

Министерство образования и науки РФ
Нижегородский государственный университет им. Н.И. Лобачевского
Национальный исследовательский университет

Ю.И. Неймарк

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КАК НАУКА И ИСКУССТВО

Учебник

Издание второе, исправленное и дополненное

Нижний Новгород
Издательство Нижегородского университета
2010

УДК 519.8
ББК В143
Н 45

Неймарк Ю.И.

Н 45 **Математическое моделирование как наука и искусство:** Учебник. –
2-е изд., испр. и доп. – Н. Новгород: Изд-во Нижегородского госуниверсите-
та, 2010. – 420 с.

ISBN 978–5–91326–145–8

В учебнике рассказывается о разнообразных конкретных математических эволюционных моделях и их исследовании в механике, физике, биологии, технике и управлении, моделях игр и поиска решения, волновых явлениях, фундаментальных законах макромира и микромира, пространстве и времени и др.

Для студентов, аспирантов, преподавателей, инженеров и всех интересующихся математическим моделированием.

Печатается по решению Ученого совета ННГУ

ISBN 978–5–91326–145–8

ББК В143

© Ю.И. Неймарк, 2010
© Нижегородский госуниверситет
им. Н.И. Лобачевского, 2010

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	X
ВВЕДЕНИЕ	XIV
Математика как язык. Математическая модель. Математическое моделирование как метод изучения окружающего мира	
1. ДИНАМИЧЕСКАЯ СИСТЕМА	1
Детерминизм Лапласа. Динамическая система как основная математическая модель естествознания. Фазовый портрет. Примеры динамических систем и их фазовых портретов: простейшие примеры, игра «жизнь», маятник на вращающемся основании. Фазовый портрет как средство отображения наших знаний о динамике модели и как средство ее изучения	
2. ВЫТЕКАНИЕ ЖИДКОСТИ ИЗ СОСУДА	19
Закон Торричелли и простейшая модель вытекания. Эффект сжатия струи. Недостаточность простейшей модели. Уточненные модели, учитывающие разгон жидкости. Фазовый портрет и отображение на нем быстрой фазы разгона и медленного вытекания. Уточнение фазы разгона	
3. РАВНОВЕСИЕ И АВТОКОЛЕБАНИЯ УРОВНЯ ЖИДКОСТИ В СОСУДЕ ПРИ ОДНОВРЕМЕННОМ ПРИТОКЕ И ОТТОКЕ	30
Равномерный приток и отток через дырочку и через сифон	
4. ПЕРЕХОДНЫЙ ПРОЦЕСС, СОСТОЯНИЕ РАВНОВЕСИЯ И АВТОКОЛЕБАНИЯ	34
Подводится итог полученным знаниям о возможных типах движений: переходные процессы, устойчивые равновесия и автоколебания	
5. ДИНАМИКА УРОВНЯ ЗЕРКАЛА ВОДОХРАНИЛИЩА С ГИДРОСТАНЦИЕЙ	37
Критические значения. Бифуркационная диаграмма	
6. ЭНЕРГЕТИЧЕСКАЯ МОДЕЛЬ СЕРДЦА	42
Виды кризисных состояний. Критические значения. Сужение жизненных возможностей	
7. ЗАСОЛЕНИЕ ВОДОЕМА С ЗАЛИВОМ И ЗАГАДКИ КАСПИЙСКОГО МОРЯ	46
Загрязнение водоема сточными водами с растворимыми загрязнителями. Загадка Каспийского моря. Засоление водоема с заливом. Равновесный водный режим и равновесная соленость. Возможность непредсказуемых смен равновесного уровня	
8. ЭКСПОНЕНЦИАЛЬНЫЕ ПРОЦЕССЫ	55
Математическая модель. Периоды полураспада и удвоения. Примеры экспоненциальных процессов: размножение и гибель, радиоактивность, цепные реакции, разряд конденсатора, разгон ракеты, торможение, поглощение излучения, охлаждение, распространение эпидемии и слухов, рост численности населения, производства, знаний, приближение и удаление от равновесия и др. Уточнение модели: учет насыщения, взрывной характер роста. Явление внезапного кризиса, «схлопывания» и исключительности	

