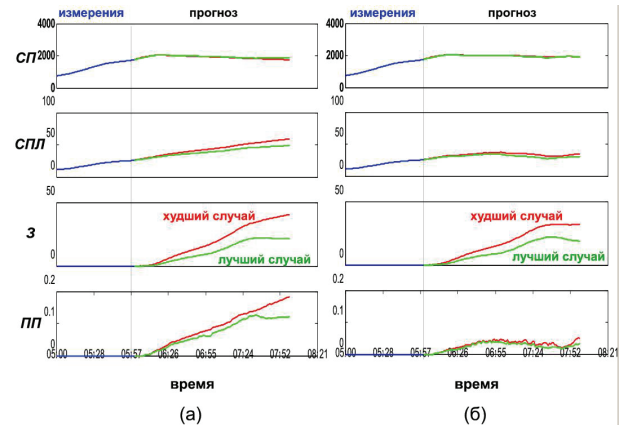


Рис. 13. Прогноз оценок эффективности западного направления автострады I-80, суммарного потока (СП), суммарной плотности (СПЛ), задержки (З), потери продуктивности (ПП) в конфигурациях (а) без управления; (б) с регулированием въездных потоков при помощи комбинации алгоритмов ALINEA и преодоления очереди. Значения оценок эффективности даны из расчёта за 1 минуту. Суммарный поток указан в автомобиле-милях

Далее введём в действие сценарий — ДТП. Место гипотетической аварии обозначено на рис. 11а как место повышенного риска ДТП. Система ReMS предоставляет статистику зарегистрированных ДТП на автострадах Калифорнии. На контурной диаграмме ДТП (см. рис. 14а) обозначено место и время, когда и где подобные происшествия происходят наиболее часто — это место мы и называем местами повышенного риска, а фотография его приведена на рис. 14б. Хотя наиболее вероятное время ДТП согласно контурной диаграмме на рис. 14а между 7 и 8 часами утра, предположим, что ДТП произойдёт несколько раньше, в 6:35, и половина полос (две из четырёх) будут блокированы в течение 15 минут, после чего движение будет полностью восстановлено.³¹ Результат прогноза для сценария ДТП приведён на рис. 14в — задержка, достигающая в худшем случае 65, а в лучшем 47 автомобиле-часов в минуту, и 14г — контурные диаграммы скорости для лучшего и худшего случаев. Данные на этих рисунках получены в предположении, что управление отсутствует. В качестве возможных мер по преодолению последствий ДТП предлагаются следующие три оперативные стратегии.

Стратегия 1 заключается в том, чтобы ещё не дожидаясь аварии, включить регуляторы ALINEA с преодолением очереди на въездах, указанных на рис. 11а и 15а, и более ничего не де-

³¹Мы выбрали время ДТП так, чтобы вред, нанесённый им дорожному движению и продуктивности транспортной сети, был максимальным. Это случится, если ДТП произойдёт перед самым началом формирования затора в обычных условиях, то есть когда интенсивность движения резко возрастает.

лать. То есть следует осуществлять обычное регулирование въездных потоков, в принципе никак не связанное с ДТП. Тем не менее это даёт хороший результат. Данные прогноза для сценария ДТП при вводе в действие стратегии 1 приведены на рис. 16а. Здесь видно, что задержка в худшем случае сократится до 47, а в лучшем до 37 автомобиль-часов в минуту, и на контурных диаграммах скорости видно, что существенно уменьшились зоны затора.

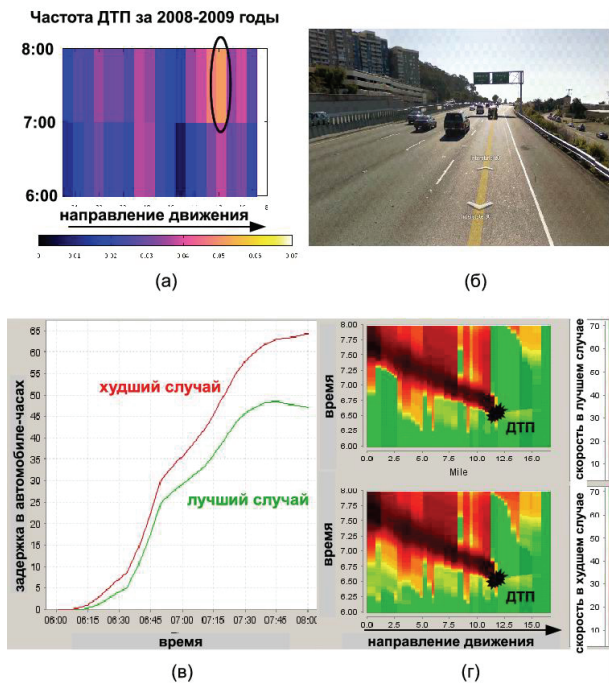


Рис. 14. Сценарий возможного ДТП на западном направлении автострады I-80 в 6:35 утра длительностью 15 минут, где (а) контурная диаграмма частоты ДТП между 6 и 8 часами утра (источник PeMS); (б) фотография места гипотетического ДТП (источник Google Maps); (в) прогноз задержки, вычисленный симулятором Auroga RNM (задержка указана в автомобиле-часах за 1 минуту); (г) контурные диаграммы скорости для двух крайних случаев, лучшего и худшего, вычисленные симулятором Auroga RNM (скорость указана в милях в час)

Стратегия 2 состоит в том, что к указанному выше регулированию при помощи ALINEA, с преодолением очереди, добавляется ограничение, требующее снижения скорости до уровня от 60 до 80 км/ч. Оно устанавливается на участке автострады непосредственно перед ДТП, обозначенном на рис. 15б. Результат прогноза, как видно из рис. 16б, оказывается лучше, чем в конфигурации без управления. Он однако хуже, чем у стратегии 1, а именно, задержка в худшем случае достигает 55, а в лучшем — 44 автомобиль-часов в минуту, причём соответствующие зоны затора на контурных диаграммах скорости оказываются больше, чем на рис. 16а.

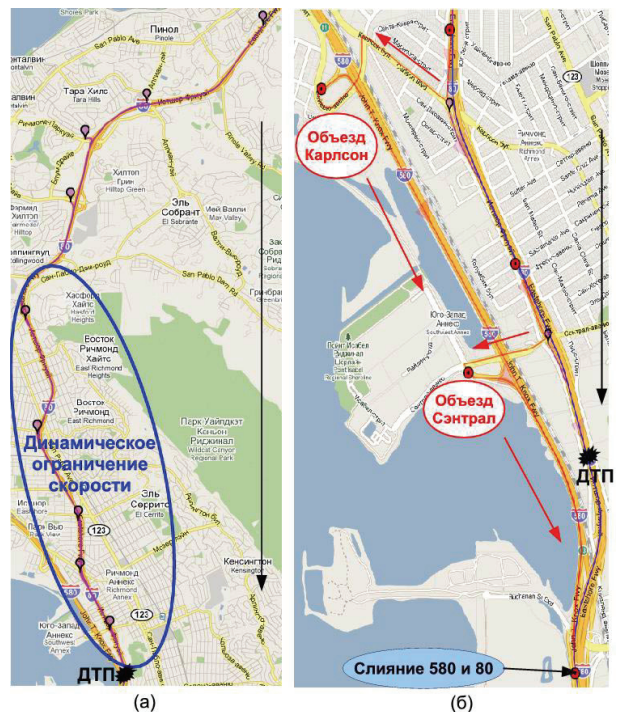


Рис. 15. Стратегии преодоления последствий ДТП на западном направлении автострады I-80, где (а) карта для стратегий 1 (регулирование въездных потоков) и 2 (регулирование въездных потоков в сочетании с ограничением скорости); (б) карта для стратегии 3 (регулирование въездных потоков в сочетании с принуждением к объезду)

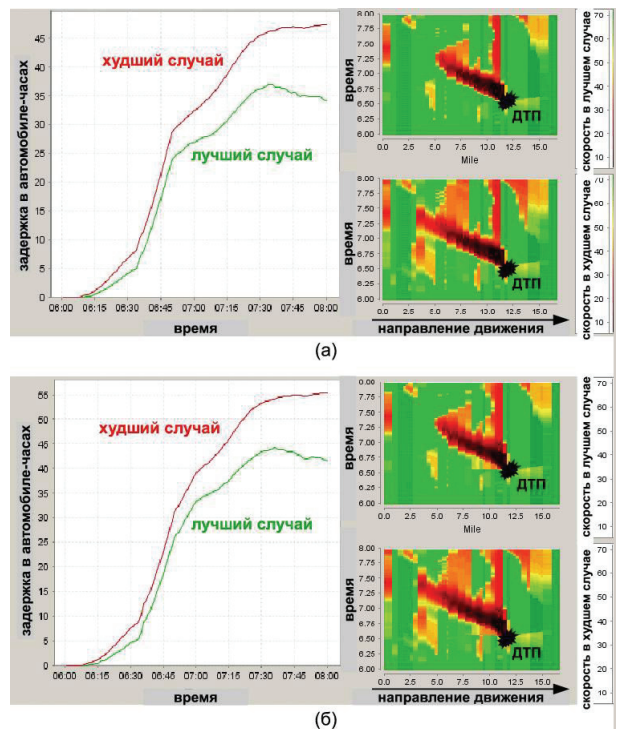


Рис. 16. Сравнение эффективности стратегий преодоления последствий ДТП на западном направлении автострады I-80, где (а) диапазон прогнозируемой задержки и скорости в случае ввода в действие стратегии 1; (б) диапазон прогнозируемой задержки и скорости в случае ввода в действие стратегии 2. На обоих графиках задержка указана в автомобиле-часах за 1 минуту, а скорость — в милях в час

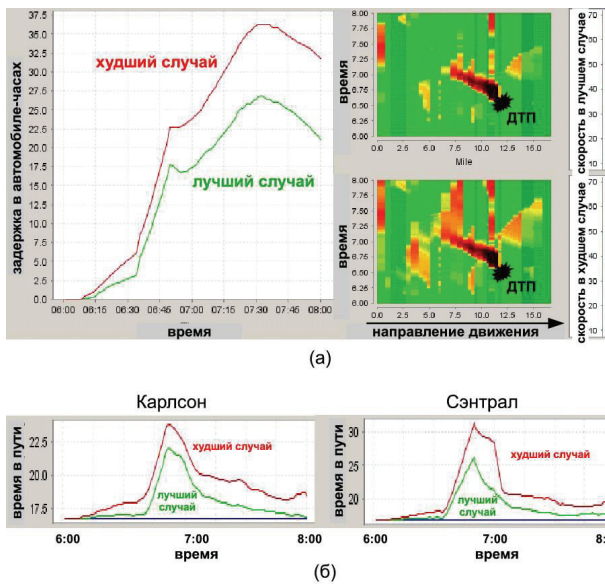


Рис. 17. Результат использования стратегии 3 для преодоления последствий ДТП на западном направлении автомагистрали I-80, где (а) диапазон прогнозируемой задержки и скорости в случае ввода в действие стратегии 3 (задержка указана в автомобиле-часах за 1 минуту, а скорость — в милях в час); (б) сравнение прогнозируемых диапазонов предполагаемого времени в пути для объездных маршрутов через Карлсон и Сэнтрал (предполагаемое время в пути указано в минутах)

Стратегия 3 интересна тем, что она специально разработана на случай ДТП на западном направлении автомагистрали I-80, в окрестности указанного места. Теперь в дополнение к регулированию ALINEA с преодолением очереди предлагается две возможности объезда зоны ДТП с последующим возвращением на I-80. Схема объездов представлена на рис. 15б. Дело в том, что место ДТП находится перед самым слиянием двух автомагистралей — западного направления I-80 и восточного направления I-580. До места соединения этих двух автомагистралей интенсивность движения на I-580 существенно ниже, чем на I-80. Поэтому её пропускную способность возможно использовать для перераспределения нагрузки. Система Augoa RNM позволяет очень быстро модифицировать сеть, добавив объездные пути и создав переключения в распределительных матрицах так, чтобы, к примеру, 25% автомобилей перенаправлялось с I-80 на I-580 через съезды соответственно на бульвар Карлсон и на Сэнтрал авеню. Оба объезда включают в себя городские улицы с регулируемыми перекрёстками, содержащими светофоры. Требуемый поток для въезда с I-580 на I-80 из первоначальной сети перенаправляется в объезд, на новый въезд к I-580 в таких размерах, чтобы не создавать затор на I-580. Как видно из рис. 17а, стратегия 3 существенно улучшает работу транспортной сети: задержка полной сети, включающей объезды, сократилась в худшем случае до 36, а в лучшем до 27 автомобиле-часов в минуту, а на контурных диаграммах скорости для маршрута без объезда зна-

чительно уменьшились зоны заторов. На рис. 17б приведены диапазоны предполагаемого времени в пути для маршрутов, включающих объезды через Карлсон и Сэнтрал: с точки зрения предполагаемых временных затрат объезд через Карлсон дешевле. Стратегия 3 является важным примером профилактики и контроля ДТП, о которых говорилось в параграфе III.

VII. Заключение

Данные измерений транспортных потоков как собранные за относительно длительный период (несколько месяцев или лет), так и поступающие в реальном времени, питающие быстрый и надёжный симулятор, позволяют создать базу для работы с аналитической моделью АУТС. Высокая скорость симуляций транспортных сетей должна позволить диспетчеру проанализировать десятки возможных оперативных стратегий в течение нескольких минут. Микромоделирование таким требованиям не отвечает. В данной статье предлагается использовать макромоделирование, которое кроме быстроты вычислений ещё и надёжно, — во-первых, потому что адекватно отображает поведение транспортной сети, и во-вторых, потому что параметры макромодели напрямую идентифицируются по данным измерений. При наличии необходимых измерений калибровка предлагаемой макромодели не требует особого искусства. Возможное отсутствие измерений на въездах и съездах конкретной автомагистрали частично компенсируется синтезом отсутствующих данных.

Поведение и продуктивность транспортной сети при различных сценариях и управляющих стратегиях оценивается при помощи таких параметров, как скорость потока, фактическое время в пути для соединений или конкретных маршрутов, а также таких, как суммарный поток, суммарная плотность, задержка и потеря продуктивности, которые могут быть вычислены как для отдельных соединений, так для конкретных маршрутов и для всей сети в целом. Для городских перекрёстков со светофорами существуют дополнительные оценки эффективности: задержка за цикл, размер очереди, мера использования фазы, нехватка цикла, отношение потока к максимальной пропускной способности и качество проезда.

Прямое управление транспортным потоком в нашей модели осуществляется в узлах сети. Локальное управление отвечает за поток, исходящий из одного конкретного входного соединения. Узловое управление отвечает за все входные потоки для данного узла. Комплексное управление координирует несколько локальных, узловых или комплексных управляющих элементов. Такая иерархия управления. Для симуляции алгоритмов управления с обратной связью и проверки их эффективности в близких к реальным услови-

ях мы подменяем обратную связь по состоянию обратной связью по измерению, используя виртуальные датчики. Виртуальные датчики моделируют работу реальных измерительных устройств, у которых качество измерений находится в диапазоне от идеального (полное отсутствие помех) до неудовлетворительного (одни сплошные помехи). Важно, что виртуальные датчики могут служить интерфейсом к реальным измерительным устройствам.

Симулятор имеет три рабочих режима. Первый — это режим оперативного планирования; он используется для проигрывания сценариев, разработки и тестирования оперативных стратегий. Второй — это режим динамического фильтра, используемый в реальном времени для улучшения качества данных в алгоритмах управления с обратной связью. Третий режим — это прогнозирование ближайшего будущего состояния системы, в котором рассчитываются ограничения на это состояние на ближайшие один-два часа. Они основываются на текущих измерениях, а также на предполагаемых значениях требуемых потоков и распределительных матриц при различных возможных оперативных стратегиях, дабы помочь диспетчеру выбрать наиболее подходящую для данной ситуации. Предлагаемый пример с автострадой I-80 в Северной Калифорнии иллюстрирует отдельные элементы АУТС в контексте данной статьи.

Динамическая модель, инструменты для построения транспортной сети, отдельные алгоритмы управления, локальные, узловые и комплексные, а также алгоритмы гарантированного прогнозирования и оценивания воплощены в бесплатном программном пакете Aurora RNM [2].

В заключение наметим возможные направления будущих исследований.

Для улучшения процесса калибровки модели необходимо разработать механизм оценки качества используемых измерений, так как параметры модели могут быть правильно идентифицированы лишь на основе данных с исправных датчиков. Даже такая качественная информационная система, как PeMS [13], которая сообщает данные об исправности датчиков, часто ошибочно выдаёт неисправные датчики за исправные.

Для оперативного планирования необходимо подразделять моделируемый в макроскопической среде транспортный поток на классы, как, например, общественный транспорт (автобусы, троллейбусы), тяжёлые грузовики, обычные автомобили и т.п., отдельные из которых могут использовать полосы специального назначения. Другие задачи состоят в быстром приблизительном расчёте эмиссий и потребления бензина, а также возможном влиянии оплаты проезда на транспортную ситуацию.

В области управления транспортным потоком стоит подробнее изучить такие существующие на

сегодняшний день противоречивые технологии, как динамическое ограничение скорости с точки зрения улучшения продуктивности транспортной сети. Также интересной задачей является координация регуляторов потоков при въездах на автострады с сигналами светофоров на ближайших крупных городских перекрёстках. Цель — выработка оперативных стратегий управления автострадой и окружающими городскими улицами как единой системой.

Важным элементом полностью функционального динамического фильтра является автоматический детектор сбоев для датчиков, работающий в реальном времени. В случае, если условие (42) не выполнено, должно быть принято решение, чему доверять — модели или измерениям. При этом датчики должны оцениваться не сами по себе, а во взаимодействии со своими соседями, с использованием данных об их исправности в прошлом, а также такой дополнительной информации, как, например, вызовы дорожно-патрульных служб и т.п.

Литература

1. *TOPL Project* <http://path.berkeley.edu/topl>
2. *Aurora RNM Homepage* <http://code.google.com/p/aurorarum>
3. *Daganzo C.F.* The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // *Transportation Research* — 1994. — V. 28, N. 4. — P. 269–287.
4. *Daganzo C.F.* The cell transmission model II: Network traffic // *Transportation Research* — 1995. — V. 29, N. 2. — P. 79–93.
5. *Greenshields B.* A study of highway capacity // *Proceedings of the 14th annual meeting of the Highway Research Board.* — 1934.
6. *Munõz L., Sun X., Horowitz R., Alvarez L.* Traffic density estimation with the cell transmission model // *Proceedings of the 2003 American Control Conference.* — 2003. — P. 3750–3755.
7. *Kerner B.S.* Experimental features of self-organization in traffic flow // *Physical Review Letters.* — 1998. — V. 81, N. 17.
8. *Orosz G., Wilson R.E., Szalai R., Stepan G.* Exciting traffic jams: Nonlinear phenomena behind traffic jam formation on highways // *Physical Review E.* — 2009. — V. 80, N. 4.
9. *Courant R., Friedrichs K., Lewy H.* Über die partiellen Differenzgleichungen der mathematischen Physik // *Mathematische Annalen.* — 1928. — V. 100, N. 1. — P. 32–74.
10. *Navteq Corporatio* <http://www.navteq.com>
11. *Tele Atlas* <http://www.teleatlas.com>
12. *OpenStreetMap Project* <http://www.openstreetmap.org>
13. *PeMS Homepage* <http://pems.eecs.berkeley.edu>

- 14.** *PORTAL* *Homepage*
http://portal.its.pdx.edu
- 15.** *Dervisoglu G., Gomes G., Kwon J., Muralidharan A., Varaiya P.* Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity // 88 Annual Meeting of the Transportation Research Board, Washington, D.C., USA. — 2008.
- 16.** *Chen C., Kwon J., Skabardonis A., Varaiya P.* Detecting errors and imputing missing data for single loop surveillance systems // Transportation Research Record. — 2003. — N. 1855. — P. 160–167.
- 17.** *Muralidharan A., Horowitz R.* Imputation of ramp data flow for freeway traffic simulation // Transportation Research Record. — 2009. — N. 2099. — P. 58–64.
- 18.** *Papageorgiou M., Hadj-Salem H., Blosseville H.M.* ALINEA: A local feedback control law for onramp metering // Transportation Research Record. — 1991. — N. 1320.
- 19.** *Papamichail I., Papageorgiou M., Vong V., Gaffney J.* HERO Coordinated Ramp Metering Implemented at Monash Freeway, Australia // 89th Annual Meeting of the Transportation Research Board, Washington, D.C., USA. — 2010.
- 20.** *Diakaki C., Papageorgiou M., McLean T.* Integrated Traffic-Responsive Urban Corridor Control Strategy in Glasgow, Scotland // Transportation Research Record. — 2000. — N. 1727. — P. 101–111.
- 21.** *Sensys* *Networks*
http://www.sensysnetworks.com
- 22.** *Papageorgiou M., Varaiya P.* Link vehicle-count — the missing measurement for traffic control // Proceedings of the 12th IFAC Symposium on Control in Transportation Systems. — 2009.
- 23.** *Lin W.H.* Spillovers, merging traffic and the morning commute // Proceedings of the 4th IEEE Intelligent Transportation Systems Conference, Oakland, CA. — 2001.
- 24.** *Okutani I., Stephanedes Y.J.* Dynamic prediction of traffic volume through Kalman filtering theory // Transportation Research, B. — 1984. — V. 18. — P. 1–11.
- 25.** *Smith B.L., Williams B.M., Oswald R.K.* Comparison of parametric and nonparametric models for traffic flow forecasting // Transportation Research, C. — 2002. — V. 10. — P. 303–321.
- 26.** *Коцеев А.С., Куржанский А.Б.* Адаптивное оценивание эволюции многошаговых систем в условиях неопределённости // Известия АН СССР. Техническая кибернетика. — 1983. — Т. 2. — С. 72–93.
- 27.** *Kurzhanski A.B.* Identification — a theory of guaranteed estimates // From Data to Model, (ed.) J.C. Willems, Springer Verlag. — 1989. — P. 135–214.
- 28.** *Milanese M., Norton J., Piet-Lahanier H., Walter É.* Bounding Approaches to System Identification // Springer Verlag. — 1996.
- 29.** *Kurzhanskiy A.A.* Set-valued estimation of freeway traffic density // Proceedings of the 12th IFAC Symposium on Control in Transportation Systems. — 2009.

Поступила в редакцию 15.10.2010.

Е.А. Нурминский^{1,2}, Н.Б. Шамрай^{1,2}

¹ Институт автоматизации и процессов управления ДВО РАН

² Дальневосточный федеральный университет

Прогнозное моделирование автомобильного трафика Владивостока

В ходе реализации масштабных федеральных программ развития Дальнего Востока России значительные средства вкладываются в инфраструктурные проекты совершенствования транспортных сетей г. Владивостока и его непосредственного окружения. В настоящей работе на основе генерального плана развития города разработаны прогнозные модели изменений транспортной ситуации и перераспределения транспортных потоков в результате реализации этих проектов. Описаны результаты вычислительных экспериментов по расчету транспортных потоков в существующей и создаваемой сетях с использованием теории экономического равновесия и модели стационарной динамики.

Ключевые слова: конкурентное равновесие, моделирование транспортных потоков, вариационные неравенства, матрица корреспонденций, проективные методы.

I. Введение

Моделирование городского автомобильного трафика имеет значительную историю и включает в себя громадный объем фактического материала, накопленного за десятилетия процессов активной автомобилизации. С 50-х годов прошлого столетия в этой области начали применяться модели, включающие различные принципы оптимальности и соответствующие численные методы. Знаковым событием было появление работы М. Beckmann, С.В. McGuire и С.В. Winsten, сначала как технического отчета корпорации RAND, а затем и монографии [1], ставшей в некотором смысле классикой моделирования автомобильного трафика. В отличие от многочисленных моделей, описывающих детали поведения потока автомобилей с той или иной степенью подробности, М. Beckmann, С.В. McGuire и С.В. Winsten взглянули на проблему с экономической точки зрения, когда в общем-то пренебрегают деталями технологического процесса, а именно, каким образом двигался поток автомобилей, а интересуются лишь конечным результатом — каковы были затраты при перемещении из начального пункта в конечный и как водитель в качестве рационального экономического агента должен вести себя в этой ситуации. Предполагая полную информированность водителя, его неспособность вступать в коалицию с другими пользователями сети и то, что влияние отдельного водителя на общий трафик пренебрежимо мало, авторы впервые математически выписали условия равновесия и показали их эквивалентность при определенных допущениях условиям оптимальности в некоторой вспомогательной искусственно построенной экстремальной задаче.

Альтернативным характером эгоистического поведения водителей является идеальная кооперация, направленная на получение максимально большого системного эффекта, например, ми-

нимизации суммарных транспортных издержек. В общем случае это приводит к совершенно иному распределению транспортных потоков по сети, и разность между бескоалиционными и кооперативными издержками можно трактовать как «цену анархии».

С тех пор в прогнозировании транспортных потоков сформировались две парадигмы моделирования, часто цитируемые как поведенческие принципы Вардропа [2]:

- 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.

Первый подход получил также название пользовательско-оптимального (user optimization, UO), второй — системно-оптимального (system optimization, SO).

Реально, по-видимому, реализуется некоторый промежуточный принцип поведения. Водителей автомобилей вынуждают при движении учитывать не только их собственные, но и системные интересы: на определенных участках транспортной сети вводят специальные режимы, перекрестки регулируются светофорами или работниками ГАИ и т.п. Это уменьшает «цену анархии», но оставляет в силе принципиальное различие между этими двумя подходами.

С тех пор UO-модели транспортных потоков развивались в различных направлениях, и состояние дел можно себе представить по обзору [3]. В той или иной степени эта методология, как правило, используется в процессах городского планирования (к сожалению, не в России) и реализована в различных программно-аппаратных

комплексах автоматизации транспортного проектирования (PTVision, AIMSUN, CONTRAM, TRANSIMS, TRANSNET и др.)

В данной работе изложены результаты проведенных исследований для прогнозирования автомобильного трафика на дорогах Владивостока. Описывается опыт моделирования транспортных потоков с позиции УО-методологии на основе двух подходов: монотонного вариационного неравенства и модели стационарной динамики [4]. Проводится сравнение полученных результатов с данными натурных обследований. Как показали численные расчеты, решение вариационного неравенства получилось весьма близким к эмпирическим данным. Модель стационарной динамики хоть и имеет ряд замечательных теоретических свойств, однако основана на некоторых экстремистских принципах, и ее применимость к реальным сетям вызывает некоторые вопросы. При помощи вариационного неравенства моделируется автомобильный трафик в модернизированной транспортной сети. Численные расчеты показали, что запланированные в генеральном плане развития Владивостока работы по строительству новых дорог улучшат транспортную ситуацию в городе.

II. Общая транспортная ситуация во Владивостоке

В данной работе улично-дорожная сеть (УДС) Владивостока представлена ориентированным графом, географическая привязка которого осуществлялась на основе данных, предоставленных геодезическим предприятием ФГУП «ПримАГП». Некоторые характеристики этого графа, отражающие топологию сети, приведены в табл. 1.

Т а б л и ц а 1

Характеристики УДС г. Владивостока

Степени вершин					Всего	Дуг
1	2	3	4	5		
1274	241	2521	245	9	4290	5172

Метрические характеристики УДС можно определить общей протяженностью дорог в пределах территории города, которая составляет 1143,37 км, и средним расстоянием между перекрестками, которое равно 412 м.

Интересной особенностью УДС является большое (2521) количество вершин степени 3, что характерно для естественных графовых структур (трещин на поверхности земли, речных сетей, кровеносных сосудов и пр.), и говорит о том, что в большей степени топология УДС города складывалась стихийно. Любопытно также, что это число практически совпадает с числом вершин бинарного дерева, имеющего такое же как в УДС количество терминальных вершин (1274). План основной части УДС с соблюдением пространственных пропорций представлен на рис. 1, где хорошо видна также неоднородность и хаотичность сети.



Рис. 1. Графовая модель УДС г. Владивостока

Помимо неразвитой дорожной сети ситуация усугубляется самой высокой в РФ плотностью автомобилей, составляющей по некоторым данным [5] 566 автомобилей на 1000 человек городского населения.

Подавляющее большинство участников дорожного движения в часы пик во Владивостоке составляют легковые автомобили и миниавтобусы, 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. Вместе с тем маятниковые поездки трудовой миграции имеют ряд особенностей, делающих их удобным объектом для моделирования. В первую очередь в силу их повторяемости характеристики таких поездок можно считать стационарными, а самих водителей — имеющих полную информацию о возможных издержках на различных маршрутах. Также разумно предполагать совершенный эгоизм участников этих поездок и стремление минимизировать потери времени на проезд. Очевидно и отсутствие коалиционного поведения, за исключением разовых акций, которые в общем-то не связаны с ежедневными регулярными поездками.

Поэтому при исследовании транспортной ситуации во Владивостоке в качестве основной загрузки УДС в утренне-вечерние часы пик были рассмотрены транспортные потоки, порождаемые легковым частным транспортом при трудовой миграции, а в качестве поведенческой модели использован принцип пользовательской оптимизации.

III. Математическая модель конкурентного равновесия

Как уже упоминалось выше, для разработки математической модели конкурентного потокового равновесия УДС Владивостока была представлена ориентированным графом $\Gamma(V,A)$, где V —

множество вершин, A — множество дуг графа. Каждая дуга соответствует реальному участку автодороги без перекрестков, вершины представляют узлы, разделяющие участки дорог. Вершины степени 2 соответствуют существенным поворотам дорог, что для условий рельефа Владивостока также имеет значение при расчете временных затрат на проезд. Направление дуги определяет ход следования автотранспорта по участку дороги. Магистралы с двусторонним движением соответственно имеют парные противоположно ориентированные дуги.

При исследовании потокообразующих факторов в множестве вершин V были выделены два подмножества: $S \subseteq V$, содержащее вершины сети, порождающие потоки, то есть такие, для которых баланс выходящих и входящих потоков положителен, и $D \subseteq V$, содержащее вершины, поглощающие потоки, то есть такие, для которых этот баланс отрицателен. Вершины, принадлежащие S , назовем источниками, принадлежащие D — стоками. Остальные вершины являются транзитными. Применительно к задаче моделирования потоков, порожденных ежедневной трудовой миграцией в утренне-вечерние часы пик, источниками являются спальные районы и пригороды, стоками — деловые и производственные районы города. Декартово произведение $S \times D$ образует множество всех потокообразующих пар в сети Γ .

Существуют различные модели, определяющие объем транспортных перевозок. В данном случае использовался простейший вариант, так называемая модель с фиксированным спросом на перевозки, когда для каждой пары источник–сток $(s, d) \in S \times D$ предполагается заданным общий объем корреспонденции ρ_{sd} , то есть поток, который выходит из узла s и должен прибыть в пункт d . Набор $\rho = \{\rho_{sd} : (s, d) \in S \times D\}$ называется матрицей корреспонденций. При моделировании транспортных потоков рассматриваются и иные варианты, так называемые модели с эластичным спросом (см., напр., [6]), где матрица корреспонденций сама зависит от затрат на проезд. Для Владивостока при достаточно устойчивой концентрации бизнеса в определенных районах города и различного рода обстоятельств, реально сильно ограничивающих мобильность населения, такие модели на данном этапе не рассматривались.

Путем (маршрутом) в сети Γ , соединяющим вершины s и d , назовем последовательность дуг $a_1 = (s \rightarrow k_1)$, $a_2 = (k_1 \rightarrow k_2)$, ..., $a_l = (k_{l-1} \rightarrow k_l)$, $a_{l+1} = (k_l \rightarrow d)$, где $a_t \in A$ при всех $t = 1, \dots, l + 1$. Конечно, в маршрутах предполагается отсутствие петель и циклов. Обозначим через P_{sd} множество альтернативных маршрутов, следуя которым для каждой пары $(s, d) \in S \times D$ исходящий из источника s поток достигает стока d . Совокупность всех путей в сети Γ обозначим через $P = \bigcup_{(s, d) \in S \times D} P_{sd}$.

Пусть x_p — это величина потока по пути $p \in P$. Для каждой пары $(s, d) \in S \times D$ объемы корреспонденций ρ_{sd} порождают потоки x_p , которые должны удовлетворять условиям баланса и неотрицательности:

$$X_{sd} = \left\{ x_p \geq 0 : p \in P_{sd}, \sum_{p \in P_{sd}} x_p = \rho_{sd} \right\}.$$

Объединим величины x_p в вектор $x = \{x_p : p \in P\}$, тогда допустимой областью для вектора x является множество

$$X = \prod_{(s, d) \in S \times D} X_{sd} = \left\{ x \geq 0 : \sum_{p \in P_{sd}} x_p = \rho_{sd}, (s, d) \in S \times D \right\}. \quad (1)$$

Проезд по каждому из путей $p \in P$ требует определенных транспортных затрат, зависящих как от характеристик маршрута, так и от интенсивности и плотности движения в сети. Обозначим через $G_p(x)$ удельные затраты на проезд по пути p при существующих потоках x , то есть затраты на перемещение одного транспортного средства по маршруту p .

Во введенных обозначениях равновесные потоки $x^* \in X$ в условиях пользовательской оптимизации характеризуются следующим образом: водители выбирают путь с наименьшими транспортными издержками, поэтому для каждой пары $(s, d) \in S \times D$ если по пути $p \in P_{sd}$ идет ненулевой поток $x_p^* > 0$, то затраты по этому пути минимальны:

$$G_p(x^*) = \min_{q \in P_{sd}} G_q(x^*) = u_{sd}(x^*), \quad (2)$$

где $u_{sd}(x^*)$ — минимальные транспортные затраты для пары (s, d) при загрузке сети потоками x^* .

Удобной и общепринятой математической переформулировкой принципа пользовательской оптимальности является вариационное неравенство (см., напр., [6–8]): найти $x^* \in X$ такой, что

$$\sum_{p \in P} G_p(x^*)(x_p - x_p^*) \geq 0, \quad \forall x \in X. \quad (3)$$

Описание равновесных потоков в виде решения вариационного неравенства позволяет использовать теоретический и алгоритмический аппараты последних для исследования задач транспортного равновесия.

III.1. Функции задержек

Сложность решения задачи (2) критическим образом зависит от характера зависимости удельных затрат $G_p(x)$ от загрузки сети x и в общем случае представляется весьма высокой. Одним из

первых упрощающих предположений является аддитивность затрат по маршруту в зависимости от затрат по дугам, входящим в его состав:

$$G_p(x) = \sum_{a \in A_p} \tau_a(x), \quad (4)$$

где $A_p \subseteq A$ — множество дуг, входящих в маршрут p , $\tau_a(x)$ — удельные затраты на проезд по дуге a при загрузке сети x . Наиболее распространенная форма задания функций $\tau_a(x)$, принятая также и в настоящей работе, выражает временные затраты на проезд по дуге a , монотонно зависящие только от потока y_a по этой дуге:

$$\tau_a(x) = t_a(y_a), \quad y_a = \sum_{p \in P_a} x_p,$$

где $P_a \subseteq P$ — множество маршрутов, проходящих по дуге a . В практике анализа транспортных сетей наиболее популярной является эмпирическая зависимость, известная как BPR-функция (Bureau of Public Road):

$$t_a(y_a) = \bar{\tau}_a(1 + \mu y_a / \bar{y}_a)^n, \quad (5)$$

где $\bar{\tau}_a$ — время проезда по свободной от трафика дороге, \bar{y}_a — пропускная способность дороги, $\mu > 0$ — масштабирующий коэффициент, n — показатель степени, отражающий эффект прогрессивного замедления движения при росте загруженности [9]. Пропускная способность определяет максимальное количество автомобилей, способных проехать по участку дороги в единицу времени. За минимальное время проезда принимают продолжительность движения одного автомобиля с максимальной разрешенной скоростью по свободной дороге. Типично рассматривают значения n порядка 4, коэффициент μ принимают равным 1.

Вместе с тем с формальной точки зрения монотонная функция затрат $\tau_a(x)$ допускает сколь угодно большие значения потоков по магистралям города, что едва согласуется с реальностью и справедливо критикуется. Одна из попыток избежать представление τ_a от упомянутого недостатка описана в работе [4]. В этой работе для каждой дуги $a \in A$ транспортной сети предлагается ввести два вполне измеримых показателя: пропускную способность \bar{y}_a и минимальное время проезда $\bar{\tau}_a$. Далее естественным образом предполагается, что в транспортной сети поток по дуге не может превышать ее пропускную способность, а потраченное на проезд время не может быть меньше, чем минимальное время свободного проезда. Согласно [4], ситуация потокового равновесия в транспортной сети теперь определяется как загрузка ее дуг $y^* = \{y_a^* : a \in A\}$ и временные задержки на дугах $\tau^* = \{\tau_a^* : a \in A\}$, удовлетворяющие ограничениям:

$$0 \leq y_a^* \leq \bar{y}_a, \quad \tau_a^* \geq \bar{\tau}_a, \quad a \in A,$$

и для которых выполнены условия пользовательской оптимальности (2):

$$\tau_a^* \begin{cases} = \bar{\tau}_a, & \text{если } y_a^* < \bar{y}_a, \\ \geq \bar{\tau}_a, & \text{если } y_a^* = \bar{y}_a. \end{cases} \quad (6)$$

Условие (6) показывает, что задержки зависят от потока, но эта зависимость не является непрерывной. В случае полной загрузки сети можно гарантировать лишь только то, что задержки на дугах будут не меньше минимального времени проезда.

Равновесные потоки y^* предлагается искать как решение задачи линейного программирования [4]:

$$\begin{aligned} \sum_{a \in A} \bar{\tau}_a y_a &\rightarrow \min, \\ y_a &= \sum_{s \in S} y_a^s \leq \bar{y}_a, \quad y_a^s \geq 0, \quad a \in A, \quad s \in S, \\ \sum_{a \in A_v^+} y_a^s - \sum_{a \in A_v^-} y_a^s &= \rho_{sv}, \quad (s, v) \in S \times V, \end{aligned} \quad (7)$$

где y_a^s — поток по дуге a , порождаемый источником $s \in S$, $A_v^+ = \{a \in A : \text{дуга } a \text{ входит в вершину } v\}$, $A_v^- = \{a \in A : \text{дуга } a \text{ выходит из вершины } v\}$.

