# Введение в математическое моделирование транспортных потоков

Под редакцией А. В. Гасникова

Предисловие руководителя Департамента транспорта и развития дорожно-транспортной инфраструктуры г. Москвы, заместителя мэра г. Москвы М. С. Ликсутова

Издание второе, исправленное и дополненное

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

Москва
Издательство МЦНМО
2013

УДК 519.1, 519.2, 519.6, 519.8(075)
ББК 22.1я73
В24

Авторский коллектив: А. В. Гасников, С. Л. Кленов, Е. А. Нурминский, Я. А. Холодов, Н. Б. Шамрай.

Приложения: М. Л. Бланк; К. В. Воронцов, Ю. В. Чехович; Е. В. Гасникова; А. А. Замятин, В. А. Малышев; А. В. Колесников; Ю. Е. Нестеров, С. В. Шпирко; А. М. Райгородский. Практическое приложение: А. В. Прохоров, В. Л. Швецов. Вступительное слово: руководитель Департамента транспорта г. Москвы М. С. Ликсутов

Рецензенты:

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

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

Научный консультант академик А. А. Петров

В24 **Введение в математическое моделирование транспортных потоков**: Учебное пособие / Издание 2-е, испр. и доп. А. В. Гасников и др. Под ред. А. В. Гасникова. — М.: МЦНМО, 2013. — ??? с.

ISBN 978-5-4439-0040-7

В книге излагается математический аппарат и некоторые физические концепции, которые могут пригодиться при создании (модернизации) интеллектуальной транспортной системы (ИТС).

Первое издание вышло в 2010 году в издательстве МФТИ. В настоящее второе издание среди прочего были добавлены материалы практического характера от компаний «А+С КонсалтПроект» (PTV Vision ®), «Яндекс.Пробки».

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

ББК 22.1я73

**ISBN 978-5-4439-0040-7**



# Оглавление

**Вступительное слово** руководителя Департамента транспорта и развития дорожно-транспортной инфраструктуры г. Москвы М. С. Ликсутова . . . . . **6**
**Предисловие к новому изданию** . . . . . **10**
**Предисловие** . . . . . **12**
**Введение** . . . . . **19**

Глава 1. **Исследование транспортных потоков с помощью теории экономического равновесия** . . . . . **25**

Введение . . . . . 25
1.1. Задача транспортного равновесия . . . . . 26
1.1.1. Моделирование транспортных потоков как задача принятия решений . . . . . 26
1.1.2. Формализация проблемы . . . . . 27
1.1.3. Сведение к вариационному неравенству . . . . . 30
1.1.4. Разрешимость задач транспортного равновесия . . . . . 33
1.1.5. Симметричные задачи транспортного равновесия . . . . . 36
1.2. Построение функций транспортных затрат . . . . . 37
1.2.1. Аддитивные функции затрат . . . . . 38
1.2.2. Неаддитивные функции затрат . . . . . 40
1.2.3. Модель стационарной динамики . . . . . 41
1.3. Соотношение между системным оптимумом и конкурентным равновесием . . . . . 42
1.4. Численные методы решения задач транспортного равновесия . . . . . 48
1.4.1. Проекционные методы решения задачи транспортного равновесия 49
1.4.2. Декомпозиция проекционных методов для поиска равновесных потоков . . . . . 51
1.4.3. Проекционный метод с генерацией маршрутов . . . . . 52
1.4.4. Ступенчатая регулировка шага проекционного метода . . . . . 55
1.5. Построение матрицы корреспонденций . . . . . 58
1.5.1. Гравитационная модель . . . . . 58
1.5.2. Энтропийная модель . . . . . 60
1.5.3. Связь между гравитационной и энтропийной моделями . . . . 64
1.6. Парадоксы транспортного равновесия . . . . . 65
1.6.1. Парадокс Браесса . . . . . 65
1.6.2. Транспортно-экологические парадоксы . . . . . 68
1.7. Практическая работа . . . . . 72

Литература . . . . . 75

Глава 2. **Математические модели транспортных потоков** . . . . . **79**

2.1. Макроскопические модели . . . . . 79
2.1.1. Модель Лайтхилла—Уизема—Ричардса (LWR) . . . . . 79
2.1.2. Модель Танака . . . . . 89
2.1.3. Модель Уизема . . . . . 90
2.1.4. Модель Пейна и ее обобщения . . . . . 98
2.1.5. Кинетические модели . . . . . 103
2.1.6. Практические приложения моделей . . . . . 105
2.2. Микроскопические модели . . . . . 107
2.2.1. Модель оптимальной скорости Ньюэлла . . . . . 107
2.2.2. Модель следования за лидером «Дженерал Моторс» . . . . . 112
2.2.3. Модель Трайбера «разумного водителя» . . . . . 113
2.2.4. Модели клеточных автоматов . . . . . 116
2.3. Модельные задачи . . . . . 122
2.3.1. Эволюции глобального затора в транспортном потоке, описываемом моделями LWR и Уизема . . . . . 122
2.3.2. Эволюции локального затора в транспортном потоке, описываемом моделями LWR и Уизема . . . . . 135
2.3.3. Задача о светофоре: при каких условиях перед светофором не будет скапливаться очередь . . . . . 141
2.4. Теория Кернера—Конхойзера движущихся локальных кластеров в моделях класса «Дженерал Моторс» . . . . . 143
2.4.1. Фундаментальные эмпирические свойства перехода от свободного транспортного потока к плотному и модели транспортного потока . . . . . 144
2.4.2. Характеристические параметры широкого движущегося кластера 147
2.4.3. Линия J Кернера . . . . . 150
Литература . . . . . 152

Глава 3. **Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий 162**

3.1. Три фазы транспортного потока . . . . . 164
3.1.1. Предварительные сведения . . . . . 164
3.1.2. Свободный транспортный поток — фаза F . . . . . 165
3.1.3. Плотный транспортный поток . . . . . 165
3.1.4. Определение фаз J и S в плотном транспортном потоке . . . . 166
3.1.5. Возникновение плотного потока — фазовый переход F → S . . 168
3.1.6. Бесконечное число значений пропускных способностей скоростной автомагистрали . . . . . 172
3.1.7. Широкие движущиеся кластеры (локальные движущиеся заторы) — фаза J . . . . . 175
3.1.8. Синхронизированный транспортный поток — фаза S . . . . . 176
3.1.9. Фазовый переход S → J . . . . . 177



3.1.10. Неоднородные пространственно-временные структуры транспортного потока, состоящие из фаз S и J . . . . . 178
3.2. Стохастические модели в рамках теории трех фаз Кернера . . . . . 180
3.2.1. Стохастическая микроскопическая трехфазная модель транспортного потока . . . . . 180
3.2.2. Моделирование свойств пространственно-временных структур в транспортном потоке вблизи въезда на скоростную автомагистраль . . . . . 185
3.2.3. Трехфазная модель клеточных автоматов для транспортного потока (KKB-модель) . . . . . 188
3.2.4. Новая трехфазная модель клеточных автоматов для транспортного потока (ККШ-модель) . . . . . 190
3.3. Применение теории трех фаз Кернера для интеллектуальных транспортных технологий . . . . . 192
Литература . . . . . 192

**Приложения** . . . . . **200**

*М. Л. Бланк*. Процессы с запретами в моделях транспортных потоков . . 200
*К. В. Воронцов*, *Ю. В. Чехович*. Интеллектуальный анализ данных в задачах моделирования транспортных потоков . . . . . 225
*Е. В. Гасникова*. О возможной динамике в модели расчета матрицы корреспонденций . . . . . 249
*А. А. Замятин*, *В. А. Малышев*. Введение в стохастические модели транспортных потоков . . . . . 272
*А. В. Колесников*. Транспортная задача и концентрация . . . . . 306
*Ю. Е. Нестеров*, *С. В. Шпирко*. Стохастическое транспортное равновесие 316
*А. М. Райгородский*. Модели случайных графов и их применения . . . . 327

**Задачи** . . . . . **351**

Задачи к главам пособия и приложениям . . . . . 351
Задача Штейнера и задачи на графах транспортных сетей . . . . . 380
Задачи от «Яндекс.Пробки» . . . . . 393

**Исследовательские вычислительные задачи, предлагавшиеся в 2011 г.** . . **400**

**Практическое приложение** . . . . . **417**

*А. В. Прохоров*, *В. Л. Швецов*. О практическом опыте моделирования транспортных потоков с помощью пакета программ PTV Vision ® . . . . 417

# Вступительное слово

Транспорт — одна из ключевых систем городского организма, которую по важности уместно сравнить с кровоснабжением. Именно транспорт позволяет городу в полной мере выполнять связующую, коммуникационную и обеспечивающую функции. Тема транспорта касается практически каждого городского жителя, и тем важнее становятся усилия по систематизации и распространению соответствующих знаний.

Для управления дорожным движением на транспортной сети городов повсеместно используются системы управления, алгоритмы работы которых основаны на моделях транспортных потоков. Требования к точности и сложности моделей чрезвычайно велики. Достаточно сказать, что на простейшем перекрестке может быть 12 направлений движения транспортных средств. Для участка улично-дорожной сети с 10 такими пересечениями речь идет уже о 120 направлениях и необходима минимизация задержек по каждому из этих направлений при условии, что интенсивность движения постоянно изменяется во времени и в пространстве.

Кроме того, без транспортного моделирования невозможно планирование строительства новых и модернизации существующих транспортных объектов, объектов жилищного и делового строительства, схем организации дорожного движения, действий при чрезвычайных ситуациях, решение целого ряда других практических задач.

Вниманию читателей предлагается издание, в котором собраны фундаментальные знания в области математического моделирования транспортных потоков, что позволяет рассматривать это издание как основу подготовки специалистов самой высокой квалификации для многих сфер городского хозяйства.

Одним из самых значительных применений транспортных моделей на перспективу не менее 10–15 лет будет проектирование интеллектуальных транспортных систем (ИТС), необходимость создания которых обусловлена коренными изменениями условий дорожного движения и задачами управления дорожным движением, вызванными, в свою очередь, интенсивным, взрывным ростом уровня автомобилизации.

С точки зрения управления движением при малой загрузке улично-дорожной сети (20–30% пропускной способности) движение фактически является свободным и управление сводится к локальному светофорному регулированию, которое вводится по критериям безопасности. Потребностей в применении каких-либо моделей и алгоритмов управления на их основе практически не возникает.

Интервал в 20–70% загрузки от пропускной способности улично-дорожной сети (УДС) — сфера традиционных автоматизированных систем управления дорожным движением (АСУДД), когда ставится и решается задача увеличения пропускной способности за счет координированного управления светофорной сигнализацией.



Основное физическое явление, за счет которого достигается выигрыш при таком управлении — формирование и пропуск пачки (группы) транспортных средств. Если пачка «рассыпается» (при дистанции между светофорами 800–1000 м и более), выигрыш за счет координированного светофорного регулирования фактически не достигается. Отсюда вытекают требования к количеству светофорных объектов, к алгоритмам координации и, соответственно, к моделям.

При загрузке в 80% и более задача управления принципиально меняется. Любая перегрузка улично-дорожной сети сверх пропускной способности приводит к фатальным последствиям.

МКАД (Московская кольцевая автомобильная дорога), например, спроектирована на пропуск не менее 8–10 тысяч единиц транспорта в час. Но в настоящее время каждый день имеются участки, где пропускная способность падает до двух тысяч и менее единиц (ниспадающая ветвь фундаментальной диаграммы транспортного потока). При этом однажды возникшая пробка подвержена явлению гистерезиса, которое состоит в том, что как только трафик превышает пропускную способность полосы, движение входит в нестабильную зону функционирования в загруженном режиме и вернуться к свободному движению возможно лишь после того, как «спрос» станет явно ниже пропускной способности. Никакие алгоритмы координации (и соответствующие потоковые модели) в условиях перегрузки положительного результата не дают.

Эффективное управление дорожным движением в этих условиях должно обеспечивать загрузку транспортной сети на грани ее пропускной способности и поддерживать непрерывное равномерное движение — пусть даже на относительно небольших скоростях (оптимальные режимы движения при такой загрузке собственно и соответствуют синхронизированной фазе по Кернеру).

Иными словами, задача пропуска возможно большего количества транспортных средств меняется на задачу достижения транспортного баланса между реальной пропускной способностью УДС и спросом на объемы движения при максимальном использовании возможностей, предоставляемых геометрическими параметрами уличной сети.

Такая постановка задачи принципиально меняет построение системы управления, алгоритмы управления и, соответственно, модели, лежащие в их основе.

Прежде всего, необходимо понимать, что пропускная способность — величина переменная, зависящая от погодных условий, аварий, производства ремонтных работ и т. д. При загрузке до 60–70% имеется резерв,

который сглаживает изменения пропускной способности. При загрузке в 90% такого резерва нет и интеллектуальная транспортная система должна в реальном масштабе времени оценивать текущую пропускную способность и перераспределять потоки.

Далее, в условиях интенсивной автомобилизации достижение транспортного баланса невозможно без введения и поддержания механизмов ограничения спроса на дорожное движение за счет информирования участников движения о загрузке УДС и возможных маршрутах движения, без развития общественного транспорта, грамотной логистики, управления парковочным пространством, перераспределения транспортных потоков в зависимости от складывающихся условий и т. д. — вплоть до введения административных запретов.

Например, в Москве в настоящее время на улицах города в «часы пик» могут одновременно находиться в движении не более 400 тысяч автомобилей. Как показывают результаты наблюдения, город «встает», когда число выехавших на дороги автомобилей достигает 500 тысяч, а при экстремальных погодных условиях — и при меньшей численности. Данную цифру можно рассматривать как предельную пропускную способность улично-дорожной сети города.

Между тем по экспертным оценкам при благоприятных условиях движения в часы пик готовы выехать на улицы около четверти от общего числа транспортных средств. Учитывая, что в Москве с пригородами насчитывается более 5 млн единиц транспорта, общий спрос на передвижение в часы пик можно оценить в 1 млн единиц транспорта, что в два раза превышает пропускную способность улично-дорожной сети. Каждый второй водитель уже не садится за руль только из-за неудовлетворительных условий движения и парковки.

На перспективу до 2016 года прогнозируется, что численность автотранспортных средств в Московском регионе может увеличиться до 8 млн единиц, т. е. спрос может четырехкратно превысить пропускную способность.

Примерно такая же ситуация, обусловленная интенсивной автомобилизацией в последние 10–15 лет, складывается в большинстве крупных и средних городов страны. В реально сложившихся условиях никакое наращивание дорожно-мостового строительства и принятие локальных мер не позволит удовлетворить «отложенный спрос».

Именно поэтому в современных условиях задача пропуска возможно большего числа транспортных средств по УДС меняется на задачу поддержания транспортного баланса между пропускной способностью существующей улично-дорожной сети и ее реальной загрузкой за счет перераспределения (а при необходимости — за счет введения ограничений на движение) транспортных потоков.



Несомненно, что инженерная реализация рассматриваемых моделей в виде реально работающих автоматизированных систем требует своих исследователей и разработчиков.

Но тот факт, что относительно небольшой набор математических идей «делает погоду» в разных областях транспортной науки, позволяет утверждать, что изучение собранных в монографии материалов может быть рекомендовано как ученым, так и инженерам, применяющим в сфере своей деятельности аппарат математического моделирования транспортных потоков.

*Максим Станиславович Ликсутов*
Руководитель Департамента транспорта и развития
дорожно-транспортной инфраструктуры города Москвы
12 апреля 2012 г.

# Предисловие к новому изданию

В 2010 г. в издательстве МФТИ небольшим тиражом (250 экз.) вышла наша книга «Введение в математическое моделирование транспортных потоков». Книга оказалась довольно востребованной и практически сразу стала библиографической редкостью. Предложение о подготовке нового издания поступило к нам от И. В. Ященко и В. В. Фурина летом 2011 г. во время ЛШСМ-2011. На этой замечательной летней школе мы как раз рассказывали продвинутым школьникам ряд сюжетов из первого издания книги. В новом издании была учтена «обратная связь» от школьников ЛШСМ-2011 и от студентов, слушавших одноименный (с названием книги) курс лекций в весеннем семестре 2011 г. в Независимом московском университете.

По сравнению с прошлым изданием в новом издании были пополнены материалы главы 1, главы 3, приложения М. Л. Бланка, Е. В. Гасниковой, А. В. Колесникова, А. М. Райгородского и задачного раздела. В новое издание также были добавлены приложения:

- «Интеллектуальный анализ данных в задачах моделирования транспортных потоков», написанное известными специалистами в области интеллектуального анализа данных профессором К. В. Воронцовым (ВЦ РАН, МФТИ, «Яндекс») и к.ф.-м.н. Ю. В. Чеховичем (ВЦ РАН, «Форексис»);
- «Стохастическое транспортное равновесие», написанное известными специалистами в области численных методов выпуклой оптимизации профессором Ю. Е. Нестеровым (CORE/INMA (UCL), ПреМоЛаб МФТИ) и к.ф.-м.н. С. В. Шпирко (ПреМоЛаб МФТИ);
- «О практическом опыте моделирования транспортных потоков с помощью пакета программ PTV Vision», написанное А. В. Прохоровым и В. Л. Швецовым («A+C КонсалтПроект») на основе собственного опыта выполнения различных транспортных проектов в ряде городов России.

Небольшие изменения были внесены и в другие части. За время, прошедшее с момента выхода первого издания, к нам обращались за консультациями по вопросам математического моделирования транспортных потоков разнообразные организации. Большинство задач, по которым требовалась консультация, имели интересную научную составляющую. Поэтому было решено включить в новое издание ряд таких исследовательских вычислительных задач, знакомство с которыми, на наш взгляд, значительно способствует лучшему усвоению основного материала.



Нельзя не отметить большой труд редакционного характера, затраченный при подготовке первого издания В. Н. Тарасовым, В. А. Дружининой, И. А. Волковой, О. П. Котовой, Л. В. Себовой. Всем им мы выражаем глубокую благодарность.

Мы также благодарим Юрия Николаевича Торхова, способствовавшего выходу настоящего издания, и Александра Ведерникова за предоставленную фотографию на обложку.

Работа выполнена при поддержке гранта РФФИ и Лаборатории структурных методов анализа данных в предсказательном моделировании (ПреМоЛаб) МФТИ, грант правительства РФ дог. 11 11.G34.31.0073.

*А. В. Гасников* (avgasnikov@gmail.com)
доцент кафедры МОУ ФУПМ МФТИ
с.н.с. ПреМоЛаб МФТИ
5 декабря 2011 г.

# Предисловие

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

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

Курс содержал дополнительные главы следующих дисциплин:

— уравнений математической физики (обобщенные решения законов сохранения, групповой анализ, автомодельная редукция, принципы максимума для квазилинейных параболических уравнений);

— теории вероятностей и случайных процессов (аппарат производящих функций, системы массового обслуживания, концентрация меры, исследование асимптотик с помощью метода перевала);

— функционального анализа (сжимающие отображения, монотонные операторы, конусные методы);

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



— теории динамических систем (методы функционалов Ляпунова) и эргодической теории (концентрация инвариантной меры, элементы статистической физики);

— кинетической теории (уравнения Колмогорова, социодинамика, динамика систем с мотивацией, самоорганизация);

— теории игр (эволюционные игры: равновесие Нэша как устойчивое положение равновесия динамики наилучших ответов);

— оптимизации в конечномерных и бесконечномерных пространствах (принцип Лагранжа, двойственность, отделимость, принцип Беллмана, элементы теории управления);

— дискретной математики (задачи на графах и эффективные (приближенные, вероятностные) алгоритмы их решения);

— численных методов выпуклой оптимизации (прямо-двойственные методы, стохастические субградиентные методы, субградиентные методы для задач огромной размерности и др.).

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

По сути, речь идет о том, как оптимальным образом использовать имеющуюся информацию. Например, в Москве сейчас установлено (в основном на крупных перекрестках) в общей сложности более 500 видеокамер и несколько тысяч различных детекторов. Порядка $10^5$ автомобилей [1)], курсирующих по Москве и области, оснащены GPS-навигаторами, что позволяет получать треки (пути следования) автомобилей с информацией о скоростях движения вдоль этих треков. Заметим, что всего в Москве ежедневно бывает более $4 \cdot 10^6$ автомобилей.

Создание КИТС на основе имеющихся данных предполагает выполнение следующих действий.

- Выработку адекватной (имеющимся данным [2)]) математической модели, описывающей транспортный поток. Например, можно в неплохом приближении уподоблять транспортный поток сжимаемой жидкости с мотивацией и использовать гидродинамические модели (или модели клеточных автоматов). Калибровка таких моделей на прямолинейных участках

[1)] Если говорить о сечении по времени, то таких автомобилей на дорогах будет на порядок меньше.

[2)] Важно подобрать модель, адекватную имеющимся данным, дабы не «забивать микроскопом гвозди».

дороги (ребрах графа транспортной сети) довольно просто осуществляется исходя из имеющейся информации.

• Для постановки начально-краевых условий: описание характеристик источников и стоков автомобилей, узлов графа транспортной сети (перекрестки, въезды, съезды и т. п.) — также требуется работа с накопленными данными. В результате такой работы получается матрица корреспонденций, на основе этой матрицы рассчитываются распределения потоков, а затем и матрицы перемешивания в узлах графа транспортной сети. На наш взгляд, адекватная постановка начально-краевых условий — это одна из самых сложных текущих задач. И наметки приведенного здесь пути — далеко не единственный способ получения краевых условий.

• Откалиброванная модель может использоваться для локального (по времени) управления на основе текущей информации, например, светофорной сигнализацией (въездами на крупные магистрали). Возникающие здесь задачи связаны с управлением сложными (гибридными) динамическими системами в условиях неопределенности. Такое локальное управление позволит несколько разгрузить складывающуюся на дорогах в данный момент ситуацию (правильным образом распределяя ресурсы транспортной сети между ее пользователями). Например, в Калифорнии коллектив, работающий в Беркли и возглавляемый П. Варайя и А. Б. Куржанским, предложил несколько лет назад способ локального управления въездами на основные магистрали. Это привело к тому, что для среднестатистического водителя время в пути уменьшилось на 30%.

• Помимо задач локального управления имеются задачи долгосрочного управления. Где и какую дорогу следует построить при заданных бюджетных ограничениях? Каким образом (в каком размере) взимать плату за проезд с трасс в центре Москвы? [1)] На каких трассах стоит в первую очередь увеличивать число полос? Где стоит в первую очередь переделывать развязки или делать новые (в частности, решать вопрос: а выгодно ли увеличивать степень непланарности графа транспортной сети)? Эти задачи, так же как и задачи предыдущего пункта, должны решаться для всей сети в целом (не локально по пространству!). Иначе говоря, сумматорные функционалы качества критериев должны в себя включать всех участников дорожного движения. Понятно, что для решения этих задач достаточно просматривать различные сценарии (в том числе предложенные руководством города) с помощью выработанной и откалиброванной модели на предмет их состоятельности, путем (разумного) перебора выбирать лучшие предложения по разгрузке ситуации на дорогах.

[1)] Нужна некая «золотая середина»: с одной стороны, плата за проезд должна разгрузить эти трассы, с другой — желательно, чтобы пропускные ресурсы трасс использовались по максимуму.



Рис. 1. Транспортный затор на одной из улиц Москвы

Некоторые рассмотренные в пособии задачи имеют также и коммерческий выход. Например, актуальной в последнее время задачей [1)] является *задача маршрутизации*: выбор оптимального (кратчайшего) маршрута следования. Понятно, что если считать веса ребер графа транспортной сети известными и не меняющимися со временем, то эта задача довольно эффективно решается. Но на практике далеко не всегда все нужные веса ребер бывают известными. В зависимости от времени суток ситуация на дорогах может кардинально меняться, поэтому возникает необходимость прогнозирования загрузки элементов сети. Примером таких изменений служит образование заторов в часы пик — за короткий промежуток времени движение может быть практически парализовано даже на многополосных магистралях (рис. 1).

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

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

выбранном уровне детализации и объеме излагаемого материала вполне уместно.

Конечно, представленный в книге материал далеко не полон [1]. Причина проста — колоссальный объем накопившегося на данный момент материала, посвященного транспортной проблематике. Достаточно сказать, что сейчас в мире существуют десятки реферируемых научных журналов, в которых регулярно публикуются материалы на транспортную тематику. Упомянем лишь некоторые из них: «Transportation Research B», «Physical Review E», «Review of modern physics», «Transportation Science», не говоря уже об электронных ресурсах, таких, например, как `http://arxiv.org/`. Раз в два года проводится крупнейшая в транспортном сообществе конференция по математическому моделированию транспортных потоков и смежным вопросам: «Traffic and granular flow», труды которой публикует известное немецкое издательство Springer. Кстати, в 2011 г. эта конференция впервые прошла в Москве (см. `http://tgf11.ru`). Четвертый номер журнала «Труды МФТИ» за 2010 г. под редакцией вице-президента РАН академика В. В. Козлова всецело посвящен транспортной проблематике. Однако, несмотря на вышесказанное, мы все же постарались собрать наиболее базовые (математически) вещи и описать текущее состояние дел.

**Об авторах.** Во многом определяющим моментом в создании этого пособия стало участие в его написании ряда ведущих специалистов в своих областях. Так, глава 1 написана профессором Е. А. Нурминским и доцентом Н. Б. Шамрай (ИАПУ ДВО РАН) и посвящена применению теории бескоалиционного равновесия для расчета транспортной сети при условии стационарности потоков и моделям построения матрицы корреспонденций. Глава 2 написана А. В. Гасниковым при участии доцента С. Л. Кленова (кафедра общей физики МФТИ) и доцента Я. А. Холодова (кафедра вычислительной математики МФТИ). С. Л. Кленовым был написан раздел 2.4, а Я. А. Холодов принял участие в написании пунктов 2.1.4 и 2.2.4. Глава 3, посвященная теории трех фаз Кернера транспортного потока, всецело написана С. Л. Кленовым (коллегой Б. С. Кернера) и содержит как упомянутые выше физические концепции, так и примеры эмпирических (измеренных) пространственно-временных структур плотного потока на скоростных автомагистралях. Как показала обратная связь от студентов, слушавших упомянутый выше курс по выбору, востребованными оказались «стохастические» приложения доцента А. А. Замятина и профессора В. А. Малышева (кафедра теории вероятностей мехмата МГУ) и профессора А. М. Райгородского (кафедра математической статистики и случайных процессов мехмата МГУ, кафедра анализа данных МФТИ). Важную роль в пособии играют эргодические приложения профессора М. Л. Бланка (лаборатория Р. Л. Добрушина ИППИ РАН), аспирантки Е. В. Гасниковой (кафедра анализа систем и решений ФУПМ МФТИ) и приложение д.ф.-м.н. А. В. Колесникова (математический факультет ВШЭ), посвященное связи задачи Монжа—Канторовича о перемещении масс и явления концентрации меры. Эти три приложения, помимо того что представляют самостоятельную ценность, также завязывают (математически) между собой многие темы этого пособия. Другими словами, знакомство с ними желательно для формирования целостного восприятия.

В конце учебного пособия приведены задачи, часть из которых в разное время предлагалась студентам [1]. При подготовке задач большую помощь оказали молодые ученые, работающие в близких направлениях. В пособии имеется целый раздел задач (написанный ассистентом кафедры МОУ ФУПМ МФТИ Е. Г. Молчановым), посвященный задачам на графах и, по сути, восполняющий нехватку в пособии темы «Транспортные потоки и Computer Science». В этом же разделе приводится довольно интересная задача, пришедшая из практических приложений от службы «Яндекс.Пробки».

**Благодарности.** В заключение хотелось бы выразить благодарность профессору А. А. Шананину, академику В. В. Козлову, академику А. А. Петрову, члену-корреспонденту А. С. Холодову. Общение и участие в мероприятиях, к которым они имели отношения, всегда доставляло большое удовольствие и иногда вдохновляло на улучшение материала данной книги. Ценную обратную связь при подготовке этого пособия получал от профессора А. П. Буслаева и доцента О. С. Розановой, а также от всех коллег, принимавших участие в его написании. Много полезных замечаний по всему тексту сделал самый активный слушатель курса — Ю. В. Дорн (студент 6-го курса ФАКИ МФТИ). Ряд ценных замечаний по темам, изложенным в пособии, сделали С. Я. Аввакумов, В. И. Аркин, Л. Г. Афанасьева, М. А. Бабенко, П. П. Бобрик, А. С. Бугаев, Е. В. Булинская, А. М. Валуев, Н. Д. Введенская, В. В. Веденяпин, И. Е. Виноградов, К. А. Волосов, А. Э. Воробьев, В. В. Вьюгин, А. И. Голиков, А. Н. Дарьин, К. Дафермос, А. В. Дмитрук, Е. Г. Дорогуш, В. А. Дородницын, В. А. Дружинина, В. Г. Жадан, А. В. Казейкина, Б. С. Кернер, А. В. Козлов, В. Ф. Колчин, Н. С. Кукушкин, А. Г. Куликовский, А. А. Куржанский, Г. Л. Литвинов, И. А. Лубашевский, А. Е. Макаров, В. П. Мар-

[1] Например, очень мало внимания уделяется важному на практике четырехстадийному способу моделирования тарнспортных потоков. Заинтересованному в практических применениях читателю мы рекомендуем книгу: *Ortúzar J. D.*, *Willumsen L. G.* Modelling transport. John Willey & Sons, 2011.

[1] Упражнения и задачи повышенной сложности отмечены звездочкой. Те задачи, полные решения которых авторам неизвестны, помечены двумя звездочками. В разделе «Исследовательские задачи» и «Задачи от „Яндекс.Пробки“» все задачи повышенной сложности.

тынов, И. С. Меньшиков, В. Д. Мильман, И. И. Морозов, Т. А. Нагапетян, А. И. Назаров, А. В. Назин, А. С. Немировский, Т. С. Обидина, В. И. Опойцев, В. П. Осипов, Я. С. Панасюк, Е. Ю. Панов, Д. И. Петрашко, Н. С. Петросян, С. А. Пирогов, В. М. Полтерович, Б. Т. Поляк, Ю. С. Попков, И. Г. Поспелов, В. Ю. Протасов, В. В. Пухначёв, В. Н. Разжевайкин, И. В. Рублев, А. Н. Рыбко, В. Ж. Сакбаев, А. Ю. Семёнов, Д. Серр, Н. Н. Смирнов, А. Н. Соболевский, П. А. Солоневич, В. Г. Спокойный, Е. О. Степанов, Н. Н. Субботина, В. Н. Тарасов, С. П. Тарасов, Г. М. Хенкин, Б. Н. Четверушкин, А. П. Чугайнова, С. В. Чуканов, Н. Г. Чурбанова, В. М. Шелкович, А. Шень, В. И. Швецов, М. В. Яшина. Ценным было общение с академиком А. Б. Куржанским и академиком В. П. Масловым.

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

*А. В. Гасников* (avgasnikov@gmail.com)
доцент кафедры МОУ ФУПМ МФТИ
12 ноября 2010 г.

# Введение

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

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

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

Продолжая аналогию с газовой динамикой, И. Пригожин полвека назад (а затем С. Павери-Фонтана, Д. Хельбинг и др.) предложил описывать транспортный поток кинетическим уравнением (типа Больцмана с «ин-

тегралом взаимодействия автомобилей» вместо «интеграла столкновения частиц газа»). При таком подходе макроскопическая модель получается из кинетической подобно тому, как система уравнений Эйлера получается из уравнения Больцмана.

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

Несмотря на то что с момента появления первых фундаментальных работ прошло более полувека, по мнению ряда известных специалистов в области математического моделирования дорожного движения (К. Нагель, Х. Махмасани, М. Шрекенберг и др.), проблема образования предзаторных и заторных ситуаций еще до конца не изучена (и сродни проблеме описания турбулентных течений). Используя терминологию, предложенную Б. С. Кернером, можно сказать, что на данный момент нет общепринятого подхода, описывающего поведение движения автотранспорта в области синхронизированного потока. Иначе говоря, если автомобильный поток уподобляется жидкости, то наиболее сложная для моделирования ситуация — это замерзающая жидкость. Подтверждением вышесказанному может служить тот факт, что разные коллективы, занимающиеся моделированием транспортных потоков, как правило, используют разные модели: начиная от модели Лайтхилла—Уизема (А. А. Куржанский и др.), заканчивая моделями, в которых каждый водитель характеризуется своим вариационным принципом (И. А. Лубашевский и др.). Отметим здесь главу 3, в которой приводится «эмпирический базис», т. е. даются свойства реальных пространственно-временных структур, возникающих в плотном транспортном потоке вблизи «узкого места»,[1)] 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 эмпирических (измеренных) данных.

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



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

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

Из-за сильной неустойчивости (при достаточно больших плотностях) решений уравнений, описывающих транспортные потоки, задача получения достоверного прогноза загрузки транспортной сети по имеющимся данным на час вперед сродни задаче получения достоверного прогноза погоды на неделю вперед. При этом вычислительные мощности современных высокопроизводительных кластеров в терафлопс и выше позволяют просчитывать реальную ситуацию по Москве (в которой, напомним, порядка четырех миллионов автомобилей) со значительным опережением реального времени. Другими словами, основной проблемой при моделировании транспортных потоков является не ограничение по вычислительным мощностям и ресурсам памяти, а большая чувствительность описываемой реальной транспортной системы к входным данным (характеристики источников и стоков автомобилей) и невозможность собрать достаточно полную информацию о входных данных.

Одним из возможных выходов из этого является рассмотрение в некотором смысле усредненных показателей транспортной системы — например, в смысле теории систем массового обслуживания. Обгоны на многополосной дороге, очереди перед светофорами и многое другое можно описывать таким образом — о чем говорится в приложении А. А. Замятина и В. А. Малышева. В основе моделей этого приложения лежат эргодические марковские процессы. При таком подходе исследователь следит лишь за трендом и «не обращает внимания» на высокочастотные случайные колебания (флуктуации), возможно большой амплитуды, вокруг этого тренда. В связи с упоминанием словосочетания *усредненные пока-*

*затели* отметим здесь также приложение А. М. Райгородского, в котором исследуются различные свойства случайных графов (транспортных графов, web-графов), например, такое важное свойство, как надежность графа транспортной сети к случайным отказам ребер (отказ ребра означает, что на ребре образовалась пробка). В этих приложениях наблюдается плотная концентрация исследуемых макровеличин (макропоказателей) в маленьких окрестностях своих математических ожиданий. Однако если в приложении Замятина—Малышева мера, которая концентрируется, порождается (как финальная = стационарная) эргодической марковской динамикой, то в приложении А. М. Райгородского она задается непосредственно, скажем, из соображений независимости и однородности (модель случайного графа Эрдёша—Реньи).

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

Вернемся теперь к тому, как все-таки ставить начально-краевые условия для целостного описания транспортного потока на полном графе транспортной сети. Будем считать, что есть лишь информация о том, сколько людей живет в том или ином районе и сколько рабочих мест есть в том или ином районе. В главе 1 и приложении Е. В. Гасниковой приведены различные способы обоснования известной на практике энтропийной (гравитационной) модели А. Дж. Вильсона (1967) расчета матрицы корреспонденций (сколько людей, проживающих в районе $i$, работают в районе $j$). По сути, матрица корреспонденций определяется как наиболее вероятное макросостояние, в окрестности которого и будет плотная концентрация, стационарной меры «разумной» эргодической марковской динамики, порождающей изучаемую макросистему. Точнее говоря, эргодическая марковская динамика приводит на больших временах к стационарной (инвариантной) пуассоновской (сложной) мере 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 — законами сохранения. Приятной особенностью такого класса задач является явная и легко выписываемая зависимость решения прямой задачи через двойственные переменные. Поскольку число ограничений, как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы базируются на решении двойственной задачи: минимизации выпуклой функции. Отметим, что описанная здесь задача энтропийно-линейного программирования имеет много общего с обычной транспортной задачей.



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

Имея информацию о том, как распределяются потоки, уже можно получать оценки матриц перемешивания в узлах графа транспортной сети, тем самым замыкать целостную модель. К сожалению, такой способ крайне чувствителен к точности (полноте) входных (обучающих) данных.

В приложении А. В. Колесникова рассматривается задача Монжа—Канторовича о перемещении масс [1)], эквивалентная, при весьма общих условиях, задаче Монжа. Оптимальный план перевозок (точнее, потенциал этого отображения) удовлетворяет уравнению в частных производных Монжа—Ампера и порождает метрику Канторовича(—Рубинштейна). С помощью этой метрики устанавливаются довольно тонкие функциональные неравенства о *концентрации меры* (М. Громов, М. Талагран, К. Мартон, М. Леду и др.). Сам термин «концентрация меры», по-видимому, был впервые предложен В. Д. Мильманом, внесшим значительный вклад в эту область. Геометрически этот принцип можно довольно просто пояснить задачей Пуанкаре—Леви (см. с. 376, 377): площадь многомерной сферы (с выделенным северным и южным полюсом) практически полностью сосредоточена в маленькой полоске вокруг экватора. Этот принцип нашел широкое применение, например, в теории вероятностей (нелинейные законы больших чисел — концентрация значений липшицевых функций в окрестности медианы), асимптотической комбинаторике (в частности, при исследовании плотной концентрации различных функций, типа числа независимости, на случайных графах; см. приложение А. М. Райгородского). Явление концентрации меры играет немаловажную роль в понимании ряда

[1)] Сильно связанная с транспортной задачей, о которой написано в главе 1, приложении Е. В. Гасниковой и которая фигурирует также в задачах, приведенных в конце пособия.

материалов пособия: концепции равновесия макросистемы (модель расчета матрицы корреспонденций), исследованиях надежности графа транспортной сети (по модели Эрдёша—Реньи), оценках скорости сходимости к равновесию и др.

# Глава 1

# Исследование транспортных потоков с помощью теории экономического равновесия

## Введение

Одним из наиболее агрегированных способов описания транспортных систем является экономический подход, суть которого заключается в соотнесении интенсивности использования тех или иных технологий, ресурсов и пр. и итогового результата, выражающего конечный выпуск продуктов, оказание определенного объема услуг и т. п. Схематически этот взгляд на экономику отраслей, в том числе и на транспорт, представлен на рис. 1, где $x$ — это усилия, предпринятые для перевозки, нагрузка на транспортную систему, $y$ — объем грузов или количество людей, перевезенные системой. Помимо этого у системы — черного ящика присутствует социально-экономическая оценка $q$ технологического процесса ($x$, $y$).

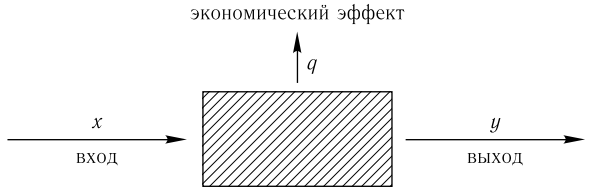



Рис. 1. Представление экономиста о транспорте и не только

Допустимые сочетания затрат $x$ и выпусков $y$ образуют технологическое множество, описанием которого экономист не занимается, в его задачу входит формирование понятия эффективного функционирования системы и отбор и анализ эффективных вариантов.

В данной главе описан один из подходов к моделированию и исследованию транспортных потоков, основанный на теории конкурентного бескоалиционного равновесия, которая позволяет описать достаточно адекватный механизм функционирования автомобильных улично-дорожных сетей (УДС). Будут рассмотрены основные элементы транспортной системы, включающие в себя УДС, факторы, определяющие потребность в перевозках, критерии эффективности транспортной системы и принципы ее функционирования.

Рассматриваемые модели применяются для получения прогнозных оценок загрузки элементов транспортной сети. Подобные задачи интересны

в частности тем, что являются одним из инструментов для объективной оценки эффективности проектов по модификации УДС с точки зрения разгрузки наиболее проблемных участков дорог и улучшения качества транспортного обслуживания.

## 1.1. Задача транспортного равновесия

### 1.1.1. Моделирование транспортных потоков как задача принятия решений

Для определения объемов загрузки УДС в первую очередь необходимо выявить правила, по которым водители выбирают тот или иной маршрут следования. Поведенческие принципы пользователей транспортной сети окончательно были сформулированы в работе [64], где постулировались следующие две возможные ситуации.

1) Пользователи сети независимо друг от друга выбирают маршруты следования, соответствующие их минимальным транспортным расходам.

2) Пользователи сети выбирают маршруты следования исходя из минимизации общих транспортных расходов в сети.

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

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

Стоит отдельно отметить, что первый поведенческий принцип предполагает определенные допущения. Во-первых, это совершенная информированность участников движения о ситуации на дорогах — каждый знает затраты на передвижения по тем или иным маршрутам. В настоящее время такое предположение не выглядит изрядной идеализацией, поскольку интенсивно развиваются и внедряются в практическое использование автоматизированные автонавигаторы и интеллектуальные транспортные системы, идет активное оповещение о ситуациях на дорогах через интернет, радио и другие средства информации. Во-вторых, предполагается ничтожно малое влияние отдельного участника движения на затраты по всем маршрутам. Хотя такое предположение и заведомо неверно для крупногабаритных транспортных средств, для легковых автомобилей оно представляется достаточно разумным, исключая случаи аварийных ситуаций или неопытных водителей за рулем.



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

Несмотря на то что приведенные поведенческие принципы широко цитируются как принципы Вардропа, на самом деле чуть ранее их сформулировали Ф. Найт [48] и А. Пигу [60], утверждая, что все участники движения, направляющиеся из одного узла сети в другой, распределяются по различным маршрутам таким образом, чтобы удельные (в расчете на один автомобиль) затраты на проезд были одни и те же для всех.

В ситуации массовой автомобилизации, имеющей место практически во всех странах, подавляющее большинство участников дорожного движения любого города составляют легковые автомобили, 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 удобными 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. Более того, разумно предполагать совершенный эгоизм участников этих поездок и стремление нести минимальные потери при проезде. Очевидно также и отсутствие возможностей организовать коалиционное поведение, за исключением разовых акций, которые в общем-то не связаны с ежедневными регулярными поездками. Такое поведение явно соответствует первому принципу Вардропа, поэтому при исследовании загрузки УДС рассмотрим потоки, порождаемые именно легковым частным автотранспортом в утренне-вечерние часы пик.

### 1.1.2. Формализация проблемы

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

Транспортную сеть опишем в виде ориентированного графа $\Gamma(V, E)$, где $V$ — множество вершин, $E$ — множество дуг сети. Каждая дуга соответствует реальному участку автодороги без перекрестков. Каждая вер-

шина представляет узел (перекресток) или место существенного изменения характеристик дороги. Направление дуги определяет ход следования автотранспорта. Магистрали с двусторонним движением соответственно имеют парные противоположно ориентированные дуги.

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

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

Путем (маршрутом) в сети $\Gamma$, соединяющим вершины $i$ и $j$, назовем последовательность дуг

$$e_1 = (i \to k_1), \;\; e_2 = (k_1 \to k_2), \;\; \ldots, \;\; e_m = (k_{m-1} \to k_m), \;\; e_{m+1} = (k_m \to j),$$

где $e_l \in E$ при всех $l = 1, \ldots, m+1$. В маршрутах предполагается отсутствие петель и циклов. Обозначим через $P_w$ множество альтернативных маршрутов, связывающих пару $w \in W$. Совокупность всех путей в сети $\Gamma$ обозначим через $P = \bigcup\limits_{w \in W} P_w$.

Перемещаясь от источников к стокам, пользователи сети выбирают тот или иной маршрут следования. Обозначим через $x_p$ величину потока, идущего по пути $p \in P$; тогда загрузку всей сети задает вектор $x = (x_p \colon p \in P)$.

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

Во введенных обозначениях первый принцип Вардропа можно формализовать следующим образом. Водители выбирают путь с наименьшими транспортными расходами, поэтому для каждой пары $w$ выполнены условия: если по пути $p \in P_w$ идет ненулевой поток $x_p^\dagger$, то затраты по этому пути минимальны, то есть



$$\text{если } x_p^\dagger > 0, \;\; \text{то } G_p(x^\dagger) = \min_{q \in P_w} G_q(x^\dagger) = u_w(x^\dagger), \tag{1}$$

где $u_w(x^\dagger)$ — минимальные транспортные затраты по маршрутам, соединяющим пару $w \in W$, при загрузке сети, определяемой вектором $x^\dagger$.

Соотношения (1), определенные для каждой пары $w \in W$, задают так называемые *условия равновесия транспортных потоков*. Потоки $x^\dagger = (x_p^\dagger \colon p \in P)$, удовлетворяющие условию (1), называются *равновесными*. Для полноты картины необходимо ввести ограничения на допустимость потоков.

Каждой паре источник-сток $w = (i, j) \in W$ соответствует свой спрос на перевозку $\rho_w$ — общий объем пользователей, которые из пункта $i$ должны прибыть в пункт $j$. Набор $(\rho_w \colon w \in W)$ называется *матрицей корреспонденций*. В общем случае предполагается, что корреспонденции являются функциями от минимальных затрат на передвижения в сети, то есть $\rho_w = \rho(u_w^\dagger)$, где $u_w^\dagger = u_w(x^\dagger)$.

Традиционно для транспортных задач потоковые переменные должны быть неотрицательными и удовлетворять балансовым ограничениям. Поэтому допустимой областью для рассматриваемых потоков $x$ является множество

$$X(x^\dagger) = \Big\{ x \geqslant 0 \colon \sum_{p \in P_w} x_p = \rho_w(u_w(x^\dagger)), \; w \in W \Big\} \tag{2}$$

(запись $x \geqslant 0$ означает, что все компоненты $x_p \geqslant 0$).

Как видно из определения (2), допустимая область $X(x^\dagger)$ является «подвижной» и непосредственно зависит от равновесного распределения потоков $x^\dagger$. Однако если объемы корреспонденций известны и имеют стационарные значения $\rho_w(u_w(x^\dagger)) \equiv \rho_w$, что вполне характерно для трудовых миграций в УДС, то допустимое множество имеет фиксированную структуру:

$$X = \Big\{ x \geqslant 0 \colon \sum_{p \in P_w} x_p = \rho_w, \;\; w \in W \Big\}. \tag{3}$$

Проблема поиска равновесных потоков $x^\dagger \in X(x^\dagger)$ называется *задачей транспортного* (*потокового*) *равновесия с эластичным спросом*. При заданных корреспонденциях проблема поиска равновесных потоков $x^\dagger \in X$ называется *задачей транспортного* (*потокового*) *равновесия с фиксированным спросом*.

Допустимое множество $X$, определенное в (3), обладает замечательными свойствами. Во-первых, это полиэдральное [1] и ограниченное мно-

[1] Полиэдральным (многогранным) множеством называется множество решений конечной системы неравенств $Ax \leqslant b$, это замкнутое выпуклое множество.

жество. Во-вторых, $X$ естественным образом можно представить в виде декартова произведения непересекающихся обобщенных симплексов, т. е. $X = \prod\limits_{w \in W} X_w$, где

$$X_w = \Big\{ x_p \geqslant 0 \colon p \in P_w, \ \sum_{p \in P_w} x_p = \rho_w \Big\}. \tag{4}$$

Эти свойства $X$ качественно влияют на построение теоретического и алгоритмического аппаратов задач транспортного равновесия.

Основной подход к решению и исследованию задач транспортного равновесия состоит в сведении исходной проблемы к эквивалентной оптимизационной задаче или вариационному неравенству. Прямое влияние на то, какая из эквивалентных форм будет рассматриваться, оказывают свойства функций транспортных затрат $G_p(x)$. Для компактности последующего изложения объединим компоненты $G_p(x)$ в вектор-функцию $G(x) = (G_p(x) \colon p \in P)$.

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

### 1.1.3. Сведение к вариационному неравенству

Предположим, что имеет место непрерывная монотонная зависимость транспортных издержек от объемов загрузки УДС. Традиционно в этом случае поиск равновесных потоков сводится к решению вариационного неравенства (в частном случае к оптимизационной задаче) [23, 39, 50, 54, 59]. Обоснование для такого сведения дает следующий результат.

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

$$G(x^\dagger)(x - x^\dagger) \geqslant 0 \quad \forall x \in X(x^\dagger). \tag{5}$$

Доказательство. Пусть вектор $x^\dagger = (x_p^\dagger \colon p \in P) \in X(x^\dagger)$ является решением квазивариационного неравенства (5). Покажем, что в $x^\dagger$ выполнено условие (1). Предположим противное, а именно, что для пары $w$ существует такой путь $\bar{p} \in P_w$, что $x_{\bar{p}}^\dagger > 0$ и $G_{\bar{p}}(x^\dagger) > G_{\bar{q}}(x^\dagger)$ для некоторого $\bar{q} \in P_w$, $\bar{q} \neq \bar{p}$. Рассмотрим такой вектор $x^\varepsilon = (x_p^\varepsilon \colon p \in P)$, что

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



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

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

что противоречит тому, что $x^\dagger$ — решение неравенства (5). Следовательно, в точке $x^\dagger$ условие (1) всегда выполнено.

Покажем обратное, а именно, если $x^\dagger \in X(x^\dagger)$ удовлетворяет условию (1), то $x^\dagger$ — решение квазивариационного неравенства (5). Нетрудно видеть, что для всех $p \in P_w$, $w \in W$ и $x \in X(x^\dagger)$ выполнены соотношения

$$G_p(x^\dagger) - u_w(x^\dagger) \geqslant 0, \quad (G_p(x^\dagger) - u_w(x^\dagger))x_p^\dagger = 0, \quad (G_p(x^\dagger) - u_w(x^\dagger))x_p \geqslant 0,$$

следовательно, имеет место оценка

$$\begin{aligned} 0 \leqslant \sum_{w \in W} \sum_{p \in P_w} (G_p(x^\dagger) - u_w(x^\dagger))(x_p - x_p^\dagger) = \\ = \sum_{w \in W} \sum_{p \in P_w} G_p(x^\dagger)(x_p - x_p^\dagger) - \sum_{w \in W} \sum_{p \in P_w} u_w(x^\dagger)(x_p - x_p^\dagger) = \\ = G(x^\dagger)(x - x^\dagger) - \sum_{w \in W} u_w(x^\dagger) \Big( \sum_{p \in P_w} x_p - \sum_{p \in P_w} x_p^\dagger \Big) = \\ = G(x^\dagger)(x - x^\dagger) - \sum_{w \in W} u_w(x^\dagger)(\rho_w(u^\dagger) - \rho_w(u^\dagger)) = G(x^\dagger)(x - x^\dagger), \end{aligned}$$

то есть $x^\dagger$ — решение квазивариационного неравенства (5). □

Для задач транспортного равновесия с фиксированным спросом квазивариационное неравенство заменяется на классическое вариационное неравенство вида:

$$G(x^\dagger)(x - x^\dagger) \geqslant 0 \quad \forall x \in X. \tag{6}$$

Очевидно, что вариационное неравенство (6) представляет частный случай квазивариационного неравенства (5). Теория и методы решения обоих классов неравенств к настоящему времени уже достаточно хорошо разработаны. Познакомиться с достижениями в этой области можно, например, по работам [3, 5, 41, 49–51, 54].

Далее, для того чтобы изучать свойства задач транспортного равновесия в рамках единого математического аппарата, примем дополнительные соглашения и сведем решение задачи транспортного равновесия с эластичным спросом к решению именно вариационного неравенства.

Предположим, что для каждого маршрута $p \in P$ транспортные затраты $G_p(x)$ строго положительны, а для всех пар $w \in W$ функция спроса $\rho_w(u_w)$ принимает только неотрицательные значения.

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

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

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

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

Допустимое множество для вектора $z$ представляет собой неотрицательный ортант $Z = \{z : z \geqslant 0\}$.

**Утверждение 1.** *Вектор $z^\dagger = (x^\dagger, u^\dagger) \geqslant 0$ является решением вариационного неравенства*

$$F(z^\dagger)(z - z^\dagger) \geqslant 0 \quad \forall z \in Z \tag{7}$$

*тогда и только тогда, когда $x^\dagger$ — решение квазивариационного неравенства* (5).

Доказательство. Пусть $x^\dagger \in X(x^\dagger)$ — решение квазивариационного неравенства (5). Тогда для любых $x \geqslant 0$ и $u \geqslant 0$ выполнены условия

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

Откуда следует, что

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

Покажем обратное. Пусть $z^\dagger = (x^\dagger, u^\dagger) \geqslant 0$ — решение вариационного неравенства (7), то есть для любых $z \geqslant 0$ выполнено

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

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

$$F(z^\dagger)z^\dagger \begin{cases} \leqslant 0 & \text{при } \sigma = 0, \\ \geqslant \dfrac{F(z^\dagger)z^\dagger}{\sigma} \to 0 & \text{при } \sigma \to +\infty. \end{cases}$$

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

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

$$F(z^\dagger) \geqslant 0, \quad z^\dagger \geqslant 0, \quad F(z^\dagger)z^\dagger = 0, \tag{8}$$



известной как *нелинейная задача дополнительности* (см., например, [23, 41]).

Перепишем условия (8) в виде

$$G(x^\dagger) - \Xi u^\dagger \geqslant 0, \qquad x^\dagger \geqslant 0, \qquad (G(x^\dagger) - \Xi u^\dagger)x^\dagger = 0, \tag{9}$$
$$\Xi^T x^\dagger - \rho(u^\dagger) \geqslant 0, \qquad u^\dagger \geqslant 0, \qquad (\Xi^T x^\dagger - \rho(u^\dagger))u^\dagger = 0. \tag{10}$$

Система (9) показывает, что вектор $u^\dagger$ соответствует минимальным транспортным затратам в сети при загрузке, определяемой потоками $x^\dagger$. При условии положительности транспортных затрат из (10) следует, что $\Xi^T x^\dagger - \rho(u^\dagger) = 0$, тогда неравенство (7) можно переписать в виде

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

что окончательно доказывает утверждение. □

Очевидно, что допустимая область $Z$ вариационного неравенства (7) является полиэдральным множеством, однако в отличие от допустимой области $X$ вариационного неравенства (6) множество $Z$ неограничено. Это критическим образом влияет на условия разрешимости транспортных задач с эластичным и фиксированным спросом.

### 1.1.4. Разрешимость задач транспортного равновесия

Критерии существования равновесных транспортных потоков формулируются на базе теории разрешимости вариационных неравенств. Рассмотрим вариационное неравенство (6) как общую форму задачи, специально не оговаривая свойства вектор-функции $G$ и допустимого множества $X$. Для изложения основных результатов данного раздела понадобятся следующие определения.

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

**Определение 2.** Точка $x^*$ называется неподвижной точкой отображения $T : X \to X$, если $x^* = T(x^*)$.

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

$$(p - y)(x - p) \geqslant 0 \quad \forall x \in X. \tag{11}$$

Решение вариационного неравенства (6) тесно связано с поиском неподвижных точек проекционного отображения

$$H(x) = \pi_X(x - \lambda G(x)),$$

где $\lambda > 0$ — некоторое фиксированное число.

**Утверждение 2.** *Множество решений $X^\dagger \subseteq X$ вариационного неравенства* (6) *совпадает с множеством неподвижных точек отображения $H(x)$, то есть $X^\dagger = \{x^\dagger \in X\colon x^\dagger = H(x^\dagger)\}$.*

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

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

следовательно, в силу свойства (11) имеем

$$x^\dagger = \pi_X(x^\dagger - \lambda G(x^\dagger)) = H(x^\dagger).$$

Пусть $x^\dagger = H(x^\dagger)$, тогда в силу свойства (11) для любого $x \in X$ выполнено условие

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

то есть $x^\dagger \in X^\dagger$. □

**Теорема 2.** *Пусть вектор-функция $G$ непрерывна, множество $X$ непусто, выпукло и замкнуто. Если $X$ ограничено, то вариационное неравенство* (6) *разрешимо.*

Доказательство. Для выпуклого множества $X$ отображение $H(x)\colon X \to X$ является непрерывным и однозначным. Множество $X$ по условию теоремы компактно, следовательно, по теореме Брауэра (см., например, [2, 14, 24]) у $H(x)$ существует неподвижная точка $x^\dagger = H(x^\dagger)$, которая в силу утверждения 2 одновременно является решением вариационного неравенства (6). □

Из теоремы 2 следует, что если затраты на передвижение $G_p(x)$ являются непрерывными функциями от потоков $x \in X$, то транспортная задача с фиксированным спросом всегда разрешима.

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

$$G(x_R^\dagger)(x - x_R^\dagger) \geqslant 0 \quad \forall x \in X_R. \tag{12}$$

**Теорема 3.** *Пусть вектор-функция $G$ непрерывна, множество $X$ непусто, выпукло и замкнуто. Если существует такой радиус $R > 0$, что $X_R \neq \varnothing$ и решение $x_R^\dagger \in X_R$ вариационного неравенства* (12) *удовлетворяет условию $\|x_R^\dagger\| < R$, то вариационное неравенство* (6) *разрешимо.*



Доказательство. Для произвольного $x \in X$ выберем такое $\lambda \in (0, 1]$, что точка $\bar{x} = x_R^\dagger + \lambda(x - x_R^\dagger) \in X_R$. Имеем

$$0 \leqslant G(x_R^\dagger)(\bar{x} - x_R^\dagger) = G(x_R^\dagger)(x_R^\dagger + \lambda(x - x_R^\dagger) - x_R^\dagger) = \lambda G(x_R^\dagger)(x - x_R^\dagger),$$

то есть $x_R^\dagger$ одновременно является решением вариационного неравенства (6). □

Из теоремы 3 можно получить ряд следствий (см., например, [23]).

**Следствие 1.** *Пусть вектор-функция $G$ непрерывна, множество $X$ непусто, выпукло и замкнуто. Если вектор-функция $G(x)$ коэрцитивна относительно $X$, то есть для некоторого $\bar{x} \in X$ выполнено*

$$\lim_{\|x\| \to \infty,\, x \in X} \frac{G(x)(x - \bar{x})}{\|x\|} \to \infty, \tag{13}$$

*то вариационное неравенство* (6) *разрешимо.*

Доказательство. Условие коэрцитивности (13) позволяет для каждого фиксированного $C > 0$ подобрать достаточно большое $R_C > 0$ такое, что

$$G(x)(x - \bar{x}) \geqslant C\|x\| \quad \forall x \in X, \quad \|x\| = R_C,$$

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

В силу теоремы 2 разрешимо вариационное неравенство

$$G(x_{R_C}^\dagger)(x - x_{R_C}^\dagger) \geqslant 0 \quad \forall x \in X_{R_C}.$$

Если $\|x_C^\dagger\| < R_C$, то по теореме 3 точка $x_{R_C}^\dagger$ является решением исходного вариационного неравенства (6).

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

$$G(x_{R_C}^\dagger)(\bar{x} - x_{R_C}^\dagger) \leqslant -C\|x_{R_C}^\dagger\| = -CR < 0,$$

что противоречит определению $x_{R_C}^\dagger$. □

Таким образом, для разрешимости задачи транспортного равновесия с эластичным спросом нужны более сильные, чем непрерывность, предположения о свойствах вектор-функции $F(z)$.

Вопрос единственности равновесного распределения транспортных потоков решается за счет свойств строгой монотонности функции транспортных издержек.

**Определение 3.** Вектор-функция $G\colon X \to \mathbb{R}^n$ называется строго монотонной на $X$, если для любых таких $x, y \in X$, что $x \neq y$, выполнено неравенство $(G(x) - G(y))(x - y) > 0$.

**Теорема 4.** *Если вектор-функция $G(x)$ строго монотонна, то вариационное неравенство* (6) *может иметь не более одного решения.*

Доказательство. Предположим противное, а именно, что существуют два различных решения $x^1$, $x^2 \in X$, $x^1 \neq x^2$, задачи (6). Очевидно, при этом выполнены неравенства

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

складывая которые, получаем

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

что противоречит свойству строгой монотонности $G(x)$. □

С точки зрения задачи транспортного равновесия утверждение теоремы 4 означает, что если транспортные затраты возрастают с увеличением загрузки сети — а это весьма естественное предположение, — то существует единственное равновесное распределение транспортных потоков.

### 1.1.5. Симметричные задачи транспортного равновесия

Задачу транспортного равновесия будем называть *симметричной*, если для вектор-функции транспортных издержек $G(x)$ матрица Якоби $\nabla G(x) = \left(\dfrac{\partial G_p(x)}{\partial x_q} : p, q \in P\right)$ симметрична для всех $x \in X$.

Свойство симметричности матрицы $\nabla G(x)$ является достаточным условием для того, чтобы гарантировать существование дифференцируемой функции $f : X \to \mathbb{R}$ такой, что $\nabla f(x) = G(x)$ для всех $x \in X$. Вектор-функция $G$ в таком случае называется *потенциальной*, а вариационное неравенство (6) можно переписать в виде

$$\nabla f(x^\dagger)(x - x^\dagger) \geqslant 0 \quad \forall x \in X. \tag{14}$$

Из теории оптимизации известно, что условие (14) представляет необходимый критерий оптимальности в задаче

$$f(x) \to \min, \quad x \in X, \tag{15}$$

где $X$ — выпуклое замкнутое множество.

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

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

Переходя к пределу при $\lambda \to 0$, получаем (14).

Считается, что решить оптимизационную задачу намного проще, чем вариационное неравенство [56]. Теория методов оптимизации богата разнообразными алгоритмами. Кроме того, существует множество программных пакетов для решения этого класса задач, чего нельзя сказать о вариационных неравенствах. Однако основная трудность состоит в построении



функции $f(x)$. Для потенциальных отображений такую функцию можно построить, проведя следующие рассуждения.

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

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

$$\mathcal{I} = \oint_{x^0}^{x} G(x(t))\, d(x(t)) = \int_0^1 G(x(t)) x'_t(t)\, dt = \int_0^1 \nabla f(x(t)) x'_t(t)\, dt =$$

$$= \int_0^1 df(x(t)) = f(x(t))\Big|_0^1 = f(x(1)) - f(x(0)) = f(x) - f(x^0).$$

Видим, что значение интеграла $\mathcal{I}$ не зависит от параметрического задания кривой $L$. Рассмотрим простейший пример такого задания: $x(t) = x^0 + t(x - x^0)$, тогда при $G(x) = \nabla f(x)$ вариационное неравенство (6) эквивалентно следующей оптимизационной задаче:

$$f(x) = f(x^0) + \int_0^1 G(x^0 + t(x - x^0))(x - x^0)\, dt \to \min, \quad x \in X. \tag{16}$$

Таким образом, решение симметричной задачи транспортного равновесия эквивалентно решению некоторой оптимизационной задачи. Именно этому классу задач посвящена большая часть работ по исследованию транспортного равновесия [22, 28, 29, 35, 37, 45, 46].

## 1.2. Построение функций транспортных затрат

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

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

$$y_e = \sum_{p \in P} \theta_{ep} x_p, \tag{17}$$

где

$$\theta_{ep} = \begin{cases} 1, & \text{если путь } p \text{ проходит через дугу } e; \\ 0 & \text{в противном случае.} \end{cases}$$

Определим $\Theta = (\theta_{ep}\colon e \in E,\ \ p \in P)$ — матрицу инцидентности дуг и путей, $y = (y_e\colon e \in E)$ — вектор, описывающий загрузку дуг сети $\Gamma$. В матричной форме взаимосвязь потоков по путям и дугам описывается уравнением $y = \Theta x$.

В ряде случаев рассматриваются транспортные задачи в терминах только потоковых переменных по дугам. Отметим, что в множестве $X$, определенном в (3), от потоковых переменных по путям $x$ можно легко перейти к вектору $y$; обратный переход неоднозначен.

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

### 1.2.1. Аддитивные функции затрат

Самым распространенным и простым предположением о свойствах функций транспортных затрат является аддитивная зависимость $G(x)$ от $\tau(y)$, означающая, что транспортные затраты на прохождение каждого пути $p \in P$ складываются только из затрат на проезд по дугам, составляющим этот путь [38, 39, 54]:

$$G_p(x) = \sum_{e\in E} \theta_{ep}\tau_e(y). \tag{18}$$

В результате получаем, что вектор-функция $G(x)$ вариационного неравенства (6) имеет вид

$$G(x) = \Theta^T\tau(y), \quad y = \Theta x. \tag{19}$$

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

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

Следовательно, матрица Якоби $\nabla G(x)$ симметрична для любых $x \in X$, то есть вектор-функция $G(x)$ потенциальна и равновесные транспортные потоки можно определить как решение оптимизационной задачи (16). Учитывая соотношения (19), вид целевой функции $f(x)$ определяется как:

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

Таким образом, при $\tau_e(y) \equiv \tau_e(y_e)$ задача (16) перепишется в виде:

$$\sum_{e\in E}\int_0^{y_e}\tau_e(z)\,dz \to \min, \quad y = \Theta x,\ \ x \in X. \tag{20}$$

От потоковых переменных по путям $x$ в модели (20) можно избавиться, если ввести следующие обозначения. В общем потоке $y_e$ по каждой дуге $e \in E$ отдельно выделим поток $y_e^s$, порождаемый источником $s \in S$ и идущий по $e$. Вектор $y^s = (y_e^s\colon e \in E)$ определяет загрузку дуг транспортной сети, порожденную источником $s$. Очевидно, должны быть выполнены условия:

$$y = \sum_{s\in S} y^s, \quad y^s \geqslant 0, \quad s \in S.$$

Через $\rho_{sv}$ обозначим объем потока, который из источника $s \in S$ должен достичь вершины $v \in V$. При заданной матрице корреспонденций $(\rho_w\colon w \in W = S \times D)$ величины $\rho_{sv}$ определяются по правилу:

$$\rho_{sv} = \begin{cases} \rho_w, & \text{если } v \neq s,\ w = (s, v) \in W, \\ 0, & \text{если } v \neq s,\ w = (s, v) \notin W, \\ -\rho_w, & \text{если } v = s. \end{cases}$$

Балансовые ограничения при переходе к новым переменным $y_e^s$ запишутся в виде

$$\sum_{e\in E_v^+} y_e^s - \sum_{e\in E_v^-} y_e^s = \rho_{sv}, \quad (s, v) \in S \times V, \tag{21}$$

где

$$E_v^+ = \{e \in E\colon \text{дуга } e \text{ входит в вершину } v\},$$
$$E_v^- = \{e \in E\colon \text{дуга } e \text{ выходит из вершины } v\}.$$

Определим множество $Y_s = \{y^s \geqslant 0\colon \text{выполнены условия (21)}\}$. В результате симметричная задача транспортного равновесия (20) перепишется в виде:

$$\sum_{e\in E} \int_0^{y_e} \tau_e(z)\,dz \to \min, \quad y = \sum_{s\in S} y^s, \;\; y^s \in Y_s. \tag{22}$$

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

$$\tau_e(y) = \tau_e^0\Big(1 + \mu\Big(\frac{y_e}{c_e}\Big)^n\Big),$$

где $\tau_e^0$ — время проезда по свободной дуге $e$, $c_e$ — пропускная способность дуги $e$, $\mu$ и $n$ — некоторые положительные константы. При использовании BPR-функции задача транспортного равновесия сводится к оптимизационной задаче (20).

### 1.2.2. Неаддитивные функции затрат

В общем случае, построение функции затрат $\tau_e(y)$ является задачей, требующей отдельных исследований. Здесь окажутся полезными как натурные замеры потоков и соответствующих им задержек в реальных УДС (улично-дорожных сетях), так и результаты компьютерного моделирования, например, при помощи специальных программ для агентного моделирования, так активно развивающиеся в последние годы.

Существуют ситуации, когда предположение об аддитивности функций $G_p(x)$ не подходит для описания транспортных затрат. Стремление к более адекватному моделированию автомобильных потоков привело к новым формам аналитического описания затрат [32, 43, 52]. Неаддитивные транспортные затраты возникают, например, в случаях, когда при моделировании одновременно учитываются и временные, и финансовые расходы. Так, в работе [43] предложена функция, характеризующая финансовые затраты, на которые, в свою очередь, влияют временные затраты:

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

где $\tau_e(\cdot)$ — время, потраченное на прохождение дуги $e$, $\Phi_p(\cdot)$ — функция, преобразующая временные затраты для пути $p$ в финансовые затраты, $\Psi_p(\cdot)$ — финансовые затраты, характеризующие маршрут $p$, которые могут меняться в зависимости от загрузки сети, $\eta > 0$ — эксплуатационные



расходы в единицу времени. В работе [32] предложен более общий вид неаддитивной функции затрат:

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

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

С одной стороны, неаддитивные затраты более реалистично могут описать функционирование транспортной системы, с другой — вариационное неравенство (6) (а тем более (7)) при сложных функциях $G_p(x)$ весьма трудоемко для анализа и решения.

### 1.2.3. Модель стационарной динамики

С формальной точки зрения, если объемы потоков не ограничены сверху пропускной способностью транспортной сети, то монотонные функции затрат допускают сколь угодно большие значения потоков, что едва ли согласуется с реальностью и справедливо критикуется. Однако полученный опыт моделирования транспортных потоков в реальной УДС [18, 19, 25, 26] показывает, что именно монотонное возрастание затрат является сдерживающим фактором для получения нереально больших потоков на дугах.

Одна из попыток избавить представление $\tau_e$ от упомянутого недостатка описана в работе [57]. Здесь для каждой дуги $e \in E$ транспортной сети предлагается ввести два вполне измеримых показателя: пропускную способность $c_e$ и время проезда по свободной дуге (минимальное время проезда) $\tau_e^0$. 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. Согласно [57], ситуация потокового равновесия в транспортной сети теперь определяется как загрузка ее дуг $y^\dagger = (y_e^\dagger\colon e \in E)$ и временные затраты на дугах $\tau^\dagger = (\tau_e^\dagger\colon e \in E)$, которые удовлетворяют ограничениям:

$$0 \leqslant y_e^\dagger \leqslant c_e, \quad \tau_e^\dagger \geqslant \tau_e^0, \quad e \in E, \tag{23}$$

при этом выполнены условия пользовательской оптимальности (1):

$$\tau_e^\dagger \begin{cases} = \tau_e^0, & \text{если } \; y_e^\dagger < c_e, \\ \geqslant \tau_e^0, & \text{если } \; y_e^\dagger = c_e. \end{cases} \tag{24}$$

Условие (24) показывает, что временные затраты зависят от потока, но эта зависимость не является монотонной. В случае полной загрузки сети можно гарантировать лишь только то, что временные затраты на дугах

будут не меньше минимального времени проезда. Пара $(y^\dagger, \tau^\dagger)$, удовлетворяющая условиям (23), (24), называется стационарным динамическим решением задачи транспортного равновесия.

На взгляд авторов, независимость 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. Однако этот подход сопровождается весьма интересными теоретическими результатами, о которых вкратце стоит упомянуть.

Для каждой потокообразующей пары $w \in W$ длину кратчайшего (по времени) пути при временных затратах на дугах, определяемых вектором $\tau$, задает вогнутая кусочно-линейная функция

$$u_w(\tau) = \min_{q \in P_w} \Big\{ G_q = \sum_{e \in E} \theta_{eq} \tau_e \Big\}.$$

В качестве стационарного динамического решения задачи транспортного равновесия авторы [57] предлагают брать такое решение $(y^\dagger, \tau^\dagger)$ негладкой оптимизационной задачи

$$\sum_{w \in W} d_w u_w(\tau) - \sum_{e \in E} y_e \tau_e \to \max, \quad \tau_e \geqslant \tau_e^0, \quad e \in E, \tag{25}$$

на котором бы выполнялись равенства $\eta_e^\dagger = c_e - y_e^\dagger$, где $\eta_e^\dagger$ — оптимальные значения двойственных переменных задачи (25).

Интересным результатом является тот факт, что сложную во всех отношениях задачу (25) можно заменить на двойственную к ней, которая в свою очередь является задачей линейного программирования, интерпретируемой как задача минимизации издержек в многопродуктовой транспортной задаче:

$$\tau^0 y \to \min, \quad y = \sum_{s \in S} y^s \leqslant c, \quad y^s \in Y_s, \tag{26}$$

где $\tau^0 = (\tau_e^0 \colon e \in E)$ — вектор минимальных временных затрат в сети, $c = (c_e \colon e \in E)$ — вектор пропускной способности сети, переменные $y^s$ и множество $Y_s$ определены в разделе 1.2.1.

Практика применимости задачи (26) к УДС Владивостока описана в работе [19].

## 1.3. Соотношение между системным оптимумом и конкурентным равновесием

Очевидно, что общие затраты при системной оптимизации не могут превышать общих затрат при пользовательской оптимизации. Поэтому разность между совокупными транспортными затратами, которые несут



пользователи сети, перемещаясь согласно либо только первому, либо только второму поведенческим принципам Вардропа, можно рассматривать как цену анархии, и существуют примеры, когда эта цена составляет существенную долю от общих расходов.

На принципиальную разницу между конкурентным транспортным равновесием и системным оптимумом одним из первых обратил внимание А. Пигу [60]. Он рассмотрел простейшую транспортную сеть, состоящую из двух дуг, соединяющих два пункта, скажем, спальный район $A$ и бизнес-зону $B$ (см. рис. 2).

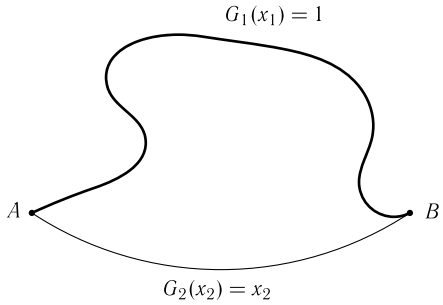



Рис. 2. Пример транспортной сети Пигу

Жители пункта $A$ вольны выбирать, по какой из двух дорог им лучше добираться до работы. Обозначим через $x_1$ и $x_2$ доли общего объема трудового потока, едущего по первой и второй дорогам соответственно. Дороги в рассматриваемой сети неравноценны. Первая представляет магистральное шоссе, которое способно принять весь поток автомобилей из пункта $A$ в пункт $B$ без всякого замедления движения. Однако эта дорога достаточно длинная и проезд по ней требует определенного времени $G_1$, которое будем считать равным, например, одному часу, то есть $G_1(x_1) = 1$. По второй дороге путь существенно короче, но это дорога узкая и движение сильно замедляется при наличии на ней потока автомобилей. Чтобы подчеркнуть суть примера, будем считать, что время проезда по второй дороге $G_2$ линейно зависит от потока $x_2$ по этой дороге и задается соотношением $G_2(x_2) = x_2$. Тогда в соответствии с первым принципом Вардропа (Пигу—Найта—Вардропа) равновесному состоянию будет соответствовать такое распределение потоков $(x_1^\dagger, x_2^\dagger)$, что

$$G_1(x_1^\dagger) = G_2(x_2^\dagger), \quad x_1^\dagger + x_2^\dagger = 1, \quad x_1^\dagger,\ x_2^\dagger \geqslant 0,$$

откуда немедленно следует, что $x_1^\dagger = 0,\ x_2^\dagger = 1$, при этом системные затраты $c(x_1^\dagger, x_2^\dagger) = 1 \cdot x_1^\dagger + x_2^\dagger \cdot x_2^\dagger = 1$.

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

$$x_1 + x_2^2 \to \min\colon\ x_1 + x_2 = 1,\ \ x_1,\, x_2 \geqslant 0, \tag{27}$$

минимум которой достигается в точке $x_1^\star = x_2^\star = 0{,}5$, минимальные затраты $c(x_1^\star,\, x_2^\star) = 0{,}75$, что на 25% уменьшает системные издержки в сети.

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

Из условия равновесия (1), очевидно, следует, что если $x_p^\dagger > 0$ и $x_q^\dagger > 0$ для путей $p,\ q \in P_w$, то $G_p(x^\dagger) = G_q(x^\dagger) = u_w^\dagger$. Поскольку вклад в суммарные системные затраты (обозначим их $c(x^\dagger)$) при равновесном распределении $x^\dagger$ вносят только ненулевые потоки $x_p^\dagger > 0$, а для них все удельные затраты в пределах одной пары $w$ одинаковы и равны $u_w^\dagger$, то значение $c(x^\dagger)$ можно рассчитать следующим образом:

$$c(x^\dagger) = \sum_{p\in P} G_p(x^\dagger)x_p^\dagger = \sum_{w\in W}\sum_{p\in P_w} G_p(x^\dagger)x_p^\dagger = \sum_{w\in W} u_w^\dagger \sum_{p\in P_w} x_p^\dagger = \sum_{w\in W} u_w^\dagger \rho_w. \tag{28}$$

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

$$c(x) = \sum_{p\in P} G_p(x)x_p \to \min, \quad x \in X. \tag{29}$$

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

**Теорема 5.** *Пусть $x^\star$ — решения задачи* (29), *то есть оптимальное распределение потоков в сети. Тогда для всякой пары $w \in W$ верно следующее*: *если $x_p^\star > 0,\ \ p \in P_w$, то $\dfrac{\partial c(x^\star)}{\partial x_p} \leqslant \dfrac{\partial c(x^\star)}{\partial x_q}$ для всех $q \in P_w$.*

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

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



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

$$c(x^\varepsilon) - c(x^\star) \leqslant \nabla c(x^\varepsilon)(x^\varepsilon - x^\star) = \varepsilon\Big(\frac{\partial c(x^\varepsilon)}{\partial x_q} - \frac{\partial c(x^\varepsilon)}{\partial x_p}\Big) < 0,$$

что противоречит оптимальности $x^\star$.

□

В частном случае, когда $G_p(x) \equiv G_p(x_p)$, из теоремы 5 непосредственно следует, что в оптимальном распределении потоков $x^\star$ для всякой пары $w \in W$ верно следующее: если $x_p^\star > 0,\ \ p \in P_w$, то $\dfrac{\partial c_p(x^\star)}{\partial x_p} \leqslant \dfrac{\partial c_q(x^\star)}{\partial x_q}$ для всех $q \in P_w$. При этом, как и в случае равновесных потоков, для оптимальных потоков справедливы равенства $\dfrac{\partial c_p(x^\star)}{\partial x_p} = \dfrac{\partial c_q(x^\star)}{\partial x_q}$ для тех $p,\ q \in P_w$, для которых $x_p^\star > 0,\ \ x_q^\star > 0$. Эти условия известны также как условия Гиббса (см., например, [11]).

Рассмотрим случай, когда транспортные затраты на прохождение каждого пути $p \in P$ складываются только из затрат на проезд по дугам, составляющим этот путь, то есть $G_p(x)$ определяются по формуле (18), при этом затраты по дугам $\tau_e(y)$ описываются аффинными функциями $\tau_e(y) = a_e y_e + b_e$, где $a_e$ и $b_e$ — неотрицательные коэффициенты для всех $e \in E$. При этом функция системных затрат и ее частные производные определяются как

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

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

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

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

$$\sum_{e\in E} \theta_{ep}(a_e y_e^\dagger + b_e) \leqslant \sum_{e\in E} \theta_{eq}(a_e y_e^\dagger + b_e); \tag{30}$$

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

$$\sum_{e\in E} \theta_{ep}(2a_e y_e^\star + b_e) \leqslant \sum_{e\in E} \theta_{eq}(2a_e y_e^\star + b_e). \tag{31}$$

Любопытно, что при линейных функциях задержек $\tau_e(y) = a_e y_e$ из неравенств (30) и (31) следует совпадение равновесных и оптимальных потоков.

Следуя работе [61], через тройку $[\Gamma, \rho, G(x)]$ обозначим транспортную модель, определенную на сети $\Gamma$, с матрицей корреспонденций $\rho = (\rho_w : w \in W)$ и затратами $G(x) = (G_p(x) : p \in P)$. Везде далее будем полагать

$$G_p(x) = \sum_{e \in E} \theta_{ep} \tau_e(y) = \sum_{e \in E} \theta_{ep}(a_e y_e + b_e). \tag{32}$$

Имеет место следующий результат.

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

Доказательство. Если $x^\dagger$ является допустимым решением равновесной модели $[\Gamma, \rho, G(x)]$, то, очевидно, $\frac{1}{2}x^\dagger$ — допустимое решение оптимизационной модели $\left[\Gamma, \frac{1}{2}\rho, G(x)\right]$, при этом неравенства (31) для $\frac{1}{2}x^\dagger$ переходят в (30). □

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

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

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

$$c(x^\delta) \geqslant c(x^\star) + \delta \sum_{w \in W} v_w(x^\star)\rho_w, \tag{33}$$

*где $\delta \geqslant 0$, $v_w(x^\star) = \min\limits_{p \in P_w} \dfrac{\partial c(x^\star)}{\partial x_p}$.*

Доказательство. Рассмотрим допустимые относительно модели $[\Gamma, (1+\delta)\rho, G(x)]$ потоки $x^\delta$. При затратах $G_p(x)$, определенных в (32), где все коэффициенты $a_e \geqslant 0$, функция $c(x)$ выпукла, отсюда

$$c(x^\delta) \geqslant c(x^\star) + \frac{\partial c(x^\star)}{\partial x_p}(x^\delta - x^\star) = c(x^\star) + \sum_{p \in P} \frac{\partial c(x^\star)}{\partial x_p}(x_p^\delta - x_p^\star) = c(x^\star) +$$

$$+ \sum_{w \in W} \sum_{p \in P_w} \frac{\partial c(x^\star)}{\partial x_p}(x_p^\delta - x_p^\star) = c(x^\star) + \sum_{w \in W} \left( \sum_{p \in P_w} \frac{\partial c(x^\star)}{\partial x_p} x_p^\delta - \sum_{p \in P_w} \frac{\partial c(x^\star)}{\partial x_p} x_p^\star \right).$$



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

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

то

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

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

$$c(x^\delta) \geqslant c(x^\star) + \sum_{w \in W} \left( \sum_{p \in P_w} v_w(x^\star) x_p^\delta - \sum_{p \in P_w} v_w(x^\star) x_p^\star \right) =$$

$$= c(x^\star) + \sum_{w \in W} v_w(x^\star) \left( \sum_{p \in P_w} x_p^\delta - \sum_{p \in P_w} x_p^\star \right) =$$

$$= c(x^\star) + \sum_{w \in W} v_w(x^\star)((1+\delta)\rho_w - \rho_w) = c(x^\star) + \delta \sum_{w \in W} v_w(x^\star)\rho_w. \quad \square$$

Итоговый результат текущего раздела устанавливает следующая теорема.

**Теорема 6.** *Для транспортной модели $[\Gamma, \rho, G(x)]$ с аффинными функциями задержек (32) для оптимального $x^\star$ и равновесного $x^\dagger$ распределений потоков выполняется соотношение*

$$\frac{c(x^\dagger)}{c(x^\star)} \leqslant \frac{4}{3}.$$

Доказательство. Согласно (28) системные затраты для равновесного распределения $x^\dagger$ рассчитываются как

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

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

Положим $\delta = 1$ в оценке (33). Тогда для произвольного потока $x$, допустимого в модели $\left[\Gamma, 2 \cdot \frac{1}{2}\rho, G(x)\right] = [\Gamma, \rho, G(x)]$, имеем

$$c(x) \geqslant c\left(\frac{1}{2}x^\dagger\right) + \sum_{w \in W} \frac{1}{2} v_w\left(\frac{1}{2}x^\dagger\right)\rho_w =$$

$$= c\left(\frac{1}{2}x^\dagger\right) + \frac{1}{2}\sum_{w \in W} u_w(x^\dagger)\rho_w = c\left(\frac{1}{2}x^\dagger\right) + \frac{1}{2}c(x^\dagger). \tag{34}$$

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

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

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

В результате, продолжая оценку (34), получим

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

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

$$\frac{c(x^\dagger)}{c(x^\star)} \leqslant \frac{4}{3}. \qquad \square$$

Относительно общего случая нелинейных функций транспортных затрат нетрудно убедиться в том, что цена анархии может быть сколь угодно большой.

## 1.4. Численные методы решения задач транспортного равновесия

Эквивалентность задачи транспортного равновесия вариационному неравенству, а в частном случае оптимизационной задаче, позволяет адаптировать численные методы решения последних для поиска равновесных потоков. В данном разделе рассматриваются подходы к решению задачи транспортного равновесия с фиксированным спросом.

В зависимости от того, в каком пространстве переменных рассматривается исходная задача, выделяют два основных подхода к построению алгоритмических схем. Если равновесие моделируется только через потоковые переменные по дугам $y_e$, то применяют так называемые *дуговые алгоритмы* (arc-based algorithms). Если основными переменными задачи являются потоки по путям $x_p$ и, соответственно, итерирование ведется по допустимым маршрутам, то такие алгоритмы называются *маршрутными* (path-based algorithms) [34, 59].

Основным преимуществом поиска равновесия через переменные $x_p$ является возможность «убить двух зайцев одним выстрелом»: зная распределение потоков по маршрутам и используя соотношение (17), всегда можно определить загрузку дуг транспортной сети. Обратное преобразование, очевидно, неоднозначно. Информация о распределении потоков по путям сама по себе является важной при моделировании других задач, например, проблем загрязнения окружающей среды, оценки матрицы корреспонденций, планирования транспортных развязок и модернизации улично-дорожной сети, эффективного регулирования движения и т. п.



Несомненным плюсом в пользу работы в пространстве потоковых переменных по путям с алгоритмической точки зрения является возможность естественной проверки выполнения условия равновесия (1) и поиска распределения потоков, удовлетворяющих заданной точности. Информация только о потоках по дугам такой возможности не дает. Кроме того, структура допустимого множества $X$, определенного в (3), представляет собой декартово произведение непересекающихся симплексов $X_w$, и такое свойство может породить целый класс методов, использующих принципы декомпозиции и идеи параллелизации итерационных схем.

Последним аргументом в пользу исследования задачи транспортного равновесия именно в терминах потоковых переменных по путям является тот простой факт, что при общем задании функции издержек $G_p(x)$, не обязательно складывающихся из затрат на передвижение по дугам, переформулировка условия равновесия (1) в терминах переменных $y_e$ невозможна.

Основной недостаток работы с потоковыми переменными по путям — это необходимость априорного задания множества всех допустимых маршрутов $P$. Такая задача является очень трудоемкой, особенно для реальных транспортных сетей. Как вариант, для каждой потокообразующей пары можно ограничиться рассмотрением $k$ кратчайших маршрутов, заведомо исключить неперспективные пути, но от этого проблема проще не становится. На практике используется не так много вариантов движения, поэтому нет необходимости знать все элементы множества $P$. Более того, существует стандартная техника, часто называемая методом генерации столбцов, когда входные данные непосредственно строятся в процессе решения задачи. Применение такой техники к проблеме поиска транспортного равновесия позволит строить множество допустимых и перспективных для использования маршрутов непосредственно в процессе решения задачи.

Таким образом, из приведенных аргументов видно, что потоковое равновесие предпочтительней искать как решение вариационного неравенства (6) в пространстве потоковых переменных по путям.

### 1.4.1. Проекционные методы решения задачи транспортного равновесия

Среди существующих методов решения вариационных неравенств отдельно можно выделить проекционные алгоритмы, отличающиеся просто-

той своих итерационных схем и гибкостью к различного рода модификациям.

В основу проекционных методов положена связь между множеством решений вариационного неравенства и неподвижными точками проекционного отображения, установленная в утверждении 2.

Далее понадобятся следующие определения.

**Определение 4.** Отображение $G\colon X \to \mathbb{R}^n$ на множестве $X$ называется:

- липшицевым, если существует константа $L > 0$ такая, что $\|G(x) - G(y)\| \leqslant L\|x - y\|$ для всех $x, y \in X$;
- сильно монотонным с константой $\tau > 0$, если $(G(x) - G(y))(x - y) \geqslant \geqslant \tau\|x - y\|^2$ для всех $x, y \in X$;
- обратно сильно монотонным (ко-коэрцитивным) с константой $\tau > > 0$, если $(G(x) - G(y))(x - y) \geqslant \tau\|G(x) - G(y)\|^2$ для всех $x, y \in X$;
- монотонным, если $(G(x) - G(y))(x - y) \geqslant 0$ для всех $x, y \in X$;
- псевдомонотонным, если из неравенства $G(x)(y - x) \geqslant 0$ следует $G(y)(y - x) \geqslant 0$ для всех $x, y \in X$;
- строго псевдомонотонным, если из неравенства $G(x)(y - x) \geqslant 0$ следует $G(y)(y - x) > 0$ для всех $x, y \in X$, $x \neq y$.

В самой простой форме проекционный метод строит последовательность $\{x^k\} \in X$, генерируемую рекуррентным соотношением

$$x^{k+1} = \pi_X(x^k - \lambda_k G(x^k)), \quad \lambda_k > 0, \quad k = 0, 1, 2, \ldots \tag{35}$$

На текущий момент известно, что сходимость процесса (35) гарантируется при выполнении одного из следующих условий (см. [50] и библиографию в ней):

1) отображение $G$ сильно монотонно с константой $\tau$ и липшицево с константой $L$, $\lambda_k \in (0, 2\tau/L^2)$;
2) отображение $G$ ко-коэрцитивно с константой $\mu$, $\lambda_k \in (0, 2\mu)$;
3) для любых $x \in X \backslash X^*$ и $x^* \in X^*$ выполнено

$$G(x)(x - x^*) > 0, \tag{36}$$

где $X^*$ — множество решений вариационного неравенства, $\lambda_k = \dfrac{\alpha_k}{\|G(x^k)\|}$, $\sum\limits_{k=0}^{\infty} \alpha_k = \infty$, $\sum\limits_{k=0}^{\infty} \alpha_k^2 < \infty$.

Нетрудно показать, что любое сильно монотонное липшицево отображение является ко-коэрцитивным и влечет выполнение неравенства (36). Обратное, очевидно, неверно, поэтому второе и третье условия являются менее ограничительными, чем первое, однако остаются достаточно сильными предположениями, что существенно сужает круг задач, для которых



метод (35) гарантированно сходится. Кроме того, правила выбора шагового множителя $\lambda_k$ являются весьма общими для всех трех случаев, что на практике приводит к медленной скорости сходимости всего процесса.

Одной из попыток улучшить ситуацию было предложение о замене процесса (35) на так называемый экстраградиентный метод [1, 13]:

$$u^k = \pi_X(x^k - \lambda_k G(x^k)), \quad x^{k+1} = \pi_X(x^k - \lambda_k G(u^k)), \quad \lambda_k > 0, \quad k = 0, 1, 2, \ldots \tag{37}$$

Первую проекцию $u^k$ можно трактовать как предиктор, вторую $x^{k+1}$ — как корректор.

Сходимость экстраградиентного метода гарантируется при выполнении одного из следующих условий [1, 13, 31, 53]:

1) $G$ монотонно и липшицево с константой $L$, $\lambda_k \in (0, 1/L)$;
2) $G$ псевдомонотонно, локально липшицево, динамическая регулировка шага $\lambda_k \in (0, \min\{\bar{\lambda}, \gamma\|x^k - u^k\|/\|G(x^k) - G(u^k)\|\})$, $\gamma \in (0, 1)$.

Для ускорения сходимости экстраградиентного метода величину шага $\lambda_k$ у предиктора и корректора можно сделать разной, при этом отображение $G$ может удовлетворять всего лишь свойству псевдомонотонности [44, 62, 65]. Хороший обзор, посвященный исследованиям методов проекционного типа, приведен в [66].

Нетрудно видеть, что условия сходимости экстраградиентного метода менее ограничительные по сравнению с проекционным. Однако плата за такое послабление — это многократное решение задачи проекции на этапе определения шагового множителя.

Основной трудностью при реализации процессов (35) и (37) является вычисление проекции на множество $X$. В общем случае требуется решение вспомогательной оптимизационной задачи, однако при поиске транспортного равновесия как решения вариационного неравенства (6) операцию проектирования $\pi_X(\cdot)$ можно свести к более простым вычислениям $\pi_{X_w}(\cdot)$.

### 1.4.2. Декомпозиция проекционных методов для поиска равновесных потоков

Как уже было отмечено, множества $X_w$, определенные в (4), не пересекаются по переменным, поэтому вектор $x \in X$ можно разделить на подвектора $x^w = (x_p\colon p \in P_w) \in X_w$ — потоки по путям, соединяющим пару $w \in W$. Аналогично выделим вектор-функции $G^w(x) = (G_p(x)\colon p \in P_w)$ — издержки по путям, соединяющим пару $w \in W$, при загрузке сети потоками $x$. Таким образом $x = (x^w\colon w \in W)$ и $G(x) = (G^w(x)\colon w \in W)$.

Отмеченная специфика множества $X$, определенного в (3), естественным образом позволяет свести операцию $\pi_X(\cdot)$ к вычислениям $\pi_{X_w}(\cdot)$ для каждого $w \in W$. Поэтому для поиска транспортного равновесия общая

схема проекционного метода (35) трансформируется в следующий процесс:

$$x^k = ((x^w)^k : w \in W), \quad (x^w)^{k+1} = \pi_{X_w}((x^w)^k - \lambda_k G^w(x^k)),$$
$$x^{k+1} = ((x^w)^{k+1} : w \in W), \quad \lambda_k > 0, \quad k = 0, 1, 2, \ldots \tag{38}$$

Аналогично можно преобразовать и процесс (37).

Для вычисления $\pi_{X_w}(\cdot)$ существует замечательный алгоритм поиска проекции точки $z \in \mathbb{R}^n$ на симплекс $\Delta = \{x \in \mathbb{R}^n_+ : ex = \rho\}$, где $e$ — единичный вектор размерности $n$. Общая схема алгоритма следующая [47]:

Шаг 0. Вычислить $x^0 = z + ((\rho - ez)/n)e$, положить $k = 0$.

Шаг 1. Если $x^k \geqslant 0$, то проекция $z$ на множество $\Delta$ найдена, алгоритм заканчивает работу. В противном случае выполняется шаг 2.

Шаг 2. Вычислить $x^{k+1}$ по правилу:

$$x_i^{k+1} = \begin{cases} 0, & \text{если } i \notin I, \\ x_i^k + \dfrac{1}{|I|}\left(\rho - \sum\limits_{j \in I} x_j^k\right), & \text{если } i \in I. \end{cases}$$

где $I = \{i : x_i^k > 0\}$, положить $k = k + 1$ и перейти на шаг 1.

Приведенный алгоритм за не более чем $n$ шагов приведет к искомой точке $\pi_\Delta(z)$.

Отметим, что подвекторы $(x^w)^k$ определяются независимо друг от друга, поэтому при программной реализации процесс (38) легко поддается параллелизации, что несомненно улучшает его вычислительные свойства и делает проекционные методы привлекательными для использования.

### 1.4.3. Проекционный метод с генерацией маршрутов

Специфика допустимого множества $X$ задачи транспортного равновесия позволяет без труда вычислять операцию проектирования. Однако проекционные алгоритмы предполагают априорное задание полного множества допустимых маршрутов $P$, чего хотелось бы избежать по следующим трем причинам. Во-первых, трудоемкость алгоритмов построения всех путей, соединяющих пару вершин, растет экспоненциально с увеличением размерности графа. Поэтому для реальных сетей большой размерности, в которых рассматриваются сотни или тысячи потокообразующих пар, задача построения полного множества маршрутов может потребовать больших вычислительных ресурсов, превосходящих даже возможности суперкомпьютеров. Во-вторых, мощность множества $P$ определяет размерность решаемой задачи, поэтому чем больше элементов в $P$, тем труднее с вычислительной точки зрения будет проходить поиск равновесного распределения потоков в сети. И наконец, в-третьих, скорее всего, бóльшая часть путей из множества $P$ не будет использоваться при переносе заданного трафика, поэтому их включение в $P$ бессмысленно.



В [46] для решения симметричной задачи транспортного равновесия предложен метод, который наряду с решением оптимизационной задачи (20) последовательно строит множество допустимых маршрутов в сети. Такой подход можно обобщить и для несимметричного случая.

Обозначим через $x^\dagger(\mathcal{P})$ равновесное распределение потоков на множестве путей $\mathcal{P}$, через $\mathbf{0}_{|\mathcal{P}|}$ — нулевой вектор размерности, равной количеству элементов в множестве $\mathcal{P}$, через $VI(\mathcal{P})$ — вариационное неравенство вида:

$$\sum_{p \in \mathcal{P}} G_p(x^\dagger(\mathcal{P}))(x_p(\mathcal{P}) - x_p^\dagger(\mathcal{P})) \geqslant 0,$$
$$x(\mathcal{P}) = (x_p : p \in \mathcal{P}) \in X(\mathcal{P}) = \prod_{w \in W} X_w(\mathcal{P}_w),$$

где $X_w(\mathcal{P}_w) = \left\{x_p \geqslant 0 : p \in \mathcal{P}_w, \sum\limits_{p \in \mathcal{P}_w} x_p = \rho_w\right\}$ — допустимое множество потоков для пары $w$ на множестве путей $\mathcal{P}_w$ и $\mathcal{P} = \bigcup\limits_{w \in W} \mathcal{P}_w$. В силу теоремы 1 вектор $x^\dagger(\mathcal{P}) \in X(\mathcal{P})$ является решением вариационного неравенства $VI(\mathcal{P})$.

Пусть для каждой пары $w \in W$ задано некоторое непустое подмножество допустимых маршрутов $P_w^0 \subseteq P_w$. Тогда текущее множество путей в сети определяется как $P^0 = \bigcup\limits_{w \in W} P_w^0$.

Предположим, что для $P^0$ в процессе решения вариационного неравенства $VI(P^0)$ найдены равновесные потоки $x^\dagger(P^0)$. При этом по всем путям $p \in P \setminus P^0 = \bar{P}^0$ движения нет, поэтому $x(\bar{P}^0) = \mathbf{0}_{|\bar{P}^0|}$. Вектор $x^0 = (x^\dagger(P^0), \mathbf{0}_{|\bar{P}^0|})$ содержится в $X$, т. е. является допустимым распределением потоков в сети. Осталось проверить, является ли $x^0$ равновесным.

Потоки $x^0$ характеризуют загрузку сети и приводят к определенным задержкам на ее дугах. Если для некоторой пары $w \in W$ при текущем состоянии сети найден маршрут $q_w \notin P_w^0$ такой, что

$$G_{q_w}(x^0) < u_w(x^0), \tag{39}$$

то, очевидно, распределение $x^0$ не является равновесным и необходимо перераспределить потоки уже с учетом $q_w$. Если ни для одной из потокообразующих пар такого пути не существует, то потоки $x^0$ удовлетворяют условиям равновесия (1).

Обнаружить наименее затратный путь $q_w$ можно при помощи специальных алгоритмов. Возможно, что такие пути существуют для нескольких потокообразующих пар, поэтому при перераспределении потоков предпочтительно учитывать все такие пути. Новую загрузку сети можно получить в результате решения вариационного неравенства $VI(P^1)$, где $P^1 = P^0 \bigcup\limits_{w \in W} q_w$. Снова получаем допустимое распределение потоков

$x^1 = (x^\dagger(P^1), \mathbf{0}_{|\bar{P}^1|}) \in X$, которое необходимо проверить на соответствие условиям равновесия (1). Процесс будет повторяться до тех пор, пока существуют пути $q_w$, удовлетворяющие (39).

Приведенные выше рассуждения и декомпозированный проекционный метод (38) приводят к следующему алгоритму поиска потокового равновесия в транспортных сетях.

**Алгоритм 1**

Шаг 0. Сформировать множество $P^0 = \bigcup\limits_{w\in W} P_w^0$, где $P_w^0 \subseteq P_w$ и $P_w^0 \neq \varnothing$ для всех $w \in W$. Положить $k = 0$.

Шаг 1. Найти решение $x^\dagger(P^k) \in X(P^k)$ вариационного неравенства $VI(P^k)$ проекционным методом (38).

Шаг 2. Построить множество $Q^k = \{q_w \colon w \in W,\ \ G_{q_w}(x^k) < u_w(x^k)\}$. Если $Q^k = \varnothing$, то текущее распределение потоков $x^k$ является равновесным, алгоритм заканчивает работу. В противном случае сформировать множество $P^{k+1} = P^k \cup Q^k$, положить $k = k + 1$ и повторить шаг 1.

Так как $|P^{k+1}| > |P^k|$ и множество всех путей в графе $\Gamma(V, E)$ конечно, то приведенный алгоритм 1 за конечное число шагов сойдется к равновесному распределению потоков $x^\dagger \in X$.

Заметим, что с каждой следующей итерацией алгоритма 1 размерность решаемого вариационного неравенства $VI(P^k)$ увеличивается, что может весьма усложнить численные расчеты. Для преодоления «проклятия размерности» предлагается следующая модификация алгоритма 1. При $Q^k \neq \varnothing$ на шаге 2 множество $P^{k+1}$ строится по правилу

$$P^{k+1} = [P^k]_+ \cup Q^k, \quad [P^k]_+ = \{p \in P^k \colon x_p^\dagger > 0\}, \tag{40}$$

т. е. к началу следующей итерации допустимое множество маршрутов $P^{k+1}$ формируется только из путей множества $P^k$, которые участвуют в переносе трафика, и найденных кратчайших путей, не учтенных в $P^k$.

Несмотря на то что применение правила (40) исключает последовательное накопление допустимых маршрутов и есть вероятность появления в $P^k$ ранее рассмотренного пути, алгоритм 1 остается конечным, что гарантирует следующий результат [26].

Рассмотрим оценочную функцию $\varphi(x) = \max_{\eta\in X} G(x)(x - \eta) \geqslant 0$ для вариационного неравенства (6). Известно [41, 49, 50], что точка $x^\dagger \in X$ является решением $VI(P)$ тогда и только тогда, когда $x^\dagger = \operatorname{argmin}\{\varphi(x) \colon x \in X\}$ и $\varphi(x^\dagger) = 0$.

**Утверждение 3.** *Если $x^k \in X$ не является равновесной загрузкой сети и множество $P^{k+1}$ формируется по правилу* (40), *то* $\varphi(x^{k+1}) < \varphi(x^k)$.



Доказательство. Пусть для множества маршрутов $P^k$ определено равновесное распределение потоков $x^*(P^k)$ и множество $Q^k \neq \varnothing$. Следуя алгоритму 1, строим новое множество $P^{k+1}$ по правилу (40) и для него находим равновесное распределение потоков $x^*(P^{k+1})$.

Рассмотрим оценочную функцию

$$\varphi_{P^{k+1}}(x) = \max_{\eta\in X(P^{k+1})} G(x)(x - \eta),$$

определенную для вариационного неравенства $VI(P^{k+1})$. Следовательно,

$$x^*(P^{k+1}) = \operatorname{argmin}\{\varphi_{P^{k+1}}(x) \colon x \in X(P^{k+1})\}, \quad \varphi_{P^{k+1}}(x^*(P^{k+1})) = 0.$$

Построим вектор $z = (z_p \colon p \in P^{k+1})$ по правилу: $z_p = x_p^*$ при $p \in [P^k]_+$ и $z_p = 0$ при $p \in Q^k$. Очевидно, что $z \in X(P^{k+1})$, поэтому имеет место соотношение

$$0 = \varphi_{P^{k+1}}(x^*(P^{k+1})) \leqslant \varphi_{P^{k+1}}(z).$$

Оценочную функцию $\varphi_{P^{k+1}}(z)$ можно переписать в виде

$$\varphi_{P^{k+1}}(z) = \sum_{w\in W} \sum_{p\in P_w^{k+1}} z_p G_p(z) - \sum_{w\in W} G_{q_w}(x^*(P^k)) d_w = \\ = \sum_{w\in W} \sum_{p\in P_w^{k+1}} (G_p(z) - G_{q_w}(x^*(P^k)))z_p > 0.$$

Следовательно, $\varphi_{P^{k+1}}(x^*(P^{k+1})) < \varphi_{P^{k+1}}(z)$.

Рассмотрим два вектора

$$x^k = (x^*(P^k), \mathbf{0}_{|\bar{P}^k|}) \in X, \quad x^{k+1} = (x^*(P^{k+1}), \mathbf{0}_{|\bar{P}^{k+1}|}) \in X.$$

Имеет место оценка

$$\varphi(x^{k+1}) = \sum_{p\in P^{k+1}} G_p(x^{k+1}) x_p^{k+1} - \sum_{w\in W} u_w(x^{k+1}) d_w < \\ < \sum_{p\in [P^k]_+} G_p(x^k) x_p^k = \sum_{p\in P} G_p(x^k) x_p^k = \varphi(x^k),$$

что доказывает сходимость алгоритма. □

Из утверждения 3 следует, что если $k \neq l$, то $P^k \neq P^l$, поэтому модификация (40) не нарушает конечности алгоритма 1.

### 1.4.4. Ступенчатая регулировка шага проекционного метода

Проекционные методы относятся к классу так называемых фейеровских алгоритмов [8, 12], получивших широкое распространение при решении задач допустимости, оптимизации, дополнительности, равновесия и др. Обобщение свойства фейеровости было предложено в [16], что позволило

направить порождаемый фейеровским оператором процесс к некоторому заданному подмножеству данного множества.

С целью ускорения сходимости процесса (38) к равновесному распределению потоков предлагается использовать теорию фейеровских процессов с адаптивным шагом [17, 58]. Через $U(y, \varepsilon) = \{x\colon \|x - y\| < \varepsilon\}$ обозначим $\varepsilon$-окрестность точки $y$. Будем использовать следующие определения.

**Определение 5.** Отображение $R\colon X \to \mathbb{R}^n$ называется локально сильным аттрактантом подмножества $Z \subset X$, если для любой точки $x' \in X\backslash Z$ найдутся такие $\varepsilon > 0$ и $\gamma > 0$, что для любых $x \in U(x', \varepsilon)$ и $z \in Z$ выполнено $R(x)(z - x) \geqslant \gamma$.

**Определение 6.** Оператор $F$ называется локально сильно фейеровским относительно множества $X$, если $F(x) = x$ для любых $x \in X$ и для любого $y' \notin X$ существуют окрестность $U(y', \varepsilon)$ и число $\alpha \in [0, 1)$ такие, что для всех $y \in U(y', \varepsilon)$ и $x \in X$ выполнено $\|F(y) - x\| \leqslant \alpha\|y - x\|$.

Обозначим через $D(k, m) = \operatorname{conv}\{d^k, d^{k+1}, \ldots, d^m\}$ выпуклую оболочку векторов $\{d^k, d^{k+1}, \ldots, d^m\}$, через $\mathcal{N}_X(z) = \{v\colon v(x - z) \geqslant 0,\ \ x \in X\}$ — нормальный конус к множеству $X$ в точке $z$.

Рассматривается итерационный процесс

$$x^{k+1} = x^k + \lambda_k d^k,$$
$$d^k = \frac{F(x^k + \lambda_k R(x^k)) - x^k}{\lambda_k}, \quad k = 0, 1, 2, \ldots, \tag{41}$$

где размер шагового множителя $\lambda_k$ выбирается на основе накопленной информации за предыдущие итерации. Для заданной последовательности $\theta_m \to +0$ при $m \to \infty$ определим последовательность индексов $\{k_m\}$ и шаговые множители $\lambda_k$ по следующим правилам:

1) Для $m = 0$ определим $k_m = 0$ и зафиксируем произвольное положительное $\lambda_0$. Выберем $\alpha \in (0, 1)$.

2) Для данных $m$ и $k_m$ определим такой индекс $k_{m+1}$, что $0 \notin D(k_m, k) + \theta_m B$, $k_m \leqslant k < k_{m+1}$, $0 \in D(k_m, k_{m+1}) + \theta_m B$ при $\lambda_k = \lambda_{k_m}$ для $k_m \leqslant k < k_{m+1}$.

3) Положим $\lambda_{k_{m+1}} = \alpha\lambda_{k_m}$.

4) Увеличиваем номер итерации $m = m + 1$ и повторяем (41) для текущего значения $\lambda_k$.

Обоснование сходимости описанного процесса к некоторому стационарному множеству $Z^\dagger = \{z^\dagger \in X\colon 0 \in \mathcal{N}_X(z^\dagger) + R(x^\dagger)\}$ дает следующая теорема.

**Теорема 7** (см. [17]). *Пусть оператор $F$ — локально сильно фейеровский относительно $X$, отображение $R(x)$ — локально сильный аттрактант множества $Z^\dagger$. Если последовательность $\{x^k\}$, порожденная процессом* (41), *ограничена, то все ее предельные точки принадлежат $Z^\dagger$.*



Перепишем рекуррентные соотношения (38) в эквивалентной форме:

$$x^k = ((x^w)^k\colon w \in W), \quad d^k = ((d^w)^k\colon w \in W),$$
$$(d^w)^k = \frac{\pi_{X_w}((x^w)^k - \lambda_k G^w(x^k)) - (x^w)^k}{\lambda_k}, \tag{42}$$
$$x^{k+1} = x^k + \lambda_k d^k, \quad k = 0, 1, 2, \ldots,$$

где размер шага $\lambda_k$ регулируется по правилам 1–4.

Операция проектирования $\pi_X(\cdot)$ является ярким представителем локально сильно фейеровских относительно $X$ операторов. Геометрическая интерпретация решения вариационного неравенства (6) говорит о том, что $G(x^\dagger) \in \mathcal{N}_X(x^\dagger)$. Так как последовательность $\{x^k\}$, генерируемая процессом (42), принадлежит компактному множеству $X$, то из теоремы 7 непосредственно вытекает следующий результат.

**Теорема 8.** *Если $(-G(x))$ — локально сильный аттрактант множества решений $X^\dagger$ вариационного неравенства* (6), *то все предельные точки последовательности $\{x^k\}$, порожденной процессом* (42) *и правилами* 1–4, *принадлежат $X^\dagger$.*

Существует связь между свойствами аттрактантности и псевдомонотонности.

**Утверждение 4.** *Если отображение $G(x)$ вариационного неравенства* (6) *строго псевдомонотонно на $X$, то $(-G(x))$ является локально сильным аттрактантом единственного решения $x^\dagger \in X^\dagger$.*

Доказательство. Произвольно выберем точку $x' \in X$ и определим такую ее окрестность $U(x', \varepsilon)$, что $x^* \notin U(x', \varepsilon)$. В силу псевдомонотонности и непрерывности $G$ имеет место оценка

$$\inf_{x \in U(x, \varepsilon)} (-G(x)(x^* - x)) \geqslant \inf_{x \in \bar{U}} (-G(x)(x^* - x)) = \gamma \geqslant 0.$$

Значение $\gamma = 0$ противоречит строгой псевдомонотонности $G$, поэтому $\gamma > 0$ и для любых $x \in U(x', \varepsilon)$ выполнено

$$-G(x)(x^* - x) \geqslant \gamma > 0,$$

отсюда $(-G)$ — локально сильный аттрактант $X^*$. □

Нетрудно показать, что если $G$ — строго псевдомонотонно, то вариационное неравенство (6) имеет единственное решение. Действительно, пусть $x^1, x^2 \in X^\dagger$, $x^1 \neq x^2$. Имеют место неравенства:

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

что противоречит строгой псевдомонотонности $G$, следовательно, не может существовать более одного решения. Учитывая это свойство и утверждение 4, приходим к следующему выводу.

**Следствие 2.** *Если $G(x)$ — строго псевдомонотонно на $X$, то последовательность $\{x^k\}$, порожденная процессом* (42) *и правилами* 1–4, *имеет единственную предельную точку $x^\dagger \in X^\dagger$.*

Таким образом, полученые на основе теории фейеровских процессов условия сходимости проекционного метода (35) для решения задач транспортного равновесия с фиксированным спросом менее ограничительные, чем ранее приведенные.

## 1.5. Построение матрицы корреспонденций

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

Существуют разные методики для вычисления элементов матрицы $\rho = (\rho_{ij} : i \in S, j \in D)$, в том числе с применением математических моделей. Рассмотрим наиболее часто используемые, а именно гравитационную и энтропийную модели построения матрицы корреспонденций. Описание указанных моделей для транспортных сетей можно найти, например, в работах [6, 7, 9, 20, 21, 28, 40].

### 1.5.1. Гравитационная модель

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

$$\rho_{ij} = \varkappa \frac{s_i d_j}{c_{ij}^2}, \quad i \in S, \quad j \in D, \tag{43}$$

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

Система (43) обладает существенным недостатком. Нетрудно видеть, что при увеличении объемов $s_i$ и $d_j$, например, в два раза модель (43) приведет к увеличению корреспонденции $\rho_{ij}$ в четыре раза, что совершенно нелогично. Поэтому вместо классической гравитационной модели (43) на практике используют ее модификацию, в которой к условию (43) добавляют дополнительные условия, например, балансовые ограничения на выезд и въезд. Кроме того, квадрат расстояния (затрат) $c_{ij}^2$ заменяют на так называемую функцию тяготения $f(c_{ij})$, характеризующую предпочтения индивидуумов при выборе пары источник-сток $(i, j)$ для передвижения. В результате модифицированная гравитационная модель имеет вид

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

или, что то же самое:

$$\rho_{ij} = \alpha_i \beta_j s_i d_j f(c_{ij}), \quad i \in S, \quad j \in D, \tag{44}$$

где калибровочные коэффициенты $\alpha_i$ и $\beta_j$ определяются из системы

$$\alpha_i = \left[ \sum_{j \in D} \beta_j d_j f(c_{ij}) \right]^{-1}, \quad \beta_j = \left[ \sum_{i \in S} \alpha_i s_i f(c_{ij}) \right]^{-1}. \tag{45}$$

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

Выбор функции тяготения $f$ осуществляется либо в процессе калибровки модели на основе сопоставления расчетных данных по модели и эмпирических наблюдений, либо на основе некоторых соображений о предпочтениях при выборе пары источник-сток. Одна из аппроксимаций функции имеет следующий вид: $f(c_{ij}) = \exp(-\gamma c_{ij}^\theta)$, где при расчете корреспонденций трудовых миграций полагают $\gamma \approx 0{,}065$, $\theta \approx 1$ (см., например, [28] и ссылки там).

Важно отметить, что величины $\alpha_i$ и $\beta_j$ зависят от всего набора $s_i$ и $d_j$, а следовательно, и объемы корреспонденций $\rho_{ij}$ зависят от загрузки всей системы.

Численные значения $\alpha_i$ и $\beta_j$ определяют специальной итеративной процедурой. В отечественной литературе такая процедура известна как метод балансировки Шацкого—Шелейховского [27, 30]. В зарубежной литературе метод балансировки имеет свою независимую историю развития. Например, в работе [33] описана следующая процедура: начиная с матрицы

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

каждая итерация метода состоит из последовательности операций:

$$\rho_{ij}^k = \begin{cases} \rho_{ij}^k d_j \Big[\sum\limits_{i\in S} \rho_{ij}^k\Big]^{-1}, & \text{если } \sum\limits_{i\in S} \rho_{ij}^k > d_j, \\ \rho_{ij}^k & \text{в противном случае,} \end{cases}$$

$$q_i = s_i - \sum_{j\in D} \rho_{ij}^k, \quad r_j = d_j - \sum_{i\in S} \rho_{ij}^k; \tag{46}$$

$$\rho_{ij}^{k+1} = \rho_{ij}^k + q_i r_j f(c_{ij}) \Big[\sum_{l\in D} r_l f(c_{il})\Big]^{-1}.$$

Вычислительные эксперименты по расчету матрицы корреспонденций на примере УДС Владивостока [18, 19] (строилась матрица размерности $638 \times 638$) показали высокую скорость сходимости процесса (46) к искомой матрице корреспонденций — сбалансированная матрица была получена всего за 4 итерации.

### 1.5.2. Энтропийная модель

Как и в случае гравитационного подхода, идею построения энтропийной модели подсказала физика, а именно второй закон термодинамики, утверждающий, что любая замкнутая физическая система стремится достичь устойчивого равновесного состояния, которое характеризуется максимумом энтропии этой системы. Впервые концепция энтропии для определения матрицы корреспонденций была использована в работе [63].

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

Пусть задано фиксированное пространственное распределение населения по зонам, порождающим потоки, — как и ранее, назовем такие зоны *источниками* и объединим их в множество $S$ — и по зонам, поглощающим потоки, — назовем их *стоками* и объединим в множество $D$. Источниками, например, могут служить районы жилых массивов, стоками — места приложения труда. Индивидуумы в транспортной системе перемещаются от источников к стокам. Предположим, что каждый индивидуум имеет уникальный идентификатор, например номер паспорта. Состояние транспортной системы определяется распределением «помеченных» индивидуумов между парами источник-сток.