9. ДИНАМИКА СОСУЩЕСТВОВАНИЯ ПОПУЛЯЦИЙ	65
Математические модели сосуществования: «хищник – жертва», конкуренция (противостояние), симбиоз. Фазовые и бифуркационные портреты	
10. ПРОТОЧНЫЙ БИОЛОГИЧЕСКИЙ РЕАКТОР	74
Простейшая модель биологического реактора (хемостат): фазовый портрет, оптимизация	
11. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ИММУННОГО ОТВЕТА ОРГАНИЗМА НА ВТОРЖЕНИЕ ИНФЕКЦИИ	77
Рассматривается упрощенная феноменологическая модель иммунного ответа организма на инфекцию в виде системы дифференциальных уравнений четвертого порядка, которая обнаруживает основные варианты течения заболевания: выздоровление и выздоровление через обострение, гибель и гибель от истощения ресурсов, хроническое заболевание с равновесным и периодическим течением, бактерионосительство, их зависимости	
12. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СООБЩЕСТВА «ПРОИЗВОДИТЕЛИ–ПРОДУКТ–УПРАВЛЕНЦЫ»	83
Возможные виды фазовых портретов. Эволюция сообщества с ростом технологии производства в зависимости от остальных параметров	
13. ЛИНЕЙНЫЙ ОСЦИЛЛЯТОР	92
Математическая модель линейного осциллятора. Возможные типы движений. Фазовый и бифуркационный портреты. Что описывает линейный осциллятор: типы равновесий, гармонические, затухающие и нарастающие колебания	
14. ЭЛЕКТРОМЕХАНИЧЕСКИЕ АНАЛОГИИ. УРАВНЕНИЯ ЛАГРАНЖА – МАКСВЕЛЛА	103
Электромеханические аналогии. Уравнения Лагранжа–Максвелла механических, электрических и электромеханических систем. Инвариантность уравнений Лагранжа. Принцип наименьшего действия как вариационная форма математических моделей. Примеры и уравнения Лагранжа – Максвелла с обобщенными силами	
15. ЧАСЫ ГАЛИЛЕЯ – ГЮЙГЕНСА	113
Как и почему появились часы Галилея–Гюйгенса, что в них принципиально нового, что определяет точность часов? Анализ причин погрешностей и пути их уменьшения. Простейшая математическая модель часов Галилея–Гюйгенса. Фазовый портрет. Отображение А. Пуанкаре и диаграмма точечного отображения Кенигса – Ламерея. Часы как автоколебательная система, как система с обратной связью	
16. ГЕНЕРАТОР ЭЛЕКТРИЧЕСКИХ КОЛЕБАНИЙ	124
Генератор электрических колебаний как электрический аналог часов Галилея – Гюйгенса. Математическая модель. Уравнение Ван-дер-Поля и его фазовый портрет. Приближенное исследование уравнения Ван-дер-Поля и сведение к точечному отображению. Мягкий режим возбуждений автоколебаний	
17. МЯГКИЙ И ЖЕСТКИЙ РЕЖИМЫ ВОЗБУЖДЕНИЯ АВТОКОЛЕБАНИЙ	130
Фазовые портреты и бифуркационные портреты мягкого и жесткого возбуждения автоколебаний на основе исследования точечного отображения	