Условие (6) независимости времени проезда по дуге от загрузки вплоть до достижения предельного значения выглядит весьма идеализированным, поскольку реальная практика вождения показывает, что с увеличением числа автомобилей на дороге скорость движения все-таки уменьшается. Поэтому вопрос о степени применимости этого подхода к транспортной системе Владивостока представлял самостоятельный интерес и также рассмотрен в данной работе.

III.2. Фактографическое обеспечение моделей

Дальнейшая работа над задачей (2) заключалась в следующих этапах:

- 1) построении графовой модели УДС Владивостока;
- 2) выделении в ней множеств источников S и стоков D и определении значений элементов матрицы корреспонденций ρ_{sd} ;
- 3) определении множества маршрутов P ;
- 4) вычисления равновесных потоков x^* как решения двух альтернативных задач — вариационного неравенства (3) и задачи линейного программирования (7).

Исходная графовая модель УДС Владивостока, изображенная на рис. 1, была построена по данным геоинформационных систем и включала дороги четырех категорий, в том числе внутриквартальные подъезды к домам, грунтовые дороги через лесные массивы. Хотя на практике и эти пути используются для сквозного движения или объезда пробок, но все же они не играют особой роли в переносе основного потока автотранспорта

и при расчетах использованы не были. В результате для расчетов рассмотрена сеть, состоящая из 537 вершин и 1128 ориентированных дуг, изображенная на рис. 2. В рассматриваемой транспортной сети было выделено 37 потокообразующих вершин, каждая из которых одновременно являлась источником и стоком.

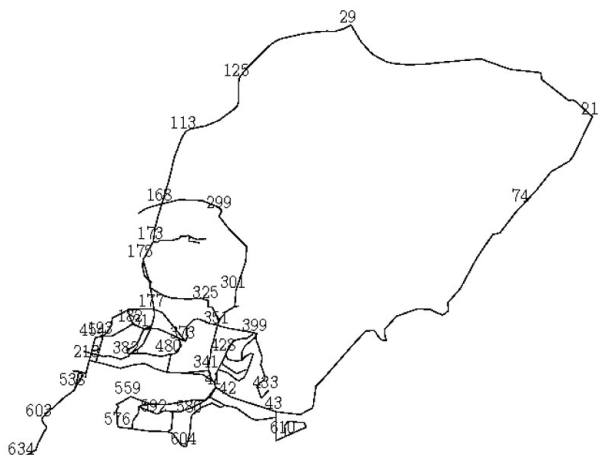


Рис. 2. Графовая модель УДС Владивостока, дороги категорий 3 и 4. Пронумерованы потокообразующие вершины

Для определения объемов корреспонденций между выделенными вершинами использовалась так называемая модифицированная гравитационная модель (см., напр., [10, 11]), представляющая матрицу корреспонденций в виде

$$\rho_{sd} = \alpha_s \beta_d Q_s R_d f(c_{sd}),$$

где параметры α_s и β_d находятся из решения системы уравнений

$$\alpha_s = \left(\sum_{d \in D} \beta_d R_d f(c_{sd}) \right)^{-1},$$

$$\beta_d = \left(\sum_{s \in S} \alpha_s Q_s f(c_{sd}) \right)^{-1}, \quad (8)$$

Q_s — общий объем исходящего потока из источника $s \in S$, R_d — общий объем входящего потока в сток $d \in D$, c_{sd} — мера удельных затрат на передвижение для пары $(s, d) \in S \times D$, $f(c_{sd})$ — функция, характеризующая предпочтение индивидуумов по перемещению из пункта s в пункт d .

Для определения объемов Q_s и R_d территория Владивостока была поделена на квадратные зоны 800×800 м. Интерактивная карта города, фрагмент которой изображен на рис. 3, позволила получить детальную информацию о застройке каждой зоны: тип здания (жилое, производственное предприятие, образовательное учреждение, торговый комплекс, бизнес-центр, медицинское учреждение и т.п.) и его этажность, что помогло подсчитать количество жилых домов разной этажности, выявить предприятия, организации и учреждения, формирующие рабочие места в каждой из рассматриваемых зон.

По таблице соответствия этажности здания и проживающих в нем людей был определен объем населения в каждой зоне. В результате таких подсчетов, суммируя количества проживающих по всем зонам, общий объем населения Владивостока получился равным 587 768 человекам, что всего лишь на 1,5% отличается от данных переписи 2008 года. Согласно официальной статистике уровень занятости населения в экономике Владивостока составляет 60% от общего числа жителей. Предполагалось, что именно эта часть населения города порождает трудовую миграцию. Учитывая полученные с помощью интерактивной карты данные по типам предприятий, организаций, учреждений и используя информацию из статистических сборников, аналитических записок и деловой прессы Владивостока, было определено количество рабочих мест в каждом квадрате и построена зональная матрица отправлений и прибытий для трудовой миграции по 638 квадратам. Далее, за каждой из 37 выделенных вершин в графе были закреплены близлежащие зоны, что дало объемы исходящих Q_s и входящих R_d потоков для этих вершин.



Рис. 3. Пример одной квадратной зоны на интерактивной карте г. Владивостока

В качестве удельных затрат c_{sd} на передвижение из пункта s в пункт d рассмотрено время, необходимое для того, чтобы добраться из s в d по кратчайшему пути со скоростью 40 км/ч. Функция $f(c_{sd})$ определена в традиционной для трудовой миграции форме: $f(c_{sd}) = \exp(-\gamma c_{sd})$, где параметр калибровки $\gamma = 0,065$.

Значения параметров α_s и β_d , удовлетворяющих системе уравнений (8), находились с помощью алгоритма балансировки [12], что завершило построение матрицы корреспонденций $\rho = (\rho_{sd} : (s, d) \in S \times D)$.

Для проверки достоверности сделанных расчетов в «узких» местах УДС Владивостока, то есть на дугах, движение по которым является един-

ственно возможным способом попасть из одной агрегированной территориальной зоны в другую, были сделаны натурные замеры потока автотранспорта. Максимальная абсолютная ошибка в подсчетах была равна 382 автомобилям, или 14% от наблюдаемого значения, что при таких приближенных оценках можно считать вполне приемлемым.

Поиск транспортного равновесия как решения вариационного неравенства (3) предполагает задание множества допустимых маршрутов для каждой пары источник–сток. В развитых транспортных сетях реальных масштабов для одной пары источник–сток можно построить комбинаторно большое количество возможных вариантов передвижения, но учитывая специфику УДС Владивостока, имеющую протяженную (в одном направлении) древовидную структуру и не имеющую большого количества дублирующих маршрутов, множество маршрутов P строилось с помощью алгоритма Йена поиска k -кратчайших путей, где максимально возможное значение k принято равным 10. В результате было выделено 13 202 пути.

Пропускная способность \bar{y}_a и минимальное время свободного проезда $\bar{\tau}_a$ для каждой дуги $a \in A$ соответствовали нормативной пропускной способности одной полосы в 1500 автомобилей в час и скорости движения 40 км/ч.

III.3. Алгоритмическое обеспечение

Эквивалентность задачи транспортного равновесия вариационному неравенству, а в частном случае оптимизационной задаче, позволяет адаптировать численные методы решения последних для поиска равновесных потоков.

К настоящему времени можно выделить два основных подхода к построению алгоритмических схем: в первом поиск равновесных потоков ведется по дугам сети (дуговые алгоритмы), во втором — по допустимым маршрутам (маршрутные алгоритмы). И в первом, и 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 с искусственной целевой функцией $f(y) = \sum_{a \in A} \int_0^{y_a} \tau_a(z) dz$ и линейными ограничениями-балансами (1). Наиболее распространенным дуговым алгоритмом является метод Франка–Вульфа [13], 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 линей-

ного программирования, так и неравномерная сходимость потоков к равновесным значениям или так называемый эффект «застревающих потоков» [14–16] — в процессе решения формируется некоторый набор дуг, по которым потоки сильно отличаются от равновесных, и такая ситуация медленно меняется при последующем итерировании.

Маршрутные алгоритмы распределяют корреспонденции непосредственно по множеству альтернативных путей, причем это множество, как правило, формируется в процессе решения [17, 18]. Перераспределение потоков не по дугам, а сразу по маршрутам позволяет своевременно уйти от «застревающих потоков», поэтому алгоритмы данного класса не обладают отмеченным недостатком метода Франка–Вульфа и сходятся более 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 потокообразующих 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 [14, 18], 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. Хотя алгоритмический аппарат для вариационных неравенств разработан достаточно хорошо, о чем свидетельствуют монографии [19, 20], но большое количество переменных, непотенциальные и неаддитивные функции транспортных затрат на практике делают задачу (3) труднорешаемой. Среди существующих методов решения вариационных неравенств отдельно можно выделить проективные методы, отличающиеся простотой своих итерационных схем:

$$x^{k+1} = \pi_X(x^k - \lambda_k G(x^k)), \quad k = 0, 1, 2, \dots, \quad (9)$$

где $\pi_X(y) = \operatorname{argmin}\{\|y - x\| : x \in X\}$ — проекция точки y на множество X , $\lambda_k > 0$ — шаговый множитель. 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 равно-
весному распределению при весьма общих предпо-
ложениях о свойствах задачи, однако на практи-
ке такой выбор ведет к очень медленной скорости
сходимости.

Несмотря на то, что при моделировании тра-
фика в УДС Владивостока предполагалось, что
транспортные задержки по дуге зависят от за-
грузки только этой дуги и задача транспортного
равновесия эквивалентна оптимизационной зада-
че, для численных расчетов использовался проек-
тивный метод (9) с применением идеи декомпози-
ции допустимого множества X . Такой выбор мето-
да (9) сделан в расчете на будущие исследования,
поскольку более адекватное моделирование все-
таки требует рассмотрения общего случая непо-
тенциальной функции $G(x) = (G_p(x) : p \in P)$ и
решения именно вариационного неравенства (3).

Для декомпозиции и ускорения сходимости
процесса (9) предполагается применить подходы,
основанные на теории фейеровских процессов с
малыми возмущениями [21, 22] с использованием
адаптивной регулировки шага [23]. Основная идея
этого подхода заключается в следующем. Допу-
стимое множество X из (1) можно представить
в виде пересечения конечного числа гиперплоско-
стей H_{sd} и неотрицательного ортанта H^+ :

$$X = \bigcap_{(s,d) \in S \times D} H_{sd} \cap H^+,$$

где $H_{sd} = \{x_p : \sum_{p \in P_{sd}} x_p = \rho_{sd}\}$, $H^+ = \{x_p \geq 0 : p \in P\}$. Объединим супермножества H_{sd} и H^+ в семейство множеств $\mathcal{H} = \{H^+, H_{sd} : (s,d) \in S \times D\} = \{\mathcal{H}_i : i = 1, 2, \dots, |S \times D| + 1\}$. Опера-
ция проектирования $\pi_{\mathcal{H}_i}(\cdot)$ для любого элемента \mathcal{H}_i вычисляется аналитически. Поэтому для чис-
ленных расчетов транспортных потоков исполь-
зовалась следующая модификация процесса (11),
получившая название метода последовательных
проекции [24]:

$$x^{k+1} = x^k + \lambda_k v^k, \quad v^k = (\pi_{\mathcal{H}_i}(\bar{x}^k) - x^k) / \lambda_k,$$

$$\mathcal{H}_i \in \mathcal{H}, \bar{x}^k = x^k - \lambda_k G(x^k) \notin \mathcal{H}_i, \quad (10)$$

$$\lambda_k > 0, k = 0, 1, 2, \dots$$

Значительного ускорения сходимости процес-
са (10) к равносному решению удается достичь
за счет адаптивного 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 $k_t = 0$, $\lambda_0 > 0$ — произ-
вольное, $q \in (0,1)$;
- 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 по-
вторить вычисление (10) для текущего зна-
чения λ_k .

Другими словами, по условию п. 2 первый пе-
реход 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, согласно п. 3,
в $q < 1$ раз.

IV. Численные эксперименты

IV.1. Расчеты для существующей транспортной сети

Размерность решаемых задач поиска транс-
портного равновесия в рассматриваемой транс-
портной сети приведена в табл. 2, откуда видно,
что «габариты» задач даже для сокращенной сети
весьма велики.

Т а б л и ц а 2

Размерность решаемых задач

	Задача (3)	Задача (7)
Неотрицательные переменные	13202	41736
Балансовые ограничения	1332	19869
Ограничения пропускной способности	—	1128

Транспортные затраты $G_p(x)$ предполагались
аддитивными (4), задержки по дугам имели вид
BPR-функции (5). Для решения вариационно-
го неравенства (3) средствами CNU Octave [25]
был реализован метод последовательных проек-
ций (10). Для решения задачи линейного програм-
мирования (7) использовался пакет GNU Linear
Programming Kit [26].

Прежде всего отметим, что на исходных дан-
ных задача (7) оказалась недопустимой. Причи-
ной недопустимости являлось ограничение на про-
пускную способность сети — максимальная про-
пускная способность дуг \bar{y}_a не позволила пропу-
стить рассчитанные для часа пик корреспонден-
ции трудовой миграции ρ_{sd} . Для выявления про-
блемных мест в сети задача (7) решалась без учета
ограничений $y_a \leq \bar{y}_a$, а именно

$$\sum_{a \in A} \bar{\tau}_a \sum_{s \in S} y_a^s \rightarrow \min,$$

$$\sum_{a \in A_v^+} y_a^s - \sum_{a \in A_v^-} y_a^s = \rho_{sv}, \quad (s, v) \in S \times V,$$

$$y_a^s \geq 0, \quad a \in A, s \in S. \quad (11)$$

На рис. 4 непрерывным черным цветом выделены наиболее проблемные участки УДС Владивостока, для которых, чтобы пропустить корреспонденции ρ_{sd} , необходимо расширить дорогу более чем на половину полосы движения в одну сторону.



Рис. 4. Наиболее проблемные участки УДС г. Владивостока. Модель стационарной динамики

Всего получено 97 проблемных участков, среди которых присутствуют дороги, требующие расширения на целых три полосы движения в одну сторону. Учитывая тот факт, что все эти 97 дуг принадлежат к магистралям центральной части города, то такое масштабное расширение практически невозможно. После того как пропускную способность всей сети увеличили в три раза, решение задачи (7) было найдено, при этом среди условий на пропускную способность остались активные ограничения.

Вариационное неравенство (3) для рассматриваемой задачи транспортного равновесия (2) всегда разрешимо, и алгоритм последовательных проекций приводит с определенной точностью к равновесным потокам. Характер сходимости процесса (10) проиллюстрирован на рис. 5, откуда видно, что сходимость имеет степенной характер, то есть точность решения ведет себя как $O(k^{-\alpha})$, где k — количество итераций, α — показатель степени скорости сходимости.

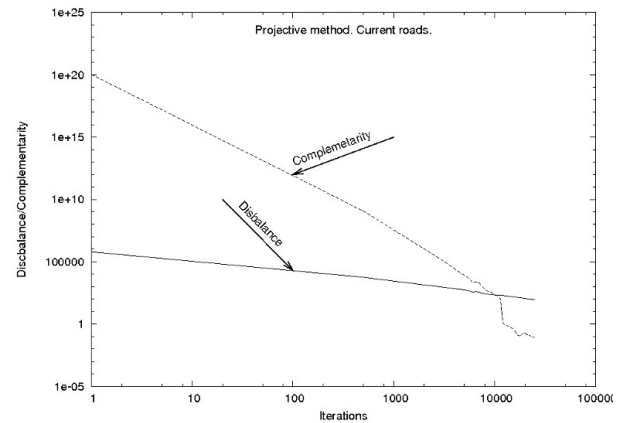


Рис. 5. Сходимость метода последовательных проекций

Сравнение полученных значений загрузки дуг сети y_a^* с их пропускной способностью \bar{y}_a также выявило наличие элементов, на которых поток превышает пропускную способность. Однако в отличие от решения задачи (7) таких элементов всего три, все они также принадлежат центральной части города.

Последующий анализ состоял в сравнении рассчитанных потоков с натурными замерами на магистралях города. В утренний час пик на 16 участках УДС Владивостока в течение 60 минут подсчитывались проезжающие легковые автомобили. На рис. 6 представлена гистограмма, отображающая рассчитанные и наблюдаемые потоки для каждого из участков.

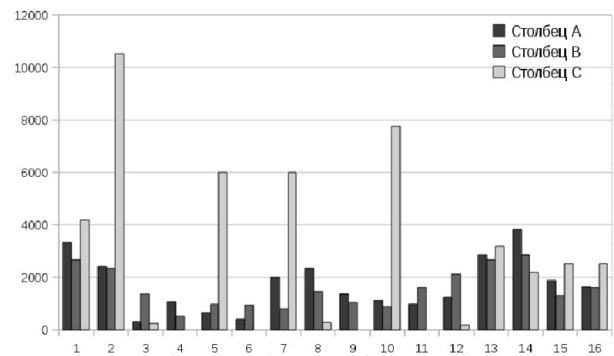


Рис. 6. Гистограмма для рассчитанных и наблюдаемых потоков

Столбец *A* характеризует рассчитанные транспортные потоки, полученные путем решения вариационного неравенства (3) с непрерывными монотонными функциями задержек (5). Столбец *B* характеризует наблюдаемые потоки в час пик на улицах города. Столбец *C* характеризует потоки, полученные в результате решения задачи линейного программирования (11), функции задержек определяются условием (6). Как видно из гистограммы, результаты расчетов для вариационного неравенства более близки к натурным замерам, чем результаты, полученные вследствие решения задачи (11).

Такой вывод подтверждает и гистограмма, изображенная на рис. 7, где столбец *A* и столбец

B представляют соответственно отклонения произведенных расчетов при решении вариационного неравенства (3) и задачи линейного программирования (11) от натуральных замеров.

Максимальная абсолютная ошибка вычислений, полученная при решении вариационного неравенства, равна 1298 автомобилям, что означает превышение в 2.5 раза рассчитанного потока над наблюдаемым, или относительную ошибку 150%. Максимальная абсолютная ошибка вычислений, полученная при решении задачи линейного программирования (11), равна 8161 автомобилю, что означает превышение в 4.5 раза рассчитанного потока над наблюдаемым, или относительную ошибку 350%.

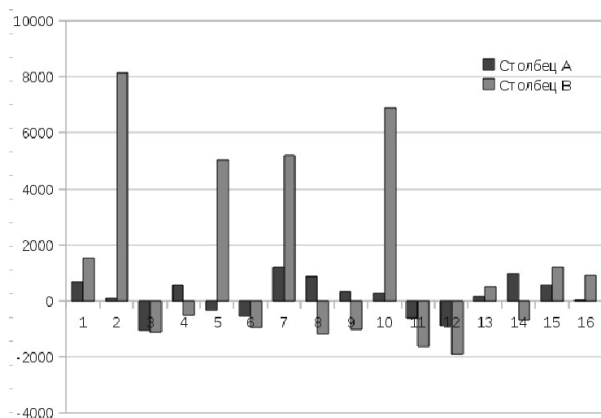


Рис. 7. Гистограмма отклонения рассчитанных потоков от наблюдаемых

Минимальная абсолютная ошибка вычислений, полученная при решении вариационного неравенства, равна 34 автомобилям, что означает превышение в 1.02 раза рассчитанного потока над наблюдаемым, или относительную ошибку 2%. Минимальная абсолютная ошибка вычислений, полученная при решении задачи линейного программирования (11), равна 498 автомобилям, что означает превышение в 1.19 раза рассчитанного потока над наблюдаемым, или относительную ошибку 19%.

Среднее отклонение рассчитанных потоков от натуральных замеров составило 540.38 автомобиля при решении вариационного неравенства (3) и 2551.4 автомобиля — при решении задачи линейного программирования (11).

IV.2. Прогноз загрузки перспективной сети

Генеральный план развития транспортной инфраструктуры Владивостока до 2025 года включает три очередности модернизации УДС города — первая очередность, на расчетный срок (до 2012 г.) и дальнейшие изменения вплоть до 2025 года [27]. При анализе влияния на транспортную ситуацию ввода в эксплуатацию новых дорог учитывались только изменения первой очереди. На рис. 8 представлена УДС Владивостока ближайшего будущего, где черным цветом выделены как

новые, так и существующие участки магистралей, которые подлежат ремонту и расширению.

По произведенным подсчетам общая протяженность модернизированных дорог равна 33.72 км, в том числе будет введено в эксплуатацию 16.65 км новых магистралей. В настоящее время идет активное строительство выделенных участков.

Обозначим через $\tilde{\Gamma} = (\tilde{A}, \tilde{V})$ граф, описывающий перспективную УДС первой очередности строительства. Предварительный анализ показал, что для множества вершин V суммарная длина кратчайших путей в графе $\tilde{\Gamma}$ уменьшилась на 4%. В среднем кратчайшие пути в новой сети стали короче на 415 м. Кратчайший путь сократился для 17% пар вершин из общего числа 288 369.

Поскольку в существующей транспортной сети определение равновесных потоков как решения вариационного неравенства (3) привело к более реалистичным результатам, чем расчеты по модели стационарной динамики, для вычисления загрузки новой сети $\tilde{\Gamma}$ далее использовался метод последовательных проекций (10).



Рис. 8. Перспективная УДС Владивостока

Безусловно новые дороги станут побудительным мотивом для развития инфраструктуры города, появятся новые бизнес-зоны, жилые массивы. Однако предположим, что это произойдет не так быстро и в ближайшие пару лет сформированная матрица корреспонденций ρ не претерпит существенных изменений. Поэтому загрузку перспективной сети $\tilde{\Gamma}$ будем определять для тех же пар потокообразующих вершин $(s, d) \in S \times D$ с фиксированными объемами корреспонденций ρ_{sd} . Вид функций транспортных затрат также оставим прежним (5).

На рис. 9 показано, как изменится загрузка существующих дуг A в перспективной сети $\tilde{\Gamma}$: черным цветом выделены участки, на которых поток уменьшится, серым — участки, на которых поток увеличится, пунктиром обозначены новые дороги.

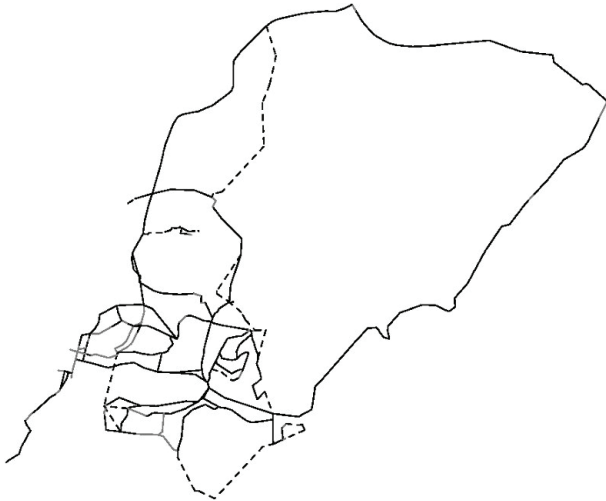


Рис. 9. Изменение загрузки дуг перспективной сети

Из рисунка видно, что практически на всех дугах объемы потоков уменьшились. Однако существуют и участки, по которым поток увеличился, а значит, возросли транспортные затраты. Но это не портит общей положительной картины, поскольку системные задержки в новой сети сократятся на 3090 часов. Проведенные расчеты показывают, что строительство новых дорог не приведет к транспортным парадоксам и ситуация на дорогах города существенно улучшится.

В перспективной сети осталось четыре дуги, по которым поток превышает пропускную способность, однако превышение получилось незначительным: минимальное составило 170, максимальное — 755 автомобилей.

V. Заключение

По результатам выполненной работы можно сделать следующие выводы. Модель равновесных конкурирующих потоков с непрерывными монотонными функциями задержек лучше описывает усредненные значения транспортных потоков трудовой миграции, чем модель, в которой задержки удовлетворяют условию (6). Отсутствие ограничения на пропускную способность дуг транспортной сети при решении вариационного неравенства (3) не приводит к абсурдным решениям, поскольку регулирующую роль на пропускную способность оказывает свойство монотонной зависимости задержек от величины потока. Есть полное основание считать, что дальнейшее уточнение исходных данных, таких как полосность и направление всех элементов УДС, объемов корреспонденций, вида функций задержек $G_p(x)$ и $\tau_a(x)$, а также увеличение наблюдений за автотрафиком, позволит повысить точность прогнозов с использованием модели конкурентного равновесия.

Численные расчеты транспортных потоков в перспективной УДС на основе теории равновесия показали, что первый этап модернизации сети, запланированный в генеральном плане развития города, снизит нагрузку на большинстве участков

дорог и сократит общие транспортные задержки на проезд.

Работа выполнена при частичной поддержке гранта РФФИ 09-01-00042 и проекта ДВО РАН 09-III-A-01-004.

Литература

1. Beckmann M., McGuire C.B., Winsten C.B. Studies in the economics of transportation. — New Haven: Yale University Press, 1956.
2. Wardrop J. Some Theoretical Aspects of Road Traffic Research // Proceedings of the institute of Engineers. — 1952.
3. Boyce D. A retrospective on Beckmann, McGuire and Winsten «Studies in the Economics of Transportation» // Papers in Regional Science. — 2005. — V. 84. — P. 85–103.
4. Nesterov Yu., De Palma A. Stationary dynamic solutions in congested transportation networks: summary and perspectives // Networks and Spatial Economics. — 2003. — V. 3. — P. 371–395.
5. Autonews om 24.09.2008. <http://www.autonews.ru/autobusiness>
6. Nagurney A. Network Economics: A Variational Inequality Approach (second revised edition). — Dordrecht: Kluwer Academic Publishers, 1999.
7. Dafermos S. Traffic Equilibrium and Variational Inequalities // Transportation Science. — 1980. — V. 14. — P. 42–54.
8. Попов Л.Д. Введение в теорию, методы и экономические приложения задач о дополнительнойности. — Екатеринбург: Изд-во Урал. ун-та, 2001.
9. Стенбринк П. Оптимизация транспортных сетей. — М.: Транспорт, 1981.
10. Васильева Е.В., Игудин Р.В., Лившиц В.Н. Оптимизация планирования и управления транспортными системами. — М.: Транспорт, 1987.
11. Вильсон А.Д. Энтропийные методы моделирования сложных систем. — М.: Наука, 1978.
12. Arrowsmith G.A. A behavioural approach to obtaining a doubly constrained trip distribution model // Operational Research Quarterly. — 1973. — V. 24, N. 1. — P. 101–111.
13. Frank M., Wolfe P. An algorithm for quadratic programming // Naval Research Logistics Quarterly. — 1956. — V. 3. — P. 95–110.
14. Швецов В.И. Алгоритмы распределения транспортных потоков // Автоматика и телемеханика. — 2009. — № 10. — С. 148–157.
15. 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.
16. Janson B., Zozaya-Gorostiza C. The problem of cyclic flows in traffic assignment // Transportation Research Part B. — 1987. — V. 21. — P. 299–310.

17. *Patriksson M.* The traffic assignment problem — models and methods // Utrecht. — Netherlands: VSP, 1994.

18. *Bar-Gera H.* Origin-based algorithm for the traffic assignment problem // Transportation Science. — 2002. — V. 36, N 4. — P. 398–417.

19. *Facchinei F., Pang J.-S.* Finite-Dimensional Variational Inequalities and Complementarity Problems — Springer, 2003.

20. *Konnov I.V.* Combined relaxation methods for variational inequalities. — Berlin: Springer, 2001.

21. *Нурминский Е.А.* Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов // Журн. вычисл. матем. и матем. физики. — 2008. — Т. 48, вып. 12. — С. 2121–2128.

22. *Нурминский Е.А.* Фейеровские процессы с малыми возмущениями // Доклады АН. — 2008. — Т. 422, вып. 5. — С. 601–605.

23. *Nurminski E.A.* Envelope stepsize control for iterative algorithms based on Fejer processes with attractants // Optimization Methods and Software. — 2010. — V. 25, N. 1. — P. 97–108.

24. *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.

25. *GNU Octave.*
<http://www.gnu.org/software/octave/>

26. *GNU Linear Programming Kit.*
<http://www.gnu.org/software/glpk/>

27. *Генеральный план Владивостокского городского округа* <http://vlc.ru/architect/genplan.htm>

Поступила в редакцию 15.10.2010.

А.М. Райгородский

Московский государственный университет им. М.В. Ломоносова
Московский физико-технический институт (государственный университет)
ООО «Яндекс»

Модели случайных графов и их применения

В статье дается обзор основных современных направлений в теории случайных графов. Делается акцент на связь моделей случайного графа с транспортной проблематикой.

Ключевые слова: случайный граф, Интернет, транспортные сети.

I. 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 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, так как и здесь наука разрослась безгранично), мы постараемся выделить лишь самые основные и принципиальные моменты теории случайных графов.

II. Модель Эрдеша–Реньи

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. Реньи (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. Реньи также сыграл значительную роль в формировании венгерской вероятностной школы, и его именем назван математический институт в Будапеште.

II.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, стало быть, 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–Реньи.

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

$$\begin{aligned} \Omega_n &= \{G = (V_n, E)\}, \mathcal{F}_n = 2^{\Omega_n}, P_{n,p}(G) = \\ &= p^{|E|} (1-p)^{C_n^2 - |E|}. \end{aligned} \quad (1)$$

Здесь через $|A|$ обозначена мощность множества A (число элементов в нем), 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 связанных 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 $n \rightarrow \infty$.

II.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 (то есть 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, сопоставим 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–Реньи $G(n,p)$ вероятность связности графа больше половины (граф, скорее, связан, чем несвязен).

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 параграфе II.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*, чересчур сильное, и в следующем параграфе мы приведем модификацию модели Эрдеша–Реньи, в рамках которой это предположение можно будет адекватно ослабить.

II.3. Обобщения модели Эрдеша–Реньи

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

$$\begin{aligned} \Omega_n &= \{G = (V_n, E)\}, \mathcal{F}_n = 2^{\Omega_n}, P_{n, p_{ij}}(G) = \\ &= \prod_{(i,j) \in E} p_{ij} \cdot \prod_{(i,j) \notin E} (1 - p_{ij}). \end{aligned}$$

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 случайном 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 случайном 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 $G(H_n, p)$ и 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.

II.4. Некоторые математические результаты о надежности сети

Прежде всего справедлива следующая теорема Эрдеша–Реньи.

Теорема 1. *Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c \ln n}{n}$. Если $c > 1$, то почти всегда случайный граф связан. Если $c < 1$, то почти всегда случайный граф не является связным.*

Этот довольно простой с точки зрения доказательства факт мы обоснуем в параграфе II.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 $\approx 0,993$, так что в результате с вероятностью, близкой к единице, перемещение между любыми двумя городами останется возможным. Поначалу это кажется противоречащим интуиции, но, при здравом размышлении, становится понятно, в чем здесь смысл. Дорог у нас $C_n^2 = \Theta(n^2)$, вероятность износа дороги равна $1 - \Theta(\ln n/n)$ (мы пишем $f = \Theta(g)$ для функций f и g , если существуют константы $c_1, c_2 > 0$, с которыми выполнено $c_1 g \leq f \leq c_2 g$). 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 параграфе II.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 $n > 100$*

$$P_{n,p}(G) \geq 1 - \frac{1}{n}.$$

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 Реньи (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 $\sim \gamma n$. А В.Е. Степанов доказал, что, опять же 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, однако же, если не вся страна, то значительная ее часть окажется консолидированной, то есть не лишенной инфраструктуры — возможности сообщения между любыми двумя городами.

Глубокий интерес представляет, конечно, устройство мира «внутри» фазовых переходов, то есть 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

$$P_{n,p}(G) \rightarrow e^{-e^{-c}}, n \rightarrow \infty.$$

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 $G(H_n,p)$. Иными словами, функция p такова, что для любой функции $p_1 \leq c_1 p$, где $c_1 < 1$, случайный граф в модели $G(H_n,p_1)$ почти всегда не связан, но для любой функции $p_2 \geq c_2 p$, где $c_2 > 1$, случайный граф в модели $G(H_n,p_2)$ почти всегда связан. Под реберной связностью графа понимается минимальное количество его ребер, удаление которых влечет потерю связности.

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 железными дорогами: они крайне редки и служат единственными лазейками с одной стороны полотна на другую.

В следующем параграфе мы докажем теорему 1, а в параграфе II.6 мы обсудим основные идеи доказательства теоремы 2. Отметим, что дополнительную информацию о поведении случайных графов в модели Эрдеша–Реньи можно почерпнуть из книг [7–10].

II.5. Доказательство теоремы 1

Сперва обсудим случай $c > 1$.

Введем случайную величину на пространстве $G(n,p)$:

$$X_n = X_n(G) = \begin{cases} 0, & \text{если } G \text{ связан,} \\ k, & \text{если у } G \text{ ровно } k \text{ компонент.} \end{cases}$$

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 $P_{n,p}(X_n \geq 1) \leq MX_n$, 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

$$X_{n,k} = X_{n,k,1} + \dots + X_{n,k,C_n^k},$$

когда 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

$$MX_n = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} MX_{n,k,i}.$$

Очевидно,

$$MX_{n,k,i} = P_{n,p}(K_i \text{ образует компоненту в } G) \leq P_{n,p}(\text{из } K_i \text{ в } V_n \setminus K_i \text{ нет ребер в } G).$$

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,

$$MX_n \leq \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} (1-p)^{k(n-k)} = \sum_{k=1}^{n-1} C_n^k (1-p)^{k(n-k)}.$$

Последняя сумма симметрична в том смысле, что ее слагаемые при k и $n-k$ равны. Рассмотрим $k=1$:

$$\begin{aligned} n(1-p)^{n-1} &\leq ne^{-p(n-1)} = ne^{-\frac{c(\ln n)(n-1)}{n}} = \\ &= n \left(\frac{1}{n}\right)^{c(1+o(1))} = o(1), \end{aligned}$$

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 случайном 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

$$MX_n = MX_{n,1} + \dots + MX_{n,n}.$$

В свою очередь

$$MX_{n,k} = P_{n,p}(kG) = (1-p)^{n-1}.$$

Таким образом,

$$\begin{aligned} MX_n &= n(1-p)^{n-1} = n(1-p)^n(1+o(1)) = \\ &= (1+o(1))ne^{-c \ln n} = (1+o(1))n^{1-c}. \end{aligned}$$

Заметим, что ввиду неравенства $c < 1$ выполнено $MX_n \rightarrow \infty$.

Посчитаем дисперсию случайной величины X_n :

$$\begin{aligned} DX_n &= MX_n^2 - (MX_n)^2 = M(X_{n,1} + \dots + X_{n,n})^2 - \\ &\quad - (MX_n)^2 = MX_{n,1}^2 + \dots + MX_{n,n}^2 + \\ &\quad + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2 = MX_{n,1} + \dots \\ &\quad \dots + MX_{n,n} + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2 = \\ &= MX_n + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2. \end{aligned}$$

Далее,

$$\begin{aligned} MX_{n,i}X_{n,j} &= P_{n,p}(i \text{ и } j \text{ изолированы в } G) = (1-p)^{2n-1} = \\ &= (1+o(1))(1-p)^{2n}, \text{ то есть} \end{aligned}$$

$$\begin{aligned} \sum_{i \neq j} MX_{n,i}X_{n,j} &= n(n-1)(1+o(1))(1-p)^{2n} = \\ &= (1+o(1))n^{2-2c} = (1+o(1))(MX_n)^2. \end{aligned}$$

В итоге

$$\begin{aligned} DX_n &= MX_n + (1+o(1))(MX_n)^2 - (MX_n)^2 = \\ &= o((MX_n)^2). \end{aligned}$$

По неравенству Чебышёва

$$\begin{aligned} P_{n,p}(G) &\leq P_{n,p}(X_n = 0) = P_{n,p}(X_n \leq 0) = \\ &= P_{n,p}(-X_n \geq 0) = P_{n,p}(MX_n - X_n \geq MX_n) \leq \\ &\leq \frac{DX_n}{(MX_n)^2} = o(1), \end{aligned}$$

и вторая часть теоремы доказана.

II.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.

Простейший ветвящийся процесс

Пусть Z_1, \dots, Z_t, \dots — независимые пуассоновские величины с одним и тем же средним λ . Положим

$$Y_0 = 1, Y_t = Y_{t-1} + Z_t - 1.$$

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, так как 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, то есть $P(\exists t : Y_t \leq 0) = 1$.

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 $1 - \gamma$, то есть $P(\exists t : Y_t \leq 0) = 1 - \gamma$.

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.

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 в аккурат 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,

$$Y_0 = 1, Y_t = Y_{t-1} + Z_t - 1.$$

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. Дело в том, что она имеет биномиальное распределение с «числом испытаний» N_t и вероятностью «успеха» p . Правда, число испытаний само случайно. По счастью, это не проблема. Удастся доказать, что Y_t имеет вид

$$Y_t = \xi_t + 1 - t, \xi_t \sim \text{Binom}(n - 1, 1 - (1 - p)^t).$$

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.

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:

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \rightarrow 0, n \rightarrow \infty.$$

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 неаккуратными, но при желании их можно сделать безукоризненно строгими. Мы же хотим максимально прояснить основную суть подхода. Итак,

$$\begin{aligned} P_{n,p}(Y_{t_0} > 0) &= P_{n,p}(\xi_{t_0} \geq t_0) \approx \\ &\approx P_{n,p}(\text{Binom}(n, 1 - (1 - p)^{t_0}) \geq t_0) \approx \end{aligned}$$

(с учетом асимптотики $1 - (1 - p)^{t_0} \sim pt_0$)

$$\approx P_{n,p}(\text{Binom}(n, pt_0) \geq t_0) \approx$$

(с учетом центральной предельной теоремы)

$$\approx \int_{\frac{t_0 - npt_0}{\sqrt{npt_0(1-pt_0)}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Поскольку $c < 1$, нижний предел интегрирования имеет порядок $\sqrt{t_0}$. Стало быть, 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.

Случай $c > 1$

Этот случай гораздо сложнее предыдущего. Здесь ветвящийся процесс на графе нужно «запускать» не один раз, а многократно. Только так удастся 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

$$P_{n,p}(Y_t \leq 0) \rightarrow 0, t \approx \gamma n, n \rightarrow \infty.$$

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 .

III. 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 первом параграфе мы расскажем о результатах Барабаши–Альберт. Во втором параграфе мы опишем модель Б. Боллобаша и О. Риордана, которая весьма неплохо ложится на статистики Барабаши–Альберт. В третьем параграфе мы обсудим возможные уточнения модели Боллобаша–Риордана.