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

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

$$P(\rho) = \nu(\rho) Q(\rho). \tag{47}$$

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

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

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

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

$$Q(\rho) = C_{\mathcal{R}}^{\rho_{11}} \cdot C_{\mathcal{R}-\rho_{11}}^{\rho_{12}} \cdot C_{\mathcal{R}-\rho_{11}-\rho_{12}}^{\rho_{13}} \cdot \ldots \cdot C_{\mathcal{R}-\sum\limits_{i=1}^{n-1}\sum\limits_{j=1}^{m-1}\rho_{ij}}^{\rho_{nm}} =$$

$$= \frac{\mathcal{R}!}{(\mathcal{R}-\rho_{11})!\,\rho_{11}!} \cdot \frac{(\mathcal{R}-\rho_{11})!}{(\mathcal{R}-\rho_{11}-\rho_{12})!\,\rho_{12}!} \times$$

$$\times \frac{(\mathcal{R}-\rho_{11}-\rho_{12})!}{(\mathcal{R}-\rho_{11}-\rho_{12}-\rho_{13})!\,\rho_{13}!} \cdot \ldots \cdot \frac{\Big(\mathcal{R}-\sum\limits_{i=1}^{n-1}\sum\limits_{j=1}^{m-1}\rho_{ij}\Big)!}{\rho_{nm}!} = \frac{\mathcal{R}!}{\prod\limits_{i=1}^{n}\prod\limits_{j=1}^{m}\rho_{ij}!}.$$

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

Подставив рассчитанные значения $\nu(\rho)$ и $Q(\rho)$ в формулу (47), получаем критерий выбора наиболее вероятного состояния системы:

$$P(\rho) = \mathcal{R}! \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{\nu_{ij}^{\rho_{ij}}}{\rho_{ij}!} \to \max. \tag{48}$$

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

$$\sum_{j=1}^{m} \rho_{ij} = s_i, \quad \sum_{i=1}^{n} \rho_{ij} = d_j, \quad \rho_{ij} \geqslant 0, \quad i \in S, \ \ j \in D. \tag{49}$$

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

$$\sum_{i=1}^{n} s_i = \sum_{j=1}^{m} d_j = \mathcal{R}. \tag{50}$$

Дополнительно к условиям баланса (49) введем ограничение на общие затраты при проезде:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \rho_{ij} = C, \tag{51}$$

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

Таким образом, проблема построения матрицы корреспонденций $\rho = (\rho_{ij} : i \in S, j \in D)$ сводится к задаче условной оптимизации (48), (49), (51).

Нет сомнений, что в заданной форме (48) функция $P(\rho)$ весьма неприятна для оптимизации. Для удобства максимизации можно воздействовать на $P(\rho)$ любым монотонным оператором, например, прологарифмировать $P(\rho)$ и вместо (48) использовать критерий

$$\ln P(\rho) = \ln \mathcal{R}! + \sum_{i=1}^{n} \sum_{j=1}^{m} (\rho_{ij} \ln \nu_{ij} - \ln \rho_{ij}!) \to \max. \tag{52}$$

Проводя параллель между физической и транспортной системами, отметим наличие большого количества неуправляемых элементов, что позволяет предположить, что значения $\rho_{ij}$ достаточно велики. Поэтому вполне правомерно для дальнейшего преобразования критерия (52) использовать формулу Стирлинга $\ln z! = z \ln z - z$, которая справедлива при больших $z$. Имеем

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

При фиксированных объемах выездов $s_i$ и въездов $d_j$ и выполнении равенства (50) величина $\mathcal{R} \ln \mathcal{R}$ постоянна и может быть исключена из критерия.

В результате проведенных преобразований наиболее вероятное состояние транспортной системы будет соответствовать такой матрице корреспонденций $\rho$, элементы которой удовлетворяют условиям (49), (51) и критерию

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \rho_{ij} \ln \frac{\nu_{ij}}{\rho_{ij}} \to \max. \tag{53}$$

При построении энтропийной модели (49), (51), (53) предполагалось, что известна априорная информация о предпочтении индивидуумом одной коммуникации другой. Если же любое состояние система принимает с равной вероятностью, то есть для любых пар $(i, j)$ значение $\nu_{ij}$ постоянно и определяется как $\nu_{ij} = \dfrac{1}{mn}$, то вместо критерия (53) рассматривают

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \rho_{ij} \ln \frac{1}{\rho_{ij}} \to \max. \tag{54}$$

Допустимая область, задаваемая условиями (49), (51), образует полиэдральное множество. Целевая функция критерия (53) на допустимой области является строго вогнутой. В самом деле, матрица Гессе для (53) имеет вид диагональной матрицы размерности $mn \times mn$ с элементами на главной диагонали $\left\{ -\dfrac{1}{x_{ij}} \right\}$. Такая матрица отрицательно определена для любых $m$, $n$ и $x_{ij} \geqslant 0$. Таким образом, задача (49), (51), (53) относится к классу задач выпуклой гладкой оптимизации. Строгая вогнутость целевой функции гарантирует единственность ее решения. Несмотря на свои хорошие свойства, для реальных транспортных сетей задача (49), (51), (53) имеет большую размерность, что в свою очередь серьезно усложняет применение на практике стандартных для этого класса задач численных методов. Так, например, для расчета трудовых миграций в УДС Владивостока [18, 19] территория города была поделена на зоны $800 \times 800$ метров. В результате получилась сетка $22 \times 29$ квадратов, каждый из которых одновременно являлся зоной-источником и зоной-стоком, при этом размерность задачи (49), (51), (53) составила 407 044 переменных, 1277 ограничений равенств (49), (51).

Для решения задачи (49), (53) разработана простая итерационная схема [27,30]: начиная с матрицы $\rho^0 = (\rho_{ij}^0 = \nu_{ij} : i \in S,\ j \in D)$ на каждой итерации метода попеременно достигается выполнение балансовых ограничений для выездов и въездов:

$$\rho_{ij}^k = \rho_{ij}^k s_i \left[\sum_{j\in D} \rho_{ij}^k\right]^{-1}, \quad \rho_{ij}^{k+1} = \rho_{ij}^k d_j \left[\sum_{i\in S} \rho_{ij}^k\right]^{-1}. \tag{55}$$

В работе [4] доказана сходимость процесса (55) к оптимальному решению задачи (49), (53). Существуют и другие подходы к решению энтропийных моделей (см., например, [10, 40]).

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

### 1.5.3. Связь между гравитационной и энтропийной моделями

Количество переменных в задаче (49), (51), (53), как правило, во много раз превышает число ограничений. Традиционно в такой ситуации вместо исходной решается двойственная задача, которая в данном случае заключается в максимизации функции Лагранжа:

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

где $\lambda = (\lambda_i : i \in S)$ — вектор двойственных переменных, соответствующих балансовым ограничениям (49) для источников, $\mu = (\mu_j : j \in D)$ — вектор двойственных переменных, соответствующих балансовым ограничениям (49) для стоков, $\gamma$ — двойственная переменная, соответствующая ограничению по затратам (51).

Точка максимума для $L(\rho, \lambda, \mu, \gamma)$ должна удовлетворять условиям (49), (51) и системе уравнений

$$\ln\frac{\nu_{ij}}{\rho_{ij}} - 1 - \lambda_i - \mu_j - \gamma c_{ij} = 0, \quad i \in S, \quad j \in D. \tag{56}$$

Из системы (56) можно выразить корреспонденции

$$\rho_{ij} = \nu_{ij}\exp(-1 - \lambda_i - \mu_j - \gamma c_{ij}). \tag{57}$$

Видим, что для $\nu_{ij} \geqslant 0$ условие неотрицательности корреспонденций $\rho_{ij}$ выполнено автоматически, поэтому может не учитываться при построении двойственной задачи и применении к ней численных методов. Однако заметим, что случай, когда $\nu_{ij} = 0$, означает отсутствие корреспонденции между парой $(i, j)$, следовательно, $\rho_{ij} = 0$ и максимизация функции Лагранжа должна рассматриваться в пространстве меньшей размерности.



Введем обозначения

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

Тогда выражение (57) перепишется в виде

$$\rho_{ij} = \alpha_i\beta_j s_i d_j \nu_{ij}\exp(-\gamma c_{ij}). \tag{58}$$

При подстановке (58) в балансовые ограничения (49) определяются параметры $\alpha_i$ и $\beta_j$:

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

Отметим, что величина $C$ на практике, как правило, неизвестна, поэтому лагранжевый множитель $\gamma$ нельзя определить из решения уравнения (51). Значение $\gamma$ определяется обычными методами калибровки.

Сравнивая выражение (58) с гравитационной моделью (44), видим, что отличие между ними состоит только в аналитическом задании функции тяготения $f(c_{ij})$. При $f(c_{ij}) = \nu_{ij}\exp(-\gamma c_{ij})$ гравитационная (44) и энтропийная (49), (51), (53) модели эквивалентны. Таким образом, при однородной цели поездок, при заданных объемах выездов $s_i$, въездов $d_j$, затратах на передвижение $c_{ij}$, при фиксированных полных затратах $C$ существует наиболее вероятное распределение поездок между зонами $(i, j)$ и это распределение совпадает с тем, которое задается гравитационной моделью с экспоненциальной функцией притяжения.

## 1.6. Парадоксы транспортного равновесия

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

### 1.6.1. Парадокс Браесса

Пример Пигу (см. раздел 1.3) заставляет усомниться в эффективности «невидимой руки рынка» Адама Смита, которая, направляя эгоистичные действия пользователей сети, позволяет достичь общественного блага. Последующий пример Браесса показывает, что конкурентное бескоалиционное равновесие может не только отклоняться от системного оптимума, но и ухудшать ситуацию для всех участников движения.

Рассмотрим появление парадокса Браесса в результате последовательных весьма вероятных трансформаций транспортной сети окрестностей г. Владивостока, которые представлены в серии рис. 3.

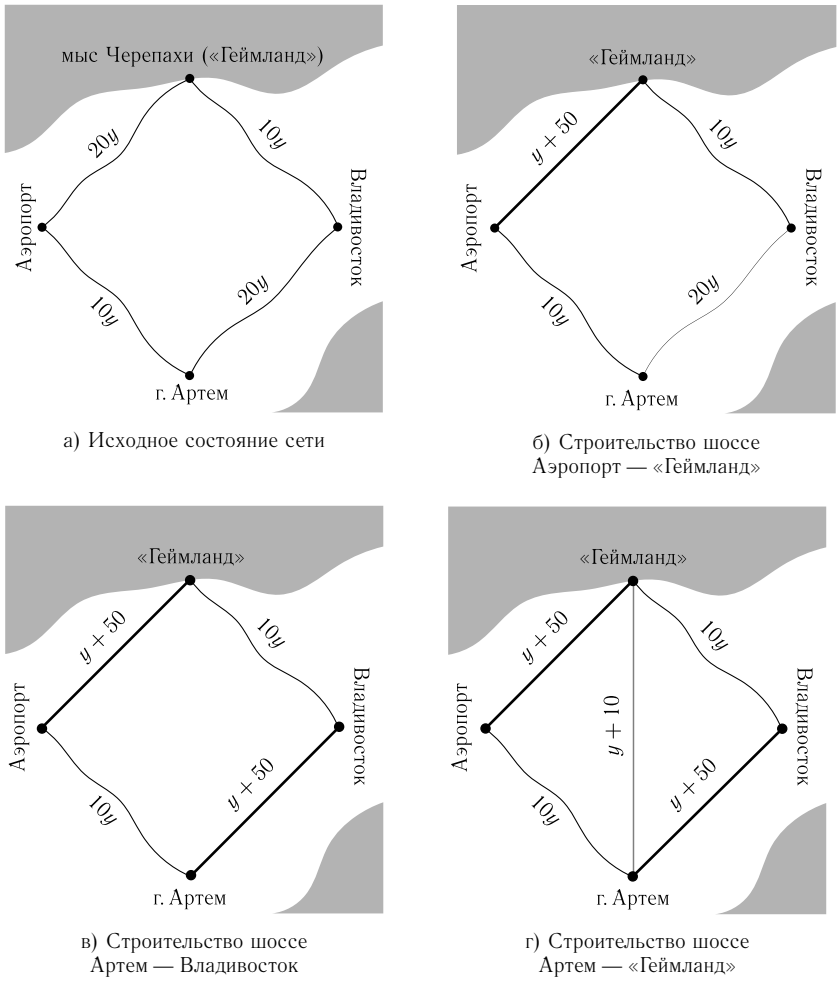



Рис. 3. Парадокс Браесса

Будем рассматривать ситуацию с точки зрения перевозок Аэропорт — Владивосток, с общей потребностью в перевозках 6 условных единиц. На рис. 3 изображены воображаемые этапы изменения участка транспортной сети в окрестности Владивостока, которые могут быть связаны с созданием игровой зоны на мысе Черепахи.

**а) Начальное состояние.** Аэропорт и Владивосток соединены двумя дорогами, одна из которых проходит через г. Артем, а другая через мыс Черепахи — место, где будет построена игровая зона «Геймланд».



В начальный момент обе дороги невысокого качества и, как показано на рис. 3 а), время проезда по ним сильно зависит от нагрузки $y$.

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

**б) Построено шоссе Аэропорт — «Геймланд».** Снизилась зависимость времени проезда из Аэропорта в «Геймланд», в затратах на проезд появилась постоянная составляющая, которая может представлять собой время проезда по пустой дороге. Часть равновесного трафика переместилась на направление Аэропорт — мыс Черепахи — Владивосток (3,17), соответственно поток по другому маршруту Аэропорт — Артем — Владивосток упал до 2,83. Пользовательские затраты на проезд составили 84,88, системные — $84{,}88 \cdot 6 = 509{,}28$.

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

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

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

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

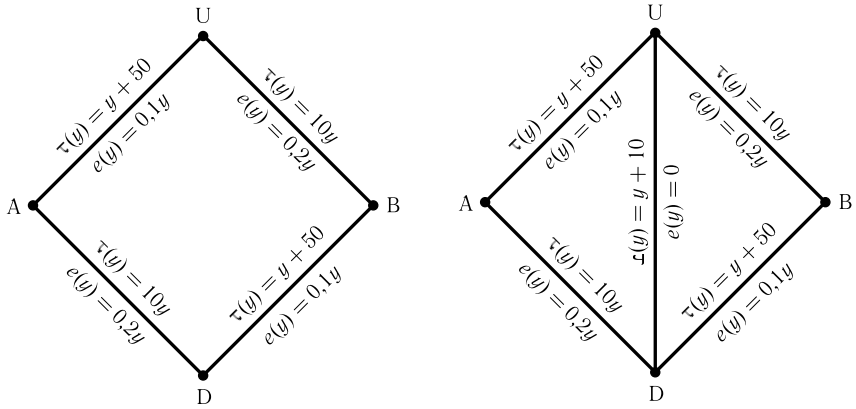
Причиной этого эффекта является то, что постройка на этапе г) шоссе Артем — «Геймланд» создала оппортунистическую возможность проехать по маршруту Аэропорт — Артем — «Геймланд» — Владивосток. При нулевом потоке на маршруте Артем — «Геймланд» временные затраты составляют 70 и провоцируют водителей на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

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

### 1.6.2. Транспортно-экологические парадоксы

Существует ряд парадоксов [55], связанных с транспортными ситуациями, в которых помимо времени проезда учитываются и дополнительные критерии. Одним из таких критериев, важных в настоящее время, является загрязнение окружающей среды (ЗОС). ЗОС является сложным многокомпонентным понятием, включающим различные виды ущерба для окружающей среды: газовое и тепловое загрязнение, разрушение сложившихся природных ландшафтов, мест обитания редких животных и пр. В данном случае будем все же считать, что ЗОС измеряется некоторым универсальным показателем, связанным с данным участком дороги и зависящим, вообще говоря, от потока транспорта по этой дороге. ЗОС от различных участков дороги суммируются, образуя итоговый ЗОС либо от маршрутов, либо от всей транспортной сети в целом.

**1.6.2.1. Экологический парадокс Браесса.** В своей схематической форме сеть, реализующая парадокс Браесса, представлена на рис. 4, где на каждом ребре показаны как временные затраты $\tau(y)$, так и экологический ущерб $e(y)$ как функции потоков по этим ребрам $y$. Взяв для предполага-




а) Исходное состояние сети

б) После строительства новой дороги

Рис. 4. Парадокс Браесса

емого потока те же данные, что в предыдущем примере, оценим ЗОС до и после строительства новой дороги. Для исходного состояния сети ЗОС оценивается как

$$E = 2 \cdot 3 \cdot (0{,}2 + 0{,}1) = 1{,}8.$$



После строительства новой дороги с нулевым экологическим ущербом ЗОС становится равным

$$E = 4 \cdot 0{,}2 + 2 \cdot 0{,}1 + 2 \cdot 0 + 4 \cdot 0{,}2 + 2 \cdot 0{,}1 = 2!$$

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

**1.6.2.2. Экологический треугольник.** Рассмотрим теперь еще более простую транспортную сеть, представленную на рис. 5. Так же как и ранее,

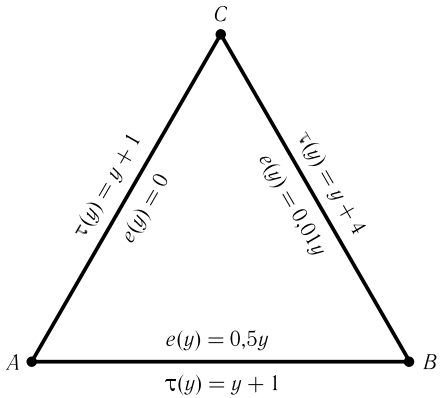



Рис. 5. Уменьшение перевозок вызывает увеличение ЗОС

на дугах этой сети приведены формулы, описывающие временные затраты $\tau(y)$ и экологический ущерб $e(y)$ как функции потока $y$ по этой дуге. Пусть требуется перевезти 2 единицы груза из $C$ в $B$ и одну единицу из $C$ в $A$. Очевидно, что достаточно рассмотреть 3 маршрута: $p_1 = C \to A$, $p_2 = C \to A \to B$ и $p_3 = C \to B$. Условия равновесия совместно с условиями удовлетворения спроса на перевозки дают систему уравнений

$$(x_1 + x_2) + 1 + x_2 + 1 = x_3 + 4,$$
$$x_2 + x_3 = 2, \quad x_1 = 1,$$

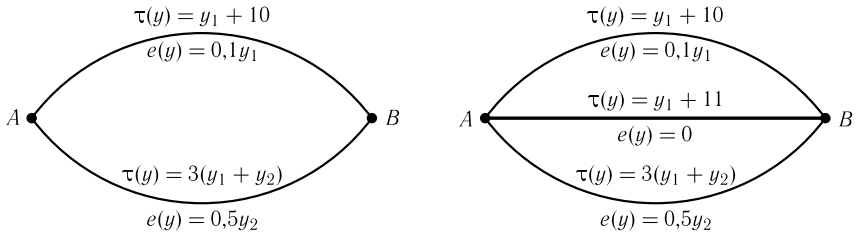
где $x_i$, $i = 1, 2, 3$, — это потоки по маршрутам $p_i$, $i = 1, 2, 3$. Решение этой системы дает равновесные потоки $x_1^\dagger = x_2^\dagger = x_3^\dagger = 1$ с общим экологическим ущербом $E = 0{,}51$.

Теперь предположим, что спрос на перевозки по маршруту $C \to A$ упал до $1/2$. Тогда решение аналогичной системы дает $x_1^\dagger = 1/2$, $x_2^\dagger = 7/6 > 1$, $x_3^\dagger = 5/6 < 1$ с общим экологическим ущербом $E = 0{,}5 \cdot 7/6 + 0{,}01 \cdot 5/6 \approx 0{,}591 > 0{,}51$.

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

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

Рассмотрим дорожную сеть, изображенную на рис. 6, часть а). Предположим, что эта сеть предназначена для перемещения автомобиль-




а) Исходная сеть б) Построена рокадная дорога

Рис. 6. Новая рокадная дорога вызывает увеличение ЗОС

ного трафика в объеме 5 условных единиц из точки $A$ в точку $B$ по двум дублирующим дорогам различного качества, временные затраты по которым задаются соотношениями на соответствующих дугах. В этих соотношениях $y_1$ означает поток по верхней дуге, $y_2$ — по нижней. Зависимость времени проезда по нижней дуге от потока по верхней может быть вызвана указанием приоритета на соответствующем перекрестке.

Определяющая система уравнений для равновесных потоков имеет вид

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

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

Если, как показано на правой части рис. 6, построена новая дорога с нулевым ущербом для окружающей среды и временными характеристиками $\tau(y) = y + 11$, то новая определяющая система будет иметь вид

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

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



**1.6.2.4. Перераспределение спроса.** Еще один пример показывает, что перенос части пассажиров с транспортного средства экологически более вредного (скажем, автобус) на менее вредный (например, трамвай) может в действительности увеличить ЗОС.

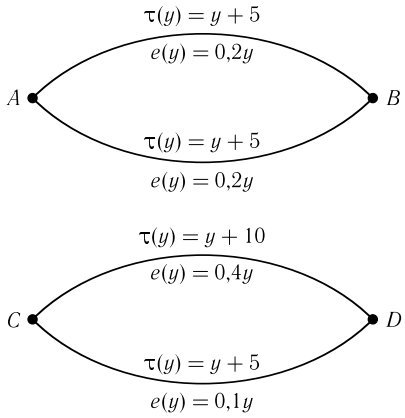



Рис. 7. Сети автомобильных дорог и трамвайных линий

На рис. 7 изображена транспортная сеть, состоящая из двух изолированных подсетей: автомобильной $AB$ и трамвайной $CD$. В пунктах $A$ и $C$ происходит разделение потока пассажиров: 10 единиц потока выбирают автобус, 5 единиц выбирают трамвай. Это распределение объемов перевозки считается заданным. Далее пассажиры, пользуясь свободой выбора, определяют, каким из двух автобусных и двух трамвайных маршрутов воспользоваться, и делают это основываясь на теории равновесия. В данном случае равновесие будет заключаться в равенстве временных затрат по каждой паре маршрутов и соответствующие потоки будут равны: в автомобильной подсети в силу симметрии общий объем перевозок поделится поровну: $y^1_{\text{авт}} = y^2_{\text{авт}} = 5$ с временными затратами для каждого пассажира в 10 временных единиц. Суммарный ЗОС в автомобильной подсети будет равен 2. В трамвайной подсети возникнет ситуация, описанная фактически в примере Пигу (см. раздел 1.3), и поток $y^1_{\text{трам}}$ по верхней дуге будет равен нулю, а по нижней дуге $y^2_{\text{трам}} = 5$ также с одинаковыми временными затратами на пассажира, равными 10. Суммарный ЗОС в трамвайной сети будет равен 0,5 и существенно ниже автомобильного. Ориентируясь на эти объемы ЗОС, может возникнуть идея перенести часть потока из автомобильной сети в трамвайную и уменьшить тем самым суммарный ЗОС, который равен первоначально 2,5. Новое равновесное решение будет иметь вид:

**Автомобильная сеть.** В силу симметрии потоки равны и составляют $y^1_{\text{авт}} = y^2_{\text{авт}} = 3{,}75$. Экологический ущерб $E = 1{,}5$.

**Трамвайная сеть.** $y^1_{\text{трам}} = 1{,}25$, $y^2_{\text{трам}} = 6{,}25$ с одинаковыми временными затратами 11,25 и суммарным экологическим ущербом $0{,}4 \cdot y^1_{\text{трам}} + 0{,}1 \cdot y^2_{\text{трам}} = 1{,}125$.

Суммарный экологический ущерб составляет 2,625, что превосходит (!) ущерб в предыдущем варианте распределения нагрузки для двух видов трафика.

То, что перенос трафика осуществляется в размере 2,5 единиц, на самом деле несущественно, любое уменьшение объема перевозок по автомобильной сети вызывает появление ненулевого объема перевозок по экологически затратной дуге 3 в трамвайной сети, что не компенсируется уменьшением объемов перевозок по дугам 1, 2 (считаем дуги пронумерованными сверху вниз). Увеличение объемов перевозок по автомобильной сети также не приводит к уменьшению ЗОС, так как вызывает в два раза больший экологический ущерб, чем уменьшение ЗОС от трамвайного трафика, который будет продолжать концентрироваться на дуге 4.

## 1.7. Практическая работа

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

1) задержка в дверях одинакова для входящих и выходящих и пропорциональна произведению интенсивностей входящего и выходящего потоков;

2) задержка в дверях пропорциональна интенсивности потока студентов, двигающихся в том же направлении, плюс задержка, пропорциональная произведению интенсивностей входящего и выходящего потоков.

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

**Упражнение 2.** Рассматривается транспортная сеть $\Gamma = (V, E)$ (пример сети взят из работы [36]), состоящая из 25 вершин ($|V| = 25$) и 40 ориентированных дуг ($|E| = 40$). Топология сети с направлением дуг представлена на рис. 8.

Дороги (дуги) транспортной сети $\Gamma$ поделены на четыре категории:

1) магистрали $E_{\text{h}} = \{(6 \to 7), (8 \to 9), (10 \to 11), (12 \to 13), (14 \to 15), (17 \to 18), (19 \to 20), (21 \to 22), (23 \to 24), (25 \to 16)\}$;



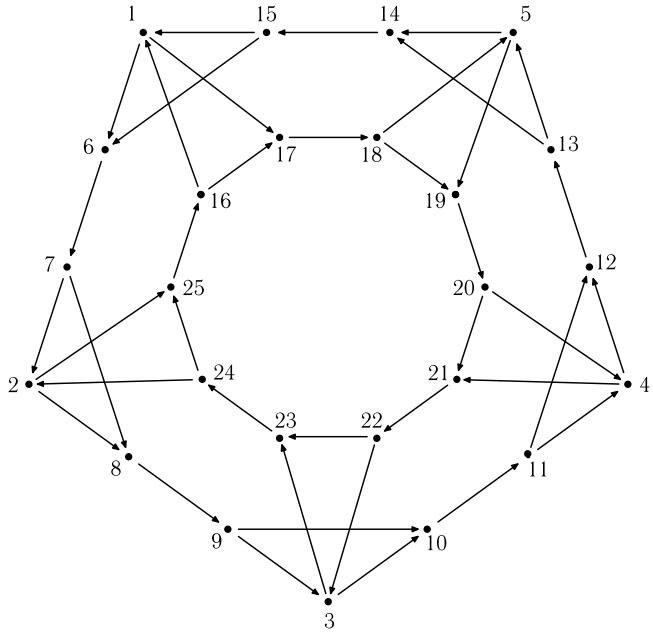



Рис. 8. Транспортная сеть $\Gamma = (V, E)$

2) выезды $E_{\text{ex}} = \{(16 \to 1), (15 \to 1), (24 \to 2), (7 \to 2), (22 \to 3), (9 \to 3), (20 \to 4), (11 \to 4), (18 \to 5), (13 \to 5)\}$;

3) въезды $E_{\text{en}} = \{(1 \to 6), (1 \to 17), (2 \to 25), (2 \to 8), (3 \to 23), (3 \to 10), (4 \to 21), (4 \to 12), (5 \to 19), (5 \to 14)\}$;

4) второстепенные дороги $E_{\text{s}} = \{(15 \to 6), (7 \to 8), (9 \to 10), (11 \to 12), (13 \to 14), (16 \to 17), (18 \to 19), (20 \to 21), (22 \to 23), (24 \to 25)\}$.

Пропускная способность магистралей равна 140 единицам потока, остальных дуг — 70 ед.

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

$$\varkappa_e = \begin{cases} 0{,}011, & e \in E_{\text{h}}, \\ 0{,}025, & e \in E_{\text{ex}} \cup E_{\text{en}}, \\ 0{,}033, & e \in E_{\text{s}}. \end{cases}$$

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

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

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

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

При моделировании необходимо учесть, что предпочтения при выборе пары $w$ определяются функцией $f(c_w) = \exp(-0{,}065 c_w)$, где $c_w$ — минимальные транспортные затраты на передвижение для пары $w$. Предполагается, что величина $c_w$ характеризует длину кратчайшего пути для каждой пары $w$ и определяется из соотношений: $c_w = 0{,}05$, если источник и сток совпадают, в противном случае

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

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

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

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



## Литература


1. *Антипин А. С.* Равновесное программирование: проксимальные методы // Автоматика и телемеханика. 1997. № 8. С. 125–137.
2. *Ашманов С. А.* Математические модели и методы в экономике. М.: Изд-во МГУ, 1980.
3. *Байокки К., Капело А.* Вариационные и квазивариационные неравенства. Приложения к задачам со свободной границей. М.: Наука, 1984.
4. *Брэгман Л. Д.* Доказательство сходимости метода Шелейховского для задачи с транспортными ограничениями // ЖВМ и МФ. 1967. Т. 7, № 1. С. 147–156.
5. *Бенсусан А., Лионс Ж.-Л.* Импульсное управление и квазивариационные неравенства. М.: Наука, 1987.
6. *Васильева Е. М., Левит Б. Ю., Лившиц В. Н.* Нелинейные транспортные задачи на сетях. М.: Финансы и статистика, 1981.
7. *Васильева Е. М., Игудин Р. В., Лившиц В. Н.* Оптимизация планирования и управления транспортными системами. М.: Транспорт, 1987.
8. *Васин В. В., Еремин И. И.* Операторы и итерационные процессы фейеровского типа. Теория и приложения. Екатеринбург: УрО РАН, 2005.
9. *Вильсон А. Дж.* Энтропийные методы моделирования сложных систем. М.: Наука, 1978.
10. *Гасникова Е. В.* Двойственные мультипликативные алгоритмы для задачи энтропийно-линейного программирования // ЖВМ и МФ. 2009. Т. 49, № 3. С. 453–464.
11. *Данскин Дж.* Теория максимина и ее приложение к задачам распределения вооружений. М.: Сов. радио, 1970.
12. *Еремина И. И., Мазурова В. Д.* Нестационарные процессы математического программирования. М.: Наука, 1979.
13. *Корпелевич Г. М.* Экстраградиентный метод для отыскания седловых точек и других задач // Экономика и матем. методы. 1976. Т. 1, № 4. С. 747–756.
14. *Никайдо Х.* Выпуклые структуры и математическая экономика. М.: Мир, 1972.
15. *Нурминский Е. А.* Использование дополнительных малых воздействий в фейеровских моделях итеративных алгоритмов // ЖВМ и МФ. 2008. Т. 48, № 12. С. 2121–2128.
16. *Нурминский Е. А.* Фейеровские процессы с малыми возмущениями // Доклады РАН. 2008. Т. 422, № 5. С. 601–605.

17. *Нурминский Е. А.* Фейеровские алгоритмы с адаптивным шагом // ЖВМ и МФ. 2011. Т. 51, вып. 5. С. 1–11.
18. *Нурминский Е. А., Шамрай Н. Б.* Моделирование транспортных потоков г. Владивостока на основе теории равновесия // Sisteme de transport si logistica: Materialele Conf. Int., Chisinau, 22–23 octombrie 2009; red. Resp. Dumitru Solomon; Acad. de Transporturi, Informatica si Comunicatii. Ch.: Evrica, 2009. P. 334–348.
19. *Нурминский Е. А., Шамрай Н. Б.* Прогнозное моделирование автомобильного трафика Владивостока // Труды МФТИ. 2010. Т. 2, № 4(8). С. 119–129.
20. *Попков Ю. С., Посохин М. В., Гутнов А. Э., Шмульян Б. Л.* Системный анализ и проблемы развития городов. М.: Наука, 1983.
21. *Попков Ю. С.* Макросистемные модели пространственной экономики. М.: КомКнига, 2008.
22. *Стенбринк П. А.* Оптимизация транспортных сетей. М.: Транспорт, 1961.
23. *Попов Л. Д.* Введение в теорию, методы и экономические приложения задач о дополнительности. Екатеринбург: Изд-во Урал. ун-та, 2001.
24. *Тодд М. Дж.* Вычисление неподвижных точек и приложения к экономике. М.: Наука, 1983.
25. *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.
26. *Шамрай Н. Б.* Поиск потокового равновесия проективными методами с использованием декомпозиции и генерации маршрутов // Автоматика и телемеханика. 2012. № 3. С. 150–165.
27. *Шацкий Ю. А.* Расчет схемы расселения и трудовых корреспонденций при разработке генерального плана города // Развитие системы городского транспорта. Киев, 1971. № 4. С. 3–14.
28. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.
29. *Швецов В. И.* Алгоритмы распределения транспортных потоков // Автоматика и телемеханика. 2009. № 10. С. 148–157.
30. *Шелейховский Г. В.* Транспортные основания композиции городского плана. Л., 1936.
31. *Хоботов Е. Н.* О модификации экстраградиентного метода для решения вариационных неравенств и некоторых задач оптимизации // ЖВМ и МФ. 1987. Т. 27, № 10. С. 1462–1473.
32. *Agdeppa R. P., Yamashita N., Fukushima M.* The traffic equilibrium problem with nonadditive costs and its monotone mixed complementarity problem formulation // Transportation Research Part B. 2007. № 41. P. 862–874.
33. *Arrowsmith G. A.* A behavioural approach to obtaining a doubly constrained trip distribution model // Operational Research Quarterly. 1973. V. 24, № 1. P. 101–111.
34. *Bar-Gera H.* Origin-based algorithm for the traffic assignment problem // Transportation Science. 2002. V. 36, № 4. P. 398–417.



35. *Beckmann M., McGuire C. B., Winsten C. B.* Studies in the economics of transportation. RM-1488. Santa Monica: RAND Corporation, 1955.
36. *Bertsekas D., Gafni E.* Projection methods for variational inequalities with application to the traffic assignment problem // Mathematical Programming Study. 1982. № 17. P. 139–159.
37. *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.
38. *Chen M., Bernstein D. H., Chien S. I. J., Mouskos K.* Simplified formulation of toll design problem // Transportation Reseach Record. 1999. № 1667. P. 88–95.
39. *Dafermos S.* Traffic equilibrium and variational inequalities // Transportation Science. 1980. V. 14, № 1. P. 42–54.
40. *Fang S.-C., Rajasekera J. R., Tsao H.-S. J.* Entropy optimization and mathematical programming. Kluwer Academic Publisher, 1997.
41. *Facchinei F., Pang J.-S.* Finite-Dimensional Variational Inequalities and Complementarity Problems (V. I, II). Springer, 2003.
42. *Frank M., Wolfe P.* An algorithm for quadratic programming // Naval Research Logistics Quarterly. 1956. V. 3. P. 95–110.
43. *Gabriel S. A., Bernstein D.* The traffic equilibrium problem with nonadditive path costs // Transportation Science. 1997. V. 31, № 4. P. 337–348.
44. *Iusem A. N.* An iterative algorithm for the variational inequality problem // Comput. and Appl. Mathematics. 1994. V. 13, № 2. P. 103–114.
45. *Janson B., Zozaya-Gorostiza C.* The problem of cyclic flows in traffic assignment // Transportation Research Part B. 1987. V. 21, № 4. P. 299–310.
46. *Leventhal T., Nemhauser G. L., Trotter L. Jr.* A column generation algorithm for optimal traffic assignment // Transportation Science. 1973. № 7. P. 168–176.
47. *Michelot C.* A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbf{R}^n$ // J. of Optimization Theory and Appl. 1986. V. 50, № 1. P. 195–200.
48. *Knight F. H.* Some fallacies in the interpretation of social cost // The Quarterly Journal of Economics. 1924. V. 38, № 4. P. 582–606.
49. *Konnov I. V.* Combined relaxation methods for variational inequalities. Berlin: Springer, 2001.
50. *Konnov I. V.* Equilibrium Models and Variational Inequalities. Elsevier Science, 2007.
51. *Kravchuk A. S., Neittaanmaki P. J.* Variational and Quasi-Variational Inequalities in Mechanics. Springer, 2007.
52. *Lo H. K., Chen A.* Traffic equlibrium problem with rout-specific costs: formulation and algorithms // Transportation Research Part B. 2000. V. 34, № 6. P. 493–513.
53. *Marcotte P.* Application of Khobotov's algorithm to variational inequalities and network equilibrium problems // INFOR. 1992. V. 29, № 4. P. 258–270.
54. *Nagurney A.* Network Economics: A Variational Inequality Approach. Dordrecht: Kluwer Academic Publishers, 1999.
55. *Nagurney A., Dong J.* Paradoxes in networks with zero emission links: implications for telecommunications versus transportation // Transportation Research Part D. 2001. V. 6, № 4. P. 283–296.

56. *Nemirovsky A.*, *Yudin D.* Informational complexity and efficient methods for solution of convex extremal problems. N.Y.: Wiley, 1983.
57. *Nesterov Yu.*, *de Palma A.* Stationary dynamic solutions in congested transportation networks: summary and perspectives // Networks and Spatial Economics. 2003. № 3. P. 371–395.
58. *Nurminski E. A.* Envelope stepsize control for iterative algorithms based on Fejer processes with attractants // Optimization Methods and Software. 2010. V. 25, № 1. P. 97–108.
59. *Patriksson M.* The traffic assignment problem — models and methods. Utrecht, Netherlands: VSP, 1994.
60. *Pigou A. C.* The economics of welfare. London: MacMillan, 1932. 4-th edition. (Русский перевод: *Пигу А. С.* Экономическая теория благосостояния. Т. 1–2. Сер. Экономическая мысль Запада. М.: Прогресс, 1985).
61. *Roughgarden T.*, *Tardos E.* How bad is selfish routing? // Journal of the ACM. 2002. V. 49, № 2. P. 236–259.
62. *Sun D.* A projection and contraction method for the nonlinear complementarity problem and its extensions // Mathematica Numerica Sinica. 1994. V. 16. P. 183–194.
63. *Wilson A. G.* A statistical theory of spatial distribution models // Transportation Research. 1967. V. 1. P. 253–270.
64. *Wardrop J.* Some Theoretical Aspects of Road Traffic Research // Proceedings of the Institute of Civil Engineers. 1952.
65. *Wang Y. J.*, *Xiu N. H.*, *Wang C. Y.* Unified framework of extragradient-type methods for pseudomonotone variational inequalities // J. of Optimization Theory and Appl. 2001. V. 111, № 3. P. 641–656.
66. *Xiu N.*, *Zhang J.* Some recent advances in projection-type methods for variational inequalities // J. of Comput. and Appl. Mathematics. 2003. V. 152. P. 559–585.

# Глава 2

# Математические модели транспортных потоков

## 2.1. Макроскопические модели

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

### 2.1.1. Модель Лайтхилла—Уизема—Ричардса (LWR)

Во второй половине 40-х годов и в 50-е годы XX века в СССР и США интенсивно занимались исследованием процессов, возникающих при взрыве бомбы (см., например, монографии [1,2]). В частности, большое внимание было уделено изучению начально-краевых задач для уравнения типа закона сохранения и систем таких уравнений. В это же время наблюдался и рост приложений, в которых встречаются схожие уравнения [3, 4]. Так в 1955 г. независимо в работах [5, 6] (см. также [7]) была предложена, по-видимому, первая макроскопическая (гидродинамическая) модель однополосного[1)] транспортного потока, названная впоследствии *моделью Лайтхилла—Уизема* (*Уитема*)*—Ричардса* (*LWR*), в которой поток АТС (автотранспортных средств) рассматривается как поток одномерной сжимаемой жидкости. Часто эту модель называют *моделью Лайтхилла—Уизема*. Отметим, что вместо термина «автомобиль» и тем более «машина» в транспортной литературе принято использовать термин АТС.

В модели LWR предполагается, что

а) существует взаимно однозначная зависимость между скоростью $v(t, x)$ и погонной плотностью $\rho(t, x)$ потока — *уравнение состояния*;

б) выполняется *закон сохранения массы* количества АТС.

[1)] Полоса бесконечная в обе стороны, движение происходит слева направо (для определенности), нет источников и стоков автотранспортных средств.

Запись $\rho(t, x)$ обозначает число АТС на единицу длины в момент времени $t$ в окрестности точки трассы с координатой $x$. Аналогично, $v(t, x)$ — скорость АТС в момент времени $t$ в окрестности точки трассы с координатой $x$. Везде в дальнейшем предполагается, что пространственные масштабы, на которых транспортный поток описывается макроскопическими (гидродинамическими) моделями, значительно превышают характерный размер АТС, т. е. составляют не менее сотни метров. В таком предположении мы будем интерпретировать $\rho(t, x)$, $v(t, x)$ не как некоторые, должным образом усредненные, величины (см., например, [5,8]), а как функции, получающиеся при переходе от *микроскопического описания* (в том числе и описания с помощью *клеточных автоматов*, см. п. 2.2) к макроскопическому. Иначе говоря, мы считаем, что транспортный поток подчиняется некоторой микроскопической модели, в которой детально описывается поведение АТС в зависимости от обстановки впереди, и эта модель является разностным или дифференциально-разностным аналогом рассматриваемой нами макроскопической модели. Таким образом, корректность предложенного здесь подхода к определению $\rho(t, x)$, $v(t, x)$ основывается на устойчивой аппроксимации макроскопической модели микроскопической. При этом необходимость рассмотрения макроскопических моделей обусловлена в первую очередь более простой техникой их исследования и большей наглядностью ввиду гидродинамических параллелей.

Первое предположение выразим условием:

$$v(t, x) = V(\rho(t, x)). \tag{1}$$

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

$$V'(\rho) < 0. \tag{2}$$

Обозначим через

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

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

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

Это условие можно понимать следующим образом: движение по двум одинаковым и независимым полосам с разными плотностями менее «эффективно», чем движение по этим полосам с одинаковой плотностью, равной среднему арифметическому первоначальных плотностей. Однако если агрегировать несколько полос в одну — иначе говоря, заменить несколько полос одной агрегированной, к которой и применять рассматриваемую модель, — то, как показывают наблюдения за реальными транспортными потоками, от вогнутости функции $Q(\rho)$, вообще говоря, придется отказаться.



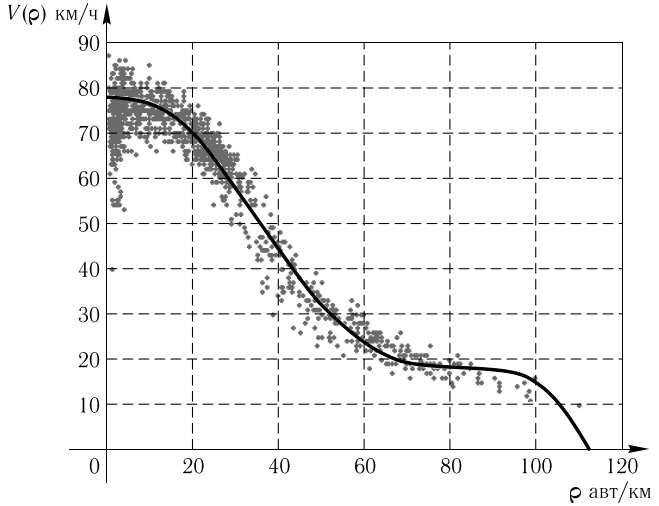



Рис. 1. Уравнение состояния транспортного потока

Так, на рис. 1, 2 отображены экспериментальные данные «Центра исследования транспортной инфраструктуры» г. Москвы, собранные (в течение одного дня в 2005 г.) по четырем полосам на участке третьего транспортного кольца от Автозаводской улицы до Варшавского шоссе и сагрегированные на одну полосу. Заметим, что в действительности измерялась зависимость $V(Q)$.

Объяснить небольшой провал интенсивности потока $Q(\rho)$ при плотностях $\rho \sim 60$–$115$ АТС/км можно, по-видимому, тем, что при этих плотностях существенное влияние на интенсивность потока оказывают перемещения АТС с одной полосы на другую. Перестраивания АТС из одной полосы в другую при этих плотностях снижают интенсивность потока. С одной стороны, за счет перемещения из полосы в полосу можно двигаться быстрее — так оно и происходит при плотностях $\rho \sim 30$–$50$ АТС/км. С другой стороны, в среднем такие перемещения при $\rho \sim 50$–$120$ АТС/км приводят к дополнительным затратам на само перестраивание и к замедлению тех АТС, перед которыми встраивается новое АТС [9]. Другое объяснение этого наблюдения, данное в главе 3, связано с тем, что при $\rho \sim 50$–$120$ АТС/км само понятие «фундаментальная диаграмма» не

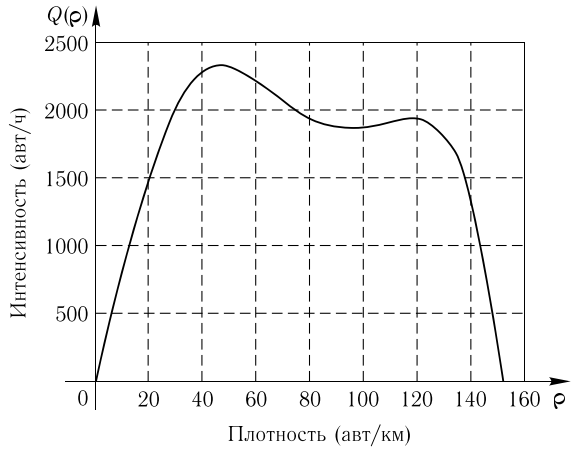



Рис. 2. Фундаментальная диаграмма

совсем корректно [10]. Иначе говоря, при этих плотностях нет четкой зависимости величины потока (скорости) от плотности. Одному значению плотности соответствует целый промежуток возможных значений потока (скорости).

Второе предположение выразим законом сохранения

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

Отсюда следует, что для любого прямоугольного контура $\Gamma$ в полуплоскости $t \geqslant 0$, $x \in \mathbb{R}$, со сторонами, параллельными осям, выполняется:

$$\oint\limits_\Gamma \rho(t,\, x)\,dx - Q(\rho(t,\, x))\,dt = 0. \tag{3}$$

Легко показать, что это соотношение справедливо для произвольного кусочно-гладкого контура $\Gamma$. В точках гладкости $\rho(t,\, x)$:

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

т. е.

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

Поставим начальное *условие типа Римана*

$$\rho(0,\, x) = \begin{cases} \rho_-, & x < x_-, \\ \rho_0(x), & x_- \leqslant x < x_+, \\ \rho_+, & x \geqslant x_+. \end{cases} \tag{5}$$



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

$$\rho_0'(x) \geqslant 0, \quad \rho_+ = \rho_{\max},$$

где $\rho_{\max}$ — максимально возможная плотность: ситуация «бампер к бамперу». Требуется определить, как по транспортному потоку будет распространяться информация о заторе впереди. Решение этой задачи позволит ответить, например, на следующий вопрос: если движение АТС с утра на Дмитровском шоссе в сторону Москвы «встало» в районе г. Долгопрудный, то через какое время затор дойдет до г. Дмитров? Ряд интересных модельных задач (задача о светофоре, об эволюции локального затора и др.) для закона сохранения (4) рассмотрен в главах 2 и 3 книги [7]; см. также раздел 2.3.

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

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

Предположим, что на плоскости $(t;\, x)$ этому разрыву соответствует кривая $L$. Возьмем в окрестности точки $(t,\, x) \in L$ прямоугольный контур (для определенности зададим ориентацию по часовой стрелке) так, как показано на рис. 3. Будем считать, что ширина контура $\Gamma$ настолько мала по сравнению с длиной, что интегралом по участкам контура $\Gamma$, поперечным к $L$, можно пренебречь (см. рис. 3).

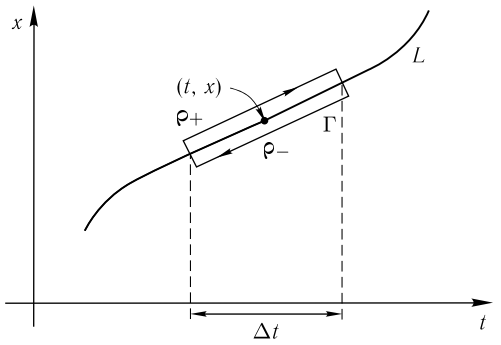



Рис. 3. RRH-условие на разрыве

Тогда из (3) следует, что

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

где $c = dx/dt$ соответствует наклону касательной к $L$ в точке $(t, x)$, $\Delta t$ — длина проекции контура на ось $t$. При $\Delta t \to 0$ это равенство переходит в следующее условие для *скорости движения разрыва* $c$, которое мы будем называть, следуя П. Лаксу [11], *условием Римана—Ранкина—Гюгонио*:

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

Похожее условие встречалось в работе Стокса (1848).

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

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

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

и *начальное условие Римана*:

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

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

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

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

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



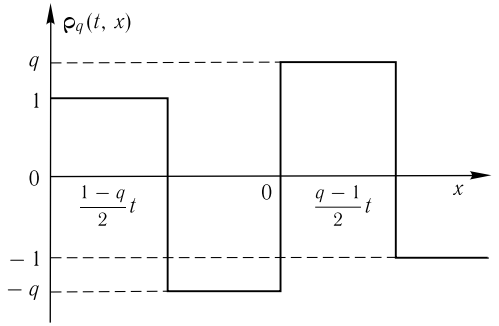



Рис. 4. Пример О. А. Олейник (1957)

Единственное решение выделяется условием отбора, по-видимому, впервые предложенным О. А. Олейник в 1958 г. [15,16] и И. М. Гельфандом в 1959 г. [17] [1)] как *условие устойчивости* (*допустимости*) *разрыва*. Частный случай выпуклой (вогнутой) функции $Q(\rho)$ см. в [13], а также в [11].

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

$$\forall \rho \in (\rho_-, \rho_+) \quad \sigma(\rho_-, \rho_+) \leqslant \sigma(\rho_-, \rho), \quad \text{если} \quad \rho_- < \rho_+;$$
$$\forall \rho \in (\rho_+, \rho_-) \quad \sigma(\rho_-, \rho_+) \geqslant \sigma(\rho_-, \rho), \quad \text{если} \quad \rho_- > \rho_+.$$

Это условие также называют *энтропийным условием*, *энтропийным условием Олейник*, *E-условием Олейник*. Объяснение (основанное на методе исчезающей вязкости, см. п. 2.1.3) того, откуда возникли E-условие и RRH-условие, будет приведено в п. 2.3.1. Заметим также, что в классе кусочно-постоянных начальных условий, аппроксимирующих класс ограниченных измеримых начальных условий, добавление E-условия как условия отбора возможных разрывов к соотношению (3) однозначно и конструктивно определяет динамику $\rho(t, x)$; нужно также оговориться, что кусочно-гладкая функция $Q(\rho)$ не имеет точек сгущения нулей второй производной. Отмеченная конструктивность активно использовалась в 70-х и 80-х годах XX века, например, при исследовании модельных задач раздела 2.3.

E-условие имеет наглядную геометрическую интерпретацию (рис. 5; для определенности считаем $\rho_- < \rho_+$): график функции $Q(\rho)$ при $\rho \in (\rho_-, \rho_+)$

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

лежит не ниже прямой, проходящей через точки $(\rho_-,\ Q(\rho_-))$ и $(\rho_+,\ Q(\rho_+))$. При этом скорость движения разрыва $c$ равна наклону этой прямой.

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

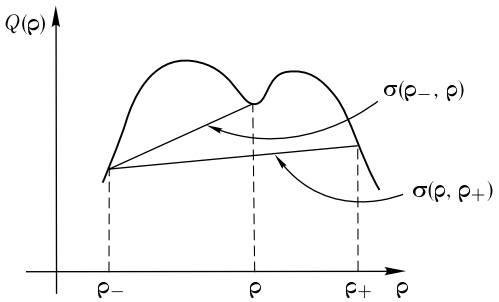



Рис. 5. E-условие

**Пример** [18]. Снова возьмем уравнение Хопфа и начальное условие

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

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

$$\rho_1(t,\ x)=\begin{cases}-1, & x\leqslant -t,\\ \dfrac{x}{t}, & -t<x\leqslant t,\\ 1, & x>t;\end{cases}\qquad \rho_2(t,\ x)=\begin{cases}-1, & x\leqslant 0,\\ 1, & x>0.\end{cases}$$

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

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

То есть неклассическое решение $\rho_2(t,\ x)$ не является устойчивым решением. Несложно проверить, что на разрыве решения $\rho_2(t,\ x)$ не выполняется E-условие, поэтому $\rho_2(t,\ x)$ не является решением. Если посмотреть на поведение характеристик системы, то можно заметить, что для решения



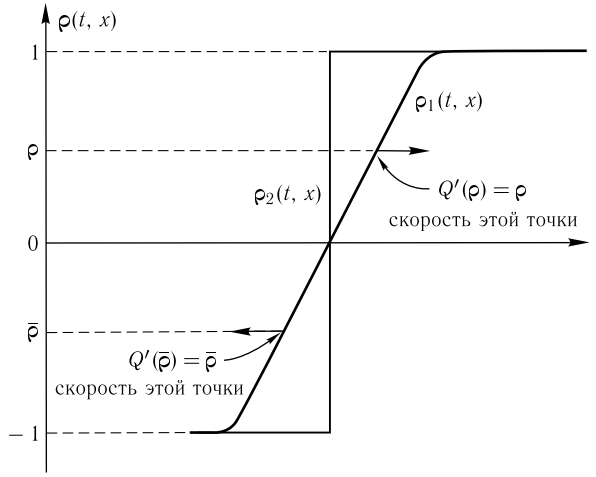



Рис. 6

$\rho_2(t,\ x)$ разрыв надуман — он не вызван пересечением характеристик. Вместо того чтобы пересекаться на разрыве, характеристики «расходятся» от разрыва.

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

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

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

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

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

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

Заметим также, что процесс, описываемый разрывным решением (4), необратим во времени (см., например, [18, 19] и п. 2.3.2). Причем условие

разрывности процесса существенно для необратимости. Так, в примере О. А. Олейник $\rho_q(t,\,x)$ при $q=1$ является разрывным решением (4), (5), для которого выполняется E-условие (как строгое неравенство). Если решать (4), (5) в попятном времени, то неравенство в E-условии поменяется на противоположное, поэтому если функция $\rho_q(t,\,x)$ при $q=1$ удовлетворяла «прямому» E-условию, то она точно не может удовлетворять «попятному» E-условию. Заметим, однако, что уравнение (4) выглядит симметричным относительно обращения времени ($t\to -t$), поскольку при попятном течении времени величина потока $Q(\rho)$ изменяет знак на противоположный. Однако, как уже отмечалось, уравнение (4), понимаемое в слабом смысле, определяет эволюцию системы, вообще говоря, не единственным образом. Выделение единственного решения является необратимой по времени процедурой.

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

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

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

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

$$C_1=\frac{\rho_+^2/2-\rho_-^2/2}{\rho_+-\rho_-}=\frac{1}{2}(\rho_++\rho_-)\overset{?}{=}\frac{2}{3}\frac{\rho_+^2+\rho_+\rho_-+\rho_-^2}{\rho_++\rho_-}=\frac{\rho_+^3/3-\rho_-^3/3}{\rho_+^2/2-\rho_-^2/2}=C_2,$$

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

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

на разрыве решения уравнения

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

с $\eta'(\rho)>0$. □

**Замечание** [20]. Пусть

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



Тогда для любого $k$, для которого

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

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

### 2.1.2. Модель Танака

Приведем один из способов определения зависимости $V(\rho)$, изложенный в 1963 г. Танака и др. [8, 21], но известный и раньше.

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

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

где

$$d(v)=L+c_1v+c_2v^2$$

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

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

$$d(v)[\text{м}]=5{,}7[\text{м}]+0{,}504[\text{с}]\cdot v[\text{м/с}]+0{,}0285[\text{с}^2/\text{м}]\cdot v^2[\text{м}^2/\text{с}^2],$$

для мокрого асфальта [22]

$$d(v)[\text{м}]=5{,}7[\text{м}]+0{,}504[\text{с}]\cdot v[\text{м/с}]+0{,}0570[\text{с}^2/\text{м}]\cdot v^2[\text{м}^2/\text{с}^2],$$

а для обледенелой дороги [22]

$$d(v)[\text{м}]=5{,}7[\text{м}]+0{,}504[\text{с}]\cdot v[\text{м/с}]+0{,}1650[\text{с}^2/\text{м}]\cdot v^2[\text{м}^2/\text{с}^2].$$

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

[1)] Величину $d(v)$ также называют *динамическим габаритом* или *дистанцией видимости*; под $d(v)$ понимают длину части полосы, содержащей АТС вместе с дистанцией экстренного торможения.

### 2.1.3. Модель Уизема

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

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

Откуда, с учетом закона сохранения количества АТС

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

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

$$\frac{\partial\rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \frac{\partial}{\partial x}\Big(D(\rho)\frac{\partial\rho}{\partial x}\Big). \tag{6}$$

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

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

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

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

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

С помощью замены (Форсайта—)Флорина—Хопфа—Коула [4, 7, 19]

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



задача (4), (5) для уравнения Бюргерса сводится к задаче Коши для уравнения теплопроводности: [1)]

$$\frac{\partial w}{\partial t} = \varepsilon\frac{\partial^2 w}{\partial\tilde x^2}, \quad w(0,\,\tilde x) = \exp\Big(-\frac{1}{2\varepsilon}\int\limits_0^{\tilde x}\tilde\rho(0,\,\xi)\,d\xi\Big);$$

см. также замечание 2 в конце этого пункта. Используя этот факт, Э. Хопф в 1950 г. изучал поведение решения начальной задачи Коши для уравнения Бюргерса [7, 14]. Так, например, им был обоснован предельный переход, получивший название *метода исчезающей вязкости* [2)], при $\varepsilon \to 0+$ от уравнения Бюргерса к уравнению Хопфа:

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

В связи с вышесказанным напомним, что для нелинейного закона сохранения (4) гладкое решение задачи Коши (4), (5) существует, как правило, только в малой окрестности линии, где заданы начальные условия. По разрывным начальным условиям решение задачи Коши для нелинейных уравнений, вообще говоря, не определяется однозначно даже в сколь угодно малой окрестности линии, где заданы начальные условия. Для того чтобы задача Коши для нелинейных уравнений с гладкими или разрывными начальными условиями была однозначно разрешима в бóльшей области, необходимо рассматривать разрывные решения уравнения и по-новому ставить задачу Коши. Казалось бы, достаточно, следуя идеям Н. М. Гюнтера, С. Л. Соболева, Л. Шварца в линейном случае, равенства (4), (5) понимать в слабом смысле, т. е. понимать (4) в смысле соотношения (3). Однако, как показывает пример О. А. Олейник из п. 2.1.1, такое определение решения не обеспечивает его единственности. *Корректный способ заключается в том, чтобы понимать решение* $\rho(t,\,x)$ *задачи Коши* (4), (5) *как предел почти всюду по* $x$ *при любом фиксированном значении* $t > 0$ *решений* $\rho_\varepsilon(t,\,x)$ *задач Коши* (6), (5) *при* [3)]

$$\varepsilon \to 0+, \quad D(\rho) := \varepsilon D(\rho), \quad D(\rho) > 0.$$

[1)] Причина, по которой уравнение Бюргерса линеаризуется, объясняется в [23]. Это связано с тем, что уравнение Бюргерса достаточно симметрично, т. е. допускает бесконечномерную алгебру Ли (группу преобразований). Интересно заметить, что есть определенная техника, позволяющая по заданному эволюционному уравнению определять, линеаризуется ли оно или нет. Для более подробного ознакомления с групповым анализом дифференциальных уравнений можно рекомендовать монографии [24] и [25]. Заметим также, что при определенных специально подобранных начальных условиях могут быть получены точные формулы для решений ряда важных в приложениях существенно нелинейных уравнений параболического типа [26–28].

[2)] Хотя в рассматриваемом нами случае речь идет скорее о диффузии, чем о вязкости.

[3)] Подчеркнем, что независимость предела от вида диффузионного слагаемого обусловлена тем, что диффузия входит дивергентным образом в правую часть уравнения (6). Если бы,

Более того, имеется оценка Н. Н. Кузнецова (1975):

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

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

Это представляется естественным. Ведь оба уравнения (4) и (6) возникли при описании одного явления на разных уровнях детализации. Обоснованием метода исчезающей вязкости интенсивно занимались в 50-е годы XX века Э. Хопф, О. А. Олейник, А. Н. Тихонов и А. А. Самарский, П. Лакс, О. А. Ладыженская, И. М. Гельфанд и др. Наиболее общие результаты получил С. Н. Кружков в конце 1960-х гг.; подробности см. в [11, 29–38], а также в конце п. 2.3.2.

Заметим, что уравнение (6) параболического типа, следовательно, решение можно понимать в обычном (классическом) смысле даже при разрывных начальных условиях. Для таких нелинейных уравнений развит достаточно эффективный аппарат. Прежде всего это различные варианты принципа максимума и основанные на них методы априорных оценок старших производных [39, 40], позволяющие довольно тонко исследовать различные свойства решений начально-краевых задач для уравнений параболического типа. Важные работы в этой области принадлежат С. Н. Бернштейну, И. Г. Петровскому, О. А. Олейник, О. А. Ладыженской. Напомним, в чем заключается принцип максимума: решение уравнения параболического типа достигает максимального и минимального значений на параболической границе (основание и боковые стороны) области, в которой рассматривается это решение (вместо уравнений можно рассматривать и неравенства). Так, например, решение начальной задачи Коши для уравнения параболического типа ограничено теми же постоянными, что и начальное условие. Установив различные свойства решения задачи Коши (6), (5) и осуществив предельный переход при $\varepsilon\to 0+$, можно получать разнообразные свойства решения задачи Коши (4), (5).

Посредством «хорошего» уравнения (6) попытаемся теперь установить связь между законом сохранения и *уравнением Гамильтона—Якоби*.

скажем, уравнение (6) имело вид

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

то предел, вообще говоря, уже зависел бы от вида $D(\rho)$. В этом легко убедиться, рассмотрев пример И. М. Гельфанда, см. п. 2.1.1. Отметим тем не менее, что если в пределе при $D(\rho)\equiv 1$ получается функция без разрывов, то предел по-прежнему не будет зависеть от $D(\rho)>0$.



Для этого рассмотрим две приводимые ниже задачи Коши:

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

$$\begin{cases}\dfrac{\partial U_\varepsilon}{\partial t}+Q\Big(\dfrac{\partial U_\varepsilon}{\partial x}\Big)=\varepsilon\dfrac{\partial^2 U_\varepsilon}{\partial x^2},\\ U_\varepsilon(0,x)=\displaystyle\int_0^x\rho(0,y)\,dy.\end{cases}\tag{8}$$

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

$$\frac{\partial\rho_\varepsilon}{\partial t}+\frac{\partial Q(\rho_\varepsilon)}{\partial x}=\varepsilon\frac{\partial}{\partial x}\Big(D(\rho_\varepsilon)\frac{\partial\rho_\varepsilon}{\partial x}\Big),\quad D(\rho_\varepsilon)>0,$$
$$\rho_\varepsilon(0,x)=\rho(0,x),$$

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

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

Используя схожую технику, можно показать, что

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

равномерно на любом компакте в полуплоскости $t>0$, $x\in\mathbb{R}$. При этом $U(t,x)$ — ограниченная непрерывная функция, которая слабо удовлетворяет *уравнению Гамильтона—Якоби*:

$$\frac{\partial U}{\partial t}+Q\Big(\frac{\partial U}{\partial x}\Big)=0\tag{9}$$

и начальному условию

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

Так определенную функцию $U(t,x)$ называют *вязкостным решением* (Крэндалла—Лионса) задачи Коши (9), (10), см. замечание 1. Поскольку решения задач (7) и (8) классические, то имеет место следующая формула:

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

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

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

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

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

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

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

Используя теорему о дифференцировании (по направлению) под знаком супремума [1)], можно проверить, что для вязкостного решения $U(t,\, x)$ справедливы следующие формулы, встречавшиеся в работах Э. Хопфа (1965), которые принято называть *формулами Хопфа(—Лакса)*: [2)]

1) пусть $U_0(x)$ — выпуклая функция, тогда

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

где

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

2) пусть $Q(\rho)$ — вогнутая функция, тогда

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

где

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

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

Проверим, например, формулу из 1). Для этого сразу заметим ввиду выпуклости $U_0(x)$, что по теореме Фенхеля—Моро [47]

$$U(0,\, x) = U_0^{**}(x) = U_0(x).$$

По теореме о дифференцировании по направлению под знаком супремума

$$\frac{\partial U(t,\, x)}{\partial t} = -Q(s(t,\, x)), \qquad \frac{\partial U(t,\, x)}{\partial x} (= \rho(t,\, x)) = s(t,\, x),$$

[1)] См. [45, 46]. Необходимые для понимания факты выпуклого анализа имеются, например, в книгах [47, 48]. Теорему о дифференцировании по направлению под знаком супремума иногда называют *теоремой Демьянова—Данскина*, поскольку оба автора независимо пришли к аналогичному утверждению в конце 60-х годов XX века [45]. Однако эта теорема была известна и раньше. Так, в середине 60-х годов XX века А. Я. Дубовицкий и А. А. Милютин при получении принципа максимума для задач с фазовыми ограничениями доказали и фактически уже использовали теорему о дифференцировании по направлению под знаком супремума [41, 49].

[2)] Аналогичные формулы возникали ранее, например, в работах О. А. Олейник и П. Лакса; приблизительно в то же время, что и в работах Э. Хопфа [46], эти формулы использовал С. Н. Кружков; близкие идеи «о сведении решения задачи Коши для обыкновенного нелинейного дифференциального уравнения или для нелинейного уравнения в частных производных к решению задач оптимизации» предлагались в 1965 г. и ранее Р. Беллманом и Р. Калабой [50]. Упомянем здесь также работу Б. Н. Пшеничного и М. И. Сагайдака [51].



где

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

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

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

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

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

терпит разрыв. Собственно, формула

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

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

Используя далее теорему о дифференцировании под знаком супремума и соотношение (11), можно также получить и формулы для энтропийного решения $\rho(t,\, x)$ задачи Коши (4), (5); заметим, что формула, полученная из пункта 2), называется *формулой Лакса—Олейник* [11, 44]. Упомянутые здесь формулы использовались, например, при исследованиях задач раздела 2.3. В частности, С. Н. Кружков и Н. С. Петросян в 1982 г. получили доказательство утверждения, очень близкого к теореме 1 п. 2.3.1, исходя из формулы 1).

**Замечание 1.** Рассмотрим следующую каноническую задачу (Лагранжа) оптимального управления понтрягинского типа [41, 42]:

$$J(t,\, x;\, u(\cdot)) = \int_{t_0}^{t} L(\tau,\, x(\tau),\, u(\tau))\, d\tau + \varphi(t_0,\, x_0),$$

$$\dot{x}(\tau) = f(\tau,\, x(\tau),\, u(\tau)), \quad u(\tau) \in M, \quad x(t) = x.$$

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

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

Тогда с некоторыми оговорками справедлива следующая теорема [43, 44].

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

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

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

**Замечание 2** (идемпотентный принцип соответствия Литвинова—Маслова [52–54]). Рассмотрим уравнение теплопроводности

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

и уравнение Гамильтона—Якоби—Хопфа:

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

Несложно проверить, что замена

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

переводит одно уравнение в другое; схожую замену также делал Э. Шрёдингер. Таким образом, кстати сказать, и была получена замена Флорина—Хопфа—Коула русским механиком В. А. Флорины в 1948 г., на два года раньше Э. Хопфа и на три года раньше С. Коула. Используя при $\varepsilon\to 0+$ замену $w_\varepsilon \rightsquigarrow -2\varepsilon\ln w_\varepsilon = U_\varepsilon$, т. е.

$$\lim_{\varepsilon\to 0+} w_\varepsilon \stackrel{\text{def}}{=} w \rightsquigarrow \lim_{\varepsilon\to 0+}(-2\varepsilon\ln w_\varepsilon) \stackrel{\text{def}}{=} U,$$

В. П. Маслов и В. П. Белавкин [52] в конце 80-х годов XX века предложили ввести *идемпотентное* (или *тропическое*) полуполе [1)] с операциями сложения и умножения $\oplus$, $\odot$, которые определяются следующим образом:

$$w^1 \rightsquigarrow U^1, \quad w^2 \rightsquigarrow U^2;$$

$$w^1 + w^2 \rightsquigarrow \lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^1/(2\varepsilon)} + e^{-U^2/(2\varepsilon)}\right) =$$

$$= \min\left\{\lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^1/(2\varepsilon)}\right), \lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^2/(2\varepsilon)}\right)\right\} =$$

$$= \min\{U^1, U^2\} \stackrel{\text{def}}{=} U^1 \oplus U^2;$$

$$w^1 \cdot w^2 \rightsquigarrow \lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^1/(2\varepsilon)}\cdot e^{-U^2/(2\varepsilon)}\right) =$$

$$= \lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^1/(2\varepsilon)}\right) + \lim_{\varepsilon\to 0+}(-2\varepsilon)\ln\left(e^{-U^2/(2\varepsilon)}\right) =$$

$$= U^1 + U^2 \stackrel{\text{def}}{=} U^1 \odot U^2.$$

Заметим, что в 60-е годы XX века Н. Н. Воробьев рассматривал аналогичные объекты, именуя их экстремальными алгебрами, см. также работы С. Н. Н. Пандита. В современном контексте идемпотентные полуполя изучал Б. Карре [54]. Далее строился функциональный анализ над идемпотентным полуполем подобно тому, как строился обычный функциональный анализ над полем действительных или комплексных чисел. Построение такого анализа выявляет много связей, иногда называемых *принципами соответствия* (Литвинова—Маслова), между классическими понятиями. Скажем, идемпотентным аналогом преобразования Фурье будет преобразование Юнга—Фенхеля—Лежандра [47], а вариационные принципы механики — это идемпотентный вариант фейнмановского подхода к квантовой механике через интегралы по траекториям. Кстати сказать, все это тесно связано с формулами типа Лакса—Олейник. Однако здесь мы хотим обратить внимание прежде всего на то, что для некоторых нелинейных уравнений, например, Гамильтона—Якоби, Гамильтона—Якоби—Беллмана, справедлив принцип суперпозиции (В. П. Маслова), правда, не над обычными полями действительных или комплексных чисел, а над идемпотентным полуполем: если $U^1$, $U^2$ — решения, то для любых действительных чисел $\lambda^1$, $\lambda^2$

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

также является решением.

В заключение этого пункта, в котором мы пояснили, что понимается под решением задачи Коши (4), (5), приведем, следуя монографии [35], оценку устойчивости решения

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

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

[1)] От обычного поля полуполе отличается тем, что в нем отсутствует операция, обратная сложению (вычитание).

$$\|\rho(t,\cdot;\rho_1(0,x),Q_1(\rho))-\rho(t,\cdot;\rho_2(0,x),Q_2(\rho))\|_{L_1(\mathbb{R})}\leqslant$$
$$\leqslant\|\rho_1(0,\cdot)-\rho_2(0,\cdot)\|_{L_1(\mathbb{R})}+$$
$$+t\min\{\mathrm{T.V.}(\rho_1(0,\cdot)),\mathrm{T.V.}(\rho_2(0,\cdot))\}\|Q_1(\cdot)-Q_2(\cdot)\|_{\mathrm{Lip}},$$

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

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

Отметим, что приводимая выше оценка при

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

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

### 2.1.4. Модель Пейна и ее обобщения

Следующим важным шагом стала *модель Пейна* (1971) [7, 55]. Эту модель можно понимать как закон сохранения

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

в котором уже не предполагается зависимость скорости от плотности, т. е. уже не предполагается, что желаемая скорость устанавливается мгновенно. Выписывается уравнение [1)]

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

стремления реальной скорости $v$ к желаемой

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

причем $\tau$ ($\tau\sim 1$ с) характеризует скорость стремления. Заметим, что в электротехнической терминологии $\tau$ — время релаксации; если же уподоблять

[1)] В литературе принято называть моделью Пейна частный случай описанной модели:

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



транспортный поток сжимаемой неньютоновской (максвелловской) жидкости, то параметр $\tau$ характеризует максвелловское затухание [56]. Полученную систему уравнений запишем в виде

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

из которого легко следует строгая гиперболичность этой системы, т. е. матрица при $\dfrac{\partial}{\partial x}$ имеет различные вещественные собственные значения.

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

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

где «давление»

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

Отмеченное обстоятельство представляется естественным. Ведь «транспортная жидкость» — это жидкость с мотивацией (стремление двигаться с желаемой скоростью), которая и присутствует в правой части. Это замечание позволяет использовать в расчетах по гидродинамическим моделям транспортных потоков хорошо разработанные более чем за полвека вычислительные алгоритмы, например, схемы П. Лакса, С. К. Годунова, сеточно-характеристический метод (Магомедова—Холодова) и др., см. [4, 11, 18, 35, 56–58] и цитированную там литературу.

**Замечание.** Хорошим тестом на устойчивость выбранной разностной схемы является разложение конечных разностей по пространственной переменной в ряд Тейлора до второго порядка включительно и исследование матрицы при вторых производных на положительную определенность [56, 57] — метод *аппроксимационной вязкости*.

Напомним также вкратце (нам это понадобится в п. 2.2.4), следуя книге П. Лакса [11] (см. также [56,57]), в чем заключается метод численного решения начально-краевой задачи для закона сохранения, предложенный С. К. Годуновым в конце 50-х годов XX века. Начальное условие $\rho(0,x)$ аппроксимируется кусочно-постоянной функцией

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

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

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

Задача с начальными данными $\rho^\delta(0, x)$ может быть решена точно. В каждой точке $k\delta$ мы должны решить *задачу Римана о распаде разрыва*, см. п. 2.3.1. «Волны», выходящие из двух соседних точек разрыва $k\delta$ и $(k+1)\delta$, не пересекаются, пока $t \cdot c_{\max} \leqslant \delta/2$, где $c_{\max} = |\max Q'(\rho)|$ — максимальная скорость распространения возмущения. Итак, объединив решения задач Римана, можно получить точное решение. Выбор шага по времени вида $\tau = \delta/(2c_{\max})$ иногда называют *правилом Лакса*, причем автоматически выполняется необходимое *условие Куранта—Фридрихса—Леви* [2] $\tau/\delta \leqslant (c_{\max})^{-1}$ сходимости разностных схем при численном решении гиперболических уравнений. В момент времени $\tau = \delta/(2c_{\max})$, равный этому шагу, мы опять заменим это точное решение приближенной кусочно-постоянной функцией и повторим процесс. Численные эксперименты говорят о том, что метод Годунова дает хорошее приближение точных решений уравнений LWR и систем типа Пейна. Тем не менее строгое доказательство устойчивости схемы Годунова имеется, насколько нам известно, лишь для конкретных систем. Если схема Годунова сходится, то непременно к энтропийному решению [36], поскольку пространственная переменная всего одна, см. текст после замечания. Отметим сильную «качественную» связь (которую, впрочем, можно обосновать и теоретически) между описанной схемой Годунова, схемами бегущего счета, схемой потенциального сглаживания [18] и сеточно-характеристическим методом [59]. Отметим также в чем-то схожий метод «front tracking» [35, 36], в котором аппроксимируется не решение (кусочно-постоянной функцией), а вектор-функция потока (в скалярном случае $Q(\rho)$) кусочно-линейной функцией. С помощью этого метода недавно были получены продвижения в вопросах корректности начальной задачи Коши для системы законов сохранения [35, 36]. В заключение заметим, что схема Годунова для LWR-модели может быть содержательно проинтерпретирована, см. п. 2.2.4. Другими словами, можно было ничего не знать про LWR-модель и из естественных соображений «напрямую» прийти к разностной схеме С. К. Годунова; в транспортной литературе принято разностные схемы называть *моделями клеточных автоматов* (см. раздел 2.2). Как показывает практика, при гидродинамическом описании транспортного потока, по сути, «дискретного» объекта, очень важно выбирать разностную схему таким образом, чтобы она могла быть самостоятельно содержательно проинтерпретирована.

Заметим также [19, 31], что уже для системы двух законов сохранения — система Х. Пейна (12) как раз представляет пример такой системы, причем имеется еще и нелинейная правая часть, — в общем случае неизвестно, как корректно определять глобальное по времени обобщенное решение. Метод исчезающей вязкости для систем оказывается уже чувствительным к выбору положительно определенной матрицы $D(\rho)$ в правой части (проблема неединственности решения) [18]. Тем не менее для строго



гиперболической системы законов сохранения с одной пространственной переменной за последние 15 лет был достигнут определенный прогресс [11, 31–38]: в общем случае построена глобальная теория существования, единственности и устойчивости по начальным данным [1)]. Отметим, что теория была построена разными способами, в том числе и с помощью метода исчезающей вязкости:

$$\varepsilon \to 0+, \quad D(\rho) := \varepsilon I,$$

где $I = \operatorname{diag}\{1, \ldots, 1\}$ — единичная матрица. Так построенное обобщенное решение часто называют *энтропийным*.

Желание «размазать разрывы решений» привело от модели LWR к модели Уизема. Этим же мотивировано и введение в модель Пейна диффузионных поправок. Заметим, что, вводя в правую часть системы диффузионные (дисперсионные) поправки, мы, как правило, решаем вопрос о том, что мы понимаем под решением [56]. Иначе говоря, для корректной и адекватной «физике процесса» постановки начальной (начально-краевой) задачи Коши для системы законов сохранения важно знать, как эта система «была приготовлена», откуда и как она возникла: что огрубили, чем пренебрегли и т. д. На таком пути получаются новые модели: Р. Кюна (1993), Кернера—Конхойзера (1994) (см. раздел 2.4) и др. В [60] приведен достаточно подробный обзор работ (более 100 моделей). Здесь мы также упомянем российских ученых, в работах которых обобщается подход Х. Пейна: Н. Н. Смирнов, А. Б. Киселев и др. (МГУ) [61–63]; А. С. Холодов и др. (МФТИ) [64].

Несколько недостатков модели Пейна так же как и многих впоследствии предложенных моделей, в том числе с диффузионными поправками, были указаны К. Даганзо (Даганцо) (1995) [58, 60, 65]; см. также критику Б. С. Кернера в [10], отчасти приведенную в разделе 2.4 и главе 3. В частности, было показано, что при сильных пространственных неоднородностях начальных условий могут возникать отрицательные значения скоростей — затор «рассасывается назад» как результат действия вязкости. При определенных значениях параметров могут возникать плотности, превышающие максимально допустимые («бампер к бамперу»). Также, согласно этим моделям, на движение АТС заметное влияние оказывают

[1)] Особо отметим в этой связи работу Д. Глимма (1965) [11, 36], предложившего стохастическую модификацию метода Годунова ($\rho_k = \rho(0, k\delta + \xi\delta)$, где $\xi \in R[0, 1]$ — равномерно распределенная на отрезке $[0, 1]$ случайная величина), с помощью которой была установлена теорема существования для начальных данных, близких к константе (имеющих малую полную вариацию). Отметим также в этой связи, что в методе Годунова «законы сохранения» выполняются точно (и эта «консервативность» очень важна, как отмечал в 2004 г. во время доклада в МИАН РАН С. К. Годунов, иначе довольно быстро могут накапливаться ошибки), а в методе Глимма точно «в среднем». Выступление С. К. Годунова можно посмотреть здесь `http://www.mathnet.ru/php/presentation.phtml?option_lang=rus`.

АТС, находящиеся сзади, что в случае одной полосы вряд ли возможно в реальном транспортном потоке. В начале XXI века А. Эу и М. Раскль [66], Дж. М. Гринберг [67], Х. М. Чжан [68] показали, как можно устранить недостатки, отмеченные К. Даганзо. Основная идея заключается в изменении второго уравнения в системе Пейна:

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

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

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

В конце статьи [66] указано, что можно оставить релаксационное слагаемое в правой части второго уравнения системы типа Пейна:

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

При этом все положительные приобретения сохраняются, но добавляются и новые. Заметим [69], что модель типа Пейна—Уизема с

$$P(\rho) = \int_0^\rho \left(\tilde{\rho}\frac{\partial p(\tilde{\rho})}{\partial \tilde{\rho}}\right)^2 d\tilde{\rho}$$

переходит в модель Эу—Раскля и этот факт допускает обобщение: большинство гидродинамических моделей второго порядка могут быть приведены к модели типа Пейна—Уизема с определенной функцией $P(\rho)$.

За последние десять лет появилось большое количество статей (А. Эу, А. Клар, П. Гоатен, Р. Коломбо, М. Гаравелло, Б. Пикколи, Ф. Сибель и В. Маузер, Д. Хельбинг и др., см. сайт `http://arxiv.org/`), в которых модель Эу—Раскля обобщалась в различных направлениях. Например, в [70] — для того чтобы объяснить экспериментально обнаруженные Б. С. Кернером (1996) [10] три фазы транспортного потока: «газ, жидкость (свободное движение) — жидкость, замерзающая жидкость (синхронизированный режим движения) — замерзающая жидкость, лед (широко движущиеся кластеры)». Эти исследования, по-видимому, мотивированы желанием так обобщить гидродинамическую модель Эу—Раскля в классе систем из двух уравнений гиперболического типа, чтобы предложенная модель объясняла все основные наблюдаемые свойства транспортного потока.

Естественно теперь задаться вопросом: «переходят» ли модели типа Пейна при $\tau \to 0+$ (предел нулевой релаксации) в модель Уизема или модель LWR? Интуиция подсказывает, что должны переходить; так же, как и в п. 2.1.3, можно ссылаться на то, что описывается одно и то же явление на разных уровнях детализации. Однако в общем случае, насколько нам известно, нет строгого обоснования такого перехода. Ссылки на различные успешно исследованные случаи имеются, например, в работах [32, 66].



В заключение этого пункта упомянем модель третьего порядка Хельбинга—Эйлера—Навье—Стокса (1995) [60,71], в которой к системе уравнений Пейна добавляется третье уравнение — «закон сохранения энергии» для вариации (дисперсии) скорости $\theta$, характеризующей «разброс скоростей» относительно среднего значения. При этом во второе уравнение, которое понимается как уравнение для среднего значения скорости, вводится дополнительное слагаемое, зависящее от $\theta$. Заметим при этом, что для системы уравнений Навье—Стокса, описывающей движение вязкой ньютоновской жидкости, неизвестно в общем случае, как поставить начальную (начально-краевую) задачу Коши, чтобы глобальное решение, определенное при всех значениях времени, было единственным. За решение этой проблемы Математический институт Клэя в 2000 г. назначил премию в один миллион долларов. Естественно считать, что если уравнение описывает реальный процесс, то оно обязано решаться, притом единственным образом. Однако выбранные уравнения — это лишь некоторая модель описываемого явления, возможно, не всегда вполне адекватная, и «механизм», выбирающий единственное решение, мог быть «огрублен» при выводе уравнений, особенно если описываемый процесс чувствителен к малейшим возмущениям, флуктуациям. Интересные взгляды на проблему имеются в статьях О. А. Ладыженской [72] и В. И. Юдовича [73]. Так, в статье [72] проблема единственности переформулирована следующим образом: «Дают ли уравнения Навье—Стокса с начальными и краевыми условиями детерминистское описание динамики несжимаемой жидкости или не дают?» Сложности, возникающие при описании транспортного потока, во многом схожи со сложностями, возникающими при описании турбулентного движения жидкостей (см. статью А. Майда в сборнике [32], а также [74]).

### 2.1.5. Кинетические модели

Продолжая аналогию с газовой динамикой, будущий нобелевский лауреат по химии И. Пригожин (при участии Ф. Эндрюса и Р. Хермана) в 1960 г. предложил описывать транспортный поток кинетическим уравнением типа Больцмана с «интегралом взаимодействия АТС» вместо «интеграла столкновения частиц газа» [58,60,75]. Подход И. Пригожина был впоследствии развит в работах С. Павери-Фонтана (1975), Д. Хельбинга (1995) и др. [58,60].

Из кинетических моделей транспортного потока (в основном многополосного) можно получать макроскопические (гидродинамические) модели подобно тому, как в кинетической теории получаются уравнения газовой динамики (гидродинамики), т. е. с помощью умножения на различные функции от скорости и последующего интегрирования по скоростям кинетического уравнения для плотности в расширенном (на скорости) фазовом пространстве $(t; x; v)$. При этом, вообще говоря, будет получаться цепочка зацепляющихся уравнений. [1)] Так, если умножить кинетическое уравнение на единицу и проинтегрировать, получим уравнение для плотности («закон сохранения массы»), в которое будет входить средняя скорость. Если умножить кинетическое уравнение на скорость и проинтегрировать, получим уравнение для средней скорости («закон сохранения импульса»), в которое будет входить вариация скорости $\theta$, по сути, определяющаяся средним значением квадрата скорости. Если умножить кинетическое уравнение на квадрат скорости и проинтегрировать, получим уравнение для среднего значения квадрата скорости (откуда можно получить уравнение для вариации скорости), в которое будет входить среднее значение куба скорости, и т. д. Приходится в какой-то момент обрывать (замыкать) цепочку, привлекая, как правило, дополнительные «физические» соображения (гипотезы). Например, постулировать на основе экспериментов или другим способом для замыкания моментной цепочки некоторые соотношения, так называемые определяющие уравнения, между величинами, входящими в эти уравнения. Так, для газа в зависимости от этих соотношений получается модель идеального газа или модель Навье—Стокса—Фурье для вязкого теплопроводного газа [38].

В связи с вышесказанным уместно заметить, что классической задачей статистической физики, восходящей к работам Максвелла [38, 78], является исследование перехода от уравнения Больцмана к уравнениям газодинамики (гидродинамики). Центральным местом здесь является проблема замыкания моментной цепочки для решения уравнения Больцмана. Однако не менее важным является изучение перехода от стохастической марковской динамики (например, транспортных потоков), лежащей в основе движения (см. п. 2.2.4), к кинетической динамике. При этом стохастическая марковская динамика, заданная, как правило, линейной полугруппой, порождает за счет скейлинга или перехода к «динамике средних» нелинейные

[1)] Заметим, что современное понимание классической неравновесной статистической механики основывается на во многом схожей теории цепочек уравнений Боголюбова—Борна—Грина—Кирквуда—Ивона (ББГКИ) [76]. Впрочем, в последнее время появился новый интересный подход (см. работы В. В. Козлова с коллегами [77]), восходящий к работам А. Пуанкаре и Дж. Гиббса. Много интересных идей и разнообразных связей собрано в записях курса лекций В. В. Веденяпина, прочитанного несколько лет назад студентам МФТИ и МГОУ и посвященного кинетической теории [78].



кинетические уравнения (например, типа Больцмана—Пригожина), которые в свою очередь порождают нелинейные гидродинамические уравнения. Важно заметить, что без понимания этих «переходов» невозможно, на наш взгляд, правильно объяснить экспериментальные данные — те самые три фазы транспортного потока.

В заключение отметим, что имеются также модели, промежуточные между кинетическими и гидродинамическими, так называемые *мезоскопические*. Такой моделью двухполосного движения пользуется, например, коллектив, возглавляемый Б. Н. Четверушкиным [79, 80].

### 2.1.6. Практические приложения моделей

Несмотря на элементарность, модель LWR (а также ее дифференциально-разностные и разностные аналоги) достаточно популярна в прикладных расчетах. Во многом это связано с недостаточным объемом данных для использования моделей более высокого уровня (поправки, привносимые более тонкими моделями, нивелируются неточностью данных). Ряд современных коллективов исследователей сосредотачивается на решении начально-краевых задач для уравнения (4) на графе транспортной сети. Основные сложности при этом возникают при постановке краевых условий в узлах графа транспортной сети (см., например, [81, 82]). Модель LWR (точнее, ее разностные аналоги) хорошо подходит и для управления транспортными потоками. В подтверждение этих слов приведем некоторые идеи, использованные, например, в подходе Берклиевской группы (А. Б. Куржанский, А. А. Куржанский, П. Варайя, Р. Хоровитц и др. [83,84]) к управлению дорожным движением.

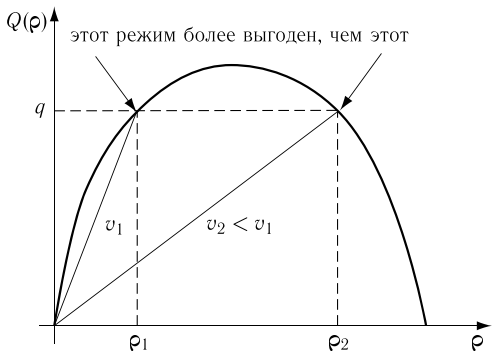



Рис. 7

Из фундаментальной диаграммы следует, что одному и тому же значению потока АТС соответствуют разные (как правило, две) плотности и,

как следствие, разные скорости.[1)] Очевидно, что более выгодным режимом является режим с большей скоростью (см. рис. 7): потоки «будут удовлетворены» в том же количестве, однако среднее время движения снизится, поскольку движение будет проходить при больших скоростях (и, как следствие, с меньшими плотностями).

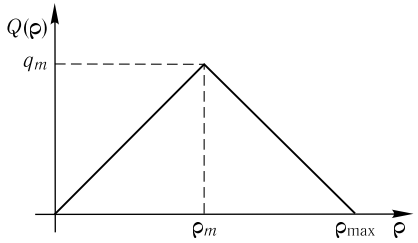



Рис. 8. Треугольная фундаментальная диаграмма

Задача управления (скажем, светофорами или въездами на основные магистрали) заключается в том, чтобы как можно большую часть времени среднестатистический водитель проводил именно в таких режимах. В статьях [85, 86] К. Даганзо была предложена модель клеточных автоматов CTM (Cell Transmission Model, см. п. 2.2.4), которую можно понимать как разностную схему Годунова для уравнения (4) с треугольной фундаментальной диаграммой[2)] (см. рис. 8). На основе этой модели в работах [83, 84] был предложен способ оптимального управления светофорами и въездами на магистрали, а также способ «оптимального разрыхления» однородного потока АТС на магистрали (с помощью светофоров) с целью уменьшения среднего времени в пути.

Несмотря на то, что с момента появления первых фундаментальных работ прошло более полувека, по мнению ряда ведущих специалистов в области математического моделирования дорожного движения (К. Нагель, Х. Махмасани, М. Шрекенберг и др.), проблема образования предзаторных и заторных ситуаций еще до конца не изучена. Используя терминологию, предложенную Б. С. Кернером [10], можно сказать, что на текущий момент нет общепринятого подхода, описывающего поведение АТС в области «синхронизированного потока» (см. также главу 3). Иначе говоря, если поток АТС уподобляется жидкости, то наиболее сложная для моделирования ситуация — это «замерзающая жидкость». Подтверждением вышесказанному может служить тот факт, что разные коллективы, занимающиеся моделированием транспортных потоков, как правило, используют разные модели: начиная от модели LWR (М. Гаравелло и Б. Пикколи [36, 81]; А. А. Куржанский и др. [83,84]) и заканчивая моделями, в которых каждый водитель описывается своим вариационным принципом (И. А. Лубашевский и др. [89,90]). Отметим также, что большое количество исследований сосредотачивается на изучении транспортного потока на отдельном прямолинейном участке транспортной сети с простейшими начально-краевыми условиями, в то время как причиной заторов, согласно К. Даганзо [65], часто являются «узкие места» — перекрестки, въезды и т. п. Поэтому особенно важно, в частности, для приложений, создать целостную модель транспортных потоков, адекватную имеющимся данным, включающую описание источников, стоков АТС и поведение АТС в вершинах графа транспортной сети (перекрестки, въезды, выезды и т. п.).

[1)] Это обстоятельство является также причиной сложностей, возникающих при постановке краевых условий в узлах (вершинах) графа транспортной сети [36, 81, 82]. Знание характеристик источников и стоков и *матриц перемешивания* в узлах (матриц, характеризующих расщепление потоков в узлах) недостаточно для корректной постановки начально-краевой задачи.

[2)] Также часто используется трапецеидальная фундаментальная диаграмма. Например, диаграмму на рис. 2 более естественно аппроксимировать именно трапецеидальной диаграммой, а не треугольной.



## 2.2. Микроскопические модели

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

### 2.2.1. Модель оптимальной скорости Ньюэлла

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

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

В микроскопической *модели Ньюэлла*, которая была предложена в 1961 г. и которая является одной из первых нелинейных *моделей оптимальной скорости* [7, 91], постулируется, что для каждого водителя существует «безопасная» скорость движения, зависящая от дистанции до

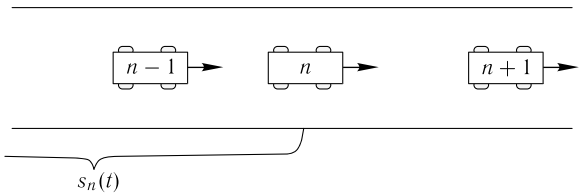



Рис. 9. Микроскопическая модель

лидера:

$$h_n(t+\tau) = V\Big(\frac{1}{h_n(t)}\Big),$$

где $\tau$ — время, характеризующее реакцию водителей, и

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

Заметим, что по зависимости интенсивности потока

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

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

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

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

$$V(\rho) = \frac{1/\rho - L}{\tau}.$$

Откуда имеем в левой окрестности точки $\rho_{\max}$

$$Q(\rho) = -\frac{L}{\tau}(\rho - \rho_{\max}).$$

Иногда в этих формулах вместо средней длины АТС $L$ фигурирует среднее расстояние между соседними АТС в заторе $d \sim 7{,}5$ м (из рис. 2 следует, что $d \sim 6{,}5$ м). Приведенные в этом абзаце формулы активно используются при исследовании роста затора [10] (см. также раздел 2.4). Отметим также, что если известна средняя длина АТС, время, характеризующее реакцию водителей, и желаемая скорость свободного потока, определяющая наклон левой ветви фундаментальной диаграммы, то треугольная фундаментальная диаграмма однозначно строится (рис. 8).



Вернемся к модели. Введем функции двух переменных

$$h(t, x), \quad \rho(t, x) = \frac{1}{h(t, x)}, \quad v(t, x),$$

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

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

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

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

получим

$$v(t, s_n(t)) + \big(v_t(t, s_n(t)) + v(t, s_n(t))v_x(t, s_n(t))\big)\tau \simeq$$
$$\simeq V(\rho(t, s_n(t))) + V'(\rho(t, s_n(t)))\rho_x(t, s_n(t))\frac{1}{2}h(t, s_n(t)),$$
$$h(t, s_n(t)) + v(t, s_n(t))h_x(t, s_n(t)) \simeq v_x(t, s_n(t))h(t, s_n(t)).$$

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

$$v + (v_t + vv_x)\tau \simeq V(\rho) + \frac{V'(\rho)}{2\rho}\rho_x,$$
$$\rho_t + (v\rho)_x \simeq 0.$$

Таким образом, мы «вывели», следуя [7], модель Пейна, получив новую интерпретацию для $\tau$ и

$$D(\rho) = -\frac{V'(\rho)}{2}.$$

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

Использованный выше прием называют *автомодельной редукцией*. Хорошим примером автомодельной редукции является вывод Даламбером (1780) волнового уравнения исходя из модели колебания струны с закрепленными концами, состоящей из множества одинаковых шариков известной массы, соединенных одинаковыми пружинками известной длины и жесткости, имеющими нулевую массу. Поведение каждого шарика описывается вторым законом Ньютона и законом Гука (для отклонения

шарика от положения равновесия $u$, см. рис. 10) и зависит только от положений соседних шариков.

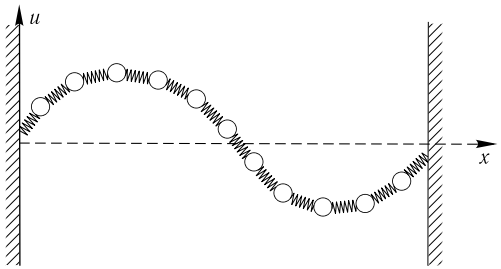



Рис. 10. Колебания струны (Даламбер, 1780)

Таким образом, получается система обыкновенных дифференциальных уравнений, из которой с помощью предельного перехода (шарики выбираются все меньше и меньше, пружинки, соединяющие шарики, становятся все меньше и меньше) получается одно уравнение в частных производных — волновое уравнение, описывающее колебание струны. Другой, более современный (середина 50-х годов XX века), пример автомодельной редукции: задача Ферми—Паста—Улама (вывод уравнения Кортевега—де Фриза, 1895) — см. [92] и цитированную там литературу. Следует заметить, что прием автомодельной редукции, достаточно популярный в математической физике, является эвристическим; для приведенных выше выкладок это особенно очевидно, если вспомнить, что, например, плотность может терпеть разрывы — и ни о каких оценках малости отбрасываемых членов ряда Тейлора не может идти и речи.

Из результатов [18,35,93–114] следует равномерная по времени «схожесть» в поведении решений начальных задач, полученных по исходной модели Ньюэлла и по моделям LWR и Уизема, связанных с моделью Ньюэлла с помощью автомодельной редукции.

Пусть поток АТС однороден и стационарен:

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

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

$$D(\rho) = -\frac{V'(\rho)}{2},$$

см. замечание в конце этого пункта. В обоих случаях ответ одинаков:

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



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

Обратим также внимание на одно интересное обстоятельство. Оказывается, модель Ньюэлла при $\tau = 0$ соответствует хорошо известной экономистам и подробно изученной модели Полтеровича—Хенкина, описывающей распространение новых технологий [100–104, 107–109]. Для обнаружения соответствия достаточно выписать, исходя из модели Ньюэлла, цепочку дифференциальных уравнений для скоростей $v_n(t)$. Если интерпретировать скорость $v_n(t)$ как функцию распределения $F_k(t)$ предприятий в отрасли производства по уровням эффективности $k = -n$, то получим цепочку уравнений Полтеровича—Хенкина.

Заметим, что модель Ньюэлла тесно связана со схемами бегущего счета и дивергентными схемами С. К. Годунова (см., например, [18, 35, 56, 57], а также п. 2.1.4). Но в отличие от разностных схем в модели Ньюэлла время течет непрерывно. Если же дискретизировать время по схеме Эйлера в этой модели, то получится разностная схема, принадлежащая упомянутым выше классам разностных схем, устойчиво аппроксимирующая закон сохранения.

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

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

и выбрать физически осмысленное решение.

**Замечание.** Рассмотрим систему Пейна с $D(\rho) = -\frac{1}{2}V'(\rho)$. Будем искать решение в виде

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

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

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

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

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

Из условия разрешимости системы

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

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

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

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

$$\int\limits_{\Lambda}\vec{z}(k,\,\omega)\exp(ikx - i\omega t)\,d\mu(k,\,\omega),$$

где $\mu(k,\,\omega)$ и $\vec{z}(k,\,\omega)\in\operatorname{Ker}A(k,\,\omega)$ подбираются так, чтобы выполнялось начальное условие (см., например, фундаментальный труд Л. Хёрмандера [115]). Из такого представления следует, что для устойчивости стационарного решения (возмущения стационарного решения со временем затухают) достаточно, а с незначительными уточнениями и необходимо, потребовать, чтобы многообразие $\Lambda$, которое можно представить в виде двух кривых (корней квадратного уравнения — дисперсионного соотношения):

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

лежало в полупространстве $\operatorname{Im}\omega < 0$. Приходим к условию устойчивости: $2\rho^2|V'(\rho)|\tau < 1$.

### 2.2.2. Модель следования за лидером «Дженерал Моторс»

Другим важным классом микроскопических моделей наряду с моделями оптимальной скорости являются *модели следования за лидером* [8,10,58, 60,116].

В 1959 г. сотрудники концерна «Дженерал Моторс» Д. Газис, Р. Херман, Р. Потс [60, 116] предложили одну из первых [1)] нетривиальных микроскопических моделей однополосного транспортного потока, с помощью которой можно получить фундаментальную диаграмму. Простейшим вариантом предложенной модели является следующая модель:

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

обозначения те же, что и выше. Ускорение $n$-го АТС $s_n''(t)$ прямо пропорционально разности скоростей:

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

с коэффициентом пропорциональности (чувствительности), обратно пропорциональным расстоянию до впереди идущего АТС:

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

[1)] Первыми, по-видимому, были модели А. Рёшеля (1950) и Л. Пайпса (1953).



Если $\Delta v_n(t) > 0$, то $s_n''(t) > 0$ — ускорение $n$-го АТС; $\Delta v_n(t) < 0$ — торможение; $\Delta v_n(t) = 0$ — стационарный режим (ускорение равно нулю). Перепишем эту модель следующим образом:

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

или

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

Положим

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

Тогда (в обозначениях предыдущего пункта)

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

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

В 1961 г. Д. Газис, Р. Херман и Р. Розэри предложили следующую модель [58, 116]:

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

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

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

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

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

### 2.2.3. Модель Трайбера «разумного водителя»

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

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

Как показали численные эксперименты, наиболее «удачной» моделью этого класса является [1] *модель М. Трайбера* (1999) [10, 58, 60, 117, 118]:

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

Первое слагаемое

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

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

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

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

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

естественно называть моделью следования за лидером.

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

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

причем желаемая дистанция определяется следующим образом:

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

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

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

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

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

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

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

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

В равновесном потоке одинаковых АТС, когда

$$s_n''(t) \equiv 0, \quad s_{n+1}'(t) - s_n'(t) \equiv 0, \quad s_n'(t) \equiv V,$$

имеем:

$$d(V) \stackrel{\text{def}}{=} s_{n+1}(t) - s_n(t) = d^*(V, 0)\left(1 - \left(\frac{V}{V^0}\right)^{\delta}\right)^{-1/2}.$$

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

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

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

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

### 2.2.4. Модели клеточных автоматов

В моделях клеточных автоматов (CA-моделях) дорога разбивается на клетки, дискретным считается и время. Часто, но далеко не всегда [83–86], считается, что в клетке может находиться не больше одного АТС. Таким образом, получаются разностные аналоги рассматриваемых ранее макроскопических уравнений. Заметим также, что часто и множество возможных значений скорости АТС считают дискретным в таких моделях.

Концепция клеточных автоматов была введена Дж. фон Нейманом в 50-е годы XX века [119] в связи с разработкой теории самовоспроизводящихся машин. Применять клеточные автоматы для моделирования транспортных потоков предлагалось в работе [120]. Однако активное использование этой концепции началось только после работы К. Нагеля и М. Шрекенберга [121]; подробности см. в обзорах [122, 123].

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

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

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

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

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

где $d \sim 7{,}5$ м (см. п. 2.2.3).

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

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

Шаг 4. *Движение*:

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

Все четыре приведенных шага необходимы для воспроизведения основных свойств транспортного потока. Так, например, шаг 3 учитывает неустойчивость транспортного потока при достаточно больших плотностях.

В работах [124, 125] описан переход от моделей типа Нагеля—Шрекенберга к гидродинамическим моделям (типа Бюргерса, Хопфа) — гидродинамический предельный переход. А в работе [126] описан обратный переход — ультраметрический предельный переход.



Продемонстрируем, следуя М. Л. Бланку [127–129] (см. также приложение М. Л. Бланка и цитированную там литературу), один из способов вывода уравнения состояния $V(\rho)$, исходя из довольно простой CA-модели.

Рассмотрим кольцевую дорогу, состоящую из $n$ ячеек (клеток). В каждой ячейке может находиться не более одного АТС. Длины всех ячеек одинаковы и равны (условной) единице. Будем также считать, что $n$ достаточно большое. Если брать не кольцевую топологию, а, скажем, бесконечную прямолинейную дорогу (полосу), то по $n$ необходимо будет сделать «*термодинамический предельный переход*» (это понятие пришло из статистической физики, см., например, [130] и цитированную там литературу) — устремить $n$ к бесконечности, «сохраняя пропорции». Пусть в начальный момент времени в некоторые из ячеек поместили АТС. Обозначим через $0 < \rho \leqslant 1$ долю занятых ячеек. Будем считать, что сначала все АТС «смотрят» в следующую по ходу движения ячейку, а потом те из АТС, для которых эти ячейки оказались свободными, независимо от остальных двигаются в свободную ячейку с вероятностью $0 < p \leqslant 1$. И так происходит на каждом шаге по времени.

Определим *среднюю пространственную* скорость:

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

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

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

При $p = 1$ см. [127].

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

Для дальнейшего знакомства с CA-моделями см. [131]. Оказывается, что простые модификации рассматриваемых здесь моделей демонстрируют [122, 129, 131], что одной плотности соответствует целый набор средних скоростей, или последнее понятие может не быть корректно определено. С точки зрения фазовых переходов описанное поведение соответствует возникновению новой «гистерезисной» фазы, которая, по-видимому, соответствует режиму синхронизированного движения (см. [10] или главу 3).

Интересные идеи исследования многополосности, которые также объясняют экспериментальную «размазанность» фундаментальной диаграм-

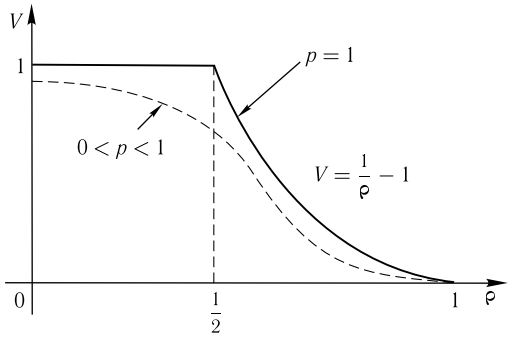



Рис. 11. «Фазовый переход» при $\rho = 1/2$

мы, недавно были предложены А. П. Буслаевым и др. (кафедра высшей математики МАДИ) [132–134]. Для описания транспортного потока использовалась модель Танака, рассмотренная в п. 2.1.2. В ней помимо плотности и скорости еще вводился «параметр регулярности», с помощью которого часть водителей АТС «разрыхляет регулярное движение» для того, чтобы осуществить относительное перемещение внутри потока. Отметим также, что с помощью моделей такого типа удалось теоретически объяснить невогнутость фундаментальной диаграммы (рис. 2) в случае многополосного движения.

Перейдем теперь к другому типу клеточных моделей, которые, по сути, являются разностными аналогами рассматриваемых ранее уравнений — СТМ-моделям. Приведем схему, в которую «ложатся» многие модели этого класса. Ограничимся ситуацией магистрали (вообще говоря, многополосной) с въездами и выездами. Разобьем магистраль на клетки (ячейки) — прямолинейные участки дороги длиной не менее сотни метров. Будем считать (не ограничивая общности), что в каждую клетку имеется только один въезд и из каждой клетки имеется только один выезд (см. рис. 12). Тогда

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

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

Так, в работах [87, 88] (рассматривалась кольцевая топология транспортной сети без въездов и выездов) полагали

$$q_i(m) = (1+\alpha_i)^{-1} Q_i(n_i(m))$$

— аналог схемы бегущего счета для модели LWR. Очевидным недостатком этой схемы является возможность следующих «неправдоподобных» ситуаций (считаем $\alpha_i \equiv 0$ и $r_i(m) \equiv 0$). Предположим, что есть две клетки (без въездов и выездов). Первая клетка полностью загружена (в ней максимальная плотность АТС — «стоячая пробка»), а следующая по ходу движения клетка полностью свободна (в ней нет АТС). Тогда согласно выбранному способу описания потока АТС ничего происходить не будет, т. е. ситуация со временем меняться не будет. В то время как из опыта известно, что АТС начнут «перетекать» в свободную клетку. Предположим теперь, что первая клетка загружена, например, наполовину, а в следующей по ходу движения клетки — стоячая пробка. Тогда $q_i(m) \equiv 0$, а модель говорит об обратном. Тем не менее исследование «до критических режимов» — «левая (возрастающая) ветка» фундаментальной диаграммы — с помощью этой модели вполне корректно.



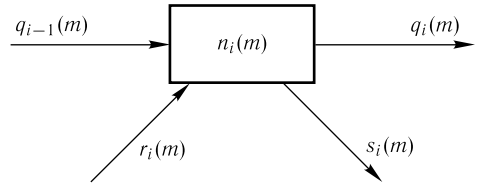



Рис. 12

В работах [83–86] рассматривался случай треугольной фундаментальной диаграммы (см. рис. 8), «вершина» которой имеет координаты $(\tilde{n}_i,\ Q_i^{\max})$, и использовалась расчетная схема

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

— аналог схемы Годунова (см. п. 2.1.4) для модели LWR с треугольной фундаментальной диаграммой (СТМ-модель К. Даганзо, 1994). Поясним обозначения:

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

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

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

Можно обобщить СТМ-модель на графы транспортной сети более сложной топологии, чем кольцевая или линейная магистраль. Правда,

для этого потребуется знать матрицу перемешивания в каждом узле графа транспортной сети. Кроме того, нужно дополнительно (по сравнению с магистральной топологией) прописать разрешение «конфликтных ситуаций» — знания матрицы перемешивания может оказаться недостаточно. Например, рассмотрим такую ситуацию: пусть имеется перекресток. Две трети водителей с входящей в перекресток дороги $A$ хотят повернуть на выходящую с перекрестка дорогу $C$, и две трети водителей с входящей дороги $B$ также хотят повернуть на дорогу $C$. Будем считать, что двухполосная дорога $C$ может пропустить 4000 АТС/час (максимальная пропускная способность). Дороги $A$ и $B$ также двухполосные, и на обеих из них поток АТС, входящий в перекресток, по 4000 АТС/км. Очевидно, что ситуация недоопределена. Для понимания того, что будет происходить, необходимо еще знать, например, режим работы светофора в этом перекрестке (в случае его наличия). Таким образом, реальное расщепление потоков зависит не только от матрицы перемешивания, но и, например, от режима работы светофора. Зная матрицу перемешивания и «беря на вооружение» правило работы светофора, можно получить упомянутое обобщение рассмотренных моделей на графы транспортной сети общего вида. Аналогично можно рассмотреть и более сложные развязки. Здесь также можно упомянуть, что еще в 1936 году молодой профессор Московского университета А. Н. Колмогоров в письме в журнал «Строительство Москвы», по сути, обсуждал вопрос, связанный с правильной организацией перекрестка [135].

Как уже отмечалось, несмотря на свою относительную простоту, СТМ-модель является одной из наиболее востребованных в приложениях (см., например, [83–86] и цитированную там литературу).

Что касается изучения аналитических свойств этих моделей, полученных с помощью разностных схем для LWR-модели, то, по-видимому, проще исследовать все-таки дифференциально-разностные аналоги LWR-моделей, т. е. считать, что время течет непрерывно [87,88]. Разностные схемы (в том числе и упомянутые выше) очевидным образом переделываются в дифференциально-разностные, при этом автоматически выполняется необходимое условие Куранта—Фридрихса—Леви корректности схемы, аппроксимирующей уравнение LWR: шаг по времени достаточно мал по сравнению с шагом по пространственной переменной, см. п. 2.1.4.

В связи со всем вышесказанным возникает задача: *исследовать положения равновесия — стационарные режимы динамической системы, а также бассейны их притяжения и отталкивания, на графе транспортной сети общего вида, полученной с помощью дифференциально-разностной схемы Годунова из LWR-модели (с треугольной или параболической фундаментальной диаграммой)*. В 2004 г. для схемы бегущего счета исследования в этом направлении были предприняты А. П. Буслаевым и др. [87] для графов специальной структуры (кольцевой, «цветочной» и т. п.). В 2006 г. А. И. Назаров обобщил результаты статьи [88] на графы общего вида. Однако, как уже отмечалось выше, схема бегущего счета — не самый подходящий вариант для описания транспортного потока во всевозможных состояниях. В диссертации 2007 г. А. А. Куржанского [83] исследовалась асимптотическая устойчивость (глобальная) положений равновесия, образующих многообразие с довольно простой и полностью описанной линейной структурой, для СТМ-модели (схема Годунова + треугольная фундаментальная диаграмма) транспортных потоков на магистрали, т. е. с простой топологией графа транспортной сети.



Выпишем систему обыкновенных дифференциальных уравнений на графе транспортной сети [1)], используя лишенную деталей модификацию правила «пропорциональных приоритетов» [84, 86]:

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

здесь каждое ребро ориентированного графа транспортной сети пронумеровано, запись $j \to i$ означает, что с ребра $j$ можно повернуть на ребро $i$; $\rho_i$ — плотность потока на $i$-м ребре, $l_i$ — длина $i$-го ребра, $\beta_i^j$ — доля потока АТС на ребре $j$, ответвляющаяся на ребро $i$. Обратим внимание, что в общем случае следует считать $\beta_i^j(t, \rho_1, \rho_2, \ldots)$ функцией от $\rho_1, \rho_2, \ldots$ Причем если учитывать задержки в узлах графа транспортной сети, связанные, например, с наличием светофоров, то, вообще говоря, $\sum\limits_{k:\, i\to k} \beta_k^i(t, \rho_1, \rho_2, \ldots) < 1$. Имеются и другие «правила обработки» узлов графа транспортной сети (см., например, [36, 81, 82]).

**Упражнение**** (мотивированное работами [83, 87, 88]). Для каждой замкнутой транспортной сети стационарный режим будет устойчивым, если значения стационарных плотностей «лежат» на левых (возрастающих) ветках соответствующих треугольных фундаментальных диаграмм. Это утверждение можно обобщить и на открытые сети.

**Упражнение** (о стационарном распределении потоков; А. М. Валуев). Рассмотрим поток автомобилей, движущийся по кольцевой дороге. Будем считать, что на ней заданы точки входа (источники) и выхода (стоки), причем в каждом стоке доля автомобилей, уходящих с кольца, по отношению к количеству подъезжающих к нему постоянна. Обоснуйте следующее утверждение: при постоянных интенсивностях потоков, поступающих из

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

каждого источника, на кольце устанавливается стационарное распределение потоков по отдельным участкам, если их пропускная способность достаточно велика. Можно ли обобщить (при каких условиях?) это утверждение на ориентированный граф дорожной сети общего вида? Установите зависимость между стационарным распределением потоков и входными величинами — интенсивностями входных потоков и коэффициентами выбытия в стоках.

## 2.3. Модельные задачи

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

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

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

Напомним необходимые соотношения (по модели LWR):

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

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

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

где $\rho_0(x)$ — ограниченная измеримая функция (см. рис. 13, 14). Для определенности будем считать в формулах (14), (15) $\rho_- < \rho_+$. Случай $\rho_- > \rho_+$ рассматривается аналогично и может быть заменой $\rho \to -\rho$ сведен к случаю $\rho_- < \rho_+$ (с $Q(\rho) := Q(-\rho)$). Задача Коши (13), (14) называется *задачей* (*Римана*) *о распаде* (*произвольного*) *разрыва*.



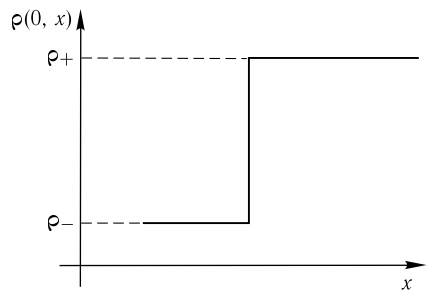



Рис. 13. Начальное условие Римана

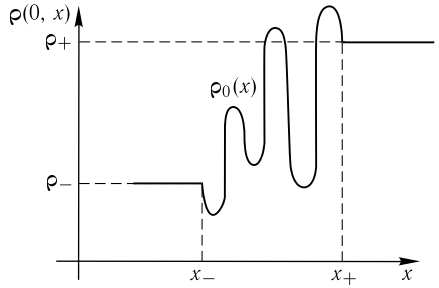



Рис. 14. Начальное условие типа Римана

Мы уже можем частично представить себе ее решение, основываясь на примерах, разобранных в п. 2.1.1. Для того чтобы точнее понять, какое решение будет у этой задачи, заметим сначала, что:

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

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

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

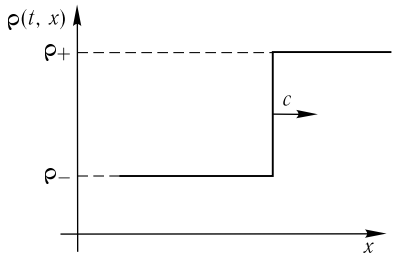



Рис. 15. Ударная волна

- закон сохранения (13) имеет двухпараметрическое ($t \to t + \tau$; $x \to x + a$) семейство автомодельных решений вида *волны разре-*

*жения* (см. рис. 16):

$$g\left(\frac{x}{t}\right)=\begin{cases}\rho_-, & x<Q'(\rho_-)t,\\ Q'^{-1}\left(\dfrac{x}{t}\right), & Q'(\rho_-)t\leqslant x<Q'(\rho_+)t,\\ \rho_+, & x\geqslant Q'(\rho_+)t\end{cases}\qquad(16)$$

($Q'^{-1}(\cdot)$ означает обратную функцию к функции $Q'(\cdot)$), тогда и только тогда, когда $Q''(\rho)>0$ при $\rho\in(\rho_-,\rho_+)$, за исключением, быть может, конечного числа точек, в которых имеет место равенство (напомним, что $\rho_-<\rho_+$).

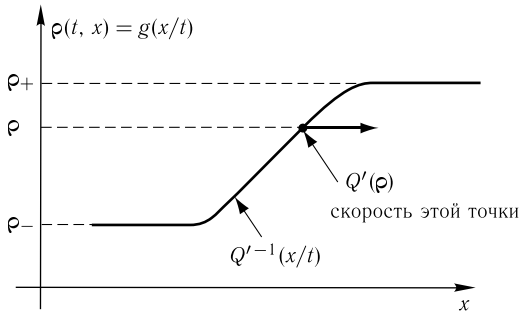



Рис. 16. Волна разрежения

**Замечание.** По поводу автомодельных решений см., например, [1, 136, 137], а также более ранние книги этих авторов. Грубо говоря, автомодельные решения — это решения рассматриваемого уравнения в частных производных, которые могут быть представлены как функции (в данном случае) одного аргумента, зависящего от пространственных и временных переменных, т. е. от $t$ и $x$. Групповой анализ указывает на то, что если автомодельные решения существуют, то прежде всего их следует искать в виде неизвестной функции от инвариантов (сконструированных из аргументов неизвестной функции, т. е. $t$ и $x$) группы преобразований, допускаемой рассматриваемым уравнением. Поясним сказанное немного подробнее. Хорошо известно, что сфера, например, допускает группу ортогональных преобразований (всевозможных поворотов), т. е. сфера будет переходить сама в себя при таких преобразованиях. С другой стороны, на сферу можно смотреть как на многообразие, заданное хорошо известным уравнением. Групповой анализ предлагает смотреть на уравнение в частных производных как на многообразие, заданное этим уравнением в продолженном (на различные частные производные) пространстве. Правда, в отличие от обычных многообразий, на класс возможных групп преобразований налагается условие, по сути определяющее действие группы на продолженных переменных, через действие группы на неизвестную функцию и ее аргументы.