18. СТОХАСТИЧЕСКИЙ ОСЦИЛЛЯТОР («ЧАСЫ НАОБОРОТ»)	135
Осциллятор с непредсказуемым поведением. Математическая модель, фазовый портрет, точечное отображение. Непредсказуемость и случайность. Два типа поведения динамических систем: устойчивость и неустойчивость, предсказуемость и непредсказуемость. Осциллятор (динамическая система) как генератор стохастических колебаний	
19. НЕУСТОЙЧИВОСТЬ И АВТОКОЛЕБАНИЯ, ВЫЗЫВАЕМЫЕ ТРЕНИЕМ	143
Математическая модель и ее фазовый портрет. Классическая кулоновская модель сухого трения и необходимость ее уточнения. Возбуждение колебаний трением и возникновение автоколебаний, исчезновение трения	
20. ВЫНУЖДЕННЫЕ КОЛЕБАНИЯ ЛИНЕЙНОГО ОСЦИЛЛЯТОРА	152
Математическая модель. Амплитудно-фазовая частотная характеристика. Явления резонанса и фазового сдвига. Примеры килевой качки корабля и динамического гасителя колебаний. Спектральный подход и принцип суперпозиции. Обобщенная δ -функция	
21. ПАРАМЕТРИЧЕСКОЕ ВОЗБУЖДЕНИЕ И СТАБИЛИЗАЦИЯ	162
Математическая модель. Явления параметрического возбуждения и резонанса. Особенности параметрического резонанса, отличающие его от обычного резонанса	
22. НОРМАЛЬНЫЕ КОЛЕБАНИЯ И БИЕНИЯ	169
Математическая модель двух взаимодействующих осцилляторов. Нормальные колебания и перекачка энергии, вызывающая явление биений	
23. СТАБИЛИЗАЦИЯ ПЕРЕВЕРНУТОГО МАЯТНИКА	174
Управление как могучее средство изменения поведения и свойств динамических объектов и систем. Жонглирование вертикально стоящей на опоре и двумя стоящими друг на друге палочками. Математические модели. Стратегия управления, условия стабилизации, роль запаздывания	
24. УПРАВЛЯЕМЫЙ МАЯТНИК И ДВУНОГАЯ ХОДЬБА	184
Перевернутый управляемый маятник как математическая модель автоколебательной двуногой ходьбы	
25. ДИНАМИЧЕСКИЕ МОДЕЛИ ИГР, ОБУЧЕНИЯ И ЦЕЛЕСООБРАЗНОГО ПОВЕДЕНИЯ	192
Автоматные модели игроков и динамические детерминированные и стохастические модели игр	
26. ПЕРСЕПТРОН И РАСПОЗНАВАНИЕ ОБРАЗОВ	207
Рассказывается о принципе работы перцептрона Розенблата, трактуемого как динамическая система. Описываются математические модели объекта, образа, распознавания и обучения распознаванию. Формулируется и доказывается теорема о конечности числа ошибок. Обнаруживается связь между алгоритмом обучения перцептрона и стохастической аппроксимацией	
27. ЗАКОНЫ КЕПЛЕРА И ПРОБЛЕМА ДВУХ ТЕЛ, РЕШЕННАЯ НЬЮТОНОМ	215
Рассказывается о проблеме двух тел и ее связи с некоторыми вопросами астрономии: черные дыры, расширение Вселенной, эволюция Солнечной системы	

28. РАСПРЕДЕЛЕННЫЕ ДИНАМИЧЕСКИЕ МОДЕЛИ МЕХАНИКИ И ФИЗИКИ	228
Определяется понятие распределенной динамической системы. Описываются распределенные модели классической математической физики, механики, электродинамики и квантовой механики: уравнения Эйлера и Навье – Стокса, Максвелла и Шредингера	
29. ФУНДАМЕНТАЛЬНОЕ РЕШЕНИЕ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ	237
Рассказывается о расплывании поля температур, диффундирующего вещества или плотности вероятности блуждающей случайной частицы, поначалу сосредоточенных в некоторой точке. Решается задача о прогреве полупространства	
30. БЕГУЩИЕ ВОЛНЫ И ДИСПЕРСИОННОЕ УРАВНЕНИЕ	245
Рассказывается о бегущих гармонических волнах. Решается задача о температуре поверхностного слоя земли под влиянием суточных и годовых колебаний над ним. Решается задача о скорости намерзания льда на поверхности воды	
31. ТЕОРИЯ ЭЛЕКТРОМАГНЕТИЗМА ФАРАДЕЯ – МАКСВЕЛЛА И ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ МАКСВЕЛЛА – ГЕРЦА	253
Показывается наличие у уравнений Максвелла решения в виде бегущей со скоростью света гармонической волны	
32. ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ ВОЛН	257
Находятся отраженная и преломленная волны при набегании на границу раздела сред бегущей гармонической волны	
33. СТОЯЧИЕ ВОЛНЫ И КОЛЕБАНИЯ ОГРАНИЧЕННОЙ СТРУНЫ	260
Излагается метод Фурье изучения колебаний струны. Решается задача о колебаниях струны, вызванных сосредоточенным ударом. Выясняется спектр и его связь с «окраской» звука	
34. МИКРОЧАСТИЦА	265
Рассказывается об особенностях микромира. О необходимости пересмотра привычных представлений классической физики. Излагается формализм квантовой механики и приводятся примеры его применения. Описываются понятия квантового состояния, оператора его изменения и его связь с измеряемыми и наблюдаемыми физическими величинами. Рассматривается поведение квантовой частицы в потенциальной яме, в радиально-симметричном электрическом поле заряда	
35. ПРОСТРАНСТВО И ВРЕМЯ	284
Рассказывается о преобразованиях Галилея и Лоренца. Об инвариантности уравнений Ньютона и неинвариантности уравнений Максвелла по отношению к преобразованиям Галилея. Выводятся преобразования Лоренца из свойства сохранения волнового уравнения, вытекающего из уравнений Максвелла, и представлений однородности и изотропности пространства и однородности времени. Рассматривается связь между событиями, наблюдаемыми в разных системах координат: сокращение расстояний и замедление часов, сложение скоростей. Вводится четырехмерное пространство – время. Обсуждаются понятие одновременности и принцип причинности. Указываются необходимые уточнения уравнений Ньютона в случае очень больших скоростей движения, приближающихся к скорости света	