III.1. Наблюдения Барабаши–Альберт

В своих работах [14–16] Барабаши и Альберт, а также Х. Джеонг описали те статистики Интернета, которые легли в основу науки о росте этой сети — науки, имеющей глубокие приложения как в собственно проблематике Интернета, 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 дру-

гую его страницу). Это такой «псевдомультиторграф». Сразу понятно, что для подобного «зверя» модель Эрдеша–Реньи 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 (см. [16]) 5–7. 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: «кликакая» по ссылкам, можно с любого сайта на любой другой перейти за 5–7 нажатий клавиши компьютерной мыши. Конечно, тут есть важная оговорка. Некоторые едва появившиеся сайты 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–Реньи 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 мало-мальски гарантируют разреженность (то есть при $p = \Theta(1/n)$), указанное биномиальное распределение аппроксимируется пуассоновским, а вовсе не степенным.

Сами 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 идее, называются моделями *предпочтительного присоединения*. В своих работах Барабаш 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 удастся доказать даже такой забавный факт (см. [17]).

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 Барабаш–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 Барабаш–Альберт предложили в начале двухтысячных годов Б. Боллобаш и О. Риордан. В следующем параграфе мы ее обсудим.

III.2. Модель Боллобаш–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.

Динамическая модификация

Сперва 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 $G_1^1 = (\{1\}, \{(1,1)\})$, то есть 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 $\frac{\text{deg} i}{2n-1}$, где $\text{deg} i$ — степень вершины i в графе G_1^{n-1} . Очевидно, что распределение вероятностей задано корректно, поскольку

$$\sum_{i=1}^{n-1} \frac{\text{deg} i}{2n-1} + \frac{1}{2n-1} = \frac{2n-2}{2n-1} + \frac{1}{2n-1} = 1.$$

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 :

$$\{1, \dots, k\}, \{k+1, \dots, 2k\}, \dots, \{k(n-1)+1, \dots, kn\}.$$

Объявляем каждый кусок «вершиной», а ребра сохраняем, то есть 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.

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, короче, LCD по-английски). 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 первые 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 порождает 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, то есть 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.

Некоторые результаты

Замечательна модель Боллобаша–Риордана не только тем, что с ее помощью наводится порядок в «каше», которую «заварили» Барабаша и Альберт, но еще и тем, что она полностью адекватна эмпирическим наблюдениям. Прежде всего справедлива

Теорема 8. Для любого $k \geq 2$ и любого $\varepsilon > 0$

$$P\left((1-\varepsilon)\frac{\ln n}{\ln \ln n} \leq \text{diam} G_k^n \leq (1+\varepsilon)\frac{\ln n}{\ln \ln n}\right) \rightarrow 1, n \rightarrow \infty.$$

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 $d \leq n^{1/15}$

$$M\left(\frac{|\{i = 1, \dots, n : \deg_{G_k^n} i = d\}|}{n}\right) \sim \frac{2k(k+1)}{(d+k+1)(d+k+2)(d+k+3)}. \quad (2)$$

Поскольку k — константа, выражение в правой части (2) имеет вид const/d^3 . Да это же в точности степенной закон! Правда, в формулировке теоремы написано математическое ожидание, а не вероятность, но одно из другого получается за счет маргингальных 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 параграфе III.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 симпатичных следствий, которые мы выпишем и здесь.

Теорема 10. Пусть $k \geq 2$. Пусть также K_3 — полный граф на трех вершинах. Тогда

$$M(\sharp(K_3, G_k^n)) = (1+o(1)) \cdot \frac{k \cdot (k-1) \cdot (k+1)}{48} \cdot (\ln n)^3$$

при $n \rightarrow \infty$.

Теорема 11. Пусть фиксированы $k \geq 2$ и $l \geq 3$. Пусть также C_l — цикл на l вершинах. Тогда

$$M(\sharp(C_l, G_k^n)) = (1+o(1)) \cdot c_{k,l} \cdot (\ln n)^l$$

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 $\sharp(d_i = t)$ число вершин в H , степень каждой из которых равна t . Тогда

$$M(\sharp(H, G_k^n)) = \Theta(n^{\sharp(d_i=0)} \cdot (\sqrt{n})^{\sharp(d_i=1)} \cdot (\ln n)^{\sharp(d_i=2)}).$$

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

$$\sharp(d_i = 0) = \sharp(d_i = 1) = 0, \quad \sharp(d_i = 2) = 3.$$

По теореме 12

$$M(\sharp(K_3, G_k^n)) = \Theta((\ln n)^3),$$

и это прекрасно согласуется 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.

III.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 Барабаша–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 III.2. Однако здесь процесс будет устроен сильно по-другому.

В качестве начального графа возьмем любой d -регулярный граф (граф, у которого степень каждой вершины равна d). Пусть построен граф с номером t . Обозначим его $G_t = (V_t, E_t)$. Здесь $V_t = \{u_1, \dots, u_s\}$, где s отличается от t на число вершин начального графа, то есть 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 вышла 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, так как 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 полно. Это особенно важно ввиду того, что спамерские структуры, о которых мы вскользь говорили в конце пункта III.2, зачастую образуют именно двудольные графы с плотной перелинковкой.

Работа 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.

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, вып. 1. — С. 55–67.
5. Степанов В. Е. Фазовый переход в случайных графах // Теория вероятностей и ее применения. — 1970. — Т. 15, вып. 2. — С. 187–203.
6. Степанов В. Е. Структура случайных графов $g_n(x|h)$ // Теория вероятностей и ее применения. — 1972. — Т. 17, вып. 3. — С. 227–242.
7. Колчин В. Ф. Случайные графы. — М.: Физматлит, 2004. — 256 с.
8. Bollobás B. Random Graphs. — Cambridge Univ. Press, 2001. — 520 с.
9. Алон Н., Спенсер Дж. Вероятностный метод. — М.: Бином. Лаборатория знаний, 2007. — 320 с.
10. Janson S., Luczak T., Rucinski A. Random graphs. — New York: Wiley, 2000. — 333 с.
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. — 536 с.

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*. — 2000. — V. A281. — 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: Регулярная и хаотическая динамика, 2009. — 120 с.

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.

Поступила в редакцию 15.10.2010.

Н.Н. Смирнов, А.Б. Киселев, В.Ф. Никитин, А.В. Кокорева

Московский государственный университет им. М.В. Ломоносова

Математическое моделирование движения автотранспортных потоков методами механики сплошной среды. Двухполосный транспортный поток: модель Т-образного перекрестка, исследование влияния перестроений транспортных средств на пропускную способность участка магистрали

Представлено дальнейшее развитие модели движения автотранспортных потоков методами механики сплошной среды. Построена новая модель двухполосного транспортного потока, учитывающая перестроения транспортных средств и основанная на подходе механики многофазных сред. В качестве примера использования модели двухполосного движения проведено численное моделирование Т-образного перекрестка, образованного примыканием второстепенной двухполосной дороги к основной также двухполосной трассе. Движение на этом перекрестке регулируется светофором с тремя фазами работы.

Ключевые слова: математическое моделирование движения автотранспортных потоков, светофор, Т-образный перекресток, перестроение транспортных средств, пропускная способность, механика сплошной среды, многофазные потоки.

I. Введение

Первые математические модели движения автотранспортных потоков с точки зрения механики сплошной среды опубликованы Лайтхиллом и Уиземом [1], а также Ричардсом [2]. Анализ основных полученных ими результатов дается в известной монографии Уизема [3]. С развитием вычислительной техники для изучения движения автотранспортных потоков стало широко применяться численное моделирование. С современным состоянием вопроса можно ознакомиться, например, прочитав учебное пособие по математическому моделированию транспортных потоков [4].

Изучение динамики автотранспортных потоков началось в нашей стране в конце 70-х годов XX века на механико-математическом факультете МГУ им. М.В. Ломоносова в связи с подготовкой к Олимпийским играм 1980 г. в Москве. Результаты этих исследований неоднократно докладывались В.Н. Беловым на научно-исследовательском семинаре И.Н. Зверева. В настоящее время изучение автотранспортных потоков на факультете ведется под руководством Н.Н. Смирнова и А.Б. Киселева на кафедре газовой и волновой динамики и в лабораториях волновых процессов и динамики деформируемых сред. Эти исследования неоднократно поддерживались грантами Правительства г. Москвы и Правительства региона г. Брюсселя (Бельгия).

Существуют три традиционных «механических» подхода к моделированию движения транспортного потока. Микроскопические модели описывают воздействие предыдущего автомобиля на следующий за ним при помощи обыкновенных дифференциальных уравнений, основанных на ньютоновской механике [5, 6]. Макроскопические

модели основываются на уравнениях газовой динамики [1, 7, 8]. Мезоскопические модели представляют собой промежуточное звено между двумя предыдущими моделями. Они основаны на кинетических уравнениях бальмановского типа [9]. В рамках макроскопического, или континуального, подхода для описания движения потока транспорта используется уравнение неразрывности. В качестве второго уравнения ряд исследователей использовали эмпирическое соотношение, связывающее плотность и расход [2, 10]. Такая система позволяет описывать движение локально-равновесного потока. Для описания неравновесного потока было предложено уравнение движения, учитывающее стремление водителя привести свою скорость в соответствие с некой равновесной скоростью [11]. Однако такой подход не позволяет адекватно описывать возникновение ударных волн плотности потока автомашин. Существующие газодинамические модели не учитывают влияние движения впереди идущих транспортных средств на движение автомобилей, следующих позади, и стремление водителя привести свою скорость в соответствие с максимально безопасной скоростью. На решение этих проблем и направлена предлагаемая модель. Ранее проведенные исследования, использующие ее, опубликованы в работах [12–19].

II. Модель движения транспортного потока по автомагистрали

Рассмотрим однонаправленный поток машин по однополосной дороге. Пересечение с другими дорогами и наличие светофоров будет учитываться соответствующими граничными условиями. Введем Эйлерову систему координат x вдоль автомагистрали в направлении движения потока

и время t . Среднюю плотность потока $\rho(x,t)$ определим как отношение площади полосы движения, занятой транспортными средствами, к площади всего рассматриваемого участка полосы движения:

$$\rho = \frac{S_{tr}}{S} = \frac{hn\ell}{hL} = \frac{n\ell}{L}, \quad (1)$$

где h — ширина полосы движения, L — длина контрольного участка дороги, ℓ — средняя длина транспортного средства с минимальным расстоянием между стоящими автомобилями, n — количество транспортных средств на контрольном участке. Так введенная плотность является безразмерной величиной и изменяется в интервале $0 \leq \rho \leq 1$.

Введем скорость потока $v(x,t)$, которая может изменяться в пределах $0 \leq v \leq v_{\max}^0$, где v_{\max}^0 — максимально разрешенная скорость движения на магистрали вне систем регулирования дорожного движения. Из определений следует, что максимальная плотность $\rho = 1$ соответствует ситуации, когда машины располагаются практически вплотную («бампер в бампер»). В этом случае естественно принять $v = 0$, то есть на дороге образовалась «неподвижная пробка».

Условно называя «массой», сосредоточенной на участке длины L , величину

$$m = \int_0^L \rho dx,$$

можно записать закон изменения «массы» на автомагистрали. Для непрерывного потока машин будет иметь место уравнение неразрывности:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0. \quad (2)$$

Запишем уравнение динамики транспортного потока, точнее, уравнение изменения режима движения. Режим движения транспортных средств на дороге определяется следующими основными факторами: реакцией водителя на изменение дорожной обстановки и предпринимаемыми им активными действиями, откликом транспортного средства на действия водителя и техническими характеристиками транспортных средств. При разработке модели динамики транспортных средств были сделаны следующие основные предположения:

- Модель призвана описывать усредненное движение множества транспортных средств, а не движение каждого автомобиля в отдельности. Вследствие этого модель оперирует с усредненными характеристиками транспортных средств, не учитывая индивидуальных различий мощности, инерции, тормозного пути и т.п.

- Модель предполагает в среднем адекватную реакцию всех водителей на изменение дорожной обстановки, а именно: предполагается, что, видя красный сигнал светофора или знак ограничения скорости, например, перед «лежащим полицейским», или скопление машин перед ним, водитель замедляет движение до последующей полной остановки или до допустимой скорости, а не продолжает ускоряться, чтобы впоследствии применить режим экстренного торможения.
- Предполагается, что подавляющее большинство водителей соблюдают правила дорожного движения, в частности, не превышают максимально допустимого скоростного режима, разрешенного на дороге, и выдерживают безопасный интервал между транспортными средствами в зависимости от скорости движения. Кроме того, движения транспорта задним ходом нет.

Тогда уравнение изменения скорости запишется в следующем виде:

$$\frac{dv}{dt} = a; \quad a = \max\{-a^-; \min\{a^+; a'\}\};$$

$$a' = \sigma a_\rho + \frac{(1-\sigma)}{\Delta} \int_x^{x+\Delta} a_\rho(t,y) dy + \frac{V(\rho) - v}{\tau}; \quad (3)$$

$$a_\rho = -\frac{k^2}{\rho} \frac{\partial \rho}{\partial x}.$$

Здесь a — ускорение транспортного потока; a^+ — максимально возможное ускорение разгона; a^- — ускорение экстренного торможения; величины a^+ и a^- положительны и определяются техническими характеристиками транспортного средства. Δ — «расстояние принятия решения», то есть длина участка магистрали перед участником движения, где находятся транспортные средства, на характер движения которых реагирует рассматриваемое транспортное средство, σ — безразмерный параметр ($0 \leq \sigma \leq 1$), характеризующий «вес» локальной ситуации по сравнению с ситуацией на некотором расстоянии впереди автомобиля. Параметр $k > 0$ является, как было показано в работах [8, 12, 13], скоростью распространения возмущений («скоростью звука») в транспортном потоке. Параметр τ имеет смысл времени задержки, обусловленной конечностью скорости реакции водителя на изменение дорожной обстановки и техническими характеристиками транспортного средства. Этот параметр отвечает за стремление водителя привести скорость автомобиля в соответствие с максимально безопасной скоростью движения $V(\rho)$ для плотности потока ρ [8, 12, 13]. В зависимости от того, требуется ли для достижения максимальной безопасной скорости $V(\rho)$ притормаживать или разгоняться, значение параметра τ

может быть различным:

$$\tau = \begin{cases} \tau^+, V(\rho) < v, \\ \tau^-, V(\rho) \geq v. \end{cases}$$

В выражении (3) для ускорения транспортного потока a' первое слагаемое отвечает за влияние на поведение водителя локальной ситуации, второе — за влияние ситуации вперед по потоку, а третье — за стремление водителя привести свою скорость в соответствие с максимально безопасной скоростью. Если принять $\sigma = 0$, то выражение для ускорения (3) примет следующий вид:

$$a' = \frac{1}{\Delta} \int_x^{x+\Delta} a_\rho(t,y) dy + \frac{V(\rho) - v}{\tau}.$$

То есть ускорение зависит только от величины ускорения на участке длины Δ , где Δ — расстояние принятия решения, и от стремления водителя привести свою скорость в соответствие с максимально безопасной скоростью. При $\sigma = 1$ выражение (3) принимает следующий вид:

$$a' = a_\rho + \frac{V(\rho) - v}{\tau}.$$

Теперь ускорение зависит от локальной ситуации, а также от третьего слагаемого. Конкретные значения описанных выше параметров, использованных при математическом моделировании, будут приведены ниже.

Таким образом, для описания динамики автотранспортного потока по однополосной дороге из (1) и (2) получается система двух квазилинейных уравнений в частных производных в дивергентной форме:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} &= 0; \\ \frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho v^2)}{\partial x} &= \rho a. \end{aligned} \quad (4)$$

где ускорение a определяется тремя последними формулами (3).

II.1. Исследование основных характеристик системы уравнений, предлагаемой для описания транспортного потока

Модель с постоянной скоростью распространения возмущений. Рассмотрим случай, когда параметр k в формулах (3), (4) постоянен и не зависит от скорости и плотности потока. Данная модель обладает высокой степенью точности, когда скорость потока превосходит скорость распространения возмущений. Рассмотрим случай, когда $\sigma = 1$, а $\tau = \infty$. Тогда система (4) примет вид

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial(\rho V)}{\partial x} &= 0; \\ \frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho v^2)}{\partial x} &= -k^2 \frac{\partial \rho}{\partial x}. \end{aligned} \quad (5)$$

Полученная система (5) двух квазилинейных уравнений с неизвестными ρ, v является гиперболической. Условия вдоль характеристик C^+ и C^- в плоскости (x, t) следующие:

$$\begin{aligned} C^+ : \frac{dx}{dt} &= v + k, \rho dv = -k d\rho; \\ C^- : \frac{dx}{dt} &= v - k, \rho dv = k d\rho. \end{aligned} \quad (6)$$

Полученные выше характеристики несут информацию о дорожной ситуации соответственно по потоку и против него.

Транспортные потоки и существующие системы организации дорожного движения специфичны тем, что распространение информации происходит однонаправлено — навстречу потоку. Тогда для волн, распространяющихся влево, навстречу потоку с постоянными параметрами, имеет место интеграл Римана, полученный из соотношения на характеристике C^- (6):

$$v - v_0 = k \frac{\rho_0}{\rho}. \quad (7)$$

Скорость $v(\rho)$ определяется из условия зависимости скорости автомобиля v от плотности потока ρ для условий простой волны, возникающей при начале движения потока с места при условиях $\rho_0 = 1$ и $v = 0$, с учетом ограничения на максимально допустимую скорость движения ($v \leq v_{\max}^0$) получим выражение для максимально безопасной скорости движения:

$$V(\rho) = \{-k \ln \rho, v < v_{\max}^0; v_{\max}^0, v \geq v_{\max}^0\}. \quad (8)$$

Оценка величины скорости распространения возмущений k сделана в работе [13], исходя из следующих соображений. Пусть, начиная движение из состояния покоя ($v = 0, \rho = 1$) и ускоряясь до скорости v_{\max}^0 , поток достигает при этом максимально допустимой плотности ρ_* , гарантирующей безопасность движения. Под безопасной плотностью понимается такая плотность, при которой расстояние между машинами не меньше тормозного пути $X(v)$. Тогда

$$\begin{aligned} \rho_* &= (1 + X(v_{\max}^0)/\ell)^{-1}, \\ k &= v_{\max}^0 \ln^{-1}(1 + X(v_{\max}^0)/\ell). \end{aligned}$$

При $v_{\max}^0 = 80$ км/ч тормозной путь автомобиля типа ВАЗ составляет 45 м, что при средней длине автомобиля (с учетом минимального расстояния между стоящими автомобилями) $\ell = 5$ м дает скорость распространения слабых возмущений $k = 35$ км/ч. Для такой величины v_{\max}^0 максимально возможная безопасная плотность потока равна $\rho_* = 0,1$. Величины максимальных ускорений для автомобилей данного класса составляют $a^+ = 1,63$ м/с² и $a^- = 5,5$ м/с².

Оцененная таким образом «скорость звука» k хорошо согласуется с экспериментальными данными, приведенными в работе [20].

Переменная скорость распространения возмущений. Рассмотрим случай, когда параметр k в формулах (3), (4) зависит от скорости

v и плотности потока ρ . Предположим, что транспортному средству требуется время τ_p для реакции на изменение дорожной обстановки впереди по потоку: $k = \ell'/\tau_p$, где τ_p — время реакции водителя, ℓ' — расстояние между транспортными средствами (рис. 1). По определению плотности $\ell' = \ell/\rho$, следовательно, $k = k_1/\rho$, где $k_1 = \ell/\tau_p$.

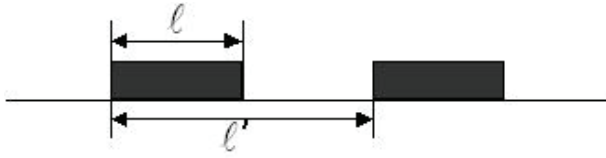


Рис. 1

Анализ наблюдений за движением потока транспорта показал [20, 3], что при небольшой плотности потока возмущения распространяются с постоянной скоростью. Тогда скорость распространения возмущений можно вычислять по формуле

$$k(\rho) = \begin{cases} k_1/\rho & \text{при } \rho \geq \rho_k = k_1/k_0, \\ k_0 & \text{при } \rho < \rho_k. \end{cases} \quad (9)$$

График этой зависимости представлен на рис. 2.

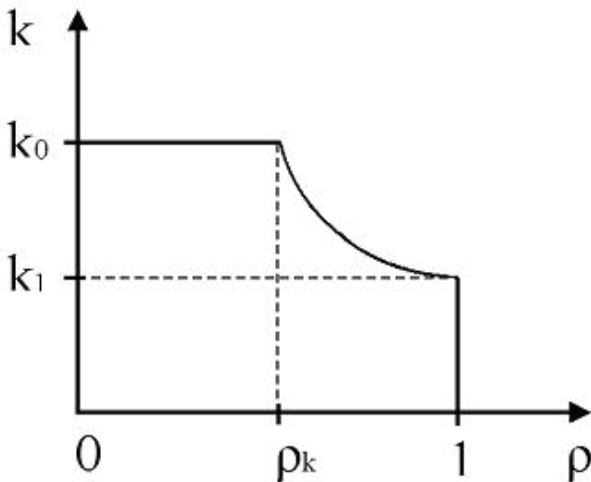


Рис. 2. Зависимость скорости распространения возмущений от плотности потока

Если в (3) принять $\sigma = 1, \tau = \infty$, то при $\rho \geq \rho_k$ получим систему двух квазилинейных уравнений в частных производных:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + v \frac{\partial \rho}{\partial x} + \rho \frac{\partial v}{\partial x} &= 0, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} + \frac{k_1^2}{\rho^3} \frac{\partial \rho}{\partial x} &= 0. \end{aligned} \quad (10)$$

Ее характеристики в плоскости $(x; t)$ — C^+ , C^- и условия вдоль них следующие:

$$\begin{aligned} C^+ : \frac{dx}{dt} &= v + \frac{k_1}{\rho}, \rho dv = -\frac{k_1}{\rho} d\rho; \\ C^- : \frac{dx}{dt} &= v - \frac{k_1}{\rho}, \rho dv = \frac{k_1}{\rho} d\rho. \end{aligned} \quad (11)$$

Для волн, распространяющихся влево, навстречу потоку с постоянными параметрами, имеет место интеграл Римана, полученный из соотношения на характеристике C^- (11):

$$v - v_0 = k_1 \left(\frac{1}{\rho} - \frac{1}{\rho_0} \right).$$

В предложенной модели предполагается, что $v = 0$ при $\rho = 1$. Таким образом, получаем зависимость для скорости потока:

$$v = k_1 \left(\frac{1}{\rho} - 1 \right).$$

В модели принимается, что $v = k_0 - k_1$ при $\rho = \rho_k$. Тогда из (11) получаем следующую зависимость для скорости потока:

$$v = k_0 - k_1 - k_0 \ln \left(\frac{\rho}{\rho_k} \right).$$

Учитывая ограничение на максимальную скорость потока ($v < v_{\max}^0$), можно получить следующую зависимость для максимальной безопасной, для данной плотности ρ , скорости потока:

$$V(\rho) = \min \left\{ v_{\max}^0; k_1 \left(\frac{1}{\rho} - 1 \right) \text{ при } \rho \geq \rho_k, \right. \\ \left. (k_0 - k_1) - k_0 \ln \frac{\rho}{\rho_k} \text{ при } \rho < \rho_k \right\}. \quad (12)$$

График этой зависимости представлен на рис. 3.

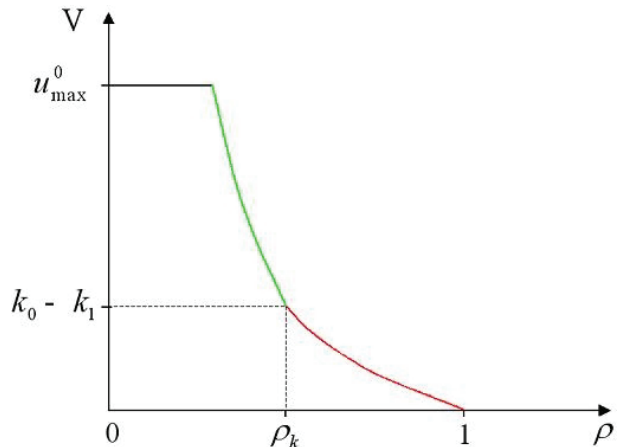


Рис. 3. Максимально безопасная для данной плотности ρ скорость потока

Экспериментальное определение значения скорости распространения возмущений k_1 при начале движения потока. Эксперимент проводился для однополосного потока машин, без учета возможности перестроений в очереди перед светофором. В начальный момент времени, когда включен красный сигнал светофора, автомобили располагаются вплотную, «бампер в бампер». Тогда, согласно представленной модели, плотность потока $\rho = 1$ и $k(1) = \frac{k_1}{1} = k_1$. Величина скорости k_1 определялась экспериментальным следующим образом: осуществлялись прямые наблюдения и замеры времени между тем, как загорался зеленый сигнал, и моментами, когда начинали движение автомобили, удаленные на расстояние $l_1 = 48,5$ м и $l_2 = 73,5$ м от светофора. Фиксировалось соответственно время начала движения в точках l_1 и l_2 , что позволяло определить среднюю скорость распространения сигнала:

$$k_{1,\text{эксп}} = \frac{l_2 - l_1}{t_2 - t_1}.$$

Результаты произведенных замеров сведены в табл. 1. В расчетах использовалось среднее арифметическое значение $k_1 = 4,1 \text{ м/с} \approx 15 \text{ км/ч}$.

Т а б л и ц а 1

Результаты эксперимента по определению скорости распространения возмущений k_1

№	$l_1, \text{ м}$	$l_2, \text{ м}$	$t_1, \text{ с}$	$t_2, \text{ с}$	$k_{1, \text{ эксп.}}, \text{ м/с}$
1	48,5	73,5	14,46	19	5,5066
2	48,5	73,5	10,99	17	4,1597
3	48,5	73,5	12,58	18	4,6125
4	48,5	73,5	9,33	17	3,2595
5	48,5	73,5	12,53	18	4,5704
6	48,5	73,5	14,17	21	3,6603
7	48,5	73,5	11,52	19	3,3422
8	48,5	73,5	12,5	20	3,3333

II.2. Модель перестроений транспортных средств между полосами

Рассмотрим однонаправленный поток машин по двухполосной дороге при подъезде к светофору. Введем эйлерову систему координат x вдоль автомагистрали в направлении движения потока и время t . Средняя плотность потока $\rho(x,t)$ вычисляется по формуле (1).

Движение транспортного потока регулируется светофором. L_1 — координата места расположения светофора (рис. 4). Основные параметры работы светофора — длительность сигналов: зеленого t_1 и красного t_2 . Во время работы зеленого сигнала разрешен проезд по обоим полосам, во время работы красного сигнала проезд по обоим полосам запрещен.

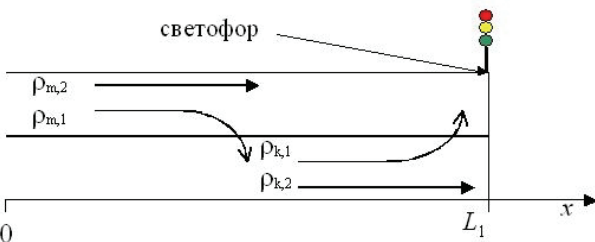


Рис. 4. Схема движения двухполосного транспортного потока с учетом перестроений между соседними полосами

Пусть в начальный момент времени транспортные средства, которые должны проехать светофор в левом и правом рядах, равномерно распределены между полосами движения. Тогда плотность потока по каждой полосе $\rho_k(x) = \rho_{1,k} + \rho_{2,k}$, где x — координата вдоль полосы по направлению движения, $\rho_{1,k}$ — плотность транспортных средств, которые не обязаны сворачивать на соседнюю полосу и $\rho_{2,k}$ — плотность тех транспортных средств, что должны перестроиться в другой ряд до светофора (рис. 4).

Тогда уравнения баланса транспортных средств на двух соседних полосах с индексами

k, m запишутся в следующем виде:

$$\begin{aligned} \frac{\partial \rho_{1,k}}{\partial t} + \frac{\partial}{\partial x}(\rho_{1,k} v_k) &= \omega_{km}(\rho_{2,m}, \rho_k), \\ \frac{\partial \rho_{2,k}}{\partial t} + \frac{\partial}{\partial x}(\rho_{2,k} v_k) &= -\omega_{mk}(\rho_{2,k}, \rho_m); \\ \frac{\partial \rho_{1,m}}{\partial t} + \frac{\partial}{\partial x}(\rho_{1,m} v_m) &= \omega_{mk}(\rho_{2,k}, \rho_m), \\ \frac{\partial \rho_{2,m}}{\partial t} + \frac{\partial}{\partial x}(\rho_{2,m} v_m) &= -\omega_{km}(\rho_{2,m}, \rho_k). \end{aligned}$$

Здесь ω_{mk} — поток транспорта с полосы k на полосу m . Он рассчитывается следующим образом: рассматривается условие перестроения по близости светофора; если

$$v_k \tau_{tr} + \Delta_{tr} > L_1 - x, \tag{13}$$

то необходимо перестраиваться. Рассматривается условие по плотности до светофора (если первое условие не выполнено); если

$$\rho_k(L_1 > x_1 > x) > \rho_{tr}, \tag{14}$$

то производится перестроение. В случае выполнения условия перестроения рассчитывается ω_{mk} :

$$\omega_{mk} = \frac{(\rho_{2,k}(1 - \rho_m))^{\delta_{tr}}}{\tau_{tr}}, \tag{15}$$

τ_{tr} — характерное время перестроения, Δ_{tr} — дополнительная дистанция, ρ_{tr} — критическая плотность на полосе, на которую должно быть выполнено перестроение, параметр δ_{tr} отвечает за интенсивность потока, L_1 — координата светофора.

Уравнения изменения скорости для транспортных средств, движущихся по соседним полосам, запишутся в следующем виде:

$$\begin{aligned} \frac{dv_k}{dt} &= -\frac{k^2}{\rho_{1,k} + \rho_{2,k}} \frac{\partial(\rho_{1,k} + \rho_{2,k})}{\partial x}; \\ \frac{dv_m}{dt} &= -\frac{k^2}{\rho_{1,m} + \rho_{2,m}} \frac{\partial(\rho_{1,m} + \rho_{2,m})}{\partial x}. \end{aligned}$$

Предложенная модель движения двухполосного транспортного потока с учетом межрядовых перестроений при подъезде к светофору учитывает как случай свободного движения потока транспорта, так и ситуацию, когда возникает подвижная пробка.

III. Двухполосный транспортный поток

III.1. Математическая модель двухполосного транспортного потока

Рассмотрим однонаправленный поток машин по двухполосной дороге. Введем эйлерову систему координат x вдоль автомагистрали в направлении движения потока и время t . Средняя плотность потока $\rho(x,t)$ вычисляется по формуле (1). Для описания динамики автотранспортного потока получается система двух квазилинейных уравнений в частных производных в дивергентной форме (4).

Воспользуемся моделью, предложенной в разделе II.1, то есть будем считать, что параметр k зависит от ρ и эта зависимость выражается формулой (8). Тогда система (4) примет вид (10). В этом случае максимально безопасная скорость движения транспортного потока для данной плотности $V(\rho)$ вычисляется по формуле (9).

В качестве начальных условий примем, что на участке длиной x_0 , считая от входа $x = 0$, магистраль занята потоком машин плотности ρ_0 , движущихся с максимально безопасной для данной плотности скоростью $V(\rho_0)$, а при $x_0 < x \leq L_1$ магистраль свободна от машин ($\rho = 0, v = 0$). Плотности транспортных средств на полосах в начальный момент времени вычисляются по следующим формулам:

$$\rho_{1,k}^0 = \left(1 - \frac{p}{100}\right) \rho_0, \quad \rho_{2,k}^0 = \frac{p}{100} \rho_0,$$

$$\rho_{1,m}^0 = \frac{p}{100} \rho_0, \quad \rho_{2,m}^0 = \left(1 - \frac{p}{100}\right) \rho_0,$$

где p — процент транспортных средств, которые должны оказаться на правой полосе k до светофора (рис. 4).

В качестве граничных условий на концах рассматриваемого участка магистрали $0 \leq x \leq L$ выберем следующие: на входе потока при $x = 0$ возможны два варианта граничного условия.

1. В условиях отсутствия «пробки» задается плотность потока и скорость, равная максимально безопасной для данной плотности:

$$\rho(0,t) = \rho_0; \quad v(0,t) = V(\rho_0).$$

1. В условиях подвижной или неподвижной «пробки», примыкающей к входному участку дороги $x = 0$, ставится условие равенства нулю градиента плотности, а скорость равна максимально безопасной для данной плотности:

$$\left. \frac{\partial \rho}{\partial x} \right|_{x=0} = 0; \quad v(0,t) = V(\rho).$$

Наличие или отсутствие «подвижной пробки», примыкающей к левой границе расчетной области $x = 0$, определяется после расчета очередного шага по времени по следующему критерию. Если

$$\left. \frac{\partial \rho}{\partial x} \right|_{x=0} > 0 \quad \text{и} \quad \rho > \rho_0,$$

то имеется «подвижная пробка».

На выходе потока при $x = L$ ставится условие «свободного выхода»:

$$\frac{\partial \rho}{\partial x} = 0; \quad \frac{\partial v}{\partial x} = 0.$$

III.2. Модель Т-образного перекрестка

На основе описанной выше модели двухполосного потока проведено численное моделирование Т-образного перекрестка, образованного примыканием второстепенной двухполосной дороги к основной также двухполосной трассе. Движение на этом перекрестке регулируется светофором с тремя фазами работы. В ходе соответствующей фазы движение на перекрестке осуществляется так, как показано на рис. 5, 6, 7. На рис. 5 видно, что с второстепенной дороги правый поворот осуществим только из правого ряда, а левый — только из левого. На рис. 6 показано, что при движении по основной дороге слева направо правый ряд предназначен для поворота направо, а левый — для движения прямо. При движении справа налево левый поворот осуществляется из левого ряда, а движение прямо — из правого ряда.

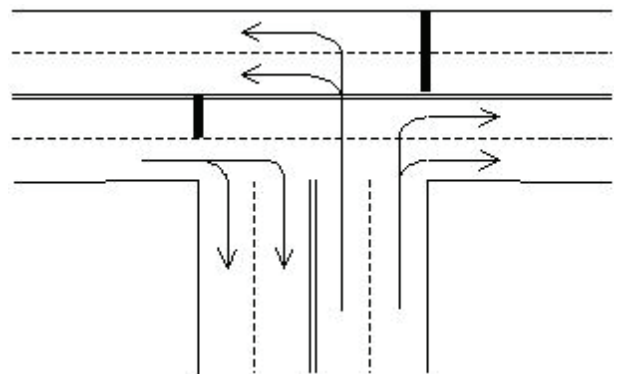


Рис. 5. Схема движения автотранспорта по Т-образному перекрестку в ходе первой фазы работы светофора

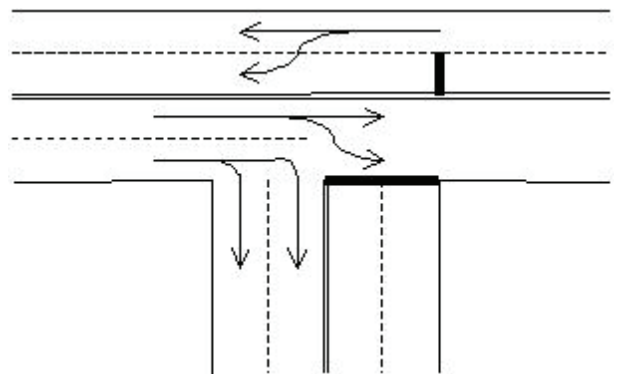


Рис. 6. Схема движения автотранспорта по Т-образному перекрестку в ходе второй фазы работы светофора

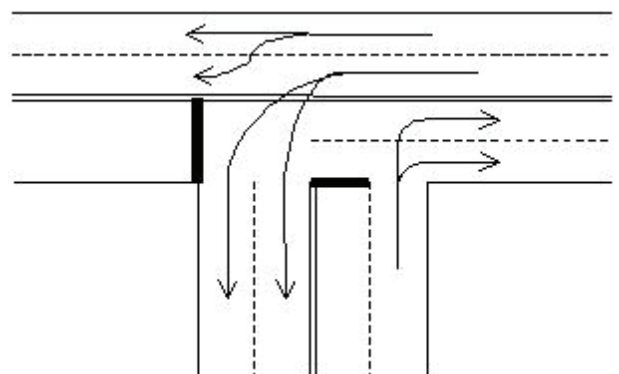


Рис. 7. Схема движения автотранспорта по Т-образному перекрестку в ходе третьей фазы работы светофора

В начальный момент времени автомобили, которые собираются поворачивать или продолжать движение прямо, равномерно рассредоточены по обоим рядам. По мере приближения к светофору автомобили начинают перестраиваться в нужный ряд, если выполнено одно из условий (13) или (14).

При расчетах использованы следующие параметры, общие для трех ветвей перекрестка:

$L_1 = 1000$ м — длина участка дороги до перекрестка;

$x_0 = 25$ м — длина участка, занятого движущимся транспортом в начальный момент времени $t = 0$;

$v_{\max}^0 = 18$ м/с — максимальная скорость движения на основном участке дороги;

$v_p = 3$ м/с — максимальная скорость автомобилей, поворачивающих в зоне перекрестка;

$k_0 = 7,9$ м/с, $k_1 = 4,1$ м/с — коэффициенты в выражении (2.2.4) для скорости распространения возмущений в потоке автотранспорта;

$a^+ = 1,5$ м/с² — максимальное ускорение потока; $a^- = 5$ м/с² — максимальное (экстренное) ускорение торможения потока;

$a_r = 1,5$ м/с² — стандартное ускорение торможения;

$\Delta = 50$ м — расстояние принятия решения;

$\sigma = 0,5$ — «вес» локальной ситуации;

$\tau^+ = 3,3$ с, $\tau^- = \infty$ — время задержки подстройки под безопасную скорость движения в выражении (2.1.3);

$\rho_{tr} = 0,8$ — критическая плотность на полосе, на которую должно быть выполнено перестроение;

$\tau_{tr} = 5$ с — характерное время перестроения;

$\Delta_{tr} = 20$ м — дополнительная дистанция;

$\delta_{tr} = 0,01$ в выражении (2.5.3); $t_1 = 60$ с, $t_2 = 45$ с, $t_3 = 5$ с — длительности фаз работы светофора;

$\rho_0^1 = 0,07$, $\rho_0^2 = 0,08$, $\rho_0^3 = 0,11$ — плотности потока автотранспорта на входе в расчетную область при $x = 0$ для левой, правой и нижней частей перекрестка соответственно;

$p_0^1 = 0,26$, $p_0^2 = 0,65$, $p_0^3 = 0,53$ — доля транспортных средств, которые должны пересечь перекресток в правом ряду для левой, правой и нижней частей перекрестка соответственно.