В качестве примера укажем, что закон сохранения (13) допускает группу трансляций по времени и по координате, группу подобных преобразований временной



и пространственной переменной. Это означает, что вид уравнения (13) не меняется при заменах:

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

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

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

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

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

в частности, следует, что решение уравнения (13) прежде всего следует попробовать искать в виде

$$\tilde\rho(x-ct)\quad\text{и}\quad g\left(\frac{x}{t}\right).$$

Роль автомодельных решений или параметрических семейств таких решений в теории эволюционных уравнений в частных производных часто аналогична роли неподвижных точек в теории обыкновенных дифференциальных уравнений [1, 92–100, 103–114, 136–147]. Параметрическое семейство автомодельных решений или семейство, полученное путем «склеивания» автомодельных решений, часто является, например, притягивающим (другие решения) семейством (см. теоремы 1, 2 в п. 2.3.1).

Зная приведенный выше в этом пункте материал и тот материал, который был изложен в п. 2.1.1, И. М. Гельфанд построил в конце 50-х годов XX века решение *задачи о распаде произвольного разрыва* [17, 30], т. е. решение задачи Коши (13), (14). Можно было также добавить (см. п. 2.1.3): «построил с помощью метода исчезающей вязкости», поскольку, как будет показано ниже в этом пункте, выполнение RRH-условия и E-условия на разрывах обобщенных решений уравнения (13) есть прямое следствие этого метода.

Приведем способ построения решения. Строгое обоснование см. в оригинальных, и отчасти монографических, записях курса лекций [30], которые С. Н. Кружков читал в конце 60-х и начале 70-х годов XX века на мехмате МГУ. Для этого введем функцию

$$\tilde Q(\rho)=F^{**}(\rho),\qquad F(\rho)=Q(\rho)+I_{[\rho_-,\rho_+]}(\rho),$$

$$I_{[\rho_-,\rho_+]}(\rho)=\begin{cases}0, & \rho\in[\rho_-,\rho_+],\\ \infty, & \rho\notin[\rho_-,\rho_+],\end{cases}$$

где

$$F^*(x)=\sup_{y\in\mathbb{R}}(xy-F(y))$$

— функция, сопряженная (по Юнгу—Фенхелю—Лежандру) к $F(y)$ [47]. Таким образом, $\widetilde{Q}(\rho)$ — нижняя граница выпуклой оболочки множества

$$\{(p, q)\colon \rho \in [\rho_-, \rho_+],\ q \geqslant Q(\rho)\}.$$

Будем искать автомодельное решение уравнения

$$\frac{\partial \rho}{\partial t} + \frac{\partial \widetilde{Q}(\rho)}{\partial x} = 0 \tag{17}$$

вида

$$\rho(t, x) := \rho\Big(\frac{x}{t}\Big),$$

удовлетворяющее начальному условию (14). Заметим, что $\xi = x/t$ является инвариантом группы растяжений: $x' \to kx$, $t' \to kt$, допускаемой законом сохранения (17)). Подстановка $\rho(x/t)$ в (17) приводит к обыкновенному дифференциальному уравнению

$$\rho'(\xi)(\xi - \widetilde{Q}'(\rho)) = 0.$$

Откуда с учетом начального условия (14) следует, что

$$\rho(t, x) = \widetilde{Q}'^{-1}\Big(\frac{x}{t}\Big),$$

где $\widetilde{Q}'^{-1}(\cdot)$ — обратная функция к $\widetilde{Q}'(\cdot)$. Для уточнений и большей наглядности представим решение немного в другом виде. Положим (см. рис. 17)

$$\{\rho \in [\rho_-, \rho_+]\colon Q(\rho) > \widetilde{Q}(\rho)\} = (\alpha_0, \beta_0) \cup (\alpha_1, \beta_1) \cup \ldots \cup (\alpha_n, \beta_n),$$
$$c_0 = \widetilde{Q}'(\alpha_0) = Q'(\beta_0), \quad c_n = Q'(\alpha_n) = \widetilde{Q}'(\beta_n),$$
$$c_k = Q'(\alpha_k) = Q'(\beta_k), \quad k = 1, \ldots, n-1,$$

где $\rho_- = \alpha_0 \leqslant \beta_0 \leqslant \alpha_1 < \beta_1 \leqslant \alpha_2 < \ldots < \beta_{n-1} \leqslant \alpha_n \leqslant \beta_n = \rho_+$; очевидно, что $c_{k-1} \leqslant c_k$.

Будем также считать (для простоты и наглядности формулировок), что выполняются условия «отсутствия прилипания» [109]:

$$\begin{aligned}
&Q''(\rho) > 0 && \text{при } \rho \in [\beta_k, \alpha_{k+1}],\ k = 0, \ldots, n-1;\\
&Q''(\alpha_0) > 0, && \text{если } \alpha_0 < \beta_0 \text{ и } Q'(\alpha_0) = c_0;\\
&Q''(\beta_n) > 0, && \text{если } \alpha_n < \beta_n \text{ и } Q'(\beta_n) = c_n.
\end{aligned}$$

Тогда (см. рис. 18 с $d_k = 0$)

$$\rho(t, x) = \begin{cases} \rho_-, & x < c_0 t,\\ Q'^{-1}(x/t), & c_{k-1} t \leqslant x < c_k t,\ k = 1, \ldots, n,\\ \rho_+, & x \geqslant c_n t. \end{cases} \tag{18}$$



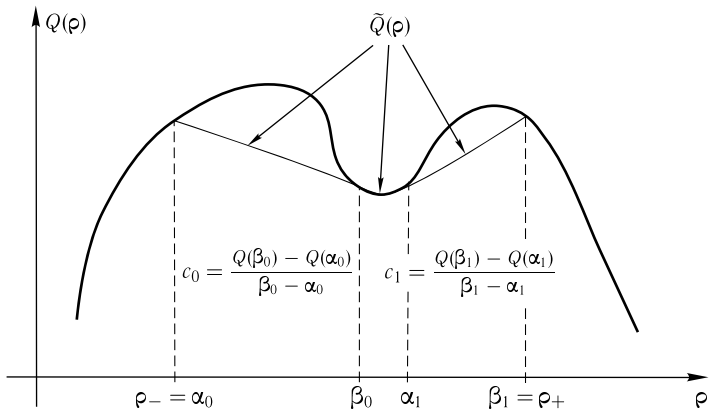



Рис. 17. Построение И. М. Гельфанда (1958)

Заметим, что по построению мы можем быть уверены лишь в нестрогих неравенствах:

$$\begin{aligned}
&Q''(\rho) \geqslant 0 && \text{при } \rho \in [\beta_k, \alpha_{k+1}],\ k = 0, \ldots, n-1;\\
&Q''(\alpha_0) \geqslant 0, && \text{если } \alpha_0 < \beta_0 \text{ и } Q'(\alpha_0) = c_0;\\
&Q''(\beta_n) \geqslant 0, && \text{если } \alpha_n < \beta_n \text{ и } Q'(\beta_n) = c_n.
\end{aligned}$$

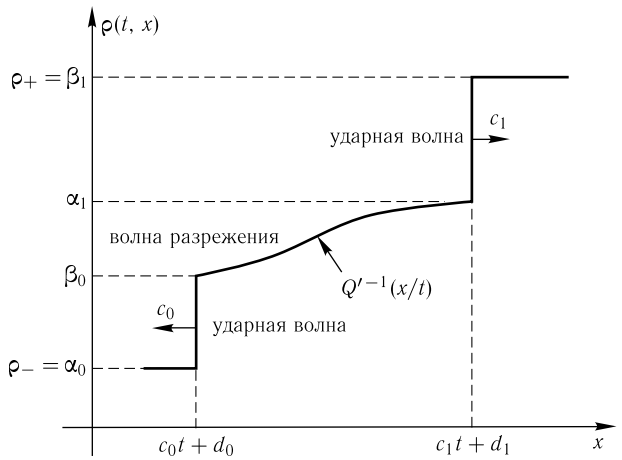



Рис. 18. Система волн

Если же условия «отсутствия прилипания» не выполняются, то все приводимые ниже результаты остаются в принципе верными (с некоторыми

несущественными для приложений оговорками), но оценки будут менее точными (см. [113, 114] и цитированную там литературу).

Заметим также, что $\rho(t, x)$ определяется с точностью до почти всюду по $x$ при любом фиксированном значении $t \geqslant 0$ (см. п. 2.1.3). Поэтому в формуле (18) можно не доопределять значения $\rho(t, x)$ на разрывах. Естественно называть разрывы решения (18) *ударными волнами*, а

$$Q'^{-1}\left(\frac{x}{t}\right), \quad c_{k-1}t < x < c_k t, \quad k = 1, \ldots, n,$$

— *волнами разрежения*; волна разрежения исчезает, если $c_{k-1} = c_k$, поскольку в этом случае $\beta_k = \alpha_k$.

Естественно ожидать, что если немного «размазать разрыв» в начальных условиях (14), то решение уравнения (13) будет иметь схожий вид. Иначе говоря, мы ожидаем, что решение задачи Коши (13), (15) будет структурно похоже на (18). Действительно, имеет место следующий факт [14, 93, 96–99, 109–114] (см. рис. 18).

**Теорема 1.** *Существует такой набор* $\{d_k\}_{k=0}^n$, *что решение задачи Коши* (13), (15) *сходится при* $t \to \infty$ *в* $L_1(\mathbb{R}_x)$ *к системе волн*

$$\tilde{\rho}(t, x; \{d_k\}_{k=0}^n) = \begin{cases} \rho_-, & x < c_0 t + d_0, \\ Q'^{-1}(x/t), & c_{k-1}t + d_{k-1} \leqslant x < c_k t + d_k,\ k = 1, \ldots, n, \\ \rho_+, & x \geqslant c_n t + d_n, \end{cases} \tag{19}$$

*причем если* $c_{k-1} = c_k$, *то* $d_{k-1} \leqslant d_k$.

Заметим, что в работах [93, 98, 99] также были найдены формулы для расчета $d_k$.

Рассмотрим теперь модель Уизема:

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \varepsilon \frac{\partial}{\partial x}\left(D(\rho)\frac{\partial \rho}{\partial x}\right), \quad D(\rho) > 0, \quad \varepsilon > 0. \tag{20}$$

«Построим» решение задачи Коши (20), (15). Поскольку задачи Коши (13), (15) и (20), (15) описывают одно и то же явление на разных уровнях детализации — эволюцию («размазанного») глобального затора в транспортном потоке, то мы вправе ожидать, что решение задачи Коши (20), (15) будет структурно похоже на решение задачи Коши (13), (15) и, стало быть, будет вести себя на больших временах подобно системе волн (19).

Действуя по тому же плану, что и при исследовании асимптотики (по времени) задачи Коши (13), (15), найдем сначала автомодельные решения уравнения (20). Согласно работам [17, 93]:

- закон сохранения с диффузией (20) имеет однопараметрическое с параметром $x_0 \in \mathbb{R}$ семейство *автомодельных решений* вида *бегущей волны* (см. рис. 19):

$$\tilde{\rho}_\varepsilon(x - ct + x_0),$$



таких, что

$$\lim_{s \to -\infty} \tilde{\rho}_\varepsilon(s) = \rho_-, \qquad \lim_{s \to \infty} \tilde{\rho}_\varepsilon(s) = \rho_+,$$

тогда и только тогда, когда выполняется RRH-условие

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

и *строгое E-условие* (напомним, что $\rho_- < \rho_+$)

$$\forall \rho \in (\rho_-, \rho_+) \qquad \sigma(\rho_-, \rho_+) < \sigma(\rho_-, \rho);$$

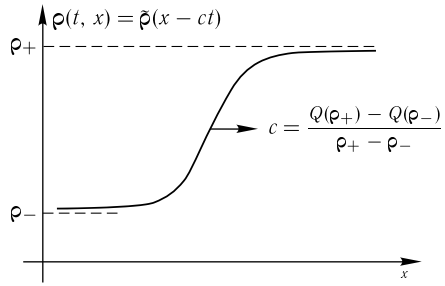



Рис. 19. Бегущая волна

- закон сохранения с диффузией (20) имеет двухпараметрическое семейство *асимптотически автомодельных решений* (при подстановке такого автомодельного «решения» в уравнение (20) почти всюду возникает невязка (в нашем случае $O(1/t^2)$, которая стремится к нулю с ростом времени) вида *волны разрежения* (16) тогда и только тогда, когда $Q''(\rho) > 0$ при $\rho \in (\rho_-, \rho_+)$, за исключением, быть может, конечного числа точек, в которых имеет место равенство.

Заметим, что для доказательства всех приведенных утверждений относительно бегущей волны достаточно рассмотреть краевую задачу для обыкновенного дифференциального уравнения с разделяющимися переменными, которое получается при подстановке

$$\tilde{\rho}_\varepsilon(x - ct + x_0)$$

в (20) и интегрировании возникшего уравнения с учетом одного из краевых условий $\lim\limits_{s \to -\infty} \tilde{\rho}_\varepsilon(s) = \rho_-$:

$$\varepsilon \frac{d\tilde{\rho}_\varepsilon(s)}{ds} = \frac{Q(\tilde{\rho}_\varepsilon(s)) - c\tilde{\rho}_\varepsilon(s) - (Q(\rho_-) - c\rho_-)}{D(\tilde{\rho}_\varepsilon(s))}.$$

Заметим также, что RRH-условие получается из оставшегося краевого условия

$$\lim_{s\to\infty} \tilde\rho_\varepsilon(s) = \rho_+.$$

А строгое E-условие равносильно требованию отсутствия (за исключением $s = \pm\infty$) особых точек (точек, в которых $\tilde\rho'_\varepsilon(s) = 0$) у этого обыкновенного дифференциального уравнения, что в свою очередь равносильно условию

$$\tilde\rho'_\varepsilon(s) > 0 \quad \text{(поскольку } \rho_- < \rho_+\text{)}.$$

Если же допустить существование особой точки $s^* \in (-\infty, \infty)$, то интеграл

$$\int (Q(\tilde\rho_\varepsilon) - c\tilde\rho_\varepsilon - (Q(\rho_-) - c\rho_-))^{-1} d\tilde\rho_\varepsilon$$

будет иметь неинтегрируемую особенность в точке $\tilde\rho_\varepsilon = \tilde\rho_\varepsilon(s^*)$. Отсюда следует, что $s^* = \infty$ или $s^* = -\infty$. Пришли к противоречию с предположением $s^* \in (-\infty, \infty)$.

Теперь можно объяснить уже анонсированную возможность обоснования RRH-условия и E-условия на разрывах обобщенных решений уравнения (13) с помощью метода исчезающей вязкости:

$$\tilde\rho_\varepsilon(x - ct + x_0) = \tilde\rho_1\Big(\frac{x - ct + x_0}{\varepsilon}\Big) \xrightarrow[\varepsilon\to 0+]{} \begin{cases} \rho_-, & x \leqslant ct - x_0, \\ \rho_+, & x > ct - x_0. \end{cases}$$

**Упражнение*** [19, 30]. Почему строгое E-условие «переходит в пределе» при $\varepsilon \to 0+$ просто в E-условие?

**Замечание** (модель Лайтхилла—Уизема—Кортевега—де Фриза—Бюргерса). Если, подобно модели Уизема (см. п. 2.1.3), положить

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

где $V(\rho(t, x))$ — желаемая скорость при данной плотности, а выражение в скобке описывает «дальнозоркость» водителей (поэтому $\varepsilon > 0, \mu > 0$), то получим, с учетом закона сохранения количества АТС:

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

*уравнение типа Кортевега—де Фриза—Бюргерса* (КдФБ):

$$\frac{\partial\rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = \varepsilon\frac{\partial^2\rho}{\partial x^2} + \mu\frac{\partial^3\rho}{\partial x^3}. \tag{21}$$

Поскольку предложенная модель описывает то же самое явление, что и модель Уизема, то мы вправе надеяться на структурную схожесть автомодельных решений



уравнений (20) и (21). Действительно, предположим, что выполняется строгое E-условие, а скорость $c$ определяется формулой RRH. Тогда если [1)]

$$4\sup_{\rho\in(\rho_-,\rho_+)}\Big(c - \frac{Q(\rho) - Q(\rho_+)}{\rho - \rho_+}\Big) \leqslant \frac{\varepsilon^2}{\mu},$$

то существует однопараметрическое семейство строго возрастающих автомодельных решений уравнения (21) вида бегущей волны $\tilde\rho_{\varepsilon,\mu}(x - ct + x_0)$ [141]. Легко показать, что уравнение (21) при $Q''(\rho) > 0$, $\rho \in (\rho_-, \rho_+)$ имеет и асимптотическое автомодельное решение вида волны разрежения (16). Интересно при этом заметить, что в отличие от уравнения типа Бюргерса для уравнения типа КдФБ обоснование метода исчезающей вязкости и дисперсии — по-прежнему открытая задача (см. статью П. Лакса в сборнике [32], а также [36, 38]).

Приведенная ниже теорема резюмирует результаты работ [14, 93, 95, 105–113], см. рис. 20. Для простоты формулировки мы также считаем, что $c_{k-1} < c_k$, $k = 1, \ldots, n$ (или, что то же самое, $\beta_k < \alpha_k$, $k = 1, \ldots, n$). Общий случай изложен в статье [113].

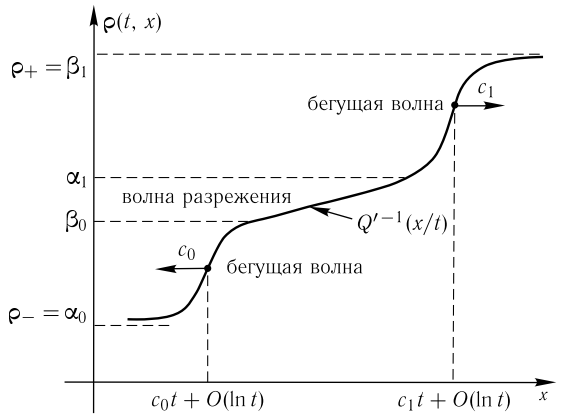



Рис. 20. Система волн

**Теорема 2.** *Существует такой набор функций*

$$\{d_k(t)\}_{k=0}^n = O(\varepsilon \ln t) + O(1),$$

[1)] Если это условие нарушается, то сначала (когда не сильно нарушается) теряется свойство монотонности бегущей волны, возникают затухающие осцилляции типа $(\sin x)/x$ при стремлении решения к предельным значениям на краях, а затем автомодельные решения вида бегущей волны вообще перестают существовать. Численное подтверждение этих фактов имеется в работах А. Г. Куликовского с коллегами (см., например, [56, 148] и цитированную там литературу). Аналитические исследования, подтверждающие сказанное выше, были недавно проведены в [149].

*что решение задачи Коши* (20), (15) *сходится при* $t \to \infty$ *в* $L_\infty(\mathbb{R}_x)$ (*равномерно по* $x \in \mathbb{R}$) *к системе волн*

$$\tilde{\tilde{\rho}}(t,\, x;\, \{d_k(t)\}_{k=0}^n) = \begin{cases} \rho_-, & x < c_0 t - \sqrt{t}, \\ \tilde{\rho}^k(x - c_k t + d_k(t)), & c_k t - \sqrt{t} \leqslant x < c_k t + \sqrt{t}, \\ & k = 0,\, \ldots,\, n, \\ Q'^{-1}(x/t), & c_{k-1} t + \sqrt{t} \leqslant x < c_k t - \sqrt{t}, \\ & k = 1,\, \ldots,\, n, \\ \rho_+, & x \geqslant c_n t + \sqrt{t}, \end{cases}$$

*где* $\tilde{\rho}^k(x - c_k t)$ *— решение уравнения* (20) *вида бегущей волны с*

$$\lim_{s \to -\infty} \tilde{\rho}^k(s) = \alpha_k, \quad \lim_{s \to \infty} \tilde{\rho}^k(s) = \beta_k.$$

**Замечания.** 1) Из работ [93, 94, 110, 113] следует, что если

$$\rho_- = \alpha_0 < \beta_0 = \rho_+ \quad \text{и} \quad Q'(\alpha_0) \neq c_0, \quad Q'(\beta_0) \neq c_0,$$

то

$$\lim_{t \to \infty} d_0(t) = d_0,$$

где $d_0$ находится из первого интеграла (20) (при $t = 0$):

$$I(t;\, d_0) = \int\limits_{-\infty}^{\infty} \{\rho(t,\, x) - \tilde{\rho}^0(x - c_0 t + d_0)\}\, dx \equiv 0.$$

Полезной находкой А. М. Ильина и О. А. Олейник стало введение функции

$$w(t,\, x) = \int\limits_{-\infty}^{x} \{\rho(t,\, s) - \tilde{\rho}^0(s - c_0 t + d_0)\}\, ds,$$

которая так же, как и $\rho(t,\, x)$, удовлетворяет уравнению параболического типа, но при этом $\lim\limits_{x \to \pm\infty} w(t,\, x) \equiv 0$.

2) Можно показать (см. [107–109]), что асимптотические оценки сдвигов фаз

$$\{d_k(t)\}_{k=0}^n = O(\varepsilon \ln t) + O(1)$$

разумно определять из предложенных в работе [107] «локализованных законов сохранения (первых интегралов)»:

$$I_k(t;\, d_k(t)) = \int\limits_{c_k t - \sqrt{t}}^{c_k t + \sqrt{t}} \{\rho(t,\, x) - \tilde{\rho}^k(x - c_k t + d_k(t))\}\, dx \equiv 0.$$

Если приравнять нулю полную производную по времени от $I_k(t;\, d_k(t))$, используя уравнение (20), то получим оценку $d_k(t)$ (приняв за гипотезу, которую апостериорно проверим, что $d_k'(t) = o(t^{-1/2})$). Точность оценки определяется значениями разностей:

$$\rho(t,\, c_k t \pm \sqrt{t}) - \tilde{\rho}^k(\pm\sqrt{t} + d_k(t)) \quad \text{и} \quad \rho_x(t,\, c_k t \pm \sqrt{t}) - \tilde{\rho}^k_x(\pm\sqrt{t} + d_k(t)).$$



Эти значения в свою очередь определяются исходя из исследования сходимости решения к системе волн на участках, соответствующих поведению «волна разрежения». Оказывается, что выбор зависимости $\sqrt{t}$ в определении системы волн и в определении локализованных законов сохранения дает наилучшую по точности оценку сдвигов фаз, исходя из имеющихся способов оценки разностей. В заключение отметим, что при $D(\rho) \equiv 1$ в работах [106, 108, 109] были найдены первые (логарифмические) члены асимптотических рядов для сдвигов фаз.

3) Из работ [111–113] следует, что в теореме 2 скорость сходимости есть [1)] $O(1/\sqrt[3]{t})$, т. е.

$$\rho(t,\, x) = \tilde{\tilde{\rho}}(t,\, x;\, \{d_k(t)\}_{k=0}^n) + O\Big(\frac{1}{\sqrt[3]{t}}\Big).$$

В теореме 1 (см. [37, 93, 98, 99]) скорость сходимости в равномерной норме (за вычетом сколь угодно малых, но фиксированных на момент рассмотрения окрестностей ударных волн решения) есть $O(1/\sqrt{t})$, причем эта оценка неулучшаема [93] и достигается на асимптотике вида волны разрежения.

4) Ввиду замечания на с. 130 можно ожидать, что результат, аналогичный теореме 2, имеет место и для задачи Коши (21), (15), если только $\mu/\varepsilon^2$ достаточно мало; насколько мало, зависит как от функции $Q(\rho)$, так и от начального условия (15). Для случая, когда $Q(\rho)$ — вогнутая парабола, локальная сходимость к одному из представителей однопараметрического семейства бегущих волн (параметр — сдвиг фазы, определяется, как в замечании 1) была установлена П. И. Наумкиным и И. А. Шишмарёвым (1991) [142, 143]. Этот результат недавно был перенесен А. В. Казейкиной (при участии А. А. Шананина) на произвольную вогнутую функцию [144]. Для случая, когда $Q(\rho)$ — выпуклая функция, глобальная сходимость к волне разрежения была установлена Р. Дуанем и Х. Чжао (2007) [145].

5) Интересно также заметить, что все сказанное выше относительно модели Уизема в точности переносится и на модель Ньюэлла: та же структура асимптотики, те же самые скорости бегущих волн, та же асимптотика у *сдвигов фаз* $d_k(t) = O(\ln t)$; подробности см. в [100–110]. Для модели Пейна получено (см., например, [7]) условие существования автомодельного решения вида «ударно-бегущей волны». Оказывается, что скорость такой волны будет удовлетворять формуле RRH.

6) Результаты о сходимости к бегущей волне и волне разрежения решения начальной задачи Коши (20), (15) могут быть «перенесены» на начально-краевую задачу для уравнения (20) (подробности см., например, в [146, 147]).

**Упражнение**** (см. п. 2.1.3). При каких условиях уравнение (20) линеаризуется? Попробуйте построить асимптотические ряды для $d_k(t)$ в этих (частных) случаях, подобно тому, как это было сделано в работе [14] для уравнения Бюргерса (в случае общих начальных условий см. [106]); см. также [7].

[1)] Несложно показать (см. [93]), рассмотрев случай сходимости к одной волне разрежения, что имеет место сходимость не быстрее $\sim 1/\sqrt{t}$. Однако пока в общем случае (при отсутствии прилипания) не доказано (и не опровергнуто), что скорость сходимости в теореме 2 оценивается как $O(1/\sqrt{t})$.

Приведем в заключение этого пункта схему доказательства теорем 1, 2 и аналогичного утверждения для модели Ньюэлла. Доказательство основывается на следствиях из *принципа максимума* для линейных параболических уравнений (с коэффициентами, зависящими от $(t,\,x)$) [39,40,139]: на *принципе сравнения*

$$\rho^1(0,\,x)\leqslant\rho^2(0,\,x)\implies\rho^1(t,\,x)\leqslant\rho^2(t,\,x),\quad t\geqslant 0,$$

и на *принципе сравнения на фазовой плоскости*

$$\{\rho^1(0,\,x)<\rho^2(0,\,x)\implies\rho^1(0,\,x')\leqslant\rho^2(0,\,x'),\ \ x'\geqslant x\}\implies$$
$$\implies\{\rho^1(t,\,x)<\rho^2(t,\,x)\implies\rho^1(t,\,x')\leqslant\rho^2(t,\,x'),\ \ x'\geqslant x\},\quad t\geqslant 0.$$

Эти принципы переносятся и на нелинейные параболические уравнения, а также на некоторые их дифференциально-разностные и просто разностные аналоги. Идея такого перенесения достаточно простая и широко используется в теории параболических уравнений [39, 40], например, для оценки старших производных неизвестной функции. Идея состоит в том, что нелинейные коэффициенты при частных производных объявляются некоторыми функциями независимых переменных, в нашем случае $t$ и $x$. Далее делаются «априорные» предположения относительно неизвестной функции — как правило, о ее равномерной ограниченности или(и) о существования у нее (равномерно ограниченных) производных. В этих предположениях коэффициенты при частных производных оказываются «настолько хорошими» функциями, что принципы сравнения применимы к возникшему уравнению, в котором нелинейные коэффициенты, зависящие от неизвестной функции, интерпретируются просто как некоторые функции независимых переменных. Затем, уже с помощью этих принципов, проверяется, что сделанные априорно предположения выполняются. В этой связи также заметим, что иногда априорные предположения выбирают единственное решение из множества возможных. Так, для обычного уравнения теплопроводности априорное предположение о том, что «решение начальной задачи Коши (с равномерно ограниченной начальной функцией) будем искать в классе равномерно ограниченных функций», приводит начальную задачу Коши в класс корректных (по Адамару), для которых существует и притом единственное решение, устойчивое по начальным данным. Но, как показывает пример А. Н. Тихонова, если не накладывать ограничение на рост решения с увеличением времени, то построенное решение уже не будет единственным [150].

Приведенные принципы сравнения также переносятся и на некоторые уравнения (скажем, на закон сохранения), которые получаются путем предельного перехода (например, с помощью метода исчезающей вязкости) из нелинейных параболических уравнений.



В доказательстве часто в качестве сравниваемых функций выбираются: решение рассматриваемой задачи и специальным образом подобранное автомодельное решение (асимптотически автомодельное решение) либо специальным образом «склеенная» функция из таких решений. При этом значения на $x=\pm\infty$, а также ординаты «склеек», в случае если таковые имелись, у таких автомодельных решений выбираются из множества точек

$$\{\alpha_0,\,\beta_0,\,\alpha_1,\,\ldots,\,\beta_{n-1},\,\alpha_n,\,\beta_n\}$$

либо из малых окрестностей этих точек, а сдвиги фаз подбираются так, чтобы в начальный момент времени выполнялись условия используемого варианта принципа сравнения.

Заметим, что основные идеи описанного подхода к доказательству теорем 1, 2 применялись ранее для исследования асимптотики (по времени) решения начальной задачи Коши для уравнения теплопроводности с нелинейным источником (такого рода уравнения возникают, например, при описании распространения генных волн и пламени). Так, еще в 1937 г. в пионерской работе А. Н. Колмогорова, И. Г. Петровского, Н. С. Пискунова [140] была исследована сходимость к одной бегущей волне. Современное состояние дел отражено в работах [139, 141] и цитированной там литературе.

Следуя монографии [139], можно называть сходимость в теореме 2 *сходимостью по форме*. А согласно терминологии работы [136], системы волн, возникающие в теоремах 1, 2, следует называть *промежуточными асимптотиками*. Заметим (см. теорему 2), что промежуточная асимптотика (система волн) в общем случае сама не обязана являться решением уравнения.

### 2.3.2. Эволюции локального затора в транспортном потоке, описываемом моделями LWR и Уизема

Предположим, что в однородном стационарном транспортном потоке (одна бесконечная полоса без въездов и выездов) образовалось локальное увеличение плотности (локальный затор). Проследим (не строго!) за эволюцией этого «бугорка» — за тем, как будет рассасываться это локальное увеличение плотности. Для этого введем $h(t)$ АТС/км — «высоту» бугорка и $l(t)$ км — «длину» бугорка в момент времени $t$ (см. рис. 21). Поскольку площадь бугорка сохраняется (что как раз и отражает закон сохранения количества АТС), то

$$h(t)l(t)\approx 2S_0,$$

где $S_0$ — площадь (размерность площади — АТС) под бугорком. Здесь стоит приближенное равенство из-за того, что бугорок, вообще говоря, не будет в точности иметь вид прямоугольного треугольника (гипотенуза, в общем случае, будет кривой линией и лишь на больших временах достаточно

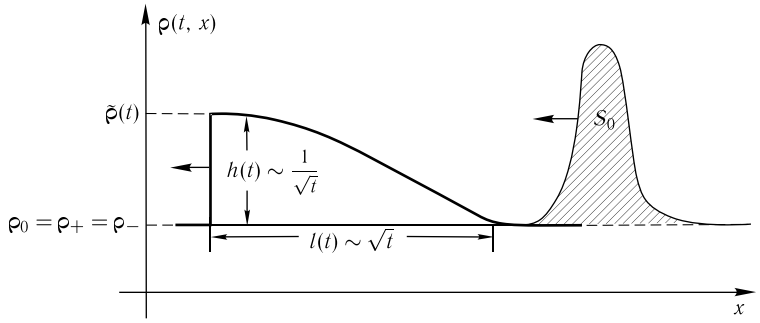



Рис. 21. Эволюция локального затора

хорошо приближается прямой). Сам факт превращения с ростом времени произвольного бугорка в криволинейный прямоугольный треугольник несложно устанавливается с помощью метода характеристик (так же, как на рис. 7, считаем, что $Q''(\rho_0) < 0$).

Перейдем теперь в систему координат, движущуюся со скоростью

$$c_0 = Q'(\rho_0),$$

где $\rho_0$ — значение плотности вдали от бугорка. В этой системе координат «скорость разбегания бугорка» можно посчитать по формуле RRH:

$$l'(t) = \Big| \frac{Q(\tilde\rho(t)) - Q(\rho_0)}{\tilde\rho(t) - \rho_0} - c_0 \Big| \approx \frac{1}{2}|Q''(\rho_0)|h(t),$$

где

$$\tilde\rho(t) = \rho_0 + h(t)$$

— максимальное значение плотности в момент времени $t$. Таким образом,

$$l'(t) \sim \frac{|Q''(\rho_0)|S_0}{l(t)} \Longrightarrow l(t) \sim \sqrt{2|Q''(\rho_0)|S_0}\cdot\sqrt{t} \Longrightarrow h(t) \sim \sqrt{\frac{2S_0}{|Q''(\rho_0)|}}\frac{1}{\sqrt{t}}.$$

Сделанные качественные выводы неплохо согласуются с практикой.

Приведенные выше рассуждения являются частным случаем более общего и более точного утверждения об асимптотике вида $N$-волны (см. работы П. Лакса, Т.-П. Лю, Р. Ди Перна, К. Дафермоса, С. Н. Кружкова и Н. С. Петросян [2, 7, 11, 37, 96, 99]).

**Упражнение** [37, 44, 99]. Пусть, как и в приведенных выше рассуждениях,

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

Используя формулу Лакса—Олейник (см. п. 2.1.3), строго обоснуйте приведенный выше результат. Выкладки будут особенно простыми, если дополнительно известно, что зависимость $Q(\rho)$ — параболическая.



Для того чтобы описать поведение локального затора согласно модели Уизема, умножим уравнение (20) на

$$\rho(t, x) - \rho_0 \in L_1(\mathbb{R}) \cap L_\infty(\mathbb{R}_x)$$

и проинтегрируем (по частям) [1] по $x$ от $-\infty$ до $\infty$:

$$\frac{d}{dt}\int\limits_{-\infty}^{\infty}\frac{1}{2}(\rho(t, x) - \rho_0)^2\,dx = -\int\limits_{-\infty}^{\infty}\Big(D(\rho(t, x))\frac{\partial\rho(t, x)}{\partial x}\Big)^2 dx \leqslant$$

$$\leqslant -D_{\min}\int\limits_{-\infty}^{\infty}\Big(\frac{\partial\rho(t, x)}{\partial x}\Big)^2 dx < 0,$$

где

$$D_{\min} = \min_{\rho\in[0,\rho_{\max}]} D^2(\rho) > 0.$$

Полученное неравенство, которое обычно называют *энергетическим неравенством* [19, 110], отражает тот факт, что процесс эволюции локального затора происходит с диссипацией «энергии». Иначе говоря, «энергия»

$$V(\rho(t, \cdot)) = \int\limits_{-\infty}^{\infty}\frac{1}{2}(\rho(t, x) - \rho_0)^2\,dx$$

есть *функционал Ляпунова* [138] для уравнения (20) на многообразии таких решений $\rho(t, x)$, что $\rho(t, x) - \rho_0 \in L_1(\mathbb{R}) \cap L_\infty(\mathbb{R}_x)$. Действительно,

$$\frac{d}{dt}V(\rho(t, \cdot)) < 0 \ \text{ при } \ \rho(t, x) \overset{L_2(\mathbb{R}_x)}{\neq} \rho_0 \ \text{ и } \ \frac{d}{dt}V(\rho_0) = 0;$$

$$V(\rho(t, \cdot)) > 0 \ \text{ при } \ \rho(t, x) \overset{L_2(\mathbb{R}_x)}{\neq} \rho_0 \ \text{ и } \ V(\rho_0) = 0.$$

Кроме того, из энергетического неравенства и из неравенства Харди—Литлвуда—Полиа, которое представляет собой частный случай неравенства колмогоровского типа для производных с $n = 2$, $k = 1$, $p = q = r = 2$ (см. замечание 1 ниже), следует существование такой константы $\tilde K > 0$, что

$$\frac{dV(\rho(t, \cdot))}{dt} \leqslant -\tilde K V(\rho(t, \cdot))^2 \Longrightarrow \|\rho(t, \cdot) - \rho_0\|_{L_2(\mathbb{R})} = \sqrt{2V(\rho(t, \cdot))} = O\Big(\frac{1}{\sqrt{t}}\Big).$$

[1] Заметим, что при интегрировании по частям мы пользовались различными следствиями из принципа максимума (см. п. 2.1.3, 2.3.1), в частности, следующими двумя:

а) $\lim\limits_{x\to\pm\infty}\rho(0, x) = \rho_\pm \Longrightarrow \forall t \geqslant 0 \ \lim\limits_{x\to\pm\infty}\rho(t, x) = \rho_\pm$;

б) частные производные $\rho(t, x)$, входящие в уравнение (20), равномерно ограничены.

Причем последнее свойство, как правило, используется вместе с неравенствами колмогоровского типа для производных (см. замечание 1).

Немного более аккуратные рассуждения (см, например, [110, 113]) позволяют установить также сходимость в среднем [1)] в $L_1(\mathbb{R})$ и равномерную сходимость в $L_\infty(\mathbb{R})$, причем так же, как и для модели LWR:

$$\|\rho(t,\,\cdot) - \rho_0\|_{L_\infty(\mathbb{R})} = O\Big(\frac{1}{\sqrt{t}}\Big).$$

**Замечание 1.** Неравенствами колмогоровского типа для производных будем называть неравенства следующего вида:

$$\exists K > 0\colon \quad \forall z(x) \in W^n_{p,r}(\mathbb{R}) \quad \|z^{(k)}(\cdot)\|_{L_q(\mathbb{R})} \leqslant K\|z(\cdot)\|^{\alpha}_{L_p(\mathbb{R})}\|z^{(n)}(\cdot)\|^{\beta}_{L_r(\mathbb{R})},$$

где $0 \leqslant k < n$ — целые числа, $0 < p,\, q,\, r \leqslant \infty$, $\alpha, \beta \geqslant 0$, $W^n_{p,r}(\mathbb{R})$ — пространство таких функций $z(x) \in L_p(\mathbb{R})$, у которых $(n-1)$-я производная локально абсолютно непрерывна на $\mathbb{R}$ и $n$-я производная $z^{(n)}(x) \in L_r(\mathbb{R})$. Константа $K$ зависит от пяти параметров $n,\, k,\, p,\, q,\, r$. Величины $\alpha$ и $\beta$ однозначно ими определяются:

$$\alpha = \frac{n - k - \frac{1}{r} + \frac{1}{q}}{n - \frac{1}{r} + \frac{1}{p}}, \quad \beta = 1 - \alpha.$$

В работах В. Н. Габушина [151, 152] было показано, что если $q \neq p$ при $k = 0$, то

$$K(n,\, k,\, p,\, q,\, r) < \infty \iff \frac{n-k}{p} + \frac{k}{r} \geqslant \frac{n}{q}, \quad r \geqslant 1.$$

Более подробно о тех случаях, в которых удалось вычислить наилучшие (точные) значения константы $K$, написано, например, в работах [47, 153].

**Замечание 2.** Неравенства колмогоровского типа для производных возникали в исследованиях А. Н. Колмогорова конца 30-х годов XX века. Для отыскания неулучшаемых значений $K$ приходилось решать, например, задачи вариационного исчисления, оптимального управления, в том числе и с фазовыми ограничениями, в то время как из приложений (в основном военных и связанных с движением ракет) задачи оптимального управления активно стали приходить в 40-е годы прошлого века. Общие же способы решения задач оптимального управления (без смешанных ограничений) появились только в середине 1950-х годов в школах Л. С. Понтрягина (принцип максимума Понтрягина) и Р. Беллмана (принцип динамического программирования).

В цикле работ, начавшихся в 1960-е годы, А. Я. Дубовицкий и А. А. Милютин предложили наиболее общую из известных на данный момент схему получения необходимых условий экстремума [41, 49], которая в задачах оптимального управления, в том числе и со смешанными ограничениями, позволяет получить принцип максимума (не путать с принципами максимума для решений уравнений параболического типа). В основе схемы лежит простая идея: пересечение конуса направлений убывания, если задача на минимум, и возрастания, если на максимум,

[1)] Из сходимости в среднем следует, что площадь бугорка (см. рис. 21) со временем уменьшается и стремится к нулю.



функционала с конусами возможных (допустимых) направлений ограничений должно быть пустым в точке оптимума. В предположении выпуклости рассматриваемых конусов это условие представляется как *уравнение Эйлера—Лагранжа*: существует нетривиальный набор элементов из сопряженных конусов, сумма которых равняется нулю. Другой подход (см., например, [47]) состоит в использовании «гладко-выпуклой» структуры задачи. При этом для «борьбы» с гладкой частью задачи, т. е. для получения необходимых условий локальной оптимальности, используется принцип Ферма (оптимум следует искать среди экстремальных значений функционала), а для «борьбы» с выпуклой частью — *принцип Лагранжа*, т. е. возможность «перенесения» ограничений (необязательно всех) задачи оптимизации в функционал с должным образом определенными множителями Лагранжа. Для того чтобы была понятна связь принципа Лагранжа с выпуклостью задачи, заметим, что этот принцип есть, по сути, теорема об отделимости — одно из следствий теоремы Хана—Банаха: в вещественном векторном пространстве граничные точки «телесного» (имеющего внутренние точки) выпуклого множества отделимы гиперплоскостью от самого множества, т. е. найдется такая гиперплоскость, проходящая через выбранную граничную точку, что рассматриваемое выпуклое множество будет «лежать» в одном из двух полупространств, на которые гиперплоскость делит пространство. Множители Лагранжа как раз и задают уравнение этой гиперплоскости. При этом выпуклое множество строится по самой задаче — функционалу и каждому ограничению, которое хотим «внести» в функционал, соответствуют свои «компоненты», — а точка, которую нужно отделять, соответствует решению задачи оптимизации. Как следствие, множители Лагранжа являются элементами (алгебраически) сопряженных пространств к пространствам, в которых «живут» ограничения.

Элементы сопряженного пространства довольно трудно описать для некоторых важных пространств, например таких, как пространство измеримых ограниченных функций. Это является, пожалуй, одним из основных сдерживающих факторов в развитии необходимых условий оптимальности. Обратим внимание на то, что для получения принципов максимума в задачах оптимального управления со смешанными ограничениями очень сложно (если вообще возможно) напрямую использовать описанный выше гладко-выпуклый формализм. В основном при получении необходимых условий оптимальности в таких задачах удобно использовать схему Дубовицкого—Милютина.

В п. 2.1.3 обобщенные энтропийные решения уравнения LWR получались как пределы решений соответствующих уравнений Уизема. Поэтому можно ожидать, что и на предельных решениях уравнений Уизема энергия также убывает. И действительно, если предельное (энтропийное) решение имеет разрывы (и только в этом случае), то на каждом разрыве (ударной волне) происходит потеря энергии в соответствии с формулой:

$$\frac{d}{dt}V(\rho(t,\,\cdot)) = \frac{Q(\rho_-) + Q(\rho_+)}{2}(\rho_+ - \rho_-) - \int\limits_{\rho_-}^{\rho_+} Q(\rho(t,\,x))\,dx < 0;$$

здесь для определенности $\rho_- < \rho_+$. Этот интересный и довольно простой факт доказан, например, в пособии [19].

Таким образом, мы получили еще одно объяснение необратимости по времени процесса, описываемого моделью LWR (при наличии ударных волн в решении), как свойства, «унаследованного» от модели более высокого уровня — модели Уизема.

**Замечание 3** (С. Н. Кружков, Е. Ю. Панов). Пусть $\Phi(\rho)$ — произвольная дважды гладкая выпуклая функция, $f(t,x) \geqslant 0$ — произвольная дважды гладкая финитная в полуплоскости $\pi = \{(t,x)\colon t > 0,\ x \in \mathbb{R}\}$ пробная функция. Умножим теперь уравнение (20) на $\Phi'(\rho(t,x))f(t,x)$ и проинтегрируем в $\pi$, перебрасывая производные на пробную функцию; для упрощения выкладок будем считать $D(\rho) \equiv 1$ (в общем случае см., например, [20,37]):

$$0 \leqslant \varepsilon \iint\limits_{\pi} \Phi''(\rho(t,x))\rho_x(t,x)^2 f(t,x)\,dt\,dx =$$
$$= \iint\limits_{\pi} \left\{ f_t(t,x) \int\limits_k^{\rho} \Phi'(\tilde\rho)\,d\tilde\rho + f_x(t,x) \int\limits_k^{\rho} \Phi'(\tilde\rho)Q'(\tilde\rho)\,d\tilde\rho + \varepsilon f_{xx}(t,x)\Phi(\rho(t,x)) \right\} dt\,dx,$$

где $k$ — произвольное действительное число. Устремив $\varepsilon \to 0+$, получим

$$\iint\limits_{\pi} \left\{ f_t(t,x) \int\limits_k^{\rho} \Phi'(\tilde\rho)\,d\tilde\rho + f_x(t,x) \int\limits_k^{\rho} \Phi'(\tilde\rho)Q'(\tilde\rho)\,d\tilde\rho \right\} dt\,dx \geqslant 0,$$

которое справедливо для любого $k$. Заметим, что в случае, когда $\Phi(\rho)$ — линейная функция, вместо неравенства можно писать равенство. Введем обозначения:

$$\eta(\rho) = \int\limits_k^{\rho} \Phi'(\tilde\rho)\,d\tilde\rho \quad \text{— энтропия,}$$
$$q(\rho) = \int\limits_k^{\rho} \Phi'(\tilde\rho)Q'(\tilde\rho)\,d\tilde\rho \quad \text{— поток энтропии.}$$

Полученное неравенство можно переписать как *энтропийное неравенство* (перебрасывая обратно производные с пробной функции) [31,37]:

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

которое следует понимать в слабом смысле (причем с неотрицательной финитной пробной функцией). Энтропийные неравенства для уравнений газовой динамики, по-видимому, впервые рассматривал Э. Жуге в начале XX века (см., например, [31, 56]). Напомним (см. примеры О. А. Олейник и И. М. Гельфанда из п. 2.1.1), что начальная задача Коши для уравнения (13), понимаемого в слабом смысле, имеет, вообще говоря, не единственное решение и что закон сохранения (13) и тот же закон (13), умноженный, скажем, на $\Phi'(\rho) > 0$, вообще говоря, имеют разные



решения. Однако было подмечено [154], что если к закону сохранения (13) добавить энтропийное неравенство [1)], то оно может отобрать единственное решение. В конце 1960-х С. Н. Кружков с помощью энтропийных неравенств построил (по сути, используя метод исчезающей вязкости, см. п. 2.1.3), вполне законченную теорию обобщенных решений начальной задачи Коши для закона сохранения (13) [29]. [2)] В начале 1970-х годов исследованием энтропийных неравенств (для систем уравнений) занимался П. Лакс [11,31,37].

### 2.3.3. Задача о светофоре: при каких условиях перед светофором не будет скапливаться очередь

В 1955 г. в работе [5] (см. также [7,157]) М. Лайтхиллом и Дж. Уиземом была поставлена и решена (на основе модели LWR) следующая задача:

*Найти такое число $k > 0$, что перед светофором (работающим в двух режимах: зеленый и красный) не будет скапливаться очередь, если*

$$\frac{T_{\text{зел}}}{T_{\text{кр}}} \geqslant k.$$

*Считать, что транспортный поток вдали от светофора имеет плотность $\bar\rho < \rho_m$ (значение потока $\bar q$), где $\rho_m$ — плотность, при которой значение потока максимально.*

Пусть загорелся красный цвет (рис. 22). Тогда от светофора навстречу транспортному потоку пойдет ударная волна со скоростью [3)]

$$c = \left| \frac{Q(\rho_{\max}) - Q(\bar\rho)}{\rho_{\max} - \bar\rho} \right| = \frac{\bar q}{\rho_{\max} - \bar\rho}.$$

[1)] Функция, удовлетворяющая (в слабом смысле) уравнению (13) и энтропийному неравенству, полученному исходя из функции $\Phi(\rho)$, называется $\Phi$-*решением*.

[2)] Основная идея заключалась в следующем. Назовем энтропийным решением функцию, которая является $\Phi$-решением, для любой дважды гладкой выпуклой функции $\Phi(\rho)$. По построению, решение, полученное с помощью метода исчезающей вязкости, необходимо является энтропийным решением. Причем в качестве функций $\Phi(\rho)$ можно брать лишь всевозможные линейные функции и функции вида $\{|\rho - k|\}_{k\in\mathbb{R}}$, поскольку любая дважды гладкая выпуклая функция раскладывается по этому «базису». Оказывается, что энтропийное решение всегда единственно. Для того чтобы это понять, заметим, что энтропийное решение в точках гладкости удовлетворяет соотношению (13) в классическом смысле, а в точках разрыва удовлетворяет RRH-условию и E-условию (простое доказательство этих фактов имеется, например, в пособии [19]). В конце 1980-х, в контексте вышенаписанного, С. Н. Кружковым был поставлен вопрос [155]: когда $\Phi$-решение единственно? Ответ на этот вопрос был получен Е. Ю. Пановым в работе [156]. Оказывается, что в случае, когда $\Phi(\rho)$ — строго вогнутая (выпуклая) функция, $\Phi$-решение единственно ($\Phi(\rho)$ — произвольная дважды гладкая строго выпуклая функция). Если $\Phi(\rho)$ не является строго вогнутой (выпуклой) функцией, то $\Phi$-решение задачи Коши при подходящем выборе начальных данных не единственно [156].

[3)] Возникшую краевую задачу удобно понимать как начальную, если договориться считать, что за светофором плотность АТС максимальна и равна $\rho_{\max}$, т. е. за светофором движения нет.

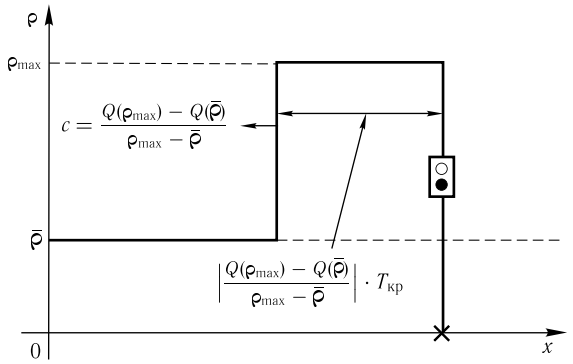



Рис. 22. Момент окончания горения красного цвета

«Излишек» АТС, скопившихся перед светофором за время горения красного цвета, равен

$$(\rho_{\max} - \bar{\rho})\frac{\bar{q}}{\rho_{\max} - \bar{\rho}} T_{\text{кр}} = \bar{q} T_{\text{кр}}.$$

Пусть теперь загорелся зеленый цвет (рис. 23, 24). Тогда до тех пор, пока весь излишек не пройдет через светофор, поток АТС через светофор будет максимальным и равным $q_m$. Это не очень очевидное утверждение может быть установлено с помощью решения соответствующей задачи о распаде произвольного разрыва (см. п. 2.3.1). Таким образом, перед светофором не будет скапливаться очередь, если

$$(q_m - \bar{q})T_{\text{зел}} \geqslant \bar{q} T_{\text{кр}} \implies k = \frac{\bar{q}}{q_m - \bar{q}}.$$

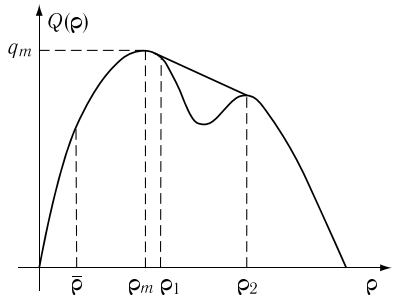



Рис. 23

Заметим, что полученное соотношение достаточно хорошо согласуется с интуитивными представлениями. Действительно, если принять, что когда



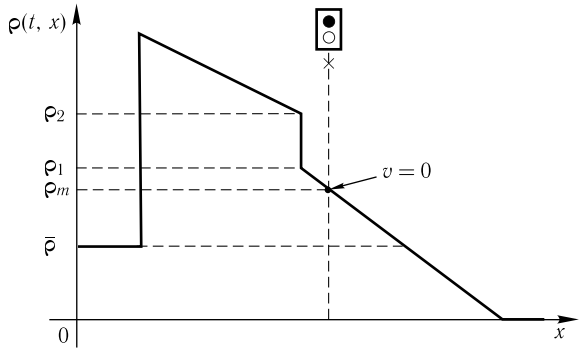



Рис. 24. Горит зеленый цвет

горит красный цвет, тогда поток АТС через светофор равен нулю, а когда горит зеленый, тогда поток максимальный (в течение всего времени горения зеленого), то получим условие: излишек АТС, скопившийся за время горения красного $\bar{q}T_{\text{кр}}$, должен быть не больше, чем та «добавка», которую получаем за время горения зеленого:

$$(q_m - \bar{q})T_{\text{зел}}.$$

«Добавка» же в свою очередь обусловлена тем, что при наличии светофора интенсивность потока АТС через светофор во время горения зеленого $q_m$ превышает интенсивность потока АТС, подъезжающих к светофору $\bar{q}$.

## 2.4. Теория Кернера—Конхойзера движущихся локальных кластеров в моделях класса «Дженерал Моторс»

В разделах 2.1–2.3 были рассмотрены в основном модели транспортного потока, в которых стационарные состояния (в стационарных состояниях АТС движутся с постоянной скоростью и плотностью) «отвечают» фундаментальной диаграмме транспортного потока. С точки зрения образования пространственно-временных структур плотного потока детальная классификация этих моделей была дана, например, в главе 10 книги [10]. Там же имеется и критическое сравнение этих модельных решений с фундаментальными эмпирическими свойствами перехода к плотному транспортному потоку [158]. В этом разделе мы рассматриваем нелинейное решение, которое возникает в результате неустойчивости в моделях класса «Дженерал Моторс» (ДМ). Чтобы понять термин *нелинейное решение*, нужно напомнить сначала смысл термина *неустойчивость* в применении к исходно

однородным состояниям транспортного потока в этих моделях. Неустойчивость означает нарастание во времени очень малых неоднородных возмущений. Окончательный результат этого нарастания приводит к структурам транспортного потока конечной и в некоторых случаях очень большой амплитуды. В последнем случае для нахождения и математического описания этих нелинейных решений уже нельзя пользоваться математическим аппаратом анализа неустойчивостей модели, который для множества моделей ДМ-класса детально рассмотрен в обзорах [60, 122]. Именно рассмотрению нелинейных пространственно-временных решений большой амплитуды, впервые найденных в 1994 г. Б. С. Кернером и П. Конхойзером [162, 163] в моделях ДМ-класса, и посвящен этот раздел [1)] данной главы. Однако прежде чем мы рассмотрим эти нелинейные решения в п. 2.4.2, необходимо коротко изложить фундаментальные эмпирические свойства перехода от свободного к плотному транспортному потоку и выяснить, могут ли модели транспортного потока, рассмотренные выше, показать эти эмпирические свойства. Более подробно об этом будет написано в главе 3.

### 2.4.1. Фундаментальные эмпирические свойства перехода от свободного транспортного потока к плотному и модели транспортного потока

Основные эмпирические свойства перехода к плотному транспортному потоку следующие [158]:

**1.** Переход к плотному транспортному потоку (traffic breakdown) является переходом F → S (буква F соответствует «free flow», т. е. свободному потоку, буква S обозначает фазу синхронизованного потока — «synchronized flow» в английской литературе).

**2.** Вероятность спонтанного перехода F → S является растущей функцией величины потока АТС.

**3.** Переход F → S может быть как спонтанным, так и индуцированным, т. е. вызванным внешним возмущением большой амплитуды, около одного и того же узкого места на дороге (bottleneck).

Как показано в главе 10 книги [10], с точки зрения перехода к плотному транспортному потоку многие модели разделов 2.1–2.3 можно разделить на два больших класса:

а) *Модели типа Лайтхилла—Уизема—Ричардса* (*LWR*), в которых переход к плотному транспортному потоку возникает не в результате неустойчивости, а за счет существования точки, в которой достигается максимум функции потока на фундаментальной диаграмме.

[1)] Б. С. Кернер предоставил для этого раздела и для главы 3 оригинальные рисунки из своих книг.



б) *Модели класса «Дженерал Моторс»* (*ДМ*), в которых переход к плотному потоку связан с неустойчивостью модельных решений начиная с некоторой критической плотности транспортного потока.

Модели LWR-типа (теория кинематических и ударных волн в транспортном потоке) не могут описывать пункты 2 и 3 фундаментальных эмпирических свойств перехода к плотному транспортному потоку. Теория, основанная на моделях типа ДМ, не может описывать никаких фундаментальных эмпирических свойств перехода к плотному транспортному потоку (пункты 1–3). Тем не менее эти модели представляют большой интерес для анализа. Одна из причин такого вывода рассматривается в этом разделе. Здесь показано, что модели ДМ-класса описывают характеристические параметры стационарного движения широкого движущегося кластера по дороге, наблюдаемого в эмпирических данных.

Неустойчивость однородного свободного транспортного потока в моделях ДМ-класса [1)] детально рассмотрена в огромном количестве работ, обзор которых можно найти в статьях Д. Кроудери с соавторами [122] и Д. Хельбинга [60]. Однако нарастание неоднородных возмущений в этих моделях, происходящее в результате упомянутой неустойчивости, приводило к волнам нереалистично большой амплитуды (в макроскопических моделях) или к столкновению АТС (в микроскопических моделях).

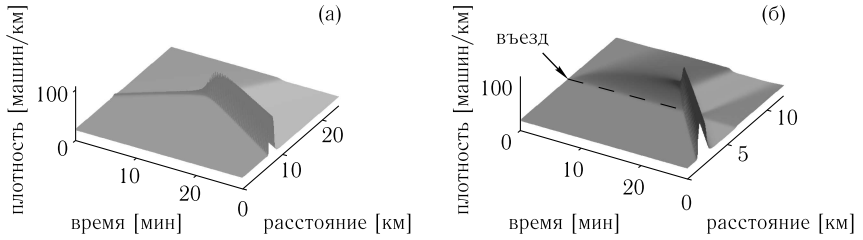



Рис. 25. «Эффект бумеранга» для нарастающего возмущения в свободном потоке, метастабильном (см. с. 173) относительно возникновения широких движущихся кластеров. Расчет на основе макроскопической модели класса Пейна (22), (23). (а) Однородная автодорога, (б) автодорога с въездом. Взято из [10]

Какие же типы структур плотного транспортного потока возникают в результате такой неустойчивости в этих и других моделях ДМ-класса? Ответ на этот вопрос был дан Б. С. Кернером и П. Конхойзером в 1993–1994 годах [162, 163] на основе численного расчета (рис. 25 и 26)

[1)] К моделям этого класса относятся, например, модель Пейна [55], оптимальной скорости Ньюэлла [91], Бандо [159], разумного водителя Трайбера [60, 118], клеточных автоматов Нагеля—Шрекенберга [121, 160], модели Видемана [161] и многие другие модели [10] (см. также раздел 2.1 и 2.2 этой главы).

и версии модели Пейна (см. п. 2.1.4):

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0, \tag{22}$$

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{V(\rho) - v}{\tau} - \frac{c_0^2}{\rho}\frac{\partial \rho}{\partial x} + \frac{\mu}{\rho}\frac{\partial^2 v}{\partial x^2}, \tag{23}$$

где $\tau$, $c_0$ и $\mu$ являются константами. По сути, по сравнению с обычной моделью Пейна произошло всего одно изменение: добавление диффузионного слагаемого $\frac{\mu}{\rho}\frac{\partial^2 v}{\partial x^2}$ в правую часть уравнения (23).

Чтобы найти реалистичные решения большой амплитуды, возникающие в результате неустойчивости, Б. С. Кернер и П. Конхойзер использовали при расчетах в модели типа Пейна специальную форму фундаментальной диаграммы (рис. 26 а). Особенность этой фундаментальной диаграммы состоит в том, что, начиная уже с не очень больших плотностей, скорость АТС на фундаментальной диаграмме экспоненциально стремилась к нулю. Это искусственная форма фундаментальной диаграммы, с помощью которой удалось в рамках макроскопической модели смоделировать задержку водителей $\tau_{\mathrm{del,jam}}^{(a)}$ при их ускорении один за другим из состояния с нулевой скоростью на заднем фронте широкого движущегося кластера (см. п. 2.4.3). Эта задержка отличается от задержки реакции водителя в других ситуациях, которая дается величиной $\tau$ в формуле (23). Идея Кернера—Конхойзера моделирования задержки при ускорении из состояния с нулевой скоростью с помощью формы фундаментальной диаграммы математически равносильна правилу slow-to-start в микроскопическом моделировании (см. работы группы М. Шрекенберга в 1998 г. [160]).

В статьях [162, 163] было получено, что в результате неустойчивости в моделях ДМ-класса образуется широкий движущийся кластер большой амплитуды (локальный движущийся затор), внутри которого плотность АТС высока, а скорость их движения близка к нулю, в то время как впереди и позади такого кластера существует свободный поток малой плотности. В англоязычной литературе для обозначения такого рода неоднородных локальных состояний большой плотности используется название «wide moving jam» (в дальнейшем широкий движущийся кластер большой амплитуды будет обозначаться буквой J). Могут возникать как отдельные движущиеся кластеры, так и последовательности таких кластеров (рис. 25 и 26).

Фазовый переход в моделях с неустойчивостью однородного состояния, по-видимому, является фазовым переходом первого рода от свободного однородного потока к широкому движущемуся кластеру (jam) и обозначается как переход F → J (напомним, что буква F соответствует «free flow», т. е. свободному потоку, а буква J соответствует «wide moving jam»).



Последующие исследования показали (см. ссылки в обзоре Д. Хельбинга [60]), что этот результат является общим для всех моделей ДМ-класса, рассмотренных выше в п. 2.1.3, 2.1.4, разделе 2.2.

Позже Б. С. Кернеру стало ясно, что этот результат,[1)] относящийся ко всем моделям ДМ-класса, противоречит фундаментальным эмпирическим свойствам, перечисленным в пунктах 1–3 выше (см. также главу 3 о теории трех фаз Кернера). Как уже было указано в пункте 1, переход от свободного к плотному транспортному потоку в эмпирических данных связан не с переходом F → J, а с переходом F → S (напомним, что буква S обозначает фазу синхронизированного потока, в английской литературе «synchronized flow»).

Хотя неустойчивость свободного потока в моделях ДМ-класса неправильно описывает переход от свободного к плотному потоку в реальном транспортном потоке, тем не менее результат нелинейного решения моделей ДМ-класса [163] (рис. 25 и 26) — стационарное движение широкого кластера по дороге — полностью соответствует эмпирическим данным и, следовательно, является важным результатом этих моделей, который остается также и в теории трех фаз Кернера (см. главу 3). Как объяснено выше, под результатом нелинейного решения здесь понимается конечное пространственно-временное распределение параметров потока, возникающее в результате неустойчивости исходно однородного решения модели транспортного потока.

Стационарное движение кластеров обладает определенными нелинейными свойствами, которые Б. С. Кернер и П. Конхойзер назвали характеристическими свойствами движения широких кластеров [163]. В дальнейшем эти характеристические свойства движения широкого кластера были подтверждены при исследовании всех других моделей класса ДМ [60]. Эти характеристические свойства движения широкого кластера рассмотрены в следующем пункте.

### 2.4.2. Характеристические параметры широкого движущегося кластера

Свойства стационарного движения широких кластеров состоят в следующем [163]: существуют характеристические параметры широкого движущегося кластера, которые при заданных внешних условиях движения транспорта (погода, день недели, процент грузовых АТС и т. п.) не зависят от параметров потока впереди широкого кластера и остаются неизменными в процессе движения кластера по дороге. Характеристические параметры являются одинаковыми для различных широких движущихся кластеров. Такими характеристическими параметрами широкого движущегося кластера являются:

[1)] Состоящий в том, что переход от свободного к плотному потоку связан с переходом F → J.

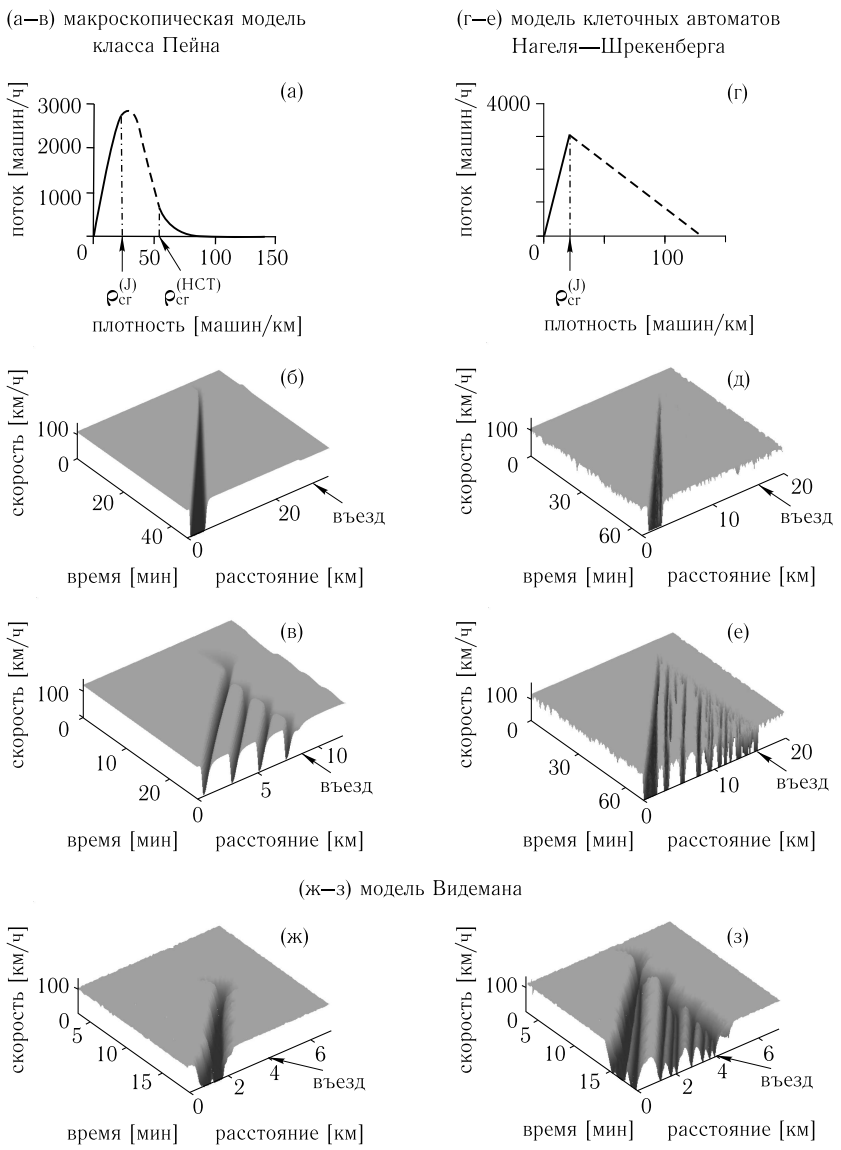



Рис. 26. Возникновение плотного потока на автодороге с въездом в моделях ДМ-класса, показывающих неустойчивость свободного потока при повышении плотности: возникновение широких движущихся кластеров в макроскопической модели (10), (11) класса Пейна (а—в), в модели клеточных автоматов Нагеля—Шрекенберга (г–е), и в модели Видемана (ж–з). (а, г) Фундаментальная диаграмма в макроскопической модели класса Пейна (а) и в модели Нагеля—Шрекенберга (г); пунктирная часть диаграмм отвечает неустойчивым состояниям. (б, в, д–з) Средняя скорость АТС в пространстве и времени. Взято из [10]



1) Средняя скорость движения заднего (по направлению потока) фронта движущегося кластера, обозначаемая как $v_g$.

2) Величина потока $q_{out}$, плотность $\rho_{min}$ и средняя скорость АТС $v_{max}$ в выходном потоке из широкого движущегося кластера. Эти величины являются характеристическими параметрами только при условии, что выходной поток из широкого движущегося кластера отвечает свободному потоку.

3) Средняя плотность АТС внутри широкого движущегося кластера обозначается как $\rho_{max}$. Здесь и далее $\rho_{max}$ совсем необязательно совпадает с максимально возможной плотностью («бампер к бамперу»).

Характеристические параметры широкого движущегося кластера качественно проиллюстрированы на рис. 27, на котором для заданного момента времени приведены распределения скорости, потока и плотности вдоль дороги, связанные с распространением широкого движущегося кластера в исходно однородном транспортном потоке. Поток $q_h$ и плотность $\rho_h$ в исходно однородном свободном потоке выбраны больше, а соответственно скорость $v_h$ меньше, чем соответствующие характеристические параметры в выходном потоке широкого движущегося кластера: $q_h > q_{out}$, $\rho_h > \rho_{min}$, $v_h < v_{max}$. Ясно, что впереди от широкого движущегося кластера остается исходный однородный транспортный поток. Однако этого не происходит позади широкого движущегося кластера из-за того, что по мере движения широкого кластера АТС, покидающие задний (в направлении движения) фронт кластера, формируют новый свободный поток с величиной потока $q_{out}$, плотностью $\rho_{min}$ и средней скоростью $v_{max}$.

Чтобы пояснить термин *выходной поток из широкого движущегося кластера*, давайте более детально посмотрим на рис. 27. Внутри широкого кластера скорость АТС равна нулю, а плотность равна $\rho_{max}$. На переднем фронте широкого кластера АТС должны резко тормозить вплоть до их остановки. На заднем фронте широкого кластера АТС ускоряются. В результате ускорения АТС из широкого кластера на его заднем фронте образуется выходной транспортный поток. Это объясняет термин *выходной поток* из широкого кластера, использованный выше. В случае, когда этот поток отвечает свободному потоку, он обозначен буквой $q_{out}$ и показан на рис. 27.

Кроме огромного количества численных исследований характеристического движения широких кластеров, опубликованных в многочисленных работах различных авторов, существует асимптотическая теория движущихся широких кластеров [164], основанная на математической теории сингулярных возмущений. Заинтересованный читатель может найти обзор численных исследований широких кластеров в обзоре [60], а асимптотическая теория широких кластеров подробно разбирается в оригинальной

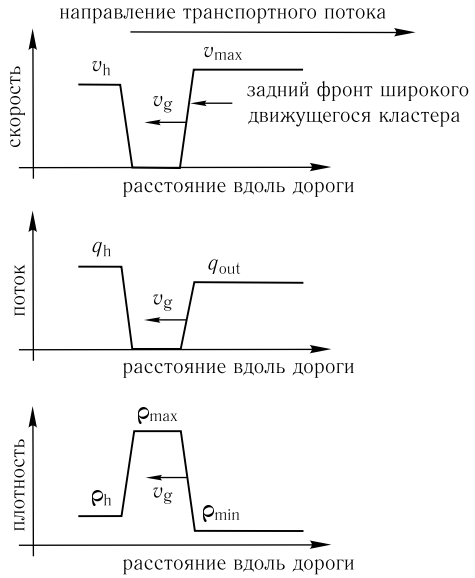



Рис. 27. Качественная иллюстрация характеристических параметров широкого движущегося кластера. Схематическое представление кластера в фиксированный момент времени. Пространственные распределения скорости АТС $v$, потока $q$ и плотности $\rho$ в широком движущемся кластере, который распространяется в исходно однородном свободном потоке, имеющем скорость $v_h$, величину потока $q_h$ и плотность $\rho_h$. Взято из [10]

статье [164]. В остальной части этого раздела будет рассмотрено одно из важнейших характеристических свойств движущегося широкого кластера — линия J.

### 2.4.3. Линия J Кернера

Характеристические параметры широкого движущегося кластера могут быть проиллюстрированы линией на плоскости поток–плотность (рис. 28). Эта линия использовалась Б. С. Кернером и была названа им *линией J* [10, 158].

Необходимо подчеркнуть, что линия J Кернера не имеет никакого отношения к линии для плотного потока на треугольной диаграмме Даганзо 1994 г. [85] (или же к любым другим фундаментальным диаграммам плотного транспортного потока). Действительно, наклон линии J Кернера определяется средней скоростью заднего фронта широкого кластера, а левая координата этой линии отвечает потоку, вытекающему из широкого кластера. Другими словами, линия J — это характеристика не плотного потока, а равномерного распространения заднего фронта широкого кластера. Эта линия фактически соединяет две точки на плоскости поток–плотность:



одну точку, отвечающую потоку и плотности внутри широкого движущегося кластера, и вторую точку, отвечающую потоку и плотности в выходном потоке после широкого кластера.

Поясним этот важный вопрос более подробно. Каждая точка на линии для плотного потока в треугольной фундаментальной диаграмме Даганзо [85] отвечает соотношению между плотностью и величиной потока в однородном плотном потоке. В то же время линия J описывает стационарное распространение заднего фронта широкого кластера, а не зависимость потока от плотности, которая отвечает фундаментальной диаграмме плотного потока. Чтобы понять это важное качество линии J и ее отличие от фундаментальной диаграммы плотного потока, можно дополнительно обратиться к рис. 9 в гл. 3. На этом рисунке можно видеть, что плотный поток в теории трех фаз Кернера отвечает не какой-то кривой или линии на плоскости поток–плотность, а двумерной области. При этом линия J разбивает эту двумерную область на две части. Таким образом, в теории трех фаз Кернера вместо прямой линии треугольной фундаментальной диаграммы Даганзо [85] для плотного потока постулируется двумерная область состояний, т. е. одному значению плотности отвечает не одно значение потока, как в диаграмме Даганзо, а бесконечное количество значений потока. В свою очередь линия J не описывает связь между потоком и плотностью в плотном потоке АТС, а как уже отмечалось, линия J — это характеристика равномерного распространения заднего фронта широкого кластера.

Чтобы объяснить линию J, рассмотрим среднюю скорость движения заднего фронта широкого кластера, где происходит ускорение АТС одного за другим. Каждое АТС, стоящее в кластере, может начать ускоряться на заднем фронте этого кластера, только если выполнены следующие условия:

- предыдущее АТС уже начало двигаться из кластера;
- в результате движения предыдущего АТС спустя некоторое время дистанция между двумя АТС превысила некоторое безопасное расстояние, которое удовлетворяет условию безопасного ускорения.

Таким образом, существует некоторая временная задержка в ускорении АТС на заднем фронте широкого движущегося кластера. Среднее время этой задержки обозначим как $\tau^{(a)}_{\mathrm{del,jam}}$. Согласно эмпирическим данным, $\tau^{(a)}_{\mathrm{del,jam}} \sim 1{,}5$–$2$ с. Движение заднего фронта широкого движущегося кластера связано с последовательным ускорением АТС, стоящих внутри кластера, на заднем фронте этого кластера. Поскольку среднее расстояние между АТС внутри кластера, включая среднюю длину АТС, равно $1/\rho_{\max}$, средняя скорость заднего фронта широкого движущегося кластера равна

$$v_g = -\frac{1}{\rho_{\max}\tau^{(a)}_{\mathrm{del,jam}}}. \qquad (24)$$

В эмпирических данных характеристическая средняя скорость заднего фронта широкого движущегося кластера по порядку величины дается формулой $v_g \approx -15$ км/ч. Наклон линии J равен характеристической скорости $v_g$. В случае, когда свободный транспортный поток формируется после кластера, характеристические величины потока $q_{out}$ и плотности $\rho_{min}$ определяют левую координату ($\rho_{min}$, $q_{out}$) линии J. Правая координата ($\rho_{max}$, 0) линии J отвечает плотности и потоку внутри широкого движущегося кластера; на рис. 28 предполагается, что средняя скорость и, следовательно, поток АТС внутри кластера равны нулю. В результате величина выходного потока $q_{out}$ из широкого движущегося кластера равна

$$q_{out} = |v_g|(\rho_{max} - \rho_{min}),$$

или, используя соотношение (24):

$$q_{out} = \frac{1}{\tau^{(a)}_{del,jam}}\left(1 - \frac{\rho_{min}}{\rho_{max}}\right).$$

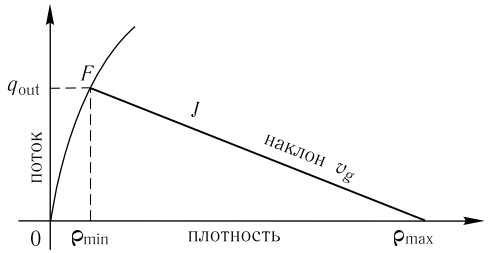



Рис. 28. Качественное представление фундаментальной диаграммы свободного потока (F) вместе с линией J Кернера, наклон которой равен средней скорости движения $v_g$ заднего фронта широкого движущегося кластера. Взято из [10]

Необходимо еще раз подчеркнуть, что, несмотря на важность моделей ДМ-класса, переход от свободного к плотному потоку в этих моделях связан с переходом к широким движущимся кластерам (переход F → J). Напротив, во всех реальных данных этот переход связан с переходом к синхронизованному потоку, т. е. с переходом F → S [158]. Последний вопрос более подробно разбирается в главе 3 о теории трех фаз Кернера.

## Литература


1. *Зельдович Я. Б.*, *Райзер Ю. П.* Физика ударных волн и высокотемпературных гидродинамических явлений. М.: Физматлит, 2008.
2. *Курант Г.*, *Фридрихс К.* Сверхзвуковое течение и ударные волны. М.: Издательство иностранной литературы, 1950.





3. *Крайко А. Н.* Краткий курс теоретической газовой динамики. М.: МФТИ, 2007.
4. *Гордин В. А.* Математика, компьютер, прогноз погоды и другие сценарии математической физики. М.: Физматлит, 2010.
5. *Lighthill M. J.*, *Whitham G. B.* On kinematic waves: II. Theory of traffic flow on long crowded roads // Proc. R. Soc. London, Ser. A. 1955. V. 229. P. 281–345.
6. *Richards P. I.* Shock Waves on the Highway // Oper. Res. 1956. V. 4. P. 42–51.
7. *Уизем Дж.* Линейные и нелинейные волны. М.: Мир, 1977.
8. Traffic flow theory: A state-of-the-art report / Editors N. H. Gartner, C. J. Messer, A. K. Rathi. Washington DC: Transportation Research Board, 2001.
9. *Луканин В. Н.*, *Буслаев А. П.*, *Трофимов Ю. В.*, *Яшина М. В.* Автотранспортные потоки и окружающая среда. М.: ИНФРА-М, Ч. 1, 2. 1998, 2001.
10. *Kerner B. S.* Introduction to modern traffic flow theory and control. The long road to three-phase traffic theory. Springer, 2009.
11. *Лакс П. Д.* Гиперболические дифференциальные уравнения в частных производных. М.–Ижевск: НИЦ «РХД», ИКИ, 2010.
12. *Ballou D. P.* Solution to nonlinear hyperbolic Cauchy problems without convexity condition // Trans. Amer. Math. Soc. 1970. V. 152, № 2. P. 441–460.
13. *Олейник О. А.* Разрывные решения нелинейных дифференциальных уравнений // УМН. 1957. Т. 12, № 3(75). С. 3–73.
14. *Hopf E.* The partial differential equation $u_t + uu_x = \mu u_{xx}$ // Comm. Pure Appl. Math. 1950. V. 3, № 3. P. 201–230.
15. *Олейник О. А.* Об одном классе разрывных решений квазилинейных уравнений первого порядка // Научные доклады высшей школы. Физико-математические науки. 1958. № 3. С. 91–98.
16. *Олейник О. А.* О единственности и устойчивости обобщенного решения задачи Коши для квазилинейного уравнения // УМН. 1959. Т. 14, № 2(86). С. 165–170.
17. *Гельфанд И. М.* Некоторые задачи теории квазилинейных уравнений // УМН. 1959. Т. 14, № 2(86). С. 87–158.
18. *Рождественский Б. Л.*, *Яненко Н. Н.* Системы квазилинейных уравнений и их приложения к газовой динамике. М.: Наука, 1978.
19. *Горицкий А. Ю.*, *Кружков С. Н.*, *Чечкин Г. А.* Уравнения с частными производными первого порядка: Учебное пособие. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 1999.
20. *Гасников А. В.* Сравнение определений обобщенного решения задачи Коши для квазилинейного уравнения. М.: ВЦ РАН, 2006.
21. *Иносэ Х.*, *Хамада Т.* Управление дорожным движением. М.: Транспорт, 1983.
22. *Бабков В. Ф.* Дорожные условия и безопасность дорожного движения. М.: Транспорт, 1982.
23. *Kumei S.*, *Bluman G. W.* When nonlinear differential equations are equivalent to linear differential equations // SIAM J. Appl. Math. 1982. V. 42, № 5. P. 1157–1173.

24. *Олвер П.* Приложения групп Ли к дифференциальным уравнениям. М.: Мир, 1989.
25. *Овсянников Л. В.* Групповой анализ дифференциальных уравнений. М.: Наука, 1993.
26. *Сидоров А. Ф.*, *Шапеев В. П.*, *Яненко Н. Н.* Метод дифференциальных связей и его приложения в газовой динамике. Новосибирск: Наука, 1984.
27. *Galaktionov V. A.*, *Svirshchevskii S. A.* Exact Solutions and Invariant Subspaces of Nonlinear Partial Differential Equations in Mechanics and Physics. Chapman & Hall/CRC applied mathematics and nonlinear science series; 10. 2007.
28. *Волосов К. А.*, *Вдовина Е. К.*, *Волосова А. К.* Новые точные решения уравнений с частными производными параболического типа: Учебное пособие. М.: МИИТ, 2010.
29. *Кружков С. Н.* Квазилинейные уравнения первого порядка со многими независимыми переменными // Матем. сб. 1970. Т. 81(123), № 2. С. 228–255.
30. *Кружков С. Н.* Нелинейные уравнения с частными производными (Лекции). Ч. 2. Уравнения первого порядка. М.: Изд-во МГУ, 1970.
31. *Serre D.* System of conservation laws: A challenge for the XXIst century // Mathematics Unlimited — 2001 and Beyond / B. Enquist, W. Schmid (Eds.). Berlin, New York: Springer-Verlag, 2001. P. 1061–1080.
32. *Лионс П.-Л.* (*Lions P.-L.*) О некоторых интригующих проблемах нелинейных уравнений в частных производных // Математика: границы и перспективы. М.: ФАЗИС, 2005. С. 193–211.
33. *Тупчиев В. А.* Обобщенные решения законов сохранения. М.: Наука, 2006.
34. *Эванс Л. К.* Методы слабой сходимости для нелинейных уравнений с частными производными. Новосибирск: Тамара Рожковская (Белая серия в математике и физике; Т. 2), 2006.
35. *Holden H.*, *Risebro N. H.* Front tracking for hyperbolic conservation laws. Springer, 2007.
36. Nonlinear conservation laws and applications. University of Minnesota, July 13–31, 2009; `http://www.ima.umn.edu/2008-2009/SP7.13-31.09/index.html#schedule`
37. *Dafermos C. M.* Hyperbolic conservation laws in continuum physics. Springer, 2010.
38. *Галкин В. А.* Анализ математических моделей: системы законов сохранения, уравнения Больцмана и Смолуховского. М.: БИНОМ. Лаборатория знаний, 2009.
39. *Ладыженская О. А.*, *Солонников В. А.*, *Уральцева Н. Н.* Линейные и квазилинейные уравнения параболического типа. М.: Наука, 1967.
40. *Крылов Н. В.* Лекции по эллиптическим и параболическим уравнениям в пространствах Гёльдера: Учебное пособие. Новосибирск: Научная книга (Университетская серия; Т. 2), 1998.
41. *Милютин А. А.*, *Дмитрук А. В.*, *Осмоловский Н. П.* Принцип максимума в оптимальном управлении. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2004; `http://www.milyutin.ru/papers.html`



42. Оптимальное управление / Под ред. Н. П. Осмоловского и В. М. Тихомирова. М.: МЦНМО, 2008.
43. *Красовский Н. Н.*, *Субботин А. И.* Позиционные дифференциальные игры. М.: Наука, 1974.
44. *Эванс Л. К.* Уравнения с частными производными. Новосибирск: Рожковская (Университетская серия; Т. 7), 2003.
45. *Демьянов В. Ф.* Минимакс, дифференцируемость по направлениям. Л.: Изд-во ЛГУ, 1974.
46. *Субботин А. И.* Обобщенные решения уравнений в частных производных. Перспективы динамической оптимизации. М.–Ижевск: НИЦ «РХД», ИКИ, 2003.
47. *Магарил-Ильяев Г. Г.*, *Тихомиров В. М.* Выпуклый анализ и его приложения. М.: УРСС, 2003.
48. *Пшеничный Б. Н.* Выпуклый анализ и экстремальные задачи. М.: Наука, 1980.
49. *Дубовицкий А. Я.*, *Милютин А. А.* Задачи на экстремум при наличии ограничений // ЖВМ и МФ. 1965. Т. 5, № 3. С. 395–453; `http://www.milyutin.ru/papers.html`
50. *Беллман Р.*, *Калаба Р.* Квазилинеаризация и нелинейные краевые задачи. М.: Мир, 1968.
51. *Пшеничный Б. Н.*, *Сагайдак М. И.* О дифференциальных играх с фиксированным временем // Кибернетика. 1970. Т. 2. С. 54–63.
52. *Maslov V. P.*, *Belavkin V. P.* Design of the optimal Dynamic Analyzer: Mathematical Aspects of Sound and Visual Pattern Recognition // Mathematical Aspects of Computer Engineering / Edited by V. P. Maslov, K. A. Volosov. M.: MIR, 1988. P. 146–237.
53. *Kolokoltsov V. N.*, *Maslov V. P.* Idempotent analysis and applications. Dordrecht: Kluwer Acad. Publ., 1997.
54. *Litvinov G. L.* Tropical mathematics, idempotent analysis, classical mechanics and geometry. AMS, Contemp. Math., 2010; `arXiv:1005.1247v1`. См. также: Семинар «Глобус», 2009. Вып. 4.
55. *Payne H. J.* Models of freeway traffic and control // Simulation Council Proc. 28, Mathematical Models of Public Systems / Edited by G. A. Bekey. 1971. V. 1. P. 51–61.
56. *Куликовский А. Г.*, *Погорелов Н. В.*, *Семенов А. Ю.* Математические вопросы численного решения гиперболических систем уравнений. М.: Физматлит, 2012.
57. *Петров И. Б.*, *Лобанов А. И.* Лекции по вычислительной математике. М.: Бином, 2006.
58. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.
59. *Чарахчьян А. А.* Об алгоритмах расчета распада разрыва для схемы С. К. Годунова // ЖВМ и МФ. 2000. Т. 40, № 5. С. 782–796.
60. *Helbing D.* Traffic and related self-driven many particle systems // Reviews of modern physics. 2001. V. 73, № 4. P. 1067–1141; `arXiv:cond-mat/0012229`

61. *Смирнов Н. Н.*, *Киселев А. Б.*, *Никитин В. Ф.*, *Юмашев М. В.* Неустановившиеся движения автотранспорта на кольцевой магистрали // Прикладная математика и механика. 2000. Т. 64, № 4. С. 651–658.
62. *Смирнов Н. Н.*, *Киселев А. Б.*, *Никитин В. Ф.*, *Юмашев М. В.* Математическое моделирование автомобильных потоков на магистралях // Вестник Московского университета. Математика. Механика. 2000. № 4. С. 39–44.
63. *Киселев А. Б.*, *Кокорева А. В.*, *Никитин В. Ф.*, *Смирнов Н. Н.* Математическое моделирование автотранспортных потоков на регулируемых дорогах // Прикладная математика и механика. 2004. Т. 68, № 6. С. 1047–1054.
64. *Холодов Я. А.*, *Холодов А. С.*, *Гасников А. В.*, *Морозов И. И.*, *Тарасов В. Н.* Моделирование транспортных потоков — актуальные проблемы и пути их решения // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 152–162.
65. *Daganzo C. F.* Fundamentals of transportation and traffic operations. N.Y.: Elsevier Science Inc., 1997.
66. *Aw A.*, *Rascle M.* Resurrection of «second order» models of traffic flow // SIAM Journal of Applied Mathematics. 2000. V. 60. P. 916–938.
67. *Greenberg J. M.* Extensions and amplifications of a traffic model of Aw and Rascle // SIAM J. Appl. Math. 2001. V. 62, № 3. P. 729–745.
68. *Zhang H. M.* A non-equilibrium traffic model devoid of gas-like behavior // Transp. Res. B. 2002. V. 36. P. 275–290.
69. *Морозов И. В.*, *Гасников А. В.*, *Тарасов В. Н.*, *Холодов Я. А.*, *Холодов А. С.* Численное исследование транспортных потоков на основе гидродинамических моделей // Компьютерные исследования и моделирование. 2011. Т. 3, № 4. С. 389–412.
70. *Siebel F.*, *Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic; `arXiv:physics/0509124v2`, 2006.
71. *Helbing D.* Improved fluid-dynamic model for vehicular traffic // Phys. Rev. E. 1995. V. 51. P. 3163–3169.
72. *Ладыженская О. А.* Шестая проблема тысячелетия: уравнение Навье—Стокса, существование и гладкость // УМН. 2003. Т. 58, № 2(350). С. 45–78.
73. *Юдович В. И.* Глобальная разрешимость — против коллапса в динамике несжимаемой жидкости // Математические события XX века. М.: Фазис, 2003. С. 519–548.
74. Проблемы турбулентности. Сборник работ. М.–Ижевск: НИЦ «РХД», ИКИ, 2006.
75. *Prigogine I.*, *Herman R.* Kinetic theory of vehicular traffic. N.Y.: Elsevier, 1971.
76. *Кац М.* Вероятность и смежные вопросы в физике. М.: Мир, 1965.
77. *Козлов В. В.* Ансамбли Гиббса и неравновесная статистическая механика. М.–Ижевск: НИЦ «РХД», ИКИ, 2008.
78. *Веденяпин В. В.* Кинетическая теория по Максвеллу, Больцману и Власову. М.: Изд-во МГОУ, 2005.



79. *Карамзин Ю. Н.*, *Трапезникова М. А.*, *Четверушкин Б. Н.*, *Чубарова Н. Г.* Двумерная модель автомобильных потоков // Матем. мод. 2006. Т. 18, № 6. С. 85–95.
80. *Сухинова А. Б.*, *Трапезникова М. А.*, *Четверушкин Б. Н.*, *Чубарова Н. Г.* Двумерная макроскопическая модель транспортных потоков // Матем. мод. 2009. Т. 21, № 2. С. 118–126.
81. *Garavello M.*, *Piccoli B.* Traffic Flow on Networks. Volume 1 of AIMS Series on Applied Mathematics. AIMS, 2006.
82. *Göttlich S.*, *Klar A.* Model hierarchies and optimization for dynamic flows on networks. Modeling and optimization of flows on networks. Cetaro (CS), June 15–19, 2009. C.I.M.E. Courses, 2009; `http://php.math.unifi.it/users/cime/Courses/2009/01/200914-Notes.pdf`
83. *Kurzhanskiy A. A.* Modeling and software tools for freeway operational planning. PhD thesis, Berkeley: University of California, 2007; (see also *Xiaotian Sun*, PhD thesis, Berkeley: University of California, 2005; *Gabriel Clemente Gomes Parisca*, PhD thesis, Berkeley: University of California, 2004); `http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-148.pdf`; `http://lihodeev.com/pubs.html`
84. *Куржанский А. А.*, *Куржанский А. Б.*, *Варайя П.* Роль макро-моделирования в активном управлении транспортной сетью // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 100–118.
85. *Daganzo C. F.* The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // Transp. Res. B. 1994. V. 28, № 4. P. 269–287.
86. *Daganzo C. F.* The cell transmission model, Part II: Network traffic // Transp. Res. B. 1995. V. 29, № 2. P. 79–93.
87. *Буслаев А. П.*, *Таташев А. Г.*, *Яшина М. В.* О свойствах решений одного класса систем нелинейных дифференциальных уравнений на графах // Владикавказский матем. жур., ВНЦ РАН. 2004. Т. 6, № 4. С. 4–18.
88. *Назаров А. И.* Об устойчивости стационарных режимов в одной системе ОДУ, возникающей при моделировании автотранспортных потоков // Вестник СПбГУ. Серия 1. Математика. Механика. Астрономия. 2006. № 3. С. 35–43.
89. *Lubashevsky I.*, *Kalenkov S.*, *Mahnke R.* Towards a variational principle for motivated vehicle motion // Phys. Rev. E. 2002. V. 65. P. 1–5.
90. *Lubashevsky I.*, *Wagner P.*, *Mahnke R.* Towards the fundamentals of car following theory; `arXiv:cond-mat/0212382v2`, 2003.
91. *Newell G. F.* Nonlinear effects in the dynamics of car flowing // Oper. Res. 1961. V. 9. P. 209–229.
92. *Новокшенов В. Ю.* Введение в теорию солитонов. М.–Ижевск: НИЦ «РХД», ИКИ, 2002.
93. *Ильин А. М.*, *Олейник О. А.* Асимптотическое поведение решений задачи Коши для некоторых квазилинейных уравнений при больших значениях времени // Матем. сб. 1960. Т. 51(93), № 2. С. 191–216.

94. *Osher S.*, *Ralston J.* $L^1$ stability of traveling waves with application to convective porous media flow // Comm. Pure Appl. Math. 1982. V. 35. P. 737–749.
95. *Weinberger H. F.* Long-time behavior for regularized scalar conservation law in absence of genuine nonlinearity // Ann. Inst. H. Poincaré, Anal. Non Linéaire. V. 7. 1990. P. 407–425.
96. *Liu T.-P.* Admissible solutions of hyperbolic conservation laws // Mem. Amer. Math. Soc. 1981. V. 30, № 240. P. 1–78.
97. *Cheng K.-S.* Asymptotic behavior of solution of a conservation law without convexity condition // J. Diff. Equat. 1981. V. 40, № 3. P. 343–376.
98. *Петросян Н. С.* Об асимптотике решения задачи Коши для квазилинейного уравнения первого порядка с невыпуклой функцией состояния // УМН. 1983. Т. 38, № 2 (230). С. 213–214.
99. *Кружков С. Н.*, *Петросян Н. С.* Асимптотическое поведение решений задачи Коши для нелинейных уравнений первого порядка // УМН. 1987. Т. 42, № 5 (257). С. 3–40.
100. *Jennings G.* Discrete shocks // Comm. Pure Appl. Math. 1974. V. 27. P. 25–37.
101. *Harten A.*, *Hyman J. M.*, *Lax P. D.* On finite-difference approximations and entropy conditions for shocks // Comm. Pure Appl. Math. 1976. V. 29. P. 297–322.
102. *Engquist B.*, *Osher S.* One-sided difference approximations for nonlinear conservation laws // Math. Comp. 1981. V. 36. P. 321–351.
103. *Henkin G. M.*, *Polterovich V. M.* Shumpetrian dynamics as non-linear wave theory // J. Math. Econom. 1991. V. 20. P. 551–590.
104. *Henkin G. M.*, *Polterovich V. M.* A difference-differential analogue of the Burgers equation and some models of economic development // Discrete and continuous dynamic systems. 1999. V. 5, № 4. P. 697–728.
105. *Mejai M.*, *Volpert Vit.* Convergence to systems of waves for viscous scalar conservation laws // Asymptotic Analysis. 1999. V. 20. P. 351–366.
106. *Engelberg S.*, *Schochet S.* Nonintegrable perturbation of scalar viscous shock profiles // Asymptotic Analysis. 2006. V. 48. P. 121–140.
107. *Henkin G. M.*, *Shananin A. A.* Asymptotic behavior of solutions of the Cauchy problem for Burgers type equations // J. Math. Purés Appl. 2004. V. 83. P. 1457–1500.
108. *Henkin G. M.*, *Shananin A. A.*, *Tumanov A. E.* Estimates for solution of Burgers type equations and some applications // J. Math. Purés Appl. 2005. V. 84. P. 717–752.
109. *Henkin G. M.* Asymptotic structure for solutions of the Cauchy problem for Burgers type equations // J. fixed point theory appl. 2007. V. 1, № 2. P. 239–291.
110. *Serre D.* $L^1$ stability of shock waves in scalar conservation laws // Evolutionary Equations. V. 1. Amsterdam: North-Holland, 2004. P. 473–553. (Handbook of Differential Equations).
111. *Гасников А. В.* О промежуточной асимптотике решения задачи Коши для квазилинейного уравнения параболического типа с монотонным начальным условием // Известия РАН. Теория и системы управления. 2008. № 3. С. 154–163.



112. *Гасников А. В.* Сходимость по форме решения задачи Коши для квазилинейного уравнения параболического типа с монотонным начальным условием к системе волн // ЖВМ и МФ. 2008. Т. 48, № 8. С. 1458–1487.
113. *Гасников А. В.* Асимптотическое по времени поведение решения начальной задачи Коши для закона сохранения с нелинейной дивергентной вязкостью // Известия РАН. Серия математическая. 2009. Т. 76, № 6. С. 39–76.
114. *Гасников А. В.* Асимптотика по времени решения задачи о распаде «размазанного разрыва» для закона сохранения // Труды МФТИ (специальный выпуск, посвященный юбилею ФУПМа). 2009. Т. 1, № 4. С. 120–125; `http://mipt.ru/nauka/trudy/N4.html`
115. *Хёрмандер Л.* Анализ линейных дифференциальных операторов с частными производными. Т. 1–4. М.: Мир, 1986–1988.
116. *Gazis D. C.* Traffic science. N.Y.: Wiley, 1974.
117. *Treiber M.*, *Helbing D.* Explanation of observed features of self-organization in traffic flow; `arXiv:cond-mat/9901239`, 1999.
118. *Treiber M.*, *Hennecke A.*, *Helbing D.* Congested traffic states in empirical observations and microscopic simulation // Phys. Rev. E. 2000. V. 62. P. 1805–1824.
119. *Фон Нейман Дж.* Теория самовоспроизводящихся автоматов. М.: УРСС, 2010.
120. *Cremer M.*, *Ludwig J.* A fast simulation model for traffic flow on the basis of Boolean operations // Math. Comp. Simul. 1986. V. 28. P. 297–303.
121. *Nagel K.*, *Schreckenberg M.* A cellular automaton model for freeway traffic // Phys. I France. 1992. V. 2. P. 2221–2229.
122. *Chowdhury D.*, *Santen L.*, *Schadschneider A.* Statistical physics of vehicular traffic and some related systems // Phys. Rep. 2000. V. 329. P. 199–329; `arXiv:cond-mat/0007053v1`
123. *Nagatani T.* The physics of traffic jams // Reports on Progress in Physics. 2002. V. 65. P. 1331–1386.
124. *Benassi A.*, *Fouque J.-P.* Hydrodynamic limit for the asymmetric simple exclusion process // Ann. of Probability. 1987. V. 15, № 2. P. 546–560.
125. *Kipnis C.*, *Olla S.*, *Varadhan S. R. S.* Hydrodynamics and large deviation for simple exclusion processes // Comm. on Pure and Applied Mathematics. 1989. V. 42. P. 115–137.
126. *Nishinari K.*, *Matsukidaira J.*, *Takahashi D.* Two-dimensional Burgers cellular automaton; `arXiv:nlin/0102027v1`, 2001.
127. *Бланк М. Л.* Точный анализ динамических систем, возникающих в моделях транспортных потоков // УМН. 2000. Т. 55(333), № 3. С. 167–168.
128. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111, № 3–4. P. 903–930; `arXiv:math.DS/0206194`
129. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120, № 3–4. P. 627–658; `arXiv:math.DS/0408240`
130. *Минлос Р. А.* Введение в математическую статистическую физику. М.: МЦНМО, 2002.
131. *Maerivoet S.*, *De Moor B.* Cellular automata models of road traffic // Physics Reports 2005. V. 419, № 1. P. 1–64; `arXiv:physics/0509082`

132. *Буслаев А. П.*, *Новиков А. В.*, *Приходько В. М.*, *Таташев А. Г.*, *Яшина М. В.* Вероятностные и имитационные подходы к оптимизации автодорожного движения. М.: Мир, 2003.
133. *Buslaev A. P.*, *Prikhodko V. M.*, *Tatashev A. G.*, *Yashina M. V.* The deterministic — stochastic flow model; `arXiv:physics/0504139v1`, 2005.
134. *Buslaev A. P.*, *Gasnikov A. V.*, *Yashina M. V.* Selected mathematical problems of traffic flow theory // International Journal of Computer Mathematics. 2012. V. 89, № 3. P. 409–432.
135. Явление чрезвычайное. Книга о А. Н. Колмогорове. М.: ФАЗИС, МИРОС, 1999. С. 236–237.
136. *Баренблатт Г. И.* Автомодельные явления — анализ размерностей и скейлинг. Долгопрудный: Издательский дом «Интеллект», 2009.
137. *Ибрагимов Н. Х.* Практический курс дифференциальных уравнений и математического моделирования. Нижний Новгород: Издательство Нижегородского университета, 2007.
138. *Шестаков А. А.* Обобщенный прямой метод Ляпунова для систем с распределенными параметрами. М.: КомКнига, 2007.
139. *Volpert A. I.*, *Volpert Vit. A.*, *Volpert Vl. A.* Traveling waves solutions of parabolic system // Translations of Mathematical Monographs. 2000. V. 140. P. 1–455.
140. *Колмогоров А. Н.*, *Петровский И. Г.*, *Пискунов Н. С.* Исследование уравнения диффузии, соединенной с возрастанием количества вещества, и его применение к одной биологической проблеме // Бюл. МГУ. Математика и механика. 1937. Т. 1, № 6. С. 1–26.
141. *Разжевайкин В. Н.* Решения типа бегущей волны для уравнения реакции — нелинейной диффузии // Труды МФТИ (специальный выпуск, посвященный юбилею ФУПМа). 2009. Т. 1, № 4. С. 99–119; `http://mipt.ru/nauka/trudy/N4.html`
142. *Наумкин П. И.*, *Шишмарев И. А.* Задача о распаде ступеньки для уравнения Кортевега—де Фриза—Бюргерса // Функц. анализ и его прил. 1991. Т. 25, № 1. С. 21–32.
143. *Наумкин П. И.*, *Шишмарев И. А.* О распаде ступеньки для уравнения Кортевега—де Фриза—Бюргерса // Функц. анализ и его прил. 1991. Т. 26, № 2. С. 88–93.
144. *Казейкина А. В.* Асимптотическое при больших временах поведение решений некоторых аналогов уравнения типа Кортевега—де Фриза. Диссертация на соискание ученой степени к.ф.-м.н. Москва: ВМиК, 2012.
145. *Duan R.*, *Zhao H.* Global stability of strong rarefacion waves for the generalized KdV—Burgers equation // Nonlinear Anal. 2007. V. 66. P. 1100–1117.
146. *Liu T.-P.*, *Nishihara K.* Asymptotic behavior for scalar viscous conservation laws with boundary effect // Journal of differential equations. 1997. V. 133. P. 296–320.
147. *Liu T.-P.*, *Matsumura A.*, *Nishihara K.* Behaviors of solutions for the Burgers equation with boundary corresponding to rarefaction waves // SIAM J. Math. Anal. 1998. V. 29, № 2. P. 293–308.



148. *Куликовский А. Г.*, *Чугайнова А. П.* Классические и неклассические разрывы в решениях уравнений нелинейной теории упрогости // УМН. 2008. Т. 63, № 2(380). С. 85–152; `http://www.mi.ras.ru/spm/pdf/007.pdf` `http://www.mi.ras.ru/noc/lectures/16kulikovskii.pdf`
149. *Казейкина А. В.* Примеры отсутствия бегущей волны для обобщенного уравнения Кортевега-де Фриза—Бюргерса // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн. 2011. № 1. С. 17–24.
150. *Шубин М. А.* Лекции об уравнениях математической физики. М.: МЦНМО, 2003.
151. *Габушин В. Н.* Неравенства между производными в метриках $L_p$ при $0 < p \leqslant \infty$ // Известия АН СССР. Серия математическая. 1976. Т. 40, № 4. С. 869–892.
152. *Габушин В. Н.* Неравенства для производных решений обыкновенных дифференциальных уравнений в метриках $L_p$ $(0 < p \leqslant \infty)$ // Дифференциальные уравнения. 1988. Т. 24, № 10. С. 1662–1670.
153. *Арестов В. В.* Приближение неограниченных операторов ограниченными и родственные экстремальные задачи // УМН. 1996. Т. 51, № 6(312). С. 89–124.
154. *Годунов С. К.* Проблема обощенного решения в теории квазилиненйых уравнений и в газовой динамике // УМН. 1962. Т. 17, № 3(105). С. 147–158.
155. *Арнольд В. И.*, *Вишик М. И.*, *Ильяшенко Ю. С.*, *Калашников А. С.*, *Кондратьев В. А.*, *Кружков С. Н.*, *Ландис Е. М.*, *Миллионщиков В. М.*, *Олейник О. А.*, *Филиппов А. Ф.*, *Шубин М. А.* Некоторые нерешенные задачи теории дифференциальных уравнений и математической физики // УМН. 1989. Т. 44, № 4(268). С. 191–202.
156. *Панов Е. Ю.* О единственности решения задачи Коши для квазилинейного уравнения первого порядка с одной допустимой строго выпуклой энтропией // Матем. заметки. 1994. Т. 55, № 5. С. 116–129.
157. *Keyfitz B. L.* Hold that light! Modeling of traffic flow by differential equations // Six themes on variation / Robert Hardt (Ed.). AMS, 2004. P. 127–153. (Student mathematical library. V. 26).
158. *Kerner B. S.* The Physics of Traffic. Berlin: Springer, 2004.
159. *Bando M.*, *Hasebe K.*, *Nakayama A.*, *Shibata A.*, *Sugiyama Y.* Dynamical model of traffic congestion and numerical simulation // Phys. Rev. E. 1995. V. 51. P. 1035–1042.
160. *Barlovic R.*, *Santen L.*, *Schadschneider A.*, *Schreckenberg M.* Metastable states in cellular automata for traffic flow // Eur. Phys. J. B. 1998. V. 5. P. 793.
161. *Wiedemann R.* Simulation des Straßenverkehrsflusses. Karlsruhe: University of Karlsruhe, 1974.
162. *Kerner B. S.*, *Konhäuser P.* Cluster effect in initially homogeneous traffic flow // Phys. Rev. E. 1993. V. 48. P. 2335–2338.
163. *Kerner B. S.*, *Konhäuser P.* Structure and parameters of clusters in traffic flow // Phys. Rev. E. 1994. V. 50. P. 54–83.
164. *Kerner B. S.*, *Klenov S. L.*, *Konhäuser P.* Asymptotic theory of traffic jams // Phys. Rev. E. 1997. V. 56. P. 4200–4216.

# Глава 3

# Теория Кернера трех фаз в транспортном потоке — новый теоретический базис для интеллектуальных транспортных технологий

## Введение

В 1996–2002 годах Б. С. Кернер с сотрудниками концерна «Даймлер» провели детальные исследования эмпирических данных, измеренных с помощью датчиков на многочисленных скоростных автомагистралях мира (в Германии, Голландии, Англии, США). Главный результат этих исследований был сформулирован в предисловии к книге [1]:

> Теории транспортного потока и математические модели, которые доминируют в настоящее время в научных журналах и учебных курсах большинства университетов, не могут объяснить ни сам переход от свободного к плотному потоку (traffic breakdown), ни основные свойства возникающих в результате этого перехода структур транспортного потока.

По этой причине Б. С. Кернер предложил и разработал альтернативную теорию транспортных потоков, названную *теорией трех фаз*, которая может предсказать и объяснить эмпирические свойства перехода к плотному потоку (traffic breakdown) и результирующих пространственно-временных структур в транспортном потоке [1–6]. Как достижения, так и критика предшествующих классических подходов к теории транспортных потоков представлены в главе 10 книги [1]. В настоящей главе кратко излагаются основные положения теории трех фаз Кернера в соответствии с [1, 2].

Цель этой главы состоит в том, чтобы дать читателю определенное представление об эмпирическом базисе и основных идеях теории трех фаз. Тем не менее эта глава не может заменить книг Б. С. Кернера, чтение которых необходимо тем, кто хочет разобраться в данной теории. По этой причине основное внимание в разделе 3.1 уделяется не математическому обоснованию тех или иных положений теории трех фаз, а в основном качественному описанию. В разделе 3.2 кратко описана стохастическая трехфазная модель в рамках теории трех фаз и некоторые ее решения. Для более подробного ознакомления с математическими результатами теории



трех фаз Кернера рекомендуется прочитать главу 11 книги [1] и часть III книги [2].

Прежде чем перейти к изложению некоторых положений теории трех фаз Кернера, важно отметить имевшиеся на тот момент фундаментальные достижения (см. разделы 2.1–2.3) по математической формулировке моделей транспортного потока, а также многочисленных эффектов взаимодействия между водителями [7–29]. К ним относится, в первую очередь, введенная в моделях ДМ-класса задержка водителей [14–17], приводящая к переторможению как реакции на замедление АТС впереди [17, 23, 24]; формулировка задержки водителей через модельные флуктуации, введенные в моделях Нагеля—Шрекенберга и А. Шашнайдера (см. обзоры [23–26]); микроскопическое описание slow-to-start rule в работах группы М. Шрекенберга (см. обзоры там же). Эти и другие математические формулировки в поведения водителя являются также важными элементами математических моделей трех фаз (см. главу 11 книги [1]).

Несмотря на эти достижения, как было сказано выше, предшествующие модели не могут объяснить ни сам переход от свободного к плотному потоку, ни основные свойства возникающих пространственно-временных структур, наблюдаемых в эмпирических данных. Детальное объяснение этого «парадокса» дано в главе 10 книги [1]. Этот парадокс объясняется очень просто: анализ эмпирических данных, который позволил выявить фундаментальные эмпирические свойства перехода от свободного к плотному 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 детально описаны в разделе 2.4 этой книги и части II книги [2]. В рамках данной главы нет возможности остановиться на рассмотрении всех этих эмпирических свойств транспортного потока. Тем не менее повторим здесь еще раз фундаментальные эмпирические свойства перехода к плотному транспортному потоку и эмпирические свойства фазовых переходов в транспортном потоке (см. раздел 2.4).

Фундаментальные эмпирические свойства перехода к плотному транспортному потоку следующие:

**1.** Переход к плотному транспортному потоку (traffic breakdown) является переходом F → S (буква F соответствует *free flow*, т. е. свободному потоку, буква S обозначает фазу синхронизированного потока, в английской литературе *synchronized flow*).

**2.** Вероятность спонтанного перехода F → S является растущей функцией величины потока АТС.

**3.** Может быть как спонтанный, так и индуцированный переход F → S около одного и того же узкого места на дороге (bottleneck).

LWR-модель кинематических и ударных волн в транспортном потоке, дискретной версией которой является CTM-модель Даганзо [13], не могут описывать пункты 2 и 3, а модели, относящиеся к ДМ-классу, не могут описывать пункты 1–3.

Фундаментальные эмпирические свойства фазовых переходов в транспортном потоке следующие:

а) В соответствии со свойствами 1–3, указанными выше, переход от свободного к плотному транспортному потоку (traffic breakdown) является фазовым переходом F → S I рода.

б) Широкие движущиеся кластеры (wide moving jams, обозначается ниже буквой J) возникают спонтанно только в синхронизированном потоке, т. е. в результате последовательности фазовых переходов F → S → J.

в) Фазовый переход S → J происходит позднее и часто совсем в другом месте, чем фазовый переход F → S.

LWR-теория кинематических волн не может описывать пункты б) и в), а теория, основанная на моделях ДМ-класса, не может описывать пункты а)–в).

## 3.1. Три фазы транспортного потока

### 3.1.1. Предварительные сведения

Теория трех фаз фокусируется главным образом на физике плотного транспортного потока на скоростных автомагистралях. В ней описывается три фазы транспортного потока, в то время как классические теории, базирующиеся на фундаментальной диаграмме транспортного потока, рассматривают две фазы: *свободный поток* и так называемый *плотный поток* (congested traffic в английской терминологии). В плотном потоке выделяются две фазы, соответственно существуют три фазы транспортного потока:

**1.** Свободный поток — фаза F.

**2.** Синхронизированный поток — фаза S.



**3.** Широкий движущийся кластер (локальный движущийся затор, в английской литературе wide moving jam) — фаза J.

*Фаза* определяется как некоторое *состояние* транспортного потока, рассматриваемое в *пространстве* и *времени*.

Следует подчеркнуть, что в теории трех фаз разделение на свободный и плотный поток точно такое же, как и в классических теориях Лайтхилла—Уизема и «Дженерал Моторс» (см. п. 3.1.3 ниже). Фундаментальное отличие теории Б. С. Кернера состоит в том, что он выделяет две фазы в плотном потоке на основе общих эмпирических пространственно-временных свойств транспортного потока, которые за все годы измерений остаются одни и те же на разных автодорогах мира. Другими словами, как определение фаз транспортного потока, так и остальные положения теории, приведенные в разделе 3.1, основаны исключительно на эмпирических данных.

### 3.1.2. Свободный транспортный поток — фаза F

В свободном транспортном потоке достаточно малой плотности водители могут практически свободно установить желаемую для них скорость. Эмпирические данные, относящиеся к свободному потоку, показывают положительную корреляцию между величиной потока $q$, измеряемой в количестве АТС в единицу времени (проходящих через данное сечение дороги), и плотностью $\rho$, измеряемой в количестве АТС на единицу длины дороги [18, 19]. Зависимость потока $q$ от плотности $\rho$ 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.

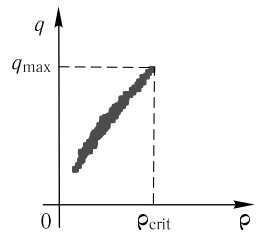



Рис. 1. Зависимость потока $q$ от плотности АТС $\rho$ в свободном потоке [18, 24]

### 3.1.3. Плотный транспортный поток

В плотном транспортном потоке, который определяется так же, как и в классических теориях Лайтхилла—Уизема и «Дженерал Моторс» (см. главу 2), скорость АТС меньше, чем минимально возможная скорость АТС в свободном потоке. Это означает, что прямая с наклоном, равным

минимальной скорости в транспортом потоке (штриховая линия на рис. 2), разделяет все эмпирические данные (точки) на плоскости поток–плотность на две области: слева от этой прямой находятся данные, относящиеся к свободному потоку, а справа — данные, относящиеся к плотному потоку.

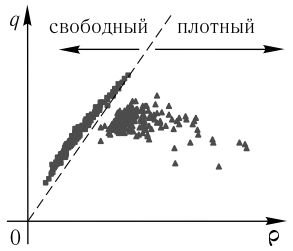



Рис. 2. Зависимость потока от плотности АТС в свободном и плотном потоке [18, 24]

Как следует из данных измерений, возникновение плотного потока обычно происходит вблизи неоднородности на автомагистрали, вызванной въездом на автомагистраль, съездом с нее, изменением числа полос, сужением дороги, подъемом и т. п. Такого типа неоднородность, вблизи которой может происходить переход к плотному транспортному потоку, в дальнейшем будем называть *узким местом* или «бутылочным горлом» [18, 24].

### 3.1.4. Определение фаз J и S в плотном транспортном потоке

Б. С. Кернер показал, что фундаментальная диаграмма и ее применения в том виде, как они используются в классических теориях транспортного потока, неадекватным образом описывают сложную динамику в плотном транспортном потоке. Он выделяет, таким образом, в плотном транспортном потоке фазу S *синхронизированного 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. Б. С. Кернер определил фазы J и S следующим образом.

*Определение фазы* [J] *широкого движущегося кластера*:

Задний по направлению движения фронт широкого движущегося кластера (локального движущегося затора), где АТС, выезжающие из кластера, ускоряются вплоть до свободного или до синхронизированного потока, движется против потока с постоянной средней скоростью $v_g$, проходя через все узкие места на скоростной автомагистрали. Это характеристическое свойство широкого движущегося кластера.