36. РАЗГОН РЕЛЯТИВИСТСКИХ ЧАСТИЦ В ЦИКЛОТРОНЕ	296
Рассматривается модель движения релятивистской частицы в циклотроне. Обнаруживается наличие квазистационарного устойчивого резонансного разгона	
37. МАТЕМАТИКА КАК ЯЗЫК, КАК ОПЕРАЦИОННАЯ СИСТЕМА И МОДЕЛИ	300
Математика как особый язык, включающий операционную систему и математические модели. Изоморфизм операционных систем и моделей. Роль изоморфизма в математическом моделировании	
38. ГЕОМЕТРИЧЕСКОЕ, ФИЗИЧЕСКОЕ, АНАЛОГОВОЕ, МАТЕМАТИЧЕСКОЕ И СТАТИСТИЧЕСКОЕ ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ	306
О геометрическом, физическом, математическом моделировании и лежащих в их основе изоморфизмах, о критериях моделируемости и правилах моделирования	
39. ОБЩАЯ СХЕМА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ	314
О двух разных взаимосвязанных частях математического моделирования (составление модели и ее исследование) и основных этапах моделирования и некоторых полезных соображениях	
40. МОДЕЛИ ВИБРОПОГРУЖЕНИЯ	317
О вибропогружении и его математическом моделировании	
41. ОСНОВНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СОВРЕМЕННОЙ НАУКИ И ТЕОРИЯ КОЛЕБАНИЙ	322
Динамическая система как основная модель точной науки, о ее роли и ее изучении. Об автоколебаниях, регулярных и хаотических движениях и гомоклинических структурах. О теории колебаний как всеобъемлющей науке об эволюционных процессах. О роли А.А. Андропова в становлении нелинейной теории колебаний	
42. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КАК ПЛОДОТВОРНАЯ ИДЕЯ НАУЧНОГО ИССЛЕДОВАНИЯ. D-РАЗБИЕНИЕ	337
На примере устойчивости линейных систем (проблемы Рауса–Гурвица, робастной устойчивости и др.) показывается возможная роль математической модели как плодотворной идеи научного использования	
43. ИДЕАЛИЗАЦИЯ, МАТЕМАТИЧЕСКАЯ КОРРЕКТНОСТЬ И РЕАЛЬНОСТЬ	344
На примерах моделирования центробежного регулятора скорости вращения и известной проблемы парадокса Пенлеве в механике иллюстрируется роль идеализации в математической корректности, в достижении требуемой адекватности результатов моделирования реальному объекту	
44. ДИНАМИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ И ГЛОБАЛЬНОЙ ПОИСКОВОЙ ОПТИМИЗАЦИИ С АДАПТИВНОЙ МОДЕЛЬЮ	355
На примерах широко известного метода наименьших квадратов и глобальной поисковой оптимизации функций многих переменных иллюстрируется полезность широкого использования модели динамической системы	
45. ИГРОВАЯ МОДЕЛЬ ЧЕЛОВЕЧЕСКОГО ОБЩЕСТВА	360
Аргументируется игровая модель человеческого общества. Обнаруживаются основные принципы его функционирования, организации	