На рис. 8 представлены результаты численного моделирования движения потока транспорта через Т-образный перекресток, регулируемый светофором. Здесь изображена плотность потоков транспорта в районе Т-образного перекрестка в моменты времени, обозначенные на рисунке. Шкала соответствия цвета значению плотности транспортного потока приведена в правом углу рисунка. В левом углу изображен график зависимости плотности потока от координаты расчетной области в левом ряду левой части перекрестка при движении слева направо, ось направлена по ходу движения транспортного потока. Момент времени $t = 412$ с соответствует первой фазе работы светофора, во время которой для левого ряда левой вет-

ви перекрестка работает запрещающий сигнал. На графике видно, что в ходе работы запрещающего сигнала начинает образовываться второй участок повышенной плотности. Момент времени $t = 465$ с соответствует второй фазе работы светофора, в ходе которой включен сигнал, разрешающий проезд перекрестка по рассматриваемому участку магистрали. На графике изображено, что к этому моменту два участка повышенной плотности объединились в один затор, распространяющийся навстречу потоку. Момент времени $t = 482$ с соответствует третьей фазе работы светофора, в ходе которой проезд перекрестка по рассматриваемой полосе запрещен. График показывает, что у перекрестка вновь начинает образовываться очередь из автомобилей. Поскольку предыдущий участок повышенной плотности за время работы разрешающего сигнала не успел миновать перекресток, то с течением времени затор, образующийся перед светофором, растет. Следовательно, для данного режима работы светофора $\rho_0^1 = 0,07$ превышает ρ_0^* , максимальную плотность входящего потока, не приводящую к затору.

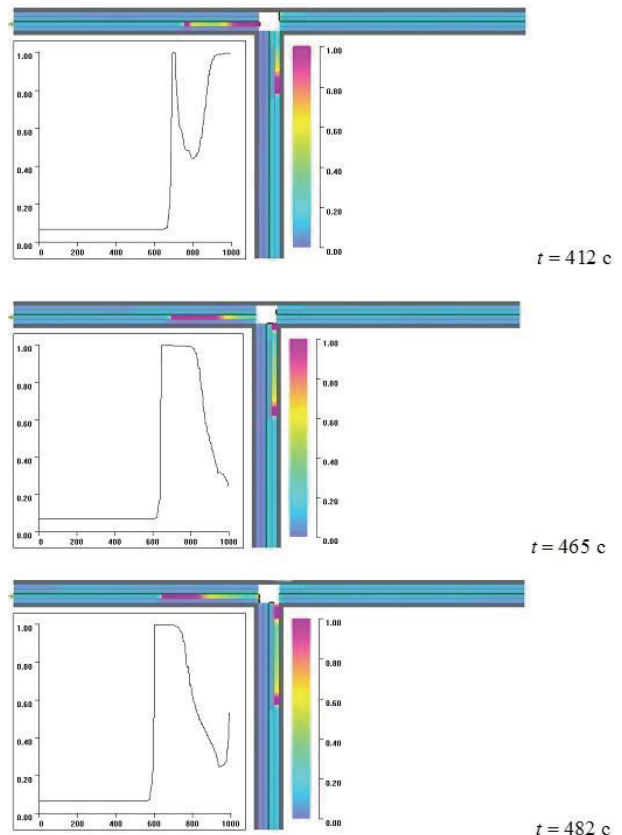


Рис. 8. Плотность потоков транспорта в районе Т-образного перекрестка в моменты времени, обозначенные на рисунке

III.3. Влияние перестроений транспортных средств из ряда в ряд на пропускную способность участка магистрали

Построенная модель движения транспортного потока через Т-образный перекресток использована для изучения влияния межрядовых перестроений на пропускную способность магистрали. Для этого нижняя ветвь перекрестка рассматривалась изолированно от остальной части, а максимальная скорость поворачивающих транспортных средств принималась равной максимально разрешенной скорости на данном участке магистрали $v_p = v_{\max}^0$.

Численное решение поставленной задачи осуществлялось методом TVD со вторым порядком точности [21]. Число узлов расчетной сетки $N = 200$.

При расчетах использованы следующие параметры:

$L_1 = 1000$ м — длина расчетной области;

$x_0 = 25$ м — длина участка, занятого движущимся транспортом в начальный момент времени $t = 0$;

$\rho_0 = 0.04 \text{ div } 0.2$ — плотность потока автотранспорта на входе в расчетную область при $x = 0$;

$v_{\max}^0 = 18$ м/с — максимальная скорость движения на рассматриваемом участке дороги;

$k_0 = 7,9$ м/с, $k_1 = 4,1$ м/с — коэффициенты в формуле (9) для скорости распространения возмущений в потоке автотранспорта;

$a^+ = 1,5$ м/с² — максимальное ускорение потока; $a^- = 5$ м/с² — максимальное (экстренное) ускорение торможения потока;

$a_r = 1,5$ м/с² — стандартное ускорение торможения;

$\Delta = 50$ м — расстояние принятия решения;

$\sigma = 0,5$ — «вес» локальной ситуации;

$\tau^+ = 3,3$ с, $\tau^- = \infty$ — время задержки подстройки под безопасную скорость движения в выражении (3);

$t_g = 40 \text{ div } 350$ с, $t_r = 50$ с — длительности разрешающего и запрещающего сигналов светофора;

$\rho_{tr} = 0,8$ — критическая плотность на полосе, на которую должно быть выполнено перестроение;

$\tau_{tr} = 5$ с — характерное время перестроения;

$\Delta_{tr} = 20$ м — дополнительная дистанция;

$\delta_{tr} = 0,01$ в выражении (15).

В расчетах варьировалась плотность входящего потока ρ_0 (а значит, и его скорость) и длительность работы зеленого сигнала светофора t_g .

Результаты расчетов представлены на рис. 9, 10, 11, 12, 13, 14, 15 и в табл. 2.

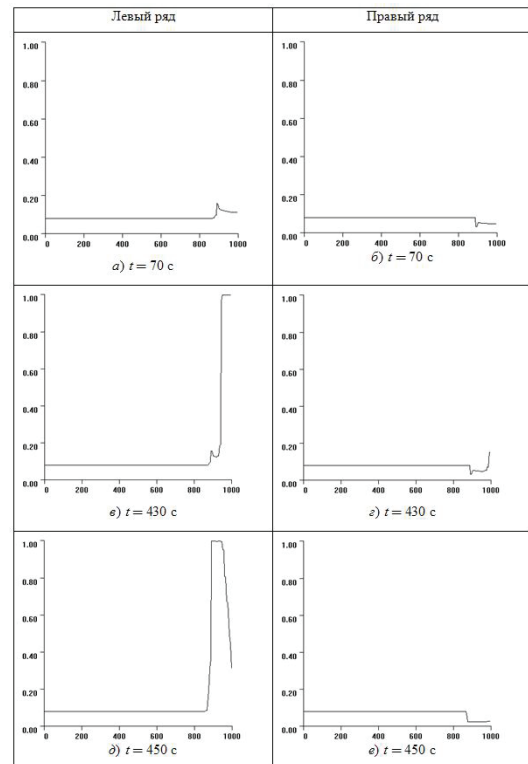


Рис. 9. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени для случая отсутствия подвижной пробки

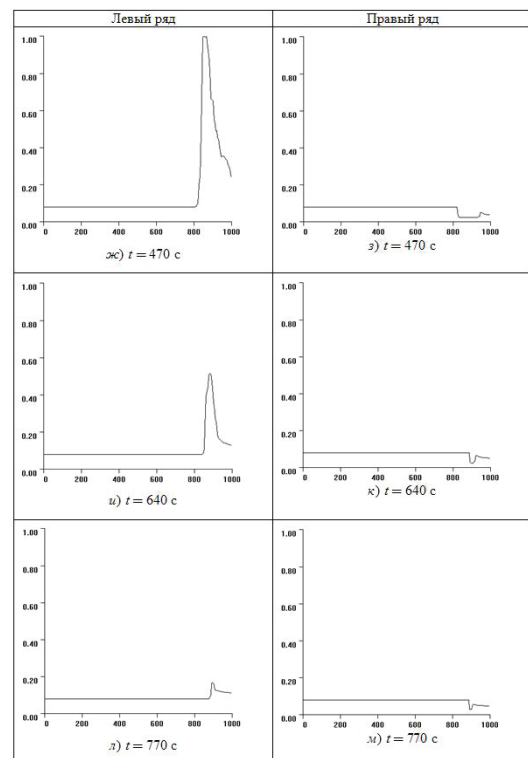


Рис. 10. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени для случая отсутствия подвижной пробки

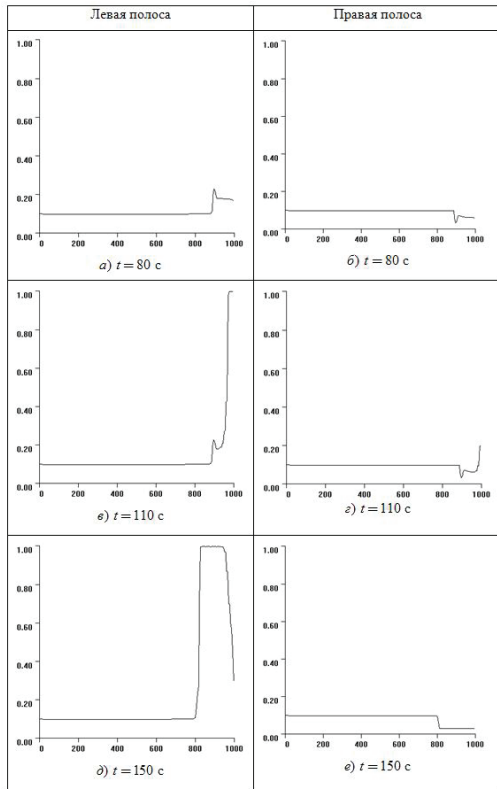


Рис. 11. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени для случая возникновения подвижной пробки

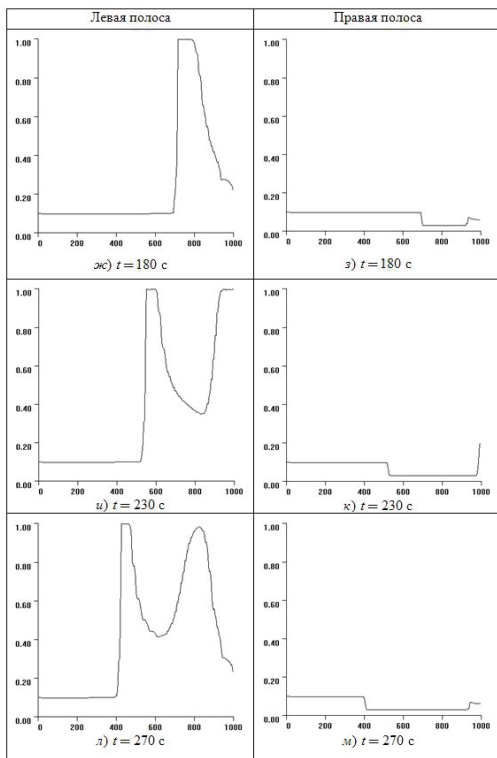


Рис. 12. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени для случая возникновения подвижной пробки



Рис. 13. Шкала соответствия цвета значению плотности транспортного потока



Рис. 14. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени для случая отсутствия подвижной пробки. Движение транспортного потока происходит слева направо

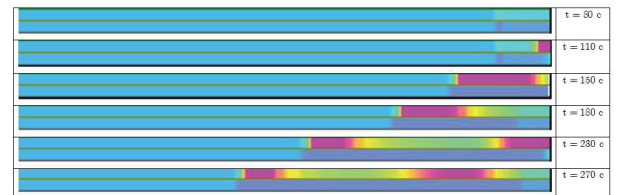


Рис. 15. Распределения плотностей ρ_m и ρ_k в левой и правой полосах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени в случае возникновения подвижной пробки. Движение транспортного потока происходит слева направо

На рис. 9, 10 и 14 даны распределения плотностей ρ_m и ρ_k в левом и правом рядах транспортного потока в зависимости от координаты расчетной области x в различные моменты времени, указанные на рисунке. При этом длительность работы зеленого сигнала светофора $t_g = 350$ с, $p = 0,3$, а начальная плотность транспортного потока равна $\rho_0 = 0,08$. В начальный момент времени работает красный сигнал светофора и до окончания его работы остается 40 с. При таких начальных условиях поток транспорта движется без образования «подвижной пробки». Время на рис. 9, 10 и 14 соответствует следующим моментам цикла работы светофора: $t = 70$ с — зеленый период первого цикла; $t = 430$ с — красный период второго цикла; $t = 450-770$ с — 4 последовательных момента зеленого периода второго цикла. Как видно из этих графиков, при $t = 70$ с в левом ряду появляется участок повышенной плотности, а в правом — пониженной, вызванные перестроениями из ряда в ряд, после включения красного сигнала плотность в обоих рядах повышается. В левом ряду плотность повышается до своего максимального значения и образуется неподвижная пробка, но за время работы зеленого сигнала зависимость плотности от координаты x возвращается к своему первоначальному виду, что видно на рис. 12л, м. В этом случае «подвижная пробка» с течением времени не образуется. Возвращение профиля плотности

к своему первоначальному виду служит критерием свободного движения потока без образования «подвижной пробки».

На рис. 11, 12, 15 приведены данные для длительности работы зеленого сигнала светофора $t_g = 50$ с, $p = 0,3$ и начальной плотности $\rho_0 = 0,1$. В начальный момент времени работает красный сигнал светофора и до окончания его работы остается 40 с. Время на рис. 11, 12, 15 соответствует следующим моментам цикла работы светофора: $t = 80$ с — зеленый период первого цикла; $t = 110$ с — красный период второго цикла; $t = 150$ с и $t = 180$ с — зеленый период второго цикла; $t = 230$ с — красный период третьего цикла; $t = 270$ с — зеленый период третьего цикла. На рис. 12ж видно, что за время работы зеленого сигнала профиль плотности не успевает вернуться к своему первоначальному виду, и с течением времени в левом ряду образуются участки повышенной плотности, где скорость движения заметно уменьшается. Другими словами, возникает затор, перемещающийся против потока. При этом в правом ряду образуется зона пониженной плотности (рис. 12е, з, к, м). Снижение плотности обусловлено тем, что в левом ряду, в который необходимо выполнить перестроение, плотность превысила критическое значение $\rho_{tr} = 0,8$ и автомобили начинают перестраиваться в левый ряд задолго до светофора.

Результаты исследования зависимости величины предельной начальной плотности потока ρ_0^* , при которой не образуется «подвижная пробка», от длительности зеленого сигнала светофора t_g и p приведены в табл. 2. Остальные исходные параметры фиксированы. Зависимость ρ_0^* от t_g хорошо описывается формулой

$$\rho_0^* = a \ln(t_g) + b, \quad (16)$$

где a, b — параметры, зависящие от многих факторов, включая длительность красного сигнала t_r . Для рассмотренных исходных данных значения a, b также приведены в табл. 2, соответствующие графики изображены на рис. 16.

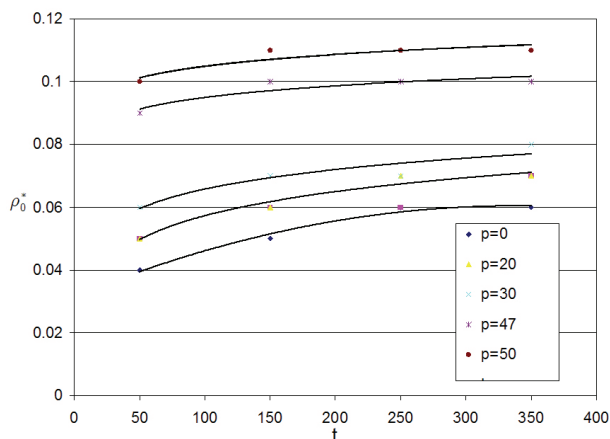


Рис. 16. Зависимость величины предельной начальной плотности потока ρ_0^* , при которой не образуется «подвижная пробка», от длительности зеленого сигнала светофора t_g и p

Проведенные расчеты позволяют сделать вывод, что предельная начальная плотность потока не зависит от процента машин, которые должны проехать перекресток в левом ряду, что объясняется тем, что совокупное количество автомобилей, выполняющих перестроение, не зависит от этого процента. Однако в целом перестроения снижают пропускную способность магистрали в 2 раза по сравнению со случаем отсутствия перестроений.

Т а б л и ц а 2

Зависимость величины предельной начальной плотности потока ρ_0^* , при которой не образуется «подвижная пробка», от длительности зеленого сигнала светофора t_g и p , а также соответствующие значения коэффициентов a и b в формуле (16)

t_1	ρ_1^+	a	b
$p = 0\%$			
50	0,04	0,011	-0,0035
150	0,05		
250	0,06		
350	0,06		
$p = 20\%$			
50	0,05	0,011	0,0065
150	0,06		
250	0,07		
350	0,07		
$p = 30\%$			
50	0,06	0,09	0,0245
150	0,07		
250	0,07		
350	0,08		
$p = 47\%$			
50	0,09	0,054	0,0703
150	0,1		
250	0,1		
350	0,1		
$p = 50\%$			
50	0,1	0,054	0,0803
150	0,11		
250	0,11		
350	0,11		

В результате численного решения рассмотренных выше задач получены следующие основные результаты:

- Для двухполосного транспортного потока построены зависимости предельной начальной плотности, не приводящей к затору, от длительности зеленого сигнала светофора и от процента автомобилей, меняющих полосу движения.
- Установлено, что при регулировании движения двухполосного потока транспорта све-

тофором максимальная пропускная способность магистрали падает почти вдвое по сравнению со случаем отсутствия перестроений. В связи с этим на особо загруженных участках дорог целесообразно ограничить в целях повышения пропускной способности магистрали возможность перестроений.

Литература

1. *Lighthill M.G., Whitham G. B.* On kinetic waves II. A theory of traffic flow on ling crowded roads // Proc. Roy. Soc. London. Ser. A. — 1955. — V. 229, N. 1178. — P. 317–345.
2. *Richards P. L.* Shock waves on the highway // Operations Research. — 1956. — V. 4, N. 1. — P. 42–51.
3. *Уизем Д.* Линейные и нелинейные волны. — М.: Мир, 1977.
4. *Гасников А.В., Кленов С.Л., Нурминский Е.А., Холодов Я.А., Шамрай Н.Б.* Введение в математическое моделирование транспортных потоков / под ред. А.В. Гасникова, с приложениями М.Л. Бланка, Е.В. Гасниковой, А.А. Замятина, В.А. Мальшева, А.В. Колесникова, А.М. Райгородского. — М.: МФТИ, 2010.
5. *Vando M., Hasebe K., Nakanish K., Nakayama A.* Delay of Vehicle Motion in Traffic Dynamics // Japan Journal of Industrial and Applied Mathematics. — 2000. — V. 17. — P. 275–290.
6. *Algers S., Bernauer E., Boero M., Breheret L., Di Taranto C., Dougherty M., Fox K., Gabard G.-F.* Review of Micro-Simulation Models. Review Report of the SMARTTEST project. 1997 (<http://www.its.leeds.ac.uk/projects/smartest/deliv3.html>).
7. *Карамзин Ю.Н., Трапезникова М.А., Четверушкин Б.Н., Чурбанова Н.Г.* Двумерная модель автомобильных потоков // Математическое моделирование. — 2006. — Т. 18, № 6. — С. 85–95.
8. *Смирнов Н.Н., Киселев А.Б., Никитин В.Ф., Юмашев М.В.* Математическое моделирование автотранспортных потоков. — М.: Механико-математический ф-т МГУ им. М.В. Ломоносова, 1999.
9. *Prigogine I., Resibois P.* On a Generalized Boltzmann-like Approach for Traffic Flow // Bull. Cl. Sci., Acad. Roy. Belg. — 1962. — V. 48, N. 9. — P. 805–814.
10. *Greenberg H.* An Analysis of Traffic Flow // Operations Research. — 1959. — V. 7. — P. 79–85.
11. *Payne H.J.* Models of Freeway Traffic and Control: Berkey G.A. Mathematical Models of Public Systems. — 1971. — V. 1. — P. 51–61.
12. *Смирнов Н.Н., Киселев А.Б., Никитин В.Ф., Юмашев М.В.* Математическое моделирование автомобильных потоков на магистралях // Вестн. МГУ. Сер. 1. Матем. механ. — 2000. — № 4. — С. 39–44.
13. *Киселев А.Б., Никитин В.Ф., Смирнов Н.Н., Юмашев М.В.* Неустановившиеся движения автотранспорта на кольцевой магистрали // Прикл. матем. и механ. (ПММ). — 2000. — Т. 64, в. 4. — С. 671–678.
14. *Kiselev A.B., Smirnov N.N., Nikitin V.F., Yumashev M.V.* Mathematical modelling of traffic flow // Proc. of 9th IFAC Symposium in Transportation Systems 2000 (June 13–15, 2000, Braunschweig, Germany). — Elsevier Science Ltd., Oxford, 2000. — P. 223–232.
15. *Смирнов Н.Н., Киселев А.Б., Никитин В.Ф., Асташова Е.Г., Асташов Н.А.* Математическое моделирование динамики автотранспортных потоков и вызываемого ими загрязнения атмосферного воздуха в автомобильных тоннелях // Научные технологии. — 2003. — Т. 4, № 9. — С. 29–43.
16. *Киселев А.Б., Кокорева А.В., Никитин В.Ф., Смирнов Н.Н.* Математическое моделирование автотранспортных потоков на регулируемых дорогах // Прикл. матем. и механ. (ПММ). — 2004. — Т. 68, в. 6. — С. 1047–1054.
17. *Киселев А.Б., Кокорева А.В., Никитин В.Ф., Смирнов Н.Н.* Математическое моделирование движения двухполосного автотранспортного потока, регулируемого светофором // Вестн. МГУ. Сер. 1. Матем. механ. — 2006. — № 4. — С. 35–40.
18. *Kiselev A.B., Kokoreva A.V., Nikitin V.F., Smirnov N.N.* Computational modelling of traffic flows // European Conf. on Computational Fluid Dynamics (Egmond aan Zee, The Netherlands, 5–8 Sept. 2006). — Proc. on CD-ROM. — 10 p.
19. *Киселев А.Б., Кокорева А.В., Никитин В.Ф., Смирнов Н.Н.* Математическое моделирование движения автотранспортных потоков методами механики сплошной среды. Исследование влияния искусственных дорожных неровностей на пропускную способность участка дороги // Современные проблемы математики и механики. Т. I. Прикладные исследования / под редакцией В.В. Александрова и В.Б. Кудрявцева. — М.: Изд-во МГУ, 2009. — С. 311–322.
20. *Greenberg H.* An analysis of traffic flow // Operations Research. — 1959. — V. 7. — P. 79–85.
21. *Оран Э., Борис Дж.* Численное моделирование реагирующих потоков. — М.: Мир, 1990.

Поступила в редакцию 15.10.2010.

Я.А. Холодов, А.С. Холодов, А.В. Гасников, И.И. Морозов, В.Н. Тарасов
Московский физико-технический институт (государственный университет)

Моделирование транспортных потоков — актуальные проблемы и перспективы их решения

Работа посвящена математическому моделированию транспортных потоков, описанию проблем, которые при этом возникают, и возможностей их решения. Предложена оригинальная макроскопическая модель для описания автомобильного движения на сложном графе транспортной сети. Макроскопическая модель основана на гидродинамическом подходе, где транспортный поток описывается уравнениями течения сжимаемой многокомпонентной жидкости с мотивацией. В качестве компоненты рассматриваются автомобили, объединённые общими характеристиками. Разработан также оригинальный алгоритм построения систем уравнений в узлах графа транспортной сети — перекрестках. Результаты численных расчётов при сравнении с экспериментальными данными показывают работоспособность предлагаемой модели.

Ключевые слова: автомобильное движение, гидродинамический подход, граф транспортной сети, многокомпонентные транспортные потоки.

I. Введение

Транспортные проблемы современных мегаполисов хорошо известны и достаточно актуальны. Решение этих проблем не всегда очевидно и уж точно всегда требует времени, при этом никто не гарантирует, что в процессе решения текущих проблем не возникнут новые. В этой связи возникает вопрос: как можно облегчить жизнь непосредственным участникам дорожного движения, а именно водителям. Ответ на этот вопрос не однозначен, поскольку требует комплексного подхода и глубокого понимания существующих на сегодняшний день проблем. Реально складывающаяся дорожно-транспортная ситуация требует комплексных решений не только в сфере развития и оптимизации транспортной инфраструктуры, но и в повседневном управлении транспортной системой для обеспечения надёжности её функционирования и безопасности её участников как в штатных, так и в кризисных ситуациях. При этом важно учитывать не только текущую дорожную обстановку, но и иметь детальный ситуативный прогноз на всю транспортную систему с учетом временных и количественных затрат реализации различных сценариев развития. Для этого необходима корректная вычислительная математическая модель, которая будет осуществлять данный прогноз по всей транспортной сети мегаполиса.

Прежде всего следует определить модель и её параметры, с которыми предстоит работать. В разрабатываемой модели следует учесть все возможные факторы, такие как периодически обновляемые данные о текущем состоянии дорожной ситуации, правила движения по полосам на перекрестках, расписания и алгоритмы работы светофоров, локальные перекрытия на участках дороги и т.д. Модель также должна включать временные

зависимости коэффициентов распределения потоков на перекрестках — матрицы перемешивания вместе с матрицей корреспонденций для заданной транспортной сети.

Транспортный поток может моделироваться в зависимости от плотности и скорости дорожного движения разными способами. Начиная от простейших макроскопических моделей, используемых в основном для однополосного беззаторможенного движения автомобильного транспорта, и заканчивая моделированием движения в многополосной пробке на основе гидродинамического приближения, когда транспортный поток уподобляется сжимаемой многокомпонентной жидкости с мотивацией, учитываемой через уравнение состояния транспортного потока. Ключевым моментом также является алгоритм построения систем уравнений в узлах графа транспортной сети — перекрестках. Данные системы уравнений обеспечивают связь моделируемых величин на всем графе транспортной сети путем задания корректных граничных условий для всех его ветвей, входящих и выходящих в каждый из его узлов.

Отдельно встает задача о выборе оптимального маршрута для каждого из водителей, которую можно рассматривать как задачу принятия решения. А именно, путем исследования множества достижимости куда и с какими характеристиками можно доехать, например, из заданной точки за данное время. При этом важно учитывать не только текущую дорожную обстановку, но и иметь детальный прогноз на всю траекторию движения с учетом временных затрат по маршруту. Мысленно себе это можно представлять как исследование растекания жидкости из вершины по транспортному графу.

В основу предлагаемой технологии положен имеющийся опыт использования сетевых вычис-

лительных моделей, основанный на решении соответствующих краевых задач для уравнений в частных производных на графах [1, 2]. Данный опыт показывает, что при современном уровне развития вычислительной математики и вычислительной техники, особенно с появлением высокопроизводительных вычислительных платформ и решений, такой подход позволяет эффективно решать достаточно сложные «глобальные» задачи данного класса [3].

II. Сетевая модель интенсивного дорожного движения в мегаполисе

Для макро моделирования автомобильного движения используются различные математические модели, в том числе основанные на дифференциальных уравнениях в частных производных и, в частности, гидродинамические модели, аналогичные уравнениям течения сжимаемой многокомпонентной жидкости с мотивацией (см. [4–11] и др.). Ниже также рассматривается вычислительная математическая модель интенсивного уличного движения в мегаполисе, основанная на решении соответствующих краевых задач для уравнений в частных производных гиперболического типа.

Будем представлять уличную сеть мегаполиса в виде перекрестков (узлов графа с номерами $l = 1, \dots, L$), связанных между собой дорогами с длиной X_k (ребрами графа с номерами $k = 1, \dots, K$, $0 \leq x_k \leq X_k$) (см., например, рис. 1). Введем следующие обозначения для используемых в модели параметров: $\rho_k(x_k, t)$ — плотность автомобильного потока (количество автомобилей на единицу длины дороги с номером k); $v_k(x_k, t)$ — средняя по сечению дороги линейная скорость потока автомобилей на k -й ветви графа в точке x_k в момент времени t ; $q_k(x_k, t) = \rho_k v_k$ — поток автомобилей.

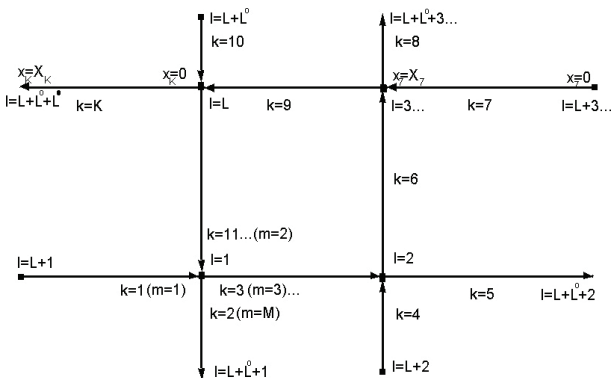


Рис. 1. Пример направленного графа для перекрестка

II.1. Система уравнений автомобильного движения на ребре графа транспортной сети

Гиперболическая система уравнений, описывающая автомобильное движение, представляет со-

бой дифференциальные законы сохранения (изменения) «массы» и «импульса» на автодорогах (по аналогии с гидродинамикой) и в дивергентной форме записывается в виде (см., например, [6]):

$$\begin{cases} \partial \rho_k / \partial t + \partial(\rho_k v_k) / \partial x_k = f_{0k}, \\ \partial(\rho_k v_k) / \partial t + \partial(\rho_k v_k^2 + p(\rho_k)) / \partial x_k = f_{1k}. \end{cases} \quad (1)$$

Здесь

$$p = p(\rho_k) \quad (2)$$

есть замыкающее систему (1) уравнение состояния (зависимость давления от плотности, рис. 2), f_{0k} — возможные источники или стоки «массы» (въезжающие на дорогу автомобили из не учитываемых явно элементов уличной сети, останавливающиеся или начинающие движение автомобили и т.п.), f_{1k} — возможные внешние импульсы, действующие на систему (1). В таких моделях основной проблемой является построение адекватного действительности уравнения состояния (2), конкретный вид которого, как и для всякой феноменологической модели, должен быть определен из экспериментальных измерений (возможно с использованием параметрических решений системы (1), (2)). Представленные на рис. 3, 4, 5, 6 для движения на одной (левой) полосе высокоскоростной автострады экспериментальные зависимости потока от плотности из [12–15] показывают (отмечены точками и соединяющими их отрезками прямых), что предельная величина потока (определяемая пропускной способностью дороги, техническими характеристиками автомобилей и правилами дорожного движения) много больше реально реализуемой максимальной его величины $q_{fs} = q_{\max}^{(free)}$ и реально реализуемое движение в фазовой плоскости $\{\rho_s, q_s\}$ (номер ветви графа k далее опущен) имеет своей границей параболу (кривая № 1 на рис. 3–6):

$$q_s(\rho_s) = \rho_s(q_{fs}/\rho_{fs} + c_{fs}(1 - \rho_s/\rho_{fs})), \quad \rho_s \leq \rho_{fs} \quad (3)$$

при плотности одной полосы дороги, меньшей критической $\rho_s < \rho_{fs}$ (область «свободного потока»). Здесь c_{fs} — аналог скорости звука при критической плотности, s — номер полосы движения ($1 \leq s \leq S$), S — число полос заданной автодороги.

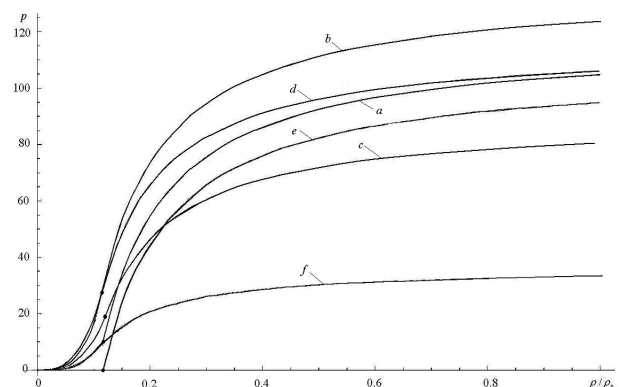


Рис. 2. Примеры уравнения состояния $p(\rho)$

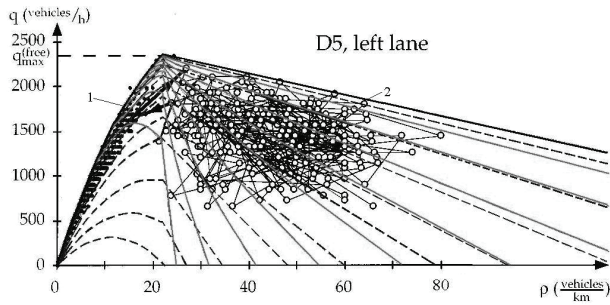


Рис. 3. Аппроксимация экспериментальных данных из [13]

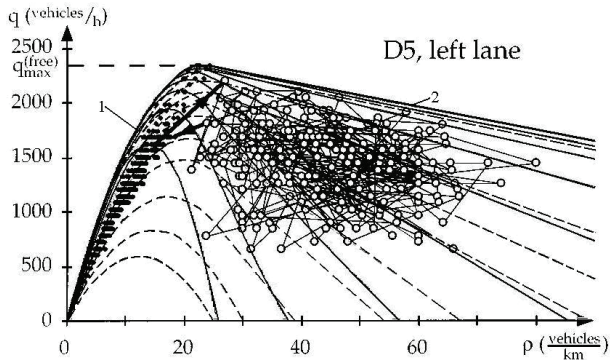


Рис. 4. Аппроксимация экспериментальных данных из [13]

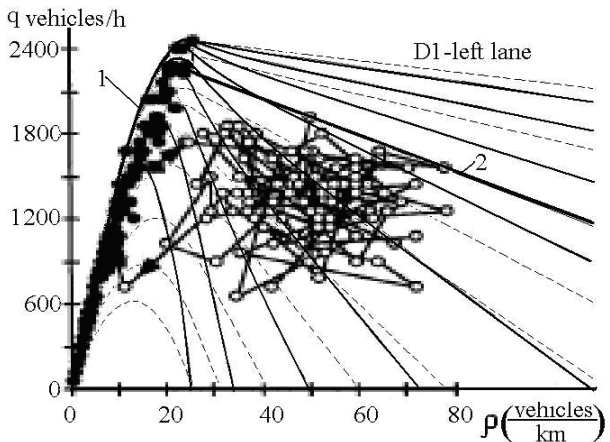


Рис. 5. Аппроксимация экспериментальных данных из [14]

Однако при дальнейшем увеличении плотности до значений $\rho_s > \rho_{fs}$ такой однозначной зависимости, как в (3), из представленных экспериментальных данных не наблюдается вследствие взаимодействия волн разгона и торможения при интенсивном (с закритической плотностью $\rho_s > \rho_{fs}$) автомобильном движении. Скорость свободного движения потока автомобилей $v_{s0}(0) = (\partial q_s(\rho_s)/\partial \rho_s)_0$ определяется правилами дорожного движения, техническими характеристиками автомобилей и стилем вождения участников дорожного движения и обеспечивает связь этого параметра со значениями $q_{fs}, \rho_{fs}, c_{fs}$:

$$v_{s0} = v_s(0) = q_{fs}/\rho_{fs} + c_{fs}. \quad (4)$$

Предполагая существование аналогичной (3) однозначной зависимости потока от плотности также и при $\rho_s > \rho_{fs}$, будем полагать ее линейной в

области реализуемых значений при $\rho_s > \rho_{fs}$ (прямая № 2 на рис. 3–6):

$$q_s(\rho_s) = c_{*s}(\rho_{*s} - \rho_s), \quad q_s(\rho_{fs}) = c_{*s}(\rho_{*s} - \rho_{fs}) = d_s q_{fs}, \quad \rho_s > \rho_{fs} \quad (5)$$

с обращающимся в ноль при максимальной плотности $\rho_s = \rho_{*s}$ потоком и с возможным его разрывом в случае $d_s < 1$ при критической плотности $\rho_s = \rho_{fs}$, коэффициент разрыва d_s определяется из измеренных данных:

$$d_s = c_{*s}(\rho_{*s} - \rho_{fs})/q_{fs} \leq 1. \quad (6)$$

Здесь ρ_{*s} — предельно допустимая плотность потока автомобилей (пропускная способность одной полосы дороги), соответственно c_{*s} — аналог скорости звука при этой плотности.