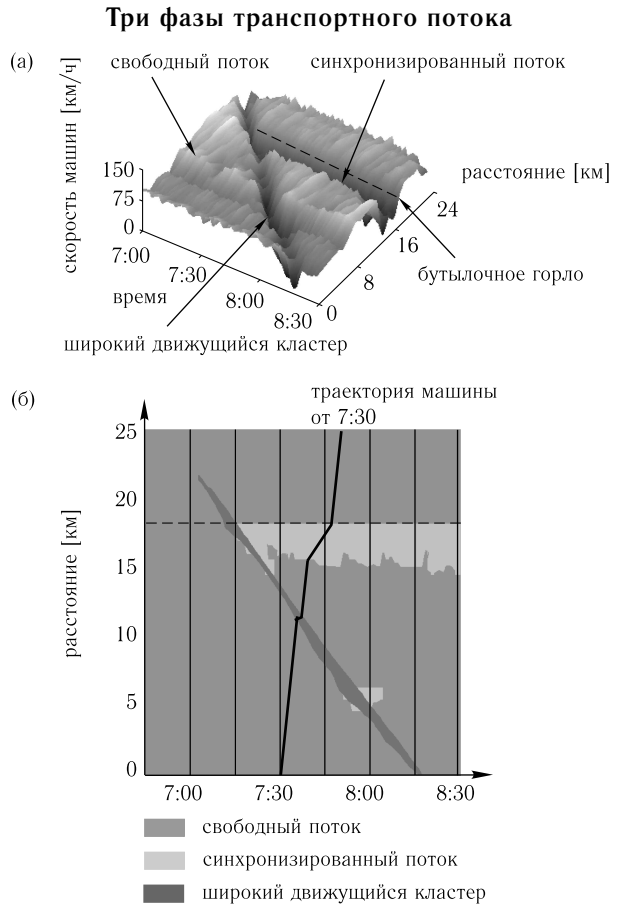



Рис. 3. Данные измерений скорости АТС в пространстве и времени (а) и их представление на координатно-временной плоскости (б). Взято из [1]

*Определение фазы* [S] *синхронизированного потока*:

Задний по направлению движения фронт области синхронизированного потока, где АТС ускоряются вплоть до свободного потока, НЕ обладает характеристическим свойством широкого движущегося кластера. В частности, задний фронт синхронизированного потока часто фиксирован вблизи узкого места на скоростной автомагистрали.

Необходимо подчеркнуть, что определение фаз [J] и [S] вытекает из эмпирических пространственно-временных свойств плотного потока, т. е. исходно не имеет никакого отношения к какой-либо математике. Теоретический смысл этих определений можно понять, прочтя раздел 6.1 книги [1].

Данные измерений средней скорости АТС (рис. 3 (а)) иллюстрируют определения [J] и [S]. На рис. 3 (а) имеются две пространственно-временные структуры плотного потока с низкой скоростью АТС. Одна из них распространяется против потока с почти постоянной скоростью заднего фронта через все узкие места на скоростной автомагистрали. Согласно определению [J], эта область плотного потока относится к фазе «широкого движущегося кластера». Напротив, задний фронт другой области плотного потока фиксирован вблизи места съезда с автомагистрали. Согласно определению [S], эта область плотного потока относится к фазе «синхронизированного потока» (рис. 3 (а) и (б)).

В секции 6.1 книги [1] было показано, что определения [S] и [J] соответственно для фаз синхронизированного потока и широкого движущегося кластера являются основой для большинства гипотез теории трех фаз и соответствующих микроскопических трехфазных моделей транспортного потока. Необходимо отметить, что определения [S] и [J] являются нелокальными и макроскопическими, и они применимы только после того, как измерены макроскопические данные в пространстве и времени, т. е. на «off-line» стадии. Это связано с тем, что для четкого разделения фаз S и J на основе определения [S] и [J] нужно рассматривать прохождение областей плотного потока через узкие места на скоростной автомагистрали. Часто это рассматривается как недостаток этих определений фаз плотного потока. Однако существуют локальные критерии для разделения фаз S и J без рассмотрения прохождения областей плотного потока через узкие места. Эти микроскопические критерии относятся к данным измерения параметров движения отдельных АТС в плотном потоке и состоят в следующем. Если в этих данных наблюдается так называемый «интервал прерывания потока», т. е. временной интервал между двумя последовательно движущимися АТС в плотном потоке значительно превышает среднее время $\tau^{(a)}_{\mathrm{del,jam}}$ задержки водителя при ускорении на заднем (по потоку) фронте широкого локального кластера (1,3–2,1 с), то эти данные отвечают фазе широкого локального кластера. После того как с помощью данного критерия выявлены все широкие кластеры в плотном потоке, остальные состояния плотного потока относятся к фазе синхронизированного потока.

### 3.1.5. Возникновение плотного потока — фазовый переход F → S

Переход от свободного к плотному потоку в англоязычной литературе известен как traffic breakdown. В теории трех фаз такой переход объясняется возникновением фазы синхронизированного потока, т. е. фазовым переходом F → S.

Такое объяснение основывается на имеющихся данных измерений, которые показывают, что после возникновения плотного потока вблизи узкого места на автомагистрали задний фронт возникшего плотного потока



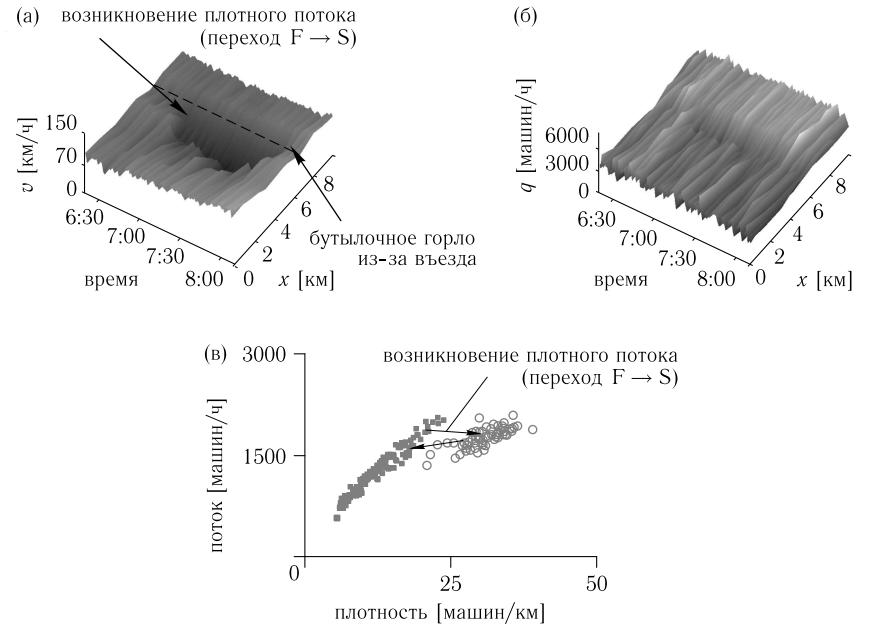



Рис. 4. Эмпирический пример возникновения плотного потока и эффект гистерезиса у бутылочного горла из-за въезда на автодорогу. Средняя скорость (а) и поток (б) на автодороге в пространстве и времени (увеличение потока после въезда на (б) связано с потоком въезжающих на дорогу АТС). (в) Эффект гистерезиса в плоскости поток-плотность обозначен двумя стрелками. Данные усреднены за 1 минуту. Взято из [1]

фиксирован вблизи этого узкого места. Таким образом, возникший плотный поток удовлетворяет определению [S] фазы синхронизированного потока. В самом деле, типичный пример перехода из свободного в синхронизированный поток вблизи въезда показан на рис. 4. Из рисунка видно, что в то время как скорость АТС резко уменьшается в процессе перехода (рис. 4 (а)), поток меняется мало (рис. 4 (б)). Скачок скорости при мало меняющемся потоке особенно наглядно виден на рис. 4 (в). В течение всего времени после перехода задний фронт между плотным и свободным потоками фиксирован у въезда на дорогу. По этой причине плотный поток соответствует определению фазы синхронизированного потока, поэтому весь плотный поток относится к фазе синхронизированного потока. Образование плотного потока примерно в 6:30 (показанное стрелкой слева направо на рис. 4 (в)) и его исчезновение примерно в 7:45 (показанное там же стрелкой справа налево) сопровождается гистерезисом, хорошо известным в теории фазовых переходов 1 рода, наблюдаемых в широком классе неравновесных физических, химических и биологических систем.

Это свойство фазового перехода F → S 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. 5 (б), координата дороги $x = 0$ км). Как можно видеть, в результате перехода к плотному потоку на узком месте, связанным с въездом на скоростную магистраль, сначала образуется фаза S синхронизированного потока. Действительно, в течение всего времени существования плотного потока на этом узком месте задний фронт плотного потока, на котором АТС ускоряются из плотного потока до свободного потока, фиксирован на этом узком месте. Поэтому по определению фаз в теории Б. С. Кернера в результате перехода к плотному потоку образуется фаза синхронизированного потока. Другими словами, плотный поток образуется в результате перехода F → S. Напротив, широкие движущиеся кластеры возникают позднее уже внутри фазы синхронизированного потока. Этот фазовый переход S → J будет рассмотрен ниже в п. 3.1.9.

Таким образом, переход от свободного к плотному потоку в эмпирических данных есть переход F → S первого рода. Это эмпирическое свойство есть общее свойство реальных транспортных потоков на скоростных магистралях. Напротив, в моделях ДМ-класса, как было объяснено в п. 2.4, переход от свободного к плотному потоку связан с возникновением широких движущихся кластеров.

Исходя из эмпирических данных, был сделан вывод, что синхронизированный поток может возникать в свободном потоке спонтанно (спонтанный переход F → S) или индуцированным образом (индуцированный переход F → S). Спонтанный переход F → S означает, что переход к синхронизированному 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 переход F → S 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 переход F → S связан с распространением в направлении против потока области синхронизированного потока или же широкого движущегося кластера, которые первоначально возникли вблизи следующего в направлении потока узкого места. Эмпирический пример индуцированного фазового перехода, приводящего к возникновению синхронизированного потока, показан на рис. 3: синхронизированный поток



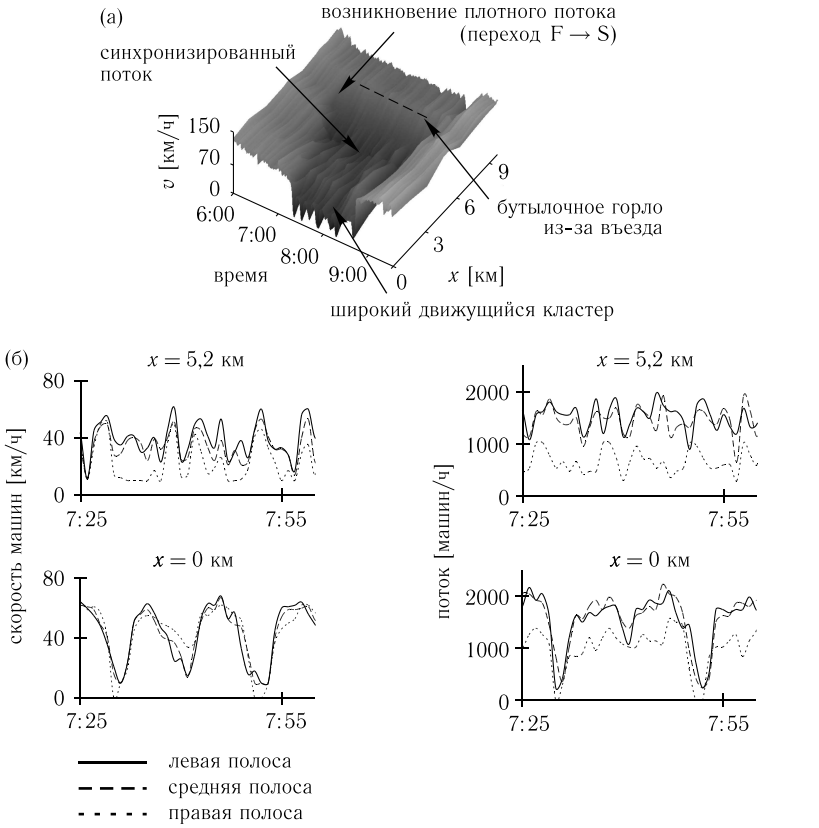



Рис. 5. Эмпирический пример возникновения широких движущихся кластеров в синхронизированном потоке: (а) Скорость АТС в пространстве и времени. (б) Скорость (слева) и поток (справа) на трех полосах дороги в области синхронизированного потока ($x = 5{,}2$ км) и в области широких движущихся кластеров ($x = 0$ км). Взято из [1]

возникает благодаря распространению против потока широкого движущегося кластера.

Природу фазового перехода F → S можно объяснить с помощью «соревнования» во времени и пространстве двух противоположных процессов: ускорения АТС при обгоне более медленного АТС впереди, названном «переускорением», и в случае, когда обгон невозможен, торможения АТС до скорости более медленного АТС, названном «адаптация скорости». «Переускорение» поддерживает дальнейшее существование свободного потока. Напротив, «адаптация скорости» ведет к синхронизированному потоку. Было постулировано, что вероятность обгона, которая совпадает

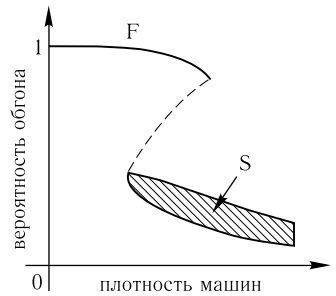



Рис. 6. Объяснение фазового перехода к плотному потоку (traffic breakdown) на основе Z-образной нелинейной функции вероятности обгона (вероятности «переускорения») в теории Б. С. Кернера. Пунктирная линия описывает критическое значение вероятности обгона как функцию плотности АТС [1]

с вероятностью «переускорения», является разрывной функцией плотности (рис. 6): при данной плотности АТС вероятность обгона в свободном потоке много больше, чем в синхронизированном потоке.

Разрывная функция вероятности обгона является одной и той же как для спонтанного, так и для индуцированного фазового перехода $F \to S$: термины *спонтанный* и *индуцированный* отличаются только источником возмущения, приводящего к фазовому переходу $F \to S$. Переход $F \to S$ происходит при условии, что вероятность обгона внутри возмущения в свободном потоке меньше, чем критическая вероятность. Эта критическая вероятность показана пунктирной линией на рис. 6. Другими словами, не имеет значения, будет ли это критическое значение вероятности обгона достигнуто благодаря возмущению в свободном потоке (спонтанный переход) или благодаря распространению до узкого места некоторого возмущения, возникшего ранее в другой области дороги (индуцированный переход).

Отметим, что фазовые переходы $F \to S$ и $S \to F$ сопровождаются гистерезисом. Этот гистерезис не имеет никакого отношения к хорошо известному гистерезису в математических моделях ДМ-класса, который был впервые найден в теории Кернера—Конхойзера. Этот известный гистерезис, описанный в огромном количестве математических работ (см. [23–26], [33, 34] и ссылки в них), описывает фазовый переход $F \to J$ и обратный переход $J \to F$. Как уже несколько раз отмечалось, спонтанный переход $F \to J$ не наблюдается в реальном транспортном потоке.

### 3.1.6. Бесконечное число значений пропускных способностей скоростной автомагистрали

Спонтанное образование плотного потока, т. е. спонтанный фазовый переход $F \to S$ может произойти в широком диапазоне значений величины



потока $q$ в свободном транспортном потоке. Основываясь на эмпирических данных измерений, был сделан вывод, что существует бесконечное число значений пропускной способности автомагистрали в свободном потоке. Это бесконечное число значений пропускной способности находится в диапазоне между минимальным $q_{\text{th}}$ и максимальным $q_{\max}$ значениями пропускной способности (см. рис. 7). Если величина потока близка к максимальному значению пропускной способности $q_{\max}$, то уже достаточно малое возмущение в свободном потоке вблизи узкого места приведет к спонтанному фазовому переходу $F \to S$. С другой стороны, если величина потока близка к минимальному значению пропускной способности $q_{\text{th}}$, то только возмущение очень большой амплитуды способно привести к спонтанному фазовому переходу $F \to S$.

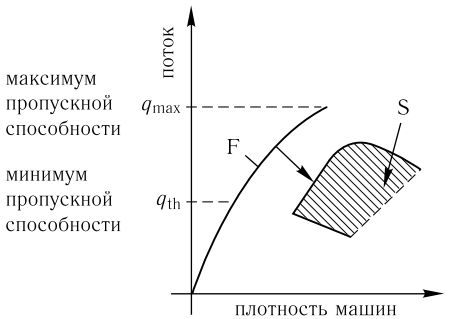



Рис. 7. Максимум и минимум пропускной способности скоростной автомагистрали в теории трех фаз. Взято из [1]

Вероятность возникновения малых возмущений в свободном транспортном потоке много выше, чем вероятность возникновения возмущений большой амплитуды. По этой причине, чем выше величина потока $q$ в свободном потоке вблизи узкого места, тем выше вероятность спонтанного фазового перехода $F \to S$.

Если величина потока $q$ меньше, чем минимальная пропускная способность $q_{\text{th}}$, то возникновение плотного потока (переход $F \to S$) невозможно.

Бесконечное число значений пропускной способности автомагистрали вблизи узкого места может быть объяснено тем, что свободный поток при значениях величины потока $q$ в диапазоне

$$q_{\text{th}} \leqslant q \leqslant q_{\max}$$

является метастабильным. Это означает, что при возникновении малых возмущений свободный поток сохраняется, т. е. является устойчивым относительно малых возмущений. Однако для больших возмущений свободный

поток оказывается неустойчивым и происходит фазовый переход F → S к синхронизированному потоку.

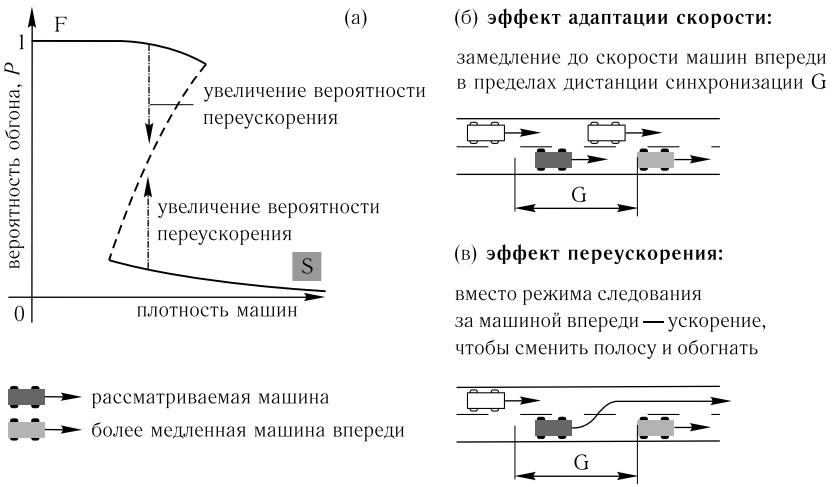



Рис. 8. Пояснение «соревнования» между переускорением и адаптацией скорости, которое объясняет бесконечное число значений пропускной способности автомагистрали. Взято из [1]

Как уже упоминалось в п. 3.1.5, природу фазового перехода F → S можно объяснить с помощью «соревнования» двух процессов: переускорения АТС при обгоне более медленного АТС впереди и торможения АТС до скорости более медленного АТС (адаптация скорости). На рис. 8 это поясняется более детально. На рис. 8 (а), который соответствует рис. 6, стрелочка вниз означает, что если в свободном потоке вблизи узкого места возникает локальное уменьшение скорости АТС, то вероятность обгона внутри этого возмущения падает. Если это уменьшение вероятности обгона становится меньше, чем критическая величина вероятности обгона, показанная пунктирной линией на рис. 8 (а), то фазовый переход F → S происходит внутри этого возмущения; в противоположном случае возмущение затухает и свободный поток остается на узком месте. Стрелочка вверх на рис. 8 (а) означает, что если исходное состояние отвечает фазе синхронизированного потока и в этом состоянии возникает случай локального увеличения скорости АТС, то внутри данного возмущения вероятность обгона возрастает. Если это возрастание вероятности обгона превышает критическое значение (как выше отмечалось, эта критическая вероятность обгона отвечает пунктирной кривой на рис. 8 (а)), то в области возмущения происходит переход S → F; в противоположном случае возмущение затухает и остается синхронизированный поток.



Бесконечное число значений пропускной способности автомагистрали вблизи узкого места в теории трех фаз не согласуется с классическими теориями транспортного потока (а также с методами управления транспортными потоками и их автоматического регулирования). Классические теории предполагают существование в любой момент времени некоторой фиксированной или случайной пропускной способности.

### 3.1.7. Широкие движущиеся кластеры (локальные движущиеся заторы) — фаза J

Широкий движущийся кластер может быть назван широким только при условии, что его ширина (вдоль дороги) заметно превышает ширину фронтов кластера. Средняя скорость движения АТС внутри широкого движущегося кластера много меньше, чем скорость АТС в свободном потоке. На заднем (в направлении потока) фронте кластера АТС могут ускоряться вплоть до свободного потока. На переднем фронте кластера АТС, подъезжающие к фронту, должны сильно уменьшать свою скорость. Согласно определению [J], широкий движущийся кластер обычно сохраняет среднюю скорость заднего фронта $v_g$, даже если кластер проходит через другие фазы транспортного потока и узкие места. Величина потока сильно падает внутри широкого движущегося кластера.

Как отмечалось в разделе 2.4, эмпирические результаты показывают, что характеристические параметры широких движущихся кластеров не зависят от величины потока на дороге и особенностей узкого места (где и когда кластер возник). Однако эти характеристические параметры зависят от погоды, дорожных условий, конструктивных характеристик АТС, процента длинных машин и т. п. Скорость заднего фронта широкого движущегося кластера $v_g$ в противоположном потоку направлении является характеристическим параметром, так же как и величина выходного потока $q_{\text{out}}$ 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. Благодаря этому эти параметры могут быть предсказаны. Движение заднего фронта широкого движущегося кластера может быть показано на плоскости поток–плотность с помощью прямой, называемой линия J Кернера (рис. 9). Наклон линии J Кернера равен скорости заднего фронта $v_g$, в то время как координата пересечения линии J Кернера с осью абсцисс (при нулевом потоке) отвечает плотности АТС $\rho_{\max}$ в широком движущемся кластере (о линии J Кернера более подробно смотри в п. 2.4.3).

Подчеркнем, что минимум пропускной способности $q_{\text{th}}$ и величина выходного потока из широкого движущегося кластера $q_{\text{out}}$ описывают два *качественно различных свойства* свободного транспортного потока.

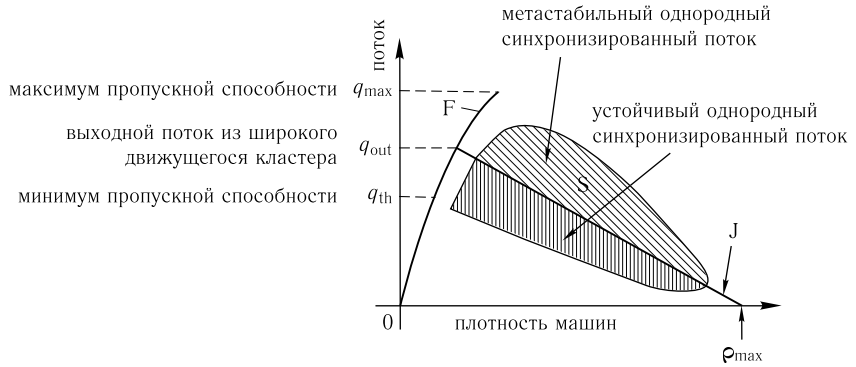



Рис. 9. Три фазы транспортного потока в плоскости поток-плотность в теории трех фаз Кернера. Взято из [1]

Минимум пропускной способности $q_{th}$ характеризует фазовый переход F → S вблизи узкого места, т. е. возникновение плотного потока (traffic breakdown). В свою очередь величина выходного потока из широкого движущегося кластера $q_{out}$ характеризует условия существования таких кластеров, т. е. фазы J. В зависимости от внешних условий, таких как погода, процент длинных машин в потоке и т. п., а также от характеристик узкого места, вблизи которого может произойти фазовый переход F → S, 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 АТС впереди, чем та дистанция, которую они принимают, ускоряясь на заднем фронте широкого движущегося кластера.

### 3.1.8. Синхронизированный транспортный поток — фаза S

В отличие от широких движущихся кластеров в синхронизированном потоке как величина потока $q$, так и скорость АТС 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 АТС сильно уменьшается.

Поскольку синхронизированный поток не имеет характеристического свойства [J] фазы широкого движущегося кластера J, в теории трех фаз предполагается, что гипотетические однородные состояния синхронизированного потока покрывают двумерную область в плоскости поток–плотность (см. заштрихованные области на рис. 9).



### 3.1.9. Фазовый переход S → J

Широкие движущиеся кластеры не возникают в свободном потоке, но они могут возникать в области синхронизированного потока. Этот фазовый переход называется фазовым переходом S → J. Эмпирический пример перехода S → J показан на рис. 5. Таким образом, образование широких движущихся кластеров в свободном потоке наблюдается в результате каскада фазовых переходов F → S → J: сначала область синхронизированного потока возникает внутри свободного потока.

Как было объяснено выше, такой фазовый переход F → S происходит в большинстве случаев вблизи узкого места. Далее внутри синхронизированного потока происходит «сжатие» потока, т. е. плотность АТС возрастает, в то время как их скорость падает. Это сжатие называется «пинч-эффект». В области синхронизированного потока, где происходит пинч-эффект, возникают узкие движущиеся кластеры. Было показано, 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.

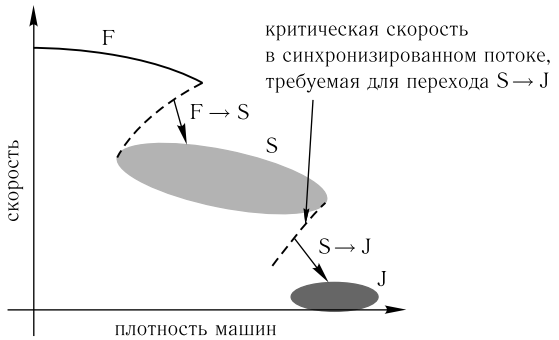



Рис. 10. Двойная Z-характеристика в теории трех фаз, поясняющая каскад фазовых переходов F → S → J. Взято из [1]

Чтобы детальнее проиллюстрировать фазовый переход S → J, следует заметить, что в теории трех фаз линия J делит все однородные состояния синхронизированного потока на две области (рис. 9). Состояния выше линии J Кернера являются метастабильными относительно образования широких движущихся кластеров, в то время как состояния ниже линии

J Кернера являются устойчивыми. Метастабильные состояния синхронизированного потока означают, что относительно малых возникающих возмущений состояние потока остается устойчивым, однако при больших возмущениях в синхронизированном потоке происходит фазовый переход $S \to J$.

Каскад фазовых переходов $F \to S \to J$ можно пояснить на основе двойной Z-характеристики в теории трех фаз (рис. 10). Пунктирная линия между фазой F и фазой S качественно соответствует критической скорости внутри локального возмущения свободного потока, при которой происходит фазовый переход $F \to S$. Другими словами, фазовый переход $F \to S$ происходит внутри локального возмущения свободного потока, в котором скорость меньше, чем критическая скорость (символически этот фазовый переход показан стрелкой между фазой F и S на рис. 10). Пунктирная линия между фазой S и фазой J качественно соответствует критической скорости внутри локального возмущения синхронизированного потока, при которой происходит фазовый переход $S \to J$. Другими словами, фазовый переход $S \to J$ происходит внутри локального возмущения синхронизированного потока, в котором скорость меньше, чем критическая скорость (символически этот фазовый переход показан стрелкой между фазой S и J на рис. 10).

### 3.1.10. Неоднородные пространственно-временные структуры транспортного потока, состоящие из фаз S и J

В эмпирических данных можно наблюдать очень сложные пространственно-временные структуры в плотном транспортном потоке, образовавшиеся в результате фазовых переходов $F \to S$ и $S \to J$.

Неоднородная пространственно-временная структура, которая состоит только из синхронизированного потока, называется *структурой синхронизированного потока* (СП). Когда задний фронт СП фиксирован вблизи узкого места на дороге, а передний фронт не распространяется против потока, такая СП называется *локализованной структурой синхронизированного потока* (ЛСП). Однако во многих случаях передний фронт структуры синхронизированного потока распространяется в направлении против потока. Если при этом задний фронт по-прежнему остается фиксированным вблизи узкого места, то ширина области синхронизированного потока увеличивается. Такая структура называется расширяющейся структурой синхронизированного потока (РСП). Возможна также ситуация, когда задний фронт синхронизированного потока уже не фиксирован вблизи узкого места, а оба фронта синхронизированного потока движутся в направлении против потока. Такая структура называется бегущей, или мигрирующей структурой синхронизированного потока (МСП).



Разница между пространственно-временными структурами, состоящими из только синхронизированного потока, и широкими движущимися кластерами становится особенно ясной, когда РСП или МСП достигают следующего узкого места, расположенного вверх по течению транспортного потока. В этом случае структура синхронизированного потока «захватывается» на этом узком месте (так называемый «catch-effect» в английской терминологии), и возникает новая пространственно-временная структура в транспортном потоке. Напротив, широкий движущийся кластер не захватывается вблизи узкого места, а распространяется дальше против потока, т. е. пробегая через узкое место на дороге. Кроме того, в отличие от широкого движущегося кластера структура синхронизированного потока, даже если она распространяется в виде МСП, не имеет характеристических параметров. В результате скорость заднего фронта МСП может заметно меняться в процессе распространения, и эта скорость может быть разной у разных МСП. Данные особенности структур синхронизированного потока и широких движущихся кластеров вытекают из определения фаз [S] и [J]. Наиболее типичная пространственно-временная структура плотного транспортного потока состоит из обеих фаз S и J. Такая структура называется *общей структурой плотного потока* (ОП).

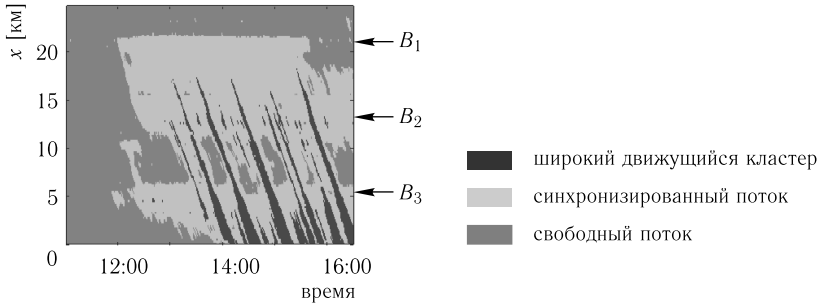



Рис. 11. ЕОП, измеренная на магистрали с тремя узкими местами $B_1$, $B_2$, $B_3$. Взято из [1]

На многих скоростных автомагистралях узкие места, связанные с въездами/выездами, располагаются очень близко друг к другу. Пространственно-временная структура, в которой синхронизированный поток охватывает два и более узких места, называется *единой структурой плотного потока* (ЕП). ЕП может состоять только из синхронизированного потока, тогда она называется ЕСП — единая структура синхронизированного потока. Однако обычно широкие движущиеся кластеры возникают в синхронизированном потоке. В этом случае ЕП называется ЕОП — *единая общая структура плотного потока* (см. рис. 11).

В данной главе были рассмотрены основные качественные положения теории трех фаз Кернера. Эти качественные положения начиная с 2002 года были использованы как теоретический базис при создании целого ряда микроскопических и макроскопических трехфазных моделей транспортного потока (см. ссылки на оригинальные работы в главе 11 книги [1]). В следующем разделе этой главы дается краткий обзор стохастических микроскопических моделей в рамках теории трех фаз и приведены некоторые результаты численных расчетов.

## 3.2. Стохастические модели в рамках теории трех фаз Кернера

Теория трех фаз Кернера является качественной теорией. Различные математические трехфазные модели транспортных потоков были разработаны в последние годы в рамках теории трех фаз. Впервые микроскопическая трехфазная модель, которая может воспроизводить эмпирические свойства перехода к плотному потоку (traffic breakdown) и результирующих пространственно-временных структур, была разработана Б. С. Кернером и С. Л. Кленовым в 2002 году [35]. Несколькими месяцами позже Б. С. Кернер, С. Л. Кленов и Д. Вольф предложили трехфазную модель на основе клеточных автоматов (ККВ-модель) [36]. Позднее были разработаны другие модели транспортного потока в рамках теории трех фаз: Л. Дэвис [37], а также Б. С. Кернер и С. Л. Кленов [38] предложили детерминистические микроскопические модели трех фаз; Х. Ли и М. Шрекенберг с соавторами [39], Р. Янг и К. Ву [40] и К. Гао с соавторами [41] разработали различные трехфазные модели клеточных автоматов (КА); Дж. Лаваль [42] и С. Хугендорн с соавторами [43] разработали макроскопические модели трех фаз. Последние результаты моделирования транспортного потока, проведенные различными научными группами в США, Германии, Голландии, Китае, Южной Корее и Японии в рамках теории трех фаз, можно найти в [44–64].

**3.2.1. Стохастическая микроскопическая трехфазная модель транспортного потока.** В данном пункте кратко рассматривается стохастическая микроскопическая трехфазная модель транспортного потока, предложенная Б. С. Кернером и С. Л. Кленовым в [35, 65]. В этой модели [65] использовалось дискретное время с шагом $\tau$, в то время как перемещение в пространстве предполагалось непрерывным. Ниже будет рассмотрена дискретная версия модели [65], сформулированная в [62], в которой используется достаточно мелкая дискретизация пространства с шагом $\delta x$ (см. также формулировку дискретной версии модели в [63]). При этом в приведенных ниже формулах координата измеряется в единицах $\delta x$, в то время как скорость АТС и ее ускорение измеряются



соответственно в единицах $\delta v = \delta x/\tau$ и $\delta a = \delta v/\tau$, где временной шаг $\tau = 1$ с.

Уравнения движения АТС в дискретной версии [62, 63] стохастической трехфазной модели [65] транспортного потока на скоростной 2-полосной автодороге в приближении идентичных АТС задаются следующими формулами:

$$v_{n+1} = \max(0,\, \min(v_{\text{free}},\, \tilde{v}_{n+1} + \xi_n,\, v_n + a\tau,\, v_{s,n})), \quad x_{n+1} = x_n + v_{n+1}\tau, \tag{1}$$

$$\tilde{v}_{n+1} = \max(0,\, \min(v_{\text{free}},\, v_{s,n},\, v_{c,n})), \tag{2}$$

$$v_{c,n} = \begin{cases} v_n + \Delta_n, & \text{если} \quad g_n \leqslant G_n, \\ v_n + a_n\tau, & \text{если} \quad g_n > G_n, \end{cases} \tag{3}$$

$$\Delta_n = \max(-b_n\tau,\, \min(a_n\tau,\, v_{\ell,n} - v_n)), \tag{4}$$

где индекс $n$ отвечает дискретному времени $t = n\tau$, $n = 0, 1, 2, \ldots$; $x_n$ и $\tilde{v}_n$ — координата и скорость АТС на временном шаге $n$; $v_{\text{free}}$ — максимальная скорость АТС в свободном потоке; $g_n = x_{\ell,n} - x_n - d$ — расстояние до АТС впереди, индекс $\ell$ относится ко всем переменным и функциям, описывающим АТС впереди, $d$ — длина АТС, которая предполагается одинаковой и включает в себя также среднее расстояние между АТС, когда они стоят внутри широкого движущегося кластера; $\tilde{v}_n$ — величина скорости АТС без шумовой компоненты $\xi_n$; $v_{s,n}$ — безопасная скорость, определенная ниже; $a_n \geqslant 0$ и $b_n \geqslant 0$; $a$ — максимальное ускорение; $G_n$ — максимальное расстояние, на котором водитель синхронизует свою скорость со скоростью АТС впереди (так называемая дистанция «синхронизации скорости»);

$$G_n = G(v_n,\, v_{\ell,n}), \tag{5}$$

где функция (5) имеет вид

$$G(u,\, w) = \max(0,\, \lfloor k\tau u + (u - w)ua^{-1} \rfloor), \tag{6}$$

константа $k > 1$, $\lfloor z \rfloor$ означает целую часть действительного числа $z$.

Решения модели (1)–(6) при стационарном и однородном движении АТС (которое отвечает потоку, где АТС находятся на одинаковом расстоянии друг от друга и движутся с постоянной скоростью) показаны на рис. 12 (а) и (б). В соответствии с фундаментальной гипотезой теории трех фаз (рис. 9) фаза синхронизированного потока (буква S) как на плоскости расстояние–скорость (рис. 12 (а)), так и на плоскости поток–плотность (рис. 12 (б)) покрывает двумерную область (заштрихованная область на рис. 12). Для того чтобы пояснить смысл этой двумерной области для стационарных состояний синхронизированного потока, рассмотрим более подробно рисунок 12 (а). Видно, что при некоторой заданной величине

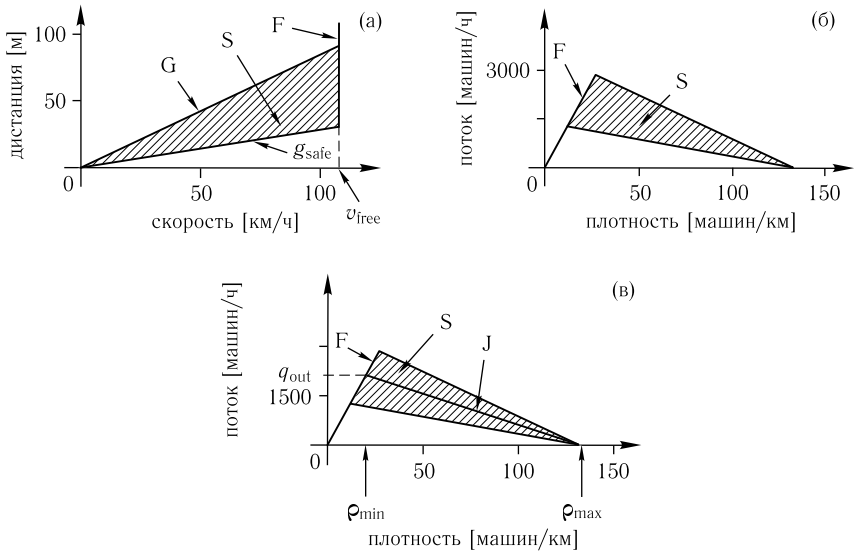



Рис. 12. Стационарные однородные состояния и линия J Кернера в стохастической трехфазной модели: (а, б) свободный поток (F) и синхронизированный поток (заштрихованная двумерная область, обозначенная буквой S) на плоскости расстояние между АТС — скорость (а) и на плоскости поток–плотность (б). На (в) показано соответствие между линией J и однородными стационарными состояниями модели на плоскости поток–плотность, взятыми из (б)

скорости $v < v_{\text{free}}$ в стационарном состоянии существует бесконечное количество расстояний между АТС в диапазоне $g_{\text{safe}} \leqslant g \leqslant G$, где $g$ — расстояние между АТС, $G = G(v, v)$ — дистанция синхронизации скорости (5), взятая при одинаковых скоростях АТС, $g_{\text{safe}}$ — безопасное расстояние между АТС, которое является решением уравнения $v = v_s(g_{\text{safe}})$, где безопасная скорость $v_s$ в стационарном однородном состоянии определяется как $v_s(g) = g/\tau$. Таким образом, стационарные однородные состояния стохастической трехфазной модели (рис. 12) принципиально отличаются от таких же состояний большинства предшествующих моделей, в которых заданной скорости отвечает одно-единственное расстояние между АТС, соответствующее точке на фундаментальной диаграмме.

Из рис. 12 (в) также можно видеть, что линия J Кернера не имеет никакого отношения к фундаментальной диаграмме стационарного однородного плотного потока в моделях, основанных на фундаментальной диаграмме. Действительно, в трехфазной стохастической модели линия J, которая находится из стационарного движения широкого кластера в этой модели, разделяет двумерную область стационарных однородных состояний



синхронизированного потока на две части. Напомним, что в соответствии с теорией трех фаз стационарные однородные состояния выше линии J являются метастабильными к образованию широких движущихся кластеров, а состояния ниже линии J являются устойчивыми к образованию широких кластеров.

Чтобы моделировать случайное время задержки водителя как при ускорении, так и при замедлении в различных транспортных ситуациях величины $a_n$ и $b_n$ в (4) и (5) задаются как случайные функции:

$$a_n = a\theta(P_0 - r_1), \quad b_n = a\theta(P_1 - r_1),$$

$$P_0 = \begin{cases} p_0, & \text{если } S_n \neq 1, \\ 1, & \text{если } S_n = 1, \end{cases} \quad P_1 = \begin{cases} p_1, & \text{если } S_n \neq -1, \\ p_2, & \text{если } S_n = -1, \end{cases}$$

где величины $1 - P_0$ и $1 - P_1$ представляют собой вероятности для времени задержки водителя соответственно при ускорении и при замедлении АТС; $p_0(v)$ и $p_2(v)$ являются функциями скорости АТС, $p_1$ — константа; $r_1 = \text{rand}(0, 1)$ представляет собой случайную величину, равномерно распределенную в интервале от 0 до 1; определение $S_n$ см. ниже.

Случайная компонента $\xi_n$ в формуле (1) описывает случайное замедление или ускорение и применяется в зависимости от того, тормозит ли АТС, или ускоряется, или не меняет свою скорость:

$$\xi_n = \begin{cases} -\xi_b, & \text{если } S_{n+1} = -1, \\ \xi_a, & \text{если } S_{n+1} = 1, \\ \xi^{(0)}, & \text{если } S_{n+1} = 0, \end{cases}$$

где $S_{n+1}$ — это состояние движения АТС в отсутствие случайной компоненты $\xi_n$,

$$S_{n+1} = \begin{cases} -1, & \text{если } \tilde{v}_{n+1} < v_n, \\ 1, & \text{если } \tilde{v}_{n+1} > v_n, \\ 0, & \text{если } \tilde{v}_{n+1} = v_n, \end{cases}$$

$\xi_a$, $\xi_b$ являются случайными источниками соответственно для ускорения и замедления АТС:

$$\xi_a = a^{(a)}\tau\theta(p_a - r), \quad \xi_b = a\tau\theta(p_b - r),$$

а случайный источник

$$\xi^{(0)} = a^{(0)}\tau \begin{cases} -1, & \text{если } r < p^{(0)}, \\ 1, & \text{если } p^{(0)} \leqslant r < 2p^{(0)} \text{ и } v_n > 0, \\ 0 & \text{в остальных случаях} \end{cases}$$

применяется в отсутствие ускорения или замедления и связан с невозможностью точно поддерживать заданную скорость. Величины $p_a$ и $p_b$ являются вероятностями соответственно случайного ускорения или торможения АТС, $p^{(0)}$ и $a^{(0)}$ — константы, $r = \mathrm{rand}(0,\,1)$, ступенчатая функция (Хевисайда) $\theta(z)$ определяется как

$$\theta(z) = \begin{cases} 0 & \text{при } z < 0, \\ 1 & \text{при } z \geqslant 0. \end{cases}$$

Безопасная скорость $v_{s,n}$ определяется следующим образом:

$$v_{s,n} = \min\left(v_n^{(\mathrm{safe})},\, v_\ell^{(a)} + \frac{g_n}{\tau}\right),$$

где

$$v_\ell^{(a)} = \max\left(0,\, \min\left(v_{\ell,n}^{(\mathrm{safe})},\, v_{\ell,n},\, \frac{g_{\ell,n}}{\tau}\right) - a\tau\right)$$

— это так называемая «ожидаемая» (прогнозируемая) скорость АТС впереди, функция $v_n^{(\mathrm{safe})} = \lfloor v^{(\mathrm{safe})}(g_n,\, v_{\ell,n})\rfloor$ задается безопасной скоростью $v^{(\mathrm{safe})}(g_n,\, v_{\ell,n})$, которая была предложена в модели С. Крауссa с соавторами [66] в 1997 году и которая в свою очередь является решением уравнения П. Гиппса [67]:

$$v^{(\mathrm{safe})}\tau + X_d(v^{(\mathrm{safe})}) = g_n + X_d(v_{\ell,n}),$$

где $X_d(u)$ — тормозной путь, проходимый АТС, движущимся с первоначальной скоростью $u$ и тормозящим с постоянным ускорением $b$ вплоть до полной остановки. В модели с дискретным временем этот путь дается формулой $X_d(u) = b\tau^2 \cdot (\alpha\beta + 0{,}5(\alpha - 1)\alpha)$, $\alpha$ и $\beta$ — соответственно целая и дробная части величины $u/b\tau$.

В рассматриваемой модели автодороги с двумя полосами смена полосы АТС происходит независимо от того, находятся ли эти АТС вдали или вблизи неоднородностей дороги, связанной с въездом/выездом. АТС меняет полосу, если некоторые необходимые условия для перехода с правой полосы на левую ($R \to L$) или с левой полосы на правую ($L \to R$) выполняются совместно с условиями безопасности при смене полосы. Необходимые для смены полосы условия имеют вид

$$R \to L\colon\ v_n^+ \geqslant v_{\ell,n} + \delta_1 \ \text{ и } \ v_n \geqslant v_{\ell,n},$$
$$L \to R\colon\ v_n^+ \geqslant v_{\ell,n} + \delta_1 \ \text{ или } \ v_n^+ > v_n + \delta_1.$$

Условия безопасности при смене полосы имеют вид

$$g_n^+ > \min(v_n\tau,\, G_n^+), \quad g_n^- > \min(v_n^-\tau,\, G_n^-), \tag{7}$$

где

$$G_n^+ = G(v_n,\, v_n^+), \quad G_n^- = G(v_n^-,\, v_n),$$



верхние индексы «+» и «−» относятся соответственно к АТС впереди и позади на соседней полосе.

Если условия (7) не выполняются, то используются более жесткие условия для «вдавливания» АТС на соседнюю полосу:

$$x_n^+ - x_n^- - d > g_{\mathrm{target}}^{(\min)}, \quad \text{где } \ g_{\mathrm{target}}^{(\min)} = \lfloor \lambda v_n^+ + d \rfloor. \tag{8}$$

В дополнение к (8) используется условие, что АТС проходит среднюю точку $x_n^{(m)} = \lfloor (x_n^+ + x_n^-)/2 \rfloor$ между двумя соседними АТС на соседней полосе, т. е. следующие условия выполняются:

$$x_{n-1} < x_{n-1}^{(m)} \ \text{ и } \ x_n \geqslant x_n^{(m)} \quad \text{или} \quad x_{n-1} \geqslant x_{n-1}^{(m)} \ \text{ и } \ x_n < x_n^{(m)}.$$

Если условия для смены полосы выполняются, АТС меняет полосу с вероятностью $p_c < 1$ на текущем шаге.

После смены полосы скорость $v_n$ устанавливается равной $\widehat{v}_n = \min(v_n^+,\, v_n + \Delta v^{(1)})$, что описывает изменение скорости после маневра по смене полосы. После смены полосы координата АТС не меняется, если выполнены условия (7), и она устанавливается равной $x_n = x_n^{(m)}$, если выполнены условия (8). Величины $\Delta v^{(1)}$, $p_c$, $\delta_1$ и $\lambda$ являются константами. Более подробно об использованных параметрах модели, условиях смены полос и модели поведения водителя на въезде/выезде со скоростной автодороги можно прочитать в главе 11 книги [1].

**3.2.2. Моделирование свойств пространственно-временных структур в транспортном потоке вблизи въезда на скоростную автомагистраль.** Предложенная Б. С. Кернером и С. Л. Кленовым модель была использована для теоретических исследований транспортного потока, а также для многих транспортных приложений, представленных в книгах [1, 2] и в статьях [54–57, 62, 63]. В частности, недавними теоретическими исследованиями транспортных потоков, выполненными с использованием этой модели, являются созданные Б. С. Кернером теории пространственно-временных структур в транспортном потоке очень высокой плотности на сильно перегруженной автомагистрали [54], а также теория пространственно-временных структур плотного потока на движущемся «бутылочном горле» и на многополосной автомагистрали [62, 63]. Недавно Б. С. Кернер применил данную модель для изучения перехода к плотному потоку на светофоре в городских сетях [68], а также для исследования предложенного в [69] принципа минимизации вероятности возникновения заторов для оптимизации и управления сложными городскими и межгородскими транспортными сетями.

Чтобы проиллюстрировать нелинейные численные результаты, получаемые с помощью предложенной Б. С. Кернером и С. Л. Кленовым модели, рассмотрим некоторые свойства пространственно-временных структур

транспортного потока, образующихся вблизи узкого места, связанного с въездом на скоростную автодорогу, которые были получены в рамках данной модели в 2002–2003 годах [35, 65].

На рис. 13 приведены результаты расчета пространственно-временных структур, возникающих вблизи въезда на автодорогу, и областей их существования (диаграммы) на плоскости, где на осях отмечены поток по основной дороге $q_{\text{in}}$, приведенный на одну из двух полос дороги, и поток со стороны въезда на дорогу $q_{\text{on}}$.

На рис. 13 (а) показана диаграмма этих пространственно-временных структур. По оси абсцисс этой диаграммы откладывается поток АТС $q_{\text{on}}$, по оси ординат указан поток АТС по основной дороге $q_{\text{in}}$. Граница $F_S^{(B)}$ на этой диаграмме разделяет свободный поток влево от границы от структур плотного потока, возникающих вблизи въезда АТС. На границе $F_S^{(B)}$ пространственно-временные структуры плотного потока возникают спонтанно. Граница $S_J^{(B)}$ разделяет структуры синхронизированного 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 $q_{\text{in}}$ и маленьких $q_{\text{on}}$ возникает *мигрирующая структура синхронизированного потока* — МСП (рис. 13 (г)). При увеличении $q_{\text{on}}$ МСП превращается в *расширяющуюся структуру синхронизированного потока* — РСП (рис. 13 (б)).

Если уменьшать $q_{\text{in}}$ и увеличивать $q_{\text{on}}$, то на границе, обозначенной буквой W, РСП превращается в *локализованную структуру синхронизированного потока* — ЛСП (рис. 13 (в)).

Укажем еще некоторые особенности общей структуры плотного потока — ОП, которая возникает выше границы $S_J^{(B)}$ на диаграмме (рис. 13 (а)). Существует граница, обозначенная буквой G внутри области ОП. Слева от этой границы, после того как широкий кластер возникает в синхронизированном потоке, новые широкие кластеры больше не формируются (рис. 13 (ж)), и возникает *рассасывающаяся общая структура плотного потока* (РОП). Правее от границы G возникает общая структура, в которой непрерывно рождаются новые кластеры внутри синхронизированного потока (рис. 13 (д)). Однако если уменьшать поток по дороге $q_{\text{in}}$, оставаясь в области ОП, и перейти к значениям $q_{\text{in}}$ меньше, чем $q_{\text{out}}$ (смысл потока $q_{\text{out}}$ объяснен в п. 2.4.3), то широкие движущиеся кластеры, спонтанно возникающие в синхронизированном потоке ОП, постепенно рассасываются, распространяясь против течения. В результате возникает ОП, показанная на рис. 13 (е). Другая особенность общей структуры состоит в том, что если поток из въезда $q_{\text{on}}$ дальше увеличивать, то возникает некоторый эффект насыщения, связанный с возникновением плотного потока на дороге, ведущей к въезду АТС. Этот эффект насыщения свя-



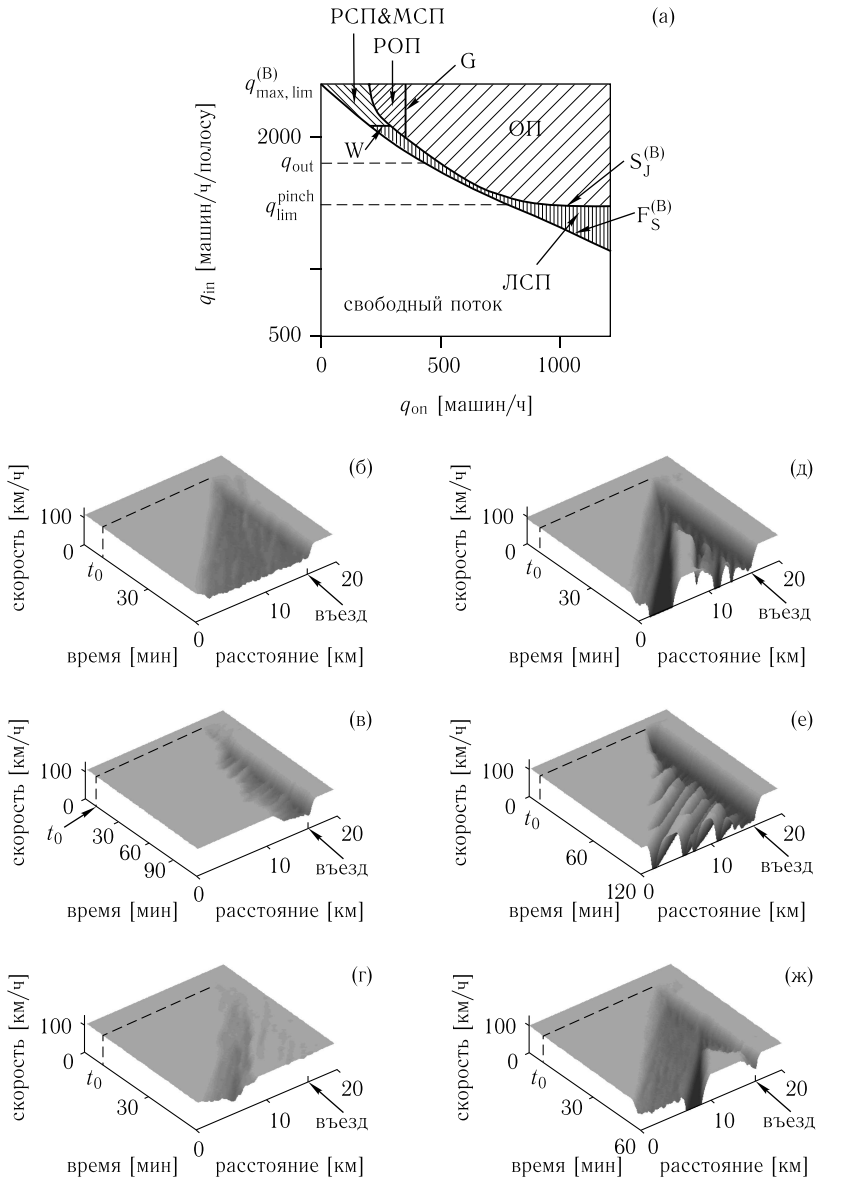



Рис. 13. Диаграмма пространственно-временных структур транспортного потока вблизи изолированного въезда (а) и соответствующие типы пространственно-временных структур (б–ж), относящиеся к диаграмме (а): СП (б–г) и ОП (д–ж). Взято из [2]

зан с тем, что поток внутри области синхронизированного потока в ОП достигает своего предельного минимального значения, обозначенного на рис. 13 (а) как $q_{\text{lim}}^{(\text{pinch})}$.

Распределение скорости и потока внутри ОП на рис. 13 (д) как функция времени при разных координатах вдоль дороги показано на рис. 14. Можно видеть, что сначала на 16,5 км возникает фазовый переход F → S из свободного в синхронизированный поток. Через 2 км против течения ($x = 14{,}5$ км) можно видеть развивающиеся узкие кластеры внутри синхронизированного потока. По мере распространения этих кластеров против течения амплитуда кластеров увеличивается и их ширина возрастает. В результате нарастания кластеров возникает последовательность широких движущихся кластеров ($x = 8$ км), которые обладают характеристическими параметрами, описанными в п. 2.4.2. В частности, распространение заднего фронта этих кластеров соответствует линии J Кернера, описанной в п. 2.4.3.

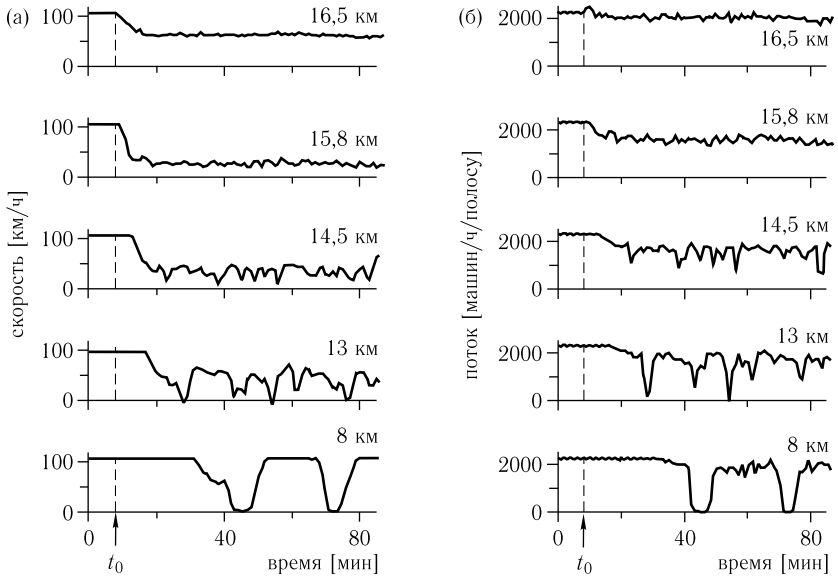



Рис. 14. Данные виртуальных детекторов, отвечающие ОП на рис. 13 (д). Въезд на скоростную дорогу отвечает координате 16 км. Данные усреднены за 1 минуту. Взято из [2]

Более подробно свойства пространственно-временных структур, полученные в результате моделирования, рассматриваются в [1,2] и [54,62,63].

**3.2.3. Трехфазная модель клеточных автоматов для транспортного потока (ККВ-модель).** Предложенная Кернером—Кленовым—Вольфом (ККВ) модель транспортного потока [36] была первой моделью трех фаз на основе клеточных автоматов, в рамках которой оказалось возможным смоделировать и объяснить эмпирические свойства перехода от свободного к плотному потоку и возникающих в результате этого перехода пространственно-временных структур транспортного потока [1,2].



Правила для движения АТС в ККВ-модели состоят в следующем [36]:

$$v_{n+1} = \max(0,\ \min(v_{\text{free}},\ \tilde{v}_{n+1} + a\tau\eta_n,\ v_n + a\tau,\ v_{s,n})), \tag{9}$$

$$x_{n+1} = x_n + v_{n+1}\tau, \tag{10}$$

$$\tilde{v}_{n+1} = \max(0,\ \min(v_{\text{free}},\ v_{c,n},\ v_{s,n})), \tag{11}$$

$$v_{c,n} = \begin{cases} v_n + a\tau, & \text{если} \quad g_n > G_n, \\ v_n + a\tau\,\text{sgn}(v_{\ell,n} - v_n), & \text{если} \quad g_n \leqslant G_n, \end{cases} \tag{12}$$

где $v_{\ell,n}$ — скорость предшествующего АТС,

$$\text{sgn}(x) = \begin{cases} 1 & \text{для } x > 0, \\ 0 & \text{для } x = 0, \\ -1 & \text{для } x < 0. \end{cases}$$

Дистанция «синхронизации скорости» $G_n$ выбирается следующим образом

$$G_n = kv_n\tau, \tag{13}$$

где постоянная $k > 1$. Так же как и в классической модели клеточных автоматов Нагеля—Шрекенберга, безопасная скорость выбиралась равной

$$v_{s,n} = \frac{g_n}{\tau}. \tag{14}$$

Случайное торможение и случайное ускорение моделировались добавлением к скорости АТС случайного слагаемого $a\tau\eta_n$, где

$$\eta_n = \begin{cases} -1, & \text{если} \quad r < p_b, \\ 1, & \text{если} \quad p_b \leqslant r < p_b + p_a, \\ 0 & \text{в остальных случаях}, \end{cases} \tag{15}$$

где $p_a$ и $p_b$ — соответственно вероятности случайного ускорения и случайного торможения, которые зависели от текущей скорости АТС, при этом $p_a + p_b < 1$; $r = \text{rand}(0,\ 1)$.

Наиболее важный новый результат теории трех фаз транспортного потока, полученный с помощью ККВ-модели, — это вычисление вероятности фазового перехода от свободного к синхронизированному потоку (переход F → S) в зависимости от величины транспортного потока $q$.

**3.2.4. Новая трехфазная модель клеточных автоматов для транспортного потока (ККШ-модель).** Как уже упоминалось, ККВ-модель была первой моделью клеточных автоматов, которая позволила описать переход к синхронизированному потоку и результирующие пространственно-временные структуры плотного потока, найденные в реальных данных. Однако моделирование эффекта «переускорения» в ККВ-модели осуществлялось неявным (модельным) образом, с помощью случайного ускорения ($\eta_n = 1$ в (15)). Это, в частности, послужило одной из причин для Кернера, Кленова и Шрекенберга предложить новую трехфазную модель клеточных автоматов [71] (ККШ-модель). С одной стороны, ККШ-модель способна описать переход от свободного к синхронизированному потоку (переход F $\to$ S) и результирующие пространственно-временные структуры плотного потока. С другой стороны, эта модель с помощью простых правил явным образом моделирует физику движения АТС в транспортном потоке в соответствии с положениями теории трех фаз Кернера.

ККШ-модель состоит из правил «ускорение», «торможение», «рандомизация» и «движение» из классической модели клеточных автоматов Нагеля—Шрекенберга [70], а также из правил «переускорение из-за перехода на быструю полосу» [1)], «сравнение расстояния с дистанцией синхронизации скорости» и «адаптация скорости в пределах дистанции синхронизации» из теории трех фаз Кернера [1, 2].

Правила движения АТС в ККШ-модели для дороги из двух полос следующие [71]:

(a) «Переускорение из-за перехода на быструю полосу»: переход на более быструю полосу (с целью обгона) выполняется с вероятностью $p_c$, когда выполнены следующие условия «намерения» (16), (17) и безопасности (18):

$$R \to L: \quad v_n^+ \geqslant v_{\ell,n} + \delta \ \text{ и } \ v_n \geqslant v_{\ell,n}, \tag{16}$$

$$L \to R: \quad v_n^+ \geqslant v_{\ell,n} + \delta \ \text{ или } \ v_n^+ \geqslant v_n + \delta, \tag{17}$$

$$g_n^+ \geqslant \min(v_n, g_c) \ \text{ или } \ g_n^- \geqslant \min(v_n^-, g_c). \tag{18}$$

(b) «Сравнение расстояния с дистанцией синхронизации скорости»: если $g_n \leqslant G(v_n)$, то выполняется правило (c) и пропускается правило (d); напротив, если $g_n > G(v_n)$, то выполняется правило (d) и пропускается правило (c).

(c) «Адаптация скорости в пределах дистанции синхронизации»:

$$v_{n+1}^{(1)} = v_n + \operatorname{sgn}(v_{\ell,n} - v_n). \tag{19}$$

[1)] В качестве «быстрой» полосы рассматривается соседняя для АТС полоса дороги, на которой скорость выше, чем на текущей полосе.



(d) «Ускорение»:

$$v_{n+1}^{(1)} = \min(v_n + 1, v_{\text{free}}). \tag{20}$$

(e) «Торможение»:

$$v_{n+1}^{(2)} = \min(v_{n+1}^{(1)}, g_n). \tag{21}$$

(f) «Рандомизация»: с вероятностью $p$

$$v_{n+1} = \begin{cases} \max(v_{n+1}^{(2)} - 1, 0), & \text{если } r < p, \\ v_{n+1}^{(2)}, & \text{если } r \geqslant p, \end{cases} \tag{22}$$

где $r = \text{rand}(0, 1)$.

(g) «Движение»:

$$x_{n+1} = x_n + v_{n+1}. \tag{23}$$

В (16)–(23) индекс $n = 0, 1, 2, \ldots$ равен номеру временного шага, $x_n$ и $v_n$ являются координатой и скоростью АТС, $v_{\text{free}}$ — это максимальная скорость, $g_n = x_{\ell,n} - x_n - d$ — это расстояние до АТС впереди, $d$ — длина АТС;

$$G(v_n) = kv_n \tag{24}$$

— это дистанция синхронизации скорости; нижний индекс $\ell$ маркирует переменные, относящиеся к АТС впереди; обозначения $R \to L$ и $L \to R$ относятся соответственно к переходу с правой полосы на левую и, наоборот, с левой на правую; в правилах смены полосы (16)–(18) верхние индексы «+» и «−» обозначают переменные и функции, относящиеся соответственно к АТС впереди и позади на соседней полосе, между которыми окажется рассматриваемая АТС после смены полосы, в частности, $g_n^+$ обозначает расстояние между рассматриваемой АТС и АТС впереди на соседней полосе, $g_n^-$ обозначает расстояние между рассматриваемой АТС и АТС позади на соседней полосе; в неравенствах (16), (17) скорость $v_n^+$ ($v_{\ell,n}$) устанавливалась $\infty$, если расстояние $g_n^+$ ($g_n$) превышало некоторое достаточно большое расстояние $L_a$; величины $v_{\text{free}}$, $L_a$, $\delta$, $k$ и $g_c$ являются константами.

В [71] было показано, что эти несколько правил ККШ-модели могут описывать фундаментальные эмпирические свойства перехода от свободного к синхронизированному потоку (traffic breakdown) и свойства пропускной способности автомагистрали, найденные в реальных данных, измеренных за многие годы в разных странах, в частности: характеристики синхронизированного потока, существование как спонтанного, так и индуцированного перехода к синхронизированному потоку (traffic breakdown) на одном и том же узком месте (бутылочном горле), а также связанные с этим вероятностные характеристики перехода к синхронизированному потоку и пропускной способности.

## 3.3. Применение теории трех фаз Кернера для интеллектуальных транспортных технологий

Б. С. Кернер с сотрудниками предложил и частично внедрил в эксплуатацию целый ряд новых методов интеллектуальных транспортных технологий. Одним из внедренных и уже установленных на скоростных автодорогах применений теории трех фаз является метод ASDA/FOTO. Метод ASDA/FOTO функционирует в работающей он-лайн системе регулирования транспортных потоков, где на основе измерений выделяются фазы S и J в плотном транспортном потоке. Распознавание, отслеживание и прогнозирование положений фаз S и J осуществляется на основе методов теории трех фаз. Метод ASDA/FOTO реализован в компьютерной системе, способной быстро и эффективно обрабатывать большие объемы данных, измеренных датчиками в сети скоростных автомагистралей (см. примеры из трех стран на рис. 15).

Дальнейшее развитие приложений теории трех фаз Кернера связано с разработкой и усовершенствованием моделей для транспортных симуляторов, методов регулирования въездного потока на автомагистраль (ANCONA), методов коллективного регулирования транспортных потоков, системы автоматического ассистента водителя и методов детектирования состояния транспортного потока, описанных в монографиях [1, 2].

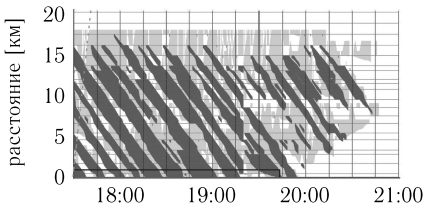
Одними из последних применений теории трех фаз являются объяснение физики образования больших заторов в городских транспортных сетях [68], а также исследование по оптимизации и управлению сложными городскими и межгородскими транспортными сетями, базирующееся на предложенном недавно Б. С. Кернером [69] принципе минимизации вероятности возникновения заторов.

## Литература

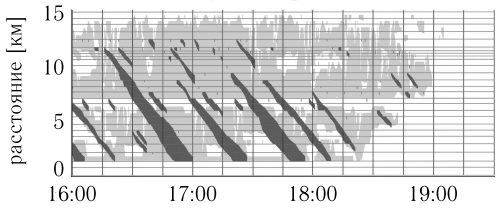

1. *Kerner B. S.* Introduction to Modern Traffic Flow Theory and Control. Berlin: Springer, 2009.
2. *Kerner B. S.* The Physics of Traffic. Berlin: Springer, 2004.
3. *Kerner B. S.* Experimental Properties of Self-Organization in Traffic Flow // Physical Review Letters. 1998. V. 81. P. 3797–3800.
4. *Kerner B. S.* The Physics of Traffic // Physics World. 1999. V. 12, № 8. P. 25–30.
5. *Kerner B. S.* Congested Traffic Flow: Observations and Theory // Transportation Research Record. 1999. V. 1678. P. 160–167.
6. *Kerner B. S.* Theory of Congested Traffic Flow: Self-Organization without Bottlenecks // In: Transportation and Traffic Theory, edited by A. Ceder. London: Elsevier Science, 1999. P. 147–171.
7. *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.




(а) Автодорога A5-Север (14 июня 2006) в Германии




(б) Автодорога M-42 (11 января 2008) в Великобритании




(в) Автодорога I405-Юг (04 марта 2003) в США

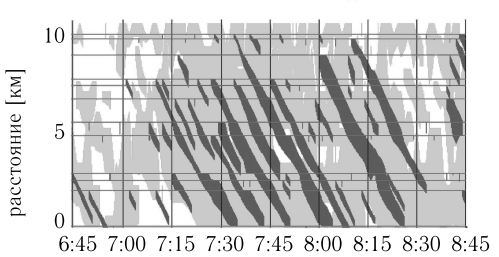



Рис. 15. Пространственно-временная структура транспортного потока, полученная методом ASDA/FOTO в трех странах. Взято из [1]


8. *Richards P. I.* Shock Waves on the Highway // Oper. Res. 1956.V. 4. P. 42–51.
9. *Уизем Дж.* Линейные и нелинейные волны. М.: Мир, 1977.
10. *Newell G. F.* Applications of Queuing Theory. London: Chapman and Hall, 1982.
11. *Newell G. F.* Nonlinear effects in the dynamics of car-following // Oper. Res. 1961. V. 9. P. 209–229.
12. *Newell G. F.* A moving bottleneck // Transp. Res. B. 1998. V. 32. P. 531–537.
13. *Daganzo C. F.* The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // Transp. Res. B. 1994. V. 28. № 4. P. 269–287.

14. *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.
15. *Gazis D. C.*, *Herman R.*, *Potts R. B.* Car following theory of steady state traffic flow // Oper. Res. 1959. V. 7. P. 499–505.
16. *Gazis D. C.*, *Herman R.*, *Rothery R. W.* Nonlinear follow the leader models of traffic flow // Oper. Res. 1961. V. 9. P. 545–567.
17. *Gazis D. C.* Traffic Theory. Berlin: Springer, 2002.
18. *May A. D.* Traffic Flow Fundamentals. Englewood Cliffs, NJ: Prentice-Hall, 1990.
19. *Leutzbach W.* Introduction to the Theory of Traffic Flow. Berlin: Springer, 1988.
20. *Daganzo C. F.* Fundamentals of Transportation and Traffic Operations. New York: Elsevier Science Inc., 1997.
21. *Muñoz J. C.*, *Daganzo C. F.* Traffic and Transportation Theory. Editor M. A. P. Taylor. Oxford: Pergamon, 2002. P. 441–462.
22. *Gartner N. H.*, *Messer C. J.*, *Rathi A.* (editors) Traffic Flow Theory. Washington, DC: Transportation Research Board, 2001.
23. *Chowdhury D.*, *Santen L.*, *Schadschneider A.* Statistical physics of vehicular traffic and some related systems // Phys. Rep. 2000. V. 329. P. 199–329.
24. *Helbing D.* Traffic and related self-driven many particle systems // Rev. Mod. Phys. 2001. V. 73. P. 1067–1141.
25. *Nagatani T.* The physics of traffic jams // Rep. Prog. Phys. 2002. V. 65. P. 1331–1386.
26. *Nagel K.*, *Wagner P.*, *Woesler R.* Still flowing: Approaches to traffic flow and traffic jam modelling // Oper. Res., 2003. V. 51. P. 681–716.
27. *Mahnke R.*, *Kaupuzs J.*, *Lubashevsky I.* Probabilistic description of traffic flow // Phys. Rep. 2005. V. 408. P. 1–130.
28. *Rakha H.*, *Pasumarthy P.*, *Adjerid S.* A simplified behavioral vehicle longitudinal motion model // Transp. Lett. 2009. V. 1. P. 95–110.
29. *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.
30. *Kerner B. S.*, *Klenov S. L.*, *Hiller A.*, *Rehborn H.* Microscopic features of moving traffic jams // Phys. Rev. E. 2007. V. 73. 046107.
31. *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.
32. *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.
33. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120. № 3–4. P. 627–658.
34. *Maerivoet S.*, *De Moor B.* Cellular automata models of road traffic // Phys. Rep. 2005. V. 419. № 1. P. 1–64.
35. *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.
36. *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.



37. *Davis L. C.* Multilane simulations of traffic phases // Phys. Rev. E, 2004. V. 69. 016108.
38. *Kerner B. S.*, *Klenov S. L.* Deterministic microscopic three-phase traffic flow models // J. Phys. A: Math. Gen. 2006. V. 39. P. 1775–1809.
39. *Lee H. K.*, *Barlovic R.*, *Schreckenberg M.*, *Kim D.* Mechanical Restriction versus Human Overreaction Triggering Congested Traffic States // Phys. Rev. Lett. 2004. V. 92. 238702.
40. *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.
41. *Gao K.*, *Jiang R.*, *Hu S.-X.*, *Wang B.-H.*, *Wu Q. S.* Cellular-automaton model with velocity adaptation in the framework of Kerner's three-phase traffic theory. Phys. Rev. E. 2007. V. 76. 026105.
42. *Laval J. A.* Linking synchronized flow and kinematic waves // In: Traffic and Granular Flow' 05. Editors A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, D. E. Wolf. 2007. P. 521–526.
43. *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.
44. *Davis L. C.* Controlling traffic flow near the transition to the synchronous flow phase // Physica A. 2006. V. 368. P. 541–550.
45. *Davis L. C.* Effect of cooperative merging on the synchronous flow phase of traffic // Physica A. 2006. V. 361. P. 606–618.
46. *Davis L. C.* Driver Choice Compared to Controlled Diversion for a Freeway Double On-Ramp in the Framework of Three-Phase Traffic Theory // Physica A. 2006. V. 379. P. 274–290.
47. *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.
48. *Jiang R.*, *Wu Q.-S.* Toward an improvement over Kerner—Klenov—Wolf three-phase cellular automaton model // Phys. Rev. E. 2005. V. 72. 067103.
49. *Jiang R.*, *Wu Q.-S.* Dangerous situations in a synchronized flow model // Physica A. 2007. V. 377. P. 633–640.
50. *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.
51. *Pottmeier A.*, *Thiemann C.*, *Schadschneider A.*, *Schreckenberg M.* Mechanical Restriction Versus Human Overreaction: Accident Avoidance and Two-Lane Traffic Simulations // In: Traffic and Granular Flow'05. Editors A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, D. E. Wolf. Berlin: Springer, 2007. P. 503–508.
52. *Siebel F.*, *Mauser W.* Synchronized flow and wide moving jams from balanced vehicular traffic // Phys. Rev. E. 2006. V. 73. 066108.
53. *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.

54. *Kerner B.S.* A theory of traffic congestion at heavy bottlenecks // J. Phys. A: Math. Theor. 2008. V. 41.
55. *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.
56. *Davis L.C.* Realizing Wardrop equilibria with real-time traffic information // Physica A. 2009. V. 388. P. 4459–4474.
57. *Davis L.C.* Predicting travel time to limit congestion at a highway bottleneck // Physica A. 2010. V. 389. P. 3588–3599.
58. *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.
59. *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.
60. *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. 2011. V. 7. P. 127–140.
61. *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.
62. *Kerner B.S., Klenov S.L.* Phase transitions in traffic flow on multi-lane roads // Phys. Rev. E. 2009. V. 80. 056101.
63. *Kerner B.S., Klenov S.L.* A theory of traffic congestion on moving bottlenecks // J. Phys. A: Math. Theor. 2010. V. 43. 42510.
64. *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.
65. *Kerner B.S., Klenov S.L.* Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks // Phys. Rev. E. 2003. V. 68. 036130.
66. *Krauß S., Wagner P., Gawron C.* Metastable states in a microscopic model of traffic flow // Phys. Rev. E. 1997. V. 55. P. 5597–5602.
67. *Gipps P.G.* A behavioural car-following model for computer simulation // Transportation Research B. 1981. V. 15. P. 105–111.
68. *Kerner B.S.* Physics of traffic gridlock in a city // Phys. Rev. E. 2011. V. 84. 045102(R).
69. *Kerner B.S.* Optimum principle for a vehicular traffic network: minimum probability of congestion // J. Phys. A: Math. Theor. 2011. V. 44. 092001.
70. *Nagel K., Schreckenberg M.* A cellular automation model for freeway traffic // J. Phys. (France) I. 1992. V. 2. P. 2221–2229.
71. *Kerner B.S., Klenov S.L., Schreckenberg M.* Simple cellular automaton model for traffic breakdown, highway capacity, and synchronized flow // Phys. Rev. E. 2011. V. 84. 046110.

# Приложения

**М. Л. Бланк**

# 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 установлены явные формулы для некоторых связанных с ними статистик. В частности, получена так называемая фундаментальная диаграмма, выражающая зависимость средней скорости движения от плотности частиц.

## Введение

Одним из естественных способов математического моделирования транспортных потоков является их реализация в виде процессов с запретами (Exclusion Processes). Последние представляют собой системы частиц, совершающих случайные блуждания и взаимодействующих по закону «исключенного объема» (hard core). Впервые простейшую решеточную модель этого типа предложил Ф. Спитцер в 1970 г., и с тех пор подобные, на первый взгляд примитивные, модели нашли весьма широкое применение в самых различных областях, начиная с моделей транспортных потоков [3, 4, 15, 18, 22], синтеза протеинов и молекулярных моторов в биологии, роста случайных поверхностей в физике (см. [11, 23]) и до анализа диаграмм Юнга в теории представлений [14].

Качественно с точки зрения порядка взаимодействий частиц имеется два типа процессов с запретами: асинхронные и синхронные. В первом случае не более одной частицы может сдвинуться в данный момент времени, а во втором — все частицы двигаются одновременно. Последний вариант приводит к необходимости анализа кратных одновременных взаимодействий, однако с точки зрения приложений представляется более естественным.

В работе будет изучено несколько простых моделей транспортных потоков этого типа (как решеточных, так и с непрерывным пространством) и установлены явные формулы для некоторых связанных с ними статистик. В частности, получена так называемая фундаментальная диаграмма, выражающая зависимость средней скорости движения от плотности частиц.

Работа была частично поддержана грантами РФФИ.



Читатель может найти серьезные обзоры по «физическим» постановкам задач по этой тематике в [13, 18].

Мой собственный интерес к такого рода задачам возник в связи со следующим практическим наблюдением. Иногда быстрее идти против движения в медленно двигающейся толпе людей (например, в переходе метро), чем по движению. Стандартная вероятностная модель диффузии частицы по или против потока явно противоречит этому наблюдению, что указывает на специальную (неслучайную) внутреннюю структуру потока в рассматриваемом случае. Одной из целей настоящей работы является обсуждение того, как подобная структура возникает из произвольных (случайных) начальных конфигураций частиц.

Мы начнем исследование с простейшей детерминированной модели на целочисленной решетке (описанной в разделе 1). Эта модель может быть полностью изучена элементарными средствами, и мы приводим ее в основном в педагогических целях. Однако даже в рамках такой модели отбрасывание условия регулярности начальной конфигурации частиц приводит к необходимости значительно более сложного математического анализа (см. [3,4]). 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 2 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. Недавно в работе [8] был предложен принципиально новый подход, позволяющий получать точные результаты о статистиках процессов с запретами с дискретным временем для целого ряда стохастических моделей. Этот подход мы вкратце опишем в двух последних разделах. Кроме того, в теореме 11 мы изучим влияние препятствий (например, светофоров) на статистики стохастических процессов с запретами.

Все известные подходы к анализу решеточных систем существенно используют комбинаторную структуру пространства конфигураций частиц (отметим, например, идею двойственности «частица–пустая позиция», используемую в разделе 1). 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 получены недавно в [6], где была разработана оригинальная техника, позволяющая изучать эргодические (статистические) свойства таких процессов. Главной технической новинкой здесь является *метод динамического каплинга* (описанный в разделе 5). 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.

## 1. Простейшая модель на целочисленной решетке

Начнем с простейшей одномерной модели транспортного потока, введенной в [22]. 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. [13, 15, 16, 18, 21, 22] и дальнейшие ссылки в них). Наиболее интересным явлением, обнаруженным в данных работах, является нетривиальная зависимость средней скорости движения частиц от их плотности $\mathrm{V}(\rho)$, равная 1 при $\rho \in \left[0,\ \frac{1}{2}\right]$ и $\frac{1}{\rho} - 1$ при $\rho \in \left(\frac{1}{2},\ 1\right]$. Ниже мы выведем этот результат при помощи техники двойственных отображений для произвольных конечных и бесконечных решеток и начальных конфигураций. Кроме того, мы дадим полное описание предельных множеств (соответствующих стационарным транспортным потокам) и точную оценку длины переходного периода. Физическая интерпретация описанного результата — это наличие фазового перехода «газ–жидкость» от свободного движения частиц (при малой плотности) к постоянному наличию транспортных пробок (при большой плотности).

С точки зрения теории динамических систем описанная выше модель может быть представлена следующим образом. Пусть $X = \{0,\ 1\}^{\mathbb{Z}}$ — множество всех возможных конфигураций — бинарных последовательностей $x = x(i),\ \ i \in \mathbb{Z}$, единицы в которой соответствуют частицам, а нули — незанятым позициям на решетке. Рассмотрим отображение $T\colon X \to X$:

$$Tx(i) := \begin{cases} 1, \text{если } x(i) = 0,\ x(i-1) = 1 \text{ или } x(i) = x(i+1) = 1, \\ 0 \ \text{в остальных случаях.} \end{cases}$$

Группу из (более одной) последовательно стоящих частиц мы назовем *кластером*; а частицу в позиции $i$ (т. е. $x(i) = 1$), позиция после которой не занята (т. е. $x(i+1) = 0$), назовем *свободной*. Будем называть конфигурацию $x \in X$ регулярной, если имеется число $\rho = \rho(X)$ (плотность частиц) и монотонная функция $\varphi(N) \to 0$ при $N \to \infty$, 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) $\mathrm{V}(x)$ понимается среднее значение (если оно корректно определено) перемещения частиц в конфигурации $x$ во время следующей итерации отображения $T$. Отметим, что в разделе 3 будет введено и изучено более тонкое понятие средней скорости индивидуальной частицы.



**Теорема 1.** *Для любой регулярной начальной конфигурации $x \in X$ с плотностью $\rho \neq \frac{1}{2}$ через не более чем $t_c(x) = \frac{1}{2}\varphi^{-1}\left(\left|\frac{1}{2} - \rho(x)\right|\right)$ итераций отображения $T$ средняя скорость станет равна $V = \min\left(1,\ \frac{1}{\rho} - 1\right)$, и при любом $t \geqslant t_c(x)$ выполняется следующая альтернатива: конфигурация $T^t x$ состоит либо только из свободных частиц, либо не имеет кластеров незанятых позиций. Более того, при любом $n$ 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 $t \geqslant t_c(x)$ 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 бóльшая плотность соответствует плотности незанятых позиций, меньшей $1/2$. Это наблюдение резко упрощает задачу, поскольку динамика в случае высокой плотности частиц нетривиальна и трудно поддается непосредственному анализу. Далее, показывая, что длина любого кластера частиц не может возрастать (т. е. в этой модели не могут возникать транспортные пробки), а число свободных частиц — убывать, мы приходим к описанной в формулировке теоремы альтернативе, что и приводит к требуемым оценкам.

Рассмотрим теперь модель движения со сверхбыстрыми частицами, отличающуюся от предыдущей тем, что на каждом шагу частица сдвигается вперед до следующей занятой позиции.

**Теорема 2.** *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. [13,14,21,22]), результаты для детерминированного случая выглядят очень похоже и для стохастической версии при $p$ достаточно близко к 1. Полный математический анализ здесь к настоящему времени проведен только для модели движения конечного набора частиц по окружности, а не бесконечной решетке (см. [16,18]). Частичный ответ в общем случае получен также при анализе динамики в непрерывном пространстве (см. следующий раздел и [6]).

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 данной задачи проведен в [3].

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. [4–6, 13, 14, 21]. С точки зрения фазовых переходов описанное поведение соответствует возникновению новой «гистерезисной» фазы.

Дадим теперь математическое описание наблюдения о движении пассивной быстрой частицы (имитирующей поведение спешащего прохожего) в медленном транспортном потоке, который мы сформулировали в начале данного раздела. Упрощая ситуацию, мы будем полагать (как обычно делают в гидродинамике), что движение нашей быстрой частицы не влияет на транспортный поток и описывается следующим образом. Положим $\tau_x^+(y) := \min(i\colon y < i \text{ и } x(i) = 1)$, $\tau_x^-(y) := \max(i\colon y > i \text{ и } x(i) = 1)$. Тогда



совместная динамика $\mathcal{T}_\pm$ конфигурации частиц $x \in X$ и положения быстрой частицы $y \in \mathbb{Z}$ определяется косым произведением отображения $T$ и одного из отображений $\tau^\pm$ (знак соответствует движению по потоку или против потока), т. е. $\mathcal{T}_\pm(x, y) := (Tx, \tau_x^\pm(y))$.

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 $\rho < 1/2$ и движении вперед* (*по потоку*), *и к $-\max(1, 1/\rho(x) - 1)$ при движении назад* (*против потока*).

## 2. 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 $r \geqslant 0$ (см. рис. 1). Предполагается, что упорядочивание по индексу соответствует естественному порядку позиций центров шаров, т. е. $\ldots \leqslant x_{-1} \leqslant x_0 \leqslant x_1 \leqslant \ldots$ Чтобы отметить зависимость от радиуса шара $r \geqslant 0$, 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

$$x_i(r) + r \leqslant x_{i+1}(r) - r \quad \forall i \in \mathbb{Z}$$

(соответствующие шары не пересекаются и могут только касаться), и обозначим через $X(r)$ пространство допустимых конфигураций.

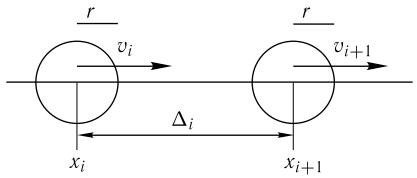



Рис. 1. Процесс с запретами в непрерывном пространстве

Динамика в пространстве конфигураций определяется следующим образом. Начнем с тривиальной конфигурации, состоящей из единственной частицы, находящейся в момент времени $t \geqslant 0$ 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. [6]).

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

$$x_i^t(r) + v_i^t + r \leqslant x_{i+1}^t(r) - r$$

перестает выполняться. В последнем случае мы будем говорить о *конфликте* между частицами $i$ и $i+1$, для разрешения которого применяется конструкция *нормализации*:

$$v_i^t \to \mathcal{N}(v_i^t,\, x^t(r)).$$

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 [6]), 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:

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

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

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

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 $\varphi$. С другой стороны, полагая $r = 1/2$, $x_i^0(r) \in \mathbb{Z}\ \forall i \in \mathbb{Z}$ и $v_i^t \in \mathbb{Z}\ \forall i \in \mathbb{Z}$, $t \geqslant 0$, мы получаем, что $x_i^t(r) \in \mathbb{Z}\ \forall i \in \mathbb{Z}$, $t \geqslant 0$. 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 $v_i^t \in [0,\, v]\ \forall i \in \mathbb{Z}$, $t \in \mathbb{Z}_0 := \mathbb{Z}_+ \cup \{0\}$ и выполнено одно из следующих (на первый взгляд противоположных) предположений:

(а) $v_i^t \equiv v_0^t\ \forall i \in \mathbb{Z}$, $t \in \mathbb{Z}_0$ и $\exists \bar{v}(\gamma) := \lim\limits_{t\to\infty} \dfrac{1}{t} \sum\limits_{s=0}^{t-1} \min(v_0^s,\, \gamma)\ \forall \gamma > 0$ почти наверное (*п.н.*);

(б) $\{v_i^t\}$ являются независимыми одинаково распределенными (*н.о.р.*), как по $i$, так и по $t$, случайными величинами (*с.в.*).

Пересечение между множествами локальных скоростей, удовлетворяющих предположениям (а) или (б), не пусто и содержит принципиально важный случай чисто детерминированных скоростей: $v_i^t \equiv v\ \forall i \in \mathbb{Z}$, $t \in \mathbb{Z}_0$. Как мы покажем, свойства всех систем, удовлетворяющих условию (а), близки к чисто детерминированному случаю. Поэтому в дальнейшем мы будем говорить о постановке (а) как о *детерминированной* [1)], а о постановке (б) — как о *стохастической*.

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

Заметим, что кажущаяся простейшей чисто детерминированная постановка $v_i^t \equiv v\ \forall i \in \mathbb{Z},\ t \in \mathbb{Z}_0$ 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.

## 3. 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 $|I_n| \xrightarrow{n\to\infty} \infty$ предел

$$\rho(x) := \lim_{n\to\infty} \rho(x, I_n)$$

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 $|x_n - x_m| \xrightarrow{|n-m|\to\infty} \dfrac{1}{\rho(x)}$.

(б) Пусть конфигурации $x(r) \in X(r)$, $r > 0$, и $x \in X$ имеют общую последовательность зазоров $\{\Delta_i\}$. Тогда $\rho_\pm(x(r)) = \dfrac{\rho_\pm(x)}{1 + 2r\rho_\pm(x)}$.

**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

$$V(x, i, t) := \frac{1}{t}\sum_{s=0}^{t-1} \mathcal{N}(v_i^s, x^s) \equiv \frac{x_i^t - x_i^0}{t}.$$

Если предел

$$V(x, i) := \lim_{t\to\infty} V(x, i, t)$$

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*

$$|V(x, j, t) - V(x, i, t)| \xrightarrow{t\to\infty} 0 \quad \textit{п.н. } \forall i, j \in \mathbb{Z}.$$

**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 $\forall t\ \exists j > t\colon \Delta_j(x^t) < v$. Тогда $\forall i\ \exists t_i < \infty\colon \Delta_i(x^t) < 2v\ \forall t \geqslant t_i$.*

## 4. 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 $t \to \infty$ последовательности $\{V(x, t)\}_{t\in\mathbb{Z}_0}$ зависит только от $\rho(x)$.*

**Теорема 5** (фундаментальная диаграмма). *В детерминированной постановке*

$$V(x) = \lim_{t\to\infty} \frac{1}{t}\sum_{s=0}^{t-1} \min\left(\frac{1}{\rho}, v_0^s\right) = \begin{cases} v, & \textit{если } \rho(x) \leqslant 1/v, \\ \dfrac{1}{\rho(x)} & \textit{в противном случае,} \end{cases}$$

*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 $v_i^t \equiv v \ \forall i, t$. Тогда*

$$V(x(r)) = \begin{cases} v, & \text{если } \rho(x) \leqslant \dfrac{1}{v+2r}, \\ \dfrac{1}{\rho(x(r)) - 2r} & \text{в противном случае.} \end{cases}$$

*В частности, для версии процесса на целочисленной решетке получаем*

$$V\left(x\left(\frac{1}{2}\right)\right) = \begin{cases} v, & \text{если } \rho(x) \leqslant \dfrac{1}{v+1}, \\ \dfrac{1}{\rho\left(x\left(\frac{1}{2}\right)\right)} - 1 & \text{в противном случае.} \end{cases}$$

**Замечание 2.** Последний результат совпадает с соответствующим утверждением о процессе на решетке, описанном в теореме 1 (см. также [3, 22]). Несмотря на это сходство, в решеточном случае имеется важное качественное отличие динамики: при высокой плотности частицы неминуемо образуют плотные кластеры (статические транспортные пробки). Доказательство же теоремы 5 в действительности показывает, что «типичное» поведение конфигураций высокой плотности качественно отлично: они также образуют кластеры частиц (т. е. наборы последовательных частиц, расстояния между которыми строго меньше $v$), но эти кластеры не стоят на месте, а передвигаются с постоянной скоростью как «эшелон». Интересно отметить, что ранее был разработан целый ряд весьма сложных решеточных моделей для имитации подобного поведения.

В чисто детерминированной постановке (т. е. $v_i^t \equiv v \ \forall i, t$) рассматриваемая система описывается детерминированным отображением $T_v : X \to X$ из пространства допустимых конфигураций в себя. Покажем, что это отображение сильно хаотическое в том смысле, что его топологическая энтропия бесконечна [1)]. Читатель может найти детальное описание конструкций, связанных с энтропией динамической системы и ее свойств, например, в [19]. 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.*

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



Доказательство этого результата основано на аналогичном утверждении для действия отображения сдвига $\sigma_v : X \to X$ в непрерывном пространстве:

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

**Лемма 4.** *Топологическая энтропия отображения сдвига $\sigma_v$ в непрерывном пространстве бесконечна.*

Идея здесь состоит в том, чтобы построить инвариантное подмножество пространства конфигураций $X$, на котором отображение $\sigma_v$ 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.

## 5. 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 пары процессов на общем вероятностном пространстве. Иными словами, каплинг — это процесс пар $(x^t, y^t)$, определенный на пространстве прямого произведения $X \times X$, удовлетворяющий следующим условиям:

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

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

Обсудим теперь конструкцию динамического каплинга между двумя копиями $x^t$, $\acute{x}^t$ рассматриваемого нами марковского процесса. Обычно при анализе систем взаимодействующих частиц на решетке с асинхронными взаимодействиями используется так называемый «равный» каплинг (см., например, [20]). В этом случае каплинг состоит в спаривании частиц процессов $x^t$, $\acute{x}^t$, занимающих одинаковые позиции. После спаривания все выборы случайных скоростей для элементов одной пары предполагаются одинаковыми. В рассматриваемом нами случае систем с синхронными взаимодействиями этот подход не работает. Действительно, произвольное число частиц может сдвинуться одновременно, что приводит к ситуации, когда частицы, принадлежащие процессам $x^t$, $\acute{x}^t$, обгоняют друг друга, но при этом ни в какой момент времени не занимают одинаковые позиции. Кроме того, имеется и более важное препятствие: может оказаться, что движение только одной частицы из пары заблокировано в момент $t$ неспаренной частицей. В результате одновременного движения всех этих

частиц получаем следующую диаграмму: $\overset{\bullet}{\bullet}\!\circ \longrightarrow \circ\,\circ\,\circ$. 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 $\dot{x}$-частицам.

Чтобы обойти это препятствие, мы и вводим *динамический* [1)] каплинг, описанный в [6, 10]. Отметим для сравнения идейно близкую конструкцию каплинга в [1, 16], 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, 16] эти расстояния могут становиться бесконечно большими).

Под *динамическим каплингом* процессов $x^t$, $\dot{x}^t$ понимается последовательное спаривание достаточно близко расположенных частиц в разных процессах, удовлетворяющее следующим условиям.

(A1) В момент $t = 0$ все частицы предполагаются неспаренными. Локальные скорости взаимно спаренных частиц всегда 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 не происходит одно из двух событий: либо нарушается условие допустимости для одной из них (т. е. ее движение заблокировано другой частицей), либо одна из частиц в паре заменяется неспаренной частицей, принадлежащей тому же процессу (см. рис. 2). Удобно представлять результат каплинга наших процесов 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 было удобнее отслеживать позиции неспаренных частиц, мы будем называть их $x$- и $\dot{x}$-*дефектами* в зависимости от процесса, к которому они принадлежат.

Практически динамический каплинг может быть реализован самыми разными способами (в частности, используя только идею обгона частиц). Чтобы продемонстрировать гибкость конструкции, мы опишем другой подход. Отметим, что в дальнейшем только свойства (A1) — (A3) используются в доказательствах.

[1)] Слово «динамический» используется для того, чтобы подчеркнуть то, что взаимное положение частиц в одной паре меняется со временем, в отличие от равного каплинга.



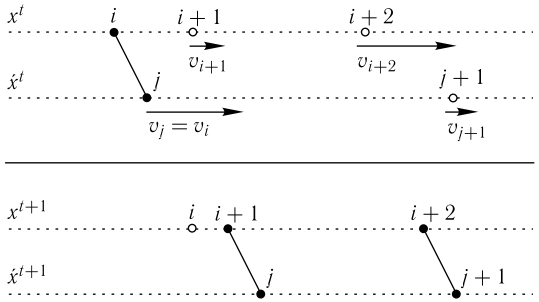



Рис. 2. Спаривание частиц. Взаимно спаренные частицы обозначены черными кружками и соединены прямыми линиями, а дефекты — белыми кружками. В момент $t$ частицы $i$ и $j$ спарены, тогда как в момент $t+1$ $x$-частица $i$ становится неспаренной, а $\dot{x}$-частица $j$ спаривается с $x$-частицей $i+1$. Неспаренные в момент $t$ частицы $i+2$ и $j+1$ спариваются в момент $t+1$

Под $x$-*тройкой* ($\overset{\bullet}{\bullet}\!\circ\,\bullet$ или $\bullet\,\circ\,\overset{\bullet}{\bullet}$) в процессе пар $(x^t, \dot{x}^t)$ понимается две взаимно спаренные частицы и $x$-дефект, находящийся между ними, индекс которого отличается на единицу от индекса спаренной $x$-частицы. $\dot{x}$-тройка ($\bullet\,\circ\,\overset{\bullet}{\bullet}$ или $\overset{\bullet}{\bullet}\,\circ\,\bullet$) определяется аналогично.

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

$x$-дефект в $x_i^t$ вместе с ближайшим к нему [1)] $\dot{x}$-дефектом в $\dot{x}_j^t$ ($\circ\,\circ$ или $\circ\,\circ$) назовем $d$-*парой*, если $|x_i^t - \dot{x}_j^t| < v$, эта пара дефектов не пересекается с другими взаимно спаренными частицами и интервал $(x_i^t, \dot{x}_j^t)$ не содержит других дефектов. Будем говорить, что $d$-пара $(i, j)$ *меньше*, чем $d$-пара $(n, m)$, если $|i| < |n|$, или $i < n$ — в случае $|i| = |n|$. Заметим, что ситуация $i = n$ и $j \neq m$ может произойти, в отличие от ситуации $i \neq n$ и $j = m$.

Приведем два примера. В наборе $\overset{\circ}{\bullet}\,\overset{\bullet}{\bullet}\,\bullet$ две первых $x$-частицы вместе с первой $\dot{x}$-частицей образуют $x$-тройку, несмотря на наличие дополнительной спаренной частицы в интервале между ними. С другой стороны, набор $\circ\,\overset{\bullet}{\bullet}\,\circ$ не содержит ни троек, ни $d$-пар.

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

Правильность пары конфигураций $(x^t, \dot{x}^t)$ в момент $t$ в общем случае не препятствует тому, что под действием динамики в момент $(t+1)$ пара

[1)] Если имеется несколько ближайших $\dot{x}$-дефектов, то выбирается тот, который имеет минимальный индекс.

$(x^{t+1}, \dot{x}^{t+1})$ может оказаться неправильной. В частности, могут возникнуть тройки обоих типов и $d$-пары, например $\bullet\ \overset{\bullet}{\circ}\ \circ \longrightarrow \bullet\ \circ\ \overset{\bullet}{\circ}$ или $\overset{\circ}{\circ}\ \circ\ \circ \longrightarrow \overset{\circ}{\circ}\ \overset{\circ}{\circ}$. Здесь важно, что ввиду сохранения порядка частиц пересекающиеся взаимно спаренные частицы не могут появиться.

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

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

$$\overset{\circ}{\bullet}\ \bullet \longrightarrow \bullet\ \overset{\bullet}{\circ}.$$

Устранение $d$-пары еще проще: дефекты «аннигилируют» друг друга, образуя спаренную пару частиц: $\overset{\circ}{\circ} \longrightarrow \overset{\bullet}{\bullet}$. 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: $\overset{\circ}{\bullet}\ \bullet \longrightarrow \bullet\ \overset{\bullet}{\circ}$.
2. Каждая $\dot{x}$-тройка рекурсивно устраняется: $\bullet\ \overset{\bullet}{\circ} \longrightarrow \overset{\bullet}{\bullet}\ \circ$.
3. Наименьшая[1)] $d$-пара рекурсивно устраняется: $\overset{\circ}{\circ} \longrightarrow \overset{\bullet}{\bullet}$.

**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 $\dot{x}$-тройки может привести к созданию новой тройки того же типа:

$$\bullet\ \overset{\circ}{\bullet}\ \bullet \longrightarrow \bullet\ \overset{\bullet}{\bullet}\ \overset{\circ}{\bullet} \longrightarrow \bullet\ \overset{\bullet}{\bullet}\ \overset{\bullet}{\circ}.$$

Заметим теперь, что при рекурсиях в процедуре каплинга дефект может сдвинуться на произвольно большое расстояние от его начальной позиции:

$$\overset{\bullet}{\circ}\ \overset{\bullet}{\bullet}\ \overset{\cdots}{\cdots}\ \overset{\bullet}{\bullet}\ \overset{\bullet}{\bullet} \longrightarrow \overset{\bullet}{\bullet}\ \overset{\bullet}{\bullet}\ \overset{\cdots}{\cdots}\ \overset{\bullet}{\bullet}\ \overset{\bullet}{\bullet}\ \circ.$$

Обозначим через $\rho_u(x, I)$ плотность $x$-дефектов в конечном сегменте $I$, а через $\rho_u(x) := \rho_u(x, \mathbb{R})$ — верхний предел величин $\rho_u(x, I_n)$, взятый по всем возможным наборам вложенных конечных сегментов $I_n$, длины которых стремятся к бесконечности.

[1)] Порядок $d$-пар может меняться после каждой процедуры рекурсии.



Говорят, что каплинг двух марковских процессов $x^t$, $\dot{x}^t$ *почти удачен*, если верхняя плотность $x$-дефектов $\rho_u(x)$ стремится к нулю по времени почти наверное. Это определение существенно отличается от принятого определения удачного каплинга (см., например, [20]), 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 $x, \dot{x} \in X$ с $\rho(x) = \rho(\dot{x}) > 0$, и предположим, что имеет место почти удачный каплинг $(x^t, \dot{x}^t)$, удовлетворяющий дополнительному условию о том, что расстояния между взаимно спаренными частицами равномерно ограничены сверху величиной $\gamma(t) = o(t)$. Тогда*

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

## 6. Схема доказательства основных эргодических результатов

Начнем с двух технических результатов.

**Лемма 8.** *Супремум $|W_{ij}^t| := x_i^t - \dot{x}_j^t$ по всем взаимно спаренным частицам при динамическом каплинге (см. раздел 5) процессов $(x^t, \dot{x}^t)$ равномерно ограничен величиной $v$ для любого $t \in \mathbb{Z}_0$.*

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

При наших предположениях (стандартный) удачный каплинг[1)] может не существовать (например, в случае чисто детерминированной постановки с двумя равнораспределенными 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 произвольных допустимых конфигурации $x$, $\dot{x}$ одинаковой плотности $\rho > 0$. Если предположить, что имеется почти удачный динамический каплинг процесса пар с начальными условиями $x$, $\dot{x}$, то по лемме 8 выполняются условия леммы 7, откуда следует, что $|V(x, 0, t) - V(\dot{x}, 0, t)| \xrightarrow{t\to\infty} 0$. Применение леммы 2 доказывает наше утверждение.

[1)] Удачный каплинг — когда почти все частицы со временем оказываются спаренными.

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

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

Тогда

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

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 следующих ситуации.

(а) $\lim\limits_{t\to\infty} W_{00}^t/t = 0$. Тогда

$$|V(x, 0, t) - V(\acute{x}, 0, t)| \leqslant \frac{|W_{00}^t|}{t} + \frac{|W_{00}^0|}{t} \xrightarrow{t\to\infty} 0,$$

что по следствию 1 влечет совпадение средних скоростей.

(б) $\limsup\limits_{t\to\infty} W_{00}^t/t > 0$. 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 превысить величины $v$. Поэтому по лемме 7 имеем $|V(x, 0, t) - V(\acute{x}, 0, t)| \xrightarrow{t\to\infty} 0$, что противоречит предположению (б).

(в) $\limsup\limits_{t\to\infty} W_{00}^t/t < 0$. Меняя роли процессов $x^t$, $\acute{x}^t$, возвращаемся к случаю (б).

Поэтому только предположение (а) 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.

## 7. Точные результаты в стохастической постановке

Обсудим теперь разработанный недавно в [8] новый подход к анализу стохастических систем с запретами, позволяющий в ряде случаев получить явное описание для фундаментальной диаграммы в стохастической постановке.



Напомним обозначения. Под *допустимой конфигурацией* $x^t$ в момент времени $t \in \mathbb{Z}_+ \cup \{0\}$ будем понимать упорядоченный счетный набор частиц (шаров) радиуса $r \geqslant 0$, центры расположены в точках $x^t := (x_i^t)_{i\in\mathbb{Z}} \subset \mathbf{R}$, $x_i^t + r \leqslant x_{i+1}^t - r$; здесь $\mathbf{R}$ — решетка. Множество всех допустимых конфигураций обозначим через $X = X(r, \mathbf{R})$. Через $v > 0$ обозначим максимально возможное перемещение отдельной частицы за единицу времени, т. е. $0 \leqslant x_i^{t+1} - x_i^t \leqslant v$. Параметр $p \in (0, 1]$ определяет вероятность передвижения отдельной частицы. Таким образом, отдельная частица без взаимодействия с остальными совершает строго асимметричное случайное блуждание со скачками величины $v$, происходящими с вероятностью $p$. *Плотность* конфигурации $x^t$ (число частиц на единицу длины) определяется как $\rho(x^t) := \lim\limits_{n\to\infty} n/(x_{n-1}^t - x_0^t)$, если последний предел имеет смысл [1].

Рассмотрим три, на первый взгляд, принципиально отличающихся типа процессов с запретами. Во всех случаях мы рассматриваем марковские процессы с дискретным временем, действующие на пространстве конфигураций $X = X(r, \mathbf{R})$, а динамика отдельной частицы в конфигурации $x^t$ определяется следующим соотношением:

$$x_i^{t+1} = \begin{cases} \min\{x_i^t + v, \ x_{i+1}^t - 2r\} & \text{с вероятностью } p, \\ x_i^t & \text{с вероятностью } 1 - p. \end{cases} \tag{1}$$

Процессы типа 1, обозначающиеся $\pi^{(1)}$, действуют на решетке $\mathbf{R} = \mathbb{Z}^1$, $r = 1/2$, $v \in \mathbb{Z}_+^1$. Один узел решетки может быть занят не более чем одной частицей. Модели этого типа широко используются для описания движения автомобилей на однорядном шоссе (см., например, [22, 24], а также раздел 2 настоящей работы).

Процессы типа 2, обозначающиеся $\pi^{(2)}$, также действуют на решетке $\mathbf{R} = \mathbb{Z}^1$, $r = 0$, $v \in \mathbb{Z}_+^1$. Принципиальное отличие состоит в том, один узел решетки может быть занят произвольным числом частиц. Модели этого типа с непрерывным временем называют zero-range процессами (см., например, [17]) и их удобно использовать для моделирования линий связи, при котором частицы соответствуют пакетам информации, ожидающим в очередях к серверам связи, расположенным в узлах решетки $\mathbf{R}$. С точки зрения квантовой статистической механики процессы типа 1 и 2 соотносятся как взаимодействующий газ Ферми и свободный газ Бозе.

Процессы типа 3, обозначающиеся $\pi^{(3)}$ и являющиеся частным случаем процессов, введенных в работе [6] (см. также раздел 2 настоящей работы), действуют уже не на решетке, а в непрерывном пространстве $\mathbf{R} = \mathbb{R}^1$, $r \geqslant 0$, $v \in \mathbb{R}_+^1$. Нетрудно понять, что при $r = 1/2$ процессы $\pi^{(3)}$ содержат все траектории процессов $\pi^{(1)}$, а при $r = 0$ — траектории процессов $\pi^{(2)}$,

[1] Односторонность роста интервалов $(x_0^t, x_{n-1}^t)$ связана с тем, что все частицы двигаются в одну сторону. Определение плотности в общем случае сложнее (см., например, [6]).

что позволяет одновременно получать аналитические результаты для всех рассматриваемых случаев.

Одной из важнейших характеристик рассматриваемых процессов является *средняя скорость* движения частиц за время $t > 0$: $V(x, i, t) := (x_i^t - x_i^0)/t$. В работе [6] было показано, что при весьма общих предположениях, включающих, в частности, динамику частиц, двигающихся в противоположных направлениях, заведомо выполненных во всех рассматриваемых в настоящей работе процессах, в пределе п.н. $t \to \infty$ (если он существует) статистика $V(x, i, t)$ зависит только от плотности $\rho$ конфигурации $x$ (см. также раздел 3). Поэтому для вычисления средней скорости $V(x) = V(\rho(x)) := \lim\limits_{t\to\infty} V(x, i, t)$ (где сходимость почти наверное) достаточно определить последнюю для специально подобранной (удобной для вычисления) начальной конфигурации той же плотности. Отметим, что существование предела по времени, не говоря уже о явных формулах для него, ранее было доказано только в детерминированной постановке (т. е. при $p = 1$), см. [6].

**Теорема 7.** *Для процесса* $\pi^{(3)}$ *для любых* $v$, $\rho \in \mathbb{R}^1_+$, $p \in [0, 1]$ *и для любой допустимой начальной конфигурации плотности* $\rho$ *средняя скорость* $V(\rho, r)$ *корректно определена и вычисляется по формулам*:

$$V(\rho, r) = (1 - 2r\rho)V\Big(\frac{\rho}{1-2r\rho}, 0\Big),$$

$$V(\rho, 0) = \frac{1}{2\rho}\Big[1 + v\rho - \sqrt{(1+v\rho)^2 - 4pv\rho}\,\Big] \xrightarrow{p\to 1} \min\Big\{\frac{1}{\rho}, v\Big\}. \qquad (2)$$

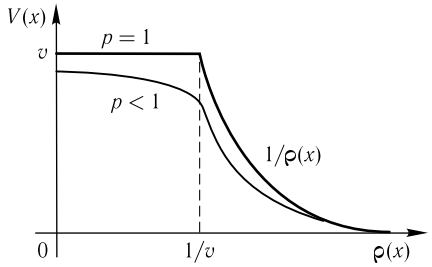



Рис. 3. Фундаментальные диаграммы (зависимость средней скорости от плотности) для процесса $\pi^{(3)}$ с $r = 0$

Метод динамического каплинга (dynamical coupling), описанный в разделе 5 и разработанный в [6, 7], позволяет получить полную информацию о свойствах средних скоростей в детерминированной постановке (т. е. при $p = 1$). Он не требует изучения (многочисленных) инвариантных мер процесса, но дает лишь условные (при условии их существования) результаты



в стохастическом случае, хотя и при существенно более широких предположениях о процессе: локальные скорости $v = v_i$ — н.о.р. случайные величины. В стохастическом случае обойти анализ инвариантных мер не удается.

Приведенный выше способ описания процессов через конфигурации упорядоченных частиц удобен для анализа динамики, но не допускает наличия инвариантных мер (стационарных распределений). Для введения последних мы одновременно с самими процессами будем рассматривать их модификации, в которых частицы неотличимы друг от друга и в которых инвариантные меры уже имеют смысл. Соответствующие множества вероятностных инвариантных мер мы обозначим через $\mathcal{M}^{(i)}_{p,v,r}$, где индекс $i \in \{1, 2, 3\}$ соответствует типу процесса. На сегодняшний день математическое описание инвариантных мер для процессов $\pi^{(i)}$ по существу отсутствует, а единственный классификационный результат [2] дает лишь формальное описание инвариантных мер для процессов $\pi^{(1)}$, $\pi^{(2)}$ при $p = v = 1$ и ничего не говорит даже о существовании нетривиальных инвариантных мер. Чтобы уточнить последнее понятие, мы будем называть меру нетривиальной или *массивной*, если она положительна на любом открытом множестве.

**Теорема 8.** *У процесса* $\pi^{(3)}$ *для любых* $v$, $r \in \mathbb{R}^1_+$, $\rho \in (0, 1/(2r))$, $p \in (0, 1]$ *имеется массивная инвариантная мера* $\mu^{(3)}_{p,r,v,\rho}$ (*являющаяся также трансляционно-инвариантной, но не эргодической*), *такая, что* $E_{\mu^{(3)}_{p,r,v,\rho}}[\pi^{(3)}] = \rho$.

**Следствие 3.** *Для процессов* $\pi^{(1)}$, $\pi^{(2)}$ *выполнено утверждение теоремы* 8.

**Теорема 9.** *У процесса* $\pi^{(1)}$ *для любых* $v \in \mathbb{Z}^1_+$, $p$, $\rho \in (0, 1]$ *при* $v = 1$ *имеется* 1*-параметрическое семейство марковских по пространству, эргодических инвариантных мер* $\{\mu^{(1)}_{p,1/2,1,\rho}\}_\rho$. *Марковский сдвиг на* $\{0, 1\}^{\mathbb{Z}}$ *с матрицей перехода* $(p_{i,j})$, $i$, $j \in \{0, 1\}$, *индуцирует меру из* $\mathcal{M}^{(1)}_{p,1,1/2}$ *тогда и только тогда, когда* $p_{00}p_{11} = (1-p)p_{10}p_{01}$. *При фиксации плотности* $\rho$ *мера* $\mu^{(1)}_{p,1/2,1,\rho}$ *единственна среди трансляционно-инвариантных мер.*

При $0 < p < 1$ инвариантных мер немного, хотя имеется целое семейство не трансляционно-инвариантных мер. Качественно отлична ситуация в детерминированном случае (при $p = 1$): имеется бесконечно много трансляционно-инвариантных мер в $\mathcal{M}^{(1)}_{1,v,1/2,\rho}$.

**Теорема 10.** *У детерминированного процесса* $\pi^{(1)}$ *при* $v = p = 1$ *имеется массивная инвариантная мера* $\mu$, *являющаяся взвешенной суммой двух мер максимальной энтропии для* $\pi^{(1)}$. *Кроме того, име-*

*ется* 1-*параметрическое семейство инвариантных мер* $\mu_\rho$ *таких, что* $E_{\mu_\rho}[x_i^t] = \rho$ *и* $\mu^{(1)}_{p,1/2,1,\rho} \xrightarrow{p\to 1} \mu_\rho$.

До сих пор мы рассматривали только процессы с запретами, действующими в однородных пространствах. В работе [7] была предложена и изучена модификация детерминированной версии процесса $\pi^{(3)}$, учитывающая наличие статических препятствий (светофоров) для движения точечных частиц (т. е. $r = 0$). Зафиксируем произвольную точечную конфигурацию $z = (z_j)_{j\in\mathbb{Z}} \in X(0, \mathbb{R})$, которая соответствует позициям препятствий. Тогда формула (1) переписывается следующим образом:

$$x_i^{t+1} = \begin{cases} \min\{x_i^t + v,\ x_{i+1}^t,\ z_{j(x_i^t)}\} & \text{с вероятностью } p, \\ x_i^t & \text{с вероятностью } 1-p. \end{cases} \tag{3}$$

где $j(x_i^t) := \min\{k \in \mathbb{Z} :\ x_i^t \leqslant z_k\}$. Таким образом, «препятствия» приостанавливают движение частиц.

При заданных $v > 0$ и конфигурации препятствий $z$ обозначим через $\tilde z$ *расширенную* конфигурацию препятствий, получаемую вставкой между каждой парой последовательных препятствий $z_i$, $z_{i+1}$ новых $\lfloor (z_{i+1} - z_i)/v \rfloor$ «виртуальных» препятствий на расстояниях $v$ друг от друга, начиная с точки $z_i$. Здесь $\lfloor u \rfloor$ — это целая часть $u$.

**Теорема 11** ([9]). *Для заданных* $v > 0$, $0 < p < 1$ *и любых конфигураций* $x$, $z \in X(0, \mathbb{R})$, *для которых плотности* $\rho(x)$, $\rho(\tilde z)$ *корректно определены*,

$$V(x, z) = \frac{\rho(x) + \rho(\tilde z) - \sqrt{(\rho(x) + \rho(\tilde z))^2 - 4p\rho(x)\rho(\tilde z)}}{2\rho(x)\rho(\tilde z)} \xrightarrow{p\to 1} \min\left\{\frac{1}{\rho(\tilde z)},\ \frac{1}{\rho(x)}\right\}. \tag{4}$$

Может показаться странным, что максимальная локальная скорость $v$ не входит явным образом в формулу для средней скорости $V(x, z)$, однако последняя явно зависит от расширенной конфигурации $\tilde z$, конструкция которой связана с $v$, в частности $\rho(\tilde z) \geqslant 1/v$.

Интересно отметить, что в [7] было показано, что в более общей постановке с невырожденным распределением значений случайных н.о.р. локальных скоростей $v_i^t$ средние скорости движения частиц могут не существовать.

## 8. Основные идеи и конструкции доказательств в стохастической постановке

Последовательность доказательств: сначала доказываем теорему 9, проверяя инвариантность меры на цилиндрических множествах. Получаем отсюда явную формулу для средней скорости для процесса $\pi^{(1)}$ с $v = 1$.



Далее эти результаты переносятся на более сложные ситуации, описанные в других утверждениях.