и управления. Изучается идеальная игра людей в обществе и выясняется существо и пути преодоления проблемы власти

46. СИНТЕЗ И ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ КВАЗИИНВАРИАНТНОГО ЛИНЕЙНОГО И НЕЛИНЕЙНОГО УПРАВЛЕНИЯ	382
На примере простейшей математической модели линейной минимально фазовой системы квазиинвариантного управления рассматриваются вопросы физической реализуемости, устойчивости, ошибки управления, на основе которых решается задача синтеза квазиинвариантного управления и возникает возможность расширения управления путем синтеза нелинейной системы, объединяющей квазиинвариантную стратегию управления с классической	
47. АБСТРАКТНАЯ И РЕАЛЬНАЯ МАТЕМАТИЧЕСКИЕ ТЕОРИИ ДИНАМИЧЕСКИХ СИСТЕМ	396
Глава содержит краткое описание современного состояния основ теории динамических систем	
СПИСОК ЛИТЕРАТУРЫ	402

Посвящаю

*моей матери Елене Николаевне
и отцу Исааку Густавовичу,
моей жене Валентине Васильевне,
без которых я не жил бы на этом свете;
а также Александру Александровичу Андронову,
открывшему мне путь в науку.
Выражаю благодарность всем своим ученикам,
которым я помог войти в науку и которые сейчас
мне служат опорой и поддержкой.*

*В науке и ее приложениях, как и в жизни,
самое главное – понимание.
Оно всегда просто, но добывается трудно.*

ПРЕДИСЛОВИЕ

Математическое моделирование (ММ) возникло скорее не как наука, а как ремесло, в основе которого лежало искусство реализации сложных вычислений и анализа их результатов с использованием, в основном, известных вычислительных методов. Однако вскоре в значительной мере благодаря появлению компьютеров и сближению с идеями математической физики, теории колебаний, теории управления и др., ММ стало фактически всеобъемлющей наукой, изучающей математические модели вне зависимости от их конкретного смысла. Не будет преувеличением сказать, что вся современная техника (авиация и наземный транспорт, ракеты и компьютеры, атомная энергетика и машиностроение, приборостроение и связь), определившая наш образ жизни, основана на успехах фундаментальной науки и является детищем математического моделирования. Сегодня цели и задачи ММ очень широки и многообразны, но могут быть кратко сформулированы как качественное и количественное изучение всего, что нас окружает, всевозможных объектов природы, техники и общества. При этом под качественным изучением подразумевается достижение понимания существа изучаемого объекта, его свойств, поведения, возможных явлений и определяющих их причин.

Общая схема ММ в значительной мере устоялась, а ее реализация опирается на фундаментальную науку, методы исследования математических моделей (аналитические, качественные и численные) и на современную вычислительную технику. Вместе с тем продумывание этой схемы обнаруживает в ней действия, не поддающиеся формализации. Это прежде всего относится к построению модели и отчасти к ее исследованию. То, что не формализуемо, можно отнести к искусству моделирования.

Требования от ММ точных количественных результатов и полного качественного исследования антагонистичны: первое требует достаточно полного учета многих факторов, усложняющих модель; второе, напротив, тем более реализуемо, чем проще модель. Конечно, всегда следует стремиться к возможно более простым моделям, простым с точки зрения возможностей их исследования, обеспечивающим требуемое понимание, пусть не очень точным и лишь частично и односторонне оценивающим рассматриваемый объект. Насколько мне известно, проблема построения простых моделей особо не ставилась, но простые модели все время спонтанно возникали в физике, химии, биологии, технических науках, экономике и играют в них существенную, если не ведущую роль. Ведь это материальная точка, точечный заряд, диполь, формулы химических соединений и химических реакций, условные и безусловные рефлексы, твердое и упругое тело, линейные и нелинейные осцилляторы, различные волны и простые поля и многое, многое другое.