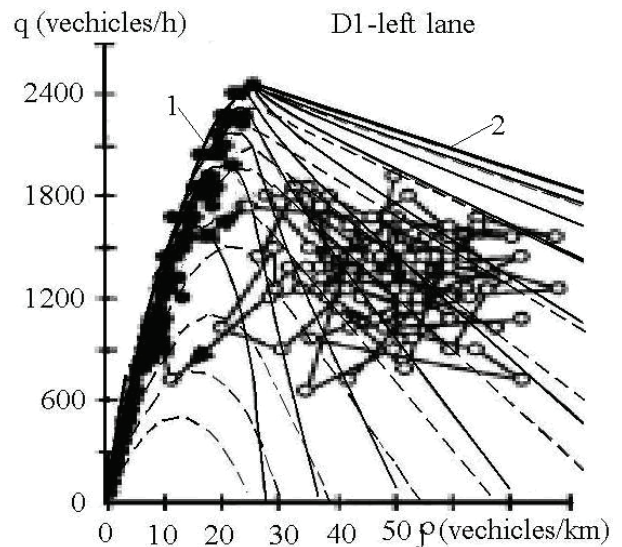


Рис. 6. Аппроксимация экспериментальных данных из [14]

Зависимости $q_s = q_s(\rho_s)$ из (3), (5) или аналогичных аппроксимаций измеренных данных вместе с первым уравнением из (1) (при $f_{0k} = 0$ дифференциальным законом сохранения числа автомобилей), так же как и система уравнений (1)–(2), могут непосредственно использоваться для моделирования дорожного движения. В этом случае движение автомобилей описывается одним нелинейным уравнением переноса для плотности (см., например, [4, 5]):

$$\partial \rho_s / \partial t + \partial q_s(\rho_s) / \partial x = \partial \rho_s / \partial t + \frac{\partial q_s(\rho_s)}{\partial \rho_s} \partial \rho_s / \partial x = 0. \quad (7)$$

Скорость v_s всегда может быть вычислена с использованием рассчитанных из (7) значений ρ_s и известной фундаментальной диаграммы $q_s(\rho_s)$:

$$v_s(x, t) = q_s(\rho_s(x, t)) / \rho_s(x, t). \quad (8)$$

В данной работе ставится тем не менее другая задача — получить из известного вида фундаментальной диаграммы аналог уравнения состояния

(зависимость давления от плотности) для заданной полосы автодороги $p_s = p_s(\rho_s)$ и использовать для решения системы (1). Использование одного уравнения не является достаточным для корректного описания всех фаз (состояний) транспортного потока, наблюдаемых на рис. 3–6 (см. [11]). Сделать это можно, если воспользоваться хорошо известными дифференциальными преобразованиями законов сохранения (см., например, [16]). Умножим (7) на $\partial q_s / \partial \rho_s$:

$$\frac{\partial q_s}{\partial \rho_s} \partial \rho_s / \partial t + \frac{\partial q_s}{\partial \rho_s} \partial q_s / \partial x = \partial q_s / \partial t + \frac{\partial q_s}{\partial \rho_s} \partial q_s / \partial x = 0 \quad (9)$$

и воспользуемся следующими дифференциальными соотношениями:

$$\partial q_s / \partial t = \partial(\rho_s v_s) / \partial t = v_s \partial \rho_s / \partial t + \rho_s \partial v_s / \partial t,$$

$$\partial q_s / \partial x = \partial(\rho_s v_s) / \partial x = v_s \partial \rho_s / \partial x + \rho_s \partial v_s / \partial x,$$

подставим их в (9), откуда получим

$$\partial v_s / \partial t + \frac{\partial q_s(\rho_s)}{\partial \rho_s} \partial v_s / \partial x = 0. \quad (10)$$

Итак, вместо системы уравнений (1) получен аналог этих законов сохранения в виде системы уравнений (7), (10) в недивергентной форме с нулевой правой частью:

$$\begin{cases} \partial \rho_s / \partial t + \frac{\partial q_s(\rho_s)}{\partial \rho_s} \partial \rho_s / \partial x = 0, \\ \partial v_s / \partial t + \frac{\partial q_s(\rho_s)}{\partial \rho_s} \partial v_s / \partial x = 0. \end{cases} \quad (11)$$

Из сопоставления уравнений (10) и второго уравнения в (1), записанного в недивергентной форме, можно получить вид уравнения состояния (2). Для этого в (10) добавим и вычтем член $v_s \partial v_s / \partial x$ и воспользуемся следствием второго дифференциального соотношения: $\partial v_s / \partial x = (\partial q_s / \partial \rho_s - q_s / \rho_s)(\partial \rho_s / \partial x) / \rho_s$. Тогда (10) приводится к виду:

$$\partial v_s / \partial t + v_s \partial v_s / \partial x + \frac{(q_s / \rho_s - \partial q_s / \partial \rho_s)^2}{\rho_s} \partial \rho_s / \partial x = 0.$$

Сравним это уравнение с недивергентной формой второго уравнения системы (1) для одной полосы с нулевой правой частью: $\partial v_s / \partial t + v_s \partial v_s / \partial x + \frac{1}{\rho_s} \frac{\partial p_s}{\partial \rho_s} \partial \rho_s / \partial x = 0$, получим соотношение, связывающее производную, давление $\partial p_s / \partial \rho_s$, аналог скорости звука c_s , поток q_s и плотность ρ_s :

$$\partial p_s / \partial \rho_s = c_s^2 = (q_s / \rho_s - \partial q_s / \partial \rho_s)^2. \quad (12)$$

Интегрируя (12), можно найти функцию $p_s(\rho_s)$ по заданной функции $q_s(\rho_s)$, если последняя описывает достаточно широкий класс дорожного движения.

Для $q_s(\rho_s)$ из (3), (5) имеем

$$p_s(\rho_s) = c_{fs}^2 \rho_s^3 / (3 \rho_{fs}^2) \quad \text{при} \quad \rho_s \leq \rho_{fs}, \quad (13)$$

$$p_s(\rho_s) = p_{fs} + \rho_{*s}^2 c_{*s}^2 (1 / \rho_{fs} - 1 / \rho_s),$$

$$p_{fs} = c_{fs}^2 \rho_{fs} / 3 \quad \text{при} \quad \rho_s > \rho_{fs}. \quad (14)$$

Здесь и далее мы полагаем вектор правой части $f = 0$ из следующих соображений. Зависимости потока от плотности, показанные на рис. 3–6, получены от датчиков, которые находились в левой полосе автобана на существенном расстоянии от каких-либо дорожных развязок (см. [13–14]), соответственно там не было никаких возможных источников или стоков «массы» автомобилей. Все остальные параметры, которые исследователи обычно так любят включать в правую часть системы уравнений (1) (см. [4–10] и др.), учитываются автоматически через полученный вид фундаментальной диаграммы для этих измерений и в правую часть не входят.

Вблизи предельных значений плотности $\rho_s \approx \rho_{*s}$ движение является слабосжимаемым, поэтому для обеспечения бесстолкновительного движения в области сильных заторов бралось $\rho_{*s} \geq \rho_s \geq \rho_\varepsilon = \rho_{*s}(1 - \varepsilon)$, $\varepsilon \sim 0,1 - 0,01$, соответственно уравнение состояния (14) модифицировалось следующим образом:

$$\begin{cases} p_s(\rho_s) = (p_{\varepsilon s} + (\rho_s - \rho_{\varepsilon s})(B + C(\rho_s + \rho_{\varepsilon s}))), \\ C = ((\partial p / \partial \rho)_{*s} - (\partial p / \partial \rho)_{\varepsilon s}) / (\rho_{*s} - \rho_\varepsilon), \\ B = (\partial p / \partial \rho)_{*s} - 2C \rho_{*s}. \end{cases} \quad (15)$$

Выбором значений ρ_ε и $(\partial p / \partial \rho)_{\varepsilon s} c_{\varepsilon s}^2$ можно также обеспечить приемлемый шаг интегрирования по времени в области автодороги, где плотность приближается к своему предельному значению ρ_{*s} , а значение $(\partial p / \partial \rho)_{*s}$ начинает неограниченно расти.

Чтобы перейти в системе уравнений (1)–(2) от многополосного к усредненному представлению дороги, для ветви графа с номером $k = 1, \dots, K$ в (1)–(2) аналогично многокомпонентной односкоростной модели баротропного газа будем полагать

$$\begin{aligned} \rho_k &= S_k \tilde{\rho}_k = \sum_{s=1}^{S_k} \rho_{sk}, & v_k &= \frac{1}{S_k} \sum_{s=1}^{S_k} v_{sk}, \\ q_k &= S_k \tilde{q}_k = \sum_{s=1}^{S_k} q_{sk}, & p_k &= S_k \tilde{p}_k = \sum_{s=1}^{S_k} p_{sk}. \end{aligned} \quad (16)$$

Переход (16) удобен тем, что его можно осуществлять как в прямую, так и в обратную сторону в любом участке дороги при возникновении такой необходимости. Индекс s можно связывать не только с номером полосы дороги, но и с маршрутом водителя, типом автомобиля, стилем вождения или опытом водителей и т.д.

При этом для различных участников движения, имеющих существенно отличающиеся друг от друга константы в уравнениях состояния (13),

(14), в дополнение к системе (1), (2), как это принято, можно ввести в рассмотрение их концентрации:

$$C_{sk} = \rho_{sk} / \sum_{s=1}^{S_k} \rho_{sk}, \quad \sum_{s=1}^{S_k} C_{sk} = 1 \quad (17)$$

в такой равновесной или неравновесной, односкоростной ($v_k = \frac{1}{S_k} \sum_{s=1}^{S_k} v_{sk}$) или многоскоростной смеси и в простейшем случае для определения этих концентраций использовать соответствующие уравнения конвективно-диффузионного переноса (в том числе между разными полосами догоги):

$$\frac{\partial C_{sk}}{\partial t} + v_{sk} \frac{\partial C_{sk}}{\partial x} = \mu \frac{\partial^2 C_{sk}}{\partial x^2} + f(C_{1k}, \dots, C_{S_k}), \quad (18)$$

где μ — «коэффициент диффузии», а $f(C_{1k}, \dots, C_{S_k})$ определяет, например, процессы перестроений и т.п.

Имея замкнутую систему уравнений (1)–(2) (хорошо известную систему уравнений Эйлера для сжимаемого баротропного газа), можно найти уравнения характеристик и соответствующие условия совместности этой системы уравнений гиперболического типа, соотношения на разрывах, возникающих при переходе от докритического («сверхзвукового») движения к закритическому («дозвуковому») в волнах торможения и т.д. Интегрируя (1), (2) по замкнутому контуру в плоскости $\{x, t\}$, содержащему фронт разрыва движущегося со скоростью D , можно получить соотношения на разрывах, в частности, уравнение ударной адиабаты, связывающее параметры перед фронтом разрыва (в докритическом потоке с $\rho_{1k} \leq \rho_{fk}$, кривая № 1 на рис. 3, 4, 5, 6, 7, 8) и за фронтом разрыва (в закритическом потоке с $\rho_{2k} > \rho_{fk}$, прямая № 2 на рис. 3–8):

$$(p_{1k}(\rho_{1k}) - p_{2k}(\rho_{2k}))(1/\rho_{2k} - 1/\rho_{1k}) - (v_{1k} - v_{2k})^2 = 0, \quad (19)$$

а также скорость движения фронта разрыва:

$$D = (\rho_{1k}v_{1k} - \rho_{2k}v_{2k})/(\rho_{1k} - \rho_{2k}) = (q_{1k} - q_{2k})/(\rho_{1k} - \rho_{2k}). \quad (20)$$

Эти соотношения и, в частности, вполне измеряемую величину D можно использовать для определения констант в уравнении состояния (13)–(14), как это делается при построении уравнений состояния сложных физических сред.

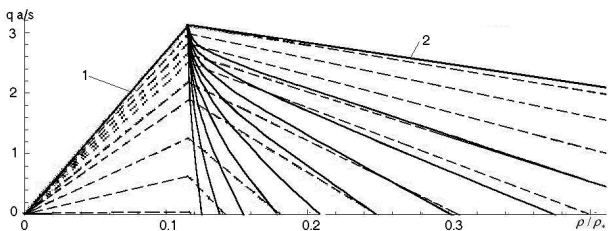


Рис. 7. $c_{fs} = 0.1$ [м/с]: предельные зависимости потока от плотности — кривые № 1, 2; ударные адиабаты — сплошные кривые без номеров; характеристики первого семейства — штриховые линии

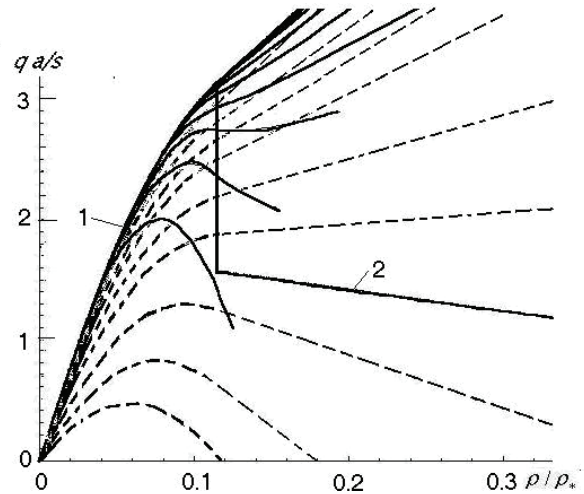


Рис. 8. $d_s = 0.5$: предельные зависимости потока от плотности — кривые № 1, 2; ударные адиабаты — сплошные кривые без номеров; характеристики первого семейства — штриховые линии

Вводя в рассмотрение векторы $V = \{\rho_k, v_k\}$, $U(V) = \{\rho_k, \rho_k v_k\}$, $F(V) = \{\rho_k v_k, \rho_k v_k^2 + p(\rho_k)\}$, $f = \{f_{0k}, f_{1k}\}$, систему уравнений (1) можно записать в векторной дивергентной форме:

$$\partial U(V)/\partial t + \partial F(V)/\partial x_k = f(V, x_k, t), \quad (21)$$

или в недивергентной форме:

$$\partial V/\partial t + A(V)\partial V/\partial x_k = \tilde{f} = (\partial U/\partial V)^{-1} f$$

с матрицей Якоби:

$$A = (\partial U/\partial V)^{-1}(\partial F/\partial V) = \begin{Bmatrix} v_k & \rho_k \\ (\partial p/\partial \rho_k)/\rho_k & v_k \end{Bmatrix}. \quad (22)$$

Собственные числа матрицы Якоби A : $\lambda_1 = v_k + c_k$, $\lambda_2 = v_k - c_k$, $c_k = \sqrt{\partial p/\partial \rho_k}$ всегда действительные и различные ($\partial p/\partial \rho_k = c_k^2 > 0$) — система (21) строго гиперболического типа, а левые собственные векторы: $\omega_1 = \{c_k, \rho_k\}$, $\omega_2 = \{c_k, -\rho_k\}$ линейно независимы. Т. е. матрица со строками из левых собственных векторов $\Omega = \begin{Bmatrix} \omega_1 \\ \omega_2 \end{Bmatrix} = \begin{Bmatrix} c_k & \rho_k \\ c_k & -\rho_k \end{Bmatrix}$ неособенная, так как $\det \Omega = -2\rho_k c_k \neq 0$ при ненулевых плотности ρ_k и «скорости звука» c_k . Для упрощения выкладок в дальнейшем будем полагать вектор правой части $f = 0$.

Эквивалентная (1) характеристическая форма уравнений совместности вдоль характеристик первого (верхний знак, индекс № 1) и второго (нижний знак, индекс № 2) семейств

$$dx = \lambda_i dt = (v_k \pm c_k) dt, \quad i = 1, 2 \quad (23)$$

имеет вид

$$\begin{cases} \omega_i \cdot \frac{dV}{dt_i} = c_k(\rho_k) d\rho_k/dt_i \pm \rho_k dv_k/dt_i = 0, \\ \frac{d}{dt_i} = \frac{\partial}{\partial t} + \lambda_i \frac{\partial}{\partial x_k}, \quad i = 1, 2. \end{cases} \quad (24)$$

Каждое из уравнений (24) является, по сути, обыкновенным дифференциальным уравнением вдоль i -й характеристики (23). Для уравнения состояния (13)–(14) условия совместности (24) интегрируются аналитически, что позволяет получить вид инвариантов Римана r_i , постоянных вдоль соответствующих характеристик λ_i , $i = 1, 2$:

$$r_i(\rho_k, v_k) = c_{fs}\rho_k / (S\rho_{fs}) \pm v_k = \text{const},$$

при $\rho_k \leq S\rho_{fs}$; (25)

$$r_i(\rho_k, v_k) = -(c_{*s}\rho_{*s}S/\rho_k) \pm v_k = \text{const},$$

при $\rho_k > S\rho_{fs}$. (26)

В фазовой плоскости $\{\rho, q\}$ кривые (25), (26) имеют вид

$$q_k(\rho_k) = \pm \rho_k (r_i - c_{fs}\rho_k / (S\rho_{fs})) = \text{const}, \quad i = 1, 2$$

при $\rho_k \leq S\rho_{fs}$; (27)

$$q_k(\rho_k) = \pm (r_i\rho_k + c_{*s}\rho_{*s}S) = \text{const}, \quad i = 1, 2$$

при $\rho_k > S\rho_{fs}$. (28)

Характеристическая форма уравнений (24) или ее модификации с использованием инвариантов Римана (25), (26) (если удастся найти точное решение (24)), так же как соотношения на разрывах (19), (20), необходимы при построении эффективных разностных схем для численного решения системы (1)–(2) (см., например, [17]), а в граничных точках они используются для замыкания краевых условий.

Для представленных на рисунках экспериментальных данных из [13] (рис. 3–4) и из [14] (рис. 5–6) (точки и соединяющие их отрезки прямых) константы $\rho_{fs}, q_{fs}, c_{fs}, d_s, \rho_{*s}, c_{*s}$ в уравнениях (3), (5), аппроксимирующих границы области реализуемых движений автомобилей, принялись следующими.

На рис. 3:

$$\begin{cases} \rho_{fs} = 0,0243[\text{АТС/М}], \\ q_{fs} = 0,656[\text{АТС/с}], \quad c_{fs} = 15,8[\text{М/с}], \\ d_s = 1, \\ \rho_{*s} = 0,21[\text{АТС/М}], \quad c_{*s} = 3,53[\text{М/с}], \\ v_{s0} = 42,8[\text{М/с}]. \end{cases} \quad (29)$$

На рис. 4:

$$\begin{cases} \rho_{fs} = 0,0243[\text{АТС/М}], \\ q_{fs} = 0,656[\text{АТС/с}], \quad c_{fs} = 27[\text{М/с}], \quad d_s = 1, \\ \rho_{*s} = 0,21[\text{АТС/М}], \quad c_{*s} = 3,53[\text{М/с}], \\ v_{s0} = 54[\text{М/с}]. \end{cases} \quad (30)$$

На рис. 5:

$$\begin{cases} \rho_{fs} = 0,0213[\text{АТС/М}], \\ q_{fs} = 0,6[\text{АТС/с}], \quad c_{fs} = 22,4[\text{М/с}], \\ d_s = 0,9, \\ \rho_{*s} = 0,21[\text{АТС/М}], \quad c_{*s} = 2,9[\text{М/с}], \\ v_{s0} = 50,6[\text{М/с}]. \end{cases} \quad (31)$$

На рис. 6:

$$\begin{cases} \rho_{fs} = 0,0213[\text{АТС/М}], \\ q_{fs} = 0,6[\text{АТС/с}], \quad c_{fs} = 22,4[\text{М/с}], \\ d_s = 0,9, \\ \rho_{*s} = 0,21[\text{АТС/М}], \quad c_{*s} = 3,17[\text{М/с}], \\ v_{s0} = 50,6[\text{М/с}]. \end{cases} \quad (32)$$

Соответствующие им граничные зависимости потока от плотности (3), (5) показаны сплошными линиями 1, 2 на рис. 3–8. Для этих же значений констант уравнения состояния (13), (14) представлены на рис. 2. Кривые a, b, c, d соответствуют константам из (29), (30), (31) и (32).

На рис. 3–6 сплошными линиями без номеров показаны ударные адиабаты (19) для ряда значений $\rho_{1k}, v_{1k} = q_{1k}/\rho_{1k}$, принадлежащих кривым (3), а штриховыми линиями показаны инварианты Римана первого семейства (27), (28), $i = 1$. Инварианты Римана второго семейства $i = 2$ симметричны первому семейству относительно оси абсцисс. Отдельно стоит обратить внимание на рис. 5. Здесь мы видим, что переход из докритического («сверхзвукового», кривая № 1) движения к закритическому («дозвуковому», прямая № 2) осуществляется через ударную волну сжатия и коэффициент разрыва $d_s < 1$. Таким образом, данный процесс не является адиабатическим. Это приводит к тому, при возвращении назад в ту же точку на кривую № 1 с прямой № 2 через волну разрежения невозможно.

Для оценки влияния констант в уравнении состояния (13), (14) на поведение давления в варианте (30) более существенно, чем в (29), (31) и (32), были изменены коэффициенты c_{fs}, d_s . На рис. 7 для случая $c_{fs} = 0,1$ [М/с] и на рис. 8 для $d_s = 0,5$ представлены аналогичные описанным выше зависимости параметров: предельные зависимости потока от плотности — кривые № 1, 2; ударные адиабаты — сплошные кривые без номеров; характеристики первого семейства — штриховые линии. Соответствующие зависимости для давления показаны на рис. 2 (кривые e и f). Видно, что существенное отличие параметра d_s от 1 приводит к заведомо неправильному поведению ударных адиабат (увеличение скорости при переходе от докритического режима к закритическому в волне торможения и т.п.). Из сравнения кривых a и e на рис. 2 видно также, что аппроксимация в (3) докритической ветви $q_s(\rho_s)$ даже линейной функцией ($c_{fs} \approx 0$) влияет на уравнение состояния сравнительно слабо. Сопоставление кривых a и c на рис. 2, для которых $\partial q_s(\rho_{fs})/\partial \rho_s > 0$,

с кривыми b и d , для которых $\partial q_s(\rho_{fs})/\partial \rho_s = 0$, показывает, что равенство нулю $\partial q_s(\rho_{fs})/\partial \rho_s = 0$ дает более гладкое сопряжение докритической и закритической ветвей $p_s(\rho_s)$ и обеспечивает принципиальную возможность перехода в закритический режим движения без экстренного торможения (аналог адиабатического безударного сжатия потока в газовой динамике). В этом случае закритическая ветвь зависимости потока от плотности (5) (прямая № 2 на рис. 6) является слабой волной торможения (19) (при переходе через которую параметры непрерывны, а разрывными являются нормальные производные) и при этом остается одной из характеристик семейства № 1.

II.2. Система уравнений автомобильного движения в узлах графа транспортной сети

Число граничных условий на свободных концах ветвей графа дорожной сети (входах $x_k = 0$ и выходах $x_k = X_k$) зависит от знаков собственных чисел матрицы Якоби A (22). Их количество на входах может быть равным двум при положительных $\lambda_1, \lambda_2 > 0$, что полностью определяет параметры такого узла, одному при $\lambda_1 > 0, \lambda_2 < 0$ или нулю при отрицательных $\lambda_1, \lambda_2 < 0$. На выходах их число может быть равным нулю при положительных $\lambda_1, \lambda_2 > 0$, единице при $\lambda_1 > 0, \lambda_2 < 0$ или двум при отрицательных $\lambda_1, \lambda_2 < 0$.

В соответствии с этим на входах–выходах из дорожной сети в качестве граничных условий могут быть заданы как функции времени значение плотности потока автомобилей $\rho_k(t)$ или величина линейной скорости $v_k(t)$ (при $\lambda_1 > 0, \lambda_2 < 0$), для расчета таких узлов в качестве второго уравнения привлекается одно из условий совместности (24) или (25), (26) вдоль идущих внутрь области интегрирования характеристик. Если необходимо задавать два граничных условия, обе эти переменные задаются одновременно. Также возможна ситуация, когда граничные узлы рассчитываются через значения во внутренних точках дороги, тогда задавать граничные условия не требуется. Вместо плотности или линейной скорости может задаваться величина автомобильного потока

$$q_k(t) = \rho_k u_k. \quad (33)$$

На выходах иногда используют неотражающие граничные условия — нулевые производные

$$(\partial q_k / \partial x_k) = 0, \quad (\partial u_k / \partial x_k) = 0, \quad (34)$$

что, вообще говоря, является корректным лишь при положительных $\lambda_1, \lambda_2 > 0$.

Помимо граничных условий, для системы (1), (2) необходимо задать также некоторые начальные условия:

$$\rho_k(x_k, 0) = \rho_{k0}(x_k), \quad v_k(x_k, 0) = v_{k0}(x_k). \quad (35)$$

В точках ветвления графа (во внутренних узлах или на перекрестках) постановка граничных условий существенно зависит от организации движения (учитывается ли многополостное движение на дороге или рассматривается осредненное по направлению движение, учитывается ли наличие светофоров или рассматриваются более длительные временные масштабы и т.д.).

В каждом внутреннем узле графа дополнительной искомой величиной, помимо значений векторов $V_{l1}, V_{l2}, \dots, V_{lM}$ в окончаниях входящих в узел l и выходящих из него $m = 1, \dots, M$ ветвей, является «давление» узла $p_l(t)$. При расчете таких граничных точек в зависимости от знаков собственных чисел λ_1, λ_2 матрицы A для различных связанных с узлом l ветвей $m = 1, \dots, M$ возможны различные ситуации. При $\lambda_1 > 0, \lambda_2 < 0$, то есть при $|v_{lk}| < c_{lk}$ (закритическое движение), в дополнение к условиям совместности (24) (или (25), (26)) вдоль направленных внутрь области интегрирования характеристик (23) в качестве граничных условий использовались соотношения

$$\psi_{lm}(t, p_l, V_{lm}) = p_{lm} - p_l - R_{lm} \rho_{lm} v_{lm} = 0. \quad (36)$$

Здесь $R_{lm} \geq 0$ — коэффициент сопротивления окончания m -й ветви, входящей в l узел (или выходящей из него), $p_{lm} = p(\rho_{lm})$ — «давление» в примыкающем к узлу l конце ветви m , связанное с плотностью ρ_{lm} в этой точке уравнением состояния (2) (или (13), (14)). С помощью коэффициентов $R_{lm}(t, \rho_{lm}, v_{lm})$, в частности, можно воспроизводить правила проезда нерегулируемых перекрестков, работу светофоров и т.п.

Помимо аналогов закона Ома (36) или аналогичной формулы Пуазейля для течений в трубах, для расчета внутренних узлов графа используется равенство нулю алгебраической суммы потоков автомобилей:

$$\varphi_l(t, V_{l1}, \dots, V_{lM}) = \sum_{m=1}^M \beta_{lm} \rho_{lm} v_{lm} = 0. \quad (37)$$

В (37) коэффициенты $\beta_{lm} = 1$ для входящих в узел ветвей и $\beta_{lm} = -1$ для выходящих из узла ветвей.

Таким образом, M дифференциальных условий совместности (24) (или ((25), (26))) и $M + 1$ нелинейных алгебраических граничных условий (36), (37) позволяют (после их аппроксимации разностными соотношениями) найти $2M + 1$ искомым параметров $V_{l1} = \{\rho_{l1}, v_{l1}\}, \dots, V_{lM} = \{\rho_{lM}, v_{lM}\}, p_l(t)$. Для решения получившейся нелинейной системы уравнений можно воспользоваться каким-либо итерационным методом. Например, аппроксимируя (24) (или ((25), (26))) по какой-либо явной схеме:

$$\begin{aligned} \varphi_{lm}(V_{lm}) &= \omega_{ilm}^n (V_{lm}^{n+1} - V_{km}^n) = c_{lm}^n (\rho_{lm}^{n+1} - \rho_{km}^n) \pm \\ &\pm \rho_{lm}^n (v_{lm}^{n+1} - v_{km}^n) = 0, \quad m = 1, \dots, M. \end{aligned} \quad (38)$$

В (38) V_{km}^n является значением сеточной вектор-функции в точке пересечения характеристики (23), приходящей в рассчитываемую точку временного слоя $t = t^{n+1}$, со слоем $t = t^n$ для k -го ребра. Эта величина определяется интерполяцией по данным вектор-функции V_k^n в узлах сетки на слое $t = t^n$. В целом, вместо (23) предпочтительнее использовать разностную аппроксимацию соответствующего инварианта Римана (25), (26), обеспечивающую более высокую точность расчетов.

Обозначив вектор неизвестных как

$$Y_l = \{\rho_{l1}, v_{l1}, \dots, \rho_{lM}, v_{lM}, p_l\},$$

получаем систему нелинейных уравнений:

$$Z_l(Y_l) = \{\varphi_l(V_{l1}, \dots, V_{lM}), \varphi_{l1}(V_{l1}), \dots, \varphi_{lM}(V_{lM}), \psi_{l1}(p_l, V_{l1}), \psi_{lM}(p_l, V_{lM})\}^T = 0 \quad (39)$$

и, задав некоторое начальное приближение Y_l^0 (по данным предыдущего шага по времени $t = t^n$) для последующих итераций в методе Ньютона из (36)–(38), имеем линейную относительно Y_l^{j+1} систему из $2M + 1$ уравнений:

$$Y_l^{j+1} = Y_l^j - \left(\frac{\partial Z(Y_l^j)}{\partial Y_l^j} \right)^{-1} Z(Y_l^j), \quad j = 0, 1, \dots, J. \quad (40)$$

В данном случае имеем

$$\begin{aligned} \left\{ \frac{\partial \varphi_{lm}}{\partial \rho_{lm}}, \frac{\partial \varphi_{lm}}{\partial v_{lm}}, \frac{\partial \varphi_{lm}}{\partial p_l} \right\} &= \{c_{lm}, \pm \rho_{lm}, 0\}, \\ \left\{ \frac{\partial \psi_{lm}}{\partial \rho_{lm}}, \frac{\partial \psi_{lm}}{\partial v_{lm}}, \frac{\partial \psi_{lm}}{\partial p_l} \right\} &= \\ &= \left\{ \frac{\partial \rho_{lm}}{\partial \rho_{lm}} - R_{lm} v_{lm}, -R_{lm} \rho_{lm}, -1 \right\} = \\ &= \{c_{lm}^2 - R_{lm} v_{lm}, -R_{lm} \rho_{lm}, -1\}, \\ \left\{ \frac{\partial \varphi_l}{\partial \rho_{lm}}, \frac{\partial \varphi_l}{\partial v_{lm}}, \frac{\partial \varphi_l}{\partial p_l} \right\} &= \{\beta_{lm} v_{lm}, \beta_{lm} \rho_{lm}, 0\}, \end{aligned}$$

и непосредственной проверкой можно убедиться, что для нелинейной системы (36)–(38) её матрица Якоби $\partial Z(Y_l^j)/\partial Y_l^j$:

$$\frac{\partial Z(Y_l^j)}{\partial Y_l^j} = \begin{pmatrix} c_{l1} & \pm \rho_{l1} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & c_{l2} & \pm \rho_{l2} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & c_{lM} & \pm \rho_{lM} & 0 \\ (c_{l1}^2 - R_{l1} v_{l1}) & -R_{l1} \rho_{l1} & 0 & 0 & \dots & 0 & 0 & -1 \\ 0 & 0 & (c_{l2}^2 - R_{l2} v_{l2}) & -R_{l2} \rho_{l2} & \dots & 0 & 0 & -1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & (c_{lM}^2 - R_{lM} v_{lM}) & -R_{lM} \rho_{lM} & -1 \\ \beta_{l1} v_{l1} & \beta_{l1} \rho_{l1} & \beta_{l2} v_{l2} & \beta_{l2} \rho_{l2} & \dots & \beta_{lM} v_{lM} & \beta_{lM} \rho_{lM} & 0 \end{pmatrix}$$

является неособенной, поскольку

$$\det(\partial Z(Y_l^j)/\partial Y_l^j) \neq 0.$$

В том случае, когда в ближайшем к узлу l конце $x_m = X_m$ какой-либо из входящих в этот узел ветвей m поток автомобилей является докритическим («сверхзвуковым»), то есть при $v_{lm} > c_{lm}$ ($\lambda_1 > 0, \lambda_2 > 0$) искомые параметры $V_{lm} = \{\rho_{lm}, v_{lm}\}$ в таких окончаниях ветви m могут быть рассчитаны так же, как и в (36)–(38), однако условия совместности (24) следует заменить разностной аппроксимацией ударной адиабаты (19):

$$\begin{aligned} \varphi_{lm}(v_{lm}) &= (p_{lm}^n(\rho_{lm}^n) - p_{lm}^{n+1}(\rho_{lm}^{n+1})) \times \\ &\times (1/\rho_{lm}^{n+1} - 1/\rho_{lm}^n) - (v_{lm}^n - v_{lm}^{n+1})^2 = 0, \quad (41) \end{aligned}$$

так как при приближении к точкам ветвления обычно происходят многочисленные затормаживающие поток междуполосные перестроения и переход движения в закритический режим движения. Аналогичным образом следует поступать при $v_{lm} > c_{lm}$ ($\lambda_1 > 0, \lambda_2 > 0$), $x_m = 0$ для выходящей из узла ветви, так как при интенсивном движении маловероятным является докритический режим движения непосредственно на выезде из перекрестка (во внутренних узлах графа). В этом случае использование (41) неявно предполагает допустимость разрывных волн ускорения («ударных волн разрежения»), что требует специальных исследований. Впрочем, возникновение таких ситуаций на перекрестках (узлах графа) с интенсивным движением маловероятно. Это замечание (об использовании (41) вместо условий совместности (24)) относится также к условиям замыкания граничных условий на входах в уличную сеть.

III. Численная реализация модели и результаты расчетов

Трудность реализации предлагаемой модели связана прежде всего со сложной структурой графа уличной сети, значительной неопределенностью определяющих параметров, констант «уравнения состояния» и процессов регулирования (многочисленность задач), поэтому полномасштабная их реализация требует значительных вычислительных ресурсов и усилий по идентификации определяющих параметров («калибровка модели»). Структура искомого решения в таких системах может иметь очень сложный характер из-за взаимодействия многочисленных волн разгона — торможения друг с другом и с узлами графа, в том числе иметь разрывной характер. Это требует специальных исследований и предъявляет очень жесткие требования к выбору численных методов, пригодных для решения данного класса задач.

Составляющая основу данной модели система уравнений в частных производных (1), (2) име-

ет гиперболический тип и для численного решения таких уравнений существует достаточно много разнообразных по своим качествам численных методов. Однако отмеченные выше особенности потребовали специальных исследований и разработки новых монотонных схем высокого порядка аппроксимации ([17] и др.).

Описанная выше математическая модель была реализована в виде комплекса программ с использованием указанных численных методов и протестирована на ряде задач. Эти тестовые расчеты показали высокую эффективность методов и работоспособность модели.

Определяющими функционирование описанной выше динамической модели уличного движения параметрами прежде всего являются характеристики графа уличной сети и уравнения состояния (2). Для оценки влияния неопределенности в их выборе (по имеющимся в литературе экспериментальным данным) были проведены тестовые расчеты автомобильного движения, примерно соответствующего условиям экспериментальных измерений [13, 14] с использованием уравнений состояния (13), (14) с константами из (29)–(32) (кривые a – d на рис. 2). Точное воспроизведение представленных в [13, 14] условий автодвижения принципиально невозможно из-за наличия на исследуемых участках пересечений с поперечными дорогами, по которым не приведено никаких данных, хотя их влияние на измеряемые параметры очевидно и может быть достаточно существенным, так как в местах пересечений обычно интенсивно происходят межполосные перестроения, что приводит к замедлению потоков, их уплотнению и даже к образованию заторов. Тем не менее по представленным в этих работах данным на входе (без учета участников движения с поперечных автодорог) были проведены расчеты и получены данные, качественно сопоставимые с экспериментальными данными на выходном измерительном пункте.

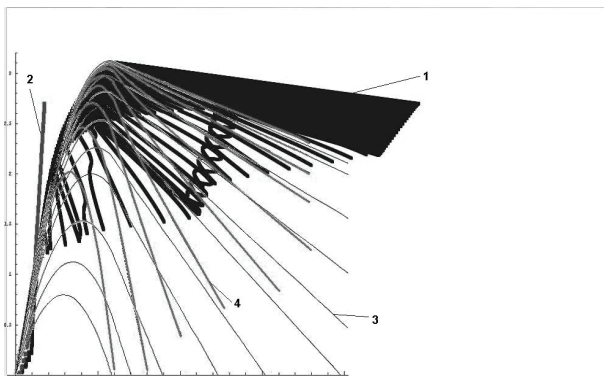


Рис. 9. Черными точками 1 показаны зависимости $q(t, x_k)$ от $\rho(t, x_k)$, полученные в различные моменты времени t . Темно-серыми точками 2 показаны те же зависимости для участка дороги, на котором реализуется свободный («сверхзвуковой») поток. Тонкие светло-серые линии 3 — характеристики первого семейства, темно-серые 4 — ударные адиабаты

В одном из серии расчетов планарный дорожный граф состоял из двух участков, на входе первого участка задавался поток как возрастающая функция времени (33) (в соответствии с экспериментальными данными), а на выходе второго участка в качестве граничных условий задавались нулевые производные от плотности и скорости (34) (неотражающие условия, так как движение на втором участке соответствовало режиму «свободного потока»). Во внутреннем узле графа одной из констант расчетной модели является коэффициент сопротивления $R_{lm} \geq 0$ (36), который варьировался в довольно широких пределах: от нуля (когда такая граничная точка ничем не отличается от внутренних узлов разностной сетки) до очень больших его значений (соответствующих полному запертию потока в этом месте).

На рис. 9 для одного из вариантов расчета в фазовой плоскости $\{q, \rho\}$ с координатами x_k черными точками нанесены значения зависимостей $q(t, x_k)$ от $\rho(t, x_k)$, полученные в узлах разностной сетки первого участка дороги с координатами x_k в различные моменты времени t , то есть параметры, которые нанесены в виде экспериментальных точек на рис. 4. Темно-серыми точками (постепенно сливающимися в линию) нанесены аналогичные данные для второго участка, на котором реализуется свободный («сверхзвуковой») поток. Тонкими черными линиями нанесены те же, что и на рис. 4, характеристики первого семейства, а более толстыми светло-серыми линиями — ударные адиабаты для использовавшегося в расчетах уравнения состояния с константами из (30), поскольку сеточных узлов было 100, а моментов времени несколько тысяч, то различить можно лишь отдельные траектории в фазовой плоскости (самые толстые — черные линии). Однако четко просматриваются движения потока как близкие характеристическим направлениям (с «безударным» режимом торможения), так и движения с переходом от докритического («сверхзвукового») режима к закритическому («дозвуковому») вдоль ударных адиабат. Для рассматриваемого здесь примера зависимость плотности автомобилей (черная кривая № 1), отнесенной к ее максимально допустимым значениям (пропускной способности первого участка), от координаты x_k , а также аналогичные распределения скорости (светло-серая кривая № 2) и потока (темно-серая кривая № 3) в один из моментов времени показаны на рис. 10. Эти два рисунка отчетливо демонстрируют, что закритический («синхронизированный») режим движения возникает в рассматриваемой здесь сугубо детерминистской модели в результате взаимодействия многочисленных волн разгона–торможения, возникающих и распространяющихся на участках дороги между локальными, замедляющими поток источниками (сужения дороги, пересечения с поперечными дорогами, различные замедляющие поток

маневры, неоднородность состава автотранспортных средств и стилей вождения и т.д.).