Как уже было отмечено, доказать даже существование средних скоростей движения частиц в стохастическом случае, не говоря уже о явных формулах для них, не получив предварительно нетривиальных инвариантных мер (которые интересны и сами по себе), не удается. Ключевым результатом здесь является доказательство теоремы 9, а именно построение марковской инвариантной меры для процесса $\pi^{(1)}$ при $v = 1$, $0 < p < 1$. Под *марковской мерой* здесь понимается единственная инвариантная мера для марковского сдвига на $\{0, 1\}^{\mathbb{Z}}$ с невырожденной матрицей перехода $(p_{ij})$, $i, j \in \{0, 1\}$, $p_{ij} \geqslant 0$, $\sum\limits_j p_{ij} = 1$. В терминах конфигураций частиц 1 здесь соответствует наличию частицы в узле решетки $\mathbb{Z}$, а 0 - ее отсутствию. Условие

$$p_{00}p_{11} = (1-p)p_{10}p_{01} \tag{5}$$

при заданном $0 < p < 1$ определяет однопараметрическое семейство марковских мер $\{\mu_a^{(1)}\}_a$, $a := p_{01} \in (0, 1)$. Инвариантность этих мер относительно динамики (1) проверяется вычислением мер конечных цилиндров и баланса вероятностей переходов между ними (собственно, отсюда и возникает приведенное условие). Из (5) получаем $p_{10} = (1-a)/(1-pa)$, а в силу равенства между плотностью частиц $\rho$ и стационарной вероятностью единиц при марковском сдвиге имеем $\rho = a/(a + p_{10}) = a(1-pa)/(1-pa^2)$. Разрешая последнее равенство относительно параметра $a$, получаем

$$a = \frac{1 - \sqrt{1 - 4p\rho(1-\rho)}}{2p(1-\rho)}. \tag{6}$$

Поэтому полученное семейство мер однозначно индексируется плотностью. Свойство массивности мер $\{\mu_a^{(1)}\}_a$ проверяется непосредственно, а их эргодичность доказывается при помощи метода динамического каплинга.

В детерминированном случае ($p = 1$) описанная конструкция не проходит (хотя бы ввиду неединственности в классе трансляционно-инвариантных мер). Однако здесь применим другой (хорошо известный в гиперболической теории) подход, связанный с построением мер максимальной энтропии (см., например, [12]). Рассматриваются равномерное распределение периодических точек периода $n$ и их пределы при $n \to \infty$. Показывается, что есть два таких предела, один из которых соответствует мере максимальной энтропии для сдвига вправо на $\{0, 1\}^{\mathbb{Z}}$ (и связан с конфигурациями плотности $\rho \leqslant 1/2$), а другой — для сдвига влево (и связан с конфигурациями плотности $\rho > 1/2$). Далее применяются соответствующие результаты для топологических марковских цепей (см., например, [12]). Для построения мер $\mu_\rho$ по мерам максимальной энтропии ис-

пользуется либо гиббсовская перестройка, либо их явное представление в терминах марковских сдвигов.

Сформулируем несколько результатов, позволяющих перейти от этого результата к более общей ситуации.

**Лемма 10.** *Для любых заданных $v$, $p$ справедливы следующие соотношения*:

1) $\pi^{(1)} = \pi^{(3)}|X(1/2, \mathbb{Z})$,

2) $\pi^{(2)} = \pi^{(3)}|X(0, \mathbb{Z})$,

3) $\forall r > 0$ *существует гомеоморфизм* $\varphi = \varphi_r \colon X(r, \mathbb{R}) \to X(0, \mathbb{R})$, *такой что* $\varphi \circ \pi_r^{(3)} = \pi_0^{(3)} \circ \varphi$,

4) $\forall u, v > 0$, $0 < p \leqslant 1$ *подрешетка* $\mathbf{R}_{u,v} := v\mathbb{Z} + u$ *инвариантна относительно процесса* $\pi^{(3)}$, *т. е.* $x^0 \subset \mathbf{R}_{u,v} \implies x^t \subset \mathbf{R}_{u,v}\ \forall t > 0$.

Последнее свойство позволяет перенести результат о существовании массивных инвариантных мер с $\pi^{(1)}$ на $\pi^{(3)}$, однако эти меры уже не являются эргодическими.

Перейдем к вычислению средних скоростей и начнем с краткой формулировки полученных ранее результатов о сравнении средних скоростей (см. также [6]).

**Теорема 12.** *Для процесса* $\pi^{(3)}$ *для любых* $v$, $r \in \mathbb{R}^1_+$, $\rho \in (0, 1/(2r))$, $p \in (0, 1]$ *справедливо следующее. Пусть* $x^0, y^0 \in X(r, \mathbb{R})$ *и* $\rho(x) = \rho(y) = \rho$, *а средняя скорость* $V(y)$ *корректно определена. Тогда* $|V(x, i, t) - V(y, j, t)| \xrightarrow{t\to 0} 0$ *для любых* $i, j \in \mathbb{Z}$.

Полученное в теореме 9 представление для марковской инвариантной меры процесса $\pi^{(1)}$ при $v = 1$ немедленно дает формулу для средней скорости $V(\rho, p, 1, 1/2) = pp_{10} = p(1-a)/(1-pa) \in [0, p]$. Подставляя значение $a = a(\rho)$, согласно формуле (6) получаем

$$V\Big(\rho, p, v = 1, r = \frac{1}{2}\Big) = \frac{1 - \sqrt{1 - 4p\rho(1-\rho)}}{2\rho}. \tag{7}$$

В силу теоремы 12 и леммы 10 этот результат переносится на процессы 3-го типа с $v = 1$, $r = 1/2$ без изменения. Заметим, что последняя формула известна в физических работах для случая процессов типа 1 с $v = 1$ (см. [24]).

Важно отметить, что наивный переход от $v = 1$ к $v > 1$ непосредственно в классе решеточных процессов (используя инвариантность подрешеток с шагом, кратным $v$) невозможен — точнее, таким образом можно изучать лишь конфигурации малой плотности $< 1/v$. Вместо этого мы воспользуемся самоподобием процессов типа 3, действующих в непрерывном пространстве. Для $\pi^{(3)}$ имеем: $1 \to v \implies \rho \to \rho/v$, $V_{v=1}(\rho, 0) \to vV_{v=1}(v\rho, 0)$.



Применяя эти преобразования подобия, из частного случая (7) мы получаем общую формулу для средней скорости (2):

$$V(\rho, p, v > 0, r = 0) = vV\Big(v\rho, p, v = 1, r = \frac{1}{2}\Big) = \frac{1 + v\rho - \sqrt{(1+v\rho)^2 - 4pv\rho}}{2\rho}.$$

В свою очередь, результаты для $\pi^{(3)}$ с произвольным $v \in \mathbb{Z}_+$ непосредственно переносятся по лемме 10 обратно на решеточные случаи.

Обсудим теперь принципиально отличающуюся ситуацию, описанную в теореме 11. Дело в том, что неоднородность пространства, в котором осуществляется коллективное случайное блуждание (наличие препятствий), в общем случае не допускает существования инвариантных мер [1)]. Поэтому основной применяемый нами шаг построения массивной инвариантной меры невозможен. Упомянутое в формулировке теоремы 11 техническое ограничение на выбор $z$ состоит в следующем. По заданной конфигурации $z$ и значению $v > 0$ определим новую *расширенную* конфигурацию $\tilde{z}$, полученную вставкой между каждой парой элементов $z_i$, $z_{i+1}$ новых $\lfloor (z_{i+1} - z_i)/v \rfloor$ «виртуальных» препятствий начиная от точки $z_i$ на расстоянии $v$ друг от друга. Здесь $\lfloor u \rfloor$ обозначает целую часть числа $u$. В этих терминах ограничение состоит в существовании плотности для расширенной конфигурации $\tilde{z}$. Отметим, что отсюда не следует даже существование плотности для самой конфигурации $z$. Используя технику, разработанную в работе [7] для детерминированной версии этой задачи, удается показать, что вычисление средней скорости $V(x, z, v, p)$ сводится к анализу марковского процесса типа 2, действующего (в отличие от уже изученной постановки) на *неоднородной* решетке $\mathbf{R} := \tilde{z}$. Существование плотности конфигурации $\tilde{z}$ позволяет перенести результаты, полученные для обычной решетки $\mathbb{Z}$, на рассматриваемый неоднородный случай, что и завершает конструкцию, детали которой описаны в [9].

### Литература


1. *Angel O.* The Stationary Measure of a 2-type Totally Asymmetric Exclusion Process // J. Combin. Theory Ser. A. 2006. V. 113, № 4. P. 625–635; `arXiv/0501005 math.CO`
2. *Belitsky V., Ferrari P. A.* Invariant Measures and Convergence for Cellular Automaton 184 and Related Processes // J. Stat. Phys. 2005. V. 118, № 3–4. P. 589–623; `arXiv:math/9811103v1`
3. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111, № 3–4. P. 903–930; `arXiv:math/0206194`
4. *Blank M.* Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys. 2005. V. 120, № 3–4. P. 627–658; `arXiv:math.DS/0408240`


[1)] Для существования инвариантных мер необходимо хотя бы выполнение условия стационарности для конфигураций препятствий $z$.

5. *Blank M.* Travelling with/against the Flow. Deterministic Diffusive Driven Systems // J. Stat. Phys. 2008. V. 133, № 4. P. 773–796; arXiv:0810.2205 math.DS
6. *Blank M.* Metric properties of discrete time exclusion type processes in continuum // J. Stat. Phys. 2010. V. 140, № 1. P. 170–197; arXiv:0904.4585 math.DS
7. *Blank M.* Exclusion type spatially heterogeneous processes in continua // J. Stat. Mech. 2011. P06016; arXiv:1105.4232 math.DS
8. *Бланк М. Л.* Нетривиальные инвариантные меры и статистики для процессов с запретами // Доклады РАН. 2012. Т. 442, № 3. С. 295–299.
9. *Blank M.* Discrete time TASEP in heterogeneous continuum. Preprint, 2011.
10. *Бланк М. Л., Пирогов С. А.* О квазиуспешном каплинге марковских процессов // Пробл. передачи информ. 2007. Т. 43, № 4. С. 51–67.
11. *Borodin A., Ferrari P. L., Sasamoto T.* Large time asymptotics of growth models on space-like paths II: PNG and parallel TASEP. arXiv:0707.4207 math-ph, 2007.
12. *Боуэн Р.* Методы символической динамики. М.: Мир, 1979.
13. *Chowdhury D., Santen L., Schadschneider A.* Statistical physics of vehicular traffic and some related systems // Physics Reports. 2000. V. 329. P. 199–329; arXiv:0007053 cond-mat
14. *Comtet A., Majumdar S. N., Ouvry S., Sabhapandit S.* Integer partitions and exclusion statistics: limit shapes and the largest parts of Young diagrams // J. Stat. Mech. 2007. P10001; arXiv:0707.2312
15. *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.
16. *Evans M. R., Ferrari P. A., Mallick K.* Matrix representation of the stationary measure for the multispecies TASEP. arXiv:0807.0327 math.PR, 2008.
17. *Evans M. R., Hanney T.* Nonequilibrium Statistical Mechanics of the Zero-Range Process and Related Models // J. Phys. A: Math. Gen. 2005. V. 38. P. R195–R239; arXiv:cond-mat/0501338
18. *Gray L., Griffeath D.* The ergodic theory of traffic jams // J. Stat. Phys. 2001. V. 105, № 3–4. P. 413–452; http://psoup.math.wisc.edu/traffic/
19. *Корнфельд И. П., Синай Я. Г., Фомин С. В.* Эргодическая теория. М.: Наука, 1980.
20. *Liggett T. M.* Interacting particle systems. N.Y.: Springer-Verlag, 1985.
21. *Maerivoet S., De Moor B.* Cellular Automata Models of Road Traffic // Physics Reports. 2005. V. 419, № 1. P. 1–64; arXiv:physics/0509082
22. *Nagel K., Schreckenberg M.* A cellular automaton model for freeway traffic // J. Physique I. 1992. V. 2. P. 2221–2229.
23. *Penrose M. D.* Existence and spatial limit theorems for lattice and continuum particle systems // Probab. Surveys. 2008. V. 5. P. 1–36.
24. *Schadschneider A., Schreckenberg M.* Cellular automation models and traffic flow // J. Phys. A: Math. Gen. V. 1993. V. 26. P. L679–L683.

**К. В. Воронцов, Ю. В. Чехович**

# Интеллектуальный анализ данных в задачах моделирования транспортных потоков

## 1. Моделирование транспортных потоков

Математическое моделирование автомобильных транспортных потоков как самостоятельная область математического моделирования ведет отсчет с середины 50-х годов XX века. Тогда, с одной стороны, применение математического аппарата при решении транспортных задач стало актуальным, так как значительно выросла автомобилизация населения и существовавшая дорожная сеть во многих городах и странах перестала справляться с возросшей нагрузкой. С другой стороны, для методов моделирования, разработанных в годы войны, в том числе использующих возможности вычислительной техники, стали искать мирные способы применения.

Среди математических моделей транспортных потоков традиционно выделяют два больших класса: модели-аналоги (макроскопические модели), в которых для моделируемого потока применяются газо- и гидродинамические аналоги, и микроскопические модели, моделирующие поведение отдельных транспортных средств в потоке в различных ситуациях. Очевидно, что разные модели строятся для решения различных задач и имеют различные зоны применимости своих решений. Чаще всего микроскопические модели применяют для решения локальных в территориальном смысле задач, например для расчетов предельных пропускных способностей различных участков дорожной сети, выбора режимов оперативного управления фазами светофорных объектов и т. п. Макроскопические модели, наоборот, используются при принятии стратегических решений, например, по изменению улично-дорожной сети, планированию маршрутов наземного городского транспорта, при строительстве новых жилых и промышленных районов, дорог, развязок и т. п.

Технология моделирования транспортных потоков, как, впрочем, и построение любых других моделей, включает такой неотъемлемый этап, как идентификация модели, то есть определение внешних параметров модели. Очень часто именно идентификация моделей оказывается наиболее дорогим и трудоемким этапом, затраты на который превосходят затраты на другие этапы на один или несколько порядков [14]. Следует отметить, что идентификация модели тесно связана с другим этапом моделирования —

верификацией модели, то есть установлением адекватности построенной модели моделируемой системе. Транспортные системы с точки зрения идентификации и верификации моделей являются чрезвычайно сложными объектами.

В первую очередь, сложность объективно присуща самой системе. Большое количество элементов системы — в крупном городе по десяткам тысяч ребер графа транспортной сети (дорогам) одновременно перемещаются сотни тысяч транспортных средств — приводит к кратным количествам внутренних и внешних переменных, используемых в моделях. Значительная территориальная распределенность, сопряженная с наличием нелокальных в пространстве эффектов, то есть ситуаций, в которых возмущение на одном участке транспортной сети (авария, поломка светофора, затор) через небольшое время может оказать влияние на очень отдаленный участок сети, часто не дает возможности провести эффективную декомпозицию модели без значительной потери адекватности.

Наличие ярко выраженных элементов контринтуитивного поведения транспортной системы, то есть ситуаций, в которых реакция системы на изменения противоречит «простой житейской логике», затрудняет конструирование и верификацию моделей. Простейшими примерами такого поведения являются ситуации, когда расширение дороги на одном участке приводит к снижению пропускной способности дороги, а непродуманное строительство новой дороги ухудшает общую эффективность транспортной системы. Более нетривиальным примером контринтуитивного поведения является, например, парадокс Доунса—Томсона, заключающийся в том, что средневзвешенная скорость движения личного автотранспорта по дорожной сети прямо зависит от скорости, с которой перемещаются пассажиры общественного внеуличного транспорта [22].

Очевидно, что сбор информации о транспортной системе сам по себе представляет значительную сложность. Судя по всему, на настоящий момент даже потенциально не существует единого источника данных, содержащего полную, актуальную и достоверную информацию о любой сколь-нибудь нетривиальной транспортной системе. При этом многие исследователи отмечают значительное влияние конкретных значений внешних параметров системы на финальные результаты моделирования.

## 2. О методологии интеллектуального анализа данных

*Интеллектуальный анализ данных* (*Data Mining*) — это современная концепция анализа данных, изначально предполагающая, что данные могут быть неточными, разнородными, содержать пропуски и при этом иметь гигантские объемы.



Согласно общепринятому определению, которое дал Г. Пятецкий-Шапиро — один из ведущих в мире экспертов в области анализа данных, *интеллектуальный анализ данных* — это *процесс обнаружения в «сырых» данных ранее неизвестных*, *нетривиальных*, *практически полезных*, *доступных интерпретации знаний*, *необходимых для принятия решений в различных сферах человеческой деятельности* [24].

На самом деле, по составу решаемых задач интеллектуальный анализ данных практически не отличается от стандартного набора средств, применяемых уже более полувека в области статистического анализа данных, поиска закономерностей и обучения по прецедентам. Основное различие заключается в эффективности алгоритмов и технологичности их применения. Подавляющее большинство классических процедур имеют квадратичное или даже кубическое (по числу объектов) время выполнения. При количестве объектов, превосходящем несколько десятков тысяч, они работают неприемлемо медленно даже на самых современных компьютерах. Специализированные алгоритмы интеллектуального анализа данных способны выполнять те же задачи за линейное или даже логарифмическое время без существенной потери точности.

## 3. Примеры задач интеллектуального анализа транспортных данных

Распространенной задачей является прогнозирование динамики характеристик транспортных потоков. Например, для построения оптимальных с точки зрения времени проезда маршрутов движения по некоторой транспортной сети требуется *прогнозирование скоростей транспортных средств* на участках этой транспортной сети [9]. Горизонт прогнозирования при этом должен быть не меньше типичного времени одной поездки, чтобы при построении машрута перед выездом можно было учитывать скорости движения на участках транспортной сети, близких к цели поездки. Исходными данными для прогнозирования являются временные ряды скоростей и информация о топологии дорожной сети, представленная, например, в виде графа. Очевидно, что прогноз зависит от времени суток, дня недели и сезона поездки. К внешним факторам можно отнести погодные условия. Кроме этого, очевидно, существуют нетривиальные закономерности между скоростями движения на различных ребрах графа дорожной сети. Например, очень низкая скорость на определенном участке (пробка) может вызывать снижение скорости на одних ребрах (распространение пробки) и одновременно повышение скорости на других ребрах за счет того, что резко уменьшается количество попадающих на них транспортных средств (экранирование).

Оценка качества прогнозирования может осуществляться, например, с использованием специальных машин-асессоров, по информации о проезде которых по заданному маршруту со скоростью потока можно оценить разницу спрогнозированного и реального времени проезда этого маршрута.

Другим характерным примером анализа данных является *задача идентификации государственных номеров транспортных средств* с помощью видеокамер. Такие камеры устанавливают, например, перед постами ГИБДД, чтобы можно было обнаруживать угнанные транспортные средства, или в составе комплексов фиксации нарушений скоростного режима вместе с радаром. Задача идентификации заключается в том, чтобы в последовательности изображений, поступающих от видеокамеры, обнаружить все участки изображений, содержащие автомобильные номера, и правильно распознать все символы на этих номерах. Сложности, возникающие при решении этой задачи, связаны прежде всего с качеством поступающих изображений — номера снимаются под различными углами, в разных условиях освещенности, с негативным влиянием погодных условий, могут быть загрязнены или частично перекрываться другими предметами. С другой стороны, каждый номер имеет строго стандартизированный вид и, как правило, представлен в видеозаписи на нескольких (или даже на нескольких десятках) последовательных кадрах. Кроме того, при решении задачи нет недостатка в количестве прецедентной информации.

Задача решается в несколько этапов. Сначала производится предобработка изображений для удаления шумов и стандартизации изображений, затем каждый кадр обрабатывается с целью поиска кандидатов на изображения номеров. На последующих кадрах производится подтверждение кандидатов. Все подтвержденные кандидаты обрабатываются дальше для выделения отдельных структурных частей автомобильного номера и определения его типа, а затем и для распознавания отдельных символов номера. Заключительным этапом может быть поиск распознанного номера в базах данных номеров ГИБДД для подтверждения правильности идентификации и получения дополнительной информации по этому номеру. В настоящее время задача решается с разной степенью качества большим количеством прикладных систем.

*Задача определения типа транспортного средства* относится к задачам классификации. Часто при моделировании транспортных систем исследователи вводят в модель ту или иную типизацию транспортных средств. При этом могут выделяться, например, такие категории: «легковые», «грузовые», «автобусы-троллейбусы». Типизация может быть и более тонкой, когда, например, среди «легковых» выделяются «джипы», а «грузовые» делятся по тоннажу. Чтобы оценить долю тех или иных типов транспортных средств, используют анализ видеоизображений с целью выделения на них автомобилей и отнесения этих автомобилей к тому



или иному типу. Также существуют задачи, например связанные с розыском, когда на основе видеоизображений требуется еще более детальная классификация, вплоть до модели транспортного средства. Прецедентами являются эталонные изображения транспортных средств, возможно снятые с различных ракурсов. Задача решается также в несколько этапов. Сначала из последовательности кадров выделяются транспортные средства, затем по последовательности изображений транспортного средства строится его трехмерная модель, которая сравнивается с трехмерными моделями, восстановленными из эталонных изображений [1, 6, 17].

Методы анализа данных дают транспортным компаниям дополнительные возможности по управлению автопарком путем решения задачи *мониторинга перемещений транспортных средств*. В настоящее время многие транспортные компании оборудуют транспортные средства устройствами для фиксирования треков перемещения. Устройство с определенной частотой получает координаты транспортного средства от спутников GPS или ГЛОНАСС и передает их в диспетчерский центр. Таким образом, в центре формируется трек транспортного средства. Очевидно, что диспетчеры не в состоянии отслеживать каждый трек непосредственно. Поэтому решается задача *обнаружения необычного поведения* транспортного средства. Объектами являются треки движения или их участки. В качестве признаков в такой задаче могут выступать значения функций, регистрирующих те или иные отклонения от обычных параметров поездки (изменение маршрута, отклонения во времени следования, наличие «разрывов» трека, вызванных временным непоступлением данных, и т. п.). В случае отнесения объекта к классу «необычных» диспетчер получает сигнал, после чего производится расследование. По результатам расследования происходит либо подтверждение, либо отклонение сигнала. Таким образом накапливается прецедентная информация. Важным аспектом решения задачи мониторинга является управление количеством сигналов. Сигналы не должны поступать слишком часто, чтобы диспетчеры могли своевременно обработать каждый поступивший сигнал.

Еще одним примером является *задача выявления заторов* в транспортной сети. Рассмотрим граф дорожной сети, на части ребер которого установлены детекторы транспорта (см. раздел 6 настоящего приложения). В случае возникновения затора требуется в течение минимального времени определить ребро, на котором возник затор, чтобы у сотрудников организации по управлению дорожным движением была возможность перенаправить транспортные потоки и принять меры к устранению затора. Такие меры, предпринятые своевременно, помогут предотвратить разрастание затора на соседние ребра. Для создания и настройки требуемого сигнального алгоритма можно использовать исторические данные, собираемые с детекторов. Прецедентами являются интервалы времени и ребра,

на которых возникали заторы. Признаки вычисляются по временным рядам данных, собираемых с детекторов.

## 4. Основные понятия теории обучения машин

Пусть существуют множество *объектов* $X$, множество *допустимых ответов* $Y$ и *целевая функция* (target function) $y^*\colon X \to Y$, значения которой $y_i = y^*(x_i)$ известны только на конечном подмножестве объектов $\{x_1, \ldots, x_\ell\} \subset X$. Пары «объект–ответ» $(x_i, y_i)$ называются *прецедентами*. Совокупность пар $X^\ell = (x_i, y_i)_{i=1}^\ell$ называется *обучающей выборкой* (training sample).

Задача *обучения по прецедентам* заключается в том, чтобы по выборке $X^\ell$ *восстановить зависимость* $y^*$, то есть построить *решающую функцию* (decision function) $a\colon X \to Y$, которая приближала бы целевую функцию $y^*(x)$, причем не только на объектах обучающей выборки, но и на всем множестве $X$. Решающая функция $a$ должна допускать эффективную компьютерную реализацию; по этой причине будем называть ее *алгоритмом*.

*Признак* (feature) $f$ объекта $x$ — это результат измерения некоторой характеристики объекта. Формально признаком называется отображение $f\colon X \to D_f$, где $D_f$ — множество допустимых значений признака.

В зависимости от природы множества $D_f$ признаки делятся на несколько типов.

Если $D_f = \{0, 1\}$, то $f$ — *бинарный* признак.

Если $D_f$ — конечное множество, то $f$ — *номинальный* признак.

Если $D_f$ — конечное упорядоченное множество, то $f$ — *порядковый* признак.

Если $D_f = \mathbb{R}$, то $f$ — *количественный* признак.

В прикладных задачах анализа транспортных потоков могут встречаться и более сложные признаки, значениями которых могут быть числовые последовательности, растровые изображения, функции, графы, результаты запросов к базе данных и т. д.

Если все признаки имеют одинаковый тип, $D_{f_1} = \ldots = D_{f_n}$, то исходные данные называются *однородными*, в противном случае — *разнородными*.

Пусть имеется набор признаков $f_1, \ldots, f_n$. Вектор $(f_1(x), \ldots, f_n(x))$ называют *признаковым описанием* объекта $x \in X$. В дальнейшем мы не будем различать объекты из $X$ и их признаковые описания, полагая $X = D_{f_1} \times \ldots \times D_{f_n}$. Совокупность признаковых описаний всех объектов выборки $X^\ell$, записанную в виде таблицы размера $\ell \times n$, называют *матрицей объектов-признаков*:



$$F = \|f_j(x_i)\|_{\ell \times n} = \begin{pmatrix} f_1(x_1) & \ldots & f_n(x_1) \\ \ldots\ldots\ldots\ldots & & \\ f_1(x_\ell) & \ldots & f_n(x_\ell) \end{pmatrix}. \qquad (1)$$

Матрица объектов-признаков является стандартным и наиболее распространенным способом представления исходных данных в прикладных задачах.

Таким образом, признаки — это характеристики объектов, которые либо измеряются непосредственно, либо вычисляются по «сырым» исходным данным. Любое отображение из множества $X$ можно рассматривать как признак. В частности, любой алгоритм $a\colon X \to Y$ также можно рассматривать как признак.

В зависимости от природы множества допустимых ответов $Y$ задачи обучения по прецедентам делятся на следующие типы.

Если $Y = \{1, \ldots, M\}$, то это задача *классификации* (classification) на $M$ непересекающихся классов. В этом случае все множество объектов $X$ разбивается на классы $K_y = \{x \in X\colon y^*(x) = y\}$, и алгоритм $a(x)$ должен давать ответ на вопрос: какому классу принадлежит $x$? В некоторых приложениях классы называют *образами* и говорят о задаче *распознавания образов* (pattern recognition).

Если $Y = \{0, 1\}^M$, то это задача *классификации на $M$ пересекающихся классов*. В простейшем случае эта задача сводится к решению $M$ независимых задач классификации с двумя непересекающимися классами.

Если $Y = \mathbb{R}$, то это задача *восстановления регрессии* (regression estimation).

Задачи *прогнозирования* (forecasting) являются частными случаями классификации или восстановления регрессии, когда $x \in X$ — описание прошлого поведения объекта $x$, а $y \in Y$ — описание некоторых характеристик его будущего поведения.

*Моделью алгоритмов* называется параметрическое семейство отображений $A$, из которого выбирается искомый алгоритм $a(x)$:

$$A = \{\varphi(x, \gamma) \mid \gamma \in \Gamma\},$$

где $\varphi\colon X \times \Gamma \to Y$ — некоторая фиксированная функция, $\Gamma$ — множество допустимых значений параметра $\gamma$, называемое *пространством параметров* или *пространством поиска* (search space).

Процесс подбора параметров модели по обучающей выборке называют *настройкой* (fitting) или *обучением* (training, learning) алгоритма [2]. В результате настройки выбирается единственный алгоритм $a \in A$, который должен приближать целевую зависимость.

*Методом обучения* (learning algorithm) $\mu\colon (X \times Y)^\ell \to A$ называется отображение, которое произвольной конечной выборке $X^\ell = ((x_1, y_1), \ldots, (x_\ell, y_\ell))$ ставит в соответствие алгоритм $a\colon X \to Y$. Говорят также, что метод $\mu$ *строит* алгоритм $a$ по выборке $X^\ell$. Метод обучения, как и сам алгоритм $a$, должен допускать эффективную программную реализацию.

В задачах обучения по прецедентам четко различаются два этапа:

- на этапе *обучения* метод $\mu$ по выборке $X^\ell$ строит алгоритм $a = \mu(X^\ell)$;
- на этапе *применения* алгоритму $a$ подаются на вход новые объекты $x$, вообще говоря, отличные от обучающих, для получения ответов $y = a(x)$.

Этап обучения наиболее сложен. Как правило, он сводится к поиску параметров модели, доставляющих оптимальное значение заданному функционалу качества.

*Функция потерь* (loss function) — это неотрицательная функция $\mathcal{L}(a, x)$, характеризующая величину ошибки алгоритма $a$ на объекте $x$. Если $\mathcal{L}(a, x) = 0$, то ответ $a(x)$ называется *корректным*.

*Функционал качества* алгоритма $a$ на выборке $X^\ell$:

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(a, x_i). \tag{2}$$

Функционал $Q$ называют также функционалом *средних потерь* или *эмпирическим риском* [2], так как он вычисляется по *эмпирическим данным* $(x_i, y_i)_{i=1}^\ell$.

Функция потерь, принимающая только значения 0 и 1, называется *бинарной*. В этом случае $\mathcal{L}(a, x) = 1$ означает, что алгоритм $a$ допускает ошибку на объекте $x$, а функционал $Q$ называется *частотой ошибок* алгоритма $a$ на выборке $X^\ell$.

Наиболее часто используются следующие функции потерь, при $Y \subseteq \mathbb{R}$:

$\mathcal{L}(a, x) = [a(x) \neq y^*(x)]$ — бинарная функция потерь, индикатор ошибки; обычно применяется в задачах классификации [1)];

$\mathcal{L}(a, x) = |a(x) - y^*(x)|$ — отклонение от правильного ответа; функционал $Q$ называется *средней ошибкой* алгоритма $a$ на выборке $X^\ell$;

$\mathcal{L}(a, x) = (a(x) - y^*(x))^2$ — квадратичная функция потерь; функционал $Q$ называется *средней квадратичной ошибкой* алгоритма $a$ на выборке $X^\ell$; обычно применяется в задачах регрессии.

Классический метод обучения, называемый *минимизацией эмпирического риска* (empirical risk minimization, ERM), заключается в том, чтобы найти в заданной модели $A$ алгоритм $a$, доставляющий минимальное значение функционалу качества $Q$ на заданной обучающей выборке $X^\ell$:

$$\mu(X^\ell) = \arg\min_{a \in A} Q(a, X^\ell). \tag{3}$$

[1)] Квадратные скобки переводят логическое значение в число по правилу [ложь] = 0, [истина] = 1.



Минимизацию эмпирического риска следует применять с известной долей осторожности. Если минимум функционала $Q(a, X^\ell)$ достигается на алгоритме $a$, то это еще не гарантирует, что $a$ будет хорошо приближать целевую зависимость на произвольной *контрольной выборке* $X^k = (x'_i, y'_i)_{i=1}^k$.

Когда качество работы алгоритма на новых объектах, не вошедших в состав обучения, оказывается существенно хуже, чем на обучающей выборке, говорят об эффекте *переобучения* (overtraining) или *переподгонки* (overfitting). При решении практических задач на реальных данных транспортных потоков с этим явлением приходится сталкиваться очень часто.

Легко представить себе метод, который минимизирует эмпирический риск до нуля, но при этом абсолютно не способен обучаться. Получив обучающую выборку $X^\ell$, он запоминает ее и строит алгоритм, который сравнивает предъявляемый объект $x$ с обучающими объектами $x_i$ из $X^\ell$. В случае совпадения $x = x_i$ алгоритм выдает правильный ответ $y_i$. Иначе выдается произвольный ответ. Эмпирический риск принимает наименьшее возможное значение, равное нулю. Однако этот алгоритм не способен восстановить зависимость вне материала обучения. Отсюда вывод: для успешного обучения необходимо не только запоминать, но и обобщать.

*Обобщающая способность* (generalization ability) метода $\mu$ характеризуется величиной $Q(\mu(X^\ell), X^k)$ при условии, что выборки $X^\ell$ и $X^k$ являются представительными. Для формализации понятия «представительная выборка» обычно принимается стандартное предположение, что выборки $X^\ell$ и $X^k$ — простые, полученные из одного и того же неизвестного вероятностного распределения на множестве $X$.

Оценки обобщающей способности позволяют предсказывать качество алгоритмов и строить более надежные методы обучения. Первые оценки были получены в конце 1960-х годов советскими математиками В. Н. Вапником и А. Я. Червоненкисом [3–5]. В настоящее время статистическая теория развивается очень активно [20], однако для многих практически интересных случаев оценки обобщающей способности либо неизвестны, либо сильно завышены. Численно точные оценки получены лишь для некоторых частных случаев [29–31].

## 5. Эвристические принципы интеллектуального анализа данных

Несмотря на большое разнообразие применяемых моделей алгоритмов и методов их настройки, общих принципов их построения не так уж много. Наиболее удачные модели совмещают в себе сразу несколько принципов.

Все эти принципы являются в той или иной степени *эвристическими* — они опираются не только на строгие математические обоснования, но в значительной степени на соображения здравого смысла. Не существует универсальных моделей, подходящих под любые задачи. Каждая эвристика хорошо работает лишь в своем классе задач. Понимание взаимосвязей между ними позволяет сочетать различные эвристики и конструировать новые методы, наиболее подходящие для конкретных случаев.

Принцип минимизации эмпирического риска был описан в предыдущем разделе. Далее приведем ряд других важных эвристик.

*Принцип сходства* предполагает, что на множестве $X$ можно так ввести функцию расстояния между объектами, что близким объектам будут, как правило, соответствовать близкие ответы. Применительно к задачам восстановления регрессии это равносильно предположению, что целевая функция $y^*$ является достаточно гладкой. Даже если она имеет резкие скачки, они не могут находиться повсюду. В случае классификации принцип сходства означает, что схожие объекты, как правило, лежат в одном классе. Граница классов может быть довольно сильно изрезана, но она не может проходить везде. В «хорошей» задаче классы представляют собой области, компактно расположенные в пространстве $X$. Это предположение называют *гипотезой компактности*. Эмпирический опыт убеждает, что слишком сложные зависимости просто не встречаются в природе — «Бог изощрен, но не злонамерен».

Принцип сходства лежит в основе метода *ближайшего соседа* (nearest neighbor), который относит произвольный объект $x$ к тому классу, которому принадлежит ближайший объект обучающей выборки $X^\ell$. Это, пожалуй, самой простой из всех алгоритмов классификации. В нем нет настраиваемых параметров, обучение сводится к элементарному запоминанию выборки. На принципе близости основаны также методы кластеризации, непараметрической регрессии, многомерного шкалирования.

Наиболее тонкий вопрос для всех метрических алгоритмов анализа данных — как построить метрику $\rho$. Если объекты представлены своими признаковыми описаниями, то можно взять евклидово расстояние между объектами:

$$\rho^2(x, x') = \sum_{j=1}^{n} (f_j(x) - f_j(x'))^2,$$



однако это далеко не единственный вариант, и часто далеко не самый лучший.

*Принцип регуляризации*. Переобучение часто возникает при использовании чрезмерно сложных моделей алгоритмов. Модели, обладающие избыточным числом свободных параметров, позволяют точнее воспроизводить ответы на материале обучения. Однако попытка описать обучающие данные точнее, чем в принципе позволяет суммарная погрешность измерений и самой модели, может привести к катастрофическому снижению обобщающей способности. На практике любая модель не точна, поэтому проблема переобучения носит всеобщий характер в машинном обучении.

Известный философский принцип *бритвы Оккама* гласит, что из множества допустимых решений всегда следует выбирать наиболее простое. В частности, модель алгоритмов не должна иметь избыточных параметров.

Справедливости ради отметим, что применение сложных моделей не всегда ведет к переобучению. Известны методы, позволяющие находить достаточно надежные алгоритмы в очень сложных алгоритмических моделях, например, методы обучения алгоритмических композиций [25]. Сложность модели — довольно тонкая характеристика. Это количество алгоритмов в модели, но не всех, а только тех, которые могут быть получены в результате обучения. То есть сложность зависит не только от модели алгоритмов, но и от восстанавливаемой зависимости, и от метода обучения, и даже от самой обучающей выборки. Увеличение числа параметров модели не влечет повышения сложности, если в процессе настройки на эти параметры накладываются определенные ограничения.

Один из способов ограничения сложности состоит в том, чтобы отойти от принципа минимизации эмпирического риска и добавить к функционалу $Q(a, X^\ell)$ штрафное слагаемое, «наказывающее» чрезмерно сложные модели:

$$\mu(X^\ell) = \operatorname*{argmin}_{a \in A} (Q(a, X^\ell) + \tau C(a)),$$

где число $\tau$ называется *параметром регуляризации*, функционал $C(a)$ выражает сложность алгоритма $a$. Это и есть принцип регуляризации некорректно поставленных задач по А. Н. Тихонову [16]: если решение существует, но оно не единственно или неустойчиво, то из множества возможных решений следует выбрать такое, которое минимизирует дополнительный критерий регуляризации $C(a)$.

Существует масса способов задать штрафное слагаемое $C(a)$. Простейшая эвристика — взять в качестве $C(a)$ число настраиваемых параметров алгоритма $a$. В алгоритмах классификации и регрессии, линейных по вектору параметров $\gamma$, часто применяется другая эвристика — взять в качестве штрафа норму этого вектора: $C(a) = \|\gamma\|$. Существуют и другие разно-

видности штрафных функционалов, основанные на теоретических оценках обобщающей способности [20].

*Принцип разделимости* относится к задачам классификации. Он предполагает, что объекты в пространстве $X$ могут быть разделены некоторой поверхностью. Например, *линейная разделимость* двух классов в евклидовом пространстве $X$ означает, что существует гиперплоскость, относительно которой точки одного класса лежат по одну сторону, а точки второго класса — по другую.

Пусть $Y = \{-1, +1\}$ и объекты описываются признаками $f_1, \ldots, f_n$. *Линейным разделяющим правилом* называется алгоритм классификации вида

$$a(x) = \operatorname{sign}(\alpha_1 f_1(x) + \ldots + \alpha_n f_n(x)),$$

где весовые коэффициенты $\alpha_1, \ldots, \alpha_n$ являются параметрами алгоритма и настраиваются по обучающей выборке $X^\ell$.

Наиболее известные методы построения линейных разделяющих правил — *метод опорных векторов* (support vector machine, SVM) и *логистическая регрессия* (logistic regression). SVM исходит из дополнительного требования, чтобы расстояние от разделяющей поверхности до ближайших объектов выборки было максимальным. Известно, что максимизация *зазора* (margin) между классами способствует более уверенной классификации и улучшает обобщающую способность.

Неявно принцип разделимости присутствует всегда, когда алгоритм классификации строится в виде $a(x) = C(b(x))$. Здесь функция $b(x)$ дает числовую оценку принадлежности объекта классу и называется *алгоритмическим оператором* или *вещественнозначным классификатором* (real-valued classifier); функция $C(b)$ переводит оценку принадлежности собственно в номер класса и называется *решающим правилом*. Обычно $C$ имеет очень простой вид. Например, в случае $Y = \{-1, +1\}$ естественно выбрать функцию $C(b) = \operatorname{sign}(b)$.

Если $Y = \{-1, +1\}$, то величина $m_i = b(x_i)y_i$ называется *отступом* (margin) объекта $x_i$ от поверхности, разделяющей классы. Отступ $m_i$ отрицателен тогда и только тогда, когда алгоритм допускает ошибку на объекте $x_i$. Распределение отступов обучающих объектов характеризует геометрию взаимного расположения классов. Аккуратный учет этой важной дополнительной информации способствует повышению качества классификации [27, 28].

*Принцип отделимости* заключается в том, чтобы в пространстве объектов $X$ строить области, каждая из которых отделяла бы объекты только какого-то одного из классов. Геометрическую форму этих областей предпочитают выбирать попроще: как правило, это шары, гиперплоскости



или гиперпараллелепипеды. Поэтому эти области называют также *эталонными*.

В общем случае *правило* класса $y \in Y$ — это предикат $\varphi_y \colon X \to \{0, 1\}$. Если $\varphi_y(x) = 1$, то говорят, что правило $\varphi_y$ *покрывает* объект $x$ и относит его к классу $y$. Если $\varphi_y(x) = 0$, то считается, что правило ничего не знает о классовой принадлежности объекта $x$, фактически отказывается от его классификации.

В отличие от принципа разделимости, здесь не ставится задача классифицировать всю выборку с помощью одной-единственной поверхности. Вместо этого строится множество правил, и каждый отделяет лишь небольшую часть своего класса.

Существует несколько способов собрать алгоритм классификации из набора правил $\varphi_{y1}, \ldots, \varphi_{yT_y}$. Чаще других используется *принцип голосования*: объект $x$ относится к тому классу, за который голосует наибольшее число правил:

$$a(x) = \arg\max_{y \in Y} \sum_{t=1}^{T_y} \varphi_{yt}(x).$$

Для построения отдельных правил применяется *принцип закономерности*. Пусть $p(\varphi_y)$ — число *позитивных* объектов $x_i \in X^\ell$, правильно покрываемых правилом $\varphi_y$, то есть для которых $\varphi_y(x_i) = 1$ и $y = y_i$. Соответственно, $n(\varphi_y)$ — число ошибочно покрываемых, или *негативных*, объектов, для которых $\varphi_y(x_i) = 1$, но $y \neq y_i$. Эталон $\varphi_y(x)$ называется *закономерностью*, если он покрывает достаточно много объектов и при этом допускает достаточно мало ошибок [12]:

$$p(\varphi_y) \geqslant \alpha, \quad n(\varphi_y) \leqslant \beta,$$

где параметры $\alpha$ и $\beta$ выбираются из априорных соображений, в зависимости от особенностей конкретной задачи.

*Принципы самоорганизации и селекции моделей*. Большой проблемой является выбор модели алгоритмов $A$, и далеко не всегда этот выбор обоснован какими-либо содержательными соображениями. Часто применяется линейный классификатор или линейная регрессия — просто потому, что их легче строить и соответствующий метод обучения находится «под рукой».

*Принцип самоорганизации моделей* предполагает, что структура модели $a(x, \gamma)$ не известна заранее и выбирается из некоторого множества альтернатив, настолько богатого, что с его помощью можно описать практически любую зависимость. Метод обучения на основе самоорганизации решает две совершенно разные задачи — выбирает структуру модели и настраивает вектор параметров $\gamma$ в выбранной модели. Первая задача

решается путем направленного перебора большого числа моделей, при этом лучшая модель выбирается по *внешнему критерию*. Вторая задача решается путем оптимизации так называемых *внутренних критериев*. Принципы самоорганизации моделей, внешних и внутренних критериев были предложены А. Г. Ивахненко еще в 1968 г. и легли в основу широко известного *метода группового учета аргументов*, *МГУА* (group method of data handling, GMDH) [11].

Поясним различие между внутренними и внешними критериями на примере двух простейших частных случаев самоорганизации.

Задача *выбора модели* (model selection) заключается в следующем. Имеется множество альтернативных методов обучения $\mu_1, \ldots, \mu_T$. Каждый метод применяется к обучающей выборке, в результате строится множество алгоритмов $a_t = \mu_t(X^\ell)$, $t = 1, \ldots, T$. Возникает вопрос: какой алгоритм лучше? Оставить алгоритм с наименьшим значением эмпирического риска $Q_t = Q(a_t, X^\ell)$ было бы неверно, так как значение $Q_t$ является заниженной оценкой ожидаемого риска. Это связано с явлением переобучения. Возможны ситуации, когда все значения $Q_t$ одинаковы и равны нулю, тогда выбрать лучший алгоритм вообще не удастся. Функционал $Q_t$ называется *внутренним критерием*, так как он используется в рамках конкретной модели для настройки ее параметров. Для выбора лучшей из $T$ моделей внутренний критерий не годится. Необходимо привлекать внешние данные, которые не были задействованы в процессе обучения. В простейшем случае исходная выборка разбивается на две части: обучающую $X^\ell$ и контрольную $X^k$, после чего лучшая модель выбирается по *внешнему критерию* минимума средней ошибки на контрольных данных:

$$t^* = \operatorname*{argmin}_{t=1,\ldots,T} Q(\mu_t(X^\ell), X^k).$$

Более сложный пример самоорганизации — задача *выбора информативных признаков* (features selection). На этапе формирования исходных данных, как правило, неизвестно, какие признаки и в каком сочетании окажутся наиболее важными. Поэтому в признаковые описания включаются все данные об объектах, которые только доступны, и задача объективного выделения наиболее значимой части информации возлагается на алгоритм обучения. К сожалению, задача поиска информативных наборов признаков является *NP*-полной, то есть в общем случае требует полного перебора $2^n$ вариантов, где $n$ — число признаков. На практике применяются эвристические схемы перебора. В некоторых методах, таких как шаговая регрессия или МГУА, формируется только один информативный набор. Другие методы используют *принцип голосования*: строится большое количество информативных наборов, и окончательный ответ получается путем усреднения их ответов.



Такая стратегия применяется в некоторых логических алгоритмах классификации и алгоритмах вычисления оценок (ABO) [7, 8]. Но это уже относится к *принципу композиции*. При решении сложных задач классификации, регрессии и прогнозирования часто возникает следующая ситуация. Одна за другой предпринимаются попытки построить алгоритм, восстанавливающий искомую зависимость, однако качество всех построенных алгоритмов оставляет желать лучшего. В таких случаях имеет смысл объединить несколько алгоритмов в композицию, в надежде на то, что погрешности этих алгоритмов взаимно скомпенсируются.

В простейшем случае *алгоритмической композицией*, составленной из *базовых алгоритмов* $a_1, \ldots, a_T\colon X \to Y$ и *корректирующей операции* $F\colon Y^T \to Y$, называется алгоритм вида $a(x) = F(a_1(x), \ldots, a_T(x))$, $x \in X$.

Выделяются два основных принципа построения алгоритмических композиций — *усреднение* и *специализация*.

Простейшим примером усреднения является среднее арифметическое. Более общий случай — взвешенное среднее:

$$a(x) = \sum_{t=1}^{T} w_t a_t(x), \qquad \sum_{t=1}^{T} w_t = 1, \qquad x \in X,$$

где $w_t$ — весовые коэффициенты. Обычно предполагается, что вес базовых алгоритмов неотрицателен и что вес $w_t$ тем больше, чем выше качество алгоритма $a_t$. Для настройки весов можно применять стандартные линейные методы классификации и регрессии, рассматривая векторы $(a_1(x), \ldots, a_T(x))$ как признаковые описания объектов $x \in X$. Существуют и специализированные методы настройки весов в линейных композициях, например, метод бустинга [25].

Согласно *принципу специализации* пространство объектов делится на области, в каждой из которых настраивается свой алгоритм. Исходная задача разбивается на более простые подзадачи по принципу «разделяй и властвуй». Формально алгоритм представляется также в виде линейной комбинации, однако теперь весовые коэффициенты $w_t$ не постоянны, а зависят от положения объекта в пространстве $X$ и называются *функциями компетентности*:

$$a(x) = \sum_{t=1}^{T} w_t(x) a_t(x), \qquad \sum_{t=1}^{T} w_t(x) = 1, \qquad x \in X.$$

Здесь предполагается, что $w_t\colon X \to [0, 1]$. Чем больше значение функции компетентности на объекте $x$, тем больше вклад алгоритма $a_t$ в результат композиции. Если функция $w_t$ принимает только два значения $\{0, 1\}$, то множество $\{x \mid w_t(x) = 1\}$ называется *областью компетентности*

базового алгоритма $a_t$ [15]. В общем случае функция $w_t$ описывает область компетентности как нечеткое множество. Методы построения таких композиций, называемых *смесями экспертов* (mixture of experts), рассматриваются в [23, 26].

На принципе композиции основаны методы алгебраического подхода [8], взвешенное голосование, бустинг и бэггинг [21, 25], метод комитетов [13], нейронные сети [19].

## 6. Источники данных для задач моделирования автомобильного транспорта

Вообще говоря, для сбора информации о транспортных системах можно использовать достаточно много различных способов. Наиболее распространенным способом является использование детекторов различного типа. Часть из них (индукционные датчики) могут только подсчитывать количество транспортных средств, проходящих мимо детектора, другие же (радарные детекторы) могут также измерять скорость транспортных средств и классифицировать их по размерным типам. К недостаткам детекторов как способу сбора данных относится то, что информация собирается в конкретных точках, при этом для получения относительно полной картины требуется достаточно развитая сеть детекторов, а также использование средств моделирования для экстраполирования получаемой информации на неохваченные участки сети.

Другим распространенным источником данных являются фото- и видеокамеры. Как правило, камеры устанавливают для визуального контроля дорожной ситуации, а также для фиксации нарушений правил дорожного движения. Если анализировать поток изображений с помощью специального программного обеспечения, то можно получать информацию о плотности и скорости потоков. В некоторых случаях можно дополнительно собирать информацию о государственных номерах и типах транспортных средств, попадающих в поле зрения камеры. Кроме того, если несколько камер снимают смежные участки дорожной сети, появляется возможность восстановления траекторий транспортных средств. Недостатки этого способа сбора данных повторяют недостатки детекторов, кроме этого, фото- и видеокамеры чувствительны к погодным условиям и недостаточной освещенности. Последний недостаток, впрочем, может компенсироваться дополнительным оборудованием, способным работать в инфракрасном диапазоне, но такое оборудование требует дополнительных расходов.

Менее распространенным способом получения данных является дистанционная, как правило авиационная или космическая, фотосъемка. Для решения некоторых задач можно использовать спутниковые изображения крупных городов, которые предоставляются общедоступными сервисами



типа «Карты Google» (`http://maps.google.ru/`) или «Яндекс.Карты» (`http://maps.yandex.ru/`). Примером такой задачи может служить определение реальной полосности дорог улично-дорожной сети (см. раздел «Исследовательские задачи...» на с. 400). Существуют примеры [18], когда аэрофотосъемку высокого разрешения использовали в том числе для оценки плотности потоков, соотношения количеств типов транспортных средств, расчета степени сужения дорог за счет припаркованных транспортных средств и т. п. Такой способ получения данных, в отличие от детекторов и камер, позволяет получать полную информацию о состоянии транспортной системы на больших площадях. Недостатками являются низкая оперативность получения информации, высокая стоимость, практически отсутствующие возможности по управлению периодичностью сбора данных и высокая зависимость от погодных условий.

Немаловажной частью требуемой для идентификации моделей информации оказывается информация о спросе на перемещения. Обычно выделяют две компоненты такого спроса: пассажирскую и грузовую. Естественно, только часть грузо- и пассажироперемещений осуществляется автомобильным транспортом. Тем не менее, в силу того, что использование различных видов транспорта для перемещений находится в динамическом равновесии, знание оценок полного спроса является очень полезным. Как правило, информация о спросе на перемещения формируется в виде матрицы, в строках и столбцах которой перечислены районы отправления и прибытия, а на пересечениях находятся оценки спроса, выражаемые в количестве людей или тонн грузов. Такая матрица обычно называется матрицей корреспонденций (подробнее см. раздел 1.5 настоящего пособия).

Необходимой информацией для построения матрицы корреспонденций являются данные о застройке: тип застройки (жилая, промышленная, офисная, торговая, развлекательная), плотность, этажность, наличие объектов притяжения грузов и населения, режим работы этих объектов. Такие данные дают возможность оценить агрегированные въездные и выездные потоки для каждого района. Информация о корреспондировании этих потоков с потоками из других районов, как правило, восстанавливается на основе данных социологических опросов. Недостатки такого способа сбора информации очевидны: используется принципиально субъективная информация; очень сложно контролировать степень смещенности получаемых оценок; требуется опрашивать значительную долю населения (речь идет о единицах процентов, что, например, для Москвы приводит к необходимости опрашивать сотни тысяч человек) и еще бóльшую долю предприятий, создающих грузовые потоки, что делает дорогостоящим регулярное проведение таких опросов.

Важно также отметить, что для России и многих других стран (в отличие, например, от США) характерной является высокая доля общественного транспорта в удовлетворении спроса на перемещения населения. Это приводит к необходимости получения данных о пассажирских перевозках с помощью общественного транспорта. Для некоторых видов транспорта, например метро, легко получить данные об объемах пассажиропотока через каждую конкретную станцию. Для наземного транспорта (автобус, троллейбус, трамвай) в принципе может быть доступна только информация о количестве входящих пассажиров, если транспортные средства оборудованы валидаторами. Для получения информации о количестве выходящих пассажиров требуется дооборудование транспортных единиц. Во всех случаях отсутствует прямая возможность получения информации о точных начальном и конечном пунктах поездки пассажиров. Она может быть восстановлена лишь частично в случае использования значительной частью пассажиров билетов на большое количество поездок. Тогда, весьма вероятно, место начала обратной поездки будет указывать на пункт назначения прямой поездки.

Относительно недавно стал использоваться способ сбора информации о транспортных потоках, основанный на получаемых от агентов обезличенных треках. Агентом называется транспортное средство, оборудованное устройством, получающим со спутника свои текущие координаты (с помощью систем GPS или ГЛОНАСС) и передающим эти координаты в центр обработки. Таким устройством может быть, например, обычный мобильный телефон, оборудованный приемником спутникового сигнала, или навигатор, совмещенный с GPRS или 3G-модемом. Современные мобильные телефоны позволяют также восстанавливать/уточнять координаты с использованием сигналов от базовых станций в условиях плотной городской сети. Коммерческие компании (грузоперевозчики, таксопарки) используют для этих целей специализированные устройства. В центре полученные координаты анализируются, на их основе восстанавливается трек движения транспортного средства, трек накладывается на координаты улично-дорожной сети и корректируется для устранения ошибок в позиционировании; вычисляются мгновенные/средние скорости транспортного средства. Наиболее известен российскому пользователю основанный на этой технологии сервис «Яндекс.Пробки» [10]. Такой способ сбора информации, в отличие от детекторов или видеокамер, позволяет получать актуальную информацию обо всей или о значительной части транспортной сети, при этом сборщик информации не несет затрат на оборудование транспортных средств или дорожной инфраструктуры, кроме того, для некоторой части транспортных средств существует возможность получения некоторой информации о пункте выезда и пункте назначения. Недостатками являются: необходимость постоянного присутствия на дорогах некоторой минимальной доли агентов относительно общего количества транспортных средств; отсутствие информации о плотности потоков; необходимость обнаруживать и исправлять ошибки в позиционировании, присущие системам спутниковой навигации.



## 7. Проблемы использования реальных данных

В прикладных задачах исходные данные отражают сложность и разнообразие реальных процессов и явлений. Поэтому данные могут обладать рядом неприятных свойств, усложняющих поиск решения, причем эти свойства сочетаются друг с другом практически в любых комбинациях. Большое разнообразие методов обучения по прецедентам во многом объясняется именно тем, что для каждого из этих случаев приходится искать свои подходы.

Следует отметить, что в последнее время наметился определенный прогресс в отношении если не качества, то, по крайней мере, объемов и типов данных, используемых при решении транспортных задач. Государственные и муниципальные органы, а иногда и частные компании стали уделять больше внимания вопросам управления автомобильными транспортными потоками, что в свою очередь приводит к внедрению различного рода систем управления, так или иначе оснащенных системами сбора исходных данных. Значительную роль играет возможность спутникового позиционирования и мобильной связи. И все же пока рано говорить о полноте и хорошем качестве этих данных.

Далее мы перечислим наиболее распространенные недостатки исходных реальных данных.

*Неточность данных*. Значения признаков $f_j(x_i)$ и целевой переменной $y_i$ могут измеряться с погрешностями. В некоторых случаях возможны грубые ошибки, приводящие к появлению редких, но больших отклонений — *выбросов* (outliers).

Очевидно, что ни один из перечисленных источников данных не обеспечивает 100% точности. Опасность зашумленных данных в том, что обучаемые алгоритмы могут настраиваться на восстановление не только целевой зависимости, но и шума. В первую очередь, это относится к сложным моделям с большим числом свободных параметров. Для обеспечения помехоустойчивости методов обучения применяют оптимизацию сложности модели по критерию *скользящего контроля*. Для корректной обработки выбросов применяют предварительную фильтрацию данных или робастные методы оценивания параметров модели.

Помимо помехоустойчивости методов может использоваться валидация источников данных, то есть проверка соответствия реальности данных из источника. Для части источников (детекторов, фото- и видеокамер)

способом валидации является организация подсчета транспортных средств (пассажиров, единиц общественного транспорта) с помощью специальных сотрудников — асессоров. Для других источников информации трекового типа валидация может проводиться с помощью машин-асессоров. Естественно, что организация процедуры валидации источников данных, во-первых, является дорогостоящей, как с точки зрения финансовых расходов, так и с точки зрения потерь времени, во-вторых, не гарантирует абсолютную точность собираемых данных.

Значения признаков $f_j(x_i)$ могут вообще отсутствовать, в этом случае говорят, что в данных имеются *пропуски* (missing data). Например, средствами сбора трековых данных оборудована только малая часть всех транспортных средств. Соответственно по значительной части улиц города в определенные моменты может просто не оказаться данных для оценки скоростей потоков.

Примером неполноты транспортных данных является ограниченность сети детекторов, измеряющих потоки на улицах города. Очевидно, что по той части дорожной сети, где детекторы не установлены, информация отсутствует. Многие алгоритмы вообще не могут работать с пропущенными данными. В таких случаях применяют предварительную обработку — заполняют пропуски «спрогнозированными» значениями. Идея заключается в том, чтобы для каждого признака, имеющего пропущенные значения, решить вспомогательную задачу обучения по прецедентам. В качестве обучающей выборки берутся все объекты, для которых значение данного признака известно. Затем строится алгоритм, восстанавливающий зависимость данного признака от остальных. Этот алгоритм применяется к оставшимся объектам для заполнения пропусков. Другой подход — использовать алгоритмы, которые умеют игнорировать отдельные пропуски, не теряя при этом всей остальной информации. Например, таким свойством обладают логические алгоритмы классификации.

*Противоречивость данных*. Прецеденты $i$ и $k$ могут противоречить друг другу. Например, $x_i = x_k$, но $y_i \neq y_k$. Они также могут противоречить априорным ограничениям. Например, может быть заранее известно, что зависимость $y^*$ является монотонной, но $(x_i < x_k) \wedge (y_i > y_k)$ для некоторой пары прецедентов $(i, k)$. Противоречивость может быть следствием неточности данных. Для обеспечения непротиворечивости применяют предварительную фильтрацию, отсеивая или заменяя противоречивые данные.

*Разнородность данных*. Признаки $f_1, \ldots, f_n$ могут иметь различные типы (измеряться в разных шкалах). Некоторые методы обучения требуют однородности пространства признаков и приводят к неадекватным решениям, если это требование нарушается. При наличии признаков разного типа данные также проходят предварительную обработку. Например, в логических алгоритмах классификации все признаки в конечном итоге



преобразуются к бинарному типу (такая обработка может выполняться «на лету» в самом алгоритме).

*Сложная структура данных*. Данные могут быть представлены в более сложной форме, чем стандартная матрица объектов-признаков. Это могут быть изображения, сигналы, тексты, графы, таблицы базы данных и т. д. Для извлечения признаков применяют различные методы предварительной обработки данных. В задачах распознавания изображений, например при определении государственных номеров транспортных средств, извлечение признаков является существенно более трудным этапом, в значительной степени предопределяющим успех распознавания.

В некоторых задачах построение признаковых описаний оказывается вообще нецелесообразным, и объекты непосредственно сравниваются друг с другом. Тогда говорят о беспризнаковом распознавании. Например, сходство контуров транспортных средств можно оценивать как работу, необходимую для преобразования одного контура в другой, если представлять их в виде пластичных проволок.

*Недостаточность данных* (*проблема малых выборок*). Объектов может оказаться существенно меньше, чем признаков. В этом случае многие классические методы статистики и аппроксимации функций становятся неустойчивыми или вообще неприменимыми. Приходится прибегать к более изощренным техникам: упрощать модель, оставляя только самые информативные признаки; искать закономерности, образованные короткими наборами признаков; привлекать дополнительную информацию в непрецедентной форме.

Как правило, проблема недостаточности данных возникает, когда получение информации об объектах требует значительного времени или серьезных финансовых расходов. Например, в описанном в [18] проекте аэрофотосъемка Москвы проводилась один раз.

*Избыточность данных* (*проблема сверхбольших выборок*). В системах с автоматическим сбором и накоплением данных возникает противоположная проблема: данных настолько много, что обычные методы обрабатывают их крайне медленно. В зависимости от целей анализа могут применяться различные методы фильтрации или агрегирования данных. Например, эффективные субквадратичные алгоритмы кластеризации способны быстро выделить в массе исторических данных группу ситуаций, схожих с текущей наблюдаемой ситуацией. Также сверхбольшие данные ставят проблему эффективного хранения. Хранилища данных должны проектироваться с учетом объемов, типов источников, состава решаемых задач и требований к производительности.

Необходимо отметить, во-первых, что в связи с вышеперечисленными проблемами для практически любых первичных транспортных данных необходимо создавать процедуру предобработки, учитывающую специфику

этих данных и специфику моделей, в которых предобработанные, очищенные, улучшенные данные будут использоваться.

Во-вторых, специфика транспортных систем состоит в том, что источники данных имеют весьма различную природу. Это, с одной стороны, создает проблемы предобработки и использования, но, с другой стороны, позволяет верифицировать данные какого-либо одного источника путем использования данных из другого источника. Такая многозначность дает возможность пространственной интерполяции данных, получаемых из локализованных источников (например, детекторов или видеокамер) или временной интерполяции (заполнения пропусков) для трековых данных.

## 8. Информационное моделирование транспортных потоков

Особенностью моделирования транспортных потоков, как, впрочем, и многих других областей, является сложность построения адекватных математических моделей при решении целого ряда задач. В отдельных случаях модель может быть даже известна, но настолько сложна, что ни адаптировать, ни просчитать ее за разумное время не представляется возможным.

В таких случаях от построения детальной физической модели вполне можно оказаться. Обычно заказчика интересует не всестороннее изучение системы или явления, а регулярное решение конкретных задач, связанных с прогнозированием и принятием управленческих решений. Но тогда имеет смысл моделировать не саму транспортную систему, а лишь некоторые ее информационные проявления. Сложность системы компенсируется ограниченностью возможных действий по управлению системой. Информационная модель, в отличие от «физической», основывается не столько на экспертных знаниях о предметной области, сколько на общих принципах преобразования информации. Такие модели называют также эвристическими, поскольку они конструируются в значительной степени исходя из соображений здравого смысла, зачастую без строгого «физического» обоснования. Вообще, процесс построения математических моделей в прикладных задачах можно разделить на два этапа.

Первый этап — формализация экспертных знаний о предметной области, в результате которой формируется структура модели. При построении физических моделей этот этап наиболее важен. В хорошей физической модели остается, как правило, небольшое число свободных параметров, имеющих четкую содержательную интерпретацию. Эти модели узко специализированы и имеют фиксированные границы применимости.

Второй этап — настройка (идентификация) параметров модели по эмпирическим (экспериментальным) данным. Он существенно более важен для информационных моделей, когда данные являются едва ли не единственным, на что можно опереться. Информационные модели имеют значительно больше свободных параметров, зачастую не поддающихся содержательной интерпретации. Среди информационных моделей немало «черных ящиков», которые способны неплохо решать практические задачи, но внутренняя логика этих решений остается загадкой даже для экспертов в данной прикладной области.



Четкого различия между физическими и информационными моделями нет. Чем больше знаний о предметной области удается привлечь на первом этапе построения модели, тем более она физична.

### Литература


1. *Арлазаров В. Л.*, *Славин О. А.*, *Хованский А. Г.* Оценка расстояния между изображениями при параллельном переносе // Доклады Академии наук. 2011. Т. 437, № 3. С. 313–315.
2. *Вапник В. Н.* Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979.
3. *Вапник В. Н.*, *Червоненкис А. Я.* О равномерной сходимости частот появления событий к их вероятностям // ДАН СССР. 1968. Т. 181, № 4. С. 781–784.
4. *Вапник В. Н.*, *Червоненкис А. Я.* О равномерной сходимости частот появления событий к их вероятностям // Теория вероятностей и ее применения. 1971. Т. 16, № 2. С. 264–280.
5. *Вапник В. Н.*, *Червоненкис А. Я.* Теория распознавания образов. М.: Наука, 1974.
6. *Григорьев А. С.*, *Николаев Д. П.*, *Ханипов Т. М.* Определение количества осей транспортного средства по видеоряду проезда // Труды 54-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук»: Часть IX. Инновации и высокие технологии. М.: МФТИ, 2011.
7. *Журавлёв Ю. И.*, *Никифоров В. В.* Алгоритмы распознавания, основанные на вычислении оценок // Кибернетика. Киев. № 3. 1971.
8. *Журавлёв Ю. И.* Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. 1978. Т. 33. С. 5–68.
9. *Ивкин Н. П.*, *Чехович Ю. В.* Краткосрочное прогнозирование скоростей транспортных потоков // Интеллектуализация обработки информации: 9-я международная конференция. Республика Черногория, г. Будва, 2012 г.: Сборник докладов. М.: Торус Пресс, 2012. С. 219–222.
10. Как работают «Яндекс.Пробки»; `http://company.yandex.ru/technologies/yaprobki/`
11. *Ивахненко А. Г.*, *Юрачковский Ю. П.* Моделирование сложных систем по экспериментальным данным. М.: Радио и связь, 1987.
12. *Лбов Г. С.* Методы обработки разнотипных экспериментальных данных. Новосибирск: Наука, 1981.
13. *Мазуров В. Д.* Метод комитетов в задачах оптимизации и классификации. М.: Наука, 1990.

14. *Павловский Ю. Н.* Имитационные модели и системы (Математическое моделирование, вып. 2). М.: ФАЗИС: ВЦ РАН, 2000, 134 с.
15. *Растригин Л. А.*, *Эренштейн Р. Х.* Коллективные правила распознавания. М.: Энергия, 1981.
16. *Тихонов А. Н.*, *Арсенин В. Я.* Методы решения некорректных задач. М.: Наука, 1986.
17. *Усилин С. А.*, *Постников В. В.*, *Николаев Д. П.* Поиск объектов в видеопотоке при известных кинематике и геометрической модели сцены // Труды 53-й научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук». Часть IX. Инновации и высокие технологии. 2010. С. 67–68.
18. *Фролов К. В.*, *Лебедев В. В.*, *Воробьев А. Ю.*, *Гаврилов В. И.*, *Харитонов В. А.* Система дистанционного мониторинга транспортных потоков основных магистралей центра Москвы // Проблемы машиностроения и надежности машин. № 5. 2000. С. 3–10.
19. *Хайкин С.* Нейронные сети: Полный курс. М.: ИД Вильямс, 2006.
20. *Boucheron S.*, *Bousquet O.*, *Lugosi G.* Theory of classification: A survey of some recent advances //ESAIM: Probability and Statistics. 2005. № 9. P. 323–375.
21. *Breiman L.* Arcing classifiers //The Annals of Statistics. 1998. V. 26, № 3. P. 801–849.
22. *Downs A.* The law of peak-hour expressway congestion // Traffic Quarterly 1962. Vol. 16, № 3. P. 393–409.
23. *Fritsch J.*, *Finke M.*, *Waibel A.* Adaptively Growing Hierarchical Mixtures of Experts // Advances in Neural Information Processing Systems. The MIT Press. 1997. V. 9. P. 459–465.
24. *Frawley W.*, *Piatetsky-Shapiro G.*, *Matheus C.* Knowledge Discovery in Databases: An Overview // AI Magazine. 1992. P. 213–228.
25. *Freund Y.*, *Schapire R. E.* Experiments with a new boosting algorithm // International Conference on Machine Learning. 1996. P. 148–156.
26. *Jacobs R. A.*, *Jordan M. I.*, *Nowlan S. J.*, *Hinton G. E.* Adaptive Mixtures of Local Experts // Neural Computation, 1991. № 3. P. 79–87.
27. *Garg A.*, *Har-Peled S.*, *Roth D.* On generalization bounds, projection profile, and margin distribution // ICML'02. 2002.
28. *Garg A.*, *Roth D.* Margin distribution and learning algorithms // ICML'03. 2003. P. 210–217.
29. *Herbrich R.*, *Williamson R. C.* Algorithmic luckiness // Journal of Machine Learning Research. 2002. № 3. P. 175–212.
30. *Langford J.* Quantitatively Tight Sample Complexity Bounds. Thesis (Ph.D.) — Carnegie Mellon Univercity. 2002.
31. *Rückert U.*, *Kramer S.* Towards tight bounds for rule learning // Proc. 21th International Conference on Machine Learning, Banff, Canada. 2004. P. 90.

**Е. В. Гасникова**

# О возможной динамике в модели расчета матрицы корреспонденций

## Введение

После работ Э. Т. Джейнса конца 50-х годов XX века [1], А. Дж. Вильсона конца 60-х годов XX века [2], И. Пригожина с коллегами, Г. Хакена в 70-е годы XX века [3, 4] в литературе достаточно прочно укрепилась концепция о плодотворности перенесения термодинамического формализма (см., например, [5–14] и цитированную там литературу) на различные макросистемы, в частности, встречающиеся в экономике, биологии, социальной сфере [2–4, 15–24]. В России систематические исследования в этом направлении были предприняты Л. Н. Розоноэром в начале 1970-х [25] (см. также [26–34] и цитированную там литературу). Упомянутая концепция часто используется для нахождения равновесия макросистемы. А именно, по аналогии с феноменологической термодинамикой, вводится вероятностное распределение на множестве состояний, в которых может пребывать макросистема. Такое распределение может, например, совпадать с *инвариантной мерой эргодической динамической системы*, порождающей рассматриваемую макросистему [11], или с *финальным* (равным *стационарному*) распределением *эргодического* (например, марковского) случайного процесса, порождающего рассматриваемую макросистему [35–40]. Если размерность макросистемы увеличивается, то, как правило, распределение сосредотачивается в окрестности наиболее вероятного макросостояния [1)]. Таким образом, с ростом времени наблюдения за макросистемой и ее размерности 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, а 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

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

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] [1)].

## 1. Возможная динамика, приводящая в асимптотике (по времени) к статической модели А. Дж. Вильсона расчета матрицы корреспонденций

Рассмотрим для начала более простой пример, иллюстрирующий формализм, описанный во введении.

**Пример 1** (кинетика социального неравенства [23,47]). В городе живет $N \gg 1$ (например, 10 000) пронумерованных жителей. У каждого $i$-го жителя есть в начальный (нулевой) момент времени целое (неотрицательное) количество рублей $s_i(0)$ (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 $t \geqslant 0$ $r$-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 $l$ $(1 \leqslant r < l \leqslant N)$ встретятся и попробуют разыграть один рубль по следующему правилу: с вероятностью $1/2$ житель с бóльшим номером отдает 1 рубль (если, конечно, он не банкрот) жителю с меньшим номером и с вероятностью $1/2$ наоборот. Будем считать, что $p_{k;m}(t) \equiv \varkappa N^{-1}$ $(\varkappa > 0)$. При этом «в среднем» в единицу времени осуществляется $\varkappa N/2$ встреч. Т. е., скажем, при $\varkappa = 1$ 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.

[1)] Укажем некоторые часто встречающиеся в приложениях [41–45] формализмы, также приводящие к задачам оптимизации энтропийно-подобных функционалов: принцип наибольшего правдоподобия при оценке неизвестных параметров по простой выборке; принцип максимума апостериорной вероятности; наименьшее отклонение в смысле расстояния Кульбака—Лейблера (энтропийное расстояние) [46]; принцип наименьшей неопределенности (энтропия — мера неопределенности) в теории информации (рассуждения опираются в ряде случаев на теорему Шеннона—Мак-Миллана—Бреймана). Важно также отметить, что энтропийно-подобный функционал часто является *функцией Ляпунова* для динамической системы (например, системы обыкновенных дифференциальных уравнений, итерационного процесса, уравнений 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]).



Пусть $c_s(t)$ — доля жителей города, имеющих ровно $s$ рублей в момент времени $t$ (заметим, что $c_s(t)$ — случайная величина). Пусть

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

Тогда по эргодической теореме для конечных однородных марковских цепей: [1)]

$$\forall q = 0, \ldots, S \;\; \exists \lambda_q > 0,\ T_q = O(N) \colon \forall t \geqslant T_q$$
$$P\left(\left|\frac{c_s(t)}{Ce^{-s/\bar{s}}} - 1\right| \leqslant \frac{\lambda_q}{\sqrt{N}},\ s = 0, \ldots, q\right) \geqslant 0{,}99,$$

где $C$ определяется из условия нормировки:

$$\sum_{s=0}^{S} Ce^{-s/\bar{s}} = 1, \quad \text{т. е.} \quad C \approx \frac{1}{\bar{s}}$$

(см. [18, 19, 35–38] и упражнение ниже). Скорость сходимости оценивается сверху, исходя из оценок в доказательстве эргодической теоремы для однородных марковских цепей с конечным числом состояний.

Как показывают численные эксперименты, [2)] 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 $\varkappa = 1$) установится «социальное неравенство». Заметим, что описанный выше случайный процесс обратим во времени. Однако наблюдается необратимая динамика $c_s(t)$.

**Замечание 1.** Для простоты формулировок в рамках этого замечания считаем время дискретным. Для оценки скорости сходимости необходимо асимптотически (по размеру макросистемы) оценить второе по величине модуля собственное значение матрицы переходных вероятностей — инфинитезимальной матрицы, определяющее основание геометрической прогрессии, мажорируемой последовательностью норм отклонений текущего состояния от стационарного в различные моменты времени [35–38]. (Первое собственное значение, которое для неотрицательных матриц часто называют *числом Фробениуса—Перрона*, равно единице, поскольку матрица стохастическая, т. е. все элементы неотрицательны и сумма всех элементов в любой строке равна единице.) Здесь нельзя не упомянуть о том, что

[1)] Эргодическая теорема используется для нахождения распределения случайных величин $c_s(t)$ на больших временах. Далее используются законы больших чисел или, другими словами, явление концентрации инвариантной (стационарной) меры, о котором мы подробнее поговорим в следующем примере. Точнее, не само это явление, а его следствие о том, что «хорошие» (например, липшицевы) функции на «хороших» компактных пространствах с мерой большого числа измерений почти везде близки к медиане [48].

[2)] В экспериментах, проведенных Т. А. Нагапетяном, получена оценка $T_q = 2N$, при довольно больших значениях $q$ и $\lambda_q \sim 1$. Недавно А. В. Колесникову удалось с помощью техники статьи [50] доказать эту оценку.

в этом направлении за последние несколько десятков лет произошла определенная революция [49], которую можно пояснить рассмотренным примером 1. Несложно проверить, что число (макро)состояний марковской цепи в этом примере растет быстрее чем экспоненциально с ростом $N$. В то время как по прошествии всего лишь $O(N)$ тактов распределение цепи будет уже довольно близко к финальному (предельному) = стационарному (инвариантному). Таким образом, если возникает потребность быстро сгенерировать дискретные случайные величины, которые могут принимать огромное число значений, то в ряде случаев удается подобрать такую марковскую цепь, которая быстро «выйдет» на стационарный режим, соответствующий желаемому распределению. Несколько интересных примеров в этом направлении (модель Изинга и др.) собрано, например, в современном курсе марковских случайных процессов [38]. Заметим, что при оценках второго по величине модуля собственного значения активно используется уже упоминавшийся принцип концентрации меры; см. [49, 50] и цитированную там литературу, а также приложение А. В. Колесникова.

Приведем отчасти схожую постановку задачи, восходящую к В. П. Маслову [33]. Ниже приведен фрагмент его интервью 2009 года, посвященного объяснению финансового кризиса 2008 года.

**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 выше написанного, с вероятностью, близкой к 1, доля банкротов будет примерно равна $1/\bar{s} = N/S$. Следовательно, количество банкротов не сильно отличается от $N/\bar{s} = N^2/S$. Если же, как в условии, $N \sim \sqrt{S}$, то можно считать, что банкротов практически нет.

**Упражнение*** (принцип сжимающих отображений и фокусирующие операторы, эргодическая теорема для конечных однородных марковских цепей [51]).

а) Покажите, что если оператор (вообще говоря, нелинейный) $A$ действует в полном метрическом пространстве $X$ и

$$\exists k \in \mathbb{N}\colon \forall x,\, y \in X \quad \rho(A^k(x),\, A^k(y)) \leqslant \theta\rho(x,\, y), \quad \theta \in (0,\, 1),$$

то

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

б) Пусть $X = \mathrm{P}\mathbb{R}^n_{++}$ — множество лучей пространства $\mathbb{R}^n$, лежащих во внутренности неотрицательного ортанта, на котором введена метрика Биркгофа:

$$\rho(x,\, y) = \ln \min\left\{\frac{\beta}{\alpha}\colon \alpha x \leqslant y \leqslant \beta x\right\} = \ln \min_{j,k=1,\ldots,n} \frac{x_j y_k}{x_k y_j}.$$

Здесь под элементами $x$ и $y$ в левой части равенства понимаются лучи, а вот в правой части уже какие-то векторы, лежащие на соответствующих лучах. Какие именно векторы — не важно. Покажите, что $X$ — полное метрическое пространство. Для матрицы $A = \|a_{ij}\|_{i,j=1}^n$ покажите, что если линейный оператор $A\colon X \to X$ положительный, т. е. $a_{ij} > 0\ \forall i,\, j = 1,\, \ldots,\, n$, то

$$\exists \theta \in (0,\, 1)\colon \forall x,\, y \in X \quad \rho(Ax,\, Ay) \leqslant \theta\rho(x,\, y).$$

в) (стохастический вариант теоремы Фробениуса—Перрона, или эргодическая теорема для конечных однородных дискретных марковских цепей (д.м.ц.)). Для стохастической матрицы $P$ размера $n \times n$ следующие условия равносильны:

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

2) $P$ — эргодическая матрица, т. е. $\exists \vec{p}^* > \vec{0}\colon \lim\limits_{m\to\infty} P^m = \big[\underbrace{\vec{p}^*,\, \vec{p}^*,\, \ldots,\, \vec{p}^*}_{n}\big]^T$, причем $\vec{p}^*$ является единственным решением системы:

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

**Замечание 2.** Из вида матрицы $\lim\limits_{m\to\infty} P^m = \big[\underbrace{\vec{p}^*,\, \vec{p}^*,\, \ldots,\, \vec{p}^*}_{n}\big]^T$ следует, что

$$\forall \vec{p}(0) \geqslant \vec{0}\left(\sum_{k=1}^n p_k(0) = 1\right) \quad \lim_{m\to\infty} \vec{p}(m) = \lim_{m\to\infty} (P^m)^T \cdot \vec{p}(0) = \vec{p}^*.$$

Это условие означает равенство финального распределения $\lim\limits_{m\to\infty} \vec{p}(m)$ стационарному $\vec{p}^*$ ($\vec{p}^* = P^T\vec{p}^*$) вне зависимости от начального распределения $\vec{p}(0)$. Заметим также, что условия 1), 2) равносильны следующим требованиям: конечная однородная д.м.ц. с матрицей переходных вероятностей $P$ — неразложимая, т. е. неприводимая (это означает, что из произвольного состояния «можно прийти» в любое наперед заданное), и непериодическая (Н.О.Д.$\{k\colon [P^k]_{11} > 0\} = 1$). Если убрать условие непериодичности, то

$$\exists!\, \vec{p}^* > \vec{0}\ (\vec{p}^* \in S)\colon \lim_{N\to\infty} \frac{1}{N} \sum_{m=1}^N P^m = \big[\underbrace{\vec{p}^*,\, \vec{p}^*,\, \ldots,\, \vec{p}^*}_{n}\big]^T \quad (\textit{сходимость по Чезаро}).$$

Если перейти к непрерывному времени, осуществляя соответствующий скейлинг, то легко показать, что необходимость в условии непериодичности исчезает. Если же цепь разложима, то система (S), вообще говоря, уже будет разрешима не единственным образом. Финальное распределение существует, но уже может зависеть

от того, с какого распределения стартуем. Соответствующий вариант эргодической теоремы приведен, например, в [37]. Доказательство в [37] также базируется на принципе сжимающих отображений. Другое, более вероятностное, доказательство эргодической теоремы для конечных однородных д.м.ц. имеется, например, в [35, 38] и базируется на каплинге (см. также приложение М. Л. Бланка).

**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), $\mathrm{x}_{ij}(t) \geqslant 0$ — число жителей, живущих в $i$-м районе и работающих в $j$-м в момент времени $t \geqslant 0$. Со временем пронумерованные жители (количество которых не меняется [1)] и равно $N = \sum\limits_{i=1}^{n} L_i = \sum\limits_{j=1}^{n} W_j$) меняют места жительства (квартиры). Для простоты будем считать, что отмеченные изменения могут происходить только за счет обмена квартирами, т. е.

$$\mathrm{x}_{ij}(t) \geqslant 0, \quad \sum_{j=1}^{n} \mathrm{x}_{ij}(t) \equiv L_i, \quad \sum_{i=1}^{n} \mathrm{x}_{ij}(t) \equiv W_j, \quad i, j = 1, \ldots, n. \qquad \text{(A)}$$

Пусть в момент времени $t \geqslant 0$ $r$-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 $p^L_{k,m;p,q}(t)\Delta t + o(\Delta t)$ есть вероятность того, что жители с номерами $r$ и $l$ ($1 \leqslant r < l \leqslant N$) 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 $t_{ij} \geqslant 0$ (вместо $t_{ij}$ в приводимую ниже формулу также осмысленно подставлять $l_{ij} \geqslant 0$ — расстояние от района $i$ до района $j$), а

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

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

$$\forall \{x_{ij}\}_{i,j=1}^n \in (\mathrm{A}) \quad \lim_{t\to\infty} P(\mathrm{x}_{ij}(t) = x_{ij},\ i, j = 1, \ldots, n) = \\ = \frac{1}{Z} \prod_{i,j=1}^{n} \exp(-\gamma t_{ij} x_{ij}) \cdot (x_{ij}!)^{-1} \stackrel{\text{def}}{=} p\big(\{x_{ij}\}_{i,j=1}^n\big),$$

[1)] Поскольку мы будем следить за системой на больших временах, то сделанное предположение кажется неестественным. Заметим, однако, что если «номер жителя» передается его потомкам (номер папы передается сыну, номер мамы — дочери), то предположение о постоянстве состава жителей выглядит разумным в первом приближении. Здесь мы также пренебрегаем миграционными потоками (город изолирован).



где статсумма $Z$ находится из условия нормировки получившейся пуассоновской вероятностной меры [52]. Распределение $p\big(\{x_{ij}\}_{i,j=1}^n\big)$ на множестве (A) сконцентрировано при $N \gg 1$ (см. ниже, с. 257) в окрестности наиболее вероятного значения $\{x^*_{ij}\}_{i,j=1}^n$, которое находится как решение задачи энтропийно-линейного программирования:

$$\sum_{i,j=1}^{n} x_{ij} \ln \frac{x_{ij}}{e} + \gamma \sum_{i,j=1}^{n} t_{ij} x_{ij} \to \min_{\{x_{ij}\}_{i,j=1}^n \in (\mathrm{A})}. \qquad (1)$$

**Замечание 3.** Нетрудно получить, воспользовавшись формулой Стирлинга, что задача поиска наиболее вероятного состояния асимптотически (по $n$) эквивалентна задаче максимизации энтропийного функционала:

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

на множестве (A). Поскольку функционал строго вогнутый и решение задачи максимизации, без предположения целочисленности $\{x_{ij}\}_{i,j=1}^n$, считаем таким, что $x^*_{ij} \gg 1$ (так как $n \gg 1$), то ограничение «целочисленности» является для данной задачи несущественным, и применение асимптотической формулы Стирлинга было законным. Обратим внимание также на то, что задача максимизации энтропийного функционала на множестве (A), т. е. по *принципу Лагранжа* [53] (балансовые ограничения (A) перенесли в функционал) задача

$$L\big(\{x_{ij}\}_{i,j=1}^n; \vec{\lambda}^L, \vec{\lambda}^W\big) = -\sum_{i,j=1}^{n} x_{ij} \ln \frac{x_{ij}}{e} - \gamma \sum_{i,j=1}^{n} t_{ij} x_{ij} + \\ + \sum_{i=1}^{n} \lambda^L_i \cdot \left(\sum_{j=1}^{n} x_{ij} - L_i\right) + \sum_{j=1}^{n} \lambda^W_j \cdot \left(\sum_{i=1}^{n} x_{ij} - W_j\right) \to \sup_{x_{ij} \geqslant 0, i,j=1,\ldots,n}$$

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

$$\frac{\partial L\big(\{x_{ij}\}_{i,j=1}^n; \vec{\lambda}^L, \vec{\lambda}^W\big)}{\partial x_{ij}} = -\ln x_{ij} + \lambda^L_i + \lambda^W_j - \gamma t_{ij} = 0, \quad i, j = 1, \ldots, n.$$

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 от *двойственных переменных*. Поскольку количество ограничений (количество двойственных переменных — *множителей Лагранжа*), как правило, на много порядков меньше числа прямых переменных, то эффективные численные методы отыскания решений базируются на решении *двойственной задачи*, представляющей собой задачу минимизации выпуклой функции на неотрицательном ортанте [30, 45, 54]. Можно показать, см. [54], что многие популярные в литературе [30, 45] численные методы решения этой двойственной задачи являются барьерно-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 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 $t_{ij} = \tau > 0\ \forall i, j = 1, \ldots, n$ все это можно сделать аналитически и в результате получить формулу (2). Отметим здесь также эффективность сепарабельных алгоритмов типа Нестерова—Немировского для задач энтропийного программирования, возникающих при нахождении равновесий макросистем [55]; в таких алгоритмах функционал декомпозируется в аддитивную сумму функций одного аргумента.

Решение задачи (1) можно представить как

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

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

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

Такой способ расчета матрицы корреспонденций в литературе часто называют *гравитационной моделью*:

$$x_{ij} = A_i B_j L_i W_j \exp(-\gamma t_{ij}),$$

где $\{A_i\}_{i=1}^n$ и $\{B_j\}_{j=1}^n$ определяются из соотношений [2, 30, 32]:

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

Описанная модель (точнее говоря, рассчитанная по этой модели матрица корреспонденций) активно используется в теоретико-игровых *моделях* (например, базирующихся на *принципах Дж. Г. Вардропа*) *равновесного распределения потоков* [32] (см. также [56]). Подробнее об этих моделях речь идет в главе 1. Один из возможных способов определения времени в пути, в зависимости от загрузки ребра, предложен в приложении М. Л. Бланка. Заметим также, что задача (1) по своим свойствам очень похожа на транспортную задачу (см. приложение А. В. Колесникова).

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



Если [1)]

$$N \sim nm, \quad L_i,\ W_j \sim m, \quad m \gg 1, \quad t_{ij} \equiv \tau > 0 \quad \forall i, j = 1, \ldots, n,$$

то распределение вероятностей $p\left(\{x_{ij}\}_{i,j=1}^n\right)$ на множестве (A) *сконцентрировано* в $O(\sqrt{m})$ окрестности (почему именно в такой окрестности, будет показано ниже) наиболее вероятного значения

$$x_{ij}^* \approx \frac{L_i W_j}{N} \sim \frac{m}{n}, \quad i, j = 1, \ldots, n, \tag{2}$$

которое ищется с помощью *метода множителей Лагранжа* [2, 53]. Формулу (2) можно интерпретировать как наличие у районов «потенциалов притяжения» [2]:

$$\frac{L_i}{\sqrt{N}} \sim \exp(\lambda_i^L), \quad i = 1, \ldots, n, \quad \text{и} \quad \frac{W_j}{\sqrt{N}} \sim \exp(\lambda_j^W), \quad j = 1, \ldots, n,$$

произведение которых дает пассажиропоток $x_{ij}^*$, $i, j = 1, \ldots, n$ (асимптотически совпадающий с математическим ожиданием и медианой).

Исследуем теперь, следуя [1,2,30], явление концентрации стационарного распределения $p\left(\{x_{ij}\}_{i,j=1}^n\right)$ в окрестности наиболее вероятного значения $\{x_{ij}^*\}_{i,j=1}^n$. Для этого прежде всего заметим, что из определения $\{x_{ij}^*\}_{i,j=1}^n$ (см. также замечание 3) следует

$$\sum_{i,j=1}^n \frac{\partial \ln p\left(\{x_{ij}^*\}_{i,j=1}^n\right)}{\partial x_{ij}} \cdot (x_{ij} - x_{ij}^*) \leqslant 0 \quad \forall \{x_{ij}^*\}_{i,j=1}^n \in \text{(A)}.$$

Поэтому

$$\forall \{x_{ij}\}_{i,j=1}^n \in \text{(A)}\ \exists \theta \in [0,1]:$$

$$\ln p\left(\{x_{ij}\}_{i,j=1}^n\right) \leqslant \ln p\left(\{x_{ij}^*\}_{i,j=1}^n\right) + \sum_{i,j=1}^n \frac{\partial^2 \ln p\left(\{x_{ij}^*\theta + x_{ij}\cdot(1-\theta)\}_{i,j=1}^n\right)}{\partial x_{ij}^2} \cdot \frac{(x_{ij} - x_{ij}^*)^2}{2}.$$

Но

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

[1)] Отметим, что хотя в этом случае динамика рассматриваемой нами макросистемы обратима по времени (так же, как и в примере 1), макросистема (в каком бы состоянии она ни 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]).

Следовательно, приходим к «неравенству о концентрации меры»: для любых $M > 0$ и любых $\{x_{ij}\}_{i,j=1}^n \in (\mathrm{A})$, для которых

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

выполняется неравенство:

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

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

$$\exists \lambda > 0\colon \lim_{t\to\infty} P\left(\left|\frac{\mathrm{x}_{ij}(t)}{x_{ij}^*} - 1\right| \leqslant \frac{\lambda}{\sqrt{m}};\ \ i,\, j = 1,\, \ldots,\, n\right) \geqslant 0{,}99.$$

**Упражнение**** (М. С. Ишманов, 2010). Верно ли, что при фиксированном $n$ и $m \to \infty$ скорость сходимости в этом соотношении оценивается как $O(m)$?

**Замечание 4** (о других возможных подходах к исследованию концентрации стационарного распределения). Один из способов восходит к методу Дарвина—Фаулера вычисления моментов [2,5,9] (метод производящих функций и анализ их асимптотического поведения методом перевала) — в этом случае концентрация наблюдается в окрестности математического ожидания; интересные приложения этого метода в комбинаторике имеются, например, в [57], см. также приложение А. А. Замятина и В. А. Малышева в этой книге и конец следующего раздела. Исследование концентрации в окрестности математического ожидания можно также проводить, например, используя предельные меры [58], метод канонических ансамблей [59] или обобщенную схему размещения [60], нашедшие применения в задачах асимптотической перечислительной комбинаторики,[1)] в исследовании случайных матриц и уравнений, в изучении статистических свойств группы перестановок с приложениями к теории разбиений (диаграммам Юнга) и асимптотической теории чисел, а также в теории предельных форм. К методу производящих функций, кроме того, тесно примыкают метод моментов [60], метод пуассоновской и гауссовской аппроксимации (метод локальной предельной теоремы) [7, 60]. Другой способ восходит к *принципу концентрации* А. Пуанкаре и П. Леви, получившему дальнейшее развитие в работах В. Д. Мильмана и др. [61] — в этом случае концентрация наблюдается в окрестности медианы[2)]. В заключение краткого обзора методов исследования концентрации меры упомянем теоремы тауберова типа [63] и мартингальные неравенства [62]. Из всего вышесказанного видно, что поиск наиболее вероятного распределения — это особенность работ, в которых изучаются равновесия макросистем. Наиболее вероятное распределение в содержательных задачах асимптотически (по размеру системы) эквивалентно медиане [61] и математическому ожиданию — в работе [2] соответствующие выкладки проделаны на примере модели расчета матрицы корреспонденций А. Дж. Вильсона, см. также следующий раздел.

[1)] В частности, в теории случайных графов (компьютерные, транспортные сети — вопросы надежности и т. п. [24,39], см. также приложение А. М. Райгородского в этой книге).

[2)] Этот принцип также нашел широкие применения в асимптотической перечислительной комбинаторике; в качестве достаточно известного примера можно упомянуть неравенство М. Талаграна и его приложения к изучению макросвойств (связность и т. п.) случайных графов [62] (см. также приложения А. М. Райгородского и А. В. Колесникова в этой книге).



**Замечание 5.** Интересно, что описанный способ изучения равновесных состояний макросистем применим к достаточно широкому классу макросистем — модель Д. Бернулли—Лапласа, модель П. и Т. Эренфестов, круговая модель М. Каца и др. (см., например, [8,27,64] и цитированную там литературу), в том числе встречающихся в экономике, биологии, социальной сфере [3,4,15–24].

## 2. Общая схема исследования равновесий макросистем

Ниже приводится (во многом под влиянием работ [18,19,65–67]) 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 $i$ рублей в момент времени $t \geqslant 0$). Будем считать, что в системе происходят случайные превращения (химические реакции). Пусть $\vec{n} \to \vec{n} - \vec{\alpha} + \vec{\beta}$, $(\vec{\alpha}, \vec{\beta}) \in J$ — все возможные типы реакций. Введем, следуя М. А. Леонтовичу (1935) [19], *интенсивность реакции* (случай дискретного времени рассматривается аналогичным образом):

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

где $K_{\vec{\beta}}^{\vec{\alpha}} \geqslant 0$ — *константы реакции* (в химической кинетике — постоянные, а в социодинамике — необязательно, см. В. Вайдлих [20]); при этом часто считают $\sum\limits_i n_i(t) \equiv M$; в частности, в примерах 1, 2 $M = N$. Т. е. $\lambda_{(\vec{\alpha},\vec{\beta})}(\vec{n})$ — вероятность осуществления в единицу времени перехода $\vec{n} \to \vec{n} - \vec{\alpha} + \vec{\beta}$: в единицу времени равновероятно выбираются любые два жителя города («приближение среднего поля») и в зависимости от того, в каких состояниях они находились, «случайно» переводятся в новые состояния (разыгрывают один рубль). На макроуровне все это соответствует принципам химической кинетики (*закон действующих масс Гульдберга—Вааге*, 1864 [18]). Таким образом, динамика макросистемы задается линейной полугруппой, представляющей однородный дискретный марковский случайный процесс. Инфинитезимальный оператор полугруппы определяется интенсивностями реакций $\lambda_{(\vec{\alpha},\vec{\beta})}(\vec{n})$.

Ниже приведены известные результаты а) В. В. Веденяпина [18], б) С. А. Пирогова и др. [19, 65] и в) результаты В. Вайдлиха и др. [20], обощенные на случай, когда рассматривается более общая схема, чем модель миграции населения:

а) $\langle\vec{\mu},\,\vec{n}(t)\rangle \equiv \langle\vec{\mu},\,\vec{n}(0)\rangle$. (inv)

тогда и только тогда, когда вектор $\vec{\mu}$ ортогонален каждому вектору семейства $\{\vec{\alpha}-\vec{\beta}\}_{(\vec{\alpha},\vec{\beta})\in J}$.

б) Пусть выполняется *условие унитарности*, которое, следуя В. В. Веденяпину, будем называть *условием Штюкельберга—Батищевой—Пирогова* $\left(K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}) \equiv K_{\vec{\beta}}^{\vec{\alpha}}\right)$:

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

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

$$E(\vec{n}) \approx -\sum_i n_i\left(\ln\frac{n_i}{\lambda_i}-1\right)$$

на множестве, задаваемом условием (inv). Отметим, что условие (ШБП), называемое также *условием унитарности* [19], обобщает хорошо известное в физике и экономике *условие детального равновесия* [20, 38]:

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

в) Пусть

$$\forall(\vec{\alpha},\,\vec{\beta})\in J\quad \sum_i \alpha_i = \sum_i \beta_i;\quad \alpha_i,\,\beta_i\in\{0,\,1\},\quad \mathrm{K}_{\vec{\beta}}^{\vec{\alpha}} = \mathrm{K}_{\vec{\alpha}}^{\vec{\beta}} \geqslant 0,$$

$$K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}) = \mathrm{K}_{\vec{\beta}}^{\vec{\alpha}}\exp\left(\sum_{i\colon\,\beta_i=1} u_i(n_i+1) - \sum_{i\colon\,\alpha_i=1} u_i(n_i)\right),\quad u_i'(n_i)\leqslant 0.$$



Тогда мера $\nu(\vec{n}) = \exp\left(\sum_i U_i(n_i)\right)\cdot\prod_i (n_i!)^{-1}$, где $U_i(n_i) = 2\sum_{\nu=1}^{n_i} u_i(\nu)$, будет инвариантной относительно предложенной стохастической марковской динамики. Эта мера экспоненциально быстро концентрируется, с ростом $M\equiv\sum_i n_i(t)$, в окрестности наиболее вероятного состояния, которое и принимается за положение равновесия макросистемы. Задача поиска наиболее вероятного макросостояния асимптотически эквивалентна задаче максимизации функционала энтропийного типа:

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

на множестве, задаваемом условием (inv).

**Замечание 6.** В пунктах б) и в) предполагалось, что марковский процесс неразложим (неприводим) в классе (inv): из любого состояния можно со временем прийти в любое другое (по-прежнему оставаясь на множестве (inv)). Отсюда следует единственность инвариантной меры. Это условие не выполняется, например, для хорошо известной модели «хищник–жертва» (кролики–трава) [3, 15, 17, 22], в которой имеется поглощающее состояние: без хищников. Нам еще будет встречаться ниже, и не один раз, эта модель.

**Замечание 7** (к пункту б)). Будем считать, что ограничения (законы сохранения) (inv) задаются СЛАУ $A\vec{n} = \vec{d}$, где $A = \|A_{kl}\|$ — матрица максимального ранга ($k = 1,\,\ldots,\,m$). Обозначим через $A$ множество неотрицательных целочисленных векторов $\vec{n}$, удовлетворяющих системе уравнений $A\vec{n} = \vec{d}$. Тогда равновесие $\vec{n}^*$ находится как решение задачи $E(\vec{n})\to\max\limits_{\vec{n}\in A}$ (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 $A\vec{n}(\vec{y}) = \vec{d}$.

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

Для этого введем *производящую функцию*

$$F(\vec{z};\,A) = \sum_{\vec{n}\geqslant\vec{0}}\mu(\vec{n})\prod_k z_k^{\sum_l A_{kl}n_l} = \prod_i\sum_{n_i\geqslant 0}\frac{1}{n_i!}\left(\lambda_i\prod_k z_k^{A_{ki}}\right)^{n_i}e^{-\lambda_i} =$$

$$= \prod_i\exp\left(\lambda_i\cdot\left(\left(\prod_k z_k^{A_{ki}}\right)-1\right)\right).$$

Тогда по формуле Коши:

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

$$Z = \frac{1}{(2\pi i)^m}\oint dz_1\ldots dz_m\left[\left\{\prod_k z_k^{-d_k-1}\right\}F(\vec{z};\,A)\right].$$

Здесь математическое ожидание $E\left[n_{r_1}^{p_1}\cdot\ldots\cdot n_{r_Q}^{p_Q}\right]$ считается по вероятностной мере, порожденной мерой Пуассона $\mu(\vec{n})$, а интегралы по $dz_k$ берутся в комплексной плоскости по замкнутым контурам, охватывающим точку ноль. Используя метод перевала [68], асимптотически оценим математическое ожидание:

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

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

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

В частности [1)],

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

где $\vec{z}^*$ определяется как решение системы уравнений:

$$\sum_i A_{ki}\left\{\lambda_i\prod_k z_k^{A_{ki}}\right\}=d_k,\quad k=1,\,\ldots,\,m.$$

Очевидна связь «точки перевала» $\vec{z}^*$ с двойственными переменными $\vec{y}^*$: $z_k^*=\exp(y_k^*)$. □

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

$$A+B^*\xrightarrow{K}A^*+B;\quad B+C^*\xrightarrow{K}B^*+C;\quad C+A^*\xrightarrow{K}C^*+A,$$

причем

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

[1)] Обратим внимание на то, что получилось $E[n_i]\approx D[n_i]\gg 1$ — это означает концентрацию распределения случайной величины $n_i$ в $\sqrt{D[n_i]}$-окрестности своего математического ожидания $E[n_i]$.



Заметим, что есть и еще один независимый закон сохранения:

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

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

$$\nu(\vec{n})\sim C_N^{n_A}\cdot C_N^{n_B}\cdot C_N^{n_C}\sim\underbrace{\frac{1^{n_A}e^{-1}}{n_A!}\cdot\ldots\cdot\frac{1^{n_{C^*}}e^{-1}}{n_{C^*}!}}_{\text{6 множителей}}$$

будет инвариантной, хотя условие (ШБП), очевидным образом, не выполняется.

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

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

Вернемся к примеру 1 «Кинетика социального неравенства», для которого (так же как и для примера 2) выполняется условие детального равновесия, следовательно, существует и единственно равновесие. Этот пример демонстрирует ситуацию, когда число состояний ($\dim\vec{n}$) и число реакций ($|J|$) растут вместе с ростом $M$. Это обстоятельство, равно как и зависимость $K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n})$, не позволяют напрямую использовать аппарат, разработанный в [18, 19, 65, 66], связанный с анализом СОДУ, возникающей при каноническом скейлинге ($M\to\infty$ так, что $\exists\lim\limits_{M\to\infty}\vec{n}(0)/M=\vec{c}>\vec{0}$) стохастической марковской динамики. Тем не менее, результаты этого примера можно перенести и на общий случай. При этом ключевым местом является существование при термодинамическом предельном переходе $M\to\infty$, $|J|\to\infty$ ненулевого финального распределения [69, 70].

Предположим теперь, что множество $J$ не зависит от $M$, и в начальный момент времени для любого $i$ существует предел $c_i(0)=\lim\limits_{M\to\infty}n_i(0)/M$, $K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}):=K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}/M)$. Тогда из результатов Т. Г. Куртца (см., например, [65]) следует, что в произвольный момент времени $t>0$ и для любого $i$ существует предел по вероятности

$$c_i(t)\overset{\text{п.н.}}{=}\lim_{M\to\infty}\frac{n_i(t)}{M};$$

заметим, что $n_i(t)$ — случайные величины, тем не менее $c_i(t)$ — уже не случайные величины. Описанный выше прием называется *каноническим скейлингом*. В результате такого скейлинга приходим к «динамике квази-

средних» (терминология В. Вайдлиха [20]):

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

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

Покажем, во многом следуя Батищевой—Веденяпину [18], что если выполняются условия (ШБП), то траектория (ДК) сходится к неподвижной точке (какой именно, зависит, вообще говоря, от «точки старта»; но можно сказать и точнее: к той единственной неподвижной точке из семейства неподвижных точек, которая принадлежит аффинному многообразию (inv), инвариантному относительно (ДК)) [2)]. Для этого, следуя второму методу Ляпунова, введем (минус) *энтропию*: $H = \sum_i c_i \cdot (\ln(c_i/\xi_i) - 1)$ и покажем, что она является функцией Ляпунова для системы (ДК). Обратим внимание, что инвариантная мера (при каноническом скейлинге) «породила» функцию Ляпунова (см. п. б)). Это не случайно. Подобные закономерности наблюдаются для рассматриваемых моделей и без условия (ШБП), и даже без предположения о том, что инвариантная мера концентрируется около единственного положения равновесия, то есть аттрактором может быть множество гораздо более сложной структуры. В частности, если инвариантная мера представляется в виде: $\nu(\vec{n}) = M \exp(-M \cdot (H(\vec{n}/M) + o(1)))$, $M \gg 1$, где $H(\vec{c})$ — строго вогнутая функция, то $H(\vec{c})$ — функция Ляпунова системы (ДК) [71].

Посчитаем полную производную $H$ в силу системы (ДК):

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

[1)] Заметим, что этот переход и возможность его использования нуждаются в строгом обосновании (и далеко не всегда правомочны). В качестве примера, укажем популярный в литературе [17,22] марковский процесс «рождения—гибели» (приводящий к системе уравнений «хищник—жертва»), для которого «флуктуации играют решающую роль, качественно меняя выводы макроскопического анализа».

[2)] Стоит заметить, что аттрактор системы (ДК) даже с постоянными коэффициентами реакции, по-видимому, в общем случае может быть сколь угодно сложным множеством [20].



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

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

Таким образом,

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

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

Естественно (ввиду примера С. А. Пирогова) теперь задаться вопросом: что будет, если условия (ШБП) не выполняются, однако система (ДК) имеет на внутренности пересечения неотрицательного ортанта и инвариантного аффинного многообразия (inv) единственную неподвижную точку? Оказывается, имеет место

**Утверждение.** *Если эта точка экспоненциально глобально устойчива и* $\sum_i n_i(t) \equiv M$, *то* 1) *все законы сохранения* (ДК) *определяются* (inv); 2) *около положения равновесия инвариантная мера будет экспоненциально быстро концентрироваться* (*с ростом* $M$); 3) *скорость сходимости к равновесию* (*mixing time* [72]) *оценивается как* $O(\mathrm{poly}(M))$; 4) *элементы корреляционной матрицы случайного вектора* $\vec{n}(t)$ *равномерно ограничены по времени*; 5) *предельные переходы* $\lim\limits_{M\to\infty}$ *и* $\lim\limits_{t\to\infty}$ *перестановочны*: $\lim\limits_{M\to\infty} \lim\limits_{t\to\infty} * = \lim\limits_{t\to\infty} \lim\limits_{M\to\infty} *$.

Обратим внимание, что модель «хищник—жертва», в изложении п. 7.3 книги [22], является хорошим примером того, что может быть, если не выполняется условие устойчивости равновесия.

Результаты [17, 20] и ряда других работ наталкивают на **гипотезу**: аттрактор динамической системы (ДК), который, как мы отмечали выше, может быть сколь угодно сложным множеством (например, в приложениях типичны случаи предельных циклов, нескольких положений равновесий и даже хаотических аттракторов), является таким множеством, в малой окрестности которого на больших временах с большой вероятностью будет пребывать рассматриваемая макросистема. Эту гипотезу удалось обосновать во многих наиболее интересных на практике случаях, см. [74]. Также в [74] при помощи неравенства Чигера показано, что если аттрактором является единственное положение равновесия, около которого инвариантная мера экспоненциально быстро концентрируется, то время выхода макросистемы на это равновесие оценивается как $O(M \cdot \ln M)$.

**Упражнение** (модель Эренфестов [8,38,64]). Рядом стоят две собаки с номерами 1 и 2. На собаках как-то расположились $M = 2n \gg 1$ блох.

Скажем, в начальный момент все блохи собрались на собаке с номером 1. На каждом шаге случайно и независимо от предыстории определяется блоха (с вероятностью $1/M$ будет выбрана любая из блох), которая перепрыгивает на другую собаку. Микросостояние системы есть способ распределения $M$ *различных* блох по двум различным собакам. Макросостояние системы есть способ распределения $M$ *одинаковых* блох между двумя различными собаками. Микросостояний будет $2^M$, а макросостояний $M+1$.

Обозначим через $P$ матрицу (размера $2^M \times 2^M$) переходных вероятностей описанной выше микроскопической динамики. Для нас в дальнейшем будет важно лишь одно свойство этой стохастической матрицы: $P = P^T$, которое следует из обратимости динамики во времени. Но поскольку $P$ — стохастическая матрица, то

$$(1, \ldots, 1) = (1, \ldots, 1)P^T \implies (1, \ldots, 1) = (1, \ldots, 1)P.$$

Отсюда с учетом нормировки распределения вероятностей на 1 имеем, что в стационарном распределении все микросостояния равновероятны, т. е. в стационарном распределении каждому микросостоянию приписана вероятность $2^{-M}$. Но тогда вероятность макросостояния $(k, M-k)$ в стационарном распределении равна $C_M^k 2^{-M}$. Покажите, что

$$\lim_{m\to\infty} P\Big(\frac{|n_1(m) - n_2(m)|}{M} \leqslant \frac{3}{\sqrt{M}}\Big) \geqslant 0{,}99,$$

где $n_1(m)$ — число блох на первой собаке на шаге $m$, а $n_2(m)$ — на второй (случайные величины). Т. е. относительная разность числа блох на собаках будет иметь порядок малости $O(1/\sqrt{M})$ на больших временах ($T \geqslant 2M$ [73]). Обратим внимание на то, что марковская цепь — периодическая с периодом 2. Однако поскольку речь идет о вычислении вероятностей относительных величин, то в данной задаче это не играет роли.

Обозначим через

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

времена соответственно первого попадания и первого возвращения в состояние $k$. Покажите, что

а) $E\sigma(k) = 2^M \dfrac{k!(M-k)!}{M!}$, и, в частности, среднее время возвращения в нулевое состояние $E\sigma(0) = 2^M$, где $E\sigma(k)$ — математическое ожидание времени первого возвращения в состояние $k$, если $n_1(0) = k$, $k = 0, \ldots, n$;

б) $E_n\tau(0) = \dfrac{1}{M}2^M(1 + o(M))$, где $E_n\tau(0)$ — математическое ожидание времени первого попадания в состояние 0, если $n_1(0) = n$;



в) $E_0\tau(n) = n\ln n + n + O(1)$, где $E_0\tau(n)$ — математическое ожидание времени первого попадания в состояние $n$, если $n_1(0) = 0$.

На примере этой модели можно говорить о том, что в макросистемах возврат к неравновесным макросостояниям вполне допустим, но происходить это может только через очень большое время (*циклы Пуанкаре*), так что нам может не хватить отведенного времени, чтобы это заметить (*парадокс Цермело*). Напомним, что описанный выше случайный процесс обратим во времени. Однако наблюдается необратимая динамика относительной разности числа блох на собаках (*парадокс Лошмидта*).

## Заключение

В приложении обсуждается концепция равновесия макросистемы. Приводятся различные подходы к обоснованию следующего принципа: *равновесие — наиболее вероятное макросостояние инвариантной* (*стационарной*) *меры динамической системы* (*марковского процесса*), *порождающей исследуемую макросистему*. Рассматриваются примеры конкретных макросистем. В частности, один из примеров «объясняет» популярную в приложениях модель А. Дж. Вильсона расчета матрицы корреспонденций.

Повторим в заключение описанную в приложении схему.

1. Макросистема состоит из огромного числа пронумерованных агентов, каждый из которых может находиться в одном из возможных состояний. Число состояний как минимум на несколько порядков меньше числа агентов (иногда можно обойтись 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.

Предполагается, что из любого возможного макросостояния можно перейти согласно такой динамике в любое другое (характерное время такого перехода определяет скорость сходимости к равновесию). Также считается, что описанная динамика имеет макрозаконы сохранения — соотношения (как правило, линейные) между 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 на пространстве макросостояний (прямое произведение распределений Пуассона). 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, как правило, аффинной структуры, заданном ограничениями — законами сохранения. Этот же энтропийный функционал, взятый со знаком минус, возникает как функция Ляпунова динамики, полученной в результате канонического скейлинга исходной марковской динамики. Отыскание предельной неподвижной точки (этой динамики), в которую придет система, сводится к решению той же самой задачи энтропийно-линейного программирования.

### Литература


1. *Jaynes E. T.* 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. *Шрёдингер 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. *Свирежев 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.* Кинетические 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/0001432v4`
24. *Baldi P.*, *Frasconi P.*, *Smyth P.* Modeling the Internet and the Web: Probabilistic methods and algorithms. 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. *Малишевский А. В.* Качественные модели в теории сложных систем. М.: Наука, 1998.
29. *Сергеев В. М.* Пределы рациональности. М.: Фазис, 1999.
30. *Попков Ю. С.* Теория макросистем: равновесные модели. М.: УРСС, 1999.
31. *Цирлин А. М.* Методы оптимизации в необратимой термодинамике и микроэкономике. М.: Физматлит, 2003.
32. *Швецов В. И.*, *Алиев А. С.* Математическое моделирование загрузки транспортных сетей. М.: УРСС, 2003.
33. *Маслов В. П.* 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. М.: МЦНМО, 2010.
39. *Вишневский 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. *Маслов В. П., Черный А. С.* О минимизации и максимизации энтропии в различных дисциплинах // ТВП. 2003. Т. 48, № 3. С. 466–486.
47. *Богданов К. Ю.* Прогулки с физикой. Библиотечка «Квант». Вып. 98. М.: Бюро Квантум, 2006 (глава 18).
48. *Зорич В. А.* Математический анализ задач естествознания. М.: МЦНМО, 2008.
49. *Diaconis P.* The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. 2009. V. 49, № 2. P. 179–205; `http://www.ams.org/journals/bull/2009-46-02/S0273-0979-08-01238-X/S0273-0979-08-01238-X.pdf`
50. *Joulin A., Ollivier Y.* Curvature, concentration and error estimates for Markov chain Monte Carlo // Ann. Prob. 2010. V. 38, № 6. P. 2418–2442; `http://www.yann-ollivier.org/rech/publs/surveycurvmarkov.pdf`
51. *Красносельский М. А., Лифшиц Е. А., Соболев А. В.* Позитивные линейные системы. Метод положительных операторов. М.: Наука, 1985.
52. *Кингман Дж.* Пуассоновские процессы / Под ред. А. М. Вершика. М.: МЦНМО, 2007.
53. *Магарил-Ильяев Г. Г., Тихомиров В. М.* Выпуклый анализ и его приложения. М.: УРСС, 2003.
54. *Гасникова Е. В.* Двойственные мультипликативные алгоритмы для задач энтропийно-линейного программирования // ЖВМ и МФ. 2009. Т. 49, № 3. С. 453–464.
55. *Нестеров Ю. Е.* Введение в выпуклую оптимизацию. М.: МЦНМО, 2010.
56. *Нурминский Е. А., Шамрай Н. Б.* Прогнозное моделирование автомобильного трафика Владивостока // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 119–129.
57. *Flajolet P., Sedgewick R.* Analytic combinatorics. Cambridge University Press, 2008; `http://algo.inria.fr/flajolet/Publications/book.pdf`
58. *Вершик А. М., Шмидт А. А.* Предельные меры, возникающие в асимптотической теории симметрических групп // ТВП. 1977. Т. 22, № 1. С. 72–88; 1978. Т. 23, № 1. С. 42–54.
59. *Синай Я. Г.* Вероятностный подход к анализу статистики выпуклых ломаных // Функц. анализ и его прил. 1994. Т. 28, № 2. С. 41–48.
60. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2004.



61. *Ledoux M.* Concentration of measure phenomenon. Providence, RI: AMS, 2001. (Math. Surveys Monogr. V. 89).
62. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.
63. *Якымив А. Л.* Вероятностные приложения тауберовых теорем. М.: Наука, 2005.
64. *Ширяев А. Н.* Вероятность 1, 2. М.: МЦНМО, 2007.
65. *Malyshev V. A., Pirogov S. A., Rybko A. N.* Random walks and chemical networks // Mosc. Math. J. 2004. V. 4, № 2. P. 441–453.
66. *Батищева Я. Г., Веденяпин В. В.* II-й закон термодинамики для химической кинетики // Мат. мод. 2005. Т. 17, № 8. С. 106–110.
67. *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.
68. *Федорюк М. В.* Метод перевала. М.: УРСС, 2010.
69. *Rybko A., Shlosman S.* Poisson hypothesis for information networks I, II. 2004; `arXiv:math-ph/0303010`, `arXiv:math-ph/0406110`, `arXiv:math-ph/0410053`.
70. *Рыбко А. Н.* Пуассоновская гипотеза для больших симметричных коммуникационных сетей // Глобус. Общематематический семинар. Вып. 4 / Под ред. М. А. Цфасмана и В. В. Прасолова. М.: МЦНМО, 2009. С. 105–126.
71. *Гасников А. В., Гасникова Е. В.* Об энтропийно-подобных функционалах, возникающих в стохастической химической кинетике при концентрации инвариантной меры и в качестве функций Ляпунова динамики квазисредних // Математические заметки, 2013 (в печати).
72. *Montenegro R., Tetali P.* Mathematical aspects of mixing times in Markov chains. 2006; `http://people.math.gatech.edu/~tetali/PUBLIS/survey.pdf`
73. *Jerrum M., Sinclair A.* The Markov chain Monte Carlo method: an approach to approximate counting and integration // Approximation Algorithms for NP-hard Problems / D.S.Hochbaum ed. Boston: PWS Publishing, 1996. P. 482–520.
74. *Гасникова Е. В.* Моделирование динамики макросистем на основе концепции равновесия. Дисс. ... канд. физ.-мат. наук. М.: МФТИ, 2012.

**А. А. Замятин, В. А. Малышев**

# Введение в стохастические модели транспортных потоков

## Введение

Математические модели автомобильного трафика могут быть весьма различны: от дифференциальных уравнений с частными производными, средств современной компьютерной физики до создания игровых моделей, где точки на видео движутся по сети улиц с перекрестками. Мы рассматриваем здесь некоторые строгие вероятностные подходы к транспортным сетям. Основная цель этого приложения — не столько представить технику решения задач, сколько представить методику (и искусство) составления адекватных моделей, которые отличаются наглядностью определений (основной объект там именно автомобиль, а не потоки) и основаны на простых интуитивных рассуждениях. Более того, все вводимые постулаты в этих моделях допускают статистическую проверку, широкие уточнения и обобщения и не используют сомнительных физических аналогий. Вообще, вероятностные модели должны связываться с психикой водителей, если водители не роботы. Такой законченной и общепринятой теории пока не существует, здесь делаются, по-видимому, первые строгие попытки установить такую связь.