Настоящая книга возникла в результате чтения автором лекций о математическом моделировании студентам, бакалаврам и магистрам, специализирующимся

по прикладной математике и информатике, и аспирантам в области точных наук Нижегородского государственного университета. Появление этого нового лекционного курса было вызвано тем, что существующая в России университетская система подготовки специалистов по прикладной математике основывалась на сочетании фундаментального математического и прикладного компьютерного образования, что ориентировало студентов на решение точно поставленных математических задач, формируя определенную отчужденность от важнейших этапов и сторон фактического решения прикладных задач, таких как осмысление и глубокое проникновение в существо конкретной задачи, и математическую подготовку, включающую упрощения, принимаемые идеализации и составление математической модели, ее коррекцию и сопоставление результатов исследования с реальным объектом. Помимо этой главной цели была еще и другая, состоящая в том, чтобы выпускники в дальнейшем занимались не только математическими вопросами и на них ориентировались, но и воспринимали и широко использовали математику как универсальный язык современной точной науки, а математическое моделирование – как могучее средство изучения природы, техники и человеческого общества.

В 1994, 1996 и 1997 годах частично эти лекции по математическим моделям естествознания и техники были изданы Нижегородским университетом в виде трех небольших книжек. Затем была написана четвертая часть, сделаны добавления к ранее опубликованным частям, содержащим материал, читаемый магистрантам и аспирантам. Получившийся труд был рекомендован Научно-методическим советом по прикладной математике и информатике университетов России в качестве учебника для студентов, бакалавров и магистров. Пожалуй, следует отметить, что, кроме перечисленных выше стимулов к написанию этой книги, важным оказалось то, что я в процессе чтения лекций осмыслил решающую роль простых математических моделей в общем понимании и познании окружающего мира и получал огромное удовольствие, строя и изучая такие простые модели, как модель автоколебательной ходьбы человека, загадок Каспийского моря, кризисов и исключительности иммунного ответа организма на вторжение инфекции, игры в отгадывание, парадокса Пенлеве и многие другие.

Автор книги – ученик А.А. Андропова, выходец из его всемирно известной школы теории нелинейных колебаний. Это определило выбор рассматриваемых моделей. В книге идет речь об эволюционных моделях, т.е. моделях процессов изменения во времени: о моделях динамики, управления, обучения, вычислительных процессов, распознавания, оптимизации, относящихся к объектам самой разнообразной природы: механической, физической, химической, биологической, технической, экологической, социальной и др. Объединяющей основой всего этого разнообразия является универсальная, всеобъемлющая математическая модель – динамическая система.

Книга может быть интересной и полезной очень широкому кругу читателей: студентам всевозможных естественных, технических и других высших учебных заведений; специалистам, сталкивающимся с задачами и проблемами, где моделирование может быть полезным, и просто людям любопытным, желающим узнать, как математика изучает окружающий мир и позволяет понять и объяснить загадочные явления, предвидеть будущие изменения, постигнуть свойства и пред-

сказать поведение не только реально существующих объектов и процессов, но и воображаемых или изобретаемых.

Моделирование в значительной мере искусство, и в силу этого в нем не последнюю роль играет красота: красота модели, красота ее исследования и, наконец, красота объяснения загадочных и непростых явлений и свойств изучаемого объекта.

Возможно, что построение и исследование некоторых моделей, содержащихся в книге, доставит вам эстетическое удовольствие.

Математические модели можно подразделить на простые, лишь в общих чертах, в самом нам интересном, описывающие реальный объект; математические модели фундаментальных законов природы – уравнения Ньютона механического движения, электродинамические уравнения Фарадея–Максвелла, уравнение Шредингера микропроцессов и т.п.; модели, от которых требуется значительная количественная адекватность, когда например моделируется полет ракеты и ее посадка на Луну. Каждый из перечисленных типов моделей имеет свою специфику. Моделирование последнего типа, как правило, требует высокой вычислительной квалификации, знания различных вычислительных методов и умения ими пользоваться, значительной вычислительной культуры и навыков, включающих навыки использования современной вычислительной техники. Для построения и исследования моделей первых двух типов многое из перечисленного не обязательно. Но далеко не всегда моделирование третьего типа может обойтись без умений и навыков, требуемых для построения и исследования моделей первых двух типов. Более того, фундаментальные модели только и обеспечивают возможность математического моделирования, а простые модели – и только они – позволяют понять функционирование изучаемого объекта и сопровождающие его явления: понять глобально и качественно основные свойства