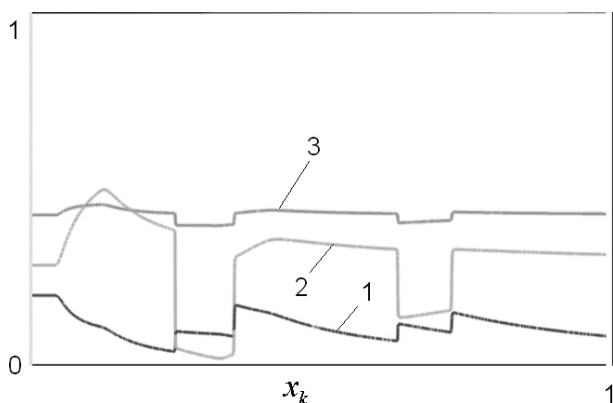


Рис. 10. Зависимости величин, отнесенные к максимальным значениям, от координаты x_k : плотности автомобилей (черная кривая № 1); скорости (светло-серая кривая № 2); потока (темно-серая кривая № 3)

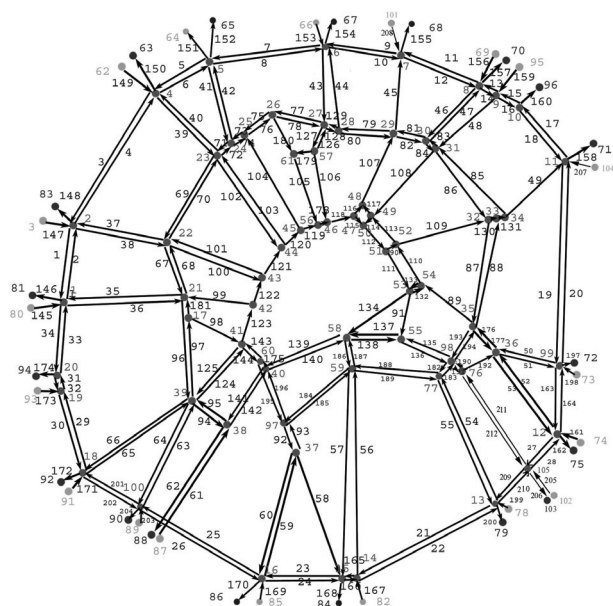


Рис. 11. Граф Садового кольца г. Москвы: 105 узлов; 212 ребер

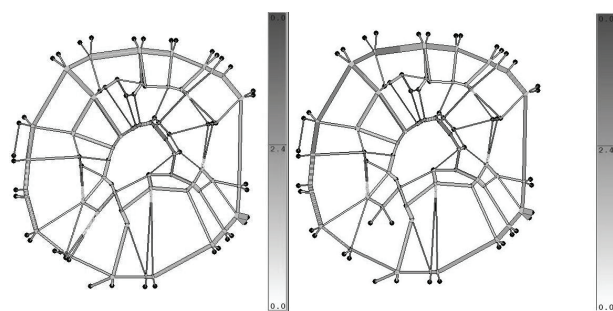


Рис. 12. Результаты рассчитанных значений потоков [АТС / с] на Садовом кольце г. Москвы: слева для двустороннего движения; справа для одностороннего

Для проверки работоспособности разработанной вычислительной модели были сделаны численные расчеты поведения транспортной системы города Москвы внутри графа Садового кольца, имеющего 105 узлов и 212 ребер и представленного на рис. 11. В начальный момент времени

на входах в Садовое кольцо задавались постоянные значения потоков (33), полученные из данных экспериментальных замеров, сделанных в 2005 г. в центральной части г. Москвы, на выходах задавались неотражающие граничные условия (34). В результате расчетов были получены стационарные распределения потоков и скоростей движения автомобилей за 3-часовой интервал с момента начала движения. Затем граф данной транспортной системы перестраивался так, чтобы движение по Садовому кольцу становилось односторонним против часовой стрелки, и расчет проводился заново, стартуя с аналогичных исходных данных по значениям транспортных потоков на входах в транспортную сеть. В результате был получен эффект увеличения пропускной способности транспортной сети на 30%, что может быть вполне объяснимо, если исходить из того, что ширина дорог выросла в два раза при том, что средний пробег между любыми двумя точками увеличился меньше, чем вдвое из-за хорошей связности графа транспортной сети внутри Садового кольца. Результаты рассчитанных значений потоков [АТС / с] приведены на рис. 12 слева для двустороннего движения и справа для одностороннего.

Довольно подробный обзор по гидродинамическим моделям транспортных потоков имеется, например, в главах 2, 3 учебного пособия [18].

Работа поддержана грантами РФФИ № 10-07-00620-а, 08-01-00959-а, 08-07-00501-а, 08-07-00158-а, 10-07-00620-а, 11-01-00494-а, РГНФ № 08-02-00347, ПФИ ОМН РАН № 3, ПФИ Президиум РАН П-2. Работа проведена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы.

Литература

1. Kholodov A.S., Kholodov Y.A. Computational models on graphs for the nonlinear hyperbolic system of equations // Proceedings of ASME 2004PVP Conference. — 2004. — V. 476. — P. 161–167.
2. Kholodov Y. A., Kholodov A.S., Kovshov N.V., Simakov S.S., Severov D.S., Bordonos A.K., Baraev A.Z. Computational models on graphs for nonlinear hyperbolic and parabolic systems of equations // Proceedings of the III European Conference on Computational Mechanics. — 2006. — P. 2279.1–2279.19.
3. Морозов И.И., Холодов Я.А. Моделирование динамики транспортных потоков // Труды 51-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». — 2008. — Т. 2. — С. 128–129.
4. Уизем Дж. Линейные и нелинейные волны. — М.: Мир, 1977.
5. Payne H.J. Models of freeway traffic and control. Mathematical Models of Public Systems // Simulation Council Proc. 28. — 1971. — V. 1. — P. 51–61.

6. *Helbing D.* Improved fluid: dynamic model for vehicular traffic // *Phys. Rev. E.* — 1995. — V. 51. — P. 3163–3169.
7. *Aw A., Rascle M.* Resurrection of «second order» models of traffic flow // *SIAM Journal of Applied Mathematics.* — 2000. — V. 60. — P. 916–938.
8. *Greenberg J. M.* Extensions and amplifications of a traffic model of Aw and Rascle // *SIAM J. Appl. Math.* — 2001. — V. 62, N. 3. — P. 729–745.
9. *Zhang H.M.* A non-equilibrium traffic model devoid of gas-like behavior // *Transp. Res. B.* — 2002. — V. 36. — P. 275–290.
10. *Siebel F., Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // e-print arXiv: physics /0509124 v2, 2006.
11. *Kerner B.S.* Introduction to modern traffic flow theory and control. The long road to three – phase traffic theory. — Springer, 2009.
12. *Kerner B.S. and H. Rehborn.* Experimental features and characteristics of traffic jams // *Phys. Rev. E.* — 1996. — V. 53, N. 2.
13. *Kerner B.S. and H. Rehborn.* Experimental properties of complexity in traffic flow // *Phys. Rev. E.* — 1996. — V. 53, N. 5.
14. *Kerner B.S. and H. Rehborn.* Experimental Properties of Phase Transitions in Traffic Flow // *Phys. Rev. Let.* — 1997. — V. 79, N. 20. — P. 4030–4033.
15. *Kerner B.S.* Experimental Features of Self-Organization in Traffic Flow // *Phys. Rev. Let.* — 1998. — V. 81, N. 20.
16. *Годунов С.К., Роменский Е.И.* Элементы механики сплошных сред и законы сохранения. — М.: Наука, 1998.
17. *Холодов А.С., Холодов Я.А.* О критериях монотонности разностных схем для уравнений гиперболического типа. // *Журнал выч. математики и мат. физики.* — 2006. — Т. 46, № 9. — С. 1560–1588.
18. *Гасников А.В., Кленов С.Л., Нурминский Е.А., Холодов Я.А., Шамрай Н.Б.* Введение в математическое моделирование транспортных потоков; учебное пособие. С приложениями Бланка М.Л., Гасниковой Е.В., Замятина А.А. и Мальшева В.А., Колесникова А.В., Райгородского А.М. — М.: МФТИ, 2010.

Поступила в редакцию 15.10.2010.

Б.Н. Четверушкин, М.А. Трапезникова, И.Р. Фурманов, Н.Г. Чурбанова

Институт прикладной математики им. М.В. Келдыша РАН

Макро- и микроскопические модели для описания движения автотранспорта на многополосных магистралях

Работа посвящена математическому моделированию автотранспортных потоков на улицах больших городов и магистралях. Разработаны оригинальные двумерные макроскопическая и микроскопическая модели для описания автомобильного движения на многополосных трассах с учётом их реальной геометрии. Макроскопическая модель синхронизованных потоков основана на кинетическом подходе по аналогии с квазигазодинамической системой уравнений. Микроскопическая модель использует теорию клеточных автоматов. Обе модели обобщены на случай «многофазного» движения для описания неоднородных потоков транспорта. Под фазой понимается группа автомобилей, объединённых общими характеристиками. Тестовые расчёты демонстрируют адекватность предложенных моделей различным дорожным ситуациям.

Ключевые слова: автотранспортные потоки, микроскопические и макроскопические модели, многополосное движение, многофазные потоки транспорта.

I. Введение

В процессе развития теории движения автотранспорта сформировалось два основных подхода к проблеме математического моделирования автотранспортных потоков [1].

Первый из них — макроскопический [2] — используется для описания движения автотранспорта на больших расстояниях, намного больших, чем размеры самих автомобилей и при этом плотность транспортного потока достаточно велика (так называемое «синхронизованное» движение). В этих условиях все водители вынуждены придерживаться одинаковых стратегий поведения и подчиняться общим правилам и закономерностям. В такой ситуации справедливо приближение «сплошной среды» и автотранспортные потоки могут рассматриваться как движение слабосжимаемой жидкости. В данной работе для вывода системы уравнений, описывающих движение транспорта, используется кинетический подход по аналогии с квазигазодинамической системой уравнений [3].

Второй подход — микроскопический [4] — рассматривает движение автомобилей как движение отдельных частиц и является наиболее эффективным при описании движения на небольших расстояниях, например, на перекрёстках. Каждый автомобиль характеризуется своей текущей скоростью, может иметь своё собственное ограничение по скорости и может иметь свою собственную цель, которые для разных автомобилей могут быть различными. Движение автомобилей происходит по законам, общим для всех участников движения, при этом должна быть обеспечена безопасность движения.

Подавляющее большинство существующих как макро-, так и микромоделей посвящено описа-

нию однополосного движения и позволяет исследовать только движение по однополосным трассам. В этом случае влияние соседних полос можно учитывать, вводя источники и стоки в правых частях уравнений. Однако такой подход слишком громоздок, а в случае четырёх и более полос вообще труднореализуем. В [5–6] была предложена двумерная, многополосная макроскопическая модель транспортных потоков, использующая реальную геометрию дороги и использующая идеологию, положенную в основу квазигазодинамических уравнений. В данной работе предложенная модель получила дальнейшее развитие.

Однако следует отметить, что при исследовании движения на малых дистанциях порядка одного километра приближение «сплошной среды» не может быть использовано, так как в этом случае следует принимать во внимание возможные различия в характеристиках каждого отдельного автомобиля. Для более детального описания таких автомобильных потоков в данной работе предлагается новая двумерная микроскопическая модель, использующая теорию клеточных автоматов. Кроме того, обе модели (макро- и микроскопическая) обобщаются на случай «многофазных» потоков, когда поток неоднороден и состоит из групп автомобилей, объединённых некоторыми общими параметрами, такими как общее ограничение скорости, общая конечная цель и т.д.

II. Макроскопическая модель

Макроскопические модели движения автотранспорта описывают потоки, скорость которых далека от скорости свободного движения. В этих условиях, по аналогии с газовой динамикой, могут быть введены некоторые усреднённые характеристики движения. Такими характеристиками являются, например, плотность потока ρ как чис-

ло машин на полосу на единицу длины, скорость u как средняя скорость автомобилей, а также поток $q = \rho u$ как функция плотности и скорости.

Система уравнений для описания автомобильного движения отличается от систем уравнений газовой динамики тем, что в ней присутствуют члены, возникающие благодаря наличию так называемого «человеческого фактора». Это, например, ускоряющая/замедляющая сила:

$$f - a\rho, \quad (1)$$

ускорение:

$$a = \frac{u_{eq} - u}{T}, \quad (2)$$

равновесная скорость:

$$u_{eq} = u_f \left(1 - \frac{\rho}{\rho_{jam}} \right), \quad (3)$$

здесь u_f — это скорость свободного движения автомобилей, ρ_{jam} — плотность, при которой возникают автомобильные «пробки», а время релаксации (реакции водителя на дорожную ситуацию):

$$T = t_0 \left(1 + \frac{r\rho}{\rho_{jam} - r\rho} \right), \quad (4)$$

t_0 и r — феноменологические константы, скоростное ограничение: $0 \leq u \leq u_{max}$, где u_{max} — максимальная разрешённая скорость.

Вводится аналог давления как реакция водителя на ситуацию впереди автомобиля:

$$P = \lambda_x \frac{\rho^{\beta_x}}{\beta_x}, \quad (5)$$

где λ_x и β_x — феноменологические константы.

Простое обобщение одномерной модели на двумерный случай невозможно из-за неэквивалентности движения вдоль и поперёк дороги. В работе [6] для описания перехода машины из одной полосы в другую было введено понятие «боковой» скорости. Таким образом, машины могут перестраиваться в полосу с большей скоростью или с меньшей плотностью, или если водитель хочет достичь определённой цели. Боковая скорость состоит из следующих трёх слагаемых:

перестроение на полосу с большей скоростью:
 $v_u = k_u \cdot \rho \frac{\partial u}{\partial y}$;

перестроение на полосу с меньшей плотностью: $v_\rho = -k_\rho \cdot u \frac{\partial \rho}{\partial y}$;

движение к определённой цели:

$$v_{des} = k_{des} \cdot \frac{u^2}{(x_{des} - x)^2} (y_{des} - y).$$

Здесь k_u , k_ρ , k_{des} — феноменологические константы, (x_{des}, y_{des}) — координаты желаемой цели.

В конечном виде двумерная система уравнений, описывающая автотранспортные потоки, была получена по аналогии с квазигазодинамической (КГД) системой уравнений [3]. Одним из критериев газодинамических течений является число Кнудсена — отношение характерной длины газовой среды (длина свободного пробега молекул) к

характерной длине течения. В газовой динамике $\text{Kn} \leq 10^{-3}$. Для автотранспортных потоков в качестве длины свободного пробега можно рассматривать среднее расстояние между автомобилями, следовательно, в этом случае $\text{Kn} \approx 0,1$. КГД система хорошо зарекомендовала себя в широком диапазоне чисел Кнудсена, в том числе и в сторону его увеличения, что явилось причиной использования её для вывода системы уравнений транспортных потоков.

По аналогии с КГД системой вводятся дополнительные потоки массы, обеспечивающие сглаживание решения на расстоянии, равном характерной длине рассматриваемой среды. Для транспортных потоков характерные длины вдоль и поперёк движения не равны друг другу. Вдоль дороги это среднее расстояние $\delta(u)$ между автомобилями, зависящее от скорости u , а поперёк дороги характерная длина — это ширина одной полосы.

В КГД системе уравнений вводится также понятие характерного времени (время преодоления молекулой длины свободного пробега). Для автотранспортных потоков можно также ввести характерное время как пересечение некоторой точки несколькими автомобилями:

$$\tau_x \approx \frac{\delta(u)}{u}, \quad \tau_y \approx \frac{1}{v}. \quad (6)$$

Для простоты τ_x и τ_y можно считать неизменными, так как при синхронизованном движении они мало зависят от скорости.

В правой части уравнения неразрывности вводятся дополнительные потоки W_x и W_y , сглаживающие решение вдоль и поперёк магистрали:

$$W_x = \frac{\tau_x}{2} \frac{\partial}{\partial x} (\rho u^2 + P), \quad (7)$$

$$W_y = \frac{\tau_y}{2} \left(\frac{\partial \rho v^2}{\partial y} + \lambda_y \rho^{\beta_y} \frac{\partial \rho}{\partial y} \right). \quad (8)$$

Аналогичные потоки вводятся и в уравнение сохранения импульса.

Учитывая (1)–(8) и обобщая всё вышесказанное, была получена следующая система уравнений, описывающая движение транспортных потоков:

$$\begin{aligned} & \frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = \\ & = \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} \frac{\partial}{\partial x} (\rho u^2 + P) \right) + \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} \frac{\partial}{\partial y} (\rho u v) \right) + \\ & + \frac{\partial}{\partial y} \left(\frac{\tau_y}{2} \left(\frac{\partial}{\partial y} (\rho v^2) + \lambda_y \rho^{\beta_y} \frac{\partial \rho}{\partial y} \right) \right) + \\ & + \frac{\partial}{\partial y} \left(\frac{\tau_y}{2} \frac{\partial}{\partial x} (\rho u v) \right) - \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} f \right), \quad (9) \\ & \frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^2}{\partial x} + \frac{\partial \rho u v}{\partial y} = \\ & = f - \frac{\partial}{\partial x} P + \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} \frac{\partial}{\partial x} (\rho u^3 + 3P u) \right) + \\ & + \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} \frac{\partial}{\partial y} (\rho u^2 v) \right) + \end{aligned}$$

$$\begin{aligned}
 & + \frac{\partial}{\partial y} \left(\frac{\tau_y}{2} \left(\frac{\partial}{\partial y} (\rho u v^2) + \lambda_y \rho^{\beta_y} \frac{\partial \rho u}{\partial y} \right) \right) + \\
 & + \frac{\partial}{\partial y} \left(\frac{\tau_y}{2} \frac{\partial}{\partial x} (\rho u^2 v) \right) - \frac{\partial}{\partial x} \left(\frac{\tau_x}{2} f u \right), \quad (10)
 \end{aligned}$$

$$v = k_u \rho \frac{\partial u}{\partial y} - k_\rho u \frac{\partial \rho}{\partial y} + k_{des} \frac{u^2}{(x_{des} - x)^2} (y_{des} - y). \quad (11)$$

В модели используется большое количество констант и параметров, которые могут быть определены эмпирическим путём из статистических или экспериментальных данных. В расчётах, приведённых ниже, использовались следующие значения:

$$\lambda_x = 60 \frac{\text{км}^2}{\text{ч}^2}, \quad \lambda_y = 4 \frac{\text{км}^2}{\text{ч}^2}, \quad \beta_x = 2, \quad \beta_y = 0,$$

$$\tau_x = 2 \cdot 10^{-3} \text{ ч}, \quad \tau_y = 3 \cdot 10^{-4} \text{ ч},$$

$$t_0 = 50 \text{ с}, \quad r = 0,95, \quad \rho_{jam} = 120 \frac{\text{авт}}{\text{км} \cdot \text{полоса}},$$

$$u_f = 90 \frac{\text{км}}{\text{ч}}, \quad u_{max} = 90 \frac{\text{км}}{\text{ч}}.$$

III. Микроскопическая модель

При построении микроскопической модели, способной адекватно описывать движение автотранспорта на небольших расстояниях, использовалась теория клеточных автоматов. Клеточные автоматы — это идеализация физической системы, в которой пространство и время предполагаются дискретными и каждая из взаимодействующих единиц имеет конечное число дискретных состояний. Для описания динамики транспортных потоков методы клеточных автоматов развиваются с 80-х годов XX века. В данном подходе трасса представляется в виде одномерной решетки. Каждая ячейка этой решетки может содержать или не содержать одну и только одну частицу, которая в данной модели означает машину. Частицы перескакивают с одной ячейки на другую (свободную) обычно в одном направлении. В случае однополосного движения они не могут обгонять друг друга. Так как время и пространство дискретны, скорость и ускорение также принимают только дискретные значения. Передвижение частиц в таких моделях регулируется правилами обновления состояний ячеек, которые обычно содержат вероятностные характеристики. Правила одинаковы для всех ячеек и применяются ко всем ячейкам параллельно, то есть за один шаг по времени меняются все ячейки.

Обычно новое состояние ячейки — функция от состояния ячейки только на предыдущем временном шаге, однако иногда необходимо принимать во внимание и состояние некоторых ячеек в другие моменты времени. Правила клеточных автоматов обладают свойством локальности, поэтому

для получения текущего состояния ячейки требуется знать только состояния некоторых ее соседей. Эти соседи называются окрестностью ячейки. Обычно окрестность ячейки невелика и состоит из нескольких соседних ячеек (от 4 до 8 соседних ячеек в двумерном случае), иначе правила становятся слишком сложными.

Длина одной ячейки равна длине участка дороги, который занимает одна машина, когда стоит в пробке, то есть длине самой машины и расстоянию между соседними машинами. Обычно используется длина ячейки в 7.5 метров. Скорость машины показывает, сколько ячеек она преодолевает за один временной шаг. Длина ячейки, максимальная скорость и временной шаг полностью описывают простейшую модель.

Одной из наиболее известных одномерных моделей микроскопического типа является модель Нагеля–Шрекенберга [4]. В этой модели для каждой машины заданы скорость от 0 до некоторой величины V_{max} , величина d — расстояние между текущей машиной и следующей перед ней, V_n — скорость текущей машины с номером n . Правило обновления состоит из следующих подшагов:

1. *Ускорение.* Если $V_n < V_{max}$, то скорость n -го автомобиля увеличивается на единицу, если $V_n = V_{max}$, то скорость не изменяется.
2. *Торможение.* Если $d < V_{n-1}$, то скорость n -го автомобиля уменьшается до $d - 1$.
3. *Случайные возмущения.* Если $V_n > 0$, то скорость n -го автомобиля может быть уменьшена на единицу с вероятностью p , скорость не изменяется, если $V_n = 0$.
4. *Движение.* Каждый автомобиль продвигается вперед на количество ячеек, соответствующее его новой скорости после выполнения предыдущих шагов.

Первый шаг отражает общее стремление всех водителей ехать как можно быстрее. Второй гарантирует отсутствие столкновений с впереди идущими автомобилями. Элемент стохастичности, учитывающий случайности в поведении водителей, вносится на третьем шаге.

Модель Нагеля–Шрекенберга воспроизводит только основные особенности реального потока. В данной работе эта модель была взята за основу для описания движения на относительно коротких участках и получила своё дальнейшее развитие путём усложнения логики поведения водителей и внесения дополнительной характеристики автомобиля — желаемая конечная цель.

Для моделирования многополосного движения приведенная выше модель была обобщена на двумерный случай. Трасса в этом случае представляет собой двумерную решетку, где количество ячеек в поперечном направлении соответствует числу полос трассы. В такой модели разрешены пе-

рестроения машин из полосы в полосу и обгоны. Процесс обновления состояний ячеек делится на два подшага:

1. Для каждой машины выясняется возможность и необходимость смены полосы. Производится смена полосы. Этот подшаг выполняется параллельно для всех машин.
2. Производится движение вперед по каждой полосе по правилам однополосного движения.

Смена полос должна происходить за один временной шаг. Если в одном направлении существуют больше, чем две полосы, то может возникнуть конфликт, когда две машины с крайних полос желают сместиться в среднюю и занять одну и ту же ячейку. Такой конфликт легко преодолеть, если разрешить перестроение вправо только на четных шагах, а влево — только на нечетных.

Для смены полос существует несколько причин: на соседней полосе выше скорость движения либо меньше плотность, перестроение на соседнюю полосу необходимо для успешного достижения цели движения. С другой стороны, перед сменой машины полосы необходимо проверить, выполнены ли условия безопасности. В целом условия для смены полосы выглядят следующим образом:

1. Машина находится в зоне, где разрешена смена полосы.
 - (а) Смена полосы ведет к увеличению скорости (снижению плотности) или необходима для достижения цели.
2. Ячейка, на которую перестраивается машина, свободна.
 - (а) Условие безопасности: расстояние на целевой полосе позади больше или равно V_{\max} , впереди больше или равно V машины, и при этом смена полосы происходит с некоторой вероятностью.

Алгоритм, предлагаемый в данной работе, обеспечивает возможность достижения цели. Целями автомобилей при многополосном движении могут быть, например, съезд с дороги или поворот в определенную сторону на светофоре. В обоих случаях машины, начиная с определенного момента времени, стремятся перестроиться в целевую полосу, игнорируя значения плотности и скорости на ней. Тем не менее водители не могут забыть про правила безопасности и вынуждены им подчиняться. Таким образом, если цель достаточно близко, машины сворачивают на целевую полосу при первом удобном случае и больше из этой полосы не уходят. Может возникнуть ситуация, когда водитель не успевает перестроиться в нужную полосу до самого достижения цели. В таких

случаях ему придется остановиться в соседней от цели полосе и ждать возможности «вклиниться» в целевую полосу. При этом он может заблокировать продвижение других машин вперед на текущей полосе.

Дистанция, начиная с которой водители пытаются намеренно перестроиться в целевой ряд, взята из общих соображений и, вообще говоря, зависит от плотности потока. В разных задачах она составляет 75–150 метров.

Таким образом, в описываемой модели для каждой машины необходимо хранить параметр цели. Цели у автомобилей обязательны и поменять их они не могут.

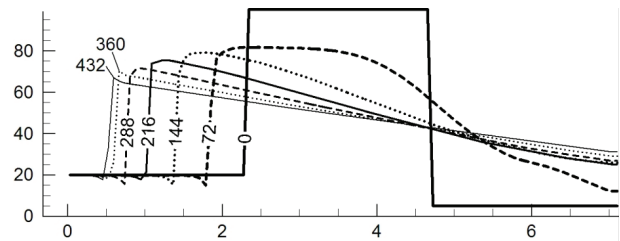


Рис. 1. Эволюция ступеньки высокой плотности — макроскопическая модель

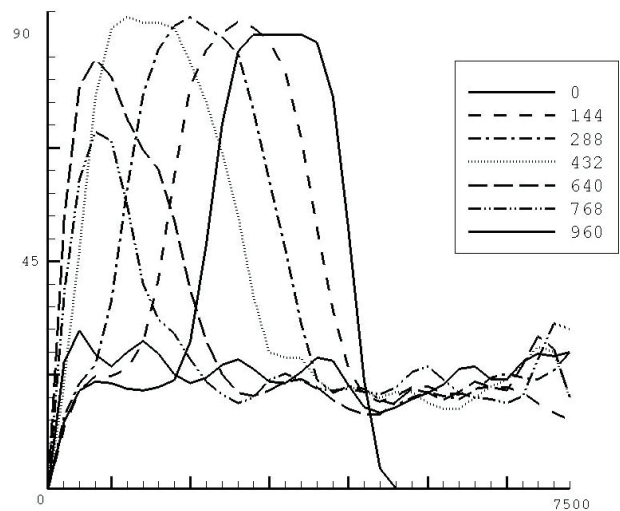


Рис. 2. Эволюция ступеньки высокой плотности — микроскопическая модель

Далее представлены результаты расчетов одномерной и двумерной тестовых задач, проведенных с использованием макроскопической и микроскопической моделей. Первая тестовая задача — одномерная — распространение скачка плотности вдоль трассы. Распределение автомобилей на трассе в начальный момент времени представляет собой ступеньку достаточно высокой плотности — порядка 80 машин на километр. На рис. 1 и 2 показан профиль плотности вдоль трассы (здесь и далее для макроскопической модели длина трассы указана в километрах, а для микроскопической модели — в метрах) на последовательные моменты времени, полученный на основе соответственно макро- и микроскопической моделей. Качественно результаты расчетов согласуются: ступенька плотности со временем «движется» назад. Это обу-

словлено тем, что поток машин в ступеньке практически достигает максимальной пропускной способности дороги. Различия в графиках связаны с разными подходами к моделированию реального поведения водителей на дороге и особенностями осреднения результатов расчетов для микроскопического метода.

Второй тест — двумерная задача о движении по трассе, имеющей локальное расширение и изобращённой на рис. 3.

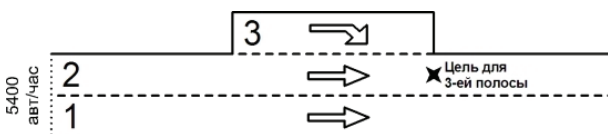


Рис. 3. Постановка задачи о движении по дороге с локальным расширением

На рис. 4, 5 показаны результаты расчётов, полученных с использованием макроскопической модели. На широком участке плотность автомобилей падает (рис. 4), а скорость возрастает (рис. 5). Однако после сужения трассы до двух полос плотность значительно возрастает, а скорость соответственно падает. Поток автомобилей на выходе с трассы становится меньше, чем на входе. В результате можно сделать вывод о том, что локальное расширение ухудшает общую ситуацию на дороге: пропускная способность трассы с локальным расширением меньше, чем была бы у трассы без расширения.

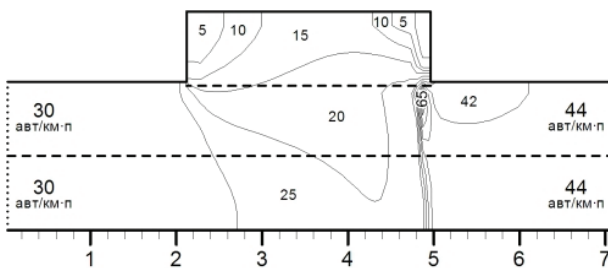


Рис. 4. Поле плотности — макроскопическая модель

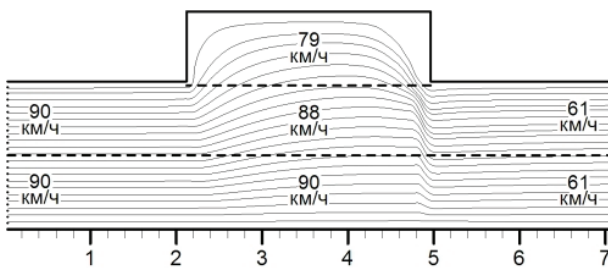


Рис. 5. Линии тока и значения скорости — макроскопическая модель

На рис. 6 приведён график осреднённой по ширине трассы плотности на разные моменты времени. Видно, что в конце расширения также наблюдается уплотнение потока.

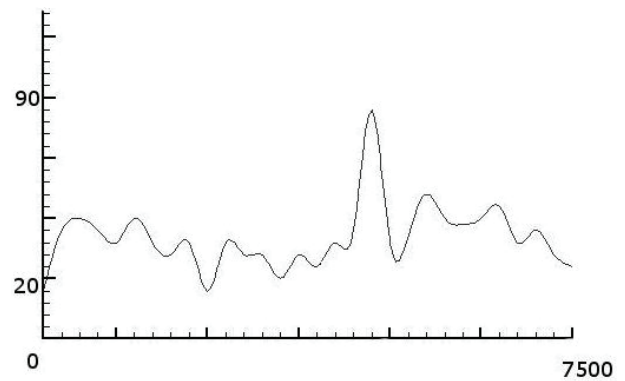


Рис. 6. Среднее значение плотности вдоль трассы — микроскопическая модель

IV. Неоднородные транспортные потоки

Приведенные выше модели описывают «однородные» транспортные потоки, то есть автомобили различаются только своими координатами. Однако в действительности на дороге присутствуют автомобили различных типов, например, грузовые и легковые. Они могут не только различаться своими характеристиками (скорость свободного движения, время релаксации потока), но и поведением на дороге. Например, грузовые автомобили могут стремиться занять правую полосу. Кроме того, транспортные средства имеют различный маршрут. Ещё большую важность представляют различные пункты назначения транспортных средств, когда изначально однородный поток сложным образом разделяется на несколько компонент.

Для описания таких сложных потоков введем понятие многофазности, под фазой будем понимать группу автомобилей, объединенных некоторыми общими характеристиками. Обе предложенные выше модели были обобщены на многофазный случай. Для макроскопической модели уравнения (9)–(11) были модифицированы с учётом различных фазовых характеристик: у каждой фазы свои плотность, скорость, боковая скорость, конечная цель и т.д. Соответственно модифицированная система уравнений содержит уравнения неразрывности, уравнения сохранения импульса и уравнения боковой скорости для каждой фазы.

В качестве тестового расчета исследовалась следующая задача. Имеются два типа автомобилей, различающиеся по своей конечной цели: фаза 1, составляющая 1/3 всего потока, — это автомобили, поворачивающие направо, и фаза 2, составляющая 2/3 потока, это автомобили, следующие прямо (рис. 7).

На рис. 8 представлены результаты расчётов с использованием макроскопической модели. Видно, что первая фаза, стремясь повернуть, на некоторое время вытесняет с первой полосы вторую фазу. Перед съездом наблюдается скачок плотности обеих фаз. После съезда первая фаза исчезает

с трассы, а автомобили второй фазы равномерно распределяются по всей ширине дороги.

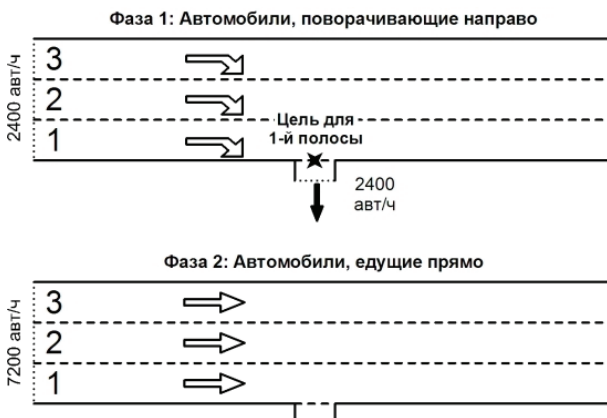


Рис. 7. Постановка задачи о двухфазном движении автомобилей

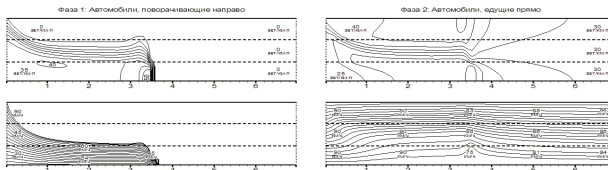


Рис. 8. Поля плотности и линии тока в задаче о двухфазном движении автомобилей — макроскопическая модель

На рис. 9 представлены результаты расчётов, полученные с использованием микроскопической модели.

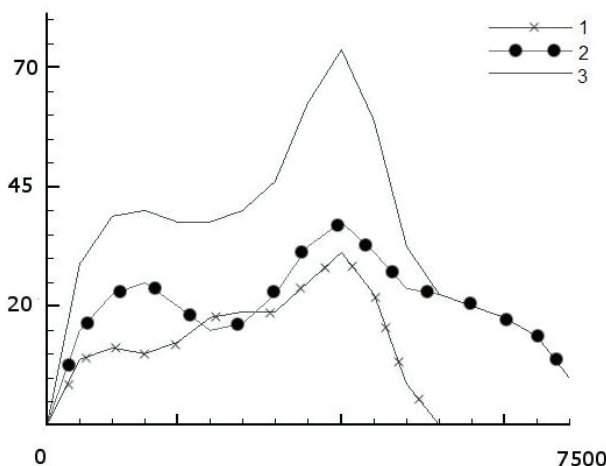


Рис. 9. Среднее значение плотности вдоль трассы в задаче о двухфазном движении автомобилей — микроскопическая модель

Здесь линией 1 изображена плотность первой фазы, линией 2 — плотность второй фазы, а линией 3 — сумма плотностей обеих фаз.

Тестовые расчёты продемонстрировали качественное совпадение результатов, полученных с использованием обеих — макро- и микроскопиче-

ской — моделей при режимах автомобильного движения, когда они обе применимы.

V. Заключение

Созданные математические модели могут быть положены в основу комплекса программ, позволяющего:

- прогнозировать возникновение заторов и пробок и давать рекомендации по способам их устранения;
- определять характер влияния геометрических параметров магистрали (сужения, расширения, изменение полосности) и режимов регулирования (светофоры) на пропускную способность транспортных систем;
- давать рекомендации по выбору оптимальной ширины трассы и расположению «карманов» вдоль дороги;
- моделировать пропускную способность сходов и въездов различного типа, оптимизировать транспортные развязки;
- изучать влияние изменения дорожных стандартов на динамику транспортных потоков.

Литература

1. Klar A., Wegener R. A hierarchy of models for multilane vehicular traffic I: Modeling // SIAM J. Appl. Math. — 1999. — V. 59. — P. 983–1001.
2. Lighthill M.H., Witham G.B. On kinematic waves: A theory of traffic flow on long crowded roads // Proc. Royal. Soc. Ser. A. — 1955. — V. 229, N. 1178. — P. 317–345.
3. Четверушкин Б.Н. Кинетические схемы и квазигазодинамическая система уравнений. — М.: МАКС Пресс, 2004.
4. Nagel K., Schreckenberg M. A Cellular automation model for freeway traffic // J. Phys. I France. — 1992. — V. 2, N. 2221.
5. Карамзин Ю.Н., Трапезникова М.А., Четверушкин Б.Н., Чурбанова Н.Г. Двумерная модель автомобильных потоков // Математическое моделирование. — 2006. — Т. 18, № 6. — С. 85–95.
6. Сугинова А.Б., Трапезникова М.А., Четверушкин Б.Н., Чурбанова Н.Г. Двумерная макроскопическая модель транспортных потоков // Математическое Моделирование. — 2009. — Т. 21, № 2. — С. 118–126.

Поступила в редакцию 15.10.2010.

В.И. Швецов

Институт системного анализа РАН

Проблемы моделирования передвижений в транспортных сетях

В работе дается краткий обзор последних направлений развития в области расчета корреспонденций. Эти направления связаны с переходом от матриц корреспонденций к рассмотрению цепочек передвижений, а также цепочек «дневной активности» (activity-based models) и применением методов Монте-Карло для моделирования корреспонденций. В работе также описана проблема, связанная с получением оценки межрайонной дальности, усредненной по способам передвижения их решения. Описан способ корректного усреднения. Далее обсуждаются различные алгоритмы поиска равновесия в транспортной сети. Принцип равновесия однозначно определяет значения потоков на дугах сети, однако оставляет большой произвол в выборе распределения корреспонденций по путям. Для устранения этой неопределенности может быть использован принцип максимизации энтропии.

Ключевые слова: транспортные потоки, межрайонные корреспонденции, модели активности, расщепление корреспонденций, равновесное распределение.

I. Введение

Основная задача комплексной математической модели транспортной сети — определение и прогноз всех параметров функционирования сети, таких как интенсивность движения на всех элементах сети, объемы перевозок в сети общественного транспорта, средние скорости движения, задержки и потери времени и т.д. Решение этой задачи основано на воспроизведении в рамках компьютерного моделирования всей структуры передвижений, осуществляемых в городе на всех видах транспорта в течение дня. Соответственно моделирование работы транспортной системы является комплексной задачей, включающей в себя большое количество связанных между собой задач.