В текст включены упражнения для лучшего усвоения материала, а также задачи посложнее, в том числе и «публикабельные».

Вероятностный подход к транспортным потокам существует уже более 50 лет, см. [1–3], однако здесь мы даем более современную трактовку и рассматриваем более сложные задачи. В то же время мы не говорим здесь о других вероятностных подходах к проблемам транспорта, например [12, 14], они отражены в других частях этой книги. Мы также ничего не говорим о гидродинамическом подходе, так как связь его со статистическим подходом пока математически плохо изучена.

Приложение состоит из трех частей. В первой дано построение случайных потоков и некоторые модели, отражающие качественные явления на автомагистрали, в том числе новая модель, основанная на сравнительно недавней теории случайных грамматик. Во второй части показано, как можно получать явные формулы с помощью техники пуассоновских потоков. В третьей рассмотрены сложные сети дорог и вычисление критической нагрузки, выше которой начинаются пробки.



## 1. Потоки автомобилей

### 1.1. 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

$$\ldots < x_i < x_{i-1} < \ldots$$

автомобилей в данный момент времени, но можно понимать его динамически как меру на множестве траекторий $\{x_i(t)\}$ автомобилей.

**Что такое конфигурация автомобилей.** Максимально детальное описание расположения автомобилей в данный момент времени таково. Автомобиль индивидуален и ему присваивается некий индекс $\alpha$. Например, пусть есть автотрасса с $k$ полосами $1, 2, \ldots, k$, представляемая $k$ прямыми, параллельными оси $x$. Тогда индекс $\alpha = (m, i)$ выделяет $i$-й автомобиль на полосе $m$. Индекс $i$ нумерует автомобиль на полосе, так что автомобиль $i$ следует за автомобилем $i-1$. Пусть $d_\alpha$ — длина этого автомобиля, $x_\alpha(t)$ — его координата (например, переднего бампера). Автомобили движутся в положительном направлении оси $x$. Далее индекс полосы опускается — читатель может его добавлять где надо — и используем только индекс $i$.

Обозначим расстояние автомобиля $i$ до предыдущего автомобиля в реальном потоке в момент $t$ через

$$d_i^+(t) = x_{i-1}(t) - x_i(t) - d_{i-1}.$$

Обозначим (тоже важная величина для водителя)

$$d_i^-(t) = d_{i+1}^+(t)$$

— расстояние до следующего автомобиля.

**Как вводятся вероятности на множестве конфигураций.** Формально точечный случайный поток на прямой задается вероятностной мерой на множестве всех счетных локально конечных (то есть конечных на каждом ограниченном интервале) подмножеств прямой. Иначе говоря, поток задается согласованной системой вероятностей

$$P(I_1, k_1; \ldots; I_n, k_n)$$

того, что в интервалах $I_j$, $j = 1, \ldots, n$, находится ровно $k_j$ частиц.

Для более конкретного задания этих распределений существует две большие науки: *теория восстановления* (см., например, [6]) и *теория гиббсовских точечных полей* [23, 24]. Первая теория существенно проще, но годится только в одномерном случае. Вторая глубоко связана с физикой, годится и для многомерных ситуаций, но довольно сложна, и мы не будем ее здесь касаться.

Самый простой случайный поток — пуассоновский, см., например, [30]. Простейший способ его понять такой. Рассмотрим интервал $[-N, N]$ и бросим на него независимо и случайно (точнее, равномерно) $M = [\rho N]$ точек, где $\rho > 0$ — некоторая константа, называемая плотностью. Легко вычислить биномиальную вероятность $P_{NM}(k, I)$ того, что в конечный интервал $I$ попадет ровно $k$ точек. Последняя при $N \to \infty$ стремится к пуассоновскому выражению

$$P(k, I) = \frac{\{\rho|I|\}^k}{k!} e^{-\rho|I|}.$$

Более общие потоки легко строятся на полупрямой $[0, \infty)$. Именно, случайные точки

$$x_0 = 0, x_1, \ldots, x_n, \ldots$$

определяются как суммы независимых одинаково распределенных случайных величин $\xi_i > 0$, $i = 1, 2, \ldots$, с распределением $G(x)$:

$$x_1 = \xi_1, \quad x_2 = \xi_1 + \xi_2, \quad \ldots$$

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 $\xi_i$ имеют плотность, то на полупрямой предел $\lim\limits_{\Delta t \to 0} \frac{1}{\Delta t} P(t, t + \Delta t)$ существует и стремится при $t \to \infty$ к $\rho = (E\xi)^{-1}$. Предельная (при $t \to \infty$) плотность вероятности того, что расстояние от точки $t$ до первой случайной точки $x_i > t$ больше $s$, равна произведению $\rho$ на вероятность того, что $\xi_i = x_{i+1} - x_i > s$, то есть равна

$$\rho \cdot (1 - G(s)).$$

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: $\xi_1, \xi_2, \ldots$ и $\eta_1, \eta_2, \ldots$ и положим

$$x_{2n} = \xi_1 + \ldots + \xi_n + \eta_1 + \ldots + \eta_n, \quad x_{2n-1} = \xi_1 + \ldots + \xi_n + \eta_1 + \ldots + \eta_{n-1}.$$

Тогда построение потока на всей прямой делается как и выше. В нашем случае чередуются длины автомобилей $d_i$ (независимые и одинаково распределенные) и функции $d_i^+$ (независимые и одинаково распределенные).

**Маркированные потоки.** Каждой точке $x_i$ точечного процесса может быть сопоставлена величина $\sigma_i$, принимающая значения в некотором множестве $S$. Эту величину в разных случаях называют маркой или спином в точке $x_i$ и говорят о случайном маркированном точечном множестве (потоке, процессе). Он определяется мерой на последовательностях пар $(x_i, \sigma_i)$. Проще всего, когда задана мера на счетных множествах, то есть задан поток без марок, а величины $\sigma_i$ объявляются независимыми и одинаково распределенными. В разделе 1.5 мы построим маркированный процесс, где марками являются скорости, причем их распределение будет сложным образом коррелировать во времени с траекториями точек.

**О марковских процессах.** Для описания (моделирования) эволюции во времени конфигураций автомобилей часто используются марковские процессы, и необходимо сказать о соответствующей терминологии. Часто определение самого процесса и его свойств (например, эргодичности) отличаются в разных источниках. Поясним это. Для простоты ограничимся случаем дискретного времени.

Рассмотрим на некотором фазовом пространстве $X$ систему мер (переходных вероятностей) $P(A|x)$, определяющих вероятности того, что в момент $t + 1$ процесс попадет в множество $A \subset X$, если в момент $t$ процесс находился в состоянии $x \in X$. Если все меры $P(A|x)$ одноточечные, то это эквивалентно заданию детерминированного отображения $T\colon X \to X$, точнее, $P(\cdot|x) = \delta(T(x))$ — единичная мера в точке $T(x)$. Тогда говорят о детерминированном отображении, задающем динамическую систему.

Заметим, что система мер $P(A|x)$ определяет очевидным образом преобразование $U$ множества вероятностных мер на $X$ в себя:

$$U\mu = \int P(\cdot|x)\, d\mu(x).$$

Важным является понятие инвариантной (относительно $U$) меры. Обычно исследуется ее существование, единственность и другие свойства.

По системе переходных вероятностей можно построить разные последовательности случайных величин $\xi_n$ со значениями в $X$ или их распределений $\mu_n$ на $X$, где $P(\xi_n \in A) = \mu_n(A)$. Вероятностным пространством

при этом служит множество траекторий $\{x_n\}$. Например, по вероятностной инвариантной мере строится стационарный марковский процесс как последовательность $\xi_n$, $n \in \mathbb{Z}_+$ случайных величин со значениями в $X$. Или по заданной начальной мере $\mu_0$ на $X$ строится последовательность $\xi_n$, $n \in \mathbb{Z}_+$.

Под марковским процессом может пониматься как одна из таких последовательностей случайных величин, так и все семейство таких последовательностей $\xi_n$. Соответственно разнится терминология, например, определение эргодичности. Динамическая система с заданной инвариантной мерой $\mu$ называется эргодической, если любое инвариантное множество имеет меру $\mu$ ноль или единицу. На любой стационарный марковский процесс можно смотреть как на динамическую систему — сдвиг в пространстве траекторий. Тогда понятие эргодичности совпадает с понятием эргодичности этой динамической системы.

Однако чаще, когда говорят о марковском процессе, имеют в виду не только стационарные процессы. Наиболее часто используемым определением эргодичности является следующее. Процесс называется эргодическим, если существует единственная инвариантная мера $\mu$ на $X$ и для любой начальной меры $\mu_0$ при $n \to \infty$ имеет место слабая сходимость $\mu_n \to \mu$.

Отметим, что марковские процессы резко разделяются на два класса. Первый класс — для которых существует положительная мера на $X$ (не обязательно вероятностная), относительно которой все меры $P(\cdot|x)$ абсолютно непрерывны. К ним относятся почти все классические марковские процессы — конечные и счетные цепи Маркова, диффузионные процессы и другие. Такие процессы называются эргодическими, если, во-первых, у преобразования нет нетривиальных инвариантных множеств, а во-вторых, существует единственная инвариантная вероятностная мера. Во многих случаях отсюда следует сходимость к этой инвариантной мере из любого начального распределения. Для счетных цепей эквивалентным условием является положительная возвратность, то есть конечность (для всех пар $x$, $y \in X$) среднего времени достижения $y$ из $x$.

Второй класс характеризуется тем, что все меры $P(\cdot|x)$ взаимно сингулярны. К этому классу относятся почти все процессы с бесконечным числом частиц. Теория таких процессов существенно сложнее.

### 1.2. 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 $v_{i-1}(t) = \dfrac{dx_{i-1}(t)}{dt}$ 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 $i$ говорит, что функции $D_i^+(v)$ разные для разных водителей). Назовем поток алгоритмическим в момент $t$, если для всех $i$

$$d_i^+(t) = D_i^+(v_{i-1}(t)),$$

то есть все скорости последовательно определяются по скорости первого автомобиля. Конечно, нас интересуют не сами функции, а их статистические характеристики. При вероятностном подходе функции $D_i^+(v)$ становятся независимыми одинаково распределенными случайными функциями, зависящими от скорости $v$ предыдущего водителя как от параметра. Распределение этих функций не может быть выведено из математических, статистических, физических и т. д. законов. Оно зависит от индивидуальной и коллективной психики водителя и должно находиться экспериментально, см. [22].



**Детерминированная динамика без обгона.** Если все автомобили, водители и скорости $v$ одинаковы, то многие задачи решаются просто. Обозначим через $d$ длину автомобиля, 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, а

$$\lambda(v) = \frac{k}{d + D^+(v)}$$

— плотность автомобилей на $k$-полосной дороге при заданной скорости $v$. Отсюда видно, что пропускная способность может уменьшаться при увеличении скорости. Этот простой вывод говорит лишь о том, что многие водители увеличивают расстояние до впереди идущего автомобиля при увеличении его скорости.

**Случайная динамика без обгона.** То же самое получится, если скорости $v$ одинаковы, функции $d_i^+$ случайны и независимы, а их средние равны (для заданного $v$) некоторому числу $d^+(v)$. Мы видим, что сам факт нетривиальной зависимости пропускной способности от скорости очевиден, и для него совершенно не нужны вероятностные модели. Однако для более тонких вопросов вероятностные модели необходимы. Сейчас мы введем довольно общую вероятностную модель с очень богатым спектром

фаз. При этом процессы с запретами (exclusion processes) появляются как вырожденный частный случай. Другие модели см. [10, 12, 22].

**Случайная динамика с обгоном (случайные грамматики).** Здесь естественно возникает связь с таким недавно открытым объектом, как случайные грамматики, см. [25]. Мы дадим краткое содержательное описание одной такой модели.

Пусть в момент $t = 0$ все автомобили находятся на левой полуоси, 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 $r \in \mathbb{Z}$, а $S$ — конечная последовательность (слово) из трех символов 0, 1, 2:

$$S = s_N \dots s_2 s_1.$$

При этом 0 соответствует пустой клетке, 1 — активному (быстрому) водителю в клетке, 2 — спокойному водителю в клетке. Длина слова $N = N(t)$ и все символы $s_k(t)$ могут меняться во времени, но так, что всегда $s_1(t) \neq \neq 0$ для всех $t \geqslant 0$. В произвольный момент $t$ каждый символ $s_k(t)$ имеет координату $x(s_k(t))$. Координаты однозначно определяются

$$x(s_k(t)) = x(s_1(t)) - k + 1 \tag{1}$$

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) $10 \to 01$ — быстрый водитель передвигается на одного вперед, освобождая за собой место, с вероятностью $\lambda_0^+ dt$ за время $dt$;

2) $120 \to 021$ — быстрый водитель обгоняет спокойного с вероятностью $\lambda_1^+ dt$;

3) $22 \to 202$, $21 \to 201$ — предусмотрительный водитель тормозит, увеличивая дистанцию перед собой, с вероятностью $\lambda_2^- dt$. Отметим, что здесь



увеличивается длина $S$ (возникает лишняя свободная ячейка), что ведет к сдвигу всех автомобилей сзади этого водителя на одного назад. Это нелокальный скачок, реально он растянут во времени, но это совместимо с правилом сложения относительных скоростей;

4) $200 \to 020$ — спокойный водитель ускоряется с вероятностью $\lambda_2^+ dt$ (если впереди, с его точки зрения, много свободного места).

Необходимо сказать, что для точной формулировки результатов, которые мы лишь обрисуем, надо делать разнообразные скейлинги параметров $t$, $N$, $\lambda$. 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 $\lambda_2^\pm$ малы по сравнению с остальными двумя параметрами, то автомобили типа 2 едут синхронно и с постоянной скоростью, а быстрые автомобили имеют дополнительную относительную скорость. Если быстрых автомобилей мало, то эта дополнительная скорость определяется движением одного автомобиля среди неподвижных препятствий и зависит от плотности $\rho_2$ автомобилей типа 2 и плотности дырок $\rho_0$ и примерно равна

$$v_{\mathrm{rel}} = \lambda_0^+ \rho_0 + 2\lambda_1^+ \rho_2.$$

Если $\lambda_2^-$ мала по сравнению с остальными двумя параметрами (нет нелокальных эффектов), а $\lambda_2^+$ имеет такой же порядок, как $\lambda_0^+$, $\lambda_1^+$, 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

$$\lambda_0^+ = \lambda_1^+, \quad \lambda_2^- = 0$$

— полностью с ним совпадающий (о TASEP см. приложение М. Бланка и ссылки в нем).

Если $\lambda_2^+$ мала, а $\lambda_2^-$ велика по сравнению с остальными двумя параметрами, то картина иная. Каждый обгон $120 \to 021$ 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.

Можно усложнять введенную динамику, например, избежать дискретизации (см. конец этого раздела), вводя вместо нулей положительные вещественные числа — расстояния между последовательными автомобилями. Это потребует существенных (см., однако, раздел 1.5) переформулировок, особенно для скачков типа 3, но сохранит грубые качественные эффекты.

### 1.3. Рост пробки

Если входной транспортный поток в некоторую фиксированную область равен $J_{\text{in}}$, а выходной $J_{\text{out}} < J_{\text{in}}$, то количество автомобилей в данной области за время $t$ увеличится на

$$t(J_{\text{in}} - J_{\text{out}}).$$

Так будет, однако, только если рассматриваемая область не находится на самом транспортном пути. Например, если автомобили скапливаются в пробке на самой дороге, то ответ другой. Дело в том, что область сама может расти за счет скапливающихся автомобилей. Чтобы уточнить эти утверждения, надо уточнить модель.

Пусть автомобили одинаковой длины $d$ едут в потоке (по одной полосе) со скоростью $v$ один за другим на одинаковом расстоянии $d^+$ между ними. Пусть в течение времени $t$ движение остановлено неким препятствием, например, красным светофором. При этом автомобили останавливаются на расстоянии $d_0^+ < d^+$ до предыдущего автомобиля.

**Упражнение 1.** Доказать, что за время $t \to \infty$ пробка (то есть максимальная длина $L(t)$ участка, где все автомобили стоят) перед препятствием будет иметь длину, асимптотически равную

$$L(t) \underset{t\to\infty}{\sim} tv\frac{d + d_0^+}{d^+ - d_0^+}. \tag{2}$$

По-видимому, этот результат зависит в действительности лишь от средних величин и остается верным при возможности обгона. Это сделано в [16] для независимого движения автомобилей (то есть когда автомобили не мешают друг другу), причем скорости автомобилей имеют флуктуации, однако средние скорости всех автомобилей одинаковы и равны $v$. Но доказательство там совсем не просто. Другие модели роста пробки см. в [17] и главе 2.

**Локальные расширения и сужения трассы.** Что происходит при переходе участка дороги с $k$ полосами в участок с $l$ полосами? Пусть этот переход происходит в точке с координатой $x = 0$.

*Случай* $k < l$. Пусть максимально разрешенная скорость равна $v_{\max}$ и предполагается дисциплинированность водителей. Пусть автомобили движутся по $k$-полосной трассе со скоростью $v < v_{\max}$ и быстрее ехать невозможно по причине фундаментального соотношения между плотностью автомобилей $\rho$ и их скоростью:

$$d + D^+(v) = \rho^{-1}k.$$

Тогда по $l$-полосной трассе длины $L$ автомобили теоретически могут сохранить $\rho$ и двигаться с такой же скоростью, но $\rho$ может скорректироваться



так, что автомобили смогут двигаться быстрее с некоторой большей скоростью $v_1$. Выгода во времени будет

$$\frac{L}{v} - \frac{L}{v_1}.$$

*Случай* $k > l$. Тогда возможны три разных ситуации.

Свободный поток. Если поток очень редкий, то автомобили будут подъезжать к точке 0 в одиночку и не заметят перехода.

Растущая пробка. Обозначим через $J_k$ текущий входящий поток и через $J_{l,\max}$ — 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 [9] автомобили задаются точками

$$\ldots < x_i(t) < x_{i-1}(t) < \ldots$$

на прямой. В начальный момент времени $t = 0$ автомобили стоят и образуют пуассоновское точечное поле с плотностью $\rho < 1$. Автомобили могут иметь две скорости: 0 или 1; обгоны запрещены. Каждый стоящий автомобиль, через независимое экспоненциально распределенное время со средним 1, начинает двигаться со скоростью 1. Может случиться так, что автомобиль с номером $i$ доедет до автомобиля $i - 1$ пока тот еще не начал двигаться. Тогда он останавливается и начинает двигаться через экспоненциальное время после того как начнет двигаться автомобиль $i - 1$. Такое правило действует всегда. Этот процесс в некотором смысле описывает выезд автомобилей из пробки.

Основной результат состоит в том, что с вероятностью 1 каждый автомобиль будет останавливаться только конечное число раз (при условии $\rho < 1$). Пусть $t_i$ — момент времени, начиная с которого автомобиль $i$ все время движется. Тогда для любых $i > k$ и любых $t_i$, $t_{i-1}$, $\ldots$, $t_{i-k}$ случайные величины

$$x_{i-1}(t_{i-1}) - x_i(t_i), \quad x_{i-2}(t_{i-2}) - x_{i-1}(t_{i-1}), \quad \ldots, \quad x_{i-k}(t_{i-k}) - x_{i-k+1}(t_{i-k+1})$$

будут независимы и экспоненциально распределены. Иначе говоря, после выезда из пробки автомобили будут образовывать пуассоновскую конфигурацию той же самой интенсивностью $\rho$, что и в начале.

Пусть теперь в момент 0 пуассоновский точечный поток с плотностью $\rho$ находится на левой полуоси. Каждая точка движется со скоростью $v > 0$, если расстояние до предыдущей точки не меньше некоторого $d_{\text{eff}} > 0$, и стоит в противном случае. Здесь очевидно, что каждая частица не останавливается, начиная с некоторого момента. Но здесь можно получить больше. Рассмотрим следующие случайные величины: $\tau_k^{(1)}$ — случайное время начала движения $k$-й точки, $\tau_k^{(2)}$ — случайное время, начиная с которого эта точка больше не останавливается, $x_k$ — расстояние до первой точки, начиная с момента $\tau_k^{(2)}$.

**Задача 1**$^{**}$**.** Найти асимптотику распределений этих случайных величин при $k \to \infty$.

Связь с задачей задержки очевидна. Пусть есть две полосы и на каждой полосе интенсивность потока $\rho$; объединенный поток, таким образом, имеет плотность $2\rho$. Автомобилям из первой полосы надо втиснуться во вторую. Алгоритмы втискивания могут быть разными. Например, любой автомобиль втискивается независимо от других, если его расстояние (по оси $x$) до предыдущего и последующего автомобиля из второй полосы было не менее некоторого числа $d^+$.

### 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. Будем говорить, что автомобиль $i$ имеет впереди себя препятствие в момент $t$, если

$$x_i(t-0) = x_{i-1}(t).$$

Процессы $w_i(t)$ взаимно независимы и определяются лишь психикой индивидуального водителя. Предположим, что существуют константы $0 < C_1 < C_2 < \infty$ такие, что для всех $t$, $i$

$$C_1 < w_i(t) < C_2.$$

Поток задается начальным положением $x_i(0)$ автомобилей, а их движение определяется как

$$x_i(t) = x_i(0) + \int_0^t v_i(s)\,ds,$$

где $v_i(t)$ — определяемая ниже скорость автомобиля в потоке. При этом начальные положения таковы, что расстояния $d_i^+(0)$ независимы и, например,



экспоненциальны с заданным параметром $\rho(0)$. Процесс будет полностью определен, если для всех $t_1, \ldots, t_n, i_1, \ldots, i_n$ мы зададим конечномерные распределения векторов:

$$(v_{i_1}(t_1), \ldots, v_{i_n}(t_n)),$$

где среди индексов $i_k$ могут быть одинаковые. Эти распределения полностью определяются следующими правилами:

1) *Правило свободной дороги*. Если ни один из автомобилей $i_1, \ldots, i_k$ при $k \leqslant n$ не имеет впереди себя препятствия, то распределение вектора $(v_{i_1}(t_1), \ldots, v_{i_k}(t_k))$ совпадает с распределением вектора $(w_{i_1}(t_1), \ldots, w_{i_k}(t_k))$ и является независимым от распределения вектора $(v_{i_{k+1}}(t_{k+1}), \ldots, v_{i_n}(t_n))$.

2) *Правило препятствия*. Если автомобиль $i$ имеет впереди себя препятствие в момент $t$, то $v_i(t) = v_{i-1}(t)$.

3) *Правила обгона*. Если автомобиль $i$ имеет впереди себя препятствие в момент $t$, то он меняется местами с предыдущим автомобилем с некоторой интенсивностью $\lambda$ в течение (случайного) интервала времени пока $w_i(t) > v_{i-1}(t)$. Смысл этого условия состоит в том, что водитель обгоняет, если его активность высока в течение некоторого промежутка времени.

Уже для этого простейшего определения транспортного потока с зависимыми от времени скоростями есть много задач. Некоторые из них мы сейчас сформулируем.

Назовем свободной фазой случай, когда автомобиль не задерживается при обгоне догоняемого автомобиля, то есть интенсивность обгона равна бесконечности. Тогда для любых автомобилей с индексами $i$, $j$ их скорости независимы, и, значит, ковариации

$$\operatorname{cov}_{ij}(t) = Ev_i(t)v_j(t) - Ev_i(t)Ev_j(t) = 0.$$

**Задача 2**$^{**}$**.** Для заданных распределений процессов $w_i(t)$ существует константа $0 < \lambda_0 < \infty$ такая, что при $\lambda < \lambda_0$ существует предельный стационарный процесс (по $i$ и по $t$), в котором ковариации $\operatorname{cov}_{ij}(t)$ убывают экспоненциально по $|j-i|$.

Назовем этот тип движения фазой с ближним порядком. Существование фазы дальнего порядка определяется следующей гипотезой.

**Задача 3**$^{**}$**.** Существует константа $0 < \lambda_{\text{cr}} < \infty$ такая, что при $\lambda > \lambda_{\text{cr}}$ существует предельный стационарный процесс, в котором ковариации $\operatorname{cov}_{ij}(t)$ не стремятся к нулю при $|j-i| \to \infty$.

Будет ли $\lambda_{\text{cr}} = \lambda_0$ предыдущей задачи?

Эти три фазы могут иметь отношение к фазам, определенным Б. С. Кернером [15].

**Задача 4**$^{**}$**.** Определить подобный процесс с длинами $d_i$, $d_i^+$, а также с дополнительными индексами, соответствующими полосам движения, и с поведением водителя, зависящим не только от предыдущего, но и от следующего автомобиля. Какие дополнительные качественные эффекты могут ловить эти модели?

## 2. Расчет средней скорости на автотрассе

Мы приводим здесь простейшую постановку задачи о снижении средней скорости движения автомобиля 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.

### 2.1. Дорога как одномерная сеть массового обслуживания

Следующая модель заимствована из [8, с. 117]. Пусть есть бесконечная дорога и два типа автомобилей, задаваемые точками на бесконечной прямой, которые движутся в одном направлении. Автомобили первого типа (быстрые) двигаются с постоянной скоростью $v_1$, автомобили второго типа (медленные) имеют постоянную скорость $v_2$, где $v_1 > v_2$.

Предположим, что быстрые автомобили в начальный момент времени образуют пуассоновскую случайную конфигурацию (пуассоновский точечный поток) на всей прямой с плотностью $\lambda_1$. Медленные автомобили расположены в момент $t = 0$ в точках

$$x_0 = 0 < x_1 < \ldots < x_n < \ldots,$$

причем расстояния $x_k - x_{k-1}$ одинаково распределены со средним $\lambda_2^{-1}$ (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 $\mu$ он обгоняет медленного, то есть начинает ехать со скоростью $v_1$. Если быстрый автомобиль догоняет группу быстрых автомобилей, следующих за медленным, то обгон



происходит в порядке очереди, точнее, в том порядке, в котором быстрые автомобили догоняли данный медленный автомобиль. Без ограничения общности, скорости медленных автомобилей можно считать равными нулю $v_2 = 0$, а скорости быстрых соответственно равными $v = v_1 - v_2$. Поэтому каждый медленный автомобиль можно представлять узлом обслуживания, на который приходят клиенты (быстрые автомобили) и в порядке очереди (то есть прибытия) ждут обслуживания (обгона), а затем обслуживаются с интенсивностью обслуживания $\mu$.

Теперь эта задача может быть сведена к линейной сети массового обслуживания, которую мы сейчас опишем. Имеется бесконечная последовательность

$$S_0 \to \ldots \to S_k \to S_{k+1} \to \ldots$$

узлов обслуживания двух типов. Каждый узел $S_k$ представляет собой систему типа $M/M/1$ с дисциплиной обслуживания FIFO (first in — first out), то есть обслуживание в порядке естественной очереди. Эти узлы соответствуют медленным автомобилям, а требования — быстрым. Например, узел $S_0$ соответствует крайнему левому медленному автомобилю. Вторая буква $M$ означает экспоненциальность времени обслуживания. Это вместе с дисциплиной FIFO отвечает формулировке нашей модели. Первая буква $M$ означает пуассоновость входящего потока прибывающих требований. Так, на узел $S_0$ поступление требований образует стационарный пуассоновский поток с интенсивностью $\lambda_1 v$. Из элементарной теории очередей известно, во-первых, что если $\lambda_1 v < \mu$, то устанавливается стационарный режим с вероятностями $P_n$ того, что длина очереди равна $n$:

$$P_n = (1 - r)r^n, \quad r = \frac{\lambda_1 v}{\mu}.$$

Во-вторых, известно (теорема Бюрке (Burke)), что в стационарном режиме выходящий поток из системы типа $M/M/1$ будет пуассоновским с интенсивностью, равной интенсивности входящего потока, то есть в нашем случае это $\lambda_1 v$.

После первого узла, со случайным, но одинаковым для всех автомобилей временным сдвигом $\frac{x_1 - x_0}{v}$, поток требований поступает на узел $S_1$, где также устанавливается стационарный режим.

Найдем среднюю скорость быстрого автомобиля на интервале $(x_0, x_N)$, $N \to \infty$. 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 равное $\lambda_2^{-1}$) быстрый автомобиль в среднем проходит за время $(\mu-\lambda_1 v)^{-1}+(\lambda_2 v)^{-1}$.

Таким образом, средняя скорость быстрого автомобиля составит

$$v_{\text{mean}}=\frac{\lambda_2^{-1}}{(\mu-\lambda_1 v)^{-1}+(\lambda_2 v)^{-1}}.$$

В следующих разделах мы рассмотрим более сложную ситуацию с более общими распределениями.

### 2.2. Снижение средней скорости из-за ремонтных работ

По длинной автотрассе едут автомобили с постоянной скоростью $v$, встречая препятствия. Препятствия обычно имеют малый размер в сравнении с расстояниями между ними, поэтому можно представлять их точками. Они возникают на произвольном участке дороге $(x,\, x+dx)\subset\mathbb{R}$ за время $(t,\, t+dt)\subset\mathbb{R}$ с вероятностью $\lambda\,dx\,dt$. Точнее говоря, пары (место и момент возникновения препятствия) $(x_j,\, t_j)\in\mathbb{R}\times\mathbb{R}_+$образуют пуассоновское точечное поле $\Pi$ на $\mathbb{R}\times\mathbb{R}_+$ с интенсивностью $\lambda$. Другое эквивалентное определение состоит в том, что для любого интервала $I\subset\mathbb{R}$ есть пуассоновский поток прибывающих препятствий интенсивности $\lambda|I|$, причем в момент прибытия препятствие выбирает точку равномерно на интервале $I$.

Предположим, что $j$-е препятствия находится на дороге некоторое случайное время $\tau_j$, после чего оно убирается с дороги. Будем считать, что $\tau_j$ — независимые одинаково распределенные случайные величины с функцией распределения $Q(t)$, не зависящие от пуассоновского точечного поля $\Pi$. Предположим, что первые два момента с.в. $\tau_j$ 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-нибудь 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 $x\to\infty$.

Пусть $b=\lambda m_Q$, $\zeta$ — с.в. с плотностью распределения

$$h(t)=m_Q^{-1}(1-Q(t)). \tag{3}$$

**Упражнение 2.** Показать, что $h$ — плотность.

Отметим, что

$$E\zeta=\frac{1}{m_Q}\int_0^\infty t(1-Q(t))\,dt=\frac{1}{m_Q}\int_0^\infty(1-Q(t))\,d\Big(\frac{t^2}{2}\Big)=\frac{1}{2m_Q}\int_0^\infty t^2\,dQ(t)=\frac{m_Q^{(2)}}{2m_Q},$$

где $m_Q^{(2)}$ — второй момент распределения $Q(t)$.

Определим с.в. $\alpha=\min(\eta,\,\zeta)$, где равенство по распределению, при этом с.в. $\eta$, $\zeta$ считаются независимыми и с.в. $\eta$ имеет функцию распределения $F(u)$. Положим

$$a=E\alpha.$$

**Теорема 1.** *С вероятностью* 1 *при* $x\to\infty$

$$\frac{x}{T(x)}\to\frac{v}{1+abv}. \tag{4}$$

Доказательство. Без ограничения общности можно считать, что автомобиль выезжает в точке $x=0$ в момент времени $t=0$. Пусть $T_0(x)$ — время простоя автомобиля. Тогда, очевидно, $T(x)-T_0(x)=v^{-1}x$ и

$$\frac{x}{T(x)}=\frac{x}{T(x)-T_0(x)+T_0(x)}=\frac{1}{v^{-1}+x^{-1}T_0(x)}.$$

Поэтому достаточно найти предел отношения $\frac{T_0(x)}{x}$ при $x\to\infty$. Мы хотим показать, что

$$T_0(x)=\sum_{i=1}^{\pi(x)}\alpha_i, \tag{5}$$

где $\alpha_i$ — н.о.р. с.в., распределенные как $\alpha$, $\pi(x)$ — с.в. с пуассоновским распределением с параметром $bx$, причем $\alpha_i$ и $\pi(x)$ независимы. Смысл этой

формулы в том, что автомобиль при прохождении расстояния $x$ встретит $\pi(x)$ препятствий и потеряет случайное время $\alpha_i$ на $i$-м препятствии.

Из (5) и усиленного закона больших чисел легко следует, что $\dfrac{T_0(x)}{x} \to ab$ п.н. при $x \to \infty$.

Докажем (5). Введем маркированное пуассоновское точечное поле $\Pi_1$ на $\mathbb{R} \times \mathbb{R}_+$ с конфигурацией $(x_j, t_j, \tau_j)$, то есть $\tau_j$ — марка в точке $(x_j, t_j)$. Следующее утверждение можно найти в [5]:

**Лемма 1.** *Маркированное точечное поле* $\Pi_1$ *эквивалентно по распределению пуассоновскому полю на* $\mathbb{R} \times \mathbb{R}_+^2$ *с интенсивностью*

$$\lambda\, dx\, dt\, dQ(t).$$

Препятствия, возникающие на дороге, удобно представлять в виде горизонтальных отрезков, изображенных на рис. 1. Координаты начальной точки определяют место и время возникновения препятствия (пара $(x_j, t_j)$). Длина отрезка — время пребывания препятствия на дороге (марка $\tau_j$).

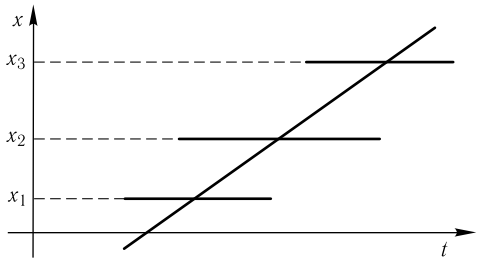



Рис. 1

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 [7].

**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.



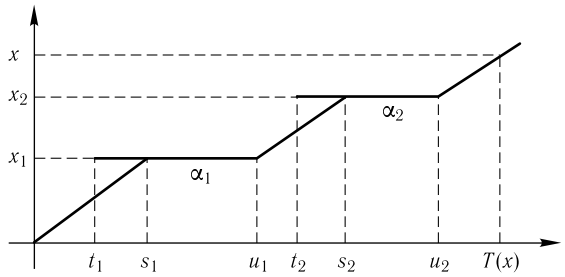



Рис. 2

Из леммы 2 и пространственно-временной однородности пуассоновского точечного поля $\Pi$ следует, что точки $x_i$ образуют пуассоновский процесс интенсивности $b$.

Под временем жизни препятствия будем понимать время его пребывания на дороге. Назовем остаточным временем жизни препятствия время его нахождения на дороге после того как его догнал автомобиль. Другими словами, это задержка автомобиля, если объезд невозможен.

**Лемма 3.** *Остаточное время жизни препятствия имеет распределение с плотностью* $h(s)$, *где* $h(s)$ *определяется формулой* (3).

В самом деле, из свойств пуассоновского точечного поля следует, что условное распределение остаточного времени жизни препятствия при условии, что полное время жизни равно $t$, совпадает с равномерным распределением на отрезке $[0, t]$. В силу леммы 2 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, т. е. $\alpha_i = \min(\eta, \zeta)$. Теорема доказана. □

Обсудим результат. Смысл константы $a$ мы уже пояснили, а константа $b$ имеет смысл стационарной плотности препятствий в пространстве.

Этот результат довольно точен при малой плотности автомобилей, так как около препятствий будет по одному автомобилю. При высокой плотности автомобилей время объезда будет пересчитываться (увеличиваться) в зависимости от средней длины очереди перед препятствием.

### 2.3. Снижение средней скорости из-за медленных автомобилей

Автотрасса описывается действительной осью $\mathbb{R}$. Потоки считаются не очень плотными, поэтому длина автомобиля роли не играет, и в данный момент времени положение автомобиля задается точкой $x_i(t) \in \mathbb{R}$, где $i$ — индекс, нумерующий автомобили. Каждый автомобиль имеет фиксированный маршрут: место и время въезда $x_{i,\text{in}}$, $t_{i,\text{in}}$, а также предписанное ему место выезда $x_{i,\text{out}}$. Но время выезда $t_{i,\text{out}}$ зависит от степени загруженности дороги. Мы определяем среднюю скорость автомобиля $i$ как

$$V_i = \frac{x_{i,\text{out}} - x_{i,\text{in}}}{t_{i,\text{out}} - t_{i,\text{in}}}.$$

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 $\tau_{ij}$ 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 $\Pi$ интенсивности $\lambda$, 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 $\rho_j$, после чего он съедет с дороги. С.в. $\rho_i$ независимы и одинаково распределены с общей функцией распределения $G(r)$. С.в. $\rho_j$ не зависят также от пуассоновского точечного поля $\Pi$. Будем предполагать существование первых двух моментов с.в. $\rho_1$. 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 $\tau_{ij}$ для того, чтобы обогнать $j$-й медленный автомобиль или всю эту группу автомобилей. При этом время обгона не зависит от размера группы. С.в. $\tau_{ij}$ предполагаются независимыми и одинаково распределенными с функцией распределения $F(u)$.

Пусть $d = \lambda m_G(v_2^{-1} - v_1^{-1})$. Введем с.в. $\beta$ 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. $\tau_{1,1}$, $\beta$ считаются независимыми. Отметим, что

$$E\beta = \frac{m_G^{(2)}}{2m_G}.$$

Положим $c = E\gamma$.

**Теорема 2.** *С вероятностью* 1 *при* $x \to \infty$

$$\frac{x}{T(x)} \to \bar{v}_1 = \frac{1 + dc}{1 + dcv_1v_2^{-1}}v_1. \tag{6}$$

Доказательство. Покажем, что этот случай сводится к рассмотренному случаю $v_2 = 0$. Введем систему координат, которая движется со скоростью $v_2$ относительно исходной. Найдем среднюю скорость быстрого автомобиля относительно новой системы координат по формуле (4), подставляя $v = v_1 - v_2$, $b = \dfrac{\lambda m_G}{v_2}$, $a = \dfrac{c}{v_2}$:

$$\frac{1}{(v_1 - v_2)^{-1} + \dfrac{\lambda mc}{v_2^2}} = \frac{v_1 - v_2}{1 + dcv_1v_2^{-1}}.$$

Тогда средняя скорость быстрого автомобиля относительно исходной системы координат составит

$$\bar{v}_1 = \frac{v_1 - v_2}{1 + dcv_1v_2^{-1}} + v_2 = \frac{1 + dc}{1 + dcv_1v_2^{-1}}v_1. \qquad \square$$

## 3. Критерии образования пробок в сложных транспортных сетях

Обязательным атрибутом транспортной сети (например, городских улиц) является граф, где множество $V$ вершин представляет перекрест-

ки (узлы или пункты обслуживания), а множество ребер $L = \{(i, j)\}$ — отрезки путей без перекрестков. Пусть число перекрестков равно $N$. Мы предполагаем, что между двумя перекрестками существует не более одного пути без перекрестков.

Наиболее разработанными являются два класса сетей. С одной стороны, это (по имени авторов и в порядке увеличения общности) сети Джексона, BCMP-сети, DB-сети (см., например, [8,26]). В них требование (сообщение, автомобиль, работа) обслуживается в каждом проходимом ими узле и затем выбирает случайно следующий узел. С другой стороны, — сети Келли (см. [8]), где каждое требование имеет заранее фиксированный маршрут. Эти два класса сетей связаны как общей техникой, так и близостью результатов. Именно они обладают замечательным свойством мультипликативности — стационарные распределения в них имеют вид так называемой продакт-формы. Мы рассматриваем только первый класс сетей.

### 3.1. 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: $P = \{p_{ij}\}_{i,j=1,\ldots,N}$, 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 $V = \{1, \ldots, N\}$. Эта цепь Маркова предполагается неразложимой. В этом случае система линейных уравнений

$$\rho P = \rho,\ \rho = (\rho_1, \ldots, \rho_N) \iff \sum_{i=1}^{N} \rho_i p_{ij} = \rho_j,\ j = 1, \ldots, N, \qquad (7)$$

имеет единственное решение (с точностью до произвольного множителя). Нормированное решение имеет вид

$$\pi_i = \frac{\rho_i}{\sum\limits_{i=1}^{N} \rho_i}, \quad i = 1, \ldots, N.$$

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 $\mu_{ik}(n_i) \leqslant \mu_i(n_i)$. При этом потребуем, чтобы

$$\sum_{k=1}^{n_i} \mu_{ik}(n_i) = \mu_i(n_i).$$

Например, общий ресурс может быть разделен в равной степени между всеми автомобилями в очереди:

$$\mu_{ik}(n_i) = \frac{\mu_i(n_i)}{n_i}.$$

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 $\mu_{ik}(n_i)$ полностью определяют дисциплину обслуживания в узлах.

Динамика сети описывается с помощью $N$-мерной марковской цепи с непрерывным временем $\xi(t) = (\xi_i(t),\ i = 1, \ldots, N)$, 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 $\bar{n} = (n_1, \ldots, n_N)$, что $n_1 + \ldots + n_N = M$.

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 $T_{ij}\bar{n} = \bar{n} - e_i + e_j$, $i \neq j$, с интенсивно-

стью

$$\alpha(\bar{n},\, T_{ij}\bar{n}) = \mu_i(n_i)p_{ij}, \tag{8}$$

при условии, что $n_i \neq 0$. Переход $\bar{n} \to T_{ij}\bar{n}$ 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 $(\mu_i(n_i),\ i = 1,\ \ldots,\ N)$.

Пусть $\rho = (\rho_1,\ \ldots,\ \rho_N)$ — решение уравнения (7), которое рассматривается как формальное уравнение для интенсивностей $\rho_i$ входящих потоков в узлы (в стационарном режиме они равны выходящим). Решая эти уравнения, находим $\rho_i$, и тогда стационарное распределение $\nu(n_1,\ \ldots,\ n_N)$ марковской цепи $\xi(t)$ будет иметь вид

$$\nu(n_1,\ \ldots,\ n_N) = \frac{1}{Z_{NM}} \prod_{i=1}^{N} \frac{\rho_i^{n_i}}{\mu_i(1)\mu_i(2)\ldots\mu_i(n_i)}, \tag{9}$$

где нормирующий множитель (малая статсумма)

$$Z_{NM} = \sum_{n_1+\ldots+n_N=M} \prod_{i=1}^{N} \frac{\rho_i^{n_i}}{\mu_i(1)\mu_i(2)\ldots\mu_i(n_i)},$$

что проверяется подстановкой ответа (9) в уравнения Колмогорова для стационарных вероятностей, см., например, [8, 29].

### 3.2. 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 $\lambda_i$, $i \in \{1,\ \ldots,\ N\}$.

Зададим матрицу маршрутизации $P = \{p_{ij}\}_{i,j=1,\ldots,N}$, где матрица $P$ неразложима и

$$\forall i\colon \sum_{j=1}^{N} p_{ij} \leqslant 1 \quad \exists i_0\colon \sum_{j=1}^{N} p_{i_0 j} < 1. \tag{10}$$

Как и в случае замкнутой сети, $p_{ij}$ — 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 $\eta(t) = (\eta_i(t),\ i = 1,\ \ldots,\ N)$, где $\eta_i(t)$ — число автомобилей в $i$-м узле в момент времени $t$. Случайный процесс $\eta(t)$ является марковской цепью с непрерывным временем и с пространством состояний $S$, где $S$ — множество $N$-мерных векторов с неотрицательными целочисленными координатами $\bar{n} = (n_1,\ \ldots,\ n_N)$.

Из состояния $\bar{n}$ марковская цепь $\xi(t)$ может перейти в одно из состояний $T_{ij}\bar{n} = \bar{n} - e_i + e_j$, $T_{i,0}\bar{n} = \bar{n} - e_i$, $T_i\bar{n} = \bar{n} + e_i$ с интенсивностями

$$\begin{aligned} \alpha(\bar{n},\, T_{ij}\bar{n}) &= \mu_i(n_i)p_{ij}, \\ \alpha(\bar{n},\, T_{i,0}\bar{n}) &= \mu_i(n_i)p_{i,0}, \\ \alpha(\bar{n},\, T_i\bar{n}) &= \lambda_i, \end{aligned} \tag{11}$$

при условии, что $T_{ij}\bar{n},\ T_{i,0}\bar{n},\ T_i\bar{n} \in S$.

Таким образом, марковская цепь $\eta(t)$ однозначно определяется триплетом $(\lambda,\ \mu,\ P)$, где $\lambda = (\lambda_1,\ \ldots,\ \lambda_N)$ — вектор интенсивностей внешних потоков, $\mu = (\mu_i(n_i),\ i = 1,\ \ldots,\ N)$ — набор интенсивностей обслуживания в узлах и $P$ — матрица маршрутизации.

Рассмотрим формальное уравнение для интенсивностей входящих потоков в узлы (в стационарном режиме они равны выходящим):

$$\rho = \lambda + \rho P \iff \rho_i = \lambda_i + \sum_{k=1}^{N} \rho_k p_{ki} \quad \forall i. \tag{12}$$

В силу условия (10) и неразложимости матрицы $P$ это уравнение имеет единственное решение, которое можно представить в виде

$$\rho = \lambda + \sum_{n=1}^{\infty} \lambda P^n.$$

Далее рассмотрим случай, когда интенсивности обслуживания $\mu_i(n_i) \equiv \mu_i$ не зависят от числа автомобилей в узлах. Введем нагрузки в узлах по формуле

$$r_i = \frac{\rho_i}{\mu_i}, \quad i = 1,\ \ldots,\ N.$$

Следующую теорему можно найти, например, в [8, 20], она называется иногда теоремой Гордона—Ньюэлла.

**Теорема 3.** *Марковская цепь $\eta(t)$ является эргодической тогда и только тогда, когда для всех $i = 1,\ \ldots,\ N$ будет $r_i < 1$. При этом*

*стационарное распределение цепи имеет вид*

$$\sigma(n_1, \ldots, n_N) = \prod_{i=1}^{N} (1 - r_i) r_i^{n_i}.$$

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 $i_1, \ldots, i_k$ нагрузка строго больше 1, то марковская цепь $\eta(t)$ является транзиентной (см., например, [33]). Это свидетельствует о том, что средние очереди в узлах $i_1, \ldots, i_k$ стремятся к бесконечности с течением времени. Подробный анализ открытых сетей дан в работе [19]. В частности, показано, что в узлах, где нагрузка больше 1, средние очереди увеличиваются линейно с ростом времени. При этом найдены скорости роста средних очередей.

### 3.3. Алгоритм вычисления критической нагрузки в замкнутых сетях

Этот раздел основан на работе [18]. Мы рассмотрим последовательность замкнутых сетей $J_N$, $N = 1, 2, \ldots$ Сеть $J_N$ состоит из $N$ узлов и $M = M(N)$ автомобилей. Интенсивности обслуживания в узлах сети $J_N$ не зависят от длины очереди: $\mu_{iN}(n_i) \equiv \mu_{iN}$. Пусть $P_N = \{p_{ijN}\}$ — матрица маршрутизации в $N$-й сети; $P_N$ предполагается неразложимой.

Пусть $\rho_N = (\rho_{1,N}, \ldots, \rho_{NN})$ — вектор с положительными компонентами, удовлетворяющий уравнению

$$\rho_N = \rho_N P_N. \tag{13}$$

Относительные нагрузки в узлах определяются как

$$r_{iN} = C_N^{-1} \rho_{iN} \tau_{iN},$$

где $\tau_{iN} = \mu_{iN}^{-1}$ и $C_N = \max\limits_{i=1,\ldots,N} \rho_{iN}\tau_{iN}$. Очевидно, что $r_{iN} \in [0, 1]$.

В соответствии с (9) стационарное распределение числа автомобилей $\xi_{iNM}$ в узлах сети $J_N$ равно

$$P_{NM}(\xi_{iNM} = n_i,\ i = 1, \ldots, N) = \frac{1}{Z_{NM}} \prod_{i=1}^{N} r_{iN}^{n_i},$$

где нормирующий множитель (малая статсумма)

$$Z_{NM} = \sum_{n_1 + \ldots + n_N = M} \prod_{i=1}^{N} r_{iN}^{n_i}. \tag{14}$$

Многие важные характеристики сети выражаются через статсумму.



**Упражнение 3.** Показать, что среднее число автомобилей в $i$-м узле в стационарном режиме равно

$$m_{iNM} = E\xi_{iNM} = \frac{r_{iN}}{Z_{NM}} \frac{\partial Z_{NM}}{\partial r_{iN}}. \tag{15}$$

Ниже мы будем требовать слабую сходимость относительных нагрузок $r_{iN}$. Точнее, определим выборочную меру на отрезке $[0, 1]$:

$$I_N(A) = \frac{1}{N} \sum_{i:\ r_{iN} \in A} 1,$$

где $A$ — произвольное борелевское множество из отрезка $[0, 1]$. Предположим, что при $N \to \infty$ меры $I_N$ слабо сходятся к некоторой вероятностной мере $I$, заданной на отрезке $[0, 1]$.

Нас будет интересовать случай больших $N$, $M$, точнее $N$, $M \to \infty$, причем так, что $\dfrac{M}{N} \to \lambda = \text{const}$, то есть удельное число автомобилей на один узел постоянно. Именно это число определяет существование пробок.

**Замечание 1.** Интересно найти конкретные последовательности растущих графов, для которых предельная мера $I$ явно описывается. Некоторые примеры, где мера $I$ одноточечна, см. в ссылках к работе [18], см. также с. 157–160 в [29].

В терминах предельной меры $I$ мы найдем критическое значение плотности $\lambda_{\text{cr}}$, так что при $\lambda < \lambda_{\text{cr}}$ средние длины очередей равномерно ограничены. Если $\lambda \geqslant \lambda_{\text{cr}}$, 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

$$\lambda_{\text{cr}} = \lim_{z \to 1-} h(z).$$

Будем предполагать, что $\lambda_{\text{cr}} > 0$.

**Теорема 4.** 1) *Если* $\lambda < \lambda_{\text{cr}}$, *то средние очереди равномерно ограничены*: *существует такая константа* $B$, *что* $m_{iN} < B$ *равномерно по* $N \geqslant 1$ *и* $1 \leqslant i \leqslant N$.

2) *Если* $\lambda \geqslant \lambda_{\text{cr}}$ *и* $i(N)$ *удовлетворяет условию* $r_{i(N),N} = 1$, *то* $m_{i(N),N} \to \infty$ *при* $N \to \infty$, *т. е. пробки будут в тех узлах, где нагрузка максимальна*.

При $z \in \mathbb{C} \setminus [1, +\infty)$ положим

$$S_N(z) = -\lambda(1+\varepsilon_N)\ln z - \frac{1}{N}\sum_{i=1}^{N}\ln(1-zr_{iN}),$$
$$S(z) = -\lambda \ln z - \int_0^1 \ln(1-zr)\,dI(r), \tag{16}$$

где $\lambda(1+\varepsilon_N) = \dfrac{M}{N}$.

Введем производящую функцию (большую статсумму):

$$\Xi_N(z) = \sum_{M=0}^{\infty} z^M Z_{NM} = \prod_{i=1}^{N} \frac{1}{1-zr_i}, \quad |z| < 1.$$

По формуле Коши и формуле (16) имеем следующее выражение для статсуммы (14):

$$Z_{NM} = \frac{1}{2\pi i}\int_\gamma \frac{\Xi_N(z)}{z^{M+1}}\,dz = \frac{1}{2\pi i}\int_\gamma \frac{\exp(NS_N(z))}{z}\,dz, \tag{17}$$

где $\gamma = \{z \in \mathbb{C}\colon |z| = \sigma < 1\}$. Для средних, согласно (15), имеем

$$m_{iN} = \frac{1}{2\pi i Z_N}\int_\gamma \frac{r_{iN}}{1-zr_{iN}}\exp(NS_N(z))\,dz. \tag{18}$$

Можно показать, что для стационарного распределения длин очередей справедлива формула

$$P_{NM}(\xi_{1,N,M} = n_1, \ldots, \xi_{K,N,M} = n_K) =$$
$$= \frac{1}{2\pi i Z_N}\int_\gamma z^{-1}\prod_{i=1}^{K}(1-zr_{iN})(zr_{iN})^{n_i}\exp(NS_N(z))\,dz. \tag{19}$$

В доказательстве теорем этого раздела существенную роль играет *метод перевала* (см. [27]), точнее его обобщение, поскольку функция в показателе экспоненты зависит от $N$. Из уравнения

$$\frac{\partial S_N(z)}{\partial z} = 0 \tag{20}$$

находятся точки перевала. Пусть $z_{0,N}$ — корень этого уравнения, лежащий в интервале $(0, 1)$.

**Упражнение 4.** Показать, что все корни уравнения (20) действительны и положительны. Всегда существует единственный корень, лежащий в интервале $(0, 1)$.



Пусть $z_0$ — корень уравнения

$$h(z) = \frac{\lambda}{z} \iff \frac{\partial S(z)}{\partial z} = 0, \tag{21}$$

лежащий в интервале $(0, 1)$.

**Упражнение 5.** 1) Доказать, что при всех $\lambda$ существует предел $\lim\limits_{N\to\infty} z_{0,N} = z_0 = z_0(\lambda) > 0$.

2) Если $\lambda < \lambda_{\mathrm{cr}}$, то $z_0(\lambda)$ — корень уравнения (21); $z_0(\lambda)$ строго возрастает по $\lambda$, $z_0(\lambda) \in (0, 1)$, $\lim\limits_{\lambda\to\lambda_{\mathrm{cr}}-} z_0(\lambda) = 1$.

3) Если $\lambda \geqslant \lambda_{\mathrm{cr}}$, то $z_0 = 1$.

В следующей теореме мы находим асимптотику статсуммы и предельное распределение для последовательности замкнутых сетей $J_N$.

**Теорема 5.** *Пусть* $\lambda < \lambda_{\mathrm{cr}}$.

1) *При* $N \to \infty$ *статсумма* $Z_N$ *и свободная энергия* $F_N = \dfrac{1}{N}\ln Z_N$ *имеют следующие асимптотики*:

$$Z_N \sim \frac{\exp(NS_N(z_{0,N}))}{z_0\sqrt{2\pi N S''(z_0)}}, \quad F_N = \frac{1}{N}\ln Z_N \sim S(z_0).$$

2) *Если при* $i = 1, \ldots, K$ *существуют пределы* $r_i = \lim\limits_{N\to\infty} r_{iN}$, *то*

$$\lim_{N\to\infty} m_{iN} = \frac{z_0 r_i}{1 - z_0 r_i},$$
$$\lim_{N\to\infty} P_{NM}(\xi_{1,N,M} = n_1, \ldots, \xi_{K,N,M} = n_K) = \prod_{i=1}^{K}(1-z_0r_i)(z_0r_i)^{n_i}.$$

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 $U_d(v) = \{z \in \mathbb{C}\colon |z - v| < d\}$. Рассмотрим контур $\gamma = \{z \in \mathbb{C} : |z| = z_0(\lambda)\}$.

**Теорема 6.** *Пусть* $\lambda < \lambda_{\mathrm{cr}}$ *и* $f(\theta, z)$, $\theta \in \Theta$, — *семейство функций, голоморфных в кольце* $\{z \in \mathbb{C}\colon z_0(\lambda) - \delta_0 < |z| < z_0(\lambda) + \delta_0\}$ *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* $\theta \in \Theta$

$$\frac{1}{2\pi i}\int_\gamma f(\theta, z)\exp(NS_N(z))\,dz = \frac{f_u\exp(NS_N(z_{0,N}))}{\sqrt{2\pi N S''(z_0)}}(1+\zeta_N),$$

*где* $|\zeta_N| < 25\varepsilon$.

Доказательство этой теоремы основано на применении метода перевала (saddle-point method, см. [27]). Отличие от стандартной ситуации состоит в том, что функция в показателе экспоненты зависит от $N$. Подробное доказательство можно найти в оригинальной статье [18].

Доказательство теоремы 5. Используя теорему 6, докажем первый пункт теоремы 5. Согласно (17) имеем

$$Z_{NM} = \frac{1}{2\pi i} \int\limits_{\gamma} \frac{\exp(NS_N(z))}{z}\, dz,$$

где $\gamma = \{z \in \mathbb{C} : |z| = z_0(\lambda)\}$. Положив $f(\theta, z) = z^{-1}$, $f_u = z_0^{-1}$ и применив теорему 6, получим, что для любого достаточно малого $\varepsilon > 0$ при достаточно больших $N$

$$Z_N = \frac{\exp(NS_N(z_{0,N}))}{z_0\sqrt{2\pi N S''(z_0)}}(1 + \zeta_N), \quad |\zeta_N| < 25\varepsilon. \tag{22}$$

Второй пункт теоремы 5 доказывается аналогично с использованием формулы (18) для средней очереди и формулы (19) для совместного распределения длин очередей.

**Упражнение 6.** Доказать третье утверждение теоремы 5, используя теорему 6 и формулы (18), (19).

□

Доказательство теоремы 4. Чтобы доказать первый пункт теоремы 4, рассмотрим семейство функций

$$f(\theta, z) = \frac{A}{z} + \frac{\theta}{1 - z\theta}, \quad \theta \in \Theta = [0, 1], \quad A > 0, \quad f_u = \frac{A}{z_0}.$$

Зафиксируем малое $\varepsilon > 0$ и выберем $\sigma_u = \dfrac{\varepsilon}{8}$, $A = \dfrac{16 z_0}{(1 - z_0)\varepsilon}$. По теореме 6 имеем для достаточно больших $N$ и всех $\theta \in \Theta$

$$\frac{1}{2\pi i} \int\limits_{\gamma} \left(\frac{A}{z} + \frac{\theta}{1 - z\theta}\right) \exp(NS_N(z))\, dz = \frac{A \exp(NS_N(z_{0,N}))}{z_0\sqrt{2\pi N S''(z_0)}}(1 + \zeta_N), \ \ |\zeta_N| < 25\varepsilon. \tag{23}$$

Разделив на $Z_N$ и применив (22) к правой части получившегося равенства, получим при достаточно больших $N$

$$A + \frac{1}{Z_N}\frac{1}{2\pi i} \int\limits_{\gamma} \frac{\theta}{1 - z\theta} \exp(NS_N(z))\, dz = A\,(1 + \zeta'_N), \quad |\zeta'_N| < 30\varepsilon.$$

Из последнего равенства и формулы (18) следует равномерная ограниченность $m_{iN}$.



Докажем второе утверждение теоремы 4. Для этого потребуется следующее свойство монотонности: при любых $M_2 \geqslant M_1 > 0$ и любом $N \geqslant 1$ выполнено $m_{i,M_2,N} \geqslant m_{i,M_1,N}$.

Поскольку $z_0(\lambda)$ строго возрастает по $\lambda$, $z_0(\lambda) \in (0, 1)$ и

$$\lim_{\lambda \to \lambda_{\text{cr}}-} z_0(\lambda) = 1,$$

то функция

$$\frac{z_0(\lambda)}{1 - z_0(\lambda)}$$

монотонно возрастает и стремится к $\infty$, когда $\lambda \nearrow \lambda_{\text{cr}}$. Поэтому для любого $m > 0$ существует такое $\lambda' = \lambda'(m) < \lambda_{\text{cr}}$, 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

$$\lim_{N \to \infty} m_{1,M'(N),N} = \frac{z_0(\lambda')}{1 - z_0(\lambda')}.$$

Следовательно, при достаточно больших $N$

$$m_{1,M'(N),N} > \frac{z_0(\lambda')}{1 - z_0(\lambda')} - 1 = m.$$

Но $M/N \to \lambda \geqslant \lambda_{\text{cr}} > \lambda'$, поэтому при достаточно больших $N$ имеем $M(N) \geqslant M'(N)$. По свойству монотонности $m_{1,N} = m_{1,M(N),N} \geqslant m_{1,M',N} > m$ для достаточно больших $N$. Это доказывает, что $m_{1,N} \to \infty$. □

**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 3.1 могут быть обобщены на случай, когда сеть содержит узлы с бесконечным числом обслуживающих устройств. Пусть, например, сеть содержит один узел такого типа ($i = 0$) и $\mu_{0,N}(n) = n\nu_N$ — интенсивность обслуживания в этом узле. Пусть $\rho_N = (\rho_{0,N}, \ldots, \rho_{NN})$ — решение уравнения (13). Тогда относительные нагрузки определим по формуле

$$r_{iN} = \frac{\mu_{0,N}}{\rho_{0,N}} \frac{\rho_{iN}}{\mu_{iN}},$$

так что $r_{0,N} = 1$. Согласно (9), стационарное распределение длин очередей имеет вид

$$P_{NM}(\xi_{iNM} = n_i,\ i = 1,\ \ldots,\ N) = \frac{1}{\hat{Z}_{NM}} \frac{1}{\left(M - \sum\limits_{i=1}^{M} n_i\right)!} \prod_{i=1}^{N} r_{iN}^{n_i},$$

где

$$\hat{Z}_{NM} = \sum_{n_1+\ldots+n_N \leqslant M} \frac{1}{\left(M - \sum\limits_{i=1}^{M} n_i\right)!} \prod_{i=1}^{N} r_{iN}^{n_i},$$

и большая статсумма сети равна

$$\hat{\Xi}_N(z) = e^z \prod_{i=1}^{N} \frac{1}{1 - z r_{iN}}.$$

Положим

$$q_{iN} = \frac{r_{iN}}{p_N}, \quad \varpi = z p_N, \quad p_N = \max_{1 \leqslant i \leqslant N} r_{iN}.$$

Тогда

$$\hat{\Xi}_N(\varpi) = e^{\varpi/p_N} \prod_{i=1}^{N} \frac{1}{1 - \varpi q_{iN}}.$$

В предположении, что $p_N N \to \alpha > 0$ при $N \to \infty$, мы можем найти критическое значение плотности $\lambda$ по формуле

$$\lambda_{\text{cr}} = \alpha^{-1} + \lim_{\varpi \to 1-} \int_0^1 \frac{q}{1 - \varpi q}\, dI(q),$$

где, как и раньше, мера $I$ есть слабый предел при $N \to \infty$ выборочных мер

$$I_N(A) = \frac{1}{N} \sum_{i:q_{iN} \in A} 1,$$

где $A$ — произвольное борелевское множество из отрезка $[0, 1]$.

В работе [21] для замкнутых сетей аналогичные результаты получаются для случая более общей зависимости интенсивностей от длин очередей в узлах.

Прием усложнения графа позволяет 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 $\tau_{id}$ в новых вершинах типа

$$\sum_d \tau_{id} = \tau_i.$$

Мы ограничились задачей, когда в системе возникала хотя бы одна пробка. Интереснее рассмотреть ситуацию, когда в разных местах графа одновременно возникает много пробок.

**Связь с практикой.** Эта модель удобна тем, что все ее параметры можно оценить. Именно, на практике статистические оценки параметров $p_{ij}$, $\mu_i$ имеют вид (например, для постоянных $\mu_i$)

$$p_{ij} \sim \frac{N_{ij}(T)}{\sum\limits_j N_{ij}(T)}, \quad \mu_i = \frac{1}{T} \sum_j N_{ij}(T),$$

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 $\tau_{id}$ и, за счет изменения матрицы $P$, подсказками о выборе маршрута. Более того, в экспоненциальной модели многие случайные величины независимы, и, значит, полностью игнорируется проблема синхронизации светофоров.

Следует сказать, что проблема светофоров в стохастическом контексте только начинает изучаться, и постановки задач там должны быть более тонкими. Ранее данная проблема изучалась в жидкостных моделях. Однако пока нет понимания (а тем более вывода) связи жидкостных транспортных моделей со стохастическими (как в статистической физике).

### Литература


1. *Хейт Ф.* Математическая теория транспортных потоков. М.: Мир, 1966.
2. *Renyi A.* On two mathematical models of the traffic on a divided highway // Journal of Applied Probability. 1964. V. 1. P. 311–320.
3. *Solomon H., Wang P.* Nonhomogeneous Poisson fields of random lines with applications to traffic flow // Proc. Sixth Berkeley Symp. on Math. Statist. and Prob. 1972. V. 3. P. 383–400.
4. *Solomon H.* Geometric Probability. Philadelphia: SIAM, 1978.
5. *Daley D., Vere-Jones D.* An Introduction to the Theory of Point Processes. V. 1. Springer, 2003.
6. *Кокс Д., Смит В.* Теория восстановления. М.: Мир, 1967.
7. *Cox D. R., Isham V.* Point processes. Chapman and Hall, 1980.
8. *Kelly F.* Reversibility and stochastic networks. N.Y.: Wiley, 1979.

9. *Caceres F.*, *Ferrari P.*, *Pechersky E.* A slow-to-start traffic model related to a M/M/1 queue // Journal of Statistical Mechanics: Theory and Experiment. 2007; arXiv:0703709 cond-mat
10. *Иносэ Х.*, *Хамада Т.* Управление дорожным движением. М.: Транспорт, 1983.
11. *Gerlough D. L.*, *Huber M. J.* Traffic flow theory: A state-of-the-art report. Washington DC: Transportation Research Board, 1975; http://www-cta.ornl.gov/cta/research/trb/tft.html
12. *Blank M.* Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys. 2003. V. 111. P. 903–930.
13. *Jost D.*, *Nagel K.* Probabilistic Traffic flow breakdown in stochastic car following models // Traffic and Granular Flow. 2005. V. 03. Part 2. P. 87–103.
14. *Lotito P.*, *Mancinelli E.*, *Quadrat J.-P.* A min-plus derivation of the fundamental car-traffic law // Automatic Control IEEE Transactions. May 2005. V. 50, № 5. P. 699–705.
15. *Kerner B. S.* Introduction to modern traffic flow theory and control. Berlin: Springer, 2009.
16. *Замятин А. А.*, *Малышев В. А.* Накопление на границе для одномерной стохастической системы частиц // Проблемы передачи информации. 2007. Т. 43, № 4. С. 68–82.
17. *Lighthill M. J.*, *Whitham G. B.* On kinematic waves. II. Theory of traffic flow on long crowded roads // Proc. R. Soc. London, Se. A. 1955. V. 229. P. 281–345.
18. *Malyshev V.*, *Yakovlev A.* Condensation in large closed Jackson networks // Ann. Appl. Probab. 1996. V. 6, № 1. P. 92–115.
19. *Botvich D. D.*, *Zamyatin A. A.* On fluid approximations for conservative networks // Markov Processes and Related Fields. 1995. V. 1, № 1. P. 113–141.
20. *Fayolle G.*, *Malyshev V.*, *Menshikov M.* Topics in the constructive theory of countable Markov chains. Cambridge University Press, 1995.
21. *Fayolle G.*, *Lasgouttes J.-M.* Asymptotics and Scalings for Large Product-Form Networks via the Central Limit Theorem // Markov Processes and Related Fields. 1996. V. 2, № 2. P. 317–349.
22. Traffic Flow Theory. A state-of-art report. Revised / Editors N. H. Gartner, C. J. Messer, A. K. Rathi. 2001; http://www.tft.pdx.edu/docs/revised_monograph_2001.pdf
23. *Рюэль Д.* Статистическая механика. Строгие результаты. М.: Мир, 1971.
24. *Малышев В. А.*, *Минлос Р. А.* Гиббсовские случайные поля. М.: Наука, 1985.
25. *Малышев В. А.* Случайные грамматики // Успехи мат. наук. 1998. Т. 53, № 2. С. 107–134.
26. *Serfozo R.* Introduction to stochastic networks. Springer, 1999.
27. *Федорюк М. В.* Метод перевала. М.: Наука, 1977.
28. *Буслаев А. П.*, *Новиков А. В.*, *Приходько В. М.*, *Таташев А. Г.*, *Яшина М. В.* Вероятностные и имитационные подходы к оптимизации автодорожного движения. М.: Мир, 2003.
29. *Афанасьева Л. Г.* Очерк исследования операций. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2007.
30. *Кингман Дж.* Пуассоновские процессы. М.: МЦНМО, 2007.



31. *Blythe R. A.*, *Evans M. R.* Nonequilibrium steady states of matrix-product form: a solver's guide // J. Phys. A. 2007. V. 40, № 46. P. R333–R441.
32. *Derrida B.* Non-equilibrium steady states: fluctuations and large deviations of the density and of the current // J. Stat. Mech. Theory Exp. 2007. № 7. P07023, 45 pp.
33. *Малышев В. А.* Кратчайшее введение в современные вероятностные модели. М.: Изд-во ЦПИ при механико-математическом факультете МГУ, 2009; http://mech.math.msu.su/~malyshev/Malyshev/Lectures/course.pdf

**А. В. Колесников**

# Транспортная задача и концентрация

Первые результаты о концентрации были получены П. Леви в его книге по функциональному анализу [9]. Само название «концентрация мер» было предложено В. Мильманом. Благодаря ему же явление концентрации приобрело большую популярность в математическом сообществе и нашло многочисленные приложения в функциональном анализе, геометрии, вероятности, комбинаторике и технических науках.

Среди сугубо математических приложений упомянем: 1) новое доказательство теоремы Дворецкого о «почти круглых» сечениях выпуклых тел, 2) изопериметрические теоремы сравнения М. Громова для многообразий положительной кривизны Риччи, 3) приложения в теории гауссовских случайных процессов (например, оценки статистического супремума), 4) применения к другим функциональным и вероятностным неравенствам (неравенства типа Соболева, неравенства типа Брунна—Минковского для выпуклых тел и т. п.). Подробнее об этом можно узнать в книгах [2,7,8,13]. О вероятностных приложениях см. статью [10]. Для ознакомления с недавними результатами о концентрации и функциональных неравенствах для логарифмически вогнутых распределений также рекомендуем статью [12]. Теоремы о концентрации также позволяют оценить скорость сходимости системы к равновесному состоянию (см. комментарий ниже и приложение Е. В. Гасниковой настоящего пособия).

Классический пример свойства концентрации дает стандартное нормальное (гауссовское) $d$-мерное распределение $\gamma$. Как известно, плотность такого распределения задается формулой

$$\rho_\gamma(x) = \frac{1}{(2\pi)^{d/2}} \exp\Big(-\frac{|x|^2}{2}\Big).$$

Для произвольного множества $A$ со свойством $\gamma(A) > \frac{1}{2}$ рассмотрим его $r$-окрестность:

$$A_r = \{x\colon \exists y \in A\colon |x-y| \leqslant r\}.$$

Тогда выполнено следующее неравенство концентрации:

$$\gamma(A_r) \geqslant 1 - \frac{1}{2} e^{-\frac{r^2}{2}}. \tag{1}$$



Как мы видим, $P(A_r)$ очень быстро (квадратично экспоненциально) стремится к единице.

Равномерное распределение $\sigma$ на единичной сфере $S^{d-1} \subset \mathbb{R}^d$ также обладает аналогичным свойством:

$$\sigma(A_r) \geqslant 1 - \Big(\frac{\pi}{8}\Big)^{\frac{1}{2}} e^{-d\frac{r^2}{2}}. \tag{2}$$

Одно из важных следствий неравенств такого типа — неравенства для колебаний липшицевых функций. Пусть $f$ — 1-липшицева функция, т. е. удовлетворяющая соотношению $|f(x) - f(y)| \leqslant |x-y|$. Используя формулу коплощади

$$\int g(x)\, d\gamma = \int\limits_{-\infty}^{\infty} \gamma(\{g > t\})\, dt,$$

из (1) можно получить неравенство вида

$$\gamma\Big(\Big\{x\colon f(x) - \int f\, d\gamma > t\Big\}\Big) \leqslant 2e^{-ct^2}$$

для некоторой универсальной константы $c$. Полученное свойство обычно формулируется в виде «липшицевы функции с большой долей вероятности мало отличаются от своего среднего значения».

Несмотря на простой вид, доказательство (1) нетривиально. Классический способ основан на описании так называемых «изопериметрических множеств» — множеств, имеющих наименьшую границу среди множеств такой же меры (вероятности). В евклидовом пространстве таким, как известно, является шар. На сфере их роль выполняют сферические «шары», а в пространстве, наделенном гауссовым распределением, — полупространства $\{x\colon \langle x, a\rangle < c\}$, $a \in \mathbb{R}^d$, $c \in \mathbb{R}$. Это — хорошо известный в теории вероятностей результат, доказанный В. Судаковым и Б. Цирельсоном (и независимо от них К. Бореллем, см. [1]). Из того факта, что $r$-окрестности полупространств являются полупространствами (т. е. опять изопериметрическими множествами), несложно извлечь следствие, что функция $r \mapsto F(A, r) = \gamma(A_r)$ среди всех множеств фиксированной вероятности $p$ растет медленнее всего для изопериметрического $A$. Для этого множества функция $F(A, r)$ явно вычисляется, и мы получаем (1).

Указанный способ плох тем, что явно найти изопериметрические множества в более общем случае невозможно. Существует несколько подходов к доказательству неравенства концентрации. В настоящем пособии мы опишем связь явления концентрации с транспортной задачей, возникшей и развившейся в совершенно другой области математики. Связь эта была найдена в работе К. Мартон [11].

Транспортная задача ведет свою историю от классической работы Г. Монжа [14], написанной в 1781 году. В этой работе задача была сформу-

лирована следующим образом: имеется куча песка и яма одинаковых объемов. Как засыпать песком яму, потратив наименьшие усилия на перевозку? Конечно, это не единственная возможная «экономическая» формулировка транспортной задачи. Речь, например, может идти о перевозе грузов со складов по заданным адресам.

В дискретной постановке мы имеем набор точек $\{x_i\}$, $1 \leqslant i \leqslant N$. Задано $N$ других точек $\{y_i\}$ и функция стоимости $c(x, y)$ (например, расстояние или квадрат расстояния). Как построить взаимно однозначное отображение $T$, сопоставляющее каждой точке из первого набора точку из второго набора так, чтобы суммарная стоимость $\sum_{i=1}^{N} c(x_i, y_i)$ была наименьшей?

В дальнейшем транспортная задача переживала как периоды забвения, так и бурного развития. На языке современной математики транспортная задача была переформулирована и решена Л. В. Канторовичем в 40-х годах XX века (см. [3]) и получила в дальнейшем название задачи Монжа—Канторовича. Важным шагом в работах Канторовича было применение развитого им в теории линейного программирования метода двойственности и формулировка транспортной задачи на языке теории меры и функционального анализа. О приложениях двойственности и задач типа транспортной в технических науках см., например, главу 2 и приложение Е. В. Гасниковой настоящего пособия.

Пусть задана пара вероятностных распределений $\mu$ и $\nu$ на пространствах $X = Y = \mathbb{R}^d$. Решением задачи Монжа—Канторовича называется распределение $m$ на $\mathbb{R}^{2d}$, удовлетворяющее следующим условиям:

1. Проекции $m$ на $X$ и $Y$ равны соответственно $\mu$ и $\nu$:

$$\mathrm{pr}_X m = \mu, \quad \mathrm{pr}_Y m = \nu. \tag{3}$$

2. Распределение $m$ реализует минимум следующего функционала:

$$\mathcal{F}(m)\colon m \to \int\limits_{X\times Y} c(x, y)\, dm,$$

где $c\colon X \times Y \to \mathbb{R}$ — некоторая функция, называемая функцией стоимости (cost function).

При весьма общих предположениях задача Монжа—Канторовича имеет решение.

В дальнейшем мы будем интересоваться только случаем $c(x, y) = |x - y|^2$.

Обратим теперь внимание на важное отличие задачи Монжа—Канторовича от исходной задачи Монжа. В задаче Монжа речь идет о *перевозе* груза, что на математическом языке соответствует задаче существования *отображения* $T\colon X \to Y$, преобразующего распределение $\mu$ в распределение $\nu$ (последнее означает, что $\nu(A) = \mu(\{x\colon T(x) \in A\})$ и реализующего



минимум функционала

$$W_2^2(\mu, \nu) = \int\limits_{\mathbb{R}^d} |x - T(x)|^2\, d\mu.$$

Оказывается, что при весьма общих условиях (например, распределения $\mu$ и $\nu$ непрерывны) эти задачи эквивалентны. Если $m$ — решение задачи Монжа—Канторовича, то $m$ сосредоточено на графике некоторого отображения $T$: $m\{(x, y)\colon y = T(x)\} = 1$. Мы будем называть $T$ оптимальным отображением. Существование $T$ было доказано Я. Бренье в [5]. Более того, имеет место следующий удивительный факт.

**Теорема 1.** *$T$ имеет вид*

$$T(x) = \nabla\varphi(x),$$

*где $\varphi$ — некоторая выпуклая функция.*

Величина

$$W_2(\mu, \nu) = \sqrt{\int\limits_{\mathbb{R}^d} |x - T(x)|^2\, d\mu}$$

называется расстоянием Канторовича (также можно встретить названия «расстояние Канторовича—Рубинштейна» и «расстояние Васерштейна»). Действительно, можно проверить, что $W_2(\mu, \nu)$ является расстоянием на пространстве вероятностных распределений.

Пусть теперь распределения $\mu$ и $\nu$ заданы плотностями $\mu = \rho_1\, dx$, $\nu = \rho_2\, dx$. Свойство $T$ отображать $\mu$ в $\nu$ аналитически записывается с помощью формулы замены переменной:

$$\rho_2(\nabla\varphi)\det D^2\varphi = \rho_1.$$

Если рассматривать $\varphi$ как неизвестную функцию, то мы получаем *уравнение Монжа—Ампера*. Под $D^2\varphi = D(\nabla\varphi)$ подразумевается матрица вторых производных (гессиан) функции $\varphi$.

Сделаем важное техническое замечание. Выполнено очевидное тождество

$$\int\limits_{\mathbb{R}^d} |x - y|^2\, dm = \int\limits_{\mathbb{R}^d} |x|^2\, d\mu - 2\int\limits_{\mathbb{R}^d} \langle x, y\rangle\, dm + \int\limits_{\mathbb{R}^d} |y|^2\, d\nu.$$

Поэтому поиск минимума функционала $\int\limits_{\mathbb{R}^d} |x - y|^2\, dm$ эквивалентен поиску максимума функционала $\int\limits_{\mathbb{R}^d} \langle x, y\rangle\, dm$.

Существование $\varphi$ может быть доказано разными способами. Стандартный подход состоит в применении метода двойственности Канторовича и работе с так называемыми циклически монотонными множествами. При

этом выпуклость $\varphi$ получается автоматически. Двойственная задача Канторовича принимает вид

$$\int \Phi(x)\,d\mu + \int \Psi(y)\,d\nu \to \max,$$

где функционал максимизируется среди функций, удовлетворяющих условию $\Phi(x) + \Psi(y) \leqslant |x-y|^2$. Отображение $T$ связано с $\Phi$ следующим образом: $T(x) = x - \nabla\Phi(x)$ (см. подробнее [16, гл. 1]).

Формальное, но поучительное доказательство того факта, что $T$ является градиентом, можно получить путем вывода уравнения Эйлера—Лагранжа (см. [6]). Пусть $T$ — произвольное отображение из $\mu$ в $\nu$. Решение задачи Монжа—Канторовича можно искать как условный экстремум функционала

$$\int\limits_{\mathbb{R}^d} \langle T(x),\, x\rangle\,d\mu$$

при условии $\rho_\nu(T)\det DT = \rho_\mu$. Составим функционал Лагранжа:

$$\int\limits_{\mathbb{R}^d} \big(\langle T(x),\, x\rangle\rho_\mu + \lambda(x)(\rho_\nu(T)\det DT - \rho_\mu)\big)\,dx.$$

Функция $\lambda$ играет роль множителя Лагранжа. Чтобы найти первую вариацию функционала Лагранжа, рассмотрим инфинитезимальную вариацию

$$T_\varepsilon(x) = T(x) + \varepsilon\omega(x)$$

отображения $T$. Здесь $\omega$ — гладкое векторное поле с компактным носителем. Очевидно,

$$\rho_\nu(T_\varepsilon) \approx \rho_\nu(T) + \varepsilon\langle\omega,\, \nabla\rho_\nu(T)\rangle.$$

Можно проверить, что

$$\begin{aligned}\det(DT + \varepsilon D\omega) &= \det DT\cdot\det(I + \varepsilon(DT)^{-1}D\omega) \approx\\ &\approx \det DT\big(1 + \varepsilon\operatorname{Tr}\big[DT^{-1}\cdot D\omega\big]\big).\end{aligned}$$

Таким образом, первая вариация функционала Лагранжа равна

$$\int\limits_{\mathbb{R}^d}\left(\langle\omega(x),\, x\rangle\rho_\mu + \lambda(x)\cdot\rho_\mu(T)\operatorname{Tr}[DT^{-1}D\omega] + \lambda(x)\langle\nabla\rho_\nu(T),\, \omega\rangle\frac{\rho_\mu}{\rho_\nu(T)}\right)dx.$$

Заметим, что $\operatorname{div}(\omega(T^{-1})) = \operatorname{Tr} D\big[\omega(T^{-1})\big] = \operatorname{Tr}\big[DT^{-1}D\omega\big](T^{-1})$.

Интегрируя по частям и применяя замену переменных, несложно убедиться в том, что

$$\int\limits_{\mathbb{R}^d}\lambda(x)\cdot\operatorname{Tr}[DT^{-1}\cdot D\omega]\rho_\mu\,dx = \int\limits_{\mathbb{R}^d}\lambda(T^{-1})\operatorname{div}(\omega(T^{-1}))\rho_\nu\,dx =$$

$$= -\int\limits_{\mathbb{R}^d}\big\langle\nabla[\lambda(T^{-1})],\,\omega(T^{-1})\big\rangle\rho_\nu\,dx - \int\limits_{\mathbb{R}^d}\lambda(T^{-1})\Big\langle\omega(T^{-1}),\,\frac{\nabla\rho_\nu}{\rho_\nu}\Big\rangle\rho_\nu\,dx.$$



Следовательно, вариация равна

$$\int\limits_{\mathbb{R}^d}\langle\omega(x),\,x\rangle\rho_\mu\,dx - \int\limits_{\mathbb{R}^d}\langle\nabla\big[\lambda(T^{-1})\big],\,\omega(T^{-1})\rangle\rho_\nu\,dx = 0.$$

Пусть $\lambda = u(T)$, где $u$ — некоторая функция. Применяя опять формулу замены переменных, получаем, что для любого гладкого поля $\omega$ выполнено

$$\int\limits_{\mathbb{R}^d}(\langle\omega(x),\,x\rangle - \langle\nabla u(T),\,\omega(x)\rangle)\rho_\mu\,dx = 0.$$

Следовательно:

$$\nabla u(T) = x \implies T^{-1} = \nabla u.$$

Таким образом, $T^{-1} = \nabla u$. В силу симметричности задачи относительно $\mu$ и $\nu$ то же утверждение можно сделать для самого отображения $T$.

В качестве иллюстрации эффективного использования оптимальной транспортировки в анализе приведем доказательство М. Громова классического изопериметрического неравенства.

**Пример 1.** Среди множеств фиксированной меры Лебега шары имеют наименьшую поверхностную меру.

Доказательство. Пусть $A\subset\mathbb{R}^d$ — борелевское множество, $B_r = \{x\colon |x|\leqslant r\}$ — шар, удовлетворяющий условию $\lambda(A) = \lambda(B_r)$, где $\lambda$ — мера Лебега на $\mathbb{R}^d$. Пусть $T = \nabla\varphi$ — оптимальная транспортировка, отображающая $\lambda|_A$ в $\lambda|_{B_r}$. По формуле замены переменных $\det D^2\varphi = 1$ на $A$ (для простоты изложения считаем, что $\varphi$ — гладкая функция, хотя аргументы ниже легко обобщаются на негладкий случай). Матрица $D^2\varphi$ симметрична и неотрицательна, поэтому $1 = \sqrt[d]{\det D^2\varphi} \leqslant \dfrac{\Delta\varphi}{d}$ в силу неравенства между средним арифметическим и средним геометрическим. Проинтегрируем это неравенство по множеству $A$ и применим теорему Остроградского—Гаусса:

$$d\lambda(A) \leqslant \int\limits_A \Delta\varphi\,dx = \int\limits_{\partial A}\langle\nabla\varphi,\,n_A\rangle\,d\mathcal{H}^{d-1} \leqslant r\mathcal{H}^{d-1}(\partial A).$$

Здесь $n_A$ — единичная нормаль к $\partial A$, $\mathcal{H}^{d-1}$ — $(d-1)$-мерная мера Хаусдорфа. Из соотношения $\lambda(A) = \lambda(B_r) = \dfrac{\pi^{\frac{d}{2}}}{\Gamma\left(1+\frac{d}{2}\right)}r^d$ получаем классическое изопериметрическое неравенство

$$\lambda^{1-\frac{1}{d}}(A) \leqslant \varkappa_d\mathcal{H}^{d-1}(\partial A),$$

где $\varkappa_d = \dfrac{\left[\Gamma\left(1+\frac{d}{2}\right)\right]^{\frac{1}{d}}}{d\sqrt{\pi}}$. Из доказательства следует, что неравенства становятся равенствами в случае $A = \{\|x - x_0\| \leqslant r\}$. Таким образом, шары

имеют наименьшую поверхностную меру среди множеств фиксированной меры Лебега. □

Пусть $\nu$ — некоторое вероятностное распределение. Энтропией вероятностного распределения $g \cdot \nu$ относительно $\nu$ называется величина $\mathrm{Ent}_\nu(g) = \int g \log g \, d\nu$ (мы считаем, что функция $x \log x$ равна нулю в точке 0).

Следующий результат, доказанный М. Талаграном [15], связывает теорию оптимальной транспортировки с функциональными неравенствами.

**Теорема 2.** *Пусть $\mu = \gamma$ — стандартное гауссовское распределение. Предположим, что $\nu = g \cdot \gamma$ — другое вероятностное распределение. Тогда квадрат расстояния Канторовича между этими распределениями оценивается относительной энтропией $g$:*

$$\frac{1}{2} W_2^2(\gamma, g \cdot \gamma) \leqslant \int\limits_{\mathbb{R}^d} g \cdot \log g \, d\gamma := \mathrm{Ent}_\gamma(g).$$

Доказательство. Для простоты изложения предположим, что $g$ и $\varphi$ — гладкие функции (это бывает не всегда, но общий случай можно свести к этому). Рассмотрим формулу замены переменной:

$$e^{-\frac{x^2}{2}} = g(\nabla\varphi) e^{-\frac{|\nabla\varphi|^2}{2}} \det D^2\varphi.$$

Прологарифмируем это соотношение:

$$-\frac{x^2}{2} = \log g(\nabla\varphi) - \frac{|\nabla\varphi|^2}{2} + \log\det D^2\varphi.$$

Перепишем его в виде

$$\frac{1}{2}|x - \nabla\varphi|^2 = \langle x, x - \nabla\varphi\rangle + \log g(\nabla\varphi) + \log\det D^2\varphi.$$

Проинтегрируем полученное неравенство по $\gamma$. Заметим, что $\nabla e^{-\frac{x^2}{2}} = -xe^{-\frac{x^2}{2}}$. Из формулы интегрирования по частям следует:

$$\int\limits_{\mathbb{R}^d} \langle x, x - \nabla\varphi\rangle \, d\gamma = \int\limits_{\mathbb{R}^d} \left(d - \mathrm{Tr}\, D^2\varphi\right) d\gamma$$

(напомним, что $d$ — размерность). Следовательно,

$$\frac{1}{2}\int\limits_{\mathbb{R}^d} |x - \nabla\varphi|^2 \, d\gamma + \int\limits_{\mathbb{R}^d} (\mathrm{Tr}\, D^2\varphi - d - \log\det D^2\varphi)\, d\gamma \leqslant \int\limits_{\mathbb{R}^d} \log g(\nabla\varphi)\, d\gamma.$$

Заметим теперь, что

$$\mathrm{Tr}\, D^2\varphi - d - \log\det D^2\varphi \geqslant 0.$$



Действительно, если $\lambda_i$ — собственные значения матрицы $D^2\varphi$, то

$$\mathrm{Tr}\, D^2\varphi - d - \log\det D^2\varphi = \sum_{i=1}^{d} \lambda_i - 1 - \log\det\lambda_i \geqslant 0$$

(в силу неотрицательности функции $x - 1 - \ln x$). Таким образом, получаем

$$\frac{1}{2}\int\limits_{\mathbb{R}^d} |x - \nabla\varphi|^2 d\gamma \leqslant \int\limits_{\mathbb{R}^d} \log g(\nabla\varphi) d\gamma = \int\limits_{\mathbb{R}^d} g \log g \, d\gamma. \quad \square$$

**Теорема 3** (К. Мартон). *Если вероятностное распределение $\mu$ удовлетворяет неравенству Талаграна*:

$$W_2^2(\mu, g \cdot \mu) \leqslant C \int\limits_{\mathbb{R}^d} g \cdot \log g \, d\mu,$$

*то $\mu$ удовлетворяет неравенству гауссовской концентрации*:

$$\mu(A_r) \geqslant 1 - 2e^{-\frac{r^2}{4C}}, \quad \mu(A) \geqslant \frac{1}{2}. \tag{4}$$

*В частности, неравенству гауссовской концентрации удовлетворяет гауссовское распределение.*

Доказательство. Положим $(A^r)^c = \mathbb{R}^d \setminus A^r$. Рассмотрим оптимальную транспортировку $\nabla\varphi$ вероятностного распределения $\mu_1 = \frac{1}{\mu(A)} \times I_A \cdot \mu$ в вероятностное распределение $\mu_2 = \frac{1}{\mu((A^r)^c)} I_{(A^r)^c} \cdot \mu$. В силу того, что расстояние между носителями $\mu_1, \mu_2$ превосходит $r$, имеем $W_2(\mu_1, \mu_2) \geqslant r$. В силу неравенства треугольника (напомним, что $W_2$ — расстояние):

$$r \leqslant W_2(\mu_1, \mu_2) \leqslant W_2(\mu_1, \mu) + W_2(\mu_2, \mu).$$

По неравенству Талаграна

$$r \leqslant \sqrt{2C\mathrm{Ent}_\mu\mu_1} + \sqrt{2C\mathrm{Ent}_\mu\mu_2}.$$

Так как $\mathrm{Ent}_\mu\mu_1 = \log\frac{1}{\mu(A)}$, $\mathrm{Ent}_\mu\mu_2 = \log\frac{1}{\mu((A^r)^c)}$, немедленно получаем

$$\frac{r^2}{4C} \leqslant \log\frac{1}{\mu(A)} + \log\frac{1}{\mu((A^r)^c)}.$$

Следовательно,

$$\mu((A^r)^c) \leqslant 2e^{-\frac{r^2}{4C}}.$$

Остается заметить, что $\mu((A^r)^c) = 1 - \mu(A^r)$. Теорема доказана. □

Несложно проверить, что приведенные выше аргументы применимы к случаю распределения с плотностью $e^{-V}$, где $D^2V \geqslant K \cdot \mathrm{Id}$, $K > 0$ (неравенство понимается в матричном смысле, эквивалентная формулировка: $\langle D^2V \cdot v, v\rangle \geqslant K$ для любого вектора $v \in \mathbb{R}^d$ единичной длины). В этом случае также получаем гауссовскую концентрацию. Те же самые аргументы работают для сферы или, более общим образом, для многообразия с положительным тензором Риччи.

Наконец, если рассмотреть риманово многообразие $M$ с метрикой $g$ с распределением $\mu = e^{-V}\,d\mathrm{vol}$, где vol — риманов объем, то «ответственным» за свойства концентрации $\mu$ является тензор Бакри—Эмери $\mathrm{R} = D^2V + \mathrm{Ric}$, где Ric — тензор Риччи многообразия $M$, а $D^2V$ (гессиан) понимается в смысле ковариантного дифференцирования на $M$. Например, согласно классическому результату Бакри и Эмери, положительность тензора Бакри—Эмери $\mathrm{R} \geqslant K \cdot g$, $K > 0$, влечет логарифмическое неравенство Соболева (см. [16]). Как мы видели, из изопериметрических неравенств следует свойство концентрации. Обратное неверно. Однако согласно совсем недавнему результату Э. Мильмана, в случае неотрицательного тензора Бакри—Эмери свойство концентрации меры оказывается равносильным некоторому изопериметрическому неравенству (см. [17]).

Напоследок кратко обсудим еще одно приложение транспортной задачи — оценку скорости сходимости к равновесному состоянию. Пусть $V$ — равномерно выпуклый потенциал:

$$D^2V \geqslant K \cdot \mathrm{Id}, \quad K > 0.$$

Рассмотрим решение уравнения Фоккера—Планка:

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \mathrm{div}(\nabla V \cdot \mu_t),$$

$\mu_t = \rho_t\,dx$ — вероятностное распределение. Оказывается, для двух решений этого уравнения выполнено неравенство

$$\frac{d}{dt} W_2^2(\mu_t, \nu_t) \leqslant -2K \cdot W_2^2(\mu_t, \nu_t)$$

(см. [16, пример 9.10]). Это можно проверить, непосредственно продифференцировав расстояние Канторовича по параметру $t$. Очевидно, эта оценка дает экспоненциальную скорость сходимости $\mu_t$ к равновесному распределению:

$$W_2(\mu_t, \nu_t) \leqslant W_2(\mu_0, \nu_0) e^{-Kt}.$$

Приложения такого рода включают в себя широкий класс уравнений, являющихся градиентными потоками относительно метрики Канторовича. Подробнее об этом см. в [4].



## Литература


1. *Богачев В. И.* Гауссовские меры. М.: Наука, 1997.
2. *Зорич В. А.* Математический анализ задач естествознания. М.: МЦНМО, 2008.
3. *Канторович Л. В.* О перемещении масс // ДАН СССР. 1942. Т. 37. С. 227–229.
4. *Ambrosio L., Gigli N., Savaré G.* Gradient flows in metric spaces and in the Wasserstein spaces of probability measures. Lectures in Math. ETH Zurich, 2008.
5. *Brenier Y.* Polar factorization and monotone rearrangement of vector valued functions // Comm. Pure Appl. Math. 1991. V. 44. P. 375–417.
6. *Evans L. C.* Partial differential equations and Monge—Kantorovich mass transfer // Current developments in mathematics. Cambridge, 1997; Boston: Int. Press, 1999. P. 65–126; `http://math.berkeley.edu/~evans/`
7. *Gromov M.* Metric structure for Riemannian and non-Riemannian spaces. V. 152. Boston: Birkhäuser, 1998.
8. *Ledoux M.* The concentration of measure phenomenon. Amer. Math. Soc., 2001. (Mathematical Surveys and Monographs. V. 89); `http://www.math.univ-toulouse.fr/ ledoux/challenge.pdf`
9. *Levy P.* Problèmes concrets d'analyse fonctionnelle. Paris: Gauthier-Villars, 1951.
10. *Lugosi G.* Concentration of measures inequalities. Barcelona, 2009; `http://www.econ.upf.edu/~lugosi/anu.pdf`
11. *Marton K.* A measure concentration inequality for contracting Markov chains // Geom. Func. Anal. 1997. V. 6. P. 556–571.
12. *Milman E.* On the role of Convexity in Isoperimetry, Spectral-Gap and Concentration // Invent. Math. 2009. V. 177. N. 1. P. 1–43.
13. *Milman V., Schechtman G.* Asymptotic theory of finite dimensional normed vector spaces. Springer, 1986. (Lect. Notes in Math. V. 1200).
14. *Monge G.* Mémoire sur la théorie des déblais et de remblais. Histoire de l'Académie Royale des science, année 1781. Paris, 1784.
15. *Talagrand M.* Transportation cost for Gaussian and other product measures // Geom. Funct. Anal. 1996. V. 6. P. 587–600.
16. *Villani C.* Topics in Optimal Transportation. Amer. Math. Soc. Providence, RI: 2003; `http://math.univ-lyon1.fr/homes-www/villani/`
17. *Milman E.* Isoperimetric and Concentration Inequalities: Equivalence under Curvature Lower Bound // Duke Math. J. 2010. V. 154, № 2. P. 207–239.
18. *Колесников А. В.* Задача о перемещении массы (Монж—Канторович), концентрация меры, скорость сходимости марковских цепей и оптимальная топология транспортной сети. 2011; `http://ium.mccme.ru/postscript/s12/gasnikov-kolesnikov.pdf`

**Ю. Е. Нестеров, С. В. Шпирко**

# Стохастическое транспортное равновесие

Во многих задачах теории игр представляется естественным ввести неопределенность в процесс принятия решения. Пусть имеется конечный набор из $M$ стратегий. Для каждой стратегии $r$ $(r = 1, \ldots, M)$ известны затраты $c_r$ ее реализации. В отсутствие полной информации игроку известно точное значение $c_r$ с некоторой погрешностью $\varepsilon_r$. Поэтому с каждой стратегией $r$ имеет смысл связать некоторую вероятность $p_r$ выбора данной стратегии. Разумеется, каждая вероятность $p_r$ зависит от распределения $\varepsilon_r$.

В дискретных моделях выбора [1] каждый игрок руководствуется принципом минимизации своих затрат (максимизации функции полезности):

$$p_r = P\{c_r + \varepsilon_r = \min_{1 \leqslant i \leqslant M} (c_i + \varepsilon_i)\}.$$

Как правило, выписать такую зависимость в явной форме бывает трудно. Однако в некоторых случаях это возможно. Так, в логит-модели в качестве распределения $\varepsilon_r$ используют двойное экспоненциальное (распределение Гумбеля). Тогда

$$p_i = e^{-c_i/\mu} \Big/ \sum_{r=1}^{M} e^{-c_r/\mu}, \tag{1}$$

где $\mu > 0$ — параметр распределения.

Введем потенциальную функцию

$$\psi(c) = \mu \ln \left( \sum_{r=1}^{M} e^{-c_r/\mu} \right). \tag{2}$$

Положим $p = (p_1, \ldots, p_M)$. Обозначим через $\nabla\psi(c)$ градиент функции $\psi(\cdot)$. Тогда (1) может быть записано в виде

$$p = -\nabla \psi(c). \tag{3}$$

Заметим, что с вычислительной точки зрения выражения (2), (3) допустимы лишь для достаточного малого числа стратегий $M$. Однако для многих важных классов моделей это не так. В частности, для транспортных задач, когда стратегией игрока является маршрут в сети. С увеличением размера сети количество маршрутов растет с экспоненциальной скоростью. В данной ситуации прямое применение формул (2), (3) становится численно не реализуемым.



Возможно ли разработать простую процедуру вычисления значения потенциальной функции типа (2) и ее градиента для некоторого множества маршрутов в сети?

Представляется интересным расширить аппарат логит-модели на модели стохастического транспортного равновесия. В данной модели фактические временные затраты складываются стихийно как результат того или иного распределения транспортного потока в сети.

Чтобы перейти к формальному описанию нашей модели, разработаем предварительно необходимый аппарат. Рассмотрим сеть $\mathcal{N}$, состоящую из $n$ узлов и $m$ направленных дуг. Обозначим через $\mathcal{A}$ множество всех дуг из $\mathcal{N}$. Для каждой дуги $\alpha \in \mathcal{A}$ введем величину временных затрат $t_\alpha \geqslant 0$. Далее, для маршрута $r$ введем функцию временных затрат

$$c_r(t) = \sum_{\alpha \in r} t_\alpha.$$

Заметим, что $c_r(t)$ линейна по $t$.

Зафиксируем два узла $p$ и $k$ и обозначим через $\mathcal{R}$ множество всех маршрутов в $\mathcal{N}$, соединяющих данные два узла. Мы можем формально ввести следующую характеристическую функцию:

$$g_{\mathcal{R}}(t) = \sum_{r \in \mathcal{R}} e^{-c_r(t)}. \tag{4}$$

Если $\mathcal{R} = \varnothing$, то положим $g_{\mathcal{R}}(t) \equiv 0$.

Заметим, что для любого конечного множества $\mathcal{R}$ данная функция корректно задана и бесконечное число раз непрерывно дифференцируема. Также $g_{\mathcal{R}}(t)$ либо положительна для всех значений $t$, либо тождественно равна нулю.

Для множества маршрутов $\mathcal{R}$ мы можем также формально определить потенциальную функцию

$$\psi_{\mathcal{R}}(t) = \ln g_{\mathcal{R}}(t).$$

Положим $\psi_\varnothing(t) \equiv -\infty$. Заметим, что для $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ при $\mathcal{R}_1 \cap \mathcal{R}_2 = \varnothing$ следует

$$\psi_{\mathcal{R}}(t) = \ln \left( e^{\psi_{\mathcal{R}_1}(t)} + e^{\psi_{\mathcal{R}_2}(t)} \right).$$

Справедлива следующая

**Лемма 1.** *Если $\mathcal{R} \neq \varnothing$, то $\psi_{\mathcal{R}}(t)$ выпукла по $t$.*

Далее, обозначим через $\alpha[i, j]$ дугу, соединяющую два соседних узла $i$ и $j$. Через $\mathcal{R}^l_{p,k}$ обозначим множество всех маршрутов в $\mathcal{N}$, соединяющих

узлы $p$ и $k$, с числом дуг, равным $l$. Заметим, что данные множества не пересекаются и их можно задать рекурсивным образом:

для $l = 1$

$$\mathcal{R}^1_{p,k} = \alpha[p,\, k], \quad p,\, k = 1,\, \ldots,\, n;$$

для $l \geqslant 1$ имеем

$$\mathcal{R}^{l+1}_{p,k} = \bigcup_{i \in I(k)} \bigcup_{r \in \mathcal{R}^l_{p,i}} \{r \cup \alpha[i,\, k]\}, \quad p,\, k = 1.., n, \tag{5}$$

где $I(k) = \{i : \alpha[i,\, k] \neq \varnothing\}$.

Поскольку число элементов в $\mathcal{R}^l_{p,k}$ конечно, то соответствующие характеристические функции корректно заданы. Образуем из них $(n \times n)$-матрицу $G_l(t)$:

$$[G_l(t)]^{k,p} = g_{\mathcal{R}^l_{p,k}}(t), \quad p,\, k = 1,\, \ldots,\, n.$$

Чтобы описать аналитическую структуру данной матрицы, введем $(n \times n)$-матрицу инциденций $E(t)$:

$$[E(t)]^{(i,j)} = \begin{cases} e^{-t_{\alpha[j,i]}}, & \text{если } \alpha[j,\, i] \neq \varnothing, \\ 0, & \text{если } \alpha[j,\, i] = \varnothing, \end{cases} \quad i,\, j = 1,\, \ldots,\, n.$$

**Теорема 1.** *Для любого $l \geqslant 1$ справедливо*

$$G_l(t) = \underbrace{E(t) \cdots E(t)}_{l \text{ раз}}.$$

Доказательство. Зафиксируем произвольный узел $p$ и рассмотрим следующую вектор-функцию

$$a^l_p(t) = G_l(t) e_p,$$

где $e_p$ — $p$-й координатный вектор из $R^n$. Для доказательства теоремы нам достаточно показать, что

$$a^l_p(t) = E^l(t) e_p. \tag{6}$$

Докажем по индукции. Для $l = 1$ с учетом определения матрицы $E^1(t) \equiv \equiv E(t)$, данное утверждение очевидно.

Далее, предположим, что оно справедливо для некоторого $l \geqslant 1$. Тогда с учетом аддитивности $c_r$ и (5) для любого $k = 1,\, \ldots,\, n$ имеем

$$[a^{l+1}_p(t)]^{(k)} = \sum_{r \in \mathcal{R}^{l+1}_{p,k}} e^{-c_r(t)} = \sum_{i \in I(k)} \sum_{r \in \mathcal{R}^l_{p,i}} e^{-t_{\alpha[i,k]} - c_r(t)} =$$

$$= \sum_{i \in I(k)} e^{-t_{\alpha[i,k]}} a^l_{i,p}(t) = \sum_{i \in I(k)} [E(t)]^{(k,i)} \cdot [a^l_p(t)]^{(i)}.$$

То есть $g^{l+1}_p(t) = E(t) g^l_p(t)$ и (6) выполнено. Теорема доказана. □



Рассмотрим теперь два важных класса маршрутов сети.

1) Множество всех маршрутов с равномерно ограниченным числом дуг (*кумулятивный класс*)

$$\widehat{\mathcal{R}}^L_{p,k} = \bigcup_{l=1}^{L} \mathcal{R}^l_{p,k}.$$

Для произвольных двух узлов $p$ и $k$ введем кумулятивную характеристическую функцию порядка $L$

$$g_{\widehat{\mathcal{R}}^L_{p,k}}(t) = \sum_{r \in \widehat{\mathcal{R}}^L_{p,k}} e^{-c_r(t)} = \sum_{l=1}^{L} g_{\mathcal{R}^l_{p,k}}(t). \tag{7}$$

Составим из данных характеристических функций матрицу $\widehat{G}_L(t)$ :

$$[\widehat{G}_L(t)]^{(k,p)} = g_{\widehat{\mathcal{R}}^L_{p,k}(t)}, \quad k,\, p = 1,\, \ldots,\, n.$$

Тогда по теореме 1 данная матрица представима в виде

$$\widehat{G}_L(t) = \sum_{l=1}^{L} E^l(t). \tag{8}$$

2) Множество всевозможных маршрутов для потокообразующей пары (*асимптотический класс*)

$$\widetilde{\mathcal{R}}_{p,k} = \bigcup_{l=1}^{\infty} \mathcal{R}^l_{p,k}$$

По аналогии с кумулятивной функцией введем асимптотическую характеристическую функцию

$$g_{\widetilde{\mathcal{R}}_{p,k}}(t) = \sum_{r \in \widetilde{\mathcal{R}}_{p,k}} e^{-c_r(t)} = \sum_{l=1}^{\infty} g_{\mathcal{R}^l_{p,k}}(t) \tag{9}$$

и матрицу $\widetilde{G}(t)$:

$$[\widetilde{G}(t)]^{(k,p)} = g_{\widetilde{\mathcal{R}}_{p,k}(t)}, \quad k,\, p = 1,\, \ldots,\, n.$$

С учетом теоремы 1 имеем

$$\widetilde{G}(t) = \sum_{l=1}^{\infty} E^l(t). \tag{10}$$

Необходимое и достаточное условие сходимости суммы (10) можно записать в терминах спектрального радиуса матрицы $E(t)$:

$$\rho(t) = \max_{1 \leqslant i \leqslant n} |\lambda_i(E(t))| < 1,$$

где числа $\lambda_i(\cdot)$ являются собственными значениями данной матрицы.

Справедлива

**Теорема 2.** 1. *Область определения матрицы $\tilde{G}(t)$ есть открытое множество*:

$$\operatorname{dom}\tilde{G} = \{t \in R^m : \rho(t) < 1\}.$$

2. *Если $\bar{t} \in \operatorname{dom}\tilde{G}$ и $t \geqslant \bar{t}$, то $t \in \tilde{G}$.*

3. *Для $\forall t \in \operatorname{dom}\tilde{G}$ матрица $\tilde{G}$ представима в виде*:

$$\tilde{G}(t) = (I - E(t))^{-1} - I = E(t)(I - E(t))^{-1} = \lim_{L\to\infty} \hat{G}_L(t). \tag{11}$$

*Более того, для таких $t$ производные по направлению $\{\mathcal{D}\tilde{G}_L(t)[h]\}$ также сходятся*:

$$\mathcal{D}\tilde{G}(t)[h] = \lim_{L\to\infty} \mathcal{D}\hat{G}_L(t)[h] \quad \forall h \in R^m. \tag{12}$$

4. *Для $\forall t \in \operatorname{dom}\tilde{G}$ спектральный радиус матрицы $\tilde{G}$ есть*

$$\tilde{\rho}(t) \equiv \lambda_{\max}(\tilde{G}(t)) = \frac{\rho(t)}{1 - \rho(t)}. \tag{13}$$

Перейдем теперь к описанию нашей модели транспортного равновесия. Зададим три ее необходимые компоненты:

**1. Стратегия водителей.**

В детерминированных транспортных моделях поведение водителей обычно описывается первым принципом Вардропа: *каждый водитель выбирает один из кратчайших путей следования из источника в сток*.

В логит-модели (1) каждый водитель выбирает маршрут $r$ из множества $\mathcal{R}$ возможных маршрутов с вероятностью

$$p_r(t) = e^{-c_r(t)/\mu} \Big/ \sum_{q\in\mathcal{R}} e^{-c_q(t)/\mu}, \quad r \in \mathcal{R}. \tag{14}$$

Рассмотрим транспортный поток из узла $p$ в $k$. Обозначим корреспонденцию для данной пары через $d$. Тогда с учетом (14) ожидаемый поток $f_r$ ($f_r = f_r(t) \in R^m$) по маршруту $r$ есть

$$f_r(t) = d \cdot e^{-c_r(t)/\mu} \Big/ \sum_{q\in\mathcal{R}} e^{-c_q(t)/\mu}, \quad r \in \mathcal{R}. \tag{15}$$

Рассмотрим для каждого маршрута $r \in \mathcal{R}$ вектор инциденций $a_r \in \{0, 1\}^m$. Тогда ожидаемый поток по дугам $f(t) \in R^m$ можно записать в виде

$$f(t) = \sum_{r\in\mathcal{R}} f_r(t) \cdot a_r. \tag{16}$$



Оказывается, (16) можно записать в терминах потенциальной функции

$$\psi_{\mathcal{R}}(t) = \ln\left(\sum_{r\in\mathcal{R}} e^{-c_r(t)}\right). \tag{17}$$

Справедлива следующая

**Лемма 2.** *Для любого $\mu > 0$ и $t \in R^m$: $t/\mu \in \operatorname{int}(\operatorname{dom}\psi_{\mathcal{R}})$ выполняется*

$$f(t) = -d \cdot \bigtriangledown\psi_{\mathcal{R}}\left(\frac{t}{\mu}\right). \tag{18}$$

Доказательство. По определению $c_r(t) = \langle a_r, t\rangle$. Следовательно, $\psi_{\mathcal{R}}(t) = \ln\left(\sum\limits_{r\in\mathcal{R}} e^{-\langle a_r,t\rangle}\right)$. Для доказательства (18) остается просто продифференцировать данную функцию. □

Таким образом, в случае логит-модели приходим к следующему принципу:

*Зафиксируем вектор $t$ временных затрат на дугах. Обозначим через $\mathcal{R}$ множество возможных маршрутов для некоторой потокообразующей пары. И пусть $d$ — корреспонденция данной пары. Тогда ожидаемый поток для пары будет определяться как*

$$f(t) = -d \cdot \bigtriangledown\psi_{\mathcal{R}}\left(\frac{t}{\mu}\right). \tag{19}$$

**2. Реализация транспортной сети.**

Обычно в моделях транспортного равновесия рассматривают временные затраты на дугах $tt$ как функции от потока: $tt = tt(f)$ [2]. Для существования равновесия в таких случаях необходимо, чтобы $tt(f)$ монотонно не убывала. В работах [3, 4] была предложена так называемая *модель стационарной динамики*.

В данной модели каждая дуга $\alpha \in \mathcal{A}$ характеризуется величиной максимального потока $\bar{f}_\alpha$ и минимальными временными затратами $\bar{t}_\alpha$. Данные величины удовлетворяют следующему предположению:

$$\text{Если } f_\alpha < \bar{f}_\alpha, \text{ то } tt_\alpha = \bar{t}_\alpha. \text{ Если } f_\alpha = \bar{f}_\alpha, \text{ то } tt_\alpha \geqslant \bar{t}_\alpha. \tag{20}$$

**3. Задание загрузки в сети.**

Обозначим через $\mathcal{OD}$ множество всех потокообразующих пар. Для каждой такой пары $(p, k) \in \mathcal{OD}$ зафиксируем величину $d_{p,k}$ корреспонденции (задана матрица корреспонденций). Обозначим через $\mathcal{R}_{p,k}$ некоторое множество маршрутов, соединяющих $p$ с $k$. Покажем, что точка стохастического транспортного равновесия может быть найдена как решение задачи выпуклой оптимизации:

$$\min\left\{\langle\bar{f}, t\rangle + \mu\psi\left(\frac{t}{\mu}\right) : t \geqslant \bar{t}\right\}, \tag{21}$$

где $\mu > 0$ и

$$\psi(t) = \sum_{(p,k)\in\mathcal{OD}} d_{p,k} \cdot \psi_{\mathcal{R}_{p,k}}(t).$$

**Теорема 3.** *Пусть* $\bar{t}/\mu \in \operatorname{dom}\psi$.

1. *Пусть корреспонденция реализована такими потоками* $f_{p,k}$, *что*

$$\sum_{(p,k)\in\mathcal{OD}} f_{p,k} < \bar{f}.$$

*Тогда задача* (21) *разрешима*.

2. *Пусть* $t^*$ *— решение* (21). *Тогда равновесные потоки удовлетворяют соотношению*

$$f^*_{p,k} = -d_{p,k} \cdot \nabla\psi_{\mathcal{R}_{p,k}}\left(\frac{t^*}{\mu}\right), \quad (p,\, k) \in \mathcal{OD}.$$

*Равновесные потоки удовлетворяют соответствующим корреспонденциям*.

3. *Равновесный поток* $f^* = \sum_{(p,k)\in\mathcal{OD}} f^*_{p,k}$ *по дуге* $\alpha$ *не превосходит* $\bar{f}_\alpha$. *Более того*, *равновесная пара* $(t^*,\, f^*)$ *удовлетворяет предположению* (20).

Доказательство. Наметим вкратце доказательство каждого из утверждений теоремы.

1. По предположению, для каждой пары $(p,\, k)$ найдется подмножество $\mathcal{R}'_{p,k} \subseteq \mathcal{R}_{p,k}$ и совокупность положительных чисел $\{\beta^r_{p,k}\}_{r\in\mathcal{R}'_{p,k}}$:

$$\sum_{r\in\mathcal{R}'_{p,k}} \beta^r_{p,k} = d_{p,k}, \quad (p,\, k) \in \mathcal{OD},$$

для которых

$$f_{p,k} = \sum_{r\in\mathcal{R}'_{p,k}} \beta^r_{p,k} a_r$$

и

$$\hat{f} \equiv \sum_{(p,k)\in\mathcal{OD}} f_{p,k} < \bar{f}.$$

Представим целевую функцию (21) в виде

$$\langle \bar{f},\, t\rangle + \mu\psi\left(\frac{t}{\mu}\right) = \langle \bar{f} - \hat{f},\, t\rangle + \langle \hat{f},\, t\rangle + \mu\psi\left(\frac{t}{\mu}\right).$$

В данном выражении рассмотрим два последних слагаемых:

$$\langle \hat{f},\, t\rangle + \mu\psi\left(\frac{t}{\mu}\right) = \langle \hat{f},\, t\rangle + \sum_{(p,k)\in\mathcal{OD}} d_{p,k} \cdot \mu\psi_{\mathcal{R}_{p,k}}\left(\frac{t}{\mu}\right).$$



Можно доказать,что данная сумма ограничена снизу некоторой константой $\mu\gamma$:

$$\langle \hat{f},\, t\rangle + \mu\psi\left(\frac{t}{\mu}\right) \geqslant \mu\gamma.$$

Таким образом, целевая функция (21) ограничена снизу линейной функцией с положительными коэффициентами:

$$\langle \bar{f},\, t\rangle + \mu\psi\left(\frac{t}{\mu}\right) \geqslant \langle \bar{f} - \hat{f},\, t\rangle + \mu\gamma, \quad t \geqslant \bar{t}.$$

Следовательно, множество уровней целевой функции ограничено и задача (21) имеет решение.

2. По предположению, $\bar{t}/\mu \in \operatorname{dom}\operatorname{int}(\psi_{\mathcal{R}_{p,k}})$. Следовательно, градиент $\nabla\psi_{\mathcal{R}_{p,k}}(t/\mu)$ корректно задан для любого $t/\mu \geqslant \bar{t}/\mu$. Остается применить лемму 2.

3. Выпишем для решения задачи (21) условия Куна—Такера:

$$\bar{f} + \nabla\psi\left(\frac{t^*}{\mu}\right) = s^*,$$

$$s^*_\alpha \cdot (t^*_\alpha - t_\alpha) = 0, \quad \text{где} \quad s^* \geqslant 0.$$

Обозначим $f^* = \bar{f} - s^*$. Если $f^*_\alpha < \bar{f}_\alpha$, то $s^*_\alpha > 0$ и мы приходим к (20). Теорема доказана. □

Чтобы численно решать задачу минимизации (21), необходимо найти эффективный способ вычисления значения соответствующей целевой функции и ее градиента. Найдем такой способ для двух рассматриваемых классов множеств маршрутов: асимптотического и кумулятивного.

1. *Асимптотический класс*.

Пусть $\mathcal{R}_{p,k} = \bar{\mathcal{R}}_{p,k}$. Предположим, что мы имеем матрицу корреспонденций $D \in R^{n\times n}$. Будем обозначать через $\langle\cdot,\cdot\rangle_M$ скалярное произведение двух матриц, стоящих в обеих частях: $\langle X,\, Y\rangle_M = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} X_{ij}Y_{ij}$, $X,\, Y \in R^{m\times n}$.