В настоящее время «классической» считается схема моделирования транспортных потоков в сети крупного города, включающая четыре основных этапа (т.н. «4-х-стадийная схема»; см., например, [1]):

- оценка общих объемов прибытия и отправления из каждого района города (Trip generation);
- расщепление по способам передвижений, таким как пешие передвижения, передвижения с использованием общественного транспорта, передвижения на личном автомобиле и др. (Modal split);
- определение матриц корреспонденций, определяющих объем передвижений между каждой парой расчетных районов города (Trip distribution);
- распределение корреспонденций по транспортной сети, то есть определение всех пу-

тей, выбираемых участниками движения, и определение количества передвижений по каждому пути (Trip assignment).

В настоящей работе обсуждаются различные аспекты этой комплексной схемы. Основным недостатком этой схемы является то, что матрицы корреспонденций для передвижений, совершаемых с разными целями, рассчитываются независимо друг от друга. Более адекватные реальности модели корреспонденций должны рассматривать цепочки связанных друг с другом передвижений. В разделе II дается краткий обзор развития моделей корреспонденций в этом направлении. Развитие так называемых «моделей активности» (activity-based models) связано с переходом от матриц корреспонденций к рассмотрению цепочек передвижений, а также схем «дневной активности» и применением методов Монте-Карло для моделирования корреспонденций.

В разделе III описаны некоторые проблемы классической схемы, связанные с получением усредненных по видам транспорта оценок межрайонной дальности.

Основой для моделирования процесса распределения корреспонденций является принцип равновесия в транспортной сети. В разделе IV обсуждаются различные алгоритмы поиска равновесия. Принцип равновесия однозначно определяет значения потоков на дугах сети, однако оставляет большой произвол в выборе распределения корреспонденций по путям. Для устранения этой неопределенности может быть использован принцип максимизации энтропии.³²

II. Модели прогноза передвижений в транспортной сети

³²Материал этого раздела несколько более подробно изложен в отдельной статье [2].

Все многообразие передвижений, совершаемое в сети, может быть разбито на группы по цели передвижения. Наиболее важными и многочисленными являются

- передвижения от мест жительства к местам приложения труда и обратно (так называемые трудовые корреспонденции);
- передвижения от мест жительства к местам культурно-бытового обслуживания (магазинам и др.) и обратно;
- передвижения, совершаемые между местами приложений труда (деловые поездки);
- передвижения, совершаемые между объектами культурно-бытового обслуживания.

Для каждой группы передвижений рассчитывается своя матрица межрайонных корреспонденций. Входной информацией к модели расчета корреспонденций являются общие объемы прибытия и отправления в каждом районе ПО. Оценка объема прибытий и отправок по разным группам связана с пространственным размещением объектов посещения и *подвижностью населения*, то есть средним количеством поездок, совершаемых с теми или иными целями. Эта оценка строится на основе имеющихся демографических и социально-экономических данных и результатов обследований. Расчет матриц корреспонденций обычно осуществляется с применением гравитационных или энтропийных моделей [3–5].

Матрицы корреспонденций расщепляются по способам передвижений в зависимости от соотношения обобщенных цен передвижений разными способами. Под различными способами передвижений понимают, например, передвижение пешком, с использованием общественного транспорта или личного автомобиля. Расщепление по способам обычно осуществляется с применением стандартных моделей дискретного выбора.

Изложенная принципиальная схема моделирования в настоящее время применяется на практике повсеместно. Основным недостатком этой схемы можно считать тот факт, что отдельные матрицы корреспонденций вычисляются и расщепляются по способам передвижений независимо друг от друга. В результате, например, может не совпадать количество людей, использующих автомобиль для поездки из дома на работу и обратно. Более того, общее количество передвижений «дом–работа» и «работа–дом» для каждой пары районов, вообще говоря, разное, поскольку жители обычно совершают попутные посещения магазинов и других мест по дороге домой (или на работу). Для корректного моделирования реальной структуры передвижений необходимо, таким образом, учитывать, что передвижения объединены в цепочки. Цепочка — это последовательность передвижений, которая начинается и закан-

чивается в одном месте (обычно дома), например, «дом–работа–магазин–дом». Как правило, способ передвижения не меняется для всех звеньев цепочки. Исключение могут составлять «промежуточные» цепочки, то есть цепочки, начинающиеся и заканчивающиеся в некотором промежуточном пункте. Например, в течение рабочего дня человек мог совершить деловую или бытовую поездку с возвращением на рабочее место.

По сложившейся терминологии модели, в которых используются отдельные матрицы корреспонденций, называются «trip-based», а модели, в которых явно рассматриваются цепочки передвижений, называются «tour-based». Среди известных практически реализованных «tour-based» моделей можно упомянуть модель Стокгольма [6] или Нью-Гемпшира [7].

Объединение отдельных передвижений в цепочке позволяет учесть влияние этих передвижений друг на друга. Например, если посещение магазина совершается из дома в рамках простейшей цепочки «дом–магазин–дом», то наиболее вероятными будут цели посещения неподалеку от дома. Если же это посещение совершается в составе цепочки «дом–работа–магазин–дом», то велика вероятность целей, находящихся на пути с работы домой, даже если они расположены далеко от дома. При таком подходе, однако, тот факт, что житель совершил попутное посещение магазина на пути с работы домой, не будет влиять на вероятность совершения дополнительной поездки в магазин после возвращения домой. Логическим шагом является рассмотрение полных «схем дневной активности» («day-pattern»), представляющих собой совокупность всех цепочек передвижений, совершаемых в течение дня. При таком подходе в качестве элементов, формирующих структуру передвижений, выступают схемы, например: 1) «дом–работа–дом»; 2) «дом–работа–магазин–дом»; 3) «дом–работа–дом; дом–магазин–дом». Каждая схема имеет свою вероятность реализации.

Модели такого типа носят общее название *моделей активности* (activity-based). Под «активностью» понимается тот или иной тип деятельности, требующий передвижения. Например, работа или посещение магазина. Одной из первых моделей активности, получившей широкое практическое применение, была модель SAMS (Sequenced Activity Mobility Simulator) [8]. Дальнейшее развитие моделей активности, происходившее в течение последнего десятилетия, было связано с попытками учета все большего числа тонких аспектов процесса формирования передвижений. Среди последних практически реализованных моделей упомянем модель Сан-Франциско [9], Нью-Йорка и ряда других городов [10], Далласа и Сиднея [11].

Математическая формулировка и программная реализация моделей активности и классической 4-стадийной модели существенно отличаются

ся. В классической модели элементами структуры являются отдельные передвижения, которые составляют матрицы передвижений (корреспонденций). В моделях активности в качестве элементов выступают целые последовательности передвижений. Для одновременного вычисления объемов совершаемых передвижений потребовались бы многоиндексные массивы, требующие чрезмерной компьютерной памяти. Проблема еще более усугубляется при учете распределения передвижений по времени суток. В этом случае у каждого передвижения в составе каждой цепочки появляется дополнительная характеристика — время начала передвижения.

При практической реализации моделей активности вместо расчета многоиндексных массивов производится микромоделирование отдельных передвижений методом Монте-Карло. Для каждого отдельного жителя производится случайный выбор характеристик передвижений — размещения целей, времени совершения, способа передвижения. При классификации цепочек передвижений различают основную и побочные цели. Каждая цепочка передвижений фактически представляет собой передвижение к основной цели с попутным посещением побочных целей. Для упрощения процедуры выбора применяют ее *декомпозицию*. Обычно применяется декомпозиция до 3 уровней [12]:

1. Выбор схемы дневной активности.
2. Моделирование цепочек: определение основной цели, основного способа и времени начала основного передвижения для каждой цепочки.
3. Моделирование звеньев: выбор расположения промежуточной цели, выбор способа и времени начала передвижения для всех звеньев.

Одной из важных особенностей моделей активности является возможность учета согласованного поведения членов семьи [13]. Например, в семьях, имеющих автомобиль, возможны передвижения типа «подвоза» членов семьи до некоторых промежуточных пунктов. Для жителей пригородов такие поездки являются типичными. С другой стороны, в семьях с единственным автомобилем факт совершения поездки на автомобиле одним из членов семьи, естественно, влияет на вероятность совершения такой же поездки другим членом.

Классификация схем активности и определение наиболее распространенных схем требует широких социологических обследований. Основной методикой таких обследований должно быть ведение «дневников передвижений». В существующих моделях количество учитываемых схем дневной активности колеблется от нескольких десятков до нескольких сотен.

Несмотря на значительные достигнутые успехи в развитии моделей активности, их практическое применение связано с трудностями:

- Для создания моделей требуется проведение масштабных и трудоемких социологических обследований.
- Реализация моделей стоит дорого. Создание известных моделей для крупных городов за рубежом занимало в типичном случае 2–3 года и стоило 600–800 тыс. долл.
- Очень большие времена счета (сравнительно с классическими моделями) — от 0.5 до 2-х суток на современном ПК.

III. Оценка межрайонных дальностей и расщепление по видам транспорта

Ввиду того, что продвинутые модели передвижений являются в настоящее время затратными и сложными в эксплуатации, сохраняет актуальность задача разработки упрощенных моделей, позволяющих адекватно воспроизводить структуру передвижений. Такую возможность предоставляет классическая 4-стадийная схема. Здесь мы обсудим некоторые проблемы, возникающие при реализации этой схемы, а также возможные способы их преодоления.

Обычно при построении комплексной модели транспортной системы города сначала вычисляются суммарные матрицы корреспонденций, а затем они расщепляются по способам передвижений. Ключевую роль в решении обеих задач играют межрайонные *дальности*, то есть показатели, характеризующие удаленность районов друг от друга с точки зрения скорости и удобства передвижения в данной транспортной системе. Обычно в качестве меры дальности используется обобщенная цена передвижения между районами. Обобщенная цена — это взвешенный критерий, включающий в себя разные факторы, такие как время передвижения, комфорт, оплата и другие. В дальнейшем для краткости будем называть обобщенную цену передвижения просто ценой.

Очевидно, передвижения, совершаемые между фиксированной парой районов разными способами, имеют разную цену. Именно на основе сопоставления этих цен определяются коэффициенты расщепления корреспонденции по способам. Однако для расчета суммарных (не расщепленных) корреспонденций требуется в качестве меры дальности использовать цену передвижения, каким-то образом усредненную по разным способам. Далее будет изложена корректная методика такого усреднения (впервые предложенная в [14]).

Корректность методики состоит в том, что должно выполняться естественное свойство *моно-*

тонности: мера дальности должна быть монотонно возрастающей (по крайней мере — неубывающей) функцией каждой из цен, вычисленных для отдельных способов передвижения. Другими словами, если цена передвижения одним из способов увеличивается, а цены для всех других способов не меняются, то межрайонная дальность не может уменьшиться. Она должна либо увеличиться, либо не измениться.

Далее везде, где рассматриваются цены передвижений или коэффициенты расщепления корреспонденций, подразумеваются передвижения между некоторой (произвольной) фиксированной парой районов. Индексы районов в формулах для краткости опускаются.

III.1. Условие монотонности оценок межрайонной дальности

Будем рассматривать n разных способов передвижения. Обозначим через t_1, \dots, t_n цены передвижения разными способами, s_1, \dots, s_n — коэффициенты расщепления: s_i — доля корреспонденции, использующая способ i . Коэффициенты расщепления предполагаются функциями цен передвижений: $s_i = s_i(t_1, \dots, t_n)$. Усредненную по способам передвижения цену обозначим через t^s . Стандартная формула усреднения с учетом коэффициентов расщепления:

$$t^s = \sum_i t_i s_i(t_1, \dots, t_n). \quad (1)$$

Тогда требуемое условие монотонности имеет вид

$$\frac{\partial t^s}{\partial t_j} = s_j + \sum_i \frac{\partial s_i}{\partial t_j} t_j \geq 0, \quad \forall j. \quad (2)$$

В корректных моделях условия (2) должны проверяться при использовании эмпирических оценок коэффициентов расщепления или специально накладываться при выводе тех или иных теоретических формул расщепления. Рассмотрим наиболее часто используемую полиномиальную (multinomial logit) модель расщепления. Модель применяется для широкого класса задач, в которых присутствует выбор из дискретного набора возможностей, для которых определена некоторая оценка «полезности» (или, наоборот, «затрат»), связанных с выбором этих возможностей. Согласно этой модели вероятность выбора одной из возможностей убывает пропорционально экспоненте от роста затрат:

$$s_i = \frac{e^{-U_i}}{\sum_j e^{-U_j}}, \quad \forall i, \quad (3)$$

где U_i — функция затрат при выборе i -й возможности. В задаче расщепления передвижений по способам естественно считать затраты связанными с ценой передвижения:

$$U_i = \alpha_i t_i + \beta_i. \quad (4)$$

Здесь коэффициенты α_i, β_i — это калибровочные параметры. Параметр α_i характеризует чувствительность населения к фактору цены при передвижениях i -м способом. Параметры β_i характеризуют асимметрию между способами передвижений. Коэффициенты β_i определены с точностью до прибавления общей константы, поэтому всегда можно положить самый маленький коэффициент равным 0.

Условие монотонности (2) с учетом (3)–(4) принимает вид

$$\sum_i (1 + \alpha_i(t_i - t_j)) e^{-U_i} \geq 0, \quad \forall j. \quad (5)$$

Ясно, что это условие выполняется не при всех значениях цен передвижений. В частности, условие нарушается, когда одна из цен существенно превосходит другие. В общем случае можно констатировать, что монотонность всегда будет нарушаться при больших t_i , если коэффициент расщепления s_i слишком быстро убывает с ростом t_i . В этом случае соответствующее слагаемое в сумме (1) убывает, когда t_i растет.

III.2. Метод вывода корректных формул оценки дальности

Предложенная в [14] методика оценки межрайонной дальности основана на следующих простых соображениях. Рассмотрим задачу определения цены передвижения одним из способов, например, на легковом автомобиле. Теоретически из одного района в другой можно проехать большим числом путей за разное время. При этом в качестве меры дальности естественно выбирается время (обобщенная цена) самого короткого пути, никакого усреднения по всем альтернативным путям не производится. Подобный подход можно было бы применить и при сравнении разных способов передвижения, то есть не усреднять цены для разных способов, а просто выбрать минимальную из них. Заметим, что такая оценка естественно удовлетворяет свойству монотонности, поскольку функция минимума из нескольких переменных всегда монотонна по каждому аргументу.

Почему же возникает необходимость в усреднении? Как представляется, причина состоит в том, что сопоставление разных способов передвижений не сводится к простому сравнению времен. Здесь учитываются разные дополнительные факторы — комфорт, денежные затраты, просто наличие или отсутствие автомобиля. Количественное влияние этих факторов различается для разных участников движения, поэтому возникает необходимость в усреднении разных индивидуальных оценок.

Предлагаемая методика основана на достаточном прямолинейной реализации изложенных выше соображений. Предположим, что каждый участник движения при сравнении способов передвижения использует индивидуальную оценку дальности

сти передвижения каждым возможным способом:

$$\tilde{t}_i = t_i + a_i, \quad (6)$$

где t_i — это «объективная» оценка цены передвижения, общая для всех участников, а a_i — это субъективное «штрафное» время, которое добавляется к цене передвижения i -м способом при сравнении разных способов. Коэффициенты a_i являются, по существу, поведенческими характеристиками участников движения, причем эти характеристики можно считать случайными величинами. Индивидуальная оценка межрайонной дальности каждым участником движения дается формулой

$$t^s(a_1, \dots, a_n) = \min(t_1 + a_1, \dots, t_n + a_n). \quad (7)$$

Оценку (7) можно назвать *микроскопической* оценкой межрайонной дальности, поскольку она описывает поведение отдельного элемента системы (участника движения). Соответственно *макроскопическая* оценка может быть получена путем усреднения микроскопической оценки по всем участникам движения. Для усреднения необходимы дополнительные предположения о виде распределения значений поведенческих параметров a_i . Пусть $\rho(a_1, \dots, a_n)$ — известная плотность этого распределения. Тогда макроскопическая оценка дальности дается интегралом

$$t^s = \int \rho(a_1, \dots, a_n) t^s(a_1, \dots, a_n) da_1 \dots da_n. \quad (8)$$

Предложенный подход может быть развит и конкретизирован в двух основных направлениях. Во-первых, вместо (6) можно применить более сложную методику вычисления индивидуальной оценки дальности: $\tilde{t}_i = \tilde{t}_i(t_i, a_i)$, где a_i не обязательно является аддитивной добавкой к цене, но играет роль поведенческого параметра, так или иначе влияющего на субъективную оценку передвижений i -м способом. Во-вторых, можно сделать разные предположения о характере распределения значений поведенческих параметров среди населения.

III.3. Формулы оценки дальности с учетом двух видов транспорта

В качестве примера выведем явные формулы для оценки межрайонных дальностей с учетом двух возможных способов передвижения: поездок на легковом автомобиле и на общественном транспорте. Соответственно используется единственный поведенческий параметр a , влияющий на сравнение цен поездок на этих видах транспорта. Будем обозначать через t_{car} цену движения на легковом автомобиле, t_{pub} — цену движения на общественном транспорте. В соответствии с (7) индивидуальная оценка межрайонной дальности имеет вид

$$t^s(a) = \min(t_{car} + a, t_{pub}), \quad (9)$$

где параметр a имеет смысл добавочного штрафного времени, которое добавляется к цене движения на автомобиле для сопоставления с ценой движения в общественном транспорте. Отметим, что параметр a может принимать как положительные, так и отрицательные значения.

Обозначим плотность распределения параметра a через $\rho(a)$, соответствующую функцию распределения — через $F(a) = \int_{-\infty}^a \rho(a') da'$. Также будем обозначать разность цен на двух видах транспорта через $\Delta t = t_{pub} - t_{car}$. Тогда формула (8) для межрайонной дальности будет иметь вид

$$t^s = t_{pub} - F(\Delta t) \Delta t + \int_{-\infty}^{\Delta t} \rho(a) a da. \quad (10)$$

Для дальнейших вычислений необходимы предположения о виде распределения значений параметра a . Рассмотрим здесь простой случай, когда параметр имеет нормальное (Гауссово) распределение:

$$\rho(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(a-\mu)^2}{2\sigma^2}}, \quad (11)$$

где μ — математическое ожидание, а σ — дисперсия случайной величины a . С учетом (11) выражение для межрайонной дальности (10) окончательно примет вид

$$t^s = t_{pub} - F(\Delta t)(\Delta t - \mu) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\Delta t - \mu)^2}{2\sigma^2}}, \quad (12)$$

где $F(\Delta t)$ — функция нормального распределения:

$$F(\Delta t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\Delta t} e^{-\frac{(a-\mu)^2}{2\sigma^2}} da. \quad (13)$$

В предыдущих формулах предполагалось, что параметр a может иметь отрицательные значения. Отрицательное значение a соответствует ситуации, когда участник движения оценивает поездку на легковом транспорте как предпочтительную, даже если t_{car} равно или немного превосходит t_{pub} . При этом индивидуальная оценка дальности может для некоторых пар районов принимать отрицательные значения (а именно, возможен случай $t_{car} + a < 0 < t_{pub}$). Этот эффект можно признать нежелательным с интуитивной точки зрения, поскольку межрайонная дальность обычно мыслится как существенно положительная величина (впрочем, такой подход имеет право на существование, если воспринимать дальность как абстрактный критерий, который, в частности, может включать в себя отрицательные «штрафные» слагаемые).

Можно обеспечить положительность оценки дальности, если внести небольшую модификацию в формулу для индивидуальной оценки (9). А

именно, в случае отрицательного значения a можно не уменьшать цену поездки на легковом транспорте, а вместо этого увеличить на ту же величину цену поездки на общественном транспорте, т. е. вместо (9) используется формула

$$t^s(a) = \begin{cases} \min(t_{car} + a, t_{pub}), & a \geq 0, \\ \min(t_{car}, t_{pub} - a), & a < 0. \end{cases} \quad (14)$$

Можно показать, что формула средней дальности (10) при этом заменяется на следующую:

$$t^s = t_{pub} - F(\Delta t)\Delta t + \int_0^{\Delta t} \rho(a)ada, \quad (15)$$

причем в случае $\Delta t < 0$ следует изменить знак перед интегралом и поменять местами пределы интегрирования в правой части (15). Поскольку модифицированная индивидуальная оценка (14) удовлетворяет требованию монотонности, это свойство сохраняется и для средней дальности (15) (это можно также проверить прямым дифференцированием).

В предположении, что параметр a имеет нормальное распределение, интегрирование в (15) дает следующее выражение для средней дальности:

$$t^s = t_{pub} - F(\Delta t)(\Delta t - \mu) - F(0)\mu + \frac{\sigma}{\sqrt{2\pi}} \left(e^{-\frac{\mu^2}{2\sigma^2}} - e^{-\frac{(\Delta t - \mu)^2}{2\sigma^2}} \right).$$

IV. Принцип равновесия

В основе алгоритмов распределения транспортных потоков по путям в сети лежит предположение о том, что каждый участник движения стремится минимизировать обобщенную цену своего пути. Цена пути представляет собой сумму цен движения по дугам, из которых состоит путь, а также цен переходов с дуги на дугу (маневров на пересечениях). В дальнейшем для сокращения формул будем опускать цены маневров и говорить только о ценах дуг, которые будем обозначать через c_a , где a — номер дуги.

Равновесным называется распределение потоков, при котором ни один из участников движения не может изменить свой путь так, чтобы уменьшить цену пути. Согласно современным представлениям распределение потоков в транспортной сети в установившейся ситуации является равновесным в указанном смысле. Сформулированное условие известно как условие Вардрупа (или условия Вардрупа, поскольку оно часто формулируется в виде двух предложений [14]).

Ключевым предположением о цене является следующее: цена движения по дуге является неубывающей функцией суммарного потока по этой дуге: $c_a = c_a(u_a)$. Другими словами, чем больше автомобилей движется по дуге, тем более

медленным и дискомфортным становится движение и соответственно возрастает цена пути. При данном предположении можно показать, что в любой транспортной системе существует равновесное состояние. Это состояние является единственным в том смысле, что условие равновесия однозначно определяет потоки на всех дугах сети. Другими словами, существует единственная загрузка элементов сети, удовлетворяющая условию равновесия. Однако равновесное распределение не является единственным в терминах распределения корреспонденций по путям в сети. Т. е. разные распределения корреспонденций по путям могут породить одинаковую загрузку всех элементов сети (на самом деле, для данной загрузки существует бесконечное количество порождающих ее распределений).

Равновесное распределение дается решением следующей задачи оптимизации с ограничениями [15]:

$$\sum_a \int_0^{u_a} c_a(v)dv \rightarrow \min. \quad (16)$$

Потоки на дугах u_a , входящие в это выражение, не являются независимыми переменными задачи. Они представляют собой суммы потоков u_{kpq} по всем путям k , использующим данную дугу a . Ограничения в задаче состоят в том, что потоки по путям u_{kpq} должны в сумме давать корреспонденции F_{pq} между всеми парами районов p и q :

$$\begin{aligned} u_a &= \sum_{p,q} \sum_{k \in K_{pq}, a \in k} u_{kpq}, \\ F_{pq} &= \sum_{k \in K_{pq}} u_{kpq}, \\ u_{kpq} &\geq 0. \end{aligned} \quad (17)$$

IV.1. Алгоритмы вычисления равновесного распределения

Хотя принцип равновесия в транспортной сети был впервые сформулирован еще в 50-е гг., практическое применение этого принципа началось значительно позже, в середине 70-х гг. [16]. Это связано с тем, что численная реализация модели требует значительного объема памяти компьютера. Для хранения всей информации о распределении корреспонденций по путям в крупном городе требуется значительный объем (до нескольких гигабайт) памяти. Компьютеры с такими возможностями стали широко распространены только в последнее время. Однако возможность моделировать равновесное распределение появилась значительно ранее, благодаря разработке алгоритма Франке–Вульфа [17]. Особенностью этого алгоритма является то, что он дает возможность вычислить итоговую равновесную загрузку транспортной сети, не сохраняя в памяти компьютера само распределение корреспонденций по путям.

Схема работы алгоритма Франке–Вульфа состоит в следующем. На начальном шаге все корреспонденции распределяются на оптимальные (по цене) пути, рассчитанные по свободной сети. На каждом последующем шаге алгоритма $n + 1$ каждая корреспонденция уже распределена на некоторое количество путей (не более n), вычисленных на предыдущих n шагах. В результате возникает загрузка дуг сети u_a^n . С учетом этой загрузки рассчитываются цены дуг c_a^n и строится новая система оптимальных путей. Данная система путей является «опорной» для перераспределения корреспонденций. При этом для некоторых корреспонденций рассчитанные пути могут оказаться действительно новыми, в то время как для других корреспонденций рассчитанный оптимальный путь уже содержится среди путей, полученных на предыдущих шагах. Далее вычисляется доля корреспонденций λ , которая будет перераспределена со всех прежних путей на новые оптимальные пути (способ вычисления λ для последующего анализа несущественен).

Особенностью алгоритма является то, что со всех прежних путей перераспределяется одна и та же доля корреспонденций λ . Это означает, что загрузка всех дуг уменьшится ровно в $(1 - \lambda)$ раз. Отсюда следует, что для вычисления новой загрузки всех дуг нет необходимости запоминать распределение по путям, а достаточно сохранять в памяти компьютера только загрузку всех дуг от предыдущего шага. С учетом того, что цены дуг зависят только от загрузки дуг $c_a = c_a(u_a)$, для вычисления самой доли λ также не требуется знать распределение по путям.

Описанное достоинство алгоритма Франке–Вульфа объясняет его «долготельство». Этот алгоритм до сих пор является наиболее распространенным в практике моделирования загрузки автомобильных сетей. Однако алгоритм обладает существенным недостатком. Хотя теоретически алгоритм всегда сходится, на практике скорость сходимости существенно ухудшается в ходе итераций. В течение первых итераций наблюдается достаточно быстрое приближение к равновесию, однако после нескольких десятков итераций приближение может практически прекратиться. Особенно сильно этот эффект проявляется на больших сетях. Ухудшение сходимости тесно связано с так называемым эффектом «застывающих потоков» (residual flows). Эффект проявляется в сильной неравномерности сходимости потоков к равновесным значениям на отдельных дугах, т. е. при общей хорошей сходимости к равновесию в сети может сохраняться небольшое количество дуг, на которых поток сильно отличается от равновесного, причем эти «выбросы» не устраняются последующими итерациями.

Причину эффекта «застывающих потоков» можно понять исходя из принципа работы алгоритма, показанного на рис. 1. В итоге работы алго-

ритма каждая корреспонденция должна быть распределена по набору альтернативных путей в таких пропорциях, чтобы цены всех путей в наборе в точности уравнились. Если минимальная и максимальная цены путей в наборе мало отличаются, то для приближения к равновесию значение λ нужно выбирать малым. Большее значение λ только отдалит распределение данной корреспонденции от равновесия. Предположим, что на ранних стадиях алгоритма в силу случайных причин небольшое количество корреспонденций «отстало» от основной массы корреспонденций по сходимости к равновесию. Для этих корреспонденций требуется перераспределить существенно большую долю λ . Поскольку значение λ выбирается как среднее (в некотором смысле) для всех корреспонденций, оно будет выбрано малым. В результате основная часть корреспонденций еще более приблизится к равновесию, в то время как «отставание» небольшой их части сохранится. При последующих итерациях значение λ , диктуемое основной частью корреспонденций, будет становиться все меньше и меньше, так что отдельные «застывшие» корреспонденции так и останутся перераспределенными.

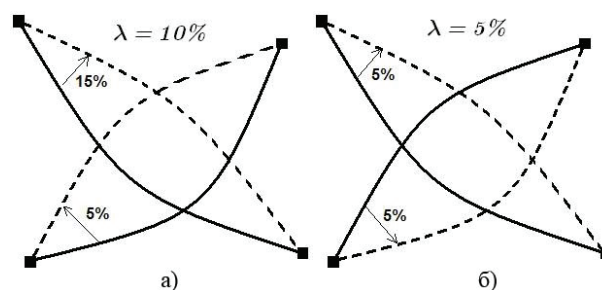


Рис. 1. Принцип работы алгоритма Франке–Вульфа

В последние годы были разработаны новые алгоритмы поиска равновесного распределения, свободные от указанных выше недостатков алгоритма Франке–Вульфа [18, 19]. Основная идея состоит в индивидуальной работе с корреспонденциями, что позволяет своевременно перераспределять отдельные «застывшие потоки». Разумеется, перераспределение одной корреспонденции среди альтернативных путей изменяет цены дуг и косвенно влияет на распределение других корреспонденций, использующих те же дуги. Отсюда возникает необходимость в многократном проходе всего массива корреспонденций для приведения системы в равновесное состояние.

Опишем простой алгоритм такого типа, использованный для моделирования Московской агломерации. Алгоритм может быть назван «балансировкой по путям». Основная идея алгоритма состоит в следующем. В равновесном состоянии каждая корреспонденция распределена среди некоторого количества альтернативных путей, причем это количество может быть разным для разных корреспонденций и заранее не известно. Если фиксировать некоторые наборы путей меж-

ду всеми парами районов, то можно сформулировать задачу поиска равновесия «в узком смысле»: найти равновесие в системе при дополнительном условии, что для передвижения используются только пути из фиксированных наборов, и никакие другие. Очевидно, что если в эти наборы включены все пути, используемые при настоящем равновесном распределении, то равновесие «в узком смысле» совпадает с равновесием в обычном смысле.

Будем называть фиксированный набор путей между парой районов «пучком» путей. На первом шаге алгоритма можно сформировать пучки, состоящие всего из одного пути — оптимального пути между соответствующей парой районов, рассчитанного по пустой сети. На каждом шаге алгоритма пучки пополняются новыми путями следующим образом. Сначала решается задача равновесия «в узком смысле» для имеющихся пучков (на первом шаге эта задача тривиальна). После этого каждый пучок пополняется оптимальным путем, рассчитанным при сложившейся загрузке дуг сети. Разумеется, в качестве «нового» может выступать один из ранее полученных путей. Фактически алгоритм закончит свою работу, когда ни один по-настоящему новый путь не будет найден. Отметим, что описанный алгоритм находит равновесное состояние за конечное число шагов, хотя оно, разумеется, не будет точным, поскольку балансировка пучков на каждом шаге ведется до достижения некоторой конечной точности.

Согласно установившейся в последние годы терминологии алгоритмы поиска равновесия в транспортной сети можно разделить на «дуговые» (link-based), то есть основанные на работе с потоками по дугам, и «маршрутные» (route-based), работающие с потоками по всем путям (или маршрутам) в сети. Типичным примером дугового алгоритма является алгоритм Франке–Вульфа, маршрутного типа — изложенный выше алгоритм балансировки по путям. Основным преимуществом маршрутных алгоритмов является существенное повышение равномерности сходимости потоков на дугах, в частности, решение проблемы «застревающих потоков». Однако это достигается за счет кардинального увеличения необходимого объема компьютерной памяти.

В этих условиях заслуживает внимания промежуточный подход, предложенный Ваг–Герга [19]. Этот подход представляет собой нечто среднее между дуговым и маршрутным подходами. Основной идеей является различение на каждой дуге представителей корреспонденций, двигающихся из общего района отправления, без различения районов прибытия. Таким образом, вместо потоков на дугах u_a или путях u_{kpq} в качестве независимых переменных используются потоки из «общего источника» (origin-based) u_{ap} , где a — номер дуги, p — номер района отправления. Очевидно, этот массив существенно (в количестве раз поряд-

ка числа районов) меньше массива потоков по путям и может помещаться в оперативную память. Вместе с тем он позволяет достаточно индивидуально работать с корреспонденциями от разных районов. Как показали численные эксперименты, алгоритм обеспечивает требуемую равномерность сходимости и решает проблему «застревающих потоков».

IV.2. Неединственность равновесного распределения и энтропия

Как уже было сказано выше, условие Вардрупа однозначно определяет значения суммарных потоков на всех дугах сети u_a . Это следует из вида целевой функции (16). Действительно, если цены дуг $c_a(u_a)$ являются возрастающими функциями потоков, то интегралы в (16) — выпуклые функции от u_a , поэтому равновесное распределение существует и единственно как точка минимума выпуклой функции при линейных ограничениях. Однако равновесное распределение не является единственным в терминах распределения корреспонденций по путям в сети. Причина этой неединственности наглядно показана на простом примере на рис. 2.

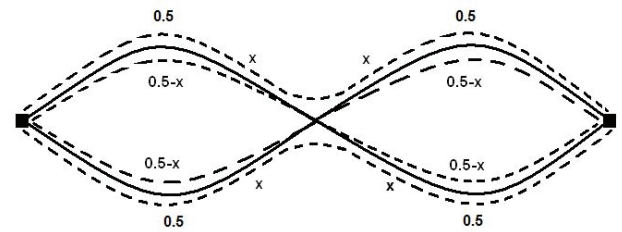


Рис. 2. Неединственность равновесного распределения

Пусть два района A и B соединены двумя парами идентичных путей с точкой пересечения посередине, как показано на рис. 2. Распределяется единственная корреспонденция из района A в район B . Ясно, что в условиях равновесия загрузки всех показанных дорог должны быть равны, то есть в обеих точках развилки корреспонденция должна делиться пополам. Существует ровно 4 разных пути из A в B . При этом существует бесконечное число способов распределения корреспонденции по этим путям, при которых корреспонденция на обеих развилках делится пополам. А именно, пусть x — это доля водителей, избравших на обеих развилках левый путь. Как показано на рисунке, при любом значении $0 \leq x \leq 0,5$ мы получаем распределение по четырем путям, порождающее равновесную загрузку. Различие этих распределений состоит только в том, сколько водителей из числа выбравших левый путь на первой развилке примет такое же решение на второй развилке.

В реальной сети с большим количеством районов ситуация более запутанная, однако природа неоднозначности распределения по путям остается такой же, как в этом примере. В сети присутствует большое количество подобных развилок,

при этом с точки зрения принципа равновесия важны только общие объемы потоков по дугам. При соблюдении нужных общих объемов совершенно не важно, из каких именно водителей состояются эти потоки.

Таким образом, приходим к заключению, что условие Вардрупа не определяет однозначно распределение корреспонденций по путям. На самом деле, каждому возможному набору значений потоков по дугам u_a соответствует многомерное линейное пространство распределений по путям u_{kpq} . Следовательно, для корректной постановки задачи поиска равновесного распределения требуется сформулировать какие-то предположения, дополнительные к условию Вардрупа.

Представляется естественным вероятностный подход к решению вопроса. Рассмотрим какую-либо пару районов p, q , соединенных пучком путей K_{pq} . Можно предположить, что водители (представители корреспонденции f_{pq}) выбирают один из путей $k \in K_{pq}$ случайно и независимо друг от друга. Если задать значения вероятности выбора того или иного пути, то несложно вычислить вероятность реализации того или иного распределения корреспонденции f_{pq} по путям. В частности, в условиях равновесия цены всех используемых путей равны, и поэтому можно принять, что каждый водитель выбирает любой из используемых путей с одинаковой вероятностью.

В терминах макросистемного подхода [20] эту ситуацию можно описать так. Водители являются элементами макросистемы, при этом выбор того или иного пути определяет состояние элемента. Микросостояние системы определяется совокупностью состояний элементов, то есть указанием пути для каждого водителя индивидуально. Макросостояние определяется «числами заполнения», то есть общим количеством элементов в том или ином состоянии. В нашем случае числа заполнения — это потоки на путях u_{kpq} . Таким образом, разные распределения корреспонденций по путям соответствуют разным макросостояниям системы. Каждое макросостояние может быть реализовано различным количеством микросостояний, что определяет его «статистический вес». В нашем случае, когда число элементов и число состояний конечны, статистические веса макросостояний просто пропорциональны вероятностям их реализации.

Если сеть находится в равновесии и известен набор равновесных значений потоков на дугах u_a^* , то это накладывает линейные ограничения на возможные значения u_{kpq} (см. ограничения (17)). Предполагая, что в реальности должно осуществляться состояние с наибольшим статистическим весом, приходим к модели максимизации энтропии (логарифма статистического веса) системы. Энтропийная модель равновесного распределения потоков может быть сформулирована в следую-

щем виде [21]:

$$\sum_{p,q} \sum_{k \in K_{pq}} \left(u_{kpq} \ln \left(\frac{u_{kpq}}{f_{pq}} \right) - u_{kpq} \right) \rightarrow \max \quad (18)$$

при ограничениях

$$\begin{aligned} u_a^* &= \sum_{p,q} \sum_{k \in K_{pq}, a \in k} u_{kpq}, \\ f_{pq} &= \sum_{k \in K_{pq}} u_{kpq}, \\ u_{kpq} &\geq 0. \end{aligned} \quad (19)$$

В связи с изложенным представляется интересным сравнить распределения по путям, возникающие при применении алгоритма Франке–Вульфа и маршрутных алгоритмов. Работа этих алгоритмов схематично показана на рис. 3. Пространство возможных наборов значений потоков на дугах u_a представлено в виде одномерной линии. В этом пространстве имеется единственная точка u_a^* , соответствующая равновесным потокам. При этом пространство возможных распределений по путям u_{kpq} оказывается расслоенным на подпространства, каждое из которых соответствует некоторому фиксированному набору u_a . В каждом из этих подпространств имеется наиболее вероятное (максимизирующее энтропию) распределение u_{kpq}^* . Алгоритмы поиска равновесия обычно начинают работу с распределения корреспонденций по оптимальным путям, вычисленным по свободной сети. Далее осуществляется итеративное перераспределение корреспонденций между путями. Логика этого перераспределения основана на стремлении привести точку u_a к точке равновесия u_a^* . При этом совершенно не учитывается степень близости u_{kpq} к u_{kpq}^* . Вследствие этого отклонение итогового распределения по путям от наиболее вероятного для всех этих алгоритмов выглядит произвольным, и сделать теоретически обоснованные выводы трудно.

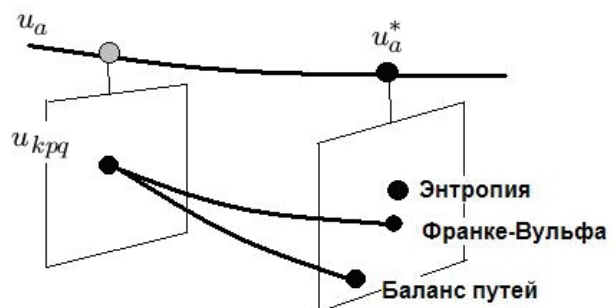


Рис. 3. Сопоставление разных алгоритмов поиска равновесия

Однако численные эксперименты свидетельствуют о том, что алгоритм Франке–Вульфа часто дает распределение по путям, более близкое к u_{kpq}^* , чем маршрутный алгоритм. В частности, в маршрутном алгоритме резко снижается расщепление корреспонденций по разным путям, то есть алгоритм стремится минимизировать число

разных путей. Такие же результаты получены в [22] при сравнении алгоритма Франке–Вульфа с алгоритмом, использующим потоки из общего источника. Возникает вывод, звучащий, на первый взгляд, несколько парадоксально: маршрутные алгоритмы по сравнению с дуговыми обеспечивают лучшую сходимость на дугах, но имеют тенденцию приводить к «худшему» распределению по маршрутам (путям).

В. Заключение

В работе дан краткий обзор моделей корреспонденций, основанных на понятии «активности». Модели этого типа могут обеспечить более реалистичное воспроизведение структуры передвижений в городах, чем классическая 4-стадийная схема. Однако применение моделей такого типа ограничивается сложностью реализации, дороговизной и слишком большими временами счета. Поэтому сохраняет актуальность задача построения упрощенных моделей передвижений, в особенности на ранних стадиях проектирования.

В работе рассмотрена также задача корректного формирования оценки межрайонной дальности на основе усреднения цен передвижений, совершаемых разными способами (разными видами транспорта или пешком). Изложен метод вывода формул для такого усреднения, который обеспечивает автоматическое выполнение естественного свойства монотонности оценки как функции цен передвижений разными способами. В рамках данного подхода могут быть получены формулы для вычисления межрайонной дальности существенно разного вида. Для практического применения результатов требуется дальнейшее изучение свойств получаемых оценок и сопоставление теоретических функций с имеющимися статистическими данными.

Также рассмотрено применение принципа равновесия для моделирования распределения корреспонденций по транспортной сети. В задаче моделирования транспортных потоков можно выделить два уровня: расчет потоков на дугах сети, то есть расчет загрузки сети, а также расчет распределения всех корреспонденций по путям в сети, которое порождает эту загрузку. Принцип равновесия однозначно определяет потоки на дугах сети, но оставляет большой произвол в выборе распределения по путям. Для устранения этой неопределенности принцип равновесия можно дополнить принципом максимизации энтропии, позволяющим рассчитать наиболее вероятное распределение по путям при известных потоках на дугах, хотя такой подход сопряжен со значительными вычислительными трудностями в связи с большой размерностью задач.

Литература

1. Швецов В.И. Математическое моделирование транспортных потоков // Автоматика и Телемеханика. — 2003. — № 11. — С. 3–46.
2. Швецов В.И. Алгоритмы распределения транспортных потоков // Автоматика и Телемеханика. — 2009. — № 10. — С. 148–157.
3. Carrothers G.A.P. An historical review of the gravity and potential concepts of human interaction // Journal of the American Institute of Planners. — 1956. — V. 22. — P. 94–102.
4. Wilson A.G. A statistical theory of spatial distribution models // Transpn. Res. — 1967. — V. 1. — P. 253–270.
5. Wilson A.G. A family of spatial interaction models and associated developments // Envir. & Plan. A. — 1971. — V. 3. — P. 255–282.
6. Algiers S., Daly A., Kjellman P., Widlert S. Stockholm model system (sims): Application // 7th World Conference of Transportation Research. — Sydney, Australia, 1995.
7. Rossi T., Shifan Y. Tour based travel demand modelling in the u.s. // Proceeding of the 8th IFAC /IFIP/ IFORS Symposium on Transportation Systems. — Chania, Greece, 1997.
8. Kitamura R., Pas E.I., Lula C.V., Lawton T.K., Benson P.E. The sequenced activity mobility simulator (sams): An integrated approach to modelling transportation, land use and air quality // Transportation. — 1996. — V. 23(3). — P. 267–291.
9. Jonnalagadda N., Freedman J., Davidson W.A., Hunt J.D. Development of a microsimulation activity-based model for san francisco – destination and mode choice models // 81st Annual Meeting of the Transportation Research Board. — Washington, 2002.
10. M. Bradley, Vovsha P. A model for joint choice of daily activity pattern types of household members // Transportation. — 2005. — V. 32. — P. 545–571.
11. Arentze T., Timmermans H.J.P. A co-evaluation approach to extracting and predicting linked sets of complex decisions rules from activity diary data // 80Th Annual Meeting of the Transportation Research Board. — Washington, D.C., 2001.
12. Shifan Y., Ben-Akiva M. A practical policy sensitive activity-based model // 11th International Conference on Travel Behaviour Research. — Kyoto, 2006.
13. Zhang J., Timmermans H.J.P., Borgers A. A model of household task allocation and time use // Transpn. Res. B. — 2005. — V. 39. — P. 91–95.
14. Котельникова А.Г., Швецов В.И. Оценка межрайонных дальностей в транспортных моделях // Сб. Трудов ИСА РАН «Прикладные про-

блемы управления макросистемами»». — 2010. — Т. 59. — С. 64–67.

15. *Wardrop J.G.* Some theoretical aspects of road traffic research // Proc. Institution of Civil Engineers II. — 1952. — P. 325–378.

16. *Sheffy Y.* Urban Transportation Networks. — Englewood Cliffs, N. J: Prentice–Hall, 1985.

17. *LeBlanc L.J., Morlok E.K., Pierskalla W.* An efficient approach to solving the road network equilibrium traffic assignment problem // Transpn. Res. B. — 1975. — V. 9. — P. 309–318.

18. *Frank M., Wolfe P.* An algorithm for quadratic programming // Naval Research Logistics Quarterly. — 1956. — V. 3. — P. 95–110.

19. *Patriksson M.* The Traffic Assignment Problem — Models and Methods. — Utrecht, Netherlands: VSP, 1994.

20. *Bar-Gera H.* Origin-based algorithm for the traffic assignment problem // Transpn. Sci. — V. 36, N 4. — P. 398–417.

21. *Popkov Yu. S.* — Macrosystems theory and its applications. — Berlin: Springer Verlag, 1995.

22. *Rossi T.F., McNeil S., Hendrickson C.* Entropy model for consistent impact fee assessment // Journal of Urban Planning and Development / ASCE. — 1989. — V. 115. — P. 51–63.

23. *Bar-Gera H.* Origin-based algorithms for transportation network modeling. — Univ. of Illinois at Chicago, 1999.

Поступила в редакцию 15.10.2010.

Памяти академика А. А. Петрова



На 78-м году жизни скончался выдающийся ученый, яркий, глубокий и очень интеллигентный человек, академик РАН, лауреат Государственной премии СССР, научный руководитель факультета управления и прикладной математики Московского физико-технического института (государственного университета), член координационного совета МФТИ, заведующий кафедрой «Анализ систем и решений» МФТИ, заслуженный профессор МФТИ и МГУ Петров Александр Александрович.

В 1957 году Александр Александрович окончил Московский физико-технический институт, получив профессию инженера-физика по специальности *термодинамика*. С 1960 по 1963 год работал ассистентом кафедры «Теоретической механики» МФТИ. С 1963 по 2011 год работал в Вычисли-

тельном центре РАН (АН СССР), с 1963 по 1968 год в должности младшего научного сотрудника, с 1968 года по 1988 год в должности заведующего сектором, с 1988 года — заведующим отделом математического моделирования экономических систем. В 1964 году в МФТИ защитил диссертацию на соискание степени кандидата физико-математических наук по специальности *вычислительная математика*. С 1965 по 2011 год работал в МФТИ. С 1965 по 1974 год — доцентом МФТИ, с 1974 по 1999 год — профессором МФТИ, с 2001 по 2011 год заведовал кафедрами МФТИ: «Теория управления и исследование операций» (до 2005), с 2005 г. — кафедрой «Анализ систем и решений». В 1974 году стал доктором физико-математических наук по специальности *математическая кибернетика*. В 1981 году присуждено звание профессора по специальности *математическая кибернетика*. В 1991 году А. А. Петров избран членом-корреспондентом РАН в секцию математики, механики, информатики по специальности *математическое моделирование*. В 1997 году Александр Александрович Петров избран академиком РАН в Отделение информатики, вычислительной техники и автоматизации по специальности *информатика*. С 1998 профессор кафедры «Информационные системы» Московского государственного университета.

Широчайшую известность получили работы А. А. Петрова по движению тел с полостями, частично заполненными жидкостью (1957—1968). В 1980 год за цикл этих работ А. А. Петров был удостоен Государственной премии СССР. В 1975 году А. А. Петровым была создана научная школа, специализирующаяся в математической экономике, и сочетающая традиционные подходы с культурой математического моделирования, сложившейся в механике и физике. В этой школе выросло несколько замечательных специалистов. Например, проф. А. В. Лотов, заведующий базовой кафедрой ФУПМ МФТИ чл.-корр. РАН И. Г. Поспелов и декан ФУПМ проф. А. А. Шананин. Последние тридцать лет Александр Александрович посвятил анализу отечественной экономики методами математического моделирования. Александру Александровичу Петрову удалось решить казавшуюся неразрешимой проблему — *проанализировать на языке математических моделей эволюцию российских экономических структур*.

Литературный талант Александра Александровича в полной мере проявился в серии блистательных статей о Физтехе и первом декане ФУПМ (своем учителе) академике Н. Н. Моисееве.

Светлая память об Александре Александровиче навсегда останется в наших сердцах.

*Н. Н. Кудрявцев, Л. А. Бекларян, А. А. Белолипецкий, Ю. Г. Евтушенко,
Ю. И. Журавлев, В. П. Иванников, В. В. Ивантер, Ю. Н. Павловский,
И. Б. Петров, Е. С. Половинкин, Ю. П. Попов, И. Г. Поспелов, К. В. Рудаков,
Г. И. Савин, А. М. Тер-Крикоров, А. Г. Тормасов, Ю. А. Флеров,
А. С. Холодов, Б. Н. Четверушкин, А. А. Шананин*

Summaries of all articles

L.G. Afanasyeva, E.V. Bulinskaya

Mathematical models of transport systems based on queuing theory

The aim of the paper is to obtain ergodicity conditions for some traffic systems involving traffic lights. The cars arrival is described by Poisson processes. The traffic lights interswitching times can have an arbitrary distribution. The optimization of traffic lights performance is also considered. The existence of a steady state for the system with two consecutive traffic lights is established, with account taken of the cars size. The impact of traffic lights and different cars velocities on the cars flow density is studied as well. The queuing theory methods turn out to be useful for investigation.

Keywords: transport flows, traffic lights, unregulated crossroads, ergodicity, optimization, flow density, heavy traffic.

M.L. Blank

Exclusion processes with synchronous updates in transport flow models

We study several simple traffic flow models represented by exclusion processes (with both lattice and continuous spaces) and obtain explicit formulas for certain associated statistics. In particular, the so-called fundamental diagram showing the dependence of the average particle speed on the particle density is derived rigorously.

Keywords: exclusion process, traffic flows, systems of interacting particles, coupling.

P.P. Bobrik

Deriving a transport correspondences gravitation model from the law of diminishing marginal utility

In this paper, we study relationships between economic behavior patterns and the famous examples of transport correspondences distribution.

Keywords: transport, gravitation model, law of diminishing marginal utility, route utility.

A.S. Bugaev, A.P. Buslaev, A.G. Tatashev, M.V. Yashina

Optimization of partially-connected flows for a deterministic-stochastic model

We study a multilane partially-connected movement model. In this model the velocity of movement is a sum of determinate and stochastic components. We solve the problem of maximizing the average velocity and intensity of movement by the parameters which determine the model.

Keywords: Partially-connected flows, multilane movement, leakage, dynamic size, stochastic model.

A.V. Gasnikov, E.V. Gasnikova

On possible dynamics in the calculation model of a correspondence matrix (A. G. Wilson)

In this paper, we present the basic technique for investigating the dynamics of macrosystems at large time values. In the core of the dynamics lies the ergodic markovian chain with a huge number of states. At large time values the macrosystem distribution on macrostates is close to the stationary one. With growing dimensions of the macrosystem (quantity of states of the markovian chain), the stationary distribution is concentrated in a small neighbourhood of the most probable macrostate which is accepted for the equilibrium of the given macrosystem. As an example of application of the described formalism, we develop the static gravitational calculation model of a correspondence matrix (one of the most popular models in practice) using the "reasonable"(individually favorable) dynamics of exchanges of residences.

Keywords: Ergodic theorem, Lyapunov's function, entropy, gravitational calculation model, correspondence matrix, invariant (stationary) measure phenomena, canonical scaling, dynamic balance condition, principle of detailed balance.

E.V. Gasnikova, Y.V. Dorn

On stochastic markovian dynamics providing the Nash–Vardrop equilibrium in a stream distribution model

In this paper, the possible dynamics providing the Nash–Vardrop equilibrium in a stream distribution model (based on the Vardrop principles) is described. It is necessary to note that the conclusions drawn in the paper are mainly based on the results of numerical experiments which show, in a number of games, a very fast convergence of proposed stochastic (Gibbs' type) dynamics (of the best responses in the corresponding evolutionary game) to the Nash equilibrium.

Keywords: Evolutionary game dynamics, Nash equilibrium, Vardrop principles, Braess paradox, Pareto optimality, Grigoriadis–Khachiyan algorithm.

A.A. Zamyatin, V.A. Malyshev

Car traffic flows - introduction to a stochastic approach

Stochastic models of transportation flows exhibiting qualitative phenomena are rather numerous: starting from the midtwentieth-century queuing theory to recent mathematical statistical physics and locally interacting processes. In this paper, we discuss and rigorously formulate many models and consider some examples in detail.

Keywords: point processes, queuing networks, transportation flows, random grammars.

S.L. Klenov

Kerner's three-phase traffic theory as a theoretical basis for intelligent transportation systems

Foundations and results of Kerner's three-phase traffic theory are briefly discussed. Kerner's theory allows one to explain and predict empirical spatio-temporal features of traffic breakdown and the resulting traffic flow patterns. The measurement results of the empirical spatio-temporal traffic flow patterns occurring on highways in Germany, Great Britain, and the USA are considered. These empirical results are a basis both for empirical phase criteria and other conclusions of Kerner's three-phase traffic theory. In Kerner's theory, traffic breakdown is explained by a first-order phase transition from free flow phase to synchronized flow phase. The phase transition can occur in a wide range of traffic flow rates. This range corresponds to an infinite number of highways capacities. The features of wide moving jams formation in a synchronized flow as well as a classification of spatio-temporal traffic flow patterns in Kerner's theory are discussed. Novel intelligent transportation systems based on Kerner's three-phase traffic theory are exemplified.

Keywords: traffic flow, Kerner's three-phase traffic theory, intelligent transportation systems, traffic breakdown, infinite number of highway capacities.

A.V. Kolesnikov

Mass transportation and contractions

According to L. Caffarelli's celebrated result, every optimal mass transportation mapping, pushing forward the standard Gaussian measure onto a log-concave measure $e^{-W} dx$ with $D^2W \geq Id$, is 1-Lipschitz. We present a short survey of related results and various applications.

Keywords: optimal transportation, Monge–Ampère equation, log-concave measures, Gaussian measures, isoperimetric inequalities, Sobolev inequalities.

A.B. Kurzhanski, A.A. Kurzhanski, P. Varaiya

Role of macromodeling in network active traffic management

Active Traffic Management (ATM) is the ability to dynamically manage recurrent and nonrecurrent congestion based on prevailing traffic conditions in order to maximize the effectiveness and efficiency of road networks. It is a continuous process of (1) obtaining and analyzing traffic measurement data; (2) operations planning — simulating various scenarios and control strategies; (3) implementing the most promising control strategies in the field; and (4) maintaining a real time decision support system that filters current traffic measurements to predict the traffic state in the near future, and to suggest the best available control strategy for the predicted situation. ATM relies on a fast and trusted traffic simulator for the rapid quantitative assessment of a large number of control strategies for the road network under various scenarios, in a matter of minutes. The open source macrosimulation tool Aurora Road Network Modeler is a good candidate for this purpose. The paper describes the underlying dynamical traffic model and what it takes to prepare the model for simulation; covers the traffic performance measures and evaluation of scenarios as part of operations planning; introduces the framework within which the control strategies are modeled and evaluated; and presents the algorithm for real time traffic state estimation and short term prediction.

Keywords: Active Traffic Management, road network, traffic congestion, dynamical traffic model, decision support system, scenarios simulation, fundamental diagram, control strategy, fast and trusted traffic simulator, performance measures, operations planning, real time traffic state estimation, prediction.

E.A. Nurminski, N.B. Shamray

Forecast modeling of car traffic in Vladivostok

Considerable funding is now allocated for improvement of transportation networks in Vladivostok and surrounding areas in federal programs for development of the Russian Far East. In the current work, the forecast models of car traffic load and redistribution of automobile flows are developed to estimate the effects of implementation of these projects. The results of numerical experiments of traffic assignments for existing and future networks based on the economic equilibrium and stationary dynamics model are described.

Keywords: competitive equilibrium, traffic assignment problem, variational inequalities, trip matrix, projective methods.

A.M. Raigorodskii

Models of random graphs and their applications

In this paper, we give a survey of the main contemporary directions of research in the theory of random graphs. It should be stressed that the models of random graphs are connected with some public transportation problems.

Keywords: random graph, Internet, transport networks.

N.N. Smirnov, A.B. Kiselev, V.F. Nikitin, A.V. Kokoreva

Mathematical modeling of traffic flows by a continua approach. Two lanes roads: modeling T-shape crossing and the effect of lane change on handling capacity

In this paper, the further development of continua models for traffic flows simulations is presented. The model for two-lanes traffic flows is developed. It takes into account changing lanes by some cars and is based on the approach of multiphase flow mechanics. Numerical simulations of traffic flows in T-shape crossing is performed as an example of taking into account the three-phase traffic lights regulation.

Keywords: mathematical modeling for traffic flows, traffic lights, T-shape crossing, lane change, handling capacity, continua mechanics, multiphase flows.

Y.A. Kholodov, A.S. Kholodov, A.V. Gasnikov, I.I. Morozov, V.N. Tarasov

Traffic flow modeling — modern problems and solution methods

This paper deals with mathematical modeling of vehicular traffic flows, the discussion of arising problems and solution methods. We develop the original macroscopic models for vehicular traffic description in the complex transport network graph. The macroscopic model is based on a hydrodynamic approach by analogy with a motivated compressible multicomponent fluid. We consider, as a component, a set of vehicles with some identical features. Also, we develop the original algorithm for deriving equation systems at the graph nodes that are crossroads. The numerical simulation results demonstrate the workability of the proposed model as compared to the experimental data.

Keywords: vehicular traffic flow, hydrodynamic approach, transport network graph, multiphase traffic flows.

B.N. Chetverushkin, M.A. Trapeznikova, I.R. Furmanov, N.G. Churbanova

Macro- and microscopic models for vehicular traffic description on multilane motorways

The paper deals with mathematical modeling of vehicular traffic flows in urban streets and on motorways. Original 2D macro- and microscopic models of multilane traffic are developed to predict flows on multilane roads, with account taken of their real geometry. The macroscopic model of synchronized flows is based on the kinetic approach by analogy with the quasi-gas-dynamic (QGD) system of equations. The microscopic model uses the cellular automata theory. Both models are generalized to the case of “multiphase” traffic to describe heterogeneous traffic flows. The phase implies a set of vehicles with some identical features. Test predictions demonstrate the adequacy of the models in different road situations.

Keywords: vehicular traffic flow, macroscopic and microscopic models, multilane traffic, multiphase traffic flow.

V.I. Shvetsov

Problems of the development of complex models of transportation systems

The paper gives a brief review of the recent approach in the development of mathematical models of traffic flows in the transportation network. This approach is characterized by transition from the trip-based to tour-based and daily activity-based models, and the use of the Monte-Carlo simulations. We also discuss the problem of evaluating the estimate of the interzone transportation distance, which is averaged over the trip modes. We propose the general correct-averaging method. Also, we discuss the different algorithms of evaluating the user equilibrium assignment in the network. The equilibrium principle defines the unique values of link flows, though the distribution of flows along different routes remains nonunique. The entropy maximization model can be used to eliminate this nonuniqueness.

Keywords: traffic flows, trip matrix, activity-based models, modal split, traffic equilibrium.

Сведения об авторах статей
(на момент подачи статьи)

Математические модели транспортных систем, основанные на теории очередей

Афанасьева Лариса Григорьевна (д.ф.-м.н., проф.), afanas@nech.math.msu.su

Буллинская Екатерина Вадимовна (д.ф.-м.н., проф.), ebulinsk@mech.math.msu.su

Синхронно обновляемые процессы с запретами в моделях транспортных потоков

Бланк Михаил Львович (д.ф.-м.н., с.н.с., н.с.), blank@iitp.ru

Обоснование гравитационной модели транспортных корреспонденций при помощи закона убывающей предельной полезности

Бобрик Петр Петрович (к.ф.-м.н., с.н.с.), bobrikpp@mail.ru

Оптимизация частично-связных потоков в детерминированно-стохастической модели

Бугаев Александр Степанович (д.ф.-м.н., акад. РАН, зав. каф.), bugaev@cos.ru

Буслаев Александр Павлович (д.ф.-м.н., проф., зав. каф.)

Таташев Александр Геннадьевич (д.ф.-м.н., доц., проф.), a-tatashev@rambler.ru

Яшина Марина Викторовна (д.т.н., доц., зав. каф.), yash-marina@yandex.ru

О возможной динамике в модели расчета матрицы корреспонденций (А.Дж. Вильсона)

Гасников Александр Владимирович (к.ф.-м.н., доц.), avgasnikov@gmail.com

Гасникова Евгения Владимировна (асп.), egasnikova@ya.ru

О стохастической марковской динамике, приводящей к равновесию Нэша–Вардропа в модели распределения потоков

Гасникова Е. В. (асп.), egasnikova@ya.ru

Дорн Ю. В. (студ.), dorn@pisem.net

Транспортные автомобильные потоки — введение в вероятностный подход

Замятин Андрей Андреевич (к.ф.-м.н., с.н.с.), andrew.zamyatin@gmail.com

Мальшиев Вадим Александрович (д.ф.-м.н., проф., зав. каф.), malyshev2@yahoo.com

Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий

Кленов Сергей Львович (к.ф.-м.н., доц.), klenov.sergei@yandex.ru

Транспортировка масс и сжимающие отображения

Колесников Александр Викторович (д.ф.-м.н., проф., доц.), sascha77@mail.ru

Роль макро моделирования в активном управлении транспортной сетью

Куржанский Александр Борисович (д.ф.-м.н., проф., зав. каф.), kurzhans@mail.ru

Куржанский Александр Александрович (к.ф.-м.н., спец.), alex@relteq.com

Варайя Правин (к.ф.-м.н., проф.), varaiya@eecs.Berkeley.edu

Прогнозное моделирование автомобильного трафика Владивостока

Нурминский Евгений Алексеевич (д.ф.-м.н., проф., зав. каф.), nurmi@dvo.ru

Шамрай Наталья Борисовна (к.ф.-м.н., н.с., доц.), shamray@dvo.ru

Модели случайных графов и их применения

Райгородский Андрей Михайлович (д.ф.-м.н., доц., проф.), mraigor@yandex.ru

Математическое моделирование движения автотранспортных потоков методами механики сплошной среды. Двухполосный транспортный поток: модель Т-образного перекрестка, исследование влияния перестроений транспортных средств на пропускную способность участка магистрали

Смирнов Николай Николаевич (д.ф.-м.н., проф., зам. зав. каф., рук. подр.), ebifsun1@mech.math.msu.su

Киселев Алексей Борисович (д.ф.-м.н., проф., зам. рук. подр.), akis2006@yandex.ru

Никитин Валерий Федорович (к.ф.-м.н., доц.), vfnikitin@rambler.ru

Кокорева Анастасия Владимировна (к.ф.-м.н., м.н.с.), kokoreva@mech.math.msu.su

Моделирование транспортных потоков — актуальные проблемы и перспективы их решения*Холодов Ярослав Александрович* (к.ф.-м.н., доц.), kholodov@crec.mipt.ru*Холодов Александр Сергеевич* (д.ф.-м.н., проф., зав. каф.), xolod@crec.mipt.ru*Гасников Александр Владимирович* (к.ф.-м.н., доц.), gasnikov@ya.ru*Морозов Иван Игоревич* (асп.), IvanMorozov@rambler.ru*Тарасов Василий Николаевич* (студ.), vtarasov@zoneos.com**Макро- и микроскопические модели для описания движения автотранспорта на многополосных магистралях***Четверушкин Борис Николаевич* (д.ф.-м.н., чл.-корр. РАН, рук. орг.), chetver@imamod.ru*Трапезникова Марина Александровна* (к.ф.-м.н., с.н.с.), marina@imamod.ru*Фурманов Илья Ростиславович* (асп.), furious.cmc@gmail.com*Чурбанова Наталья Геннадьевна* (к.ф.-м.н., с.н.с.), nata@imamod.ru**Проблемы моделирования передвижений в транспортных сетях***Швецов Владимир Иванович* (к.ф.-м.н., с.н.с., в.н.с.), vshvets@isa.ru

Ссылки на опубликованные статьи
(в соответствии с ГОСТ Р 7.0.5-2008)

Афанасьева Л.Г., Булинская Е.В. Математические модели транспортных систем, основанные на теории очередей // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 6–21.

Afanasyeva L.G., Bulinskaya E.V. Mathematical models of transport systems based on queuing theory // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 6–21.

Бланк М.Л. Синхронно обновляемые процессы с запретами в моделях транспортных потоков // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 22–30.

Blank M.L. Exclusion processes with synchronous updates in transport flow models // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 22–30.

Бобрж П.П. Обоснование гравитационной модели транспортных корреспонденций при помощи закона убывающей предельной полезности // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 31–34.

Bobrik P.P. Deriving a transport correspondences gravitation model from the law of diminishing marginal utility // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 31–34.

Бугаев А.С., Буслаев А.П., Таташев А.Г., Яшина М.В. Оптимизация частично-связных потоков в детерминированно-стохастической модели // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 35–45.

Bugaev A.S., Buslaev A.P., Tatashev A.G., Yashina M.V. Optimization of partially-connected flows for a deterministic-stochastic model // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 35–45.

Гасников А.В., Гасникова Е.В. О возможной динамике в модели расчета матрицы корреспонденций (А.Дж. Вильсона) // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 45–54.

Gasnikov A.V., Gasnikova E.V. On possible dynamics in the calculation model of a correspondence matrix (A. G.Wilson) // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 45–54.

Гасникова Е.В., Дорн Ю.В. О стохастической марковской динамике, приводящей к равновесию Нэша–Вардроп в модели распределения потоков // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 55–57.

Gasnikova E.V., Dorn Y.V. On stochastic markovian dynamics providing the Nash–Vardrop equilibrium in a stream distribution model // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 55–57.

Замятин А.А., Малышев В.А. Транспортные автомобильные потоки — введение в вероятностный подход // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 58–74.

Zamyatin A.A., Malyshev V.A. Car traffic flows - introduction to a stochastic approach // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 58–74.

Кленов С.Л. Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 75–90.

Klenov S.L. Kerner's three-phase traffic theory as a theoretical basis for intelligent transportation systems // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 75–90.

Колесников А.В. Транспортировка масс и сжимающие отображения // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 90–99.

Kolesnikov A.V. Mass transportation and contractions // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 90–99.

Куржанский А.Б., Куржанский А.А., Варайя П. Роль макро моделирования в активном управлении транспортной сетью // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 100–118.

Kurzhan'ski A.B., Kurzhanski A.A., Varaiya P. Role of macromodeling in network active traffic management // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 100–118.

Нурминский Е.А., Шамрай Н.Б. Прогнозное моделирование автомобильного трафика Владивостока // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 119–129.

Nurminski E.A., Shamray N.B. Forecast modeling of car traffic in Vladivostok // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 119–129.

Райгородский А.М. Модели случайных графов и их применения // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 130–140.

Raigorodskii A.M. Models of random graphs and their applications // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 130–140.

Смирнов Н.Н., Киселев А.Б., Никитин В.Ф., Кокорева А.В. Математическое моделирование движения автотранспортных потоков методами механики сплошной среды. Двухполосный транспортный поток: модель Т-образного перекрестка, исследование влияния перестроений транспортных средств на пропускную способность участка магистрали // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 141–151.

Smirnov N.N., Kiselev A.B., Nikitin V.F., Kokoreva A.V. Mathematical modeling of traffic flows by a continua approach. Two lanes roads: modeling T-shape crossing and the effect of lane change on handling capacity // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 141–151.

Холодов Я.А., Холодов А.С., Гасников А.В., Морозов И.И., Тарасов В.Н. Моделирование транспортных потоков — актуальные проблемы и перспективы их решения // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 152–162.

Kholodov Y.A., Kholodov A.S., Gasnikov A.V., Morozov I.I., Tarasov V.N. Traffic flow modeling — modern problems and solution methods // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 152–162.

Четверушкин Б.Н., Трапезникова М.А., Фурманов И.Р., Чурбанова Н.Г. Макро- и микроскопические модели для описания движения автотранспорта на многополосных магистралях // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 163–168.

Chetverushkin B.N., Trapeznikova M.A., Furmanov I.R., Churbanova N.G. Macro- and microscopic models for vehicular traffic description on multilane motorways // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 163–168.

Швецов В.И. Проблемы моделирования передвижений в транспортных сетях // Труды МФТИ. — 2010. — Т. 2, № 4. — С. 169–179.

Shvetsov V.I. Problems of the development of complex models of transportation systems // Proceedings of MIPT. — 2010. — V. 2, N. 4. — P. 169–179.

Содержание

<i>Афанасьева Л.Г., Булинская Е.В.</i> Математические модели транспортных систем, основанные на теории очередей	6
<i>Бланк М.Л.</i> Синхронно обновляемые процессы с запретами в моделях транспортных потоков	22
<i>Бобрин П.П.</i> Обоснование гравитационной модели транспортных корреспонденций при помощи закона убывающей предельной полезности	31
<i>Бугаев А.С., Буслаев А.П., Таташев А.Г., Яшина М.В.</i> Оптимизация частично-связных потоков в детерминированно-стохастической модели	35
<i>Гасников А.В., Гасникова Е.В.</i> О возможной динамике в модели расчета матрицы корреспонденций (А.Дж. Вильсона)	45
<i>Гасникова Е.В., Дорн Ю.В.</i> О стохастической марковской динамике, приводящей к равновесию Нэша–Вардроп в модели распределения потоков	55
<i>Замятин А.А., Малышев В.А.</i> Транспортные автомобильные потоки — введение в вероятностный подход	58
<i>Кленов С.Л.</i> Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий	75
<i>Колесников А.В.</i> Транспортировка масс и сжимающие отображения	90
<i>Куржанский А.Б., Куржанский А.А., Варайя П.</i> Роль макро моделирования в активном управлении транспортной сетью	100
<i>Нурминский Е.А., Шамрай Н.Б.</i> Прогнозное моделирование автомобильного трафика Владивостока	119
<i>Райгородский А.М.</i> Модели случайных графов и их применения	130
<i>Смирнов Н.Н., Киселев А.Б., Никитин В.Ф., Кокорева А.В.</i> Математическое моделирование движения автотранспортных потоков методами механики сплошной среды. Двухполосный транспортный поток: модель Т-образного перекрестка, исследование влияния перестроений транспортных средств на пропускную способность участка магистрали	141
<i>Холодов Я.А., Холодов А.С., Гасников А.В., Морозов И.И., Тарасов В.Н.</i> Моделирование транспортных потоков — актуальные проблемы и перспективы их решения	152
<i>Четверушкин Б.Н., Трапезникова М.А., Фурманов И.Р., Чурбанова Н.Г.</i> Макро- и микроскопические модели для описания движения автотранспорта на многополосных магистралях	163
<i>Швецов В.И.</i> Проблемы моделирования передвижений в транспортных сетях	169
<hr/>	
Summaries of all articles	180
Сведения об авторах статей	184
Ссылки на опубликованные статьи	186

*Специальный выпуск, посвящённый математическому моделированию транспортных потоков,
инициирован факультетом управления и прикладной математики МФТИ
и поддержан инновационным центром МФТИ.*

Научный редактор: *вице-президент РАН, академик Валерий Васильевич Козлов*

Ответственные за выпуск:

А.П. Буслаев, А.В. Гасников, Я.А. Холодов, Б.Н. Четверушкин, А.А. Шананин

Редактор: *В.А. Дружинина*

Корректоры: *О.П. Котова, Л.В. Себова, И.А. Волкова*

Компьютерная вёрстка: *А.В. Чудновский*

Электронная коррекция, переписка с авторами: *Е.Ю. Чиркина*

Формирование электронных изданий: *Е.А. Аникушкина*

Журнал зарегистрирован в Федеральной службе по надзору
в сфере связи, информационных технологий и массовых коммуникаций.
Свидетельство о регистрации ПИ № ФС 77-35274 от 10.02.2009

Подписной индекс 47078 в каталоге ОАО Агентства «Роспечать»

Журнал включен в Российский индекс научного цитирования (РИНЦ)
www.elibrary.ru

Подписано в печать 04.12.2010. Формат 60 × 84 ¹/₈. Бумага офсетная.
Печать офсетная. Усл. печ. л. 27,0. Тираж 500 экз. Заказ №

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

141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9
Тел.: (495) 408-41-11, факс: (495) 408-58-22
<http://www.mipt.ru/nauka/trudy/>
zhurnaltrudy@mail.ru

Отпечатано в полном соответствии с качеством предоставленных диапозитивов
Отдел автоматизированных издательских систем «ФИЗТЕХ-ПОЛИГРАФ»
141700, Московская обл., г. Долгопрудный, Институтский пер., д. 9
e-mail: rio@mail.mipt.ru



Московский физико-технический институт
(государственный университет)

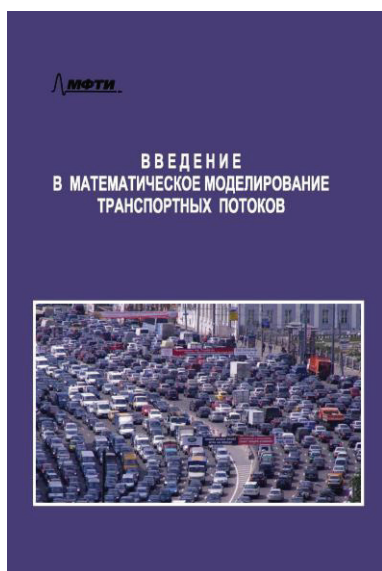
представляет учебную литературу

ISBN 5-7417-0334-2

361 с.

60×84/16

2010



Введение в математическое моделирование транспортных потоков / 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., Малышев В.А., Колесников А.В., Райгородский А.М.; Под ред. А.В. Гасникова. Учебное пособие. Гриф УМО

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

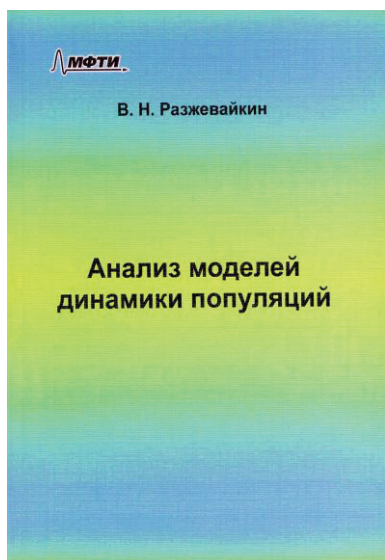
Предназначено для студентов старших курсов и аспирантов физико-математических специальностей МФТИ, МГУ им. М.В. Ломоносова. Рекомендуется научным работникам, интересующимся вопросами математического моделирования.

ISBN 5-7417-0365-6

175 с.

60×84/16

2010



Разжевайкин В.Н.

Анализ моделей динамики популяций.
Учебное пособие. Гриф УМО

Изложены методы анализа моделей динамики биологических популяций, находящие наиболее широкое применение в приложениях к задачам экологии. Представлены предметно ориентированные основы исследования параметрических систем обыкновенных дифференциальных уравнений, а также дискретных и непрерывных моделей, учитывающих возрастную структуру.

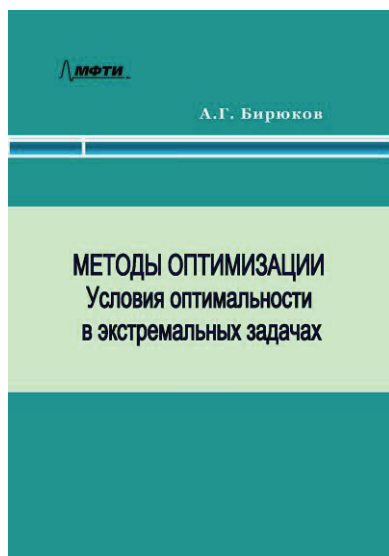
Предназначено для студентов старших курсов университетов, аспирантов и научных сотрудников.

ISBN 5-7417-0367-0

225 с.

60×84/16

2010



Бирюков А.Г.

Методы оптимизации. Условия оптимальности в экстремальных задачах. Учебное пособие. Гриф УМО

Основное внимание уделено вопросам анализа необходимых и достаточных условий оптимальности экстремальных задач конечного числа переменных. Для усвоения материала достаточно владения стандартными курсами линейной алгебры и математического анализа.

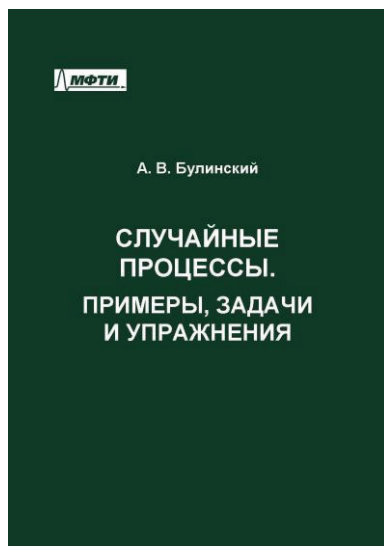
Для студентов университетов, аспирантов и научных работников в области прикладной математики.

ISBN 978-5-7417-0331-1

264 с.

60×84/16

2010



Булинский А.В.

Случайные процессы. Примеры, задачи и упражнения. Учебное пособие. Гриф УМО

Кратко изложен теоретический материал, включая необходимые сведения по теории вероятностей. Разобраны примеры и задачи по основным разделам курса «Стохастические процессы», читаемого автором студентам МФТИ. Также предлагаются упражнения для самостоятельной работы, приведена программа курса и два контрольных задания.

Предназначено для студентов, изучающих теорию случайных процессов.