Тогда задачу (21) можно записать:

$$\min\{\langle \bar{f},\, t\rangle + \langle D,\, \mu\ln\{(I - E(t/\mu))^{-1} - I\}\rangle_M :\ t \geqslant \bar{t}\}, \tag{22}$$

где через $\ln(\cdot)$ обозначена $(n\times n)$-матрица с логарифмом от своих компонент.

Обозначим через $F(t)$ нетривиальную часть данного выражения:

$$F(t) = \langle D,\, \mu\ln\{(I - E(t/\mu))^{-1} - I\}\rangle_M. \tag{23}$$

Выберем произвольное направление $h \in R^m$ и вычислим вдоль него частную производную:

$$\mathcal{D}F(t)[h] = \langle D,\, \mathcal{D}(\mu\ln\{(I - E(t/\mu))^{-1} - I\})\rangle_M$$

Обозначим через $B(t)[h]$ матрицу $n \times n$:

$$B(t)[h]^{(i,j)} = \begin{cases} e^{-t_{\alpha[j,i]}} h_{\alpha[j,i]}, & \text{если } \alpha[j,\, i] \neq \varnothing, \\ 0, & \text{если } \alpha[j,\, i] = \varnothing. \end{cases}$$

Для двух матриц $n \times n$ $B$ и $C$ обозначим через $\{B/C\}$ следующую матрицу:

$$\{B/C\}^{(i,j)} = B^{(i,j)}/C^{(i,j)}, \quad i,\, j = 1,\, \ldots,\, n.$$

Можно показать, что в терминах $B(t)[h]$ частная производная по направлению представима в виде

$$\mathcal{D}F(t)[h] = -\langle \{D/(I - E(t))^{-1}\},\, (I - E(t))^{-1} B(t)[h] (I - E(t))^{-1} \rangle_M.$$

Обозначим через $B^*[t][Y]$ оператор, сопряженный к $B(t)[h]$:

$$\langle B(t)[h],\, Y \rangle_M = \langle B^*(t)[Y],\, h \rangle \quad \forall h \in R^m, \quad \forall Y \in R^{n \times n}.$$

Тогда с учетом $\mathcal{D}F(t)[h] = \langle \bigtriangledown F(t),\, h \rangle$ получаем выражение для градиента целевой функции

$$\bigtriangledown F(t) = -B^*(t)[(I - E(t))^{-T} \{D/(I - E(t))^{-1}\} (I - E(t))^{-T}].$$

Таким образом, чтобы вычислить градиент, нам необходимо вычислить обратную матрицу к $(I - E(t))$. В худшем случае это потребует $O(n^3)$ арифметических операций.

2. *Кумулятивный класс.*

Пусть $\mathcal{R}_{p,k} = \widehat{\mathcal{R}}^L_{p,k}$. Зафиксируем узел $p$ и обозначим

$$\left.\begin{aligned} a^l_k(t) &= \mu \psi_{\mathcal{R}^l_{p,k}}(t/\mu), \\ b^l_k(t) &= \mu \psi_{\widehat{\mathcal{R}}^l_{p,k}}(t/\mu), \end{aligned}\right\} \quad k = 1,\, \ldots,\, n, \quad l = 1,\, \ldots,\, L.$$

Отметим, что некоторые из этих функций могут быть равны $-\infty$. Это означает, что соответствующее множество маршрутов пустое.

Оказывается, что данные функции можно вычислить рекурсивным образом:

для $l = 1$

$$a^1_k(t) = b^1_k(t) = \begin{cases} -t_{\alpha[p,k]}, & \text{если } \alpha[p,\, k] \neq \varnothing, \quad k = 1,\, \ldots,\, n; \\ -\infty, & \text{если } \alpha[p,\, k] = \varnothing; \end{cases}$$

для $l = 1,\, \ldots,\, L - 1$ имеем

$$\left.\begin{aligned} a^{l+1}_k(t) &= \mu \ln \Big( \sum_{i \in I(k)} e^{(a^l_i(t) - t_{\alpha[i,k]})/\mu} \Big), \\ b^{l+1}_k(t) &= \mu \ln (e^{b^l_k(t)/\mu} + e^{a^{l+1}_k(t)/\mu}), \end{aligned}\right\} \quad k = 1,\, \ldots,\, n. \tag{24}$$



На каждом шаге $l$ необходимо сделать $O(m)$ арифметических операций. Следовательно, для вычисления всех функций $b^L_k(t)$, $k = 1,\, \ldots,\, n$, необходимо $O(Lm)$ операций.

Интересно отметить, что при $\mu \to 0$ процесс (24) превращается в известный алгоритм кратчайшего пути Беллмана—Форда [6].

На практике очень трудно получить полную матрицу корреспонденций, необходимую для (21). Так что приходится исходить из предположения, что у нас есть лишь неполная информация, например, мы можем знать величину всех корреспонденций

$$\Phi = \sum_{(p,k) \in \mathcal{OD}} d_{p,k}.$$

Введем функцию

$$\theta_{\mathcal{R}_{p,k}} = -\mu \psi_{\mathcal{R}_{p,k}} \Big( \frac{t}{\mu} \Big).$$

Заметим, что если водители используют при выборе маршрута логит-модель, то данная функция есть ожидаемые минимальные затраты [1].

Введем два весовых вектора $P$ и $Q$ из $R^n$ с координатами

$$P^i > 0 \quad \forall i \in \mathcal{O} \quad \text{и} \quad P^i = 0 \quad \text{иначе},$$
$$Q^j > 0 \quad \forall j \in \mathcal{D} \quad \text{и} \quad Q^j = 0 \quad \text{иначе}.$$

В качестве весов $P$ и $Q$ можно взять величину числа жителей в узле-источнике и количества рабочих мест в узле-стоке.

Введем сначала стохастическую корреспонденцию. Будем считать, что водитель, следующий из $i$ в $k$, появляется в сети с вероятностью

$$\pi_{i,k}(t) = \frac{P_i Q_k e^{-\theta_{\mathcal{R}_{i,k}}(t)}}{\sum_{(l,j) \in \mathcal{OD}} P_l Q_j e^{-\theta_{\mathcal{R}_{l,j}}(t)}}. \tag{25}$$

Тогда ожидаемая корреспонденция для такой пары есть $\Phi \cdot \pi_{i,k}(t)$. Ожидаемый поток, с учетом леммы 2, будет

$$f_{i,k}(t) = -\Phi \pi_{i,k}(t) \bigtriangledown \psi_{\mathcal{R}_{i,k}} \Big( \frac{t}{\mu} \Big). \tag{26}$$

Отсюда общий поток в сети будет

$$f(t) = \sum_{(i,k) \in \mathcal{OD}} f_{i,k}(t) = -\Phi \sum_{(i,k) \in \mathcal{OD}} \pi_{i,k}(t) \bigtriangledown \psi_{\mathcal{R}_{i,k}} \Big( \frac{t}{\mu} \Big). \tag{27}$$

Задача поиска стохастического транспортного равновесия формулируется следующим образом:

*Найти такие векторы $t^*$ и $f^*$, для которых выполнено предположение* (20).

Оказывается, такое решение может быть найдено из следующей задачи оптимизации:

$$\min_{t \geqslant \bar{t}} \left[ \langle \bar{f},\, t \rangle + \Phi \cdot \mu \psi\left(\frac{t}{\mu}\right) \right], \tag{28}$$

где $\psi(t) \equiv \psi(P,\, Q,\, t) = \ln\left( \sum\limits_{i \in \mathcal{O}} \sum\limits_{j \in \mathcal{D}} P^{(i)} Q^{(j)} g_{\mathcal{R}_{i,j}}(t) \right)$.

В настоящей работе обсуждались основные результаты статьи [5].

**Литература**


1. *Anderson S. P.*, *de Palma A*, *Thisse J.-F.* Discrete choice theory of product differentiation. Cambridge: MIT Press, 1992.
2. *Sheffi Y.* Urban Transportation networks: Equilibrium analysis with mathematical programming methods. Englewood Cliffs: Prentice-Hall, 1985.
3. *Nesterov Y.* Stable traffic equilibria: properties and applications // Optim. Engineering. 2000. V. 1. P. 29–50.
4. *Nesterov Y.*, *de Palma A.* Static equilibrium in congested transportation networks // Networks and Spatial Economics. 2003. V. 3. P. 371–395.
5. *Nesterov Y.* Characteristic functions of directed graphs and applications to stochastic equilibrium problems // Optim. Engineering. 2007. V. 8. P. 193–214.
6. *Форд Л.*, *Фалкерсон Д.* Потоки в сетях. М.: Мир, 1966.


**А. М. Райгородский**

# Модели случайных графов и их применения

В приложении дается обзор основных современных направлений в теории случайных графов. Делается акцент на связь моделей случайного графа с транспортной проблематикой.

## 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.

Работа выполнена при финансовой поддержке гранта РФФИ 12-01-00683, гранта поддержки ведущих научных школ НШ-2519.2012.1.

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.

## 1. Модель Эрдёша—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. Эрдёш и А. Реньи (см. [1–3]). Отметим, что Эрдёш — 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.

### 1.1. Формальное описание модели

Пусть дано множество $V_n = \{1, \ldots, n\}$, 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 Эрдёша—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

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n,p}(G) = p^{|E|}(1-p)^{C_n^2 - |E|}. \qquad (1)$$

Здесь через $|A|$ обозначена мощность множества $A$ (количество 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 связен, — это величина, равная сумме вероятностей всех связных графов (на фиксированном множестве вершин). Казалось бы, все совсем просто и мы вряд ли имеем шансы наткнуться здесь на нечто особенно интересное. Однако дело обстоит прямо противоположным образом: специфика вероятностных методов, которые эффективно работают в задачах о случайных графах, позволит нам пронаблюдать весьма нетривиальные и, главное, неожиданные явления, которые возникают даже в этой простой модели и которые влекут приятные следствия для приложений.

Прежде чем двигаться дальше, сделаем еще ряд полезных замечаний. Во-первых, если $p = 1/2$, 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 $p = 1/2$ (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 $n \to \infty$.

### 1.2. Транспортная интерпретация модели

Представим себе, что в некоторой стране есть 10 городов, которые *попарно* соединены дорогами. Это довольно сильное предположение, но пока сохраним его. Допустим, каждая из дорог за определенный срок изнашивается (т. е. становится непроезжей) с известной вероятностью $q$. При этом износ данной дороги никак не зависит от совокупного износа остальных дорог. Спрашивается: какова максимальная вероятность $q$, при которой с вероятностью больше $1/2$ не исчезнет возможность перемещения между любыми двумя городами? По существу, это вопрос о *надежности* транспортной сети: чем выше искомая вероятность $q$, тем, разумеется, сеть надежнее.

Нетрудно видеть, что вопрос о надежности сети — это, в свою очередь, вопрос о *связности* случайного графа. В самом деле, 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 Эрдёша—Реньи $G(n, p)$ вероятность связности графа больше половины (граф скорее связен, 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 п. 1.4 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 п. 1.3 мы приведем модификацию модели Эрдёша—Реньи, в рамках которой это предположение можно будет адекватно ослабить.

### 1.3. Обобщения модели Эрдёша—Реньи

Пусть по-прежнему $V_n = \{1, \ldots, n\}$. Однако теперь вероятность ребра между вершинами $i$ и $j$ мы обозначим через $p_{ij}$. Иными словами, мы, как и раньше, проводим ребра независимо друг от друга, но с *разными* вероятностями. В формате аксиоматики Колмогорова мы получаем вероятностное пространство

$$G(n, p_{ij}) = (\Omega_n, \mathcal{F}_n, P_{n,p_{ij}}),$$

в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n,p_{ij}}(G) = \prod_{(i,j)\in E} p_{ij} \cdot \prod_{(i,j)\notin E} (1 - p_{ij}).$$

Важный частный случай описанного пространства получается, коль скоро мы фиксируем некоторый граф $H_n = (V_n, E_n)$ и полагаем

$$p_{ij} = \begin{cases} p, & (i, j) \in E_n, \\ 0, & (i, j) \notin E_n. \end{cases}$$

Иначе говоря, ребра графа $H_n$ возникают в 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.

### 1.4. Некоторые математические результаты о надежности сети

Прежде всего справедлива следующая теорема Эрдёша—Реньи.

**Теорема 1.** *Рассмотрим модель $G(n, p)$. Пусть $p = \dfrac{c \ln n}{n}$. Если $c > 1$, то почти всегда случайный граф связен. Если $c < 1$, то почти всегда случайный граф не является связным.*

Этот довольно простой с точки зрения доказательства факт мы обоснуем в п. 1.5. 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 $p = c \ln n/n$ 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, по 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 $c_1 g \leqslant f \leqslant c_2 g$). 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 п. 1.5, можно доказать, например, такой факт.

**Теорема 1′.** *Рассмотрим модель $G(n, p)$. Пусть $p = \dfrac{c \ln n}{n}$. Если $c > 3$, то при $n > 100$*

$$P_{n,p}(G \text{ связен}) \geqslant 1 - \frac{1}{n}.$$

Этот результат совсем замечателен своей конкретностью. Получается, что при той же тысяче городов и вероятности износа дороги $1 - \dfrac{3 \ln 1000}{1000} \approx 0{,}98$ 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 $p(n) = \ln n/n$ 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 Эрдёшем и Реньи (см. [1–3]).

**Теорема 2.** *Рассмотрим модель* $G(n, p)$. *Пусть* $p = \dfrac{c}{n}$. *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* $\geqslant \gamma n$.

И снова мы имеем фазовый переход — резкое изменение свойств случайного графа при преодолении некоторого порога. В данном случае порогом служит функция $p = 1/n$. Оказывается, что если вероятность ребра в $\dfrac{1}{c} > 1$ 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 $\geqslant \gamma n$, но и асимптотика $\sim \gamma n$. А. В. Е. Степанов доказал, что, опять же при $p = c/n$, $c > 1$, 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 $p \ll 1/n$) имеет место «феодализм» — весь граф поделен на несвязанные между собой логарифмические кусочки. Затем (при $p \gg 1/n$) возникает «империя» — гигантская компонента. Наконец, при $p \gg \ln n/n$ 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, т. е. не лишенной инфраструктуры — возможности сообщения между любыми двумя городами.



Глубокий интерес представляет, конечно, устройство мира «внутри» фазовых переходов, т. е. при $p \sim 1/n$ и при $p \sim \ln n/n$. 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* $p = \dfrac{\ln n + c + o(1)}{n}$. *Тогда*

$$P_{n,p}(G \text{ связен}) \to e^{-e^{-c}}, \quad n \to \infty.$$

*В частности, при* $p = \dfrac{\ln n}{n}$ *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. [10]).

**Теорема 4.** *Пусть* $\{H_n\}$ — *последовательность графов, реберная связность которых стремится к бесконечности при* $n \to \infty$. *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* $p_1 \leqslant c_1 p$, *где* $c_1 < 1$, *случайный граф в модели* $G(H_n, p_1)$ *почти всегда не связен, но для любой функции* $p_2 \geqslant c_2 p$, *где* $c_2 > 1$, *случайный граф в модели* $G(H_n, p_2)$ *почти всегда связен. 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, повязанная 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 пункте мы докажем теорему 1, а в п. 1.6 мы обсудим основные идеи доказательства теоремы 2. Отметим, что дополнительную информацию о поведении случайных графов в модели Эрдёша—Реньи можно почерпнуть из книг [7–9, 11, 12].

### 1.5. Доказательство теоремы 1

Сперва обсудим случай $c > 1$.

Введем случайную величину на пространстве $G(n, p)$:

$$X_n = X_n(G) = \begin{cases} 0, & \text{если } G \text{ связен,} \\ k, & \text{если у } G \text{ ровно } k \text{ компонент.} \end{cases}$$

Таким образом, $X_n$ принимает неотрицательные целые значения, причем $X_n \neq 1$. Нам нужно показать, что $P_{n,p}(X_n = 0) \to 1$ при $n \to \infty$. Это равносильно асимптотике $P_{n,p}(X_n \geqslant 1) \to 0$. По неравенству Чебышёва $P_{n,p}(X_n \geqslant 1) \leqslant EX_n$, и нам остается обосновать стремление к нулю математического ожидания.

Представим $X_n$ в виде суммы

$$X_n = X_{n,1} + \ldots + X_{n,n-1},$$

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: $K_1, \ldots, K_{C_n^k}$. Тогда, в свою очередь,

$$X_{n,k} = X_{n,k,1} + \ldots + X_{n,k,C_n^k},$$

коль скоро

$$X_{n,k,i} = X_{n,k,i}(G) = \begin{cases} 1, & \text{если } K_i \text{ образует компоненту в } G, \\ 0, & \text{иначе.} \end{cases}$$

В итоге

$$EX_n = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} EX_{n,k,i}.$$

Очевидно,

$$EX_{n,k,i} = P_{n,p}(K_i \text{ образует компоненту в } G) \leqslant$$
$$\leqslant P_{n,p}(\text{из } K_i \text{ в } V_n \setminus K_i \text{ нет ребер в } G).$$

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,

$$EX_n \leqslant \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} (1-p)^{k(n-k)} = \sum_{k=1}^{n-1} C_n^k (1-p)^{k(n-k)}.$$



Последняя сумма симметрична в том смысле, что ее слагаемые при $k$ и $n-k$ равны. Рассмотрим $k = 1$:

$$n(1-p)^{n-1} \leqslant ne^{-p(n-1)} = ne^{-\frac{c(\ln n)(n-1)}{n}} = n\left(\frac{1}{n}\right)^{c(1+o(1))} = o(1),$$

поскольку $c > 1$.

Оставшаяся часть рассуждения состоит в доказательстве того, что слагаемые с $k > 1$ и $k < n-1$ пренебрежимо малы по сравнению с первым слагаемым. Соответствующую выкладку мы пропустим. Если же поверить в ее справедливость, то получится, что вся сумма доминируется первым и последним слагаемыми, а стало быть, и она стремится к нулю.

Теорема 1 для случая $c > 1$ доказана.

Теперь рассмотрим случай $c < 1$. Обозначим через $X_n$ количество изолированных вершин в случайном графе. Запишем

$$X_n = X_{n,1} + \ldots + X_{n,n},$$

где

$$X_{n,k} = X_{n,k}(G) = \begin{cases} 1, & \text{если вершина } k \in V_n \text{ изолированная в } G, \\ 0 & \text{иначе.} \end{cases}$$

Тогда

$$EX_n = EX_{n,1} + \ldots + EX_{n,n}.$$

В свою очередь,

$$EX_{n,k} = P_{n,p}(k \text{ изолированная в } G) = (1-p)^{n-1}.$$

Таким образом,

$$EX_n = n(1-p)^{n-1} = n(1-p)^n(1+o(1)) = (1+o(1))ne^{-c\ln n} = (1+o(1))n^{1-c}.$$

Заметим, что ввиду неравенства $c < 1$ выполнено $MX_n \to \infty$.

Посчитаем дисперсию случайной величины $X_n$:

$$DX_n = EX_n^2 - (EX_n)^2 = E(X_{n,1} + \ldots + X_{n,n})^2 - (EX_n)^2 =$$
$$= EX_{n,1}^2 + \ldots + EX_{n,n}^2 + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2 =$$
$$= EX_{n,1} + \ldots + EX_{n,n} + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2 =$$
$$= EX_n + \sum_{i \neq j} EX_{n,i}X_{n,j} - (EX_n)^2.$$

Далее,

$$EX_{n,i}X_{n,j} = P_{n,p}(i \text{ и } j \text{ изолированы в } G) = (1-p)^{2n-3} = (1+o(1))(1-p)^{2n},$$

т. е.

$$\sum_{i\neq j} EX_{n,i}X_{n,j} = n(n-1)(1+o(1))(1-p)^{2n} = (1+o(1))n^{2-2c} = (1+o(1))(EX_n)^2.$$

В итоге

$$DX_n = EX_n + (1+o(1))(EX_n)^2 - (EX_n)^2 = o\big((EX_n)^2\big).$$

По неравенству Чебышёва

$$P_{n,p}(G \text{ связен}) \leqslant P_{n,p}(X_n = 0) = P_{n,p}(X_n \leqslant 0) = P_{n,p}(-X_n \geqslant 0) = \\ = P_{n,p}(EX_n - X_n \geqslant EX_n) \leqslant \frac{DX_n}{(EX_n)^2} = o(1),$$

и вторая часть теоремы доказана.

### 1.6. Идеи доказательства теоремы 2

Метод, о котором мы будем здесь говорить, восходит к Р. Карпу (см. [13]), и в таком виде он описан в книге [9]. Мы лишь перечислим ниже основные шаги рассуждения.

**1.6.1. Простейший ветвящийся процесс.** Пусть $Z_1, \ldots, Z_t, \ldots$ — независимые пуассоновские величины с одним и тем же средним $\lambda$. Положим

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Представлять себе описанный только что процесс можно так. В начальный момент времени есть одна частица. Затем она приносит $Z_1$ потомков и умирает. Заметим, что она может умереть, даже не принеся потомства, 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* $\lambda \leqslant 1$. *Тогда с вероятностью* 1 *процесс* $Y_t$ *вырождается, т. е.* $P(\exists t\colon Y_t \leqslant 0) = 1$.

**Теорема 6.** *Пусть* $\lambda > 1$. *Пусть* $\gamma \in (0, 1)$ *— единственное решение уравнения* $1 - \gamma = e^{-\lambda\gamma}$. *Тогда процесс* $Y_t$ *вырождается с вероятностью* $1 - \gamma$, *т. е.* $P(\exists t\colon Y_t \leqslant 0) = 1 - \gamma$.

Доказательства теорем 5 и 6 можно найти, например, в [9]. Впрочем, это стандартные факты теории ветвящихся процессов (см. [14]). Забегая вперед, скажем, что величина $\gamma$ 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.



**1.6.2. Ветвящийся процесс на случайном графе.** Пусть дан граф $G = (V, E)$: конкретный граф, не случайный. Зафиксируем какую-нибудь его вершину $v_1 \in V$. Назовем ее «живой», а все остальные вершины — «нейтральными». Выберем среди нейтральных вершин всех соседей вершины $v_1$. После этого объявим вершину $v_1$ «мертвой», ее соседей — живыми, а все остальные вершины — нейтральными. Снова зафиксируем какую-нибудь живую вершину $v_2$. Выберем всех ее соседей среди *нейтральных*. Вершину $v_2$ отправим в царство мертвых, а в живых останутся все, кто был жив, кроме $v_2$, и новорожденные «потомки» вершины $v_2$. Продолжая этот ветвящийся процесс, мы в конце концов получим кладбище вершин и ни одной живой вершины. Кладбище будет 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,

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Разумеется, все введенные величины зависят от графа $G$ и от последовательности выбираемых вершин $v_1, \ldots$ Если граф $G$ посчитать случайным, то при любом выборе вершин $v_1, \ldots$ 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 $v_1, \ldots$; 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. Удается доказать, что $Y_t$ имеет вид

$$Y_t = \xi_t + 1 - t, \quad \xi_t \sim \mathrm{Binom}(n-1,\, 1-(1-p)^t).$$

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.

**1.6.3. Случай $c < 1$.** Положим $t_0 = [\beta \ln n]$, где $\beta = \beta(c)$ — константа, которую мы подберем позднее. Нам хочется доказать, что с большой вероятностью каждая из компонент случайного графа имеет размер не больше $t_0$. Но размер компоненты — это момент вырождения процесса $Y_t$ на случайном графе. Значит, интересующее нас утверждение можно

записать в следующем виде:

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \to 0, \quad n \to \infty.$$

Поскольку

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \leqslant nP_{n,p}(Y_{t_0} > 0),$$

достаточно найти такое $\beta$, при котором

$$P_{n,p}(Y_{t_0} > 0) = o\Big(\frac{1}{n}\Big).$$

Дальнейшие выкладки будут слегка неаккуратными, но при желании их можно сделать безукоризненно строгими. Мы же хотим максимально прояснить основную суть подхода. Итак,

$$P_{n,p}(Y_{t_0} > 0) = P_{n,p}(\xi_{t_0} \geqslant t_0) \approx P_{n,p}(\mathrm{Binom}(n,\, 1-(1-p)^{t_0}) \geqslant t_0) \approx$$

(с учетом асимптотики $1-(1-p)^{t_0} \sim pt_0$)

$$\approx P_{n,p}(\mathrm{Binom}(n,\, pt_0) \geqslant t_0) \approx$$

(с учетом центральной предельной теоремы)

$$\approx \int\limits_{\frac{t_0-npt_0}{\sqrt{npt_0(1-pt_0)}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

Поскольку $c<1$, нижний предел интегрирования имеет порядок $\sqrt{t_0}$. Стало быть, весь интеграл не превосходит величины $e^{-\delta t_0}$. Выберем $\beta$ таким, чтобы $e^{-\delta t_0}$ оказалось меньше, чем $e^{-2\ln n} = \dfrac{1}{n^2}$, и в случае $c<1$ теорема доказана.

**1.6.4. Случай $c > 1$.** Этот случай гораздо сложнее предыдущего. Здесь ветвящийся процесс на графе нужно «запускать» не один раз, а многократно. Только так удается доказать, что почти наверняка хотя бы в одном запуске возникнет гигантская компонента. Подробности можно найти в [9], мы же лишь поясним, откуда в текущей ситуации появляется константа $\gamma$ 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

$$P_{n,p}(Y_t \leqslant 0) \to 0, \quad t \approx \gamma n, \quad n \to \infty.$$

У нас $p = \dfrac{c}{n}$. 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

$$P_{n,p}(Y_t \leqslant 0) \approx P_{n,p}(\mathrm{Binom}(n,\, 1-e^{-c\alpha}) \leqslant \alpha n).$$

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 $\alpha$, 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 $\lambda$ мы заменяем на $c$.

## 2. 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 пункте мы расскажем о результатах Барабаши—Альберт. Во втором пункте мы опишем модель Б. Боллобаша и О. Риордана, которая весьма неплохо ложится на статистики Барабаши—Альберт. В третьем, четвертом и пятом пунктах мы обсудим возможные уточнения модели Боллобаша—Риордана.

### 2.1. Наблюдения Барабаши—Альберт

В своих работах [15–17] 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 интернетской проблематике, так и в многочисленных близких дисциплинах. В действительности, большинство реальных сетей (социальные, биологические, транспортные и пр.) имеют похожую «топологию».

Итак, сперва договоримся о том, что мы понимаем под сетью интернет. Это так называемый *веб-граф*, вершины которого суть какие-либо конкретные структурные единицы в интернете: речь может идти о страницах, сайтах, хостах, владельцах и пр. Для определенности будем считать, что

вершинами веб-графа служат именно сайты. Ребрами же мы будем соединять те вершины, между которыми имеются ссылки. При этом разумно проводить столько ребер между двумя вершинами, сколько есть ссылок между соответствующими сайтами. Более того, ребра естественно считать направленными. Таким образом, веб-граф ориентирован и может иметь кратные ребра, петли и даже кратные петли (ссылки вполне могут идти с одной страницы данного сайта на другую его страницу). Это такой «псевдомультиорграф». Сразу понятно, что для подобного «зверя» модель Эрдёша—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 $k \geqslant 1$ — некоторая константа. Для сравнения, у полного графа на $t$ вершинах $C_t^2 = \Theta(t^2)$ ребер. Однако — и это во-вторых — диаметр веб-графа исключительно скромен. В 1999 году он имел величину 5–7 (см. [17]). Это хорошо всем известное свойство любой социальной сети, которое принято в обыденной речи характеризовать выражением «мир тесен». Например, говорят о том, что любые два человека в мире «знакомы через 5–6 рукопожатий». Точно так же и сайты: «кликая» по ссылкам, можно с любого сайта на любой другой перейти за 5–7 нажатий клавиши компьютерной мыши. Конечно, тут есть важная оговорка. Некоторые 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 $c/d^{\lambda}$, 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 $\lambda$.

Ввиду перечисленных наблюдений, не остается никаких сомнений в том, что модель Эрдёша—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 (т. е. 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 Барабаши и Альберт никак не конкретизировали, какую именно из этих моделей они предлагают рассматривать. А эти модели исключительно разнородны по своим свойствам. Ведь можно ставить ссылки независимо друг от друга, а можно еще и зависимости между разными ссылками с одного сайта учитывать. В итоге удается доказать даже такой забавный факт (см. [18]).

**Теорема 7.** *Пусть $f(n)$, $n \geqslant 2$, — произвольная целочисленная функция, такая, что $f(2) = 0$, $f(n) \leqslant f(n+1) \leqslant f(n) + 1$ для всех $n \geqslant 2$ и $f(n) \to \infty$ при $n \to \infty$. Тогда существует такая модель типа Барабаши—Альберт, что в ней с вероятностью, стремящейся к единице при $n \to \infty$, случайный граф содержит в точности $f(n)$ треугольников.*

Одну из наиболее правильных спецификаций модели Барабаши—Альберт предложили в начале двухтысячных годов Б. Боллобаш и О. Риордан. В следующем пункте мы ее обсудим.

### 2.2. Модель Боллобаша—Риордана

Наиболее полно эта модель описана в книге [8] и обзоре [18]. Также имеется малодоступная книга [19], которая недавно была переиздана [20]. 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.

**2.2.1. 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 $G_1^1 = (\{1\}, \{(1, 1)\})$, т. е. в начальный момент времени есть одна вершина и одна петля. Пусть теперь граф $G_1^{n-1}$ уже построен. У него вершины образуют множество $\{1, \ldots, n-1\}$, а ребер у него тоже $n-1$

штука. Добавим вершину $n$ и ребро $(n, i)$, у которого $i \in \{1, \ldots, n\}$. Ребро $(n, n)$ будет появляться с вероятностью $\dfrac{1}{2n-1}$; ребро $(n, i)$ возникнет с вероятностью $\dfrac{\deg i}{2n-1}$, где $\deg i$ — степень вершины $i$ в графе $G_1^{n-1}$. Очевидно, что распределение вероятностей задано корректно, поскольку

$$\sum_{i=1}^{n-1} \frac{\deg i}{2n-1} + \frac{1}{2n-1} = \frac{2n-2}{2n-1} + \frac{1}{2n-1} = 1.$$

Случайный граф $G_1^n$ построен, и он удовлетворяет принципу предпочтительного присоединения.

Осталось перейти к $G_k^n$. Берем $G_1^{kn}$. Это граф с $kn$ вершинами и $kn$ ребрами. Делим множество его вершин на последовательные куски размера $k$:

$$\{1, \ldots, k\}, \quad \{k+1, \ldots, 2k\}, \quad \ldots, \quad \{k(n-1)+1, \ldots, kn\}.$$

Объявляем каждый кусок «вершиной», а ребра сохраняем, т. е. если были ребра внутри куска, то будут кратные петли, а были ребра между двумя различными кусками — будут кратные ребра. Внешне — вполне себе интернет, как мы его и представляли. Вершин стало $n$, а ребер — по-прежнему $kn$. Цель реализована.

**2.2.2. Статическая модификация, или LCD-модель.** Введем такой объект, который называется *линейной хордовой диаграммой*. Вообще-то он возник в топологии и теории узлов (см., например, [21]), но его комбинаторика оказывается напрямую связана с формированием веб-графа.

Итак, зафиксируем на оси абсцисс на плоскости $2n$ точек: 1, 2, 3, … …, $2n$. Разобьем эти точки на пары, и элементы каждой пары соединим дугой, лежащей в верхней полуплоскости. Полученный объект назовем *линейной хордовой диаграммой* или LCD (по-английски *linearized chord diagram*). Дуги в нем могут пересекаться, лежать друг под дружкой, но не могут иметь общих вершин. Количество различных LCD легко считается. Оно равно

$$l_n = \frac{(2n)!}{2^n n!}.$$

По каждой LCD построим граф с $n$ вершинами и $n$ ребрами. Действуем так. Идем слева направо по оси абсцисс, пока не встретим впервые правый конец какой-либо дуги. Пусть этот конец имеет номер $i_1$. Объявляем набор $\{1, \ldots, i_1\}$ 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 $\{i_1 + 1, \ldots, i_2\}$. И так далее. Поскольку правых концов у дуг в данной диаграмме $n$ штук, получаем всего $n$ вершин. А ребра



порождаем дугами. Иными словами, две вершины соединяем ребром, коль скоро между соответствующими наборами есть дуга. Ребра ориентируем справа налево. Аналогично возникают петли. Дуг $n$, и ребер $n$.

Теперь считаем LCD случайной, т. е. полагаем вероятность каждой диаграммы равной $1/l_n$. Возникают случайные графы. Можно показать, что такие графы по своим вероятностным характеристикам практически неотличимы от графов $G_1^n$.

Графы с $n$ вершинами и $kn$ ребрами получаем тем же способом, что и в п. 2.2.1.

**2.2.3. Некоторые результаты.** Модель Боллобаша—Риордана замечательна не только тем, что с ее помощью наводится порядок в «каше», которую «заварили» Барабаши и Альберт, но еще и тем, что она полностью адекватна эмпирическим наблюдениям. Прежде всего справедлива

**Теорема 8.** *Для любого $k \geqslant 2$ и любого $\varepsilon > 0$*

$$P\left((1-\varepsilon)\frac{\ln n}{\ln\ln n} \leqslant \operatorname{diam} G_k^n \leqslant (1+\varepsilon)\frac{\ln n}{\ln\ln n}\right) \to 1, \quad n \to \infty.$$

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 [22] авторами модели. А в работе [23] была внесена ясность и в вопрос о распределении степеней вершин.

**Теорема 9.** *Для любого $k \geqslant 1$ и любого $d \leqslant n^{1/15}$*

$$E\left(\frac{|\{i = 1, \ldots, n\colon \deg_{G_k^n} i = d\}|}{n}\right) \sim \frac{2k(k+1)}{(d+k+1)(d+k+2)(d+k+3)}. \qquad (2)$$

Поскольку $k$ — константа, выражение в правой части (2) имеет вид $\mathrm{const}/d^3$. Да это же в точности степенной закон! Правда, в формулировке теоремы написано математическое ожидание, а не вероятность, но одно из другого получается за счет мартингальных неравенств и соответствующих теорем о плотной концентрации меры около среднего (см. [23]).

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 $d \leqslant n^{1/15}$, которое ставит крест на практической применимости теоремы. Даже при $n \approx 10^{12}$, чего в природе (пока) не бывает, мы имеем лишь $d \leqslant 10^{4/5}$, и это нелепо.

Последний недостаток недавно устранил Е. А. Гречников — исследователь-разработчик в «Яндексе», который получил более точный результат практически без ограничений на $d$ (см. [24]).

Первым же недостатком занимались много и, в частности, предлагали различные альтернативные модели. Две из таких моделей мы обсудим в п. 2.3 и 2.4. 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 [18] приводится громоздкая общая формула и пара ее симпатичных следствий, которые мы выпишем и здесь.

**Теорема 10.** *Пусть $k \geqslant 2$. Пусть также $K_3$ — полный граф на трех вершинах. Тогда*

$$E(\sharp(K_3,\, G_k^n)) = (1 + o(1)) \cdot \frac{k \cdot (k-1) \cdot (k+1)}{48} \cdot (\ln n)^3$$

*при $n \to \infty$.*

**Теорема 11.** *Пусть фиксированы $k \geqslant 2$ и $l \geqslant 3$. Пусть также $C_l$ — цикл на $l$ вершинах. Тогда*

$$E(\sharp(C_l,\, G_k^n)) = (1 + o(1)) \cdot c_{k,l} \cdot (\ln n)^l$$

*при $n \to \infty$, где $c_{k,l}$ — это положительная константа. Более того, при $k \to \infty$ 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 [32].

**Теорема 12.** *Пусть задан граф $H$, степени вершин которого равны $d_1, \ldots, d_s$. Обозначим через $\sharp(d_i = m)$ число вершин в $H$, степень каждой из которых равна $m$. Тогда*

$$E(\sharp(H,\, G_k^n)) = \Theta\bigl(n^{\sharp(d_i=0)} \cdot (\sqrt{n})^{\sharp(d_i=1)} \cdot (\ln n)^{\sharp(d_i=2)}\bigr).$$

*Зависимость от $k$ занесена в константу $\Theta$.*

Надо полагать, что нечто подобное было известно и авторам статьи [18], 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

$$E(\sharp(K_3,\, G_k^n)) = \Theta((\ln n)^3),$$



и это прекрасно согласуется с теоремой 10. Аналогично можно разобраться и с циклами (теорема 11). А если взять $K_4$ — полный граф на четырех вершинах, — то теорема 12 скажет, что средняя его встречаемость в веб-графе постоянна. Иными словами, «тетраэдров» в веб-графах почти не бывает.

Отметим, что в реальном вебе случаются не только тетраэдры, но и клики куда большей мощности. Это связано с деятельностью спамеров, которые искусственно расставляют ссылки, желая повысить рейтинги сайтов, заплативших за раскрутку. Спам в модели Боллобаша—Риордана не учтен, и это тоже минус.

Последний в этом пункте любопытный сюжет связан с распределением не первых, а так называемых вторых степеней вершин случайного веб-графа. Речь идет о величине

$$d_2(t) = |\{(i,\, j)\colon i \neq t,\, j \neq t,\, (i,\, t) \in G_1^n,\, (i,\, j) \in G_1^n\}|.$$

Эта величина для данной вершины $t$ графа $G_1^n$ равна числу ребер, выходящих из вершин, которые являются соседями вершины $t$, и не ведущих в $t$. Граф $G_k^n$ с $k \geqslant 2$ устроен сложнее и в этом контексте пока не рассматривался. Недавно Л. А. Остроумова при участии Гречникова установила следующий результат (см. [25]).

**Теорема 13.** *Для любого $d > 1$ выполнено*

$$E\Bigl(\frac{|\{t \in G_1^n\colon d_2(t) = d\}|}{n}\Bigr) = \frac{4}{d^2}\Bigl(1 + O\Bigl(\frac{\ln^2 d}{d}\Bigr) + O\Bigl(\frac{d^2}{n}\Bigr)\Bigr).$$

Иными словами, при полном отсутствии каких-либо ограничений мы имеем степенной закон и для вторых степеней.

### 2.3. Модель Бакли—Остгуса

Простейшая идея о том, как можно, слегка модифицировав модель Боллобаша—Риордана, получить отличный от тройки показатель в степенном законе распределения степеней вершин, пришла в голову практически одновременно сразу двум независимым группам исследователей, которые в статьях [26,27] предложили следующую конструкцию. Зафиксируем положительное вещественное число $a$. Сперва построим последовательность случайных графов $\{H_{1,a}^n\}$. Граф $H_{1,a}^1$ совпадает с графом $G_1^1$ из пункта 2.2.1. У графа $H_{1,a}^{n-1}$ вершины образуют множество $\{1, \ldots, n-1\}$, а ребер у него $n-1$ штука. Добавляем вершину $n$ и присоединяем ребро $(n,\, i)$, дабы получить граф $H_{1,a}^n$. Ребро $(n,\, n)$ появляется с вероятностью $\dfrac{a}{(a+1)n-1}$, а ребро $(n,\, i)$ — с вероятностью $\dfrac{(\deg i) - 1 + a}{(a+1)n-1}$. Граф $H_{k,a}^n$ получаем из графа $H_{1,a}^{kn}$ стандартной склейкой (см. п. 2.2.1).

Ясно, что при $a = 1$ мы возвращаемся к модели Боллобаша—Риордана. Вообще, число $a$ интерпретируют как «начальную притягательность» каждой вершины. Сами авторы модели не доказали ни одного строгого утверждения относительно тех или иных ее характеристик. Зато П. Бакли и Д. Остгус обосновали в [28] следующую важную теорему (и именно поэтому модель носит теперь имя Бакли—Остгуса).

**Теорема 14.** *Для любого $k \geqslant 1$, любого целого $a \geqslant 1$ и любого $d \leqslant n^{1/100(a+1)}$*

$$E\left(\frac{|\{i = 1, \ldots, n\colon \deg_{H^n_{k,a}} i = d\}|}{n}\right) = \Theta(d^{-2-a}).$$

На самом деле, авторы теоремы 14 писали даже асимптотику математического ожидания. Однако мы не приводим ее, так как при всей своей важности результат не вполне удовлетворителен. Дело в том, что в нем $a$ *целое*. Иными словами, мы не можем подставить вместо $a$ величину 0,1, желая получить в итоге правильный степенной закон (с показателем 2,1). Кроме того, ограничение $d \leqslant n^{1/100(a+1)}$ еще ужаснее ограничения $d \leqslant n^{1/15}$ из теоремы 9.

Обе проблемы недавно устранил Е. А. Гречников (см. [29]).

**Теорема 15.** *Для любого $k \geqslant 1$, любого $a > 0$ и любого $d \geqslant k$*

$$E\left(\frac{|\{i = 1, \ldots, n\colon \deg_{H^n_{k,a}} i = d\}|}{n}\right) \sim (a+1)\frac{\Gamma(ka + a + 1)}{\Gamma(ka)}d^{-2-a}.$$

Таким образом, модель Бакли—Остгуса и впрямь адекватнее модели Боллобаша—Риордана. О том, насколько эта адекватность удивительна, можно прочесть в совсем свежей статье [33]. Также изучены вторые степени вершин в модели [34].

### 2.4. Модель копирования

Здесь мы опишем еще одну очень интересную модель, которая также призвана объяснить феномен степенного закона в реальных сетях. Эта модель возникла практически в одно время с моделью Барабаши—Альберт. Она принадлежит Р. Кумару, П. Рагхавану, С. Раджагопалану, Д. Сивакумару, А. Томкинсу и Э. Упфалу (см. [30]).

Фиксируем $\alpha \in (0, 1)$ и $d \geqslant 1$, $d \in \mathbb{N}$. Случайный граф будет расти, и это будет похоже на процесс из п. 2.2.1. 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 $V_t = \{u_1, \ldots, u_s\}$, где $s$ отличается от $t$ на число вершин начального графа, т. е. на некоторую константу, выражаемую через $d$. Добавим к $G_t$ одну новую вершину $u_{s+1}$ и $d$ ребер, выходящих из $u_{s+1}$. Для этого сперва выберем случайную вершину $p \in V_t$ (все вершины в $V_t$ равновероятны). Одно за другим строим ребра из $u_{s+1}$ в $V_t$. На шаге с номером $i$, $i \in \{1, \ldots, d\}$, разыгрываем случайную величину, которая с вероятностью $\alpha$ 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 [30] — это теорема 16.

**Теорема 16.** *Пусть $N_{t,r}$ — это математическое ожидание числа вершин степени $r$ в графе $G_t$. Тогда*

$$\lim_{t\to\infty}\frac{N_{t,r}}{t} = \Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right).$$

Пафос теоремы в том, что в ней мы снова приходим к степенному закону. Более того, если вероятность копирования близка к 1 (а величина $\alpha$ — к нулю), то показатель степени может равняться ожидаемой величине 2,1, и это хорошо.

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 п. 2.2.3, зачастую образуют именно двудольные графы с плотной перелинковкой.

### 2.5. Ориентированная модель

Еще один существенный недостаток всех ранее рассмотренных моделей состоит в том, что по факту в них отсутствует ориентация ребер. Вернее, ориентация есть, но, скажем, в модели Боллобаша—Риордана исходящих ребер у каждой вершины $k$, что никак не соответствует реальности. Один из вариантов решения проблемы предложили в 2003 году Б. Боллобаш, К. Боргс, Дж. Чайес и О. Риордан (см. [31]).

Идея стандартная: строить случайный граф шаг за шагом. На сей раз, однако, не на каждом шаге, вообще говоря, добавляется новая вершина. А именно, с вероятностью $\alpha \in (0, 1)$ добавляется ровно одна вершина и с вероятностью $1 - \alpha$ новые вершины не появляются. Если реализуется первый вариант (с добавлением вершины), то возможны снова два случая. В первом случае (который возникает с вероятностью $\beta$) новая вершина ссылается на одну из старых, и вероятность этого пропорциональна *входящей* степени старой вершины (плюс начальная притягательность $\delta_{\text{in}} > 0$). Во втором случае, наоборот, одна из старых вершин ссылается на новую, и вероятность этого пропорциональна *исходящей* степени старой вершины (плюс «начальная тяга к простановке ссылок» $\delta_{\text{out}} > 0$). Наконец, в рамках второго варианта (когда новая вершина не возникает) ребро проводится между двумя случайными старыми вершинами. Начало ребра выбирается пропорционально исходящим степеням с учетом $\delta_{\text{out}}$; конец ребра — пропорционально входящим степеням с учетом $\delta_{\text{in}}$. Могут возникнуть и петли.

Как показали авторы модели, степенной закон в ней есть, и подбором параметров его можно сделать абсолютно адекватным. Правда, они это сделали лишь для фиксированных значений $d$ степеней вершин. Сейчас Е. А. Гречников работает над устранением этого недостатка.

### 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.
8. *Bollobás B.* Random Graphs. Cambridge Univ. Press, 2001.





9. *Алон Н.*, *Спенсер Дж.* Вероятностный метод. М.: Бином. Лаборатория знаний, 2007.
10. *Маргулис Г. А.* Вероятностные характеристики графов с большой связностью // Проблемы передачи информации. 1974. Т. 10. С. 101–108.
11. *Janson S.*, *Łuczak T.*, *Ruciński A.* Random graphs. New York: Wiley, 2000.
12. *Райгородский А. М.* Модели случайных графов. М.: МЦНМО, 2011.
13. *Karp R.* The transitive closure of a random digraph // Random structures and algorithms. 1990. V. 1. P. 73–94.
14. *Карлин С.* Основы теории случайных процессов. М: Мир, 1971. 536 с.
15. *Barabási L.-A.*, *Albert R.* Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.
16. *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.
17. *Albert R.*, *Jeong H.*, *Barabási L.A.* Diameter of the world-wide web // Nature. 1999. V. 401. P. 130–131.
18. *Bollobás B.*, *Riordan O.* Mathematical results on scale-free random graphs // Handbook of graphs and networks. Weinheim: Wiley-VCH, 2003. P. 1–34.
19. *Райгородский А. М.* Экстремальные задачи теории графов и анализ данных. М.: Регулярная и хаотическая динамика, 2009.
20. *Райгородский А. М.*Экстремальные задачи теории графов и Интернет. М.: Интеллект, 2012.
21. *Stoimenow A.* Enumeration of chord diagrams and an upper bound for Vassiliev invariants // J. Knot Theory Ramifications. 1998. V. 7, № 1. P. 93–114.
22. *Bollobás B.*, *Riordan O.* The diameter of a scale-free random graph // Combinatorica. 2004. V. 24, № 1. P. 5–34.
23. *Bollobás B.*, *Riordan O.*, *Spencer J.*, *Tusnády G.* The degree sequence of a scale-free random graph process // Random Structures Algorithms. 2001. V. 18, № 3. P. 279–290.
24. *Grechnikov E. A.* An estimate for the number of edges between vertices of given degrees in random graphs in the Bollobás—Riordan model // Moscow Journal of Combinatorics and Number Theory. 2011. V. 1, N 2. P. 40–73.
25. *Ostroumova L. A.*, *Grechnikov E. A.* The distribution of second degrees in the Bollobás—Riordan random graph model // Moscow Journal of Combinatorics and Number Theory. 2012. V. 2, iss. 2. P. 82–106.
26. *Drinea E.*, *Enachescu M.*, *Mitzenmacher M.* Variations on random graph models for the web // Technical report, Harvard University, Department of Computer Science. 2001.
27. *Dorogovtsev S. N.*, *Mendes J. F. F.*, *Samukhin A. N.* Structure of growing networks with preferential linking // Phys. rev. lett. 2000. V. 85. P. 4633.
28. *Buckley P. G.*, *Osthus D.* Popularity based random graph models leading to a scale-free degree sequence // Discrete Math. 2004. V. 282. P. 53–68.
29. *Grechnikov E. A.* The degree distribution and the number of edges between nodes of given degrees in the Buckley—Osthus model of a random web graph // Internet Mathematics. 2012. V. 8, № 3. P. 257–287.

30. *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.
31. *Bollobás B.*, *Borgs Ch.*, *Chayes J.*, *Riordan O.* Directed scale-free graphs // SODA'03 Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms. 2003.
32. *Рябченко А. А.*, *Самосват Е. А.* О числе подграфов в случайном графе Барабаши—Альберт // Изв. РАН. Сер. матем. 2012. Т. 76, № 3. С. 183–202.
33. *Grechnikov E.A. et al.* Empirical Validation of the Buckley—Osthus Model for the Web Host Graph // Proceedings of The 21st ACM Conference on Information and Knowledge Management. 2012; см. также более полную версию: *Zhukovskiy M. et al.* Empirical Validation of the Buckley—Osthus Model for the Web Host Graph: Degree and Edge Distributions; `arXiv:1208.2355`.
34. *Kupavskii A.B.*, *Ostroumova L.A.*, *Shabanov D.A.*, *Tetali P.* The distribution of second degrees in the Buckley-Osthus random graph model. Accepted to Internet Mathematics.

# Задачи

## Задачи к главам пособия и приложениям

**Задача к главе 1: устойчивость равновесия Нэша в матричной игре.** Рассматривается игра двух лиц с матрицей выигрышей

$$\begin{pmatrix} (3,\,3) & (10,\,1) \\ (1,\,10) & (7,\,7) \end{pmatrix}.$$

Найдите равновесие Нэша. Будет ли оно устойчивым (см. с. 84–98 книги [1])?

**Задача к главе 1 и приложению Ю. Е. Нестерова и С. В. Шпирко** (Нестеров—де Пальма). Из жилого района в рабочий район утром должны отправляться $N = 10^4$ автомобилей. Водитель каждого автомобиля хочет приехать ровно к 9 часам утра в рабочий район. При этом каждая минута опоздания штрафуется в $\alpha = 10$ рублей, а каждая потерянная минута в пути или в ожидании начала рабочего дня, если водитель приехал раньше времени, стоит $\beta = 3$ рубля. Время в пути по свободной дороге занимает $T = 60$ минут. Но на середине дороги есть узкое место, пропускная способность которого ограничена 3 000 автомобилей в час.

Покажите, что равновесное распределение водителей по времени выезда из жилого района $n(t)$ представимо в виде:

$$n(t) = \begin{cases} 0, & t < t_1; \\ n_1, & t_1 \leqslant t < t_2; \\ n_2, & t_2 \leqslant t < t_3; \\ 0, & t \geqslant t_3. \end{cases}$$

Найдите $n_1$, $n_2$, $t_1$, $t_2$, $t_3$. Попробуйте обобщить эту задачу.

**Задача к главе 1: сходимость к равновесию Нэша** (Малишевский—Опойцев, 1972). Рассмотрим систему обыкновенных дифференциальных уравнений (СОДУ): [1)]

$$\dot{x}_i = \gamma_i(t)(f_i(x_1,\,\ldots,\,x_n) - x_i), \tag{D}$$

[1)] Такие СОДУ возникают, например, при исследовании коллективного поведения, в частности «при нащупывании равновесия Нэша». Действительно, пусть $D_i(x_1,\,\ldots,\,x_i,\,\ldots,\,x_n)$ — функция выигрыша игрока $i$, если игроки придерживаются набора стратегий $\vec{x} = (x_1,\,\ldots\,\ldots,\,x_i,\,\ldots,\,x_n)$. Пусть «функция цели» $f_i(\vec{x})$ игрока $i$ однозначным образом определяется из

где $\gamma_i(t) > 0$ при $t \geqslant 0$ и $\int\limits_0^\infty \gamma_i(t)\,dt = \infty$, $i = 1, \ldots, n$; $f_i(\cdot)$, $i = 1, \ldots, n$ — непрерывные функции такие, что $\exists K \subset \mathbb{R}^n$ (компакт): $\vec{f}(K) \subseteq K$.

Докажите, что приводимое ниже условие обеспечивает существование равновесия СОДУ (единственность в $K$) и его асимптотическую устойчивость:

$$\sum_{j=1}^{n} \left| \frac{\partial f_i(\vec{x})}{\partial x_j} \right| < 1 \quad \forall \vec{x} \in K, \quad i = 1, \ldots, n.$$

**Указание.** В кубической норме ($\|\vec{x}\|_\square = \max\limits_{i=1,\ldots,n} |x_i|$) $\vec{f}(\vec{x})$ — сжимающее отображение компакта $K$ в себя. Устойчивость показывается с помощью второго метода Ляпунова. В качестве функции Ляпунова следует взять $V(\vec{x}) = \|\vec{x} - \vec{x}^*\|_\square$, где $\vec{x}^*$ — единственное положение равновесия в $K$.

Литература

1. *Мулен Э.* Теория игр с примерами из математической экономики. М.: Мир, 1985.
2. *Опойцев В.И.* Равновесие и устойчивость в моделях коллективного поведения. М.: Наука, 1977.
3. *Мулен Э.* Теория игр с примерами из математической экономики. М.: Мир, 1985.
4. *Малишевский А.В.* Качественные модели в теории сложных систем. М.: Наука, 1998.

**Задача к главе 1: устойчивые системы большой размерности** (В.И. Опойцев, 1985). Из курсов функционального анализа и вычислительной математики хорошо известно, что если спектральный радиус матрицы $A = \|a_{ij}\|_{i,j=1}^n$ меньше единицы: $\rho(A) < 1$, то итерационный процесс $\vec{x}^{n+1} = A\vec{x}^n + \vec{b}$ для системы уравнений $\dot{\vec{x}} = A\vec{x} + \vec{b} - \vec{x}$ вне зависимости от точки старта $\vec{x}^0$ сходится к единственному решению уравнения $\vec{x}^* = A\vec{x}^* + \vec{b}$. Скажем, если $\|A\|_\square = \max\limits_i \sum\limits_{j=1}^n |a_{ij}| < 1$, то и $\rho(A) < 1$ (обратное, конечно, не верно). Предположим, что существует такое малое $\varepsilon > 0$, что

$$\frac{1}{n} \sum_{i,j=1}^{n} |a_{ij}| < 1 - \varepsilon. \qquad \text{(S)}$$

Заметим, что отсюда не следует $\rho(A) < 1$. Тем не менее, введя на множестве S матриц, удовлетворяющих условию (S), равномерную меру, покажите, что относительная мера тех матриц из S, для которых спектральный радиус не меньше единицы, стремится к нулю с ростом $n$ ($\varepsilon$ фиксировано и от $n$ не зависит).

**Указание.** 1. Покажите, что при доказательстве можно ограничиться матрицами с неотрицательными элементами.

2. Покажите, что, не ограничивая общности, можно также считать, что в определении множества S стоит не неравенство, а равенство. Так определенное множество матриц будем называть SE.

3. Положите, например, $a_{ij} \in \mathrm{Exp}(n/(1-\varepsilon))$ — н.о.р. [1)] и покажите, что при $n \to \infty$ распределение элементов случайной матрицы $A = \|a_{ij}\|_{i,j=1}^n$ будет сходиться (уточните, в каком смысле) к равномерному распределению на SE.

4. Покажите, введя обозначение $P_n = P(\|A\|_\square \geqslant 1) \geqslant P(\rho(A) \geqslant 1)$ и используя неравенство Чебышева, что

$$P_n \leqslant nP\left(\sum_{j=1}^{n} a_{1j} \geqslant 1\right) \leqslant \frac{n}{\varepsilon^4} E\left[\left(\sum_{j=1}^{n} a_{1j} - (1-\varepsilon)\right)^4\right] = O\left(\frac{1}{n}\right) \xrightarrow[n\to\infty]{} 0.$$

Литература

1. *Опойцев В.И.* Устойчивые системы большой размерности // Автоматика и телемеханика. 1986, № 6. С. 43–49.
2. *Опойцев В.И.* Нелинейная системостатика. М.: Наука, 1986.

**Задача к главе 1: сублинейный приближенный вероятностный алгоритм для матричных игр** (Григориадис—Хачиян, 1995). Рассматривается симметричная антагонистическая игра двух лиц X и Y. Смешанные стратегии X и Y будем обозначать соответственно $\vec{x}$ и $\vec{y}$. При этом $x_k$ — вероятность того, что игрок X выберет стратегию с номером $k$, аналогично определяется $y_k$. Таким образом,

$$\vec{x}, \vec{y} \in S = \{\vec{x} \in \mathbb{R}^n : \vec{e}^T\vec{x} = 1, \vec{x} \geqslant \vec{0}\}, \quad \text{где } \vec{e} = (1, \ldots, 1)^T.$$

Выигрыш игрока X:

$$V_X(\vec{x}, \vec{y}) = \vec{y}^T A \vec{x},$$

а выигрыш игрока Y:

$$V_Y(\vec{x}, \vec{y}) = -\vec{y}^T A \vec{x}$$

(игра антагонистическая).

Каждый игрок стремится максимизировать свой выигрыш при заданном ходе оппонента. Равновесием Нэша (в смешанных стратегиях) называется

условия $D_i(x_1, \ldots, f_i(\vec{x}), \ldots, x_n) = \max\limits_{x_i \in X_i} D_i(x_1, \ldots, x_i, \ldots, x_n)$, где $X_i$ — множество возможных стратегий игрока $i$. Тогда СОДУ (D), очевидным образом, выражает стремление игроков двигаться по направлению своей цели (по мере движения цель меняется), т.е. определенную рациональность игроков. Отметим здесь нечувствительность результатов к тому, насколько сильно (это характеризуется функциями $\gamma_i(t) > 0$, $i = 1, \ldots, n$) каждый из игроков стремится к своей цели.

[1)] Независимые случайные величины, одинаково распределенные по показательному закону с параметром $n/(1-\varepsilon)$, т.е. $P(a_{ij} > x) = \exp(-(n/(1-\varepsilon))x)$, $x \geqslant 0$.

такая пара стратегий $(\vec{x}^*, \vec{y}^*)$, что

$$\vec{x}^* \in \operatorname{Arg}\max_{\vec{x}\in S} \vec{y}^{*T} A\vec{x}, \quad \vec{y}^* \in \operatorname{Arg}\min_{\vec{y}\in S} \vec{y}^T A\vec{x}^*.$$

Ценой игры называют

$$\max_{\vec{x}\in S}\min_{\vec{y}\in S} \vec{y}^T A\vec{x} = \min_{\vec{y}\in S}\max_{\vec{x}\in S} \vec{y}^T A\vec{x} = \vec{y}^{*T} A\vec{x}^*.$$

Поскольку по условию игра симметричная, то $A = -A^T$ — матрица $n \times n$. С помощью стандартной редукции можно свести к этому случаю общий случай произвольной несимметричной матричной игры. В рассматриваемом же случае цена игры (выигрыш игроков в положении равновесия Нэша) есть 0, а множества оптимальных стратегий игроков совпадают. Требуется найти с точностью $\varepsilon > 0$ положение равновесия Нэша (оптимальную стратегию), т. е. требуется найти такой вектор $\vec{x} \in S$, что $A\vec{x} \leqslant \varepsilon\vec{e}$.

Покажите, считая элементы матрицы $A$ равномерно ограниченными, скажем, единицей, что приводимый ниже алгоритм находит с вероятностью не меньшей $1/2$ (вместо $1/2$ можно взять любое положительное число, меньшее единицы) такой $\vec{x}$ за время

$$O(\varepsilon^{-2} n \log^2 n),$$

т. е. в определенном смысле даже не вся матрица (из $n^2$ элементов) просматривается. Отметим также, что в классе детерминированных алгоритмов время работы растет с ростом $n$ не медленнее, чем $n^2$ (эта нижняя оценка получается из информационных соображений [1,2]). Другими словами, никакой детерминированный алгоритм не может так же асимптотически быстро находить приближенно равновесие Нэша. Точнее говоря, описанный ниже вероятностный алгоритм дает почти квадратичное ускорение по сравнению с детерминированными.

**Алгоритм**

1. Инициализация: $\vec{X} = \vec{U} = \vec{0}$, $\vec{p} = \vec{e}/n$, $t = 0$.

2. Повторить шаги 3–7.

3. Счетчик итераций: $t := t + 1$.

4. Датчик случайных чисел: выбираем $k \in \{1, \ldots, n\}$ с вероятностью $p_k$.

5. Модификация $\vec{X}$: $X_k := X_k + 1$.

6. Модификация $\vec{U}$: $U_i := U_i + a_{ik}$, $i = 1, \ldots, n$.

7. Модификация $\vec{p}$: $p_i := p_i \exp(\varepsilon a_{ik}/2) \Big/ \left( \sum_{j=1}^{n} p_j \exp(\varepsilon a_{jk}/2) \right)$, $i = 1, \ldots, n$.



8. Критерий останова: если $\vec{U}/t \leqslant \varepsilon\vec{e}$, то останавливаемся и печатаем $\vec{x} = \vec{X}/t$.

**Указание.** Покажите, что с вероятностью не меньшей, чем $1/2$, алгоритм остановится через $t^* = 4\varepsilon^{-2} \ln n$ итераций. Для этого введите (следуя Фройнду—Шапире [3])

$$P_i(t) = \exp\left(\frac{\varepsilon U_i(t)}{2}\right) \quad \text{и} \quad \Phi(t) = \sum_{i=1}^{n} P_i(t).$$

Покажите, что

$$U(t) = Ax(t), \quad p_i(t) = \frac{P_i(t)}{\sum\limits_{j=1}^{n} P_j(t)},$$

$$\Phi(t+1) = \sum_{i=1}^{n} P_i(t) \exp\left(\frac{\varepsilon a_{ik}}{2}\right) = \Phi(t) \sum_{i=1}^{n} p_i(t) \exp\left(\frac{\varepsilon a_{ik}}{2}\right),$$

$$E[\Phi(t+1)|\vec{P}(t)] = \Phi(t) \sum_{i,k=1}^{n} p_i(t) p_k(t) \exp\left(\frac{\varepsilon a_{ik}}{2}\right)$$

и

$$\exp\left(\frac{\varepsilon a_{ik}}{2}\right) \leqslant 1 + \varepsilon\frac{a_{ik}}{2} + \frac{\varepsilon^2}{6}.$$

Используя это и кососимметричность матрицы $A$, покажите, что

$$E[\Phi(t+1)] \leqslant nE[\Phi(t)]\left(1 + \frac{\varepsilon^2}{6}\right).$$

Следовательно,

$$E[\Phi(t)] \leqslant n \exp\left(\frac{t\varepsilon^2}{6}\right) \quad \text{и} \quad E[\Phi(t^*)] \leqslant n^{5/3}.$$

Отсюда по неравенству Маркова имеем, что ($n \geqslant 8$):

$$P(\Phi(t^*) \leqslant n^2) \geqslant P(\Phi(t^*) \leqslant 2n^{5/3}) \geqslant \frac{1}{2}.$$

Тогда

$$P\left(\frac{\varepsilon U_i(t^*)}{2} \leqslant 2\ln n,\ i = 1, \ldots, n\right) \geqslant \frac{1}{2}.$$

Откуда уже следует, что

$$P(A\vec{x}(t^*) \leqslant \varepsilon\vec{e}) \geqslant \frac{1}{2}.$$

Приведенные выше идеи случайного квазиградиентного спуска сейчас активно используются в современных численных методах решения задач выпуклой оптимизации в пространствах огромной размерности [4].

Литература

1. *Хачиян Л. Г.* Избранные труды / сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.
2. *Успенский В. А.*, *Верещагин Н. К.*, *Шень А.* Колмогоровская сложность и алгоритмическая случайность. М.: МЦНМО, 2013; `ftp://ftp.mccme.ru/users/shen/kolmbook.pdf`
3. *Вьюгин В. В.* Элементы математической теории машинного обучения. М.: МФТИ, 2010; `http://www.iitp.ru/upload/publications/5759/vyugin1.pdf`
4. `http://www2.isye.gatech.edu/~nemirovs/`; `www.uclouvain.be/32349.html`; `http://elis.dvo.ru/~nurmi/`

**Задача к главе 1: стохастическая марковская динамика, приводящая к равновесию Нэша—Вардропа в BMW-модели распределения потоков*** (Аввакумов—Гасникова—Дорн, 2010). Свой путь на $(n+1)$-м шаге [1)] игрок, сидящий на корреспонденции $w$, выбирает согласно смешанной стратегии (не зависимо от других игроков): с вероятностью

$$\mathrm{Prob}_p^w(n+1) = \gamma_n \max\left(x_p(n),\, \frac{1}{n}\right) \exp\left(\frac{-G_p(\vec{x}(n))}{T}\right)/Z_n^w, \quad w \in W,$$

выбирает путь $p \in P_w$ $(0 < \gamma_n \leqslant 1)$, а с вероятностью $1-\gamma_n$ — действует 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

$$Z_n^w = \sum_{p \in P_w} \max\left(x_p(n),\, \frac{1}{n}\right) \exp\left(\frac{-G_p(\vec{x}(n))}{T}\right).$$

Множитель $\max(x_p(n),\, 1/n)$ характеризует желание имитировать, а также надежность использования этой стратегии. Параметр $\gamma$ характеризует «консерватизм» («ленивость»), чем меньше $\gamma$, 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 $\gamma_n$ наблюдается сходимость к равновесию Нэша—Вардропа при наиболее общих условиях относительно $T$ (вне зависимости от точки старта). Стоит также обратить внимание на высокую эффективность предложенной процедуры «нащупывания равновесия» с точки зрения количества итераций. Иначе говоря, на предложенный итерационный процесс можно смотреть просто как на эффективный способ численного нахождения равновесия Нэша—Вардропа. В экспериментах со студентами 5-го курса ФУПМ также наблюдалась сходимость к равновесию и колебания около него. Колебания можно объяснить, например, тем, что $\gamma_n \sim \gamma$.

Введение в динамику стохастичности сближает предложенный подход с поиском так называемых «стохастических равновесий в транспортных сетях» [1], с другой стороны, он принципиально отличается тем, что предполагает знание транспортных расходов по маршрутам (используется достоверная информация вчерашнего дня), на основе которых производится рандомизированный выбор. В стохастическом же равновесии водитель узнает лишь случайную оценку времени проезда по каждому из маршрутов и затем выбирает маршрут с минимальным временем.

Предложенную схему можно также трактовать как стохастическую динамику наилучших ответов в эволюционной (популяционной) игре [2–6], при этом имеется много общего с концепциями «quantal response equilibria» [7] (используется похожая рандомизация) и «minority games» [8] (наблюдаются похожие колебания около положения равновесия). Близкой к предложенному итерационному процессу является концепция генетических алгоритмов [9]. Однако наиболее близким к предложенной динамике является эффективный приближенный вероятностный алгоритм Григориадиса—Хачияна [10], который также может быть проинтерпретирован как стохастический вариант метода зеркального спуска поиска седловой точки [11].

Проверьте справедливость следующего утверждения.

*Пусть* $T > 0$ *достаточно мало,* $\sum\limits_{n=1}^{\infty} \gamma_n = \infty$, $\sum\limits_{n=1}^{\infty} (\gamma_n)^2 < \infty$. *Тогда вектор* $\vec{x}(n)$ *распределения потоков по путям сходится к одному из равновесий в зависимости от точки старта*:

$$\vec{x}(n) \xrightarrow[n\to\infty]{\text{п.н.}} \vec{x}^*(\vec{x}(0)).$$

*Если равновесие* $\vec{x}^*$ *единственно, то* $\vec{x}^*(\vec{x}(0)) \equiv \vec{x}^*$.

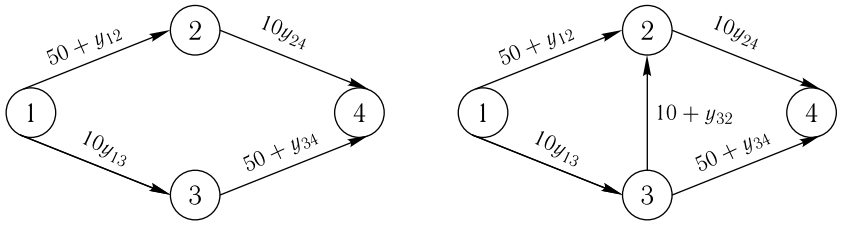
Исходя из результатов главы 1, предложите ситуацию, когда равновесие Нэша—Вардропа не единственно (см. пример В. И. Швецова ниже). Для описанной выше динамики исследуйте, к какому положению равновесия (в зависимости от точки старта) она будет сходиться.

Более подробно затронутые выше темы освещены, например, в [11–14].

В заключение рассмотрим два примера. Первый демонстрирует, что в результате строительства новой дороги новое равновесие Нэша—Вардропа окажется неэффективным по Парето и будет строго хуже, чем то, которое было до строительства. Тем не менее предложенная выше

[1)] Например, шаг с периодом в день можно проинтерпретировать как выбор утром маршрута следования (пути) из дома на работу, исходя из «опыта» вчерашнего дня. Заметим, что информацию о $G_p(\vec{x}(n))$ водители (игроки) черпают из открытых источников типа «Яндекс.Пробки», а множитель $\max(x_p(n),\, 1/n)$ определяется исходя из случайного опроса соседей, знакомых, коллег и т. п.

марковская динамика наилучших ответов (так же как и эксперимент со студентами) приводит именно к такому, не оптимальному по Парето, равновесию.




Случай 1 Случай 2

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

**Пример** (парадокс Браесса, 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: $y_{24} = x_{124} + x_{1324}$ (см. рис. 1). Естественно считать, что время движения — возрастающая функция потока.

Оба равновесия Нэша—Вардропа (в случаях 1 и 2) являются «притягивающими» положениями равновесия описанной выше динамики (для $\gamma \sim 1$, $T \sim 15 - 35$), см. рис. 2–4 (для случая 2).

Второй пример демонстрирует, что при весьма естественных условиях вектор-функция удельных затрат пользователей на проезд $\vec{G}(\vec{x})$ может не быть строго монотонной (см. п. 1.1.4 главы 1):

$$\exists \vec{x},\, \vec{y} \in X\ (\vec{x} \neq \vec{y})\colon \vec{G}(\vec{x}) = \vec{G}(\vec{y}) \Rightarrow \langle \vec{G}(\vec{x}) - \vec{G}(\vec{y}),\, \vec{x} - \vec{y} \rangle = 0.$$

Связано это может быть, например, с тем, что (см. раздел 1.2 главы 1)

$$\vec{G}(\vec{x}) = \Theta^T \vec{\tau}(\vec{y}), \quad \vec{y} = \Theta \vec{x},$$

где вектор $\vec{y} = \{y_e\}_{e \in E}$ описывает загрузку ребер (дуг) графа транспортной сети, $\vec{\tau}(\vec{y}) = \{\tau_e(y_e)\}_{e \in E}$ — вектор-функция затрат на проезд по ребрам графа транспортной сети, $\Theta$ — матрица инцидентности ребер и путей, и разные векторы распределения потоков $\vec{x}$ могут соответствовать одному и тому же вектору $\vec{y} = \Theta \vec{x}$.

**Пример: неединственность равновесия** (В. И. Швецов, 2009). На рис. 5 показано равновесное распределение потоков для любого значения параметра $x \in [0,\ 0{,}5]$. Подробности имеются в статье [15].



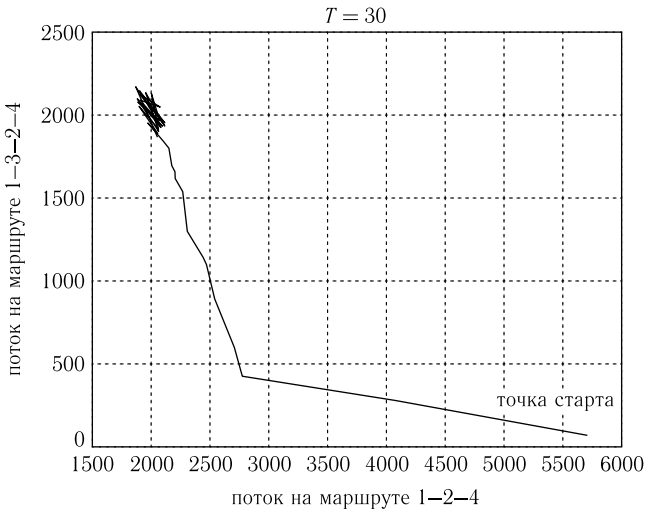



Рис. 2

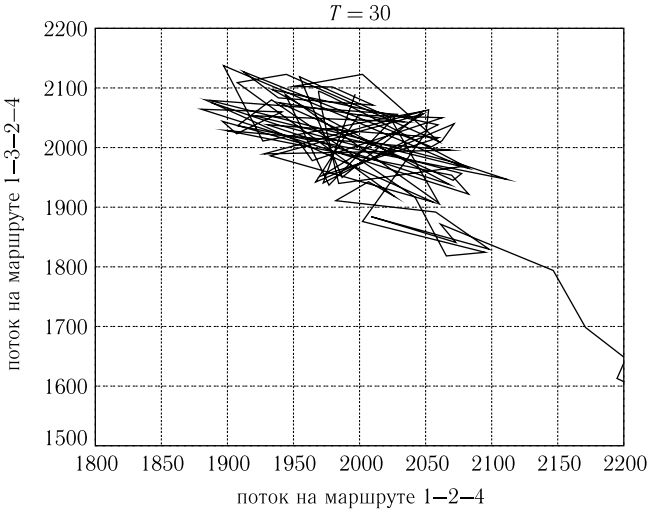



Рис. 3

В заключение напомним (см. п. 1.2.1 главы 1), что задача отыскания равновесия Нэша—Вардропа в рассматриваемом нами случае сводится

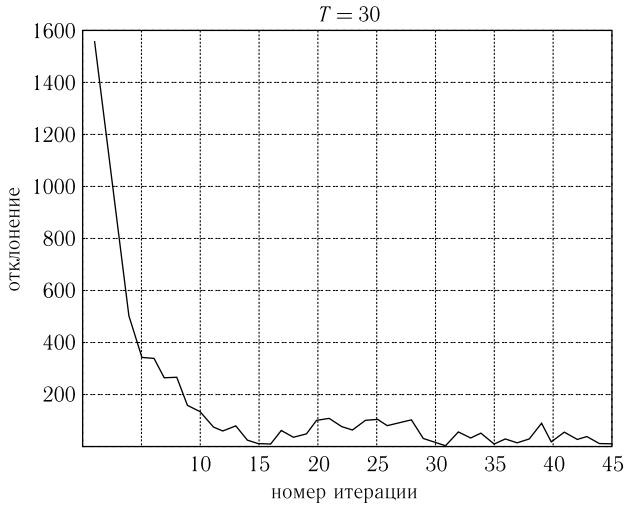



Рис. 4

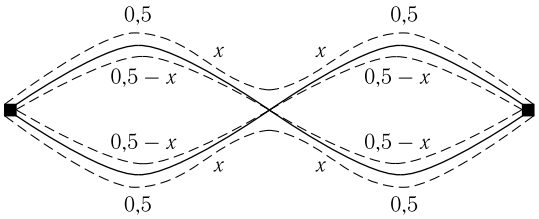



Рис. 5

к решению следующей задачи выпуклого программирования:

$$V(\vec{y}(\vec{x})) = \sum_{e\in E} \int_0^{y_e(\vec{x})} \tau_e(z)\,dz \to \min_{\vec{x}\in X}.$$

Отсюда видно, что если $\forall e \in E \quad \tau_e'(\cdot) > 0$, то $V(\vec{y})$ — строго выпуклый функционал, и равновесие $\vec{y}^*$ — единственно. Пример В. И. Швецова показывает, что это еще не означает единственность равновесия $\vec{x}^*$.

Литература

1. *Sheffi Y.* Urban transportation networks: Equilibrium analysis with mathematical programming methods. N.J.: Prentice–Hall Inc., Englewood Cliffs, 1985.
2. *Foster D., Young P.* Stochastic evolutionary game dynamics // Theoretical population biology. 1990. V. 38, № 2.



3. *Cressman R.* Evolutionary game theory and extensive form games. Cambridge, Mass.: MIT Press, 2003.
4. *Hofbauer J., Sigmund K.* Evolutionary game dynamics // Bulletin of the AMS. 2003. V. 40, № 4. P. 479–519.
5. *Васин А. А., Краснощеков П. С., Морозов В. В.* Исследование операций. М.: Издательский центр «Академия», 2008;
6. *Easley D., Kleinberg J.* Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, 2010; `http://www.cs.cornell.edu/home/kleinber/networks-book/`
7. *McKelvey R. D., Palfrey T. R.* Quantal response equilibria for extensive form games // Experimental economics. 1998. V. 1. P. 9–41.
8. *Marsili M.* Toy models of markets with heterogeneous interacting agents; `www.unifr.ch/econophysics/`, 2001.
9. *Fogel D. B.* Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. New York: IEEE Press, 2000.
10. *Хачиян Л. Г.* Избранные труды / Сост. С. П. Тарасов. М.: МЦНМО, 2009. С. 38–48.
11. *Juditsky A., Lan G., Nemirovski A., Shapiro A.* Stochastic approximation approach to stochastic programming // SIAM Journal on Optimization. 2009. V. 19, № 4. P. 1574–1609.
12. *Gasnikova E. V., Nagapetyan T. A.* About new dynamical interpretations of entropic model of correspondence matrix calculation and Nash—Wardrop's equilibrium in Beckmann's traffic flow distribution model. Ninth International Conference on Traffic and Granular Flow, 28 September–1 October 2011. Moscow: Springer, 2012; `arXiv:1112.1628`
13. *Como G., Salva K., Acemoglu D., Dahleh M. A., Frazzoli E.* Stability analysis of transportation networks with multiscale driver decisions; `arXiv:1101.2220v1`, 2011.
14. *Стенбринк П. А.* Оптимизация транспортных сетей. М.: Транспорт, 1981.
15. *Швецов В. И.* Проблемы моделирования передвижений в транспортных сетях // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 163–179.

**Задача к главе 1 и приложению Е. В. Гасниковой. 1**.** Предположим, что свой путь на $(n+1)$-м шаге игрок, сидящийг на корреспонденции $\omega$, выбирает согласно смешанной стратегии (вне зависимости от всех остальных): с вероятностью

$$\mathrm{Prob}_p^{\omega}(n+1) = \gamma \exp(-G_p(\vec{x}(n))/T)/Z_n^{\omega}, \quad \omega \in W, \tag{1}$$

выбирает путь $p \in P_\omega$ $(0 < \gamma \leqslant 1)$, а с вероятностью $1-\gamma$ — действует согласно стратегии, использованной на предыдущем $n$-м шаге. Здесь $x_p(n)$ — количество игроков, сидящих на корреспонденции $\omega$ и выбравших на $n$-м шаге стратегию $p \in P_\omega$, $Z_n^{\omega} = \sum\limits_{p\in P_\omega} \exp(-G_p(\vec{x}(n))/T)$, $T > 0$ — «темпера-

тура» (горячность) игроков. Это хорошо известное logit-распределение, или распределение Гиббса. Оно может быть проинтерпретировано как выбор каждым игроком наилучшей стратегии вчерашнего дня, если игрок «переносит» затраты вчерашнего дня $G_p(\vec{x}(n))$ на день сегодняшний, но допускает при этом случайные флуктуации $G_p(\vec{x}(n)) + \xi_p$, где $\xi_p$ — независимые случайные величины, имеющие одинаковое двойное экспоненциальное распределение (Гумбеля) с параметром $T > 0$. Тогда $\operatorname{argmin}\limits_{p\in P_w}\{G_p(\vec{x}(n)) + \xi_p\}$ как раз имеет указанное выше logit-распределение.

Пусть $\forall e \in E \quad \tau_e'(\cdot) > 0$ и $x_p^* > 0 \Rightarrow x_p^* \gg 1$. Верно ли, что стохастическая марковская динамика (1) «сходится» на больших временах к некоторому стационарному распределению вероятностей — так называемому «стохастическому равновесию в транспортной сети»? В предположении, что $T > 0$ и $\gamma > 0$ должным образом малы, верно ли, что это стационарное распределение сконцентрировано в малой окрестности такого равновесия Нэша—Вардропа:

$$\vec{x}^* = \operatorname*{argmin}_{\vec{x}\in X:\ \Theta\vec{x}=\vec{y}^*} \sum_{w\in W}\sum_{p\in P_w}\left(x_p \ln\frac{x_p}{|P_w|} - x_p\right)?$$

Здесь $\vec{y}^*$ — единственное решение задачи

$$V(\vec{y}) = \sum_{e\in E}\int_0^{y_e}\tau_e(z)\,dz \to \min_{\vec{y}=\Theta\vec{x},\ \vec{x}\in X}.$$

Описанный выше формализм позволяет пойти и дальше. В частности, можно рассмотреть модели равновесного распределения для нескольких классов пользователей (в том числе модели расщепления потоков по типу передвижений), модели равновесного распределения потоков с переменным спросом на потоки, динамические модели равновесного распределения потоков. Что касается последнего, пока имеющийся здесь задел совсем не велик и работа в этом направлении, на наш взгляд, сейчас представляет наибольший интерес в этой области.

Литература


1. *Гасников А. В.*, *Гасникова Е. В.*, *Федько О. С.* О возможной динамике в модели ранжирования web-страниц PageRank и модернизированной модели расчета матрицы корреспонденций // Труды МФТИ. 2012. Т. 4, № 2. С. 101–120.
2. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.
3. *Швецов В. И.* Проблемы моделирования передвижений в транспортных сетях // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4 (8). С. 169–179.





4. *Калинкин А. В.* Марковские ветвящиеся процессы с взаимодействием // УМН. 2002. Т. 57, № 2 (344). С. 23–84.


**Задача к главе 1 и приложению Е. В. Гасниковой. 2.** В условиях и обозначениях примера 2 приложения Е. В. Гасниковой считайте, что расстояние от района $i$ до района $j$ есть $l_{ij} > 0$, а

$$p^L_{k,m;\,p,q}(t) \equiv p^L_{k,m;\,p,q} = \lambda N^{-1}\exp\Bigg(\underbrace{R(l_{km}) + R(l_{pq})}_{\substack{\text{суммарные затраты}\\ \text{до обмена}}} - \underbrace{(R(l_{pm}) + R(l_{kq}))}_{\substack{\text{суммарные затраты}\\ \text{после обмена}}}\Bigg) > 0,$$

где $\lambda > 0$ характеризует интенсивность обменов, функция $R(l) = (\alpha l^\beta - \omega\ln l)/2$, $\alpha > 0$, $\beta > 0$, $\omega > 0$, отражает затраты в пути (первое слагаемое) и одновременно возможность найти подходящую работу на расстоянии порядка $l$ от места жительства: чем больше $l$, тем больше территория для поиска $\sim 2\pi l\Delta l$, тем больше вероятность успеха (найти подходящую работу), тем меньше должны быть затраты $R(l)$ (второе слагаемое). Напомним, что для обоснования классической энтропийной модели А. Дж. Вильсона в приложении мы брали $\omega = 0$. Однако анализ данных по Москве (на базе опыта А. В. Кулакова, МАДИ), проведенный осенью 2011 г. в Лаборатории прикладного моделирования транспортных систем ИПМ им. М. В. Келдыша РАН (зав. лаб. В. П. Осипов), показал заметно лучшее соответствие рассматриваемой модели реальным данным, если допускать $\omega > 0$ и оптимально его подбирать [1]. Отметим также, что ряд специалистов по расчету матрицы корреспонденций склонны считать, что вместо слагаемого $\alpha l^\beta$ в формуле для $R(l)$ правильнее писать:

$$R(l) = \begin{cases}\alpha l^\beta, & 0 < l \leqslant \bar{l};\\ \nu\ln l, & l > \bar{l};\end{cases}$$

где $\nu > \omega$, $\alpha\bar{l}^\beta = \nu\ln\bar{l}$.

Покажите, что равновесие $\{x_{ij}\}_{i,j=1}^n$ описанной макросистемы определяется следующими формулами:

$$x_{ij} = A_i B_j L_i W_j(l_{ij})^\omega \exp\left(-\alpha\cdot(l_{ij})^\beta\right),$$

где $\{A_i\}_{i=1}^n$ и $\{B_j\}_{j=1}^n$ определяются из соотношений:

$$A_i = \left(\sum_{j=1}^n B_j W_j(l_{ij})^\omega \exp\left(-\alpha\cdot(l_{ij})^\beta\right)\right)^{-1},$$

$$B_j = \left(\sum_{i=1}^n A_i L_i(l_{ij})^\omega \exp\left(-\alpha\cdot(l_{ij})^\beta\right)\right)^{-1}.$$

Существуют и другие модельные способы расчета матрицы корреспонденций. Например, модель С. А. Стауффера [2], также называемая моделью

промежуточных возможностей, и модель конкурирующих центров [3, 4]. Хотелось бы также обратить внимание на предложенную в статье [5] объединенную «гравитационно-конкурирующую» модель энтропийного типа. Примечательно, что обоснование таких популярных на практике моделей может быть получено подобно изложенной выше схеме, где в стохастическую динамику вводится больше специфики, в частности, вводится учет новых факторов, определяющих формирование корреспонденций [6].

Литература


1. *Гасников А. В.*, *Гасникова Е. В.*, *Федько О. С.* О возможной динамике в модели ранжирования web-страниц PageRank и модернизированной модели расчета матрицы корреспонденций // Труды МФТИ. 2012. Т. 4, № 2. С. 101–120.
2. *Stouffer S. A.* Intervening opportunities: a theory relating mobility and distance // Amer. Sociolog. Rev. 1940. V. 5. P. 845–867.
3. *Fotheringham A. S.* A new set of special-interaction models: the theory of competing destinations // Envir. & Plan. A. 1983. V. 15. P. 15–36.
4. *Fotheringham A. S.* Modelling hierarchical destination choice // Envir. & Plan. A. 1986. V. 18. P. 401–418.
5. *Gonçalves M. B.*, *Ulysséa-Neto I.* Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models // Envir. & Plan. A. 1993. V. 25. P. 817–826.
6. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.


**Задача к главе 1 и приложению Ю. Е. Нестерова и С. В. Шпирко***

(Дорн—Нестеров—Шпирко, 2012). Прежде всего, напомним основные положения модели стационарной динамики. В рамках модели предполагается, что водители действуют оппортунистически, т. е. выполнен первый принцип Вардропа. Рассмотрим некоторый граф $(V, E)$. В модели каждому ребру $i \in E$ ставятся в соответствие параметры $\bar{f}_i$ и $\bar{t}_i$. Они имеют следующую трактовку: $\bar{f}_i$ — максимальная пропускная способность ребра $i$, $\bar{t}_i$ — минимальные временные издержки на прохождение ребра $i$. Таким образом, сама модель задается графом $(V, E, \bar{f}, \bar{t})$, где $\bar{f} = \{\bar{f}_1, \ldots, \bar{f}_{|E|}\}^T$, $\bar{t} = \{\bar{t}_1, \ldots, \bar{t}_{|E|}\}^T$. Пусть $f$ — вектор распределения потоков по ребрам, инициируемый равновесным распределением потоков по маршрутам, а $t$ — вектор временных издержек, соответствующий распределению $f$. Тогда если транспортная система находится в стационарном состоянии, всегда выполняются неравенства $f \leqslant \bar{f}$ и $t \geqslant \bar{t}$. При этом считается, что если поток по ребру $f_i$ меньше, чем максимальная пропускная способность ребра $\bar{f}_i$, то все автомобили в потоке двигаются с максимальной скоростью, а их временные издержки $t_i$ минимальны и равны $\bar{t}_i$. Если же поток по ребру $f_i$ становится равным пропускной способности ребра $\bar{f}_i$, то временные издержки водителей $t_i$ могут быть сколь угодно большими.



Это удобно объяснить следующим образом. Допустим, на некоторое ребро $i$ стало поступать больше автомобилей, чем оно способно обслужить. Тогда на этом ребре начинает образовываться очередь (пробка). Временные издержки на прохождение ребра $t_i$ складываются из минимальных временных издержек $\bar{t}_i$ и времени, которое водитель вынужден отстоять в пробке. При этом, очевидно, если входящий поток автомобилей на ребро $i$ не снизится до максимально допустимого уровня (пропускной способности ребра), то очередь будет продолжать расти и система не будет находиться в стационарном состоянии. Если же в какой-то момент входящий на ребро $i$ поток снизится до уровня пропускной способности ребра, то в системе наступит равновесие. При этом пробка на ребре $i$ (если входящий поток $f_i$ будет равен $\bar{f}_i$) не будет рассасываться, т. е. временные издержки так и останутся на уровне $t_i$ ($t_i > \bar{t}_i$). Продемонстрируем это на примере из статьи [1].

Рассмотрим (в рамках модели стационарной динамики) граф $(V, E, \bar{t}, \bar{f})$ (см. рис. 6). Пункты 1 и 2 — потокообразующая пара. При этом выполнено

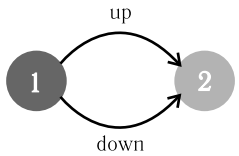



Рис. 6

следующее: $f$ — поток из 1 в 2, $\bar{t}_{\text{up}} < \bar{t}_{\text{down}}$. Тогда зависимость равновесного распределения потоков по маршрутам и издержек водителей от $f$ будет иметь вид:

$$(f_{\text{up}}, f_{\text{down}}) = \begin{cases} \text{нет стационарного равновесного распределения,} & \text{если } f > \bar{f}_{\text{up}} + \bar{f}_{\text{down}}; \\ (\bar{f}_{\text{up}}, f - f_{\text{up}}), & \text{если } \bar{f}_{\text{up}} < f \leqslant \bar{f}_{\text{up}} + \bar{f}_{\text{down}}; \\ (\bar{f}_{\text{up}}, 0), & \text{если } f \leqslant \bar{f}_{\text{up}}; \end{cases}$$

$$G_{\rho_{12}}(f) = \begin{cases} \infty, & \text{если } f > \bar{f}_{\text{up}} + \bar{f}_{\text{down}}; \\ \bar{t}_{\text{down}}, & \text{если } \bar{f}_{\text{up}} < f \leqslant \bar{f}_{\text{up}} + \bar{f}_{\text{down}}; \\ t \in [\bar{t}_{\text{up}}, \bar{t}_{\text{down}}], & \text{если } \bar{f}_{\text{up}} = f; \\ \bar{t}_{\text{up}}, & \text{если } f < \bar{f}_{\text{up}}. \end{cases}$$

Действительно, если выполнено $f < \bar{f}_{\text{up}}$, то все водители будут использовать ребро «up», причем пробок образовываться не будет, так как пропускная способность ребра больше, чем количество водителей. В момент, когда $f = \bar{f}_{\text{up}}$, возможности ребра «up» будут использоваться на пределе. Если

же в какой-то момент величина $f$ станет больше $\bar{f}_{\text{up}}$, то на ребре up начнет образовываться пробка. Она будет расти до тех пор, пока издержки от использования маршрута «up» не сравняются с издержками от использования маршрута «down». В этот момент оставшаяся часть начнет использовать маршрут «down». Если же корреспонденция из 1 в 2 превысит суммарную пропускную способность ребер «up» и «down», то пробки будут расти неограниченно (входящий поток на ребро будет больше, чем исходящий, соответственно количество автомобилей в очереди будет расти постоянно), т. е. стационарное распределение в системе никогда не установится. Более подробно данная задача (и модель) изложена в статье [1].

**а) Парадокс Браесса.** Задан граф $(V, E, \bar{t}, \bar{f})$ (см. рис. 7). При этом выполнено следующее: $\bar{t}_{13} > \bar{t}_{12} + \bar{t}_{23}$, (1,3) и (2,3) — потокообразующие пары, $f_1$ и $f_2$ — соответствующие потоки. Определите $(f_{123}, f_{13})$, $G_{\rho_{13}}(f_1, f_2)$, $G_{\rho_{23}}(f_1, f_2)$. Будет ли наблюдаться «эффект Браесса» для транспортной сети, изображенной на рис. 7, если рассматривать BMW-модель [1]?

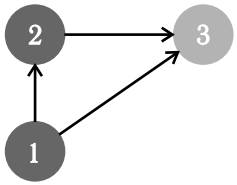



Рис. 7

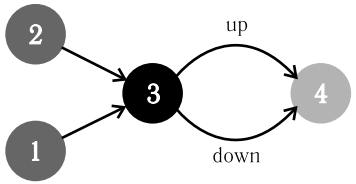



Рис. 8

**б)** Задан граф $(V, E, \bar{t}, \bar{f})$ (см. рис. 8). При этом выполнено следующее: $\bar{t}_{34_{\text{up}}} < \bar{t}_{34_{\text{down}}}$, (1,4) и (2,4) — потокообразующие пары, $f_1$ и $f_2$ — соответствующие потоки, $\bar{f}_{13} > f_1$, $\bar{f}_{23} > f_2$. Определите $(f_{134_{\text{up}}}, f_{134_{\text{down}}}, f_{234_{\text{up}}}, f_{234_{\text{down}}})$, $G_{\rho_{14}}(f_1, f_2)$, $G_{\rho_{24}}(f_1, f_2)$ [1].

**в)** Введем ряд новых обозначений. Пусть $w_k$ — потокообразующая пара графа $(V, E, \bar{t}, \bar{f})$, $P_k$ — множество соответствующих $w_k$ маршрутов, а $t$ — установившийся на графе вектор временных издержек. Тогда временные издержки, соответствующие самому быстрому маршруту из $P_k$, равны: $T_k(t) = \min\limits_{r \in P_k} \langle a_{k,r}, t \rangle$, где $a_{k,r}$ — вектор инцидентности ребер маршруту $r$. Как видно, $T_k(t)$ является функцией от вектора временных издержек $t$. Пусть корреспонденция потокообразующей пары $w_k$ равна $\rho_k$. Тогда если $f^e$ является равновесным распределением потоков для заданного графа $(V, E, \bar{t}, \bar{f})$, множества потокообразующих пар $OD$ и соответствующей им матрицы корреспонденций $W = \|w_i\| = \|\rho_i\|$, а $t^e$ — соответствующий равновесному распределению вектор временных издержек, то $f^e = \sum\limits_{i \in \mathcal{OD}} \vec{f}_i^{\,e}$, где $\vec{f}_i^{\,e}$ — вектор распределения потока, порождаемого потокообразующей парой $i$ ($i \in \mathcal{OD}$). При этом $\vec{f}_i^{\,e}$ удовлетворяет соотношению

$$\vec{f}_i^{\,e} = \rho_i g, \tag{1}$$

где $g \in \partial T(t)$.

В статье [1] приводится следующий результат.

**Теорема.** *Распределение потоков $f^e$ и вектор временных издержек $t^e$ являются равновесными для графа $(V, E, \bar{t}, \bar{f})$, заданных потокообразующих пар и соответствующих им потокам, если $t^e$ является решением оптимизационной задачи*:

$$\max_{t \geqslant \bar{t}} [C(t) - \langle \bar{f}, t \rangle], \tag{2}$$

*где $C(t) = \sum\limits_{i \in \mathcal{OD}} \rho_i T_i(t)$, $f^e = \bar{f} - s^e$, а $s^e$ — (оптимальный) вектор двойственных множителей для ограничений $t \geqslant \bar{t}$ в задаче* (2), *или, другими словами, решение двойственной задачи.*

Докажите эту теорему и проверьте, что ответы в задачах а) и б) удовлетворяют ей.

**г)** Заметим, что если в приложении Ю. Е. Нестерова и С. В. Шпирко положить в формуле (1) значение $\mu$ очень маленьким ($0 < \mu \ll \min\limits_i |c_i|$), то водители будут с равной вероятностью выбирать маршруты, имеющие одинаковые, минимальные (среди всех маршрутов данной потокообразующей пары) временные издержки, а вероятность выбора всех других маршрутов станет пренебрежимо маленькой. Покажите, что аналогичным образом из формулы (19) может быть получена формула (1) из задачи в), а задача (2) — соответственно из задачи (21).

**д)** Заметим, что если бы загрузка сети в задаче а) удовлетворяла выражению $\bar{f}_{23} + \bar{f}_{13} \geqslant f_1 + f_2 > \bar{f}_{23}$, то более эффективной (с точки зрения минимизации суммарных затрат водителей) была бы конфигурация транспортного графа, представленная на рис. 9. Аналогично, если бы в задаче б) выполнялось $\bar{f}_{34_{\text{up}}} + \bar{f}_{34_{\text{down}}} \geqslant f_1 + f_2 > \bar{f}_{34_{\text{up}}} > f_2 > f_1$, то более эффективной была бы конфигурация, изображенная на рис. 10, где

$$\bar{t}_{14}^{\,\text{new}} = \bar{t}_{13} + \bar{t}_{34_{\text{down}}}, \quad \bar{t}_{24}^{\,\text{new}} = \bar{t}_{23} + \bar{t}_{34_{\text{up}}},$$
$$\bar{f}_{14}^{\,\text{new}} = \min(\bar{f}_{13}, \bar{f}_{34_{\text{down}}}), \quad \bar{f}_{24}^{\,\text{new}} = \min(\bar{f}_{23}, \bar{f}_{34_{\text{up}}}).$$

В связи с этим интересно рассмотреть вопрос о возможности эффективно можно решить проблему неэффективности сети при определенных нагрузках с помощью светофоров.

Пусть работа светофоров в транспортной сети определяется матрицей $\Lambda = \|\alpha_{ij}\|_{i,j=1}^{|E|}$, где элемент $\alpha_{ij}$ матрицы равен доле времени от полного цикла

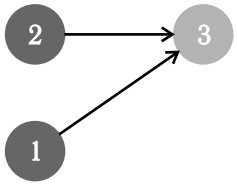



Рис. 9

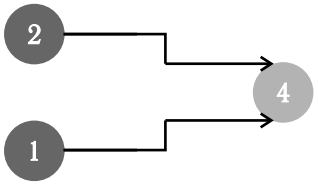



Рис. 10

светофора (на выезде с ребра $i$), во время которой возможен съезд с ребра $i$ на ребро $j$.

Пусть в вершине 3 графа из задачи б) установлен светофор. При этом $\forall i \in \{13,\ 23\}$, $j \in \{34_{\text{up}},\ 34_{\text{down}}\}$: $\alpha_{ij} > 0$. Верно ли, что какая бы матрица $\Lambda$, удовлетворяющая заданным условиям, ни была выбрана, суммарные издержки водителей снизить не удастся?

**е)** Пусть в вершинах 1 и 2 графа из задачи а) установлены светофоры. Покажите, что существует не единственная оптимальная конфигурация матрицы $\Lambda$ такая, что в сети удается избежать парадокса Браесса.

**ж)** Найдите оптимальную конфигурацию матрицы $\Lambda$ в задаче б), если считать, что в вершине 3 установлен светофор.

**з)** Рассматривая модель Нестерова—де Пальмы в терминах потоковых переменных, можно определить суммарные издержки водителей в транспортной сети следующей формулой [2]: $U(f,\ t) = \sum\limits_{i \in E} f_i t_i$, при условии, что $f$ и (соответствующий данному распределению вектор) $t$ удовлетворяют ограничениям модели. Цену анархии для транспортного графа $(V,\ E,\ \bar{t},\ \bar{f})$ и заданной матрицы корреспонденций $W$ определим как [2]:

$$PA(\Gamma,\ W) = \frac{U(f^e,\ t^e)}{\min U(f,\ t)},$$

где пара $(f^e,\ t^e)$ определяется как решение (2) для данного графа, а минимум в знаменателе берется по всем допустимым и согласованным парам $(f,\ t)$. Цена анархии является количественной характеристикой, показывающей, насколько установившееся равновесное распределение потоков по ребрам (маршрутам) далеко от оптимального (с точки зрения минимизации суммарных общественных затрат).

Определите значение $PA$ для графов и соответствующих загрузок из задач а) и б). Покажите, что с помощью оптимального управления светофорами (см. задачи е) и ж)) можно снизить цену анархии.

**и)** Базируясь на статье [3], предложите эффективный численный алгоритм поиска равновесного распределения потоков из п. в).



Литература

1. *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.
2. *Chudak F., Eleuterio V., Nesterov Y.* Static Traffic Assignment Problem. A comparison between Beckmann (1956) and Nesterov & de Palma (1998) models // Conference Paper STRC. 2007. P. 1–23.
3. *Nesterov Yu.* Primal-dual subgradient methods for convex problems // Math. Program., Ser. B. 2009. V. 120. P. 221–259;
`http://ium.mccme.ru/s12/gasnikov-sem-s12.html`

**Задача к главам 1, 2*** (В. И. Швецов, 2012). Как можно объяснить следующее противоречие: согласно п. 1.2.1 главы 1 время в пути по дуге есть строго возрастающая функция от величины потока на дуге (BPR-функция), а согласно п. 2.1.1 главы 2 эта зависимость не является однозначной — есть две ветки, одна из которых соответствует малым и умеренным плотностям и согласуется с отмеченной зависимостью, другая, соответствующая большим плотностям, приводит к убывающей зависимости времени в пути от величины потока?

**Задача к главе 1**** (PTV Vision). Подобно BPR-функциям затрат на прохождение ребер (дуг, дорог) графа транспортной сети от величины потока на ребрах, введите функции затрат на прохождение вершин (узлов, перекрестков) графа транспортной сети. Перенесите результаты главы 1 на случай, когда прохождение узлов графа транспортной сети, так же как и дуг, приводит к затратам. В каком случае задача поиска равновесного распределения потоков (Нэша—Вардропа), сводится к решению задачи выпуклой оптимизации?

Литература

1. *Швецов В. И.* Математическое моделирование транспортных потоков // Автоматика и телемеханика. 2003. № 11. С. 3–46.

**Задача к главе 2*** (О. С. Розанова, 2011). Рассмотрим транспортный поток на бесконечной магистрали без въездов и съездов. Будем считать, что поток описывается LWR-моделью:

$$\frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = 0, \quad \rho(0,\ x) = \rho_0(x).$$

Хорошо известно, что если $Q'(\rho)$ и $\rho_0(x)$ — гладкие функции, то $\rho(t,\ x)$ остается гладким в течение некоторого времени $t_* > 0$, а затем гладкость решения, вообще говоря, теряется.

а) Покажите, что если $Q'(\rho_0(x))$ имеет не более чем линейный рост при $|x| \to \infty$, то для любого $t \in [0,\ t_*)$ решение $\rho(t,\ x)$ может быть найдено как

предел при $\sigma \to 0$

$$\rho_\sigma(t,\,x) = \frac{\int\limits_{-\infty}^{\infty} \rho_0(\xi) \exp\Big(-\frac{1}{2\sigma^2 t}|Q'(\rho_0(\xi))t + \xi - x|^2\Big)d\xi}{\int\limits_{-\infty}^{\infty} \exp\Big(-\frac{1}{2\sigma^2 t}|Q'(\rho_0(\xi))t + \xi - x|^2\Big)d\xi}.$$

б) Найдите время $t_*$.

в) Найдите уравнение, которому удовлетворяет функция $\rho_\sigma(t,\,x)$ до и после момента времени $t_*$.

Литература

1. *Albeverio S.*, *Korshunova A. A.*, *Rozanova O. S.* Probabilistic model associated with the pressureless gas dynamics; `arXiv:0908.2084v2`, 2009.
2. *Albeverio S.*, *Rozanova O.* Suppression of unbounded gradients in SDE associated with the Burgers equation // Proceedings of Amer. Math. Soc. 2010. V. 138, № 1. P. 241–251.

**Задача к главе 2**.** Верно ли, что в теореме 2 п. 2.3.1 главы 2 имеет место оценка:

$$d_k(t) = \frac{2}{\beta_k - \alpha_k}\Big(\frac{D(\alpha_k)}{Q''(\alpha_k)}\theta(\alpha_k - \beta_{k-1}) - \frac{D(\beta_k)}{Q''(\beta_k)}\theta(\alpha_{k+1} - \beta_k)\Big)\varepsilon \ln t + \\ + O\big((\varepsilon \ln t)^{2/3}\big) + O(1), \quad \theta(x) = \begin{cases} 0, & x \leqslant 0; \\ 1, & x > 0? \end{cases}$$

Литература

1. *Гасников А. В.* О скорости разбегания двух подряд идущих бегущих волн в асимптотике решения задачи Коши для уравнения типа Бюргерса // ЖВМ и МФ. 2012. Т. 52, № 6. С. 1069–1071.

**Задача к главам 2, 3 и приложению М. Л. Бланка.** Предложите модель клеточных автоматов многополосного транспортного потока, в которой бы учитывались перестраивания АТС из полосы в полосу (например, для осуществления необходимых маневров в окрестности вершин графа транспортной сети) и которая бы описывала три фазы Кернера. Для проверки последнего свойства можно прибегнуть к численным экспериментам с предложенной моделью.

**Задача к приложению А. А. Замятина и В. А. Малышева: пуассоновский процесс как процесс восстановления.** В приложениях (в теории надежности, теории массового обслуживания) широко используются *процессы восстановления* (*потоки Пальма*). Основным (и наиболее удобным для анализа) представителем таких процессов является *пуассоновский процесс*, который можно определить следующим образом: [1)]

$$K(t) = \max\Big\{k\colon \sum_{i=1}^{k} T_i < t\Big\},$$

где [2)] н.о.р.с.в. $T_i \in \mathrm{Exp}(\lambda)$, т. е. [3)] $P(T_i > t) = e^{-\lambda t}$, $\lambda > 0$. Покажите, что пуассоновский процесс может быть определен таким образом. Будет ли пуассоновский процесс марковским?



**Задача к приложению А. А. Замятина и В. А. Малышева: парадокс времени ожидания.** Автобусы прибывают на остановку в соответствии с пуассоновским процессом с параметром $\lambda > 0$. Вы приходите на остановку в фиксированный момент времени (скажем, в полдень). Каково математическое ожидание времени, в течение которого вы ждете автобуса?

Литература

1. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. М.–Ижевск: ИКИ, 2002.
2. *Феллер В.* Введение в теорию вероятностей и ее приложения Т. 2. М.: УРСС, 2010.
3. *Кельберт М. Я.*, *Сухов Ю. М.* Вероятность и статистика в примерах и задачах. Т. 2. Марковские процессы как отправная точка теории случайных процессов и их приложения. М.: МЦНМО, 2010.

**Задача к приложению А. А. Замятина и В. А. Малышева: замкнутая сеть, теорема Гордона—Ньюэлла, 1967** (Л. Г. Афанасьева, 2007). Рассматривается транспортная сеть, в которой между $N$ станциями курсируют $M$ такси. Клиенты прибывают в $i$-й узел в соответствии с пуассоновским потоком с параметром $\lambda_i > 0$ ($i = 1, \ldots, N$). Если в момент прибытия в $i$-й узел там есть такси, клиент забирает его и с вероятностью $p_{ij} \geqslant 0$ направляется в $j$-й узел, по прибытии в который покидает сеть. Такси остается ждать в узле прибытия нового клиента. Длительности перемещений из узла в узел — независимые случайные величины, имеющие показательное распределение с параметром $\nu_{ij} > 0$ для пары узлов $(i,\,j)$. Если в момент прихода клиента в узел там нет такси, клиент сразу покидает узел.

Считая $p_{ij} = N^{-1}$, $\lambda_i = \lambda$, $\nu_{ij} = \nu$, покажите, что вероятность того, что клиент, поступивший в узел (в установившемся, стационарном режиме

[1)] $K(t)$ — число отказов приборов к моменту времени $t \geqslant 0$ (отказавший прибор сразу же заменяется исправным). Все приборы идентичны, т. е. имеют одинаковое распределение времени безотказной работы. Кроме того, приборы работают независимо друг от друга.

[2)] Параметр $\lambda > 0$ принято называть интенсивностью пуассоновского процесса.

[3)] Напомним важную особенность показательного распределения $\mathrm{Exp}(\lambda)$ «отсутствие последействия»: $P(T_i > t + \tau | T_i > t) = P(T_i > \tau)$. Заметим также, что общие процессы восстановления задаются аналогичной формулой с той лишь разницей, что н.о.р. $T_i \geqslant 0$ п.н. уже не обязательно распределены по показательному закону.

работы сети), получит отказ, равна

$$p_{\text{отказа}}(N,\,M)=\sum_{k=0}^{M}\frac{C_{N-2+k}^{k}\rho^{M-k}}{(M-k)!}\Bigg/\sum_{k=0}^{M}\frac{C_{N-1+k}^{k}\rho^{M-k}}{(M-k)!},\qquad \rho=\frac{N\lambda}{\nu}.$$

Методом перевала покажите справедливость следующей асимптотики при $N\to\infty$:

$$p_{\text{отказа}}(N,\,rN)=1-\frac{2r}{\frac{\lambda}{\nu}+r+1+\sqrt{\left(\frac{\lambda}{\nu}+r+1\right)^2-\frac{4\lambda r}{\nu}}}+O\Big(\frac{1}{N}\Big).$$

Литература

1. *Афанасьева Л. Г.* Очерки исследования операций. М.: Изд-во механико-математического факультета МГУ, 2007.
2. *Федорюк М. В.* Метод перевала. М.: УРСС, 2010.
3. *Афанасьева Л. Г.*, *Булинская Е. В.* Математические модели транспортных систем, основанные на теории очередей // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков, под ред. акад. В. В. Козлова). 2010. Т. 2, № 4(8). С. 6–21.

**Задача к приложению В. А. Малышева и А. А. Замятина: о переполнении серверов** (Н. Д. Введенская, 2011). Предлагается исследовать «коллективное поведение» сильно перегруженных серверов в модели системы с динамической маршрутизацией.

**Введение.** Рассматриваются системы обслуживания, состоящие из нескольких серверов, на которые поступают потоки сообщений. Каждый из потоков может обслуживаться несколькими фиксированными серверами (например, двумя). Время обслуживания сообщения пропорционально его длине. Дисциплина обслуживания FCFS: первым пришел на сервер — первым обслужился. Серверы имеют неограниченные буферы, сообщения стоят в них в очереди. Если сервер, на который сообщение направлено, занят, то оно помещается в буфер и ждет своей очереди. Время, которое сообщение проводит в буфере — время ожидания обслуживания — называется задержкой.

Каждое вновь поступившее сообщение направляется на тот из доступных ему серверов, где его задержка будет минимальной, т. е. на тот сервер, на котором меньше сумма длин сообщений, уже стоящих в очереди. Такую маршрутизацию мы называем.

Предполагается, что поступающие потоки пуассоновские и что задано распределение длин сообщений в потоках.

Мы интересуемся вероятностью больших задержек сообщений, а точнее, асимптотикой вероятности того, что задержка не меньше $n$ при $n\to\infty$.



Такими вопросами занимается теория больших уклонений (основные понятия см. в [1]).

Известно поведение вероятности большой задержки $\omega$ в системе с одним сервером, на который поступает пуассоновский поток заданной интенсивности:

$$P[\omega\geqslant n]\sim\exp(-nI). \tag{3}$$

Здесь $I$ — функционал действия — зависит от интенсивности потока $\lambda$, от скорости работы сервера и от распределения длин поступающих сообщений. Обычно смотрят именно на $I$,

$$I=\lim_{n\to\infty}\frac{-1}{n}\ln P[\omega\geqslant n]$$

Отметим, что, например, в случаях, когда распределение длин потоков экспоненциально или длина постоянна, $I$ выписывается явно.

В ряде работ (см. [2] и приведенную там литературу) рассматривалась следующая система.

Имеется два прибора и три потока. Сообщения первого потока направляются на первый прибор, сообщения второго потока — на второй, а сообщения третьего — на тот из приборов, на котором для пришедшего сообщения будет меньше задержка. Скорости работы приборов равны. Изучается поведение задержки сообщений третьего потока $\omega_3$. Оказывается, что и в этом случае вероятности больших задержек оцениваются выражением вида (3). Зная интенсивность потоков и функции распределения длин сообщений в потоках, можно выписать

$$I=\lim_{n\to\infty}\frac{-1}{n}\ln P[\omega_3\geqslant n]. \tag{4}$$

Причем оказывается, что выражение для вероятности задержки зависит от того, может ли третий поток уравновесить нагрузку на приборы или нет.

В [2] приведены результаты моделирования работы такой системы.

Неожиданно оказалось, что логарифм вероятности задержек $\omega_3$ уже при небольших значениях $n$, $n\sim 10$–$20$, хорошо аппроксимируется с точностью до константы выражением (4), т. е. что $\ln P[\omega_3\geqslant n]\sim\ln C-nI$ ($P[\omega_3\geqslant n]\sim C\exp(-nI)$ ).

**Система с $k$ серверами и $k$ потоками.** В [3] рассмотрена циклическая система с $k$ расположенными по кругу одинаковыми серверами, на которые поступает $k$ одинаковых пуассоновских потоков, каждый из которых обслуживается парой соседних серверов (так что на каждый сервер могут поступать сообщения из двух потоков). Предполагается, что система не перегружена. Например, если скорость работы приборов равна 1 и средняя длина сообщений тоже равна 1, то интенсивность каждого из потоков $\lambda<1$.

Если в некотором потоке произойдет флуктуация и за короткое время придет очень много сообщений, то на двух серверах, обслуживающих эти сообщения, образуются длинные очереди, серверы перегрузятся, а сообщения из других потоков, на которых не было флуктуаций, будут поступать на другие серверы.

Мы интересуемся вероятностью того, что в одном из потоков произошла сильная флуктуация. Конечно, задержка сообщений в этом потоке будет большой.

И в этом случае вероятности больших задержек оцениваются выражением вида (3).

Оказалось, что характер поведения такой системы (при фиксированном $k$, $k \geqslant 3$) зависит от интенсивности входных потоков: если интенсивность потоков $\lambda$ меньше некоторого «критического» значения, $\lambda < \lambda_{\text{cr}} < 1$, то, как правило (т. е. с большой вероятностью), флуктуация происходит только в этом одном потоке. А вот если $\lambda_{\text{cr}} < \lambda < 1$, то вероятнее всего, что флуктуация произошла и во всех остальных потоках, а, скажем, не в двух или в $k-1$ потоках. Т. е. либо задержка велика для сообщений из одного потока, либо она велика во всех потоках.

**Постановка задачи.** Интересно было бы промоделировать работу такой системы, например для $k = 3, 4, 5$, и посмотреть, как в зависимости от интенсивности входных потоков $\lambda$ ведут себя логарифмы вероятностей задержки бóльшей чем $1, 2, \ldots, n$ (где $n = 10, 20, 50$) на одном из серверов и какова при этом вероятность того, что минимальная задержка на серверах превосходит соответствующие значения. Конечно, вероятность даже сравнительно небольшой задержки велика и моделировать надо приход порядка $10^6$–$10^7$ сообщений.

В случаях экспоненциального распределения длин сообщений и постоянной длины в [3] приведены формулы для $I$ и значения $\lambda_{\text{cr}}$.

Данную задачу можно рассматривать как модель следующей ситуации: в некий центр ведут $k$ радиально расположенных дорог, у въездов сужения. Потоки въезжающих машин с разных сторон в среднем имеют примерно одинаковую интенсивность, но возможны флуктуации интенсивности. При этом каждый из водителей может воспользоваться одной из двух близлежащих дорог. (Например, 4 дороги: с севера, востока, юга и запада, а выбирается дорога из пары СВ, ВЮ и т. д.) Каков «режим» пробок на въездах?

Отметим, что в [4] рассматривается другое устройство системы обслуживания, в которой тоже имеет место «синхронизация» перегрузок.

Литература

1. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. М: Наука, 1972.



2. *Duffy K., Pechersky E. A., Suhov Y. M, Vvedenskaya N. D.* Using estimated entropy in a queueing system with dynamic routing // Markov Process and Related Fields. 2007. V. 13, № 1. P. 57–84.
3. *Введенская Н. Д., Печерский Е. А.* Кольцо взаимодействующих серверов; спонтанное возникновение коллективного поведения при больших флуктуациях // Проблемы передачи информации. 2008. Т. 44, № 4. С. 101–117.
4. *Введенская Н. Д.* Конфигурация перегретых серверов при динамической маршрутизации // Проблемы Передачи Информации. 2011. Т. 47, № 3. С. 80–95.

**Задача к приложению А. М. Райгородского: обобщенная схема размещений** (В. Ф. Колчин, 1984). **а)** Пусть для неотрицательных целочисленных с.в. $\eta_1, \ldots, \eta_N$ существуют независимые одинаково распределенные с.в. $\xi_1, \ldots, \xi_N$, такие, что

$$P(\eta_1 = k_1, \ldots, \eta_N = k_N) = P(\xi_1 = k_1, \ldots, \xi_N = k_N \mid \xi_1 + \ldots + \xi_N = n). \quad (*)$$

Введем независимые одинаково распределенные с.в. $\xi_1^{(r)}, \ldots, \xi_N^{(r)}$, где $r$ — целое неотрицательное число и

$$P(\xi_1^{(r)} = k) = P(\xi_1 = k \mid \xi_1 \neq r), \qquad k = 0, 1, \ldots$$

Пусть $p_r = P(\xi_1 = r)$ и $S_N = \xi_1 + \ldots + \xi_N$, $S_N^{(r)} = \xi_1^{(r)} + \ldots + \xi_N^{(r)}$. Пусть $\mu_r(n, N)$ — число тех с.в. $\eta_1, \ldots, \eta_N$, которые приняли значение $r$.

Покажите, что с.в. типа $\mu_r(n, N)$ можно изучать с помощью *обобщенной схемы размещений*: для любого $k = 0, 1, \ldots, N$

$$P(\mu_r(n, N) = k) = C_n^k p_r^k (1 - p_r)^{N-k} \frac{P(S_{N-k}^{(r)} = n - kr)}{P(S_N = n)}.$$

Напомним, что в классической схеме размещений $n$ различных частиц по $N$ различным ячейкам было доказано, что распределение заполнений ячеек $\eta_1, \ldots, \eta_N$ имеет вид

$$P(\eta_1 = k_1, \ldots, \eta_n = k_n) = \frac{n!}{k_1! \ \ldots \ k_N!\, N^n},$$

где $k_1, \ldots, k_n$ — неотрицательные целые числа, такие, что $k_1 + \ldots + k_n = n$. Если положить $\xi_1, \ldots, \xi_N \in Po(\lambda)$ — н.о.р. ($\lambda > 0$ произвольно), то получим $(*)$.

**б)** * Дан случайный граф [1] (модель Эрдёша–Реньи) $G(n, p)$. Пусть $p = (c\ln n)/n$. Покажите, что при $c > 1$ граф $G(n, p)$ почти наверное связен [2], а при $c < 1$ — почти наверное не связен.

**в)** ** Пусть $p \geqslant \sqrt{(2\ln n)/n}$, причем длина (вес) $r_{ij}$ каждого появившегося ребра есть независимая от того, какие еще пары вершин соединены ребрами и какие длины у этих ребер, случайная величина, имеющая равномерное распределение на отрезке $[0, 2r]$. Покажите, что тогда почти наверное граф $G(n, p)$ имеет гамильтонов цикл, причем длина почти всех гамильтоновых циклов стабилизируется (имеет место плотная концентрация) около $nr$.

Литература

1. *Перепелица В. А.* Асимптотический подход к решению некоторых экстремальных задач на графах // Проблемы кибернетики. 1973. Т. 26. С. 291–314.
2. *Колчин В. Ф.* Случайные графы. М.: Физматлит, 2004.
3. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.

**Задача к приложениям А. В. Колесникова, Е. В. Гасниковой: принцип концентрации площади сферы** (А. Пуанкаре, 1911). Покажите, что если в многомерном шаре задано равномерное распределение вероятностей и согласно этому распределению вероятностей сгенерировано два случайных вектора, то с вероятностью, близкой к единице, концы этих векторов будут лежать почти на границе шара и эти два случайных вектора будут почти ортогональны.

**Указание.** Нетривиально второе утверждение (про ортогональность). Для того чтобы его установить, покажите, что доля от площади всей сферы $S_r^n$ (радиуса $r$) в $\mathbb{R}^n$ ($n \geqslant 3$), которую занимает площадь сегмента, проектирующегося в отрезок $[a, b]$, скажем, оси $x_1$, равна

$$P[a, b] = \frac{\int\limits_a^b \left(1 - \left(\frac{x}{r}\right)^2\right)^{\frac{n-3}{2}} dx}{\int\limits_{-r}^{r} \left(1 - \left(\frac{x}{r}\right)^2\right)^{\frac{n-3}{2}} dx}.$$

[1] Даны $n$ вершин, любые две вершины соединены ребром с вероятностью $p$ независимо от того, какие еще пары вершин соединены ребрами. Таким образом, $q = 1 - p$ есть вероятность отказа ребра. По сути, в задаче приведен некий порог $\bar{q}$ для $q$ (по аналогии со статистической механикой иногда говорят, что этот порог характеризует «фазовый переход случайного графа»). Если в полном графе на $n$ вершинах ребра отказывают с вероятностью, «большей» $\bar{q}$, то транспортная система почти наверное разрушится, если же ребра отказывают с вероятностью, «меньшей» $\bar{q}$, то транспортная система (несмотря на то, что может потерять много ребер) почти наверное сохранит свое основное свойство — возможность добраться по ребрам из любой вершины в любую другую.

[2] Из любой вершины можно добраться в любую другую по ребрам. Словосочетание «почти наверное» означает, что вероятностная мера тех графов, для которых это не так, стремится к нулю с ростом $n$.



Фиксируя $r = 1$ и устремляя $n$ к бесконечности, получите

$$P[-\delta, \delta] \sim 1 - \sqrt{\frac{\pi}{2}} \exp\left(-\frac{\delta^2 n}{2}\right).$$

В статистической физике

$$\sum_{i=1}^{n} V_i^2 = \frac{2E_n}{m} \sim n.$$

Поэтому если известно, что вектор скоростей молекул газа равномерно распределен по поверхности постоянной энергии [1], то для того чтобы найти (следуя Максвеллу) распределение компонент вектора скорости, скажем $V_1$, нужно осуществить термодинамический скейлинг (термодинамический предельный переход) при $n \to \infty$, $r = \sigma n^{1/2} \to \infty$:

$$P[a, b] = \frac{\int\limits_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx}{\int\limits_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx} = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

Таким образом, получаем нормальный закон распределения Максвелла в статистической физике.

Литература

1. *Зорич А. В.* Математический анализ задач естествознания. М.: МЦНМО, 2008. С. 48–56.

**Задача к приложениям Е. В. Гасниковой, А. В. Колесникова, А. М. Райгородского: изопериметрическое неравенство и принцип концентрации меры** (П. Леви, 1919). Число $\mu_f$ называют медианой функции $f$, если

$$\mu(\vec{x} \in S_1^n : f(\vec{x}) \geqslant \mu_f) \geqslant \frac{1}{2} \quad \text{и} \quad \mu(\vec{x} \in S_1^n : f(\vec{x}) \leqslant \mu_f) \geqslant \frac{1}{2},$$

где $\mu(d\vec{x})$ — равномерная мера на единичной сфере $S_1^n$ в $\mathbb{R}^n$. Пусть $A$ — измеримое (борелевское) множество на сфере $S_1^n$. Через $A_\delta$ будем обозначать $\delta$-окрестность множества $A$ на сфере $S_1^n$. Предположим теперь, что в некотором царстве, расположенном на $S_1^3$, царь предложил царице Дидоне построить «огород с заданной длиной забора». Царица хочет, чтобы ее огород при заданном периметре имел наибольшую площадь. Таким

[1] Равномерное распределение на поверхности постоянной энергии возникло из-за того, что инвариантной (и предельной по эргодической гипотезе) мерой для гамильтоновой системы будет как раз равномерная мера Лиувилля (фазовый объем сохраняется). Поскольку выполняется закон сохранения энергии, то система «живет» на поверхности постоянной энергии. Следовательно, носитель инвариантной меры сосредоточен именно на этой поверхности.

образом, царице надо решить изопериметрическую задачу (такие задачи обычно рассматриваются в курсах вариационного исчисления). Решение этой задачи хорошо известно — «круглый огород». Для нас же полезно рассмотрение двойственной задачи, имеющей такое же решение: при заданной площади огорода спроектировать его так, чтобы он имел наименьшую длину забора, его ограждающего. Используя решение обобщения двойственной задачи на случай $n \geqslant 3$, покажите, что если $\mu(A) \geqslant 1/2$, то

$$\mu(A_\delta) \geqslant 1 - \sqrt{\frac{\pi}{8}} \exp\left(-\frac{\delta^2 n}{2}\right).$$

Пусть теперь на $S_1^n$ задана функция с модулем непрерывности

$$\omega_f(\delta) = \sup\{|f(\vec{x}) - f(\vec{y})| \colon \rho(\vec{x}, \vec{y}) \leqslant \delta,\ \vec{x}, \vec{y} \in S_1^n\}.$$

Тогда

$$\mu(\vec{x} \in S_1^n \colon |f(\vec{x}) - \mu_f| \geqslant \omega_f(\delta)) \leqslant \sqrt{\frac{\pi}{2}} \exp\left(-\frac{\delta^2 n}{2}\right).$$

Можно показать, что при весьма естественных условиях медиана асимптотически близка к среднему значению (математическому ожиданию). Аналогичное неравенство можно получить (М. Талагран, 1996), например, для модели случайных графов (Эрдёша—Реньи). Можно также исследовать плотную концентрацию около среднего значения различных функций на случайных графах: число независимости, хроматическое число и т. п.

Литература

1. *Ledoux M.* Concentration of measure phenomenon. Providence, RI: Amer. Math. Soc., 2001. (Math. Surveys Monogr. V. 89).
2. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.

**Задача к приложениям Е. В. Гасниковой, А. В. Колесникова, А. М. Райгородского: изопериметрическое неравенство Талаграна и его приложения** (М. Талагран, 1996). **а)*** Пусть заданы множества $\Omega_i$, $i = 1, \ldots, n$, элементарных исходов. На этих множествах заданы вероятностные меры $P_i$, $i = 1, \ldots, n$. Положим

$$\Omega = \prod_{i=1}^{n} \Omega_i, \qquad P = \prod_{i=1}^{n} P_i.$$

Введем взвешенную метрику Хэмминга:

$$d_\alpha(\vec{x}, \vec{y}) = \sum_{x_i \neq y_i} \alpha_i \Bigg/ \sqrt{\sum_{i=1}^{n} \alpha_i^2}$$



и определим $d_\alpha(\vec{x}, A) = \min\limits_{\vec{y} \in A} d_\alpha(\vec{x}, \vec{y})$, $\rho(\vec{x}, A) = \sup\limits_{\vec{x} \in \mathbb{R}^n} d_\alpha(\vec{x}, A)$. Пусть $A \subset \Omega$. Определим $t$-окрестность ($t \geqslant 0$) множества $A$ по формуле

$$A_t = \{\vec{x} \in \Omega \colon \rho(\vec{x}, A) \leqslant t\}.$$

Покажите, что тогда справедливо следующее неравенство:

$$P(A)(1 - P(A_t)) \leqslant \exp\left(-\frac{t^2}{4}\right).$$

**б)** ** Пусть в сельском районе, имеющем форму квадрата со стороной 1, находится $n$ домов ($n \gg 1$), размерами которых можно пренебречь по сравнению с линейным размером района. Будем считать, что при строительстве домов застройщик случайно (согласно равномерному распределению $R[0, 1]^2$) и независимо выбирал их местоположения. Почтальону необходимо обойти все $n$ домов ровно по одному разу (от любого дома почтальон может направиться к любому другому по прямой). Обозначим через *TSP* длину наикратчайшего из таких путей (кратчайший гамильтонов путь).

Используя п. а), покажите, что найдутся такие постоянные $c > 0$ и $\beta > 0$, не зависящие от $n$, что

$$P(|TSP - E[TSP]| \geqslant t) \leqslant \exp\left(-\frac{t^2}{4c}\right), \quad \text{где} \quad E[TSP] \sim \beta\sqrt{n}.$$

**в)** *** Пусть в условиях п. б) требуется построить систему дорог минимальной суммарной длины *SteinerTree*, по которой можно было бы добраться из любого дома в любой другой (дерево Штейнера с минимальной суммарной длиной ребер). Получите неравенство о плотной концентрации с.в. *SteinerTree* в окрестности ее математического ожидания, аналогичное неравенству п. б). Как себя асимптотически ведет $E[SteinerTree]$ при $n \to \infty$?

Литература

1. *Ledoux M.* Concentration of measure phenomenon. Providence, RI, Amer. Math. Soc., 2001. (Math. Surveys Monogr. V. 89.)
2. *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином, 2007.
3. *Dubhashi D. P., Panconesi A.* Concentration of measure for the analysis of randomized algorithms. Cambridge University Press, 2009.
4. *Gromov M.* Metric Structures for Riemannian and Non-Riemannian Spaces. With Appendices by M. Katz, P. Pansu, and S. Semmes. Boston, MA: Birkhäuser, 2007. Chapter $3\frac{1}{2+}$. (Modern Birkhäuser Classics).

# Задача Штейнера и задачи на графах транспортных сетей

**Е. Г. Молчанов**

**(при участии А. М. Валуева, Р. А. Гимадеева, Ю. В. Дорна, А. А. Зайцева, С. П. Тарасова, П. А. Шишкина)**

Приведенные задачи разделяются на два типа: задачи, связанные с задачей Штейнера на плоскости, и задачи на транспортных графах. Последние обычно являются (возможно, нетривиальными) случаями, разновидностями и усложнениями нескольких стандартных задач: задачи поиска минимального остовного дерева, кратчайшего пути, максимального потока, а также транспортной задачи. Этот задачный раздел будет состоять из нескольких частей, во главе каждой из которых будет одна из таких «стандартных» задач, далее — задачи, которые связаны с первоначальной «стандартной» задачей.

Если в задачах требуется оценить время работы программы, будем считать, что она выполняется на компьютере, совершающем порядка $10^{10}$ операций с плавающей точкой в секунду. Остальные параметры этого компьютера (например, оперативную память) будем считать неограниченными.

## Часть 1. Минимальная сумма расстояний

**Задача 1.** Согласно плану в районе строятся несколько домов, в каждом из которых будет проживать определенное количество человек. Нужно построить парковку, чтобы сумма расстояний, пройденных каждым человеком до парковки, была минимальной. Размерами домов можно пренебречь по сравнению с расстояниями между домами.

а) Где нужно построить парковку, если есть три одинаковых дома?

б) Докажите, что если все дома не лежат на одной прямой, то у задачи ровно одно решение.

**Задача 2.** Для решения задачи 1 о минимальной сумме расстояний можно воспользоваться следующей системой: на плоский немного шероховатый фанерный лист кладется карта района и в центрах домов просверливаются маленькие дырочки. Через дырочки продеты нерастяжимые нити, к которым прикреплены грузики с весом, пропорциональным количеству жителей в соответствующем доме. Начала всех нитей завязаны в один узел. Будем считать узел достаточно большим: он не будет пролезать в дырочки.

а) Докажите, что (при условии, что все точки не лежат на одной прямой) если «отпустить» все грузики, узел остановится в точке с минимальной суммой.

б) Может ли узел застрять в дырочке?

в) На основе описанной системы предложите итерационный процесс решения задачи о минимальной сумме расстояний и оцените его скорость сходимости.



**Задача 3.** Предположим теперь в задаче 1, что от домов до парковки можно строить только «вертикальные» и «горизонтальные» дорожки (т. е. дорожки направления юг—север и запад—восток).

а) Где теперь нужно строить парковку, чтобы сумма расстояний до домов по дорожкам была минимальная?

б) Может ли у пункта а) быть не единственное решение? Как в общем случае выглядит область с минимальной суммой расстояний?

**Указание.** Сведите эту задачу к задаче о поиске минимального расстояния в одномерном случае (все дома лежат на одной прямой).

## Часть 2. Минимальное остовное дерево

**Задача 4.** $N$ точек-пунктов (вершин графа) соединены между собой дорогами. Каждая дорога соединяет две точки и имеет некоторую длину. Известно, что из любой такой точки до любой другой можно добраться по дорогам. Нужно реконструировать некоторое количество дорог, так, чтобы от любой точки можно было добраться до любой другой только по реконструированным дорогам и чтобы сумма длин дорог была минимальной.

*Иными словами, в графе со взвешенными ребрами нужно найти минимальное остовное дерево.*

а) Покажите, что при решении этой задачи можно считать, что произвольные две точки соединены дорогой (т. е. граф полный).

б) Какую сложность имеют стандартные алгоритмы Прима и Краскала; как эта сложность зависит от способов хранения информации, нужной для алгоритмов (например, матрицы расстояний)?

в) Оцените время работы стандартных алгоритмов, если количество точек-вершин равно 1000.

**Задача 5.** а) В задаче 4 проектировщик решил искать нужные дороги среди гамильтоновых циклов, т. е. среди маршрутов, выходящих из какой-то вершины и последовательно проходящих через все остальные вершины. Покажите, что найденное проектировщиком решение может в любое количество раз по длине превосходить минимальное (уточнить, что вкладывается в слово «минимальное»).

б) Оцените время поиска гамильтонова цикла. Примите количество вершин равным 1000 (100, 10). Сравните его с временем поиска минимального остовного дерева.

**Задача 6.** Дорожная сеть города была построена следующим образом: каждый год строился новый микрорайон в последовательности, предусмотренной генпланом, и он соединялся дорогой с тем микрорайоном, до которого расстояние было наименьшим. После постройки города микрорайоны были связаны между собой системой дорог, образующей дерево. Будем считать, что расстояния между каждыми двумя микрорайонами

известны до постройки и никаких географических ограничений на эти расстояния нет.

Сумма длин всех построенных дорог зависит от того, в каком порядке были построены микрорайоны. Докажите, что отношение максимально возможной суммы к минимально возможной может быть сколь угодно большим числом.

**Задача 7.** В некоторых местах построены несколько микрорайонов и несколько станций метро. Будем считать, что от любой станции можно на метро добраться до любой другой. Длины возможных дорог от микрорайонов до станций метро и между микрорайонами известны. Нужно построить минимальную по суммарной длине систему дорог так, чтобы из любого микрорайона можно было добраться до любого другого, возможно, воспользовавшись метро.

Как свести эту задачу к задаче 4 о минимальном остовном дереве?

**Указание.** Будем считать, что от любой станции метро до любой другой можно добраться за нулевое время. Тогда, считая все станции метро одной вершиной, определите расстояния от метро до микрорайонов нужным образом.

## Часть 3. Задача Штейнера

**Задача 8.** Несколько городов на плоскости нужно соединить дорогами так, чтобы из любого города можно было бы доехать до любого другого по построенным дорогам и суммарная длина построенных дорог была бы минимальна. Расстояния считаются в обычной, евклидовой метрике, при построении дорог можно строить дополнительные узлы — пересечения нескольких дорог.

а) Решите задачу для четырех городов, расположенных в вершинах квадрата. Объясните, почему найденное решение не обладает всеми теми симметриями, которыми обладает исходная постановка задачи.

б) Докажите, что в оптимальной системе дорог в любом «вспомогательном» узле должны сходиться ровно три дороги под углами в 120° друг к другу.

в) Решите задачу для десяти городов, расположенных в вершинах правильного 10-угольника.

**Задача 9.** Сетью Штейнера называется система прямых дорог, такая, что любая дорога соединяет напрямую какие-то две точки, где точками могут быть города, а также вспомогательные узлы, обладающая следующими свойствами:

1. В каждом вспомогательном узле сходятся ровно три дороги под углами в 120° друг к другу.



2. Из каждого города должно выходить не более трех дорог; если из города выходят две дороги — они должны образовывать угол не менее 120°, если три — все образованные ими углы должны быть по 120° .

а) Обязана ли такая сеть обладать минимальной суммой расстояний среди всех возможных сетей дорог (таких, что из любого города можно доехать до любого другого)?

б) Покажите, что сеть с минимальной суммой расстояний нужно искать среди возможных сетей Штейнера.

**Задача 10.** Проектировщик соединил некоторые пары городов на плоскости прямыми дорогами так, чтобы из любого города в любой другой можно было бы доехать по построенным дорогам, и суммарная длина построенных дорог при этом была бы минимальной. Будем считать, что все дороги пересекаются вне города на разных уровнях — съехать с одной дороги на другую можно только в городах.

а) Докажите, что в оптимальном решении этой задачи не будет дорог, пересекающихся вне городов.

б) Докажите, что решение этой задачи по суммарной длине построенных дорог не более чем в два раза превосходит длину минимальной сети Штейнера.

**Замечание.** Отношение длины минимального остовного дерева (веса ребер — расстояния на плоскости) к длине минимальной сети Штейнера называется числом Джильберта—Поллака. В п. б) предстоит доказать, что это отношение не превосходит числа 2. Это грубая оценка сверху, точная [1] оценка — $\frac{2}{\sqrt{3}}$.

в) Приведите пример, когда оценка $\frac{2}{\sqrt{3}}$ достигается.

**Замечание.** Построение минимальной сети Штейнера является *NP*-трудной задачей, поэтому нахождение минимального остовного дерева среди графа городов (вершины — города, расстояния между парами вершин считаются в обычной, евклидовой метрике) является полиномиальным *приближенным* решением задачи Штейнера с коэффициентом приближения $\frac{2}{\sqrt{3}}$ (найденная длина дорог минимального остовного дерева не более чем в $\frac{2}{\sqrt{3}}$ превосходит оптимальное решение).

## Часть 4. Кратчайший путь

**Задача 11.** Пусть есть некоторая система дорог, каждая из которых соединяет 2 пункта-вершины, причем для каждой дороги известно время движения по этой дороге. Требуется доехать из пункта А в пункт Б за минимальное время. *Иными словами в графе со взвешенными ребрами требуется найти кратчайший путь из одной вершины в другую.*

[1] Хотя и принято считать, что «точность» этой оценки доказана около 20 лет назад, в известном нам оригинальном доказательстве имеются «лакуны».

а) Оцените сложность алгоритма Дейкстры и время его работы, если количество вершин-пунктов равно 1000.

б) Будем считать, что время движения из одного пункта до другого зависит от направления (т. е. *граф — ориентированный*). Как изменится время работы алгоритма пункта а)?

**Замечание.** Разумеется, при поиске кратчайшего расстояния на реальной сети дорог всегда рассматривают ориентированные графы. В последующих задачах этого параграфа, если не указано обратное, графы будут ориентированными.

**Задача 12.** В задаче 11 будем считать, что время проезда по определенному участку пути зависит от того, в какое время вы въехали на этот участок. Для упрощения время проезда каждого участка мы будем считать целым неотрицательным числом $T_{ij}(t_0)$: мы движемся от $i$-го перекрестка к $j$-му, начиная движение в момент времени $t_0$ (очевидно, $t_0$ также целое неотрицательное).

Как свести эту задачу к задаче 11 о кратчайшем пути в графе, веса ребер которого не изменяются со временем?

**Указание.** Постройте новый граф, сопоставив одной вершине исходного графа несколько вершин, каждая из которых отвечает своему времени.

**Замечание.** Построенный алгоритм будет работать при условии

$$\forall i,\ j,\ \ t_1 > t_0 > 0 \quad T_{ij}(t_0) - T_{ij}(t_1) \leqslant t_1 - t_0.$$

Объясните, что означает это условие и можно ли требовать выполнения этого условия в реальных транспортных графах?

**Задача 13.** $M$ автомобилям, первоначально находящимся в каких-то пунктах (вершинах графа), необходимо встретиться в одном месте (может быть, на дороге, соединяющей две вершины) и передать друг другу важную информацию. К сожалению, машинам негде припарковаться, поэтому останавливаться не разрешено, но разрешено поворачивать или разворачиваться в любом направлении в каждой вершине. Будем считать, что все машины движутся с одинаковыми скоростями, а все дороги, соединяющие некоторые пары перекрестков, требуют единицу времени для их прохождения.

Найдите минимальное время, необходимое для встречи.

**Задача 14.** Из пункта А в пункт Б (вершин транспортного графа) нужно добраться с помощью общественного транспорта, возможно, с пересадками. Каждый вид общественного транспорта ходит от какого-то одного пункта-вершины до какого-то другого строго по расписанию; расписание представляет собой таблицу времен отправлений из начального пункта и соответствующих им времен прибытий в конечный. Пересадка на другой вид транспорта в промежуточном пункте возможна, если время отправления превосходит время прибытия в этот пункт на предыдущем виде транспорта. Будем считать, что ровно в 12:00 мы оказались в пункте А. Требуется найти минимальное время, за которое мы сможем добраться до пункта Б, пользуясь общественным транспортом.



Приведите алгоритм решений этой задачи (например, метод ветвей и границ) и оцените сложность его работы.

**Задача 15.** В задаче 11 будем считать, что на некоторых вершинах могут быть установлены ограничения, запрещающие поворот с одной дороги на другую в этой вершине. Как свести эту задачу к задаче 11 о минимальном графе, в котором не существует запрещенных поворотов?

**Указание.** Как и в задаче 12, нужно превратить одну исходную вершину в несколько.

**Замечание.** См. ниже задачу «Преобразования запрещенных маневров», являющуюся обобщением этой задачи.

**Задача 16.** Пусть в условиях задачи 11 кратчайший путь ровно один. Требуется найти путь, который будет являться

а) вторым по длине (времени прохождения);

б) $n$-м по длине (времени прохождения).

**Задача 17.** Дана радиально-кольцевая транспортная сеть, состоящая из некоторого количества колец, находящихся на равном расстоянии друг от друга, и нечетного количества радиальных дорог, каждая из которых образует одинаковый угол со своими двумя «соседями».

а) Опишите как можно больше субоптимальных ациклических маршрутов (вторых, третьих по длине и т. д.), проложенных из концов «почти противоположных» радиальных дорог.

б) Рассмотрим в качестве характеристики участка транспортной сети не ее длину, а время прохождения. Определить, при каком (закономерном) уменьшении скорости движения от периферии к центру оптимальными (по затратам времени) маршрутами становится второй (по длине), третий и т. д. субоптимальный маршрут.

**Задача 18.** В городе были построены дороги, соединяющие некоторые перекрестки (вершины графа транспортной сети). Длина каждой дороги известна. Строителям для отчета требуется найти кратчайшие пути между каждой парой перекрестков.

а) Оцените сложность алгоритма и время работы соответствующей программы (1000 перекрестков), если разрешается пользоваться только алгоритмом Дейкстры поиска кратчайшего расстояния (правда, много раз).

б) Предложите более быстрый алгоритм (например, алгоритм Флойда—Уоршола) и также оцените его время работы при тех же условиях.

в) В каком смысле поиск кратчайших путей (алгоритмом Флойда) эквивалентен вычислению в идемпотентном полуполе (см. п. 2.1.3 главы 2) некоторой степени (укажите степень и предложите эффективный способ возведения в эту степень) матрицы, заполненной длинами всевозможных ребер-дорог? Как действовать, если какие-то вершины не соединены ребром?

**Задача 19.** Сопоставим каждому перекрестку (вершине графа транспортной сети) некоторое число — потенциал $P_i$ и определим новое расстояние (время прохождения) между перекрестками $i$ и $j$:

$$\mathrm{New}_{ij} = \mathrm{Old}_{ij} + P_j - P_i,$$

где $\mathrm{Old}_{ij}$ — «старое» расстояние между вершинами $i$ и $j$.

а) Докажите, что кратчайший путь в графе с «новыми» расстояниями совпадает с кратчайшим путем в графе со «старыми» расстояниями.

б) Какие потенциалы нужно расставить в вершины, чтобы алгоритм Дейкстры нашел нужный кратчайший путь, не добавляя в множество рассмотренных вершин «лишние» вершины, т. е. не лежащие на кратчайшем пути?

**Задача 20.** В городе были выделены несколько перекрестков (вершин графа транспортной сети) и были подсчитаны все расстояния от выделенных перекрестков до всех остальных.

а) Как, используя эту информацию, оценить снизу кратчайшее расстояние между любыми двумя перекрестками (вершинами)?

б) Какие перекрестки (вершины) мы должны выделить (их количество заранее оговорено), чтобы оценка из п. а) была более точной?

**Задача 21.** Так же как и в задаче 12, будем считать, что время проезда по определенному участку пути зависит от того, в какой момент времени мы въехали на этот участок. Однако в данной задаче мы не знаем точных прогнозов на время проезда каждого участка в будущем. Вместо этого мы будем считать время проезда участка известной дискретной случайной величиной, принимающей 3 натуральных значения, соответствующих времени движения в свободном режиме, предзаторном и заторном.

а) Покажите, как с помощью модификаций известных алгоритмов нахождения кратчайших путей можно посчитать математическое ожидание времени в пути между двумя точками.

б) Как необходимо переписать условие на корректность работы алгоритма (см. замечание к задаче 12) и можно ли требовать выполнения этого модифицированного условия для реальных транспортных графов?

## Часть 5. Максимальный поток

**Задача 22** (о максимальном потоке). Будем считать, что из пункта А в пункта Б едет (возможно, через другие пункты) непрерывный поток



автомобилей величиной $X$ автомобилей в единицу времени. Также будем считать, что у каждой дороги, соединяющей два пункта, пропускная способность ограничена числом $c_{ij}$ машин в единицу времени. Необходимо найти, какой максимальный поток можно запустить из пункта А в пункт Б с учетом пропускных способностей дорог.

а) Запишите эту задачу как задачу линейного программирования, напишите двойственную к ней задачу.

**Указание.** Считайте граф ориентированным. Используйте матрицу инциденций графа.

б) Пусть $c_{ij}$ — неотрицательные целые числа. Покажите, что в оптимальном решении задачи поток, проходящий через каждую дорогу, будет также целым.

**Указание.** Используйте вид матрицы инциденций графа (в каждом столбце только два числа не равны нулю; эти числа: 1 и $-1$).

в) Как при доказательстве целочисленности решения (см. пункт б)) можно использовать теорему Гофмана—Краскала [6] о том, что полиэдр, задаваемый ограничениями вида $A\vec{x} \leqslant \vec{b}$, $\vec{x} \geqslant \vec{0}$ с абсолютно унимодулярной матрицей $A$ и произвольным вектором $\vec{b}$ (таким, что система $A\vec{x} \leqslant \vec{b}$, $\vec{x} \geqslant \vec{0}$ совместна), имеет все вершины с целочисленными компонентами?

**Замечание.** Отметим, что по теореме Пуанкаре [5] матрица инциденций произвольного ориентированного графа абсолютно унимодулярна, т. е. определитель любой ее квадратной подматрицы равняется одному из трех чисел: $-1$, 0, 1.

г) Оцените сложность алгоритма Форда—Фалкерсона ($c_{ij}$ — неотрицательные целые числа) и время работы соответствующей программы для поиска максимального потока (количество пунктов — 1000).

д) Сойдется ли за конечное время к оптимальному решению алгоритм Форда—Фалкерсона, если $c_{ij}$ — неотрицательные вещественные числа?

е) Модифицируйте алгоритм Форда—Фалкерсона для решения задачи максимального потока при неотрицательных вещественных $c_{ij}$ (например, алгоритмы Эдмонса—Карпа и Диница), оцените их сложность и время работы (количество вершин — 1000).

ж) Оцените пространственную и временную сложность алгоритма Карзанова нахождения максимального потока. [1] В чем заключаются преимущества алгоритма Карзанова, например, перед алгоритмом Форда—Фалкерсона?

[1] Описание этого алгоритма имеется, например, в лекции М. Г. Фуругяна, прочитанной студентам второго курса ФУПМ МФТИ весной 2009 г.: `http://www.intuit.ru/department/algorithms/algomodex/2/`.

**Задача 23.** В контексте предыдущей задачи будем считать, что суммарный поток машин через вершину также ограничен.

Как эту задачу свести к задаче 22 поиска максимального потока?

**Задача 24.** Во вражеской стране некоторые города соединены прямыми дорогами. Партизаны хотят сорвать перевозку оружия от города А до города Б: для этого им нужно взорвать минимальное количество дорог так, чтобы от города А стало невозможно доехать до города Б, т. е. нужно найти *минимальный разрез*.

а) Предложите алгоритм нахождения дорог, которые предполагается взорвать, и оцените его время работы (количество вершин — 1000).

б) Докажите, что количество взорванных дорог равняется максимальному потоку из вершины А в вершину Б (веса всех ребер равны 1).

**Задача 25.** Чиновник хочет проинспектировать все дороги города. Для этого он составил маршрут, проходящий по каждой дороге хотя бы один раз. Будем считать, что чиновнику нужна единица времени, чтобы проинспектировать одну дорогу (проезжая дорогу несколько раз, он ее соответствующее число раз и инспектирует). Необходимо найти минимальное время, нужное чиновнику для инспекции всех дорог (как минимум по разу).

Как свести эту задачу к задаче 22 о максимальном потоке?

**Указание.** Когда в графе нельзя найти эйлеров цикл? Какие ребра нужно сделать двойными, чтобы эйлеров цикл нашелся?

**Задача 26.** Некоторые объекты (подмножество вершин) в городе назовем стратегически важными. Найдите минимальное по суммарной длине подмножество дорог, таких, что из каждого стратегически важного объекта до каждого стратегически важного можно добраться по дорогам из выбранного множества.

а) Частным случаем каких вышеупомянутых задач является данная задача? Что известно про сложность решения частных случаев?

б) Докажите *NP*-полноту этой задачи путем сведения к ней задачи 3-SAT о выполнимости булевых формул в 3-конъюнктивной нормальной форме.

**Замечание.** Условие задачи 3-SAT см. в книге [16].

**Задача 27*: транспортная задача.** Имеется $m$ производителей и $n$ потребителей некоторого товара, расположенных в узлах транспортной сети. Пусть $x_{ij} \geqslant 0$ обозначает количество продукта, перевозимого из $i$-го узла в $j$-й, $c_{ij} \geqslant 0$ — стоимость перевозки, $a_i \geqslant 0$ — объем производства в $i$-м узле, $b_j \geqslant 0$ — суммарную потребность в $j$-м (при этом $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$). Поставим задачу о составлении плана перевозок, минимизирующего общую стоимость перевозок от производителей к потребителям:



$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min, \quad \sum_{j=1}^{n} x_{ij} = a_i, \quad \sum_{i=1}^{m} x_{ij} = b_j.$$

а) Напишите двойственную задачу к вышеприведенной задаче.

б) Дополнительно введя источник и сток и пользуясь матрицей инциденций, запишите задачу нахождения максимального потока минимальной стоимости и двойственную к ней.

в) Рассматривая эту задачу как задачу линейного программирования, получите оценку для времени работы известных вам алгоритмов в наихудшем случае и в среднем. Примите $m = n = 500$.

**Задача 28.** Рассматривается та же самая транспортная задача с целыми неотрицательными параметрами

$$a_i \in \mathbb{Z}, \quad i = 1, \ldots, n; \quad b_j \in \mathbb{Z}, \quad j = 1, \ldots, m.$$

а) Покажите, что решения $x_{ij}$ задачи в классе неотрицательных действительных чисел будут целыми.

**Указание.** Запишите задачу нахождения максимального потока минимальной стоимости (см. задачу 26 б)), затем воспользуйтесь теоремой Гофмана—Краскала (см. задачу 22 в)).

б) Как нужно модифицировать алгоритм Форда—Фалкерсона нахождения максимального потока для решения транспортной задачи? Оцените время работы получившегося алгоритма при $m = n = 500$.

в) Сойдется ли за конечное время модифицированный алгоритм из пункта а), если параметры задачи могут быть не целыми?

**Замечание.** Вообще говоря, нахождение целочисленных решений, как правило, значительно усложняет алгоритм, иногда делая задачу, решаемую за полиномиальное время, *NP*-трудной (например, в задаче размещения производства, см. [15]). Однако приведенная задача как раз является исключением из этого правила.

**Задача 29.** Пусть есть граф со взвешенными ребрами, причем веса ребер могут быть отрицательными числами.

а) Как в данном графе найти цикл, сумма длин ребер которого отрицательна?

б) Как нахождение таких циклов использовать при решении транспортной задачи 26?

в) Оцените время работы полученной программы при $m = n = 500$.

### Часть 6. Построение оптимальной транспортной сети

**Задача 30** (П. П. Бобрик, 2000). Распределение мест проживания населения по городу, имеющему форму квадрата со стороной $A$ км и не имеющему дорог, равномерное. Условное распределение мест работы жителей

(при условии, что зафиксировано место жительства) также равномерное (и, как следствие, не зависит от места жительства). Руководство города решило построить сеть дорог в виде квадратной решетки со стороной квадратной ячейки, равной $a = A/n$. Известно, что затратами на строительство сети дорог можно пренебречь по сравнению с последующими суммарными затратами на ее поддержание. Пусть $C_L$ — стоимость поддержания 1 км дороги (в одну сторону) в течение 1 года, $C_T$ — стоимость 1 часа, потраченного одним человеком в дороге. Число жителей в городе $N$. Каждый человек $Q = 300$ раз в год направляется из дома на работу и обратно с работы домой. Скорость движения человека не по дороге $v$ км/ч, скорость движения человека по дороге $V$ км/ч ($V \gg v$), время, потраченное на прохождение одного перекрестка, есть $t$ мин, $t \geqslant 0$.

а) Постройте целевую функцию, отвечающую за эффективность построенной транспортной сети.

б) Какое значение $n$ «оптимально» выбрать руководству города?

**Задача 31.** В условиях предыдущей задачи было решено построить радиально-кольцевую систему дорог, состоящую из $k$ колец и $m$ радиусов.

а) Постройте целевую функцию, отвечающую за эффективность построенной транспортной сети.

б) При каких условиях эта сеть окажется эффективней оптимальной квадратной сети из предыдущей задачи?

**Задача 32.** В условиях задачи 30 был предложен альтернативный план транспортной сети — регулярная сеть дорог, с элементарной «ячейкой» в виде правильного (равностороннего) треугольника со стороной $b$ (из расчета, что длина сети совпадает с «квадратной»). При этом оказалось, что время, потраченное на прохождение перекрестка в вершине графа транспортной сети, $t = m_i\tau$, пропорционально степени $m_i$ этой вершины. Определить, при каком значении $\tau$ эффективность сетей совпадает (если такое $\tau$ существует).

**Задача 33** (А. П. Буслаев, 2001). В литературе имеется довольно большое количество различных характеристик графов транспортных сетей. Интересной задачей является исследование этих макрохарактеристик на различных типах случайных графов. Например, во второй части монографии [25] вводятся такие показатели «хорошести» графа транспортной сети (отметим, что терминология пока не устоялась, поэтому в других местах аналогичные термины могут иметь иное содержание):

$$r_G(p) = \frac{N - E_p\xi_G}{N-1} \ \text{— } \textit{надежность},$$

где $N$ — число вершин в графе $G$, $\xi_G$ — число связных компонент в графе $\widetilde{G}_p$, полученном из исходного графа $G$ путем случайного и независимого



разыгрывания его ребер (с вероятностью $p > 0$ ребро оставляют и с вероятностью $1 - p$ ребро стирают), $E_p\xi_G$ — математическое ожидание с.в. $\xi_G$;

$$s_G(p) = \frac{E_p\eta_G - 1}{N-1} \ \text{— } \textit{устойчивость},$$

где $\eta_G$ — размер (число вершин) максимальной связной компоненты в графе $\widetilde{G}_p$;

$$rt_G(p) = \frac{1}{(N-1)N}\sum_{i,j\neq i} E_p\rho_{ij}^G \ \text{— } \textit{доступность},$$

где $\rho_{ij}^G$ — длина кратчайшего пути в графе $\widetilde{G}_p$, ведущего из вершины $i$ в вершину $j$; доопределим $\rho_{ij} = \mathrm{i}$ ($\mathrm{i}$ — мнимая единица), если из вершины $i$ нельзя добраться по графу $\widetilde{G}_p$ в вершину $j$.

а) Исследуйте качественно зависимость приведенных показателей от $p$ для графов транспортной сети со структурой задач 30 и 32 (считайте, что $N \gg 1$).

б) Сравните эффективность транспортных сетей задач 30 и 32 с помощью приведенных показателей.

в) **Гигантская компонента** (см. приложение А. М. Райгородского). Покажите, что если взять в качестве $G$ полный граф на $N$ вершинах, то $\widetilde{G}_p \equiv G(N, p)$ (модель Эрдёша—Реньи) и для любого $c > 1$ найдется такое $0 < \gamma_c < 1$, что

$$P\left(s_G\left(\frac{c}{N}\right) \geqslant \gamma_c\right) \xrightarrow[N\to\infty]{} 1.$$

**Задача 34*: показатель качества графа транспортной сети** (Уотс—Строгатц, 1998). Рассматривается такой ориентированный взвешенный граф транспортной сети, что из любой его вершины можно добраться по ребрам (с учетом ориентации ребер) в любую другую. Обозначим через $d_{ij}$ длину кратчайшего пути по рассматриваемому графу транспортной сети из вершины $i$ в вершину $j$. Показателем качества исследуемого графа назовем величину:

$$e = (n-1)^{-1}n^{-1}\sum_{i,j\in V:\ i\neq j} d_{ij}^{-1},$$

где $V = \{1, \ldots, n\}$ — множество вершин графа. Оцените значение этого показателя для графа транспортной сети из задачи 30, считая $n \gg 1$.

**Задача 35**: нахождение минимума функционала среднего расстояния** (Е. О. Степанов, 2005). На множестве всех замкнутых связных множеств $\Sigma \subset \mathbb{R}^2$ хаусдорфовой размерности 1 ($H^1(\Sigma) < \infty$), удовлетворяющих ограничению на длину, $H^1(\Sigma) \leqslant l$, требуется найти решение $\Sigma^*$

следующей задачи:

$$F_\varphi(\Sigma) = \int\limits_{\mathbb{R}^2} A(\operatorname{dist}(\vec{x},\,\Sigma))\,d\varphi(\vec{x}) \to \min_{\Sigma:\ H^1(\Sigma)\leqslant l},$$

где $\varphi(\vec{x})$ — мера, отвечающая распределению населения в окрестности точки с координатой $\vec{x}$, число $A(d)$ задает затраты жителя на проезд на расстояние $d$. Транспортная сеть $\Sigma$, таким образом, проектируется исходя из соображений минимизации полных затрат населения на достижение сети из мест проживания. Покажите, что при весьма общих условиях решение $\Sigma^*$ этой задачи существует, не содержит петель, имеет конечное число концевых точек и точек ветвления, а также обладает некоторыми свойствами регулярности (см. раздел, посвященный задаче Штейнера). Предложите эффективный способ численного поиска $\Sigma^*$.

Литература

К частям 1, 3:

1. *Hwang F. K.*, *Richards D.*, *Winter P.* The Steiner tree problem. Elsevier Science Publishers, 1992.
2. *Гордеев Э. Н.*, *Тарасцов О. Г.* Задача Штейнера. Обзор // Дискретная математика. 1993. Т. 5, № 2. С. 3–28; `http://www.mathnet.ru/` [1)]
3. *Иванов А. О.*, *Тужилин А. А.* Теория экстремальных сетей. Москва–Ижевск: Институт компьютерных исследований, 2003.
4. *Протасов В. Ю.* Максимумы и минимумы в геометрии. № 31. М.: МЦНМО, Библиотечка «Математическое просвещение», 2005;
   `http://www.mccme.ru/mmmf-lectures/books/books/book.31.pdf`

К частям 2, 4, 5:

5. *Берж К.* Теория графов и ее приложения. М.: ИЛ, 1962.
6. *Филлипс Д.*, *Гарсиа-Диас А.* Методы анализа сетей. М.: Мир, 1984.
7. *Гвишиани А. Д.*, *Гурвич В. А.* Динамические задачи классификации и выпуклое программирование в приложениях. М.: Наука, 1992.
8. *Stoer Mechthild*, *Wagner Frank* A Simple Min-Cut Algorithm // Journal of the ACM. 1997. V. 44, № 4. P. 585–591.
9. *Разборов А. А.* О сложности вычислений // Матем. просв. 1999. № 3;
   `http://www.mccme.ru/free-books/matpros4.html`
10. *Смейл С.* О проблемах вычислительной сложности // Матем. просв. 2000. № 4;
    `http://www.mccme.ru/free-books/matpros5.html`
11. *Schrijver A.* On the history of the transportation and maximum flow problems // Math. Program. Ser B. 2002. V. 91. P. 437–445;
    `http://oai.cwi.nl/oai/asset/10084/10084A.pdf`

[1)] Обратим внимание, что на этом электронном ресурсе открыт свободный доступ к полным текстам статей ряда ведущих российских научных журналов («Успехи математических наук», «Математический сборник», «Математические заметки», «Функциональный анализ и его приложения», «Известия РАН», «Дискретная математика», «Математическое моделирование», «Журнал вычислительной математики и математической физики» и др.).



12. *Вялый М. Н.* Линейные неравенства и комбинаторика. М.: МЦНМО, Летняя школа «Современная математика». 2003;
    `http://www.mccme.ru/free-books/dubna/vyalyi.pdf`
13. *Зыков А. А.* Основы теории графов. М.: Вуз. книга, 2004.
14. *Diestel R.* Graph Theory. Electronic Edition, NY: Springer, 2005.
15. *Goldberg A. V.*, *Harrelson C.* SODA'05 Proceedings of the sixteenth annual ACM-SIAM symposium on discrete algorithms. USA Society for Industrial and Applied Mathematics Philadelphia. 2005. P. 156–165.
16. *Кормен Т. Х.*, *Лейзерсон Ч. И.*, *Ривест Р. Л.*, *Штайн К.* Алгоритмы: Построение и анализ. М.: Вильямс, 2005.
17. *Кузюрин Н. Н.*, *Фомин С. А.* Эффективные алгоритмы и сложность вычислений. М.: МФТИ, 2007.
18. *Емеличев В. А.*, *Мельников О. И.*, *Сарванов В. И.*, *Тышкевич Р. И.* Лекции по теории графов. М.: УРСС, 2009.
19. Annual IEEE Symposium on Foundations of Computer Science, 1–51;
    `http://theory.stanford.edu/focs2010/`
20. `http://e-maxx.ru/algo/`; `http://www.machinelearning.ru/`
21. *Шень А. Х.* Программирование: теоремы и задачи. М.: МЦНМО, 2011.

К части 6:

22. *Корте Б.*, *Виген Й.* Комбинаторная оптимизация: теория и алгоритмы. (В печати.)
23. *Стенбринк П. А.* Оптимизация транспортных сетей. М.: Транспорт, 1981.
24. *Бобрик П. П.* Сравнение эффективностей треугольной и квадратичной регулярных транспортных сетей // Транспорт: наука, техника, управление. М.: Изд-во ВИНИТИ, 2000. № 7.
25. *Луканин В. Н.*, *Буслаев А. П.*, *Трофимов Ю. В.*, *Яшина М. В.* Автотранспортные потоки и окружающая среда. Ч. 1, 2. М.: ИНФРА-М, 1998, 2001.
26. *Latora V.*, *Marchiori M.* Efficient behavior of small-world networks // Phys. Rev. Letters. 2001. V. 87, № 19;
    `http://www.w3.org/People/Massimo/papers/2001/efficiency_prl_01.pdf`
27. *Степанов Е. О.* Математические модели оптимизации транспортных сетей. СПб: СПбГУ ИТМО, 2005.
28. *Tero A. et al.* Rules for biologically inspired adaptive network design // Science. 2010. V. 327, № 5964. P. 439–442.

## Задачи от «Яндекс.Пробки»

**Преобразования запрещенных маневров** (А. И. Верещагин, В. Б. Гольдштейн, И. И. Колесниченко, М. В. Левин, Андрей А. Петров). При нахождении маршрутов проезда по городу важно учитывать не только взаимное расположение дорог, но и правила дорожного движения. Картографические компании собирают информацию о расположении дорог и дорожных знаках. Эта информация предоставляется отдельно, однако

для маршрутизации обязана учитываться вместе. Граф дорог состоит из перекрестков и улиц, их соединяющих. Встречаются улицы с односторонним и двусторонним движением, а также улицы, по которым проезд машин запрещен вовсе. Однако из-за запрещающих знаков на некоторые улицы можно выехать не всегда. Простейшим примером служит запрещенный левый поворот. Несмотря на то, что по улицам $AMB$ и $CMD$ разрешен проезд в обоих направлениях, нельзя проехать по ним последовательно. То есть проезд по пути $AMC$ и $BMD$ запрещен. Обобщая этот пример, назовем запрещенным маневром последовательность дорог, проезд по которым от начала и до конца запрещен. Запрещенные маневры могут состоять как из двух ребер (запрет левого поворота), так и из 4–5 дорог (сложный выезд на трассу).

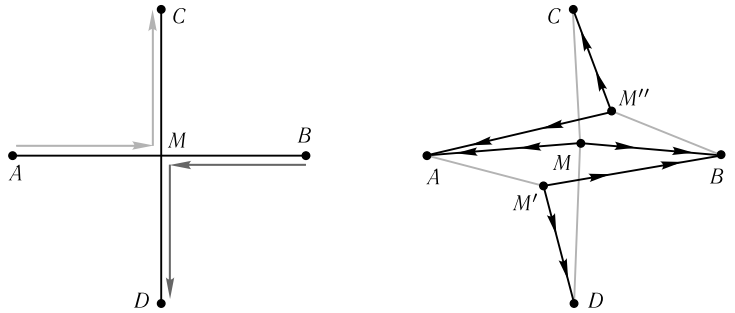



Рис. 1

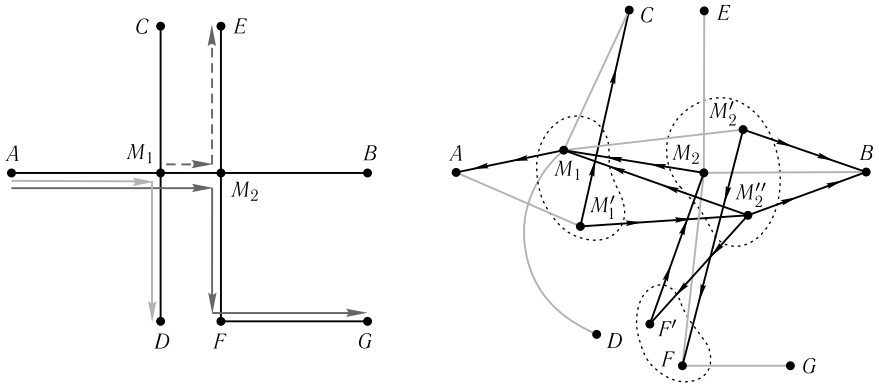



Рис. 2

Так, на рис. 1 два запрещенных маневра:

$$P_1 = (AM,\, MC) \quad \text{и} \quad P_2 = (BM,\, MD).$$



На рис. 2 три запрещенных маневра:

$$P_1 = (AM_1,\, M_1D) \quad \text{и} \quad P_2 = (AM_1,\, M_1M_2,\, M_2F,\, FG) \quad \text{и} \quad P_3 = (M_1M_2,\, M_2E).$$

Рассмотрим более формальную постановку задачи. Задан ориентированный граф $\Gamma$ без петель и набор запрещенных маршрутов $P$. Для простоты в примерах будем рассматривать неориентированный граф, каждому ребру без стрелок на рис. 1, 2 соответствуют два ребра в разных направлениях. Маршрут $P_i$ — последовательность ребер графа. Запрет проезда по маршруту $P_i$ означает запрет проезда по всему маршруту начиная с первого ребра и заканчивая последним. Проезд по любой части маршрута не запрещен.

Требуется построить такой граф $H$, в котором не будет запрещенных маневров. Каждой вершине (ребру) графа $\Gamma$ соответствует одна или несколько вершин (ребер) графа $H$. Каждому разрешенному пути (чередующаяся последовательность вершин и ребер) в графе $\Gamma$ соответствует путь в графе $H$. Запрещенному пути в графе $\Gamma$ не должно соответствовать ни одного пути в графе $H$. Разрешенный путь в графе $\Gamma$ — путь, не содержащий запрещенных маневров. Запрещенный путь в графе $\Gamma$ — путь, содержащий хотя бы один запрещенный маневр.

Все вершины графа $\Gamma$ имеют вполне конкретный географический смысл и находятся в некоторых различных точках Земли. Различные вершины графа $H$ могут находиться в одной точке Земли, однако отражать разное состояние.

На рис. 1 знания, что машина находится в точке $M$, недостаточно для понимания, куда может ехать машина. В графе $H$ вершине $M$ будет соответствовать три вершины:

1) если мы приехали из вершины $A$;
2) из вершины $B$;
3) из другой вершины.

Рассмотрим граф $H$ для графа, изображенного на рис. 2. Серым обозначены двусторонние дороги. Вершина $M'_1$ означает, что машина находится в вершине $M_1$ и запрещены следующие пути: $\{M_1M_2E,\, M_1D,\, M_1M_2FG\}$; вершина $M'_2$ находится в $M_2$ и запрещен путь $M_2E$; $M''_2$ находится в $M_2$ и запрещены пути $\{M_2E,\, M_2FG\}$; $F'$ находится в $F$ и запрещен путь $FG$.

**Задача 1** (М. А. Хохлов). Для описания транспортной ситуации на некотором участке дороги необходимо агрегировать данные от многочисленных транспортных средств, проезжающих по нему, в единый показатель (например, «среднюю скорость» движения по участку в данный момент времени). Данные об отдельных проездах характеризуются большим разбросом значений, который может быть связан с индивидуальными особенностями машин и водителей (при свободном движении), с хаотичным

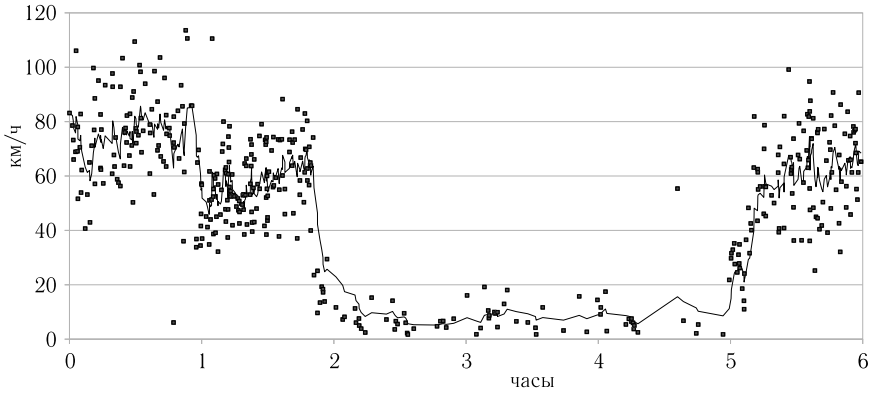



Рис. 3. Сглаживание индивидуальных скоростей итерационным алгоритмом

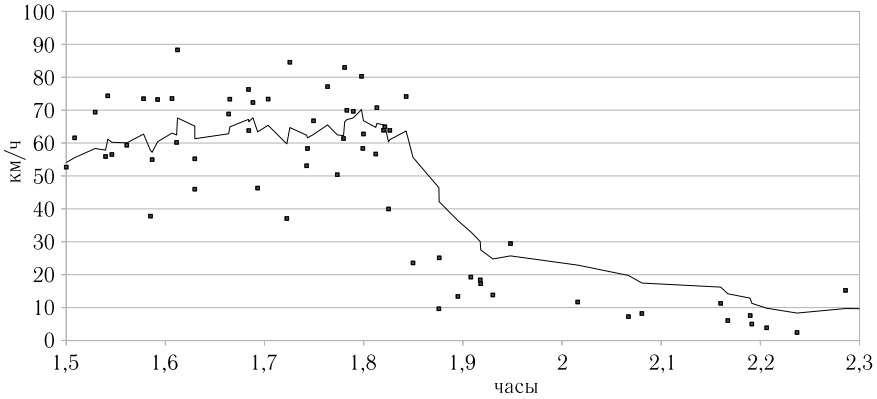



Рис. 4. Сглаживание индивидуальных скоростей итерационным алгоритмом. Момент возникновения затора

режимом движения (в условиях затруднения), а также с погрешностями самого метода измерения скорости.

Традиционный способ борьбы с указанным разбросом — сглаживание входных данных, например, итерационным алгоритмом:

$$I_{n+1} = (1 - \alpha) \cdot I_n + \alpha \cdot v_n,$$

который после каждого нового наблюдения $v_n$ корректирует выдаваемое значение $I_n$ с коэффициентом сглаживания $\alpha$. Результаты применения этого алгоритма к данным о скоростях проезда на одной из московских улиц приведены на рис. 3. Здесь отдельные точки соответствуют скоростям проезда отдельных транспортных средств по фиксированному участку дороги.



Линией показан результат сглаживания. По оси абсцисс — календарное время в часах, по оси ординат — скорость в км/ч.

Недостатки указанного метода хорошо видны при увеличении временного масштаба в момент возникновения затора (рис. 4). Сглаженное значение отстает от реального более чем на 5 мин., при этом на протяжении 20 мин. после возникновения затора показывает завышенное значение скорости.

Задача состоит в том, чтобы предложить способ агрегирования (сглаживания) индивидуальных скоростей, лишенный указанных недостатков, т. е. с одной стороны, выдающий стабильное значение скорости при устойчивом режиме движения, а с другой — оперативно реагирующий на изменение режима движения и выдающий адекватное значение скорости в пробке.

**Задача 2** (М. А. Хохлов). Изучая исторические данные о скоростях движения транспортных средств на участках дорог, можно заметить, что движение с некоторыми скоростями оказывается более устойчивым, чем с другими, т. е. существуют определенные режимы движения, сохраняющиеся в течение длительного времени. Скорости таких режимов могут сильно отличаться для различных дорог. Особенно отчетливо существование таких режимов наблюдается на гистограммах скоростей за достаточно длительный период наблюдений (месяц и более). На рис. 5 приведены гистограммы скоростей для участков двух различных московских улиц. Задача: предложить способ автоматического определения количества и типичных скоростей режимов транспортного потока на основе исторических данных (например, по гистограмме).

**Задача о краткосрочном прогнозировании** (М. А. Хохлов, Б. Н. Карпов). В Москве имеются данные о траекториях большого числа автомобилей. Данные поступают в следующем виде: координата автомобиля (с точностью до нескольких метров), время. Каждый оснащенный автомобиль посылает такие данные с небольшим периодом (несколько десятков секунд). Данные довольно полные, т. е. если рассматривать основные магистрали, то в любой момент времени по большинству из них обязательно кто-нибудь едет и сообщает свои координаты. Требуется построить прогноз значений скоростей автомобилей на определенных ребрах на час вперед, например, «внедрив» физические представления о поведении транспортного потока в регрессионные подходы к получению прогноза и(или) подходы типа метода ближайших соседей, основанные на простом принципе: «ищем в истории наиболее близкую ситуацию и говорим, что сейчас будет так, как было тогда». Предложите, как можно «внедрять эти физические представления».

**Указание.** В качестве примера наглядного отображения данных приведем рис. 6 (по оси абсцисс расстояние, по оси ординат время). Отдадим

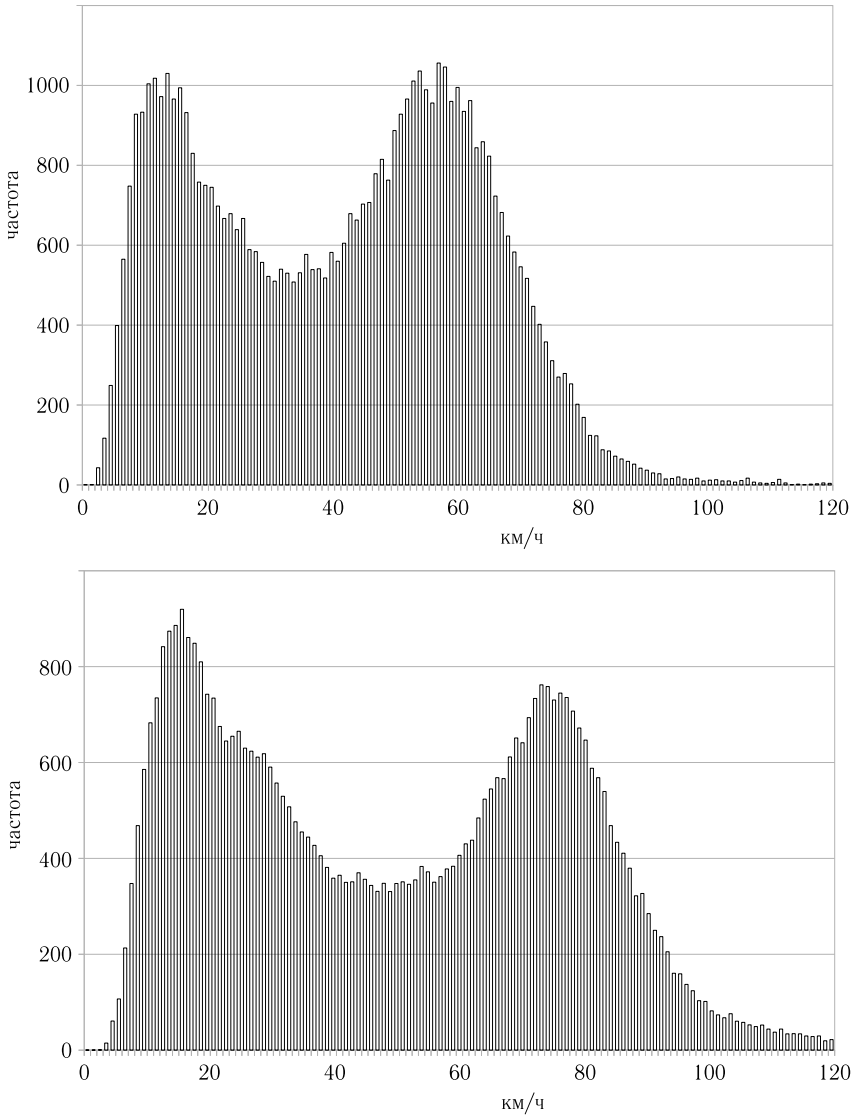



Рис. 5. Гистограммы наблюдаемых скоростей по историческим данным за 1 месяц

предпочтение СТМ-модели (п. 2.2.4 главы 2) в описании свободного движения (скорость свободного движения на заданном участке легко определяется по исторической информации) и широких движущихся кластеров



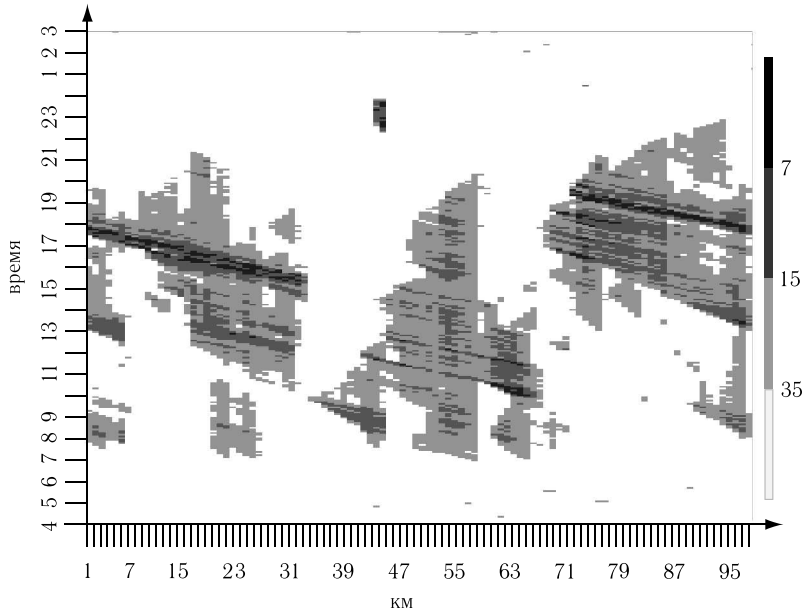



Рис. 6. Один день МКАД (будний день, весна 2011)

(раздел 2.4 главы 2 и глава 3). По исторической информации эволюцию широких движущихся кластеров можно неплохо научиться прогнозировать. Для завязки (в корреляционном подходе это будет особенно важно) характеристик возникающих синхронизированных фаз на скорости, наблюдающиеся на соседних сегментах, потребуется рассмотрение различных типовых ситуаций и их детальный просчет на микроуровне, с последующим агрегированием на нужный нам уровень детализации.

Литература

1. *Червоненкис А. Я.* Компьютерный анализ данных. М.: Яндекс, 2009.
2. `http://habrahabr.ru/company/yandex/blog/153631/`
3. *Дорогуш Е. Г.* Математическое моделирование транспортных потоков на кольцевой автостраде // Сборник статей молодых ученых факультета ВМК МГУ. 2011. № 8. С. 54–68;
`http://smu.cs.msu.su/sites/default/files/attachments/smu8.pdf`

# Исследовательские вычислительные задачи, предлагавшиеся в 2011 г.

Практически все приводимые ниже задачи являются упрощенными постановками реальных прикладных задач, с которыми к нам (нашим коллегам) обращались в последнее время различные заказчики (мы намеренно не называем их, приводя лишь фамилии людей, частично формализовавших задачи и исследовавших их). Дабы «не закопаться в деталях», мы намеренно упростили постановки, опуская детали или добавляя задачам излишнюю симметрию. Ряд задач специально сформулирован весьма общо (например, не указано, какой моделью транспортного потока нужно пользоваться при решении) — как правило, это связано с тем, что существует несколько (иногда даже много) подходов к решению и не хочется выделять какой-то один из них.

**Задача об оптимальном числе поездов в кольцевом метро** (Е. О. Черноусова). В некотором городе все $n = 15$ станций метро расположены на кольцевом маршруте. Непрекращающееся движение поездов осуществляется в оба направления. Время, затрачиваемое поездом на преодоление расстояния между любыми двумя соседними станциями, фиксировано и равно $T = 5$ мин. Время, требуемое на остановку, не учитывается (пренебрежимо мало по сравнению с $T$). Всего на кольцевом маршруте курсируют по $m$ поездов в каждом направлении с интервалом движения, равным $t$. Потоки пассажиров, приходящих на каждую из станций метро, описываются независимыми пуассоновскими процессами с фиксированной (одинаковой) интенсивностью $\lambda = 50$ чел/мин; для простоты считаем интенсивность постоянной во времени в течение всего рабочего дня (12 часов) и изо дня в день. В связи со свойствами пуассоновского процесса (теорема Григелиониса), такое описание вполне естественно и часто используется в теории массового обслуживания. Предполагается, что конечная станция «выбирается» каждым пассажиром случайно и равномерно среди $n - 1$ станций (направление движения выбирается по принципу кратчайшего пути, если по обоим направлениям время в пути одинаково, то направление выбирается случайно). Вместимость поездов одинакова и равна $K = 600$ чел. Требуется определить «оптимальное» число поездов:

$$\omega h(m) + \psi \cdot 2m \to \min_{m \in \mathbb{N}},$$

где $\omega = 2$ руб./мин — цена одной минуты, потерянной в дороге одним пассажиром, $h(m)$ — математическое ожидание суммарного времени (в мин) по всем пассажирам, потерянного в течение одного месяца (30 дней, т. е. 21600 минут) на ожидание поезда, $\psi = 5 \cdot 10^5$ руб. — затраты в месяц на содержание одного поезда.



Литература

1. *Серебровский А. П.* Курс лекций по математическим методам в теории массового обслуживания. М.: МФТИ, 2010; `http://frtk.ru/forstudents/study/studyMaterials/4kurs/TMO2010-arpggwlikuv.pdf`

**Задача об оптимальном распределении автобусов по маршрутам** (М. С. Ишманов). В некотором городе с заданным (взвешенным) графом транспортной сети $G = (V, E, \{t_e\}_{e \in E})$ (вершины $V$ отвечают остановкам, $t_e$ — время прохождения ребра $e$) имеются различные маршруты движения автобусов $P$ (считаем, что автобусы на каждом маршруте циркулируют туда и обратно по этому маршруту). Потоки пассажиров, приходящих на каждую из остановок, описываются независимыми пуассоновскими процессами с фиксированными интенсивностями $\lambda_i$, $i \in V$. Считайте, что пассажиры садятся на первый подходящий автобус, не дожидаясь следующего (с другого маршрута), который может чуть быстрее доставить их в пункт назначения. Пусть $\pi_{ij}$ — доля пассажиров, направляющихся с остановки $i$ на остановку $j$, причем $\forall i, j \in V\colon \pi_{ij} > 0\ \exists p \in P\colon i, j \in p$. Пусть $m_p$ — число автобусов, курсирующих по маршруту $p$. Считая $\sum\limits_{p \in P} m_p = M$ и априорно предполагая, что $m_p \gg 1$, предложите эффективный алгоритм решения задачи распределения имеющегося автобусного парка из $M$ автобусов по маршрутам так, чтобы математическое ожидание общего времени, потерянного всеми пассажирами в пути (ожидание автобуса на остановке + движение по маршруту) было бы минимальным.

Литература

1. *Зак Ю. А.* Прикладные задачи теории расписаний и маршрутизации перевозок. М.: Книжный дом «Либроком», 2012.

**Задача: минимальный процент водителей, предоставляющих информацию** (В. Л. Швецов). а) Некоторый город (или район города, например, Манхэттен) имеет квадратную клеточную транспортную сеть с $n^2 = 20^2$ клетками. Считаем, что независимо из каждой вершины в утренние часы пик с постоянной интенсивностью (пуассоновского процесса) $\lambda = 1000$ чел/час водители выезжают на работу на автомобилях, причем место работы с равной вероятностью находится в любой другой вершине. Маршрут также выбирается случайно и равновероятно из кратчайших (по числу ребер) маршрутов с одинаковым числом ребер. Время в пути по ребру $e$ есть экспоненциальная случайная величина с математическим ожиданием, равным $\tau = \max\{5,\ 2 \cdot 10^{-2} \cdot N_e\}$ мин, где $N_e$ — число автомобилей на ребре $e$ в данный момент времени. Какой процент автомобилистов

должен сообщать информацию о своих скоростях на ребрах, чтобы в произвольный момент времени (здесь предполагается, что к этому моменту времени город уже довольно долго «жил» в описанном выше режиме) не менее чем по $\rho = 90\,\%$ ребер была текущая информация о загрузке (возможно избыточная, т. е. несколько значений по одному ребру) с вероятностью не менее $\gamma = 0{,}9$? Информация считается текущей, если она была получена менее $t = 30$ минут назад. Время, потраченное на прохождения перекрестка, считайте равным 0,5 мин.

б) Генплан города с графом транспортной сети, аналогичным п. а) (с $n^2 = 5^2$), решил разработать электронный ресурс, на котором путем опроса населения будут собираться данные, недостающие для моделирования (прежде всего речь идет о восстановлении матрицы корреспонденций). Мотивацией пользователей, заходящих на этот ресурс, может быть учет их интересов при модернизации городской транспортной системы (обезличенный пользователь будет желать оставлять информацию о своих типичных маршрутах ровно по той же причине, что и ходить на выборы голосовать за понравившегося кандидата). Считая, что все жители города (300 000 человек) направляются утром из дома на работу, определите, какой процент (случайно выбранных) жителей должен оставить информацию о своих передвижениях утром, чтобы можно было с относительной точностью не ниже 10% с вероятностью не меньшей 0,9 восстановить каждый элемент утренней матрицы корреспонденций. Считайте также, что утренняя матрица корреспонденций состоит из одинаковых по порядку величины элементов.

**Задача о критическом числе автомобилей для заданного города** (В. А. Малышев). Пусть имеется город с определенным графом транспортной сети, $p_{ij}$ обозначает вероятность того, что произвольный автомобиль, циркулирующий по транспортному графу и оказавшийся на ребре $i$, повернет на ребро $j$. Считаем стохастическую матрицу $P = \|p_{ij}\|$ заданной. Поведение транспортного потока подчиняется модели типа TASEP (см. приложения М. Л. Бланка). Покажите (см. п. 3.3 приложения Замятина — Малышева), что при весьма общих условиях существует такое критическое число автомобилей (которые курсируют по транспортному графу), небольшое превышение которого ведет к резкому росту загруженности транспортной сети, к образованию пробок. Тем не менее, типичным также будет наличие определенной довольно большой доли ребер графа транспортной сети, загрузка которых практически не чувствительна к такому увеличению.

**Замечание.** В действительности многие крупные города как раз находятся где-то на границе этого «фазового перехода». Причина проста и имеет в своей основе принцип неподвижной точки в форме теоремы Брауэра. Если рассматривать эволюцию города с точки зрения появления новых жителей, новых рабочих мест, строительства новых дорог, то можно условно считать, что новый водитель будет пользоваться автомобилем в городе, если «комфортность» такого пользования не ниже некоторого уровня. Поскольку в малой окрестности критического значения происходит резкое падение этой комфортности, то у большинства новых потенциальных пользователей этой транспортной сети пропадает желание ими быть (и они выбирают себе альтернативы: «переходят» на общественный транспорт, выбирают соответствующим образом место работы и т. п.). Аналогично можно пойти и в обратную сторону.



Литература

1. *Neri I.*, *Kern N.*, *Parmeggiani A.* The totally asymmetric simple exclusion process on networks // `arXiv:1105.2905v2`, 2011.
2. *Furtlehner C.*, *Lasgouttes J.-M.*, *Samsonov M.* One-dimensional Particle Processes with Acceleration/Braking Asymmetry // `arXiv:1109.1761v1`, 2011.

**Задача о слабо связанной архитектуре модели поведения водителя и агрегировании микроскопической модели** (Я. С. Панасюк). Для компьютерных микроскопических систем моделирования транспортных потоков и задач, связанных с моделированием транспортных средств на микроскопическом уровне, чрезвычайно важной является модель поведения автомобиля на дороге или модель поведения водителя. Обычно требования к модели поведения водителя определяются в виде набора опорных моделей (модели следования за лидером, модели перестроения и т. п.) и набора общих правил поведения (различные запреты, приоритеты дорожных ситуаций и т. п.). Программная имплементация модели даже с небольшим количеством опорных моделей представляет непростую задачу: необходимо определить алгоритмы поведения водителя во всех возможных комбинациях дорожных ситуаций, а количество таких комбинаций находится примерно в экспоненциальной зависимости от числа опорных моделей и правил поведения. Помимо сложности создания самой модели, программную архитектуру модели поведения водителя зачастую можно характеризовать как тесно связанную («tightly coupled») или монолитную («weak cohesion») — опорные модели и имплементация правил поведения оказываются в сильной зависимости друг от друга. Внесение минорных изменений в поведение водителя отражается в изменении многих компонент модели, в худшем случае приходится пересматривать все зависимости и связи. Особенно остро проблема тесной связанности компонент модели поведения водителя наблюдается при создании больших комплексных систем микроскопического моделирования транспортных потоков, когда количество опорных моделей исчисляется десятками, а правила поведения почти совпадают с ПДД и могут меняться в зависимости от страны применения.

Предложите слабо связанную («loosely coupled») архитектуру (принцип построения) модели поведения водителя, которая бы обеспечивала гибкую настройку и простую расширяемость модели новыми опорными моделями и новыми правилами поведения. Постройте на основе предложенной архитектуры модель поведения водителя, реагирующего на многополосные дороги, сигналы светофоров и знаки ограничения скорости. В качестве основы для системы микроскопического моделирования можно использовать любую систему с открытым исходным кодом из [1].

Оцените, на основе построенной микромодели, насколько эффективней может быть движение, если имеет место синхронизация действий водителей в потоке (водители декларируют свои намерения ближайшим соседям не только с помощью сигналов фар и жестов)?

Как с помощью предложенной модели можно агрегированно описывать транспортную сеть? Агрегированное описание крайне важно в реальных приложениях, поскольку калибровать более грубую (агрегированную) модель намного проще и адекватнее (меньше переобучение).

Литература


1. SUMO, COS.SIM, EMME/3, MITSIM, PTV (VISSUM, VISSIM), TRANSIMS, TRANS-NET, SCATS, TOPL, AIMSUN, PARAMICS, AUTOBAHN, IHSDM, INTEGRATION, PLANSIM-T, FLEXSYT-II и др.; `http://sumo.sourceforge.net/`, `http://code.google.com/p/cos-sim/`, `http://www.pdfqueen.com/manual-emme3`, `http://web.mit.edu/its/dynamit.html`, `http://web.mit.edu/its/mitsimlab.html`, `http://www.ptv-vision.ru/`, `http://www.aimsun.com/wp/?p=1967`, `www.isa.ru/transnet`, `http://www.scats.com.au/`, `http://gateway.path.berkeley.edu/topl/docs.html`, `http://www.traffic-simulation.de/`
2. *Horiguchi R.*, *Kuwahara M.* The art of utilization of traffic simulation models: How do we make them be reliable tools? // Knowledge-Based and Intelligent Information and Engineering Systems. 2010. P. 308–317. (Lecture Notes in Computer Science. V. 6279). `http://www.transport.iis.u-tokyo.ac.jp/publications/2002-034.pdf`
3. *Hanabusa H. et al.* Construction of a data set for validation of traffic simulations // Journal of Japan Society of Civil Engineers. 2001. № 688/IV-53. P. 115–123.
4. Эталонные наборы данных для проведения процедур валидации систем моделирования транспортных потоков. `http://www.jste.or.jp/sim/bmdata/index.html`
5. Fundamentals of Traffic Simulation / Jaume Barceló (Ed.) Springer, 2012. (International Series in Operations Research & Management Science. V. 145).
6. *Lee H. K.*, *Kim B. J.* Dissolution of traffic jam via additional local interactions // `arXiv:1109.2191v1`, 2011.




В следующей задаче и везде в дальнейшем, если не оговорено противное, МКАД рассматривается без въездов и съездов.

**Задача: транспортный поток как многокомпонентная жидкость** (А. С. Холодов). На основе какой-нибудь модели многополосного транспортного потока на МКАД исследуйте зависимость уравнения состояния $V(\rho)$ от параметров микромодели (например, времени реакции водителя, желаемой скорости или характеристик тормозной системы). Как влияет многокомпонентность транспортного потока — сосуществование различных типов водителей (обычно выделяется несколько компонент, от 3 до 12, типа «тракторы–грузовики», «обычные водители», «мигалки–лихачи») на уравнение состояния?

Литература


1. *Холодов Я. А.*, *Холодов А. С.*, *Гасников А. В.*, *Морозов И. И.*, *Тарасов В. Н.* Моделирование транспортных потоков — актуальные проблемы и пути их решения // Труды МФТИ (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 152–162; `http://mipt.ru/nauka/trudy/N+4+%288%29.html`


**Задача: как правильно себя вести в пробке** (Д. И. Петрашко). На основе какой-нибудь многополосной микроскопической модели транспортных потоков на МКАД исследуйте влияние поведения водителей на скорость рассасывания локального затора и на среднюю установившуюся скорость движения автомобилей по МКАД в зависимости от числа автомобилей, циркулирующих по МКАД. Например, влияние того, насколько близко автомобиль в практически стоячей пробке подъезжает к впереди идущему автомобилю, или влияние времени реакции. Верно ли, что «фазовый переход» (резкое снижение скорости при прохождении плотностью критического значения) становится тем более выраженным, чем более длинная кольцевая дорога рассматривается?

Литература

1. *Богданов К. Ю.* Прогулки с физикой. Библиотечка «Квант» В. 98. М.: Бюро Квантум, 2006 (гл. 18);
http://kvant.info/k/bibl_98/181-192.pdf

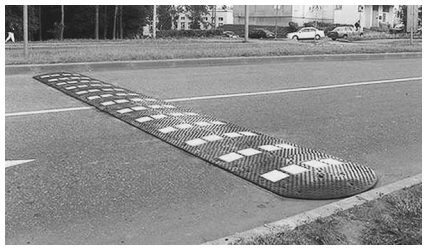
**Задача о возможном «вреде» локального расширения дороги и влиянии съезда на пропускную способность дороги** (Б. Н. Четверушкин и др.). а) На основе какой-нибудь многополосной микроскопической модели транспортных потоков исследуйте на «двухполосном МКАД» эффект наличия третьей полосы на протяжении $r\%$ пути, в зависимости от $r$ и от величины загруженности МКАД (например, от числа автомобилей, циркулирующих по МКАД). С помощью гидродинамических аналогий постарайтесь пояснить наличие (и исследовать характеристики) возникающей при определенных режимах синхронизированной фазы (см. главу 3), «зацепившейся» за место сужения трех полос до двух.

б) На основе какой-нибудь многополосной микроскопической модели транспортных потоков исследуйте (в зависимости от загрузки дороги) снижение пропускной способности дороги из-за наличия на ней съезда в окрестности этого съезда.

Ключевым местом в этих задачах является способ описания перестроения автомобилей. Например, в п. а) особенно важно описание перестроения автомобилей в месте сужения трех полос в две.

Литература

1. *Сухинова А. Б.*, *Трапезникова М. А.*, *Четверушкин Б. Н.*, *Чубарова Н. Г.* Двумерная макроскопическая модель транспортных потоков // Матем. мод. 2009. Т. 21, № 2. С. 118–126;
http://www.mathnet.ru/links/c3157afce8d7545c8864a28c55d3210e/mm2741.pdf



**Задача о «лежачем полицейском»** (Н. Н. Смирнов и др.). Руководство некоторого города решило разместить на двухполосной дороге рядом со школой два «лежачих полицейских» (с точки зрения безопасности пешеходов, автомобили не должны суметь набрать скорость в промежутке между «лежачими полицейскими» большую, чем 30 км/час). Считайте, что



лежачий полицейский рекомендуется проходить автомобилям на скорости, не превышающей 10 км/час.

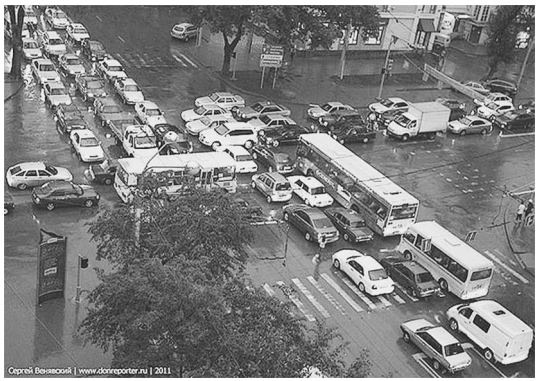
а) Используя какую-нибудь модель транспортного потока, определите, насколько далеко можно разнести «лежачие полицейские» (по постановке задачи это расстояние может быть от 10 м до 100 м).

б) Чтобы в результате строительства не снижать пропускную способность рассматриваемого участка дороги, было решено увеличить на нем число полос. На какое минимальное число полос необходимо увеличить уже имеющееся число полос, чтобы наличие «лежачих полицейских» не снижало пропускную способность?

в) Стоит ли заменить «лежачих полицейских» из п. а) на светофор ($T_{\text{зел}} = 90$ с, $T_{\text{кр}} = 30$ с, $T_{\text{жел}} = 0$ с)?

Литература

1. *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. Том I. Прикладные исследования / Под редакцией В. В. Александрова и В. Б. Кудрявцева. М.: Изд-во МГУ, 2009. С. 311–322.




**Задача об оптимальном режиме работы светофора** (М. В. Обидин, Т. С. Обидина). Пусть имеется регулируемый (светофором) перекресток, в котором пересекаются две двусторонние (по три полосы в каждую сторону) дороги. Известны входящие в перекресток потоки вдалеке от светофора, выходящие с перекрестка потоки ничем не ограничены. Светофор может работать в разных фазах (по очередности их включая). Можно менять как саму схему фаз, так и длительности фаз при уже выбранной схеме. Постройте на основе какой-нибудь микроскопической модели транспортных потоков «оптимальный» режим работы светофо-

ра, в зависимости от входящих потоков. Под оптимальностью имеется в виду минимизация суммарного времени всех водителей, потраченного на преодоление перекрестка, за достаточно большой промежуток времени. Важным местом в задаче является способ описания перестроения автомобилей (особенно, непосредственно перед перекрестком). С помощью гидродинамических аналогий постарайтесь пояснить наличие при весьма общих условиях синхронизированных фаз (см. главу 3), «зацепившихся» за перекресток.

Литература

1. *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 (специальный выпуск, посвященный математическому моделированию транспортных потоков) / Под ред. акад. В. В. Козлова. 2010. Т. 2, № 4(8). С. 141–151; `http://mipt.ru/nauka/trudy/N+4+%288%29.html`
2. *Трапезникова М. А.*, *Фурманова И. Р.*, *Чурбанова Н. Г.*, *Липп Р.* Моделирование многополосного движения автотранспорта на основе теории клеточных автоматов // Матем. мод. 2011. Т. 23, № 6. С. 133–146.

**Задача об управлении системой светофоров в небольшом городе** (К. К. Глухарев, А. М. Валуев и др.). В некотором небольшом провинциальном городе транспортная сеть имеет вид, изображенный на рис. 1.

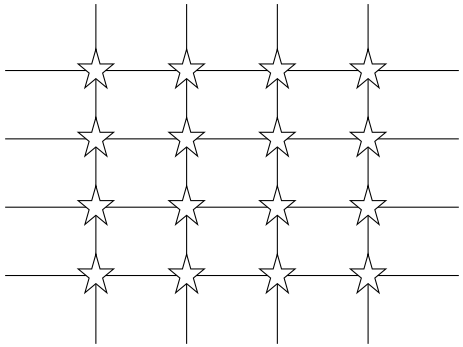


Рис. 1

Звездочками обозначены регулируемые (светофорами) перекрестки. Все дороги двусторонние трехполосные в каждую сторону, имеющие одинаковую длину $L = 3$ км участка между соседними светофорами. Входящие потоки $q = 5000$ авт/час известны и также одинаковы. Каждый автомобиль (желаемая скорость движения $v = 90$ км/час, время реакции водителя $\tau = 0{,}5$ сек), входящий в систему, с равной вероятностью выбирает своей конечной целью любую из оставшихся 15 точек входа/выхода. При этом свой маршрут он равновероятно выбирает среди наикратчайших (длина маршрута рассчитывается водителем исходя только из числа ребер в маршруте, т. е. загрузки ребер не учитываются). Предложите «оптимальный» (в смысле предыдущей задачи) способ управления такой системой светофоров.



Имеет ли смысл в этой транспортной сети (при описанном выше типичном режиме работы) делать односторонними некоторые дороги, и если делать, то какие именно? Очевидно, что это увеличит пропускную способность этих дорог в одну сторону, но одновременно увеличится и перепробег.

Литература

1. `http://lihodeev.com/pubs.html`

**Задача о выделенных полосах и расщеплении потоков** (М. С. Ишманов). **а)** По МКАД циркулируют личные автомобили и общественный транспорт. Личные автомобили появляются в $\Delta x$-окрестности произвольной точки МКАД $x$ согласно пуассоновскому процессу с интенсивностью $\lambda_{\text{Car}} x$, причем это независимо происходит во всех точках МКАД. Далее, появившийся автомобиль проезжает путь, длина которого есть независимая (ни от чего) случайная величина, равномерно распределенная от нуля до половины длины МКАД. Общественный же транспорт, не останавливаясь, ездит по кругу, собирая пассажиров, появляющихся в случайных точках на МКАД с интенсивностью $\lambda_{\text{Bus}}$ [человек/единицу длины] и проезжающих путь, длина которого также есть независимая случайная величина, равномерно распределенная от нуля до половины длины МКАД. Считайте, что средняя скорость движения по МКАД $V = 30$ км/час, значение потока $Q = 8000$ авт/час, средняя доля общественного транспорта составляет $\eta = 5\,\%$, желаемая скорость общественного транспорта (по выделенной полосе) $V_{\text{Bus}} = 60$ км/час, типичный автомобиль перевозит в среднем двух пассажиров, а типичный автобус перевозит в среднем 50 пассажиров. Стоит ли из 5 полос (в одну сторону) МКАД выделить одну полосу для общественного транспорта, если критерием является суммарное среднее время в пути всех пользователей МКАД за достаточно большой промежуток времени? Считайте, что если выделенная полоса будет введена, то это снизит среднюю скорость движения автомобилей по оставшимся 4 полосам до $V = 20$ км/час.

Заметим, что в этом пункте не учитывается, что открытие выделенной полосы может привести к переходу части водителей с личного транспорта на общественный.

**б) Метаигровой синтез.** Считаем, что на МКАД постоянно в течение месяца (30 дней, 24 часа в день) приходят пользователи (каждый из

которых имеет собственный автомобиль, но не обязательно им пользующийся) согласно условиям п. а), с известными интенсивностями. На МКАД сделали выделенную полосу для общественного транспорта. Пусть $M$ — максимально возможное число автобусов, которые потенциально можно вовлечь в перевозку пассажиров по МКАД по выделенной полосе, при этом скорость перевозки $V_{\text{Bus}}$ не зависит от $m$ — реального числа автобусов на МКАД (в одну сторону; вместимость автобусов известна и равна $K$). Зависимость $V(Q)$ ($V'(Q) < 0$) для одной «агрегированной» полосы МКАД считайте известной. Считайте, что каждый пользователь МКАД выбором типа транспортного средства $s = \{\text{Bus, Car}\}$ старается уменьшить затраты на поездку, которые рассчитываются по формуле:

$$F(s) = \begin{cases} \omega T_{\text{Bus}} + \zeta + D, & s = \text{Bus}, \\ \omega T_{\text{Car}} + \xi T_{\text{Car}}, & s = \text{Car}, \end{cases}$$

где $T_s$ — время в пути, $\omega$ — цена единицы времени в пути, $\zeta$ — плата за проезд на общественном транспорте, $D$ — цена дискомфорта, связанного с использованием общественного транспорта и отказа от возможности пользоваться собственным автомобилем, $\xi$ — цена топлива.

Целью является оптимально подобрать «правила игры» $m$ и $\zeta$:

$$\omega h(m, \zeta) + \psi m \to \min_{\substack{0 \leqslant m \leqslant M \\ \zeta \geqslant 0}}, \tag{1}$$

где $h(m, \zeta)$ — среднее время в пути всех пользователей за месяц, $\psi$ — затраты в месяц на содержание одного автобуса.

Литература

1. *Бурков В. Н.* Основы математической теории активных систем. М.: Наука, 1977.



2. *Гудвин Ф.* Решение проблемы пробок; `http://www.polit.ru/article/2009/03/24/probki/`, 2009.
3. *Вучек В. Р.* Транспорт в городах, удобных для жизни. М.: Территория будущего, 2011.

**Задача: платные дороги и метаигровой синтез** (Ю. В. Дорн). Скажем несколько слов о платных дорогах, которые, на наш взгляд, могут частично решить проблему пробок в Москве, подобно Сингапуру. В перспективе плату за проезд по дорогам можно будет рассчитывать исходя из информации о треках автомобилей. Планируется оснастить все автомобили GPS (ГЛОНАСС)-навигаторами, позволяющими определять каждые 5 минут положения автомобиля с точностью до нескольких десятков метров. Каждая дорога имеет свой тариф: за разовый проезд по дороге начисляется плата, подобно плате за разговор по мобильному телефону. В конце месяца приходит счет.

Введем Центр (государство). Обозначим через $\tau_e$ плату (штраф), взимаемую с каждого водителя за проезд по ребру $e$. Положим $\vec{\tau} = \{\tau_e \colon e \in E\}$. После того, как игрок Центр выбирает ту или иную допустимую стратегию $\vec{\tau}$, издержки на маршрутах, вообще говоря, возрастают $G_p(\vec{x}; \vec{\tau}) = \sum\limits_{e \in E} \delta_{ep} \cdot (\tau_e(y_e(\vec{x})) + \tau_e)$, что приводит к перераспределению потоков по маршрутам: изменению рядом игроков-водителей своих стратегий (маршрутов). Это перераспределение потоков происходит до тех пор, пока не будет достигнуто новое равновесное распределение по путям $\vec{x}^*(\vec{\tau})$ ($\vec{y}^*(\vec{\tau})$ по дугам). Этому равновесию соответствуют свои суммарные издержки всех пользователей сети в единицу времени

$$L(\vec{\tau}) = \sum_{p \in P} G_p(\vec{x}^*(\vec{\tau}); \vec{0}) x_p^*(\vec{\tau}).$$

Задача Центра — так подобрать $\vec{\tau}$, чтобы

$$L(\vec{\tau}) = \min_{\vec{x} \in X} \sum_{p \in P} G_p(\vec{x}; \vec{0}) x_p = \sum_{p \in P} G_p(\vec{x}^{\text{opt}}; \vec{0}) x_p^{\text{opt}}. \tag{2}$$

Вектор $\vec{x}^{opt}$ называют системным оптимумом. Естественно, что в системном оптимуме системе не хуже, чем в равновесии. Насколько большой может быть «цена анархии»

$$\frac{\sum\limits_{p \in P} G_p(\vec{x}^*; \vec{0}) x_p^*}{\sum\limits_{p \in P} G_p(\vec{x}^{\text{opt}}; \vec{0}) x_p^{\text{opt}}},$$

«не зависит» от размеров графа транспортной сети. Далее приводятся в ряде случаев неулучшаемые оценки сверху для цены анархии (оценка

цены анархии приводится для худшего случая, который во всех случаях достигается на графах с небольшим числом вершин):

| $\tau_e(y_e)$ | Цена анархии |
|---|---|
| $a + by$ | 1,333 |
| $ay^2 + by + c$ | 1,626 |
| $ay^3 + by^2 + cy + d$ | 1,896 |
| Полином степени $n \gg 1$ | $n/\log n$ |

Примечательно во всем этом то, что плату достаточно взимать с дуг (дорог), а не с путей. Это крайне важно для практики и, вообще говоря, сразу может быть и не очевидно!

В ряде случаев бороться с пробками помогают и более радикальные способы. Например, полное или частичное перекрытие нескольких дорог. Исходя их парадокса Браесса, введите определение неэффективных ребер (дорог) и предложите способ их отыскания, а также метод борьбы с такими ребрами.

Наряду с постановкой задачи об оптимальном взимании платы за проезд (2) можно также рассматривать и такую постановку:

$$\min_{\vec{\tau}\geqslant\vec{0}} \sum_{p\in P} G_p(\vec{x}^*(\vec{\tau});\, \vec{\tau})x_p^*(\vec{\tau}). \qquad (3)$$

Определите решение задачи (3) для парадокса Браесса и сравните с решением, полученным в постановке (2).

Литература

1. *Sandholm W.* Evolutionary Implementation and Congestion Pricing // Review of Economic Studies. 2002. V. 69(3). P. 667–689.
2. *Roughgarden T.* The Price of Anarchy is Independent of the Network Topology. Proceedings of the 34th Annual ACM Symposium on the Theory of Computing, May 2002.

**Задача о построении композиции алгоритмов для краткосрочного прогнозирования** (Ю. В. Чехович, Н. П. Ивкин). Рассматривается задача краткосрочного (с горизонтом от 20 до 120 минут) прогнозирования скоростей движения автомобилей по дорожной сети города Москвы. Исходные данные поступают в виде треков движения от автомобилей, оборудованных специальными устройствами. Устройства передают координаты автомобиля в привязке к моментам времени. Координаты определяются с некоторой ошибкой. Обычно ошибка определения координат находится в диапазоне от 10 до 50 метров. Устройствами оборудована небольшая часть всех автомобилей, поэтому более-менее точная и актуальная информация существует только о магистральных улицах города — примерно 20% всех ребер



графа дорожной сети. Программы предобработки преобразуют получаемые треки автомобилей в средние скорости движения на ребрах графа. Усреднение осуществляется в пределах коротких (1, 2, 4 минуты) временных интервалов. Для прогнозирования скорости движения на ребре графа (или среднего времени проезда этого ребра) используются несколько (от 10 до 20) относительно простых базовых алгоритмов, использующих исторические данные по этому ребру. При поступлении данных по каждому следующему интервалу времени каждый базовый алгоритм дает прогноз несколько интервалов времени вперед (так, чтобы обеспечивать прогнозирование на требуемый горизонт). В дополнение к историческим данным накоплены значения прогнозов, которые были рассчитаны каждым базовым алгоритмом в каждый момент времени для каждого интервала времени в пределах горизонта прогнозирования. Также рассчитаны ошибки каждого прогноза по отношению к фактическим значениям.

Таким образом, входные данные можно представить в виде пространства ⟨ребро графа, момент расчета прогноза, интервал прогнозирования⟩. Для каждой точки в этом пространстве известны фактическое значение скорости и ошибки каждого алгоритма. Объем таких данных для одной недели наблюдений при усреднении скоростей, например, на двухминутных интервалах составляет 20 000 (значимых ребер графа) × 5040 (двухминутных интервалов в неделе) × 60 (интервалов прогнозирования) × 15 (значений ошибок алгоритмов), т. е. приблизительно 360 Гб.

Необходимо предложить алгоритм, который бы позволил надежно выделить в данных «области компетентности» базовых алгоритмов и построить композицию из базовых алгоритмов, ошибка которого на исследуемой неделе должна быть меньше, чем ошибка любого базового алгоритма. Композиция может не покрывать все описанные данные, а относится только к их значимой части. Контрольная проверка полученной композиции осуществляется на данных следующей недели. Задача считается успешно решенной, если контроль подтверждает результаты обучения.

Альтернативной задачей является оценка качества имеющихся данных с точки зрения принципиальной возможности построения указанной композиции.

**Задача: получение статистических характеристик транспортных потоков по данным видеозаписей** (Ю. В. Чехович). Треки автомобилей, оборудованных специальными устройствами, позволяют хорошо оценивать скорости движения на различных участках транспортной сети, но при этом практически не дают возможности оценить плотность транспортных потоков. Одним из источников данных такого рода могут стать видеокамеры, например в сервисе «Яндекс.Пробки» (`http://maps.yandex.ru/`) или в сервисе «Пробки из окна» (`http://www.probkiizokna.ru/`).


Рис. 2. Изображения из сервиса «Яндекс.Пробки»


Рис. 3. Изображения из сервиса «Пробки из окна»

Необходимо разработать средство обработки видеоданных, позволяющее оценивать количества транспортных средств, проезжающих в единицу времени в поле зрения камеры. Средство должно позволять: (1) обособленно обрабатывать несколько проезжих частей, находящихся в поле зрения; (2) стабильно работать в условиях плохой видимости, нестабильного канала передачи видеопотока, низкой скорости транспортного потока (пробки); (3) иметь возможность обособленной обработки нескольких полос в рамках каждой проезжей части.

**Задача: расчет количества полос движения в транспортной сети мегаполиса на основе спутниковых изображений** (Ю. В. Чехович). При работе с графом дорожной сети возникает необходимость оценки предельной пропускной способности участков дорожной сети. Учитывая, что официальные данные о ширине проезжих частей и количестве полос на улицах города Москвы часто труднодоступны, предлагается разработать способ определения полосности улиц города Москвы на основе изображений спутниковой съемки, используемых в общедоступных сервисах, таких как «Карты Google» (`http://maps.google.ru/`) или «Яндекс.Карты»



Рис. 4. Спутниковое изображение улиц Москвы сервиса «Карты Google». Небольшое количество автомобилей

Рис. 5. Спутниковое изображение улиц Москвы сервиса «Карты Google». Пробка и свободная проезжая часть

(`http://maps.yandex.ru/`). Этот способ должен уметь выделять полосы как в условиях небольшого количества транспортных средств на изображении, так и в условиях высокой плотности автомобилей. Желательно оценивать также долю проезжей части, занятую припаркованными автомобилями.

### Литература к разделу

1. Multi-agent systems for traffic and transportation engineering / Ana L. C. Bazzan and Franziska Klugl (Ed.). Information science reference. New York: Hershey, 2009.

2. *Cascetta E.* Transportation systems analysis. Models and applications. Springer, 2009. (Optimization and application. V. 29).
3. Vehicular networks: from theory to practice / Stephan Olariu and Michele C. Weigle (Ed.). 2009. (Chapman & Hall/CRC Computer & Information Science Series. V. 20).
4. *Braker J. G.* Algorithms and applications in timed discrete event systems. PhD Thesis. 1993.
5. *Patriksson M.* The traffic assignment problem. Models and methods. Utrecht Netherlands: VSP, 1994.
6. Transport planning and traffic engineering / C. A. O'Flaherty (Ed.). Elsevier, 2006.
7. *Garber N. J.*, *Hoel L. A.* Traffic & highway engineering. Virginia: Nelson Engineering, 2010.
8. *Mannering F. L.*, *Washburn S. S.*, *Kilareski W. P.* Principles of highway engineering and traffic analysis. John Wiley and Sons Ltd., 2008.
9. *Daganzo C. F.* Fundamentals of Transportation and Traffic Operations. Oxford, U.K.: Pergamon–Elsevier, 1997.

# Практическое приложение

**А. В. Прохоров, В. Л. Швецов**

## О практическом опыте моделирования транспортных потоков с помощью пакета программ PTV Vision ®

В данном приложении будут рассмотрены некоторые важные, на взгляд авторов, практические аспекты моделирования транспортных потоков.

При планировании развития транспортной инфраструктуры города или региона возникает ряд определенных задач, например:

— хранение обширной базы данных транспортных и социально-экономических показателей;

— расчет величины транспортных и пассажирских потоков на текущую ситуацию;

— оценка различных транспортных ситуаций и вариантов развития транспортной инфраструктуры по заданной системе показателей, что делает возможным управление транспортными потоками на основе сравнимых количественных значений;

— прогноз объемов пассажиропотоков и интенсивности движения на участках транспортной сети;

— технико-экономическое обоснование различных инвестиционных проектов в развитие транспортной инфраструктуры моделируемого региона;

— оптимизация потоков индивидуального транспорта;

— оптимизация работы общественного транспорта.

Транспортные модели в среде современных программных комплексов транспортного планирования помогают в решении указанных задач и общем процессе управления транспортной системой на стратегическом уровне.

Наиболее яркими примерами с точки зрения практического применения являются, по мнению авторов, следующие задачи:

— принятие решения о строительстве автомобильной дороги и выбор ее параметров (место прохождения, количество полос движения, разрешенная скорость и т. д.);

— оценка целесообразности ввода нового маршрута общественного транспорта;

— оценка последствий ввода выделенных полос для движения общественного транспорта;

— оценка влияния строительства нового жилого (делового и т. д.) района (бизнес-центра и т. д.) в области моделирования;

— определение оптимальных, как с точки зрения концессионера, так и с точки зрения пользователей, тарифов за проезд по платным участкам автомобильных дорог.

Транспортную модель можно условно разделить на два основных блока — *транспортное предложение* и *транспортный спрос*. Модель транспортного предложения описывает транспортную инфраструктуру моделируемой области — транспортную сеть, состоящую из узлов (перекрестков, развязок и т. д.) и соединяющих их ребер (улиц, дорог и т. д.), которая предоставляет возможность перемещения для участников транспортного движения и описывает затраты на данные перемещения. Модели спроса на транспорт описывают качественно и количественно перемещения с учетом причин их возникновения, различных видов транспорта и путей следования.

В мировой практике классическим подходом к расчету модели транспортного спроса является так называемая четырехшаговая модель:

1) **расчет объемов прибытий и отправлений** по каждому району; результатами расчета являются сумма элементов по строкам и столбцам матриц корреспонденций (Trip generation);

2) **расчет** общих межрайонных **корреспонденций**: расчет объемов перемещений между каждой парой транспортных районов; результатами расчета являются элементы матриц корреспонденций (Trip distribution);

3) **расщепление** общих межрайонных **корреспонденций** по способам передвижений (видам транспорта): пешие передвижения, передвижения с использованием общественного транспорта, передвижения на личном автомобиле и т. д.; результатами расчета являются элементы матриц корреспонденций (Modal split);

4) **распределение корреспонденций по транспортной сети**, то есть определение всех путей, выбираемых участниками движения, и определение количества передвижений по каждому пути; результатами расчета являются модельные значения интенсивности транспортных потоков и объемы пассажиропотоков по участкам транспортной сети (Trip assignment).

На данный момент существуют и другие подходы к расчету транспортного спроса, например, activity-based models или модификации четырехшаговой модели (модель одновременного и взаимосвязанного расчета



шагов определения корреспонденций и их расщепления) и другие. Поэтому приведенное разделение достаточно условно, так как все этапы расчета взаимосвязаны. Как уже отмечалось в главе 1, расчет корреспонденций и их расщепление приводит к более точным результатам, если уже известна загрузка сети (распределение потоков), которая, в свою очередь, требует матрицу корреспонденций и информацию о расщеплении потоков. Именно поэтому возникает необходимость решать задачу последовательными приближениями, повторяя все шаги в итеративном режиме. Однако на практике данный подход (классический вариант четырехшаговой модели), особенно в случае сложных транспортных систем крупных городов и мегаполисов (например, Москвы или Санкт-Петербурга), является наиболее приемлемым с точки зрения соотношения затраты/качество (время разработки, трудозатраты и качество результатов моделирования) для создания модели и проведения последующих расчетов с ее использованием.

Далее будут последовательно рассмотрены различные примеры проектов, которые были выполнены для ряда городов России [1)] (Санкт-Петербург, Пермь, Тверь, Новый Уренгой и др.).

Одним из наиболее актуальных проектов является проект «Разработка научно обоснованных предложений по развитию транспортной системы Санкт-Петербурга и Ленинградской области на период до 2020 года». В рамках данного проекта был проведен анализ существующего состояния транспортной системы Санкт-Петербурга и Ленинградской области и прогноз объемов перевозки пассажиров в Санкт-Петербургском транспортном узле до 2020 года. Базовым годом для разработки транспортной модели являлся 2011 год, прогнозным 2020 год. Областью исследования является Санкт-Петербург и Ленинградская область в границах магистральной автомобильной дороги А120 (см. рис. 1). Модель была разработана и реализована в программном комплексе PTV Vision ® VISUM (см. рис. 2).

Основные исходные данные для работы — данные социально-экономического развития районов моделируемой области (численность населения, численность трудоспособного населения, среднесписочная численность работников, среднесписочная численность школьников, численность лиц школьного возраста, среднесписочная численность учащихся в вузах, численность лиц студенческого возраста), данные замеров интенсивности движения, данные транспортной сети. Модель можно охарактеризовать следующими параметрами:

- 12 моделируемых видов транспорта;

[1)] Более подробный список проектов и информацию по ним можно найти на сайте `ptv-vision.ru`

Рис. 1. Районирование г. Санкт-Петербург


Рис. 2. Транспортная сеть г. Санкт-Петербург

- 90 000 узлов с указанием типа регулирования и разрешенных маневров;
- 170 000 отрезков с атрибутикой: название, категория, количество полос, разрешенные системы транспорта, скорость движения, пропускная способность;
- 7 000 пунктов остановок общественного транспорта;
- 1 000 маршрутов движения общественного транспорта, в том числе автобусные (социальные и коммерческие), трамвайные, троллейбусные маршруты, маршруты метрополитена и пригородных электропоездов, водного транспорта;
- 500 транспортных районов с данными социально-экономической статистики;
- 4 900 примыканий («выходы в сеть»);
- 17 слоев транспортного спроса по различным причинам поездки (например: дом—работа, работа—дом, учеба—дом, дом—учеба и т. д.).

Для расчета транспортного спроса использовался классический вариант четырехшаговой модели. Для определения параметров и коэффициентов для каждого из шагов расчета был проведен социологический опрос населения. В рамках социологического опроса опрашиваются респонденты о совершенных ими за предыдущий день перемещениях с указанием их целей, используемого транспорта, времени и других параметров. По результатам обработки опроса были получены следующие величины, которые были использованы для расчета транспортного спроса:

— статистическое распределение количества совершаемых жителями города перемещений по различным целям и социально-экономическим группам. Из этих данных могут быть рассчитаны коэффициенты подвижности (генерации и притяжения) населения по различным целям и социально-экономическим группам, которые необходимы для расчета первого шага транспортного спроса;

— статистическое распределение совершаемых жителями города перемещений по длительности их совершения с разделением по целям перемещений, например, из дома на работу, видам используемого транспорта, например, легковой автомобиль или автобус и метро, и социально-экономическим группам, например, пенсионеры, студенты, занятые в экономике с различным уровнем дохода. Указанные данные используются на втором и третьем шагах расчета транспортного спроса для преобразования затрат между каждой парой районов в субъективные вероятности совершения поездки в зависимости от целей, социально-экономических групп и видов транспорта.



Расчет перспективной интенсивности движения основан на анализе и прогнозе показателей социально-экономического развития и развития транспортной инфраструктуры. Общий алгоритм прогнозирования с использованием транспортных моделей можно представить следующей последовательностью действий:

— сбор необходимых исходных данных для построения транспортной модели: данные транспортного предложения (сеть автомобильных и прочих дорог, данные социально-экономической статистики по транспортным районам);

— проведение обследования интенсивности движения транспорта для оценки существующей ситуации, калибровки модели и последующей оценки точности транспортной модели;

— разработка модели транспортных потоков в области тяготения исследуемого объекта на основе исходных данных;

— моделирование транспортных потоков на базовый год, калибровка модели, оценка точности;

— определение параметров для прогнозных вариантов, включая перспективную дорожную сеть на расчетные сроки и прогноз данных социально-экономической статистики;

— прогноз перспективной интенсивности движения по определенным вариантам (см. рис. 3).

В результате проведенных работ в рамках данного проекта были рассчитаны перспективные интенсивности движения и пассажиропотоки, на основе которых были построены картограммы изменения нагрузки и проведена оценка ряда общесетевых показателей. В качестве таких показателей, в том числе, использовались:

- Средняя скорость поездки на общественном транспорте. Данный показатель может быть описан как средневзвешенная по объему пассажиропотока скорость перемещения с использованием общественного транспорта (от двери до двери) в зоне моделирования за моделируемый период (сутки).
- Средняя скорость поездки на индивидуальном транспорте. Данный показатель аналогичен по своей сути и расчету соответствующему показателю для общественного транспорта и может быть описан как средневзвешенная по объему транспортного потока скорость перемещения с использованием легковых автомобилей (от двери до двери) в зоне моделирования за моделируемый период (сутки).
- Среднее время поездки на индивидуальном транспорте. Данный показатель может быть описан как средневзвешенное по объему

Рис. 3. Картограмма нагрузки на сеть индивидуального транспорта



транспортного потока время перемещения с использованием легковых автомобилей (от двери до двери) в зоне моделирования за моделируемый период (сутки).

- Среднее время поездки на общественном транспорте. Данный показатель аналогичен по своей сути и расчету соответствующему показателю для индивидуального транспорта и может быть описан как средневзвешенное по объему транспортного потока время перемещения с использованием общественного транспорта (от двери до двери) в зоне моделирования за моделируемый период (сутки).
- Среднее количество пересадок на общественном транспорте. Данный показатель может быть описан как средневзвешенное по объему пассажиропотока значение количества пересадок (частота пересадок) в поездках с использованием общественного транспорта (от двери до двери) в зоне моделирования за моделируемый период (сутки).
- Средний уровень загрузки магистральных и скоростных дорог. Данный показатель может быть описан как среднее значение уровня загрузки по всем магистральным и скоростным дорогам (представляемым в модели как набор направленных отрезков) в процентах (%).

Далее будут перечислены некоторые практические примеры других проектов по разработке транспортных моделей городов и регионов.

В 2007 г. ЗАО «Петербургский НИПИГрад» при поддержке специалистов ООО «А+С КонсалтПроект» в рамках выполнения проекта «Анализ и прогнозирование транспортных потоков на первую очередь Генерального плана г. Новый Уренгой» была разработана компьютерная модель общественного и индивидуального транспорта. Модель была разработана и реализована в программном комплексе PTV Vision ® VISUM.

В ходе построения модели территория г. Новый Уренгой была разделена на 26 внутренних транспортных районов и 5 кордонных транспортных районов. В качестве данных социальной статистики использовались следующие показатели транспортных районов:

— численность населения;
— численность экономически активного населения;
— число рабочих мест;
— численность учащихся;
— число учебных мест;
— число рабочих мест в сфере услуг

для существующего положения и те же показатели как прогноз на 2026 г.

Прогнозные оценки рассчитывались на основе предположения о том, что показатели социально-экономического положения транспортных рай-

онов будут расти пропорционально населению с различными коэффициентами роста, зависящими от типа транспортного района. Замеры интенсивности транспортного потока на отдельных сечениях УДС, предназначенные для калибровки модели и определения степени её адекватности реальной ситуации, проводились на 48 местах подсчёта. Был проведен сбор о суточных интенсивностях потоков индивидуального транспорта и пассажиропотоков общественного транспорта. Основным результатом работы с моделью является расчет изменения нагрузки ОТ и ИТ на УДС к 2026 г., рассчитанный для двух сценариев:

— нулевой сценарий (УДС г. Новый Уренгой не изменяется);

— активный сценарий — реализация трёх основных строительных проектов:

- ввод в эксплуатацию восточной магистрали, которая представляет собой новое соединение южного центра (ул. Промысловая) с северным центром (ул. № 20);
- строительство двухуровневой развязки на пересечении ул. Магистральной и Западной магистрали;
- строительство Южного обхода с Южной магистралью на автодороге на поселок Уралец.

Таким образом, модель г. Новый Уренгой дает пример успешного опыта применения транспортной модели в качестве инструмента построения долгосрочных прогнозов функционирования системы ГОПТ с учётом её взаимодействия с ИТ, необходимых для разработки Генерального плана города.

Транспортная модель г. Калуги была разработана по заказу управления городским хозяйством г. Калуги и предназначена для расчёта и калибровки матриц корреспонденций, описывающих распределение потоков общественного и индивидуального транспорта. Модель была реализована и настроена средствами программного комплекса PTV Vision ® VISUM. Исследуемая область была разделена на 27 транспортных районов. В качестве данных социальной статистики использовались следующие показатели транспортных районов:

— численность населения;

— численность экономически активного населения;

— число рабочих мест;

— численность учащихся;

— число учебных мест;

— число рабочих мест в сфере услуг.



За основу в построении транспортной модели была взята карта исследуемого региона. Данные по УДС, использованные при построении транспортной модели, включают в себя геометрические характеристики элементов сети, такие как количество и расположение узлов и отрезков. Данные по общественному транспорту включали в себя маршрутную сеть и данные об остановках. Замеры интенсивности транспортного потока, предназначенные для калибровки модели и определения степени её адекватности реальной ситуации, проводились на отдельных сечениях УДС. Был проведен сбор данных о почасовых интенсивностях потоков легковых автомобилей, троллейбусов и автобусов, причём учитывались различия автобусов по секционности (большие, средние, малые). Кроме того, для более точной картины дорожной ситуации были собраны дополнительные данные — о суточной интенсивности потока легковых автомобилей и об интенсивности потоков грузового транспорта. Основные результаты реализации проекта заключаются в следующем:

— сформирована транспортная модель г. Калуги, откалиброванная и настроенная с достаточной точностью для поддержки принятия решений в области оптимизации функционирования системы ГОПТ, а также разработки КТС города;

— проведен расчёт нагрузки на УДС г. Калуги.

Кроме отмеченных выше задач, довольно часто приходится сталкиваться и со следующими постановками: выбор оптимальных вариантов пересечений, организация дорожного движения на основе интенсивностей транспортных потоков, оптимизация светофорных циклов (см. рис. 4).

Для решения такого типа задач активно используется, как уточнение и дополнение макромоделей, имитационное моделирование на основе микроскопических моделей транспортного потока, описанных в главах 2 и 3. В качестве исходных данных используется информация из натурных наблюдений, например, замеренные значения входящих потоков на перекрестке в определенное время суток для определения граничных условий и начальных характеристик в микроскопических моделях.

В заключение хотелось бы отметить, что актуальность использования транспортных моделей на практике неуклонно растет с каждым годом. Современные реалии требуют системного подхода к процессу транспортного планирования и более согласованного взаимодействия различных структур при выработке и реализации транспортных решений. Постоянное повышение сложности и комплексности транспортных систем, особенно в крупных городах и мегаполисах, масштабность возникающих перед менеджерами и проектировщиками задач, необходимость взаимосвязанного учета колоссального количества факторов — все это обуславливает переход на новые методы транспортного планирования с применением компьютерных

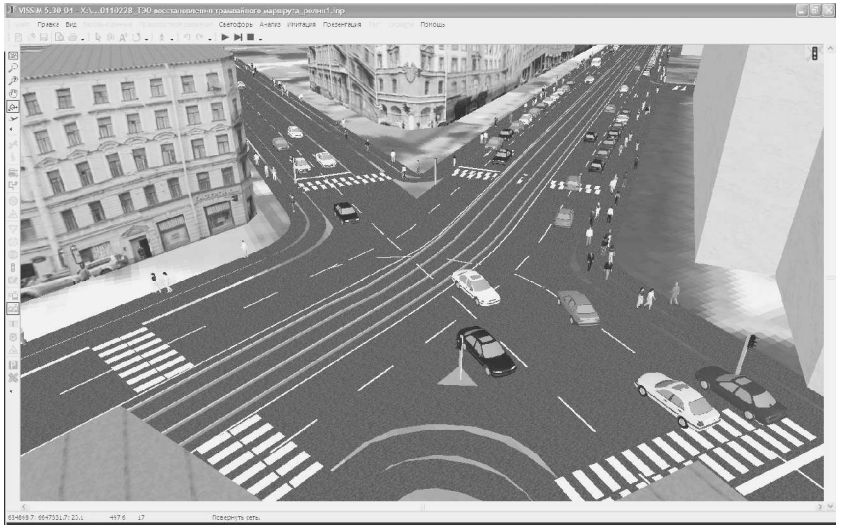


Рис. 4. Моделирование движения транспортных средств на перекрестке

транспортных моделей и комплексной оценки последствий мероприятий по развитию транспортной инфраструктуры.

Научное издание

*Гасников Александр Владимирович*
*Кленов Сергей Львович*
*Нурминский Евгений Алексеевич*
*Холодов Ярослав Александрович*
*Шамрай Наталья Борисовна*

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТРАНСПОРТНЫХ ПОТОКОВ

Редактор Т. Л. Коробкова

Подписано в печать ??.??.2012 г. Формат $60 \times 90\ ^1/_{16}$. Бумага офсетная. Печать офсетная. Печ. л. ??. Тираж 1000 экз. Заказ №

Издательство Московского центра непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241-74-83.

Отпечатано с готовых диапозитивов в ППП «Типография “Наука”»
121099, Москва, Шубинский пер., 6

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга», Большой Власьевский пер., д. 11. Тел. (499) 241–72–85. E-mail: biblio@mccme.ru
http:// biblio.mccme.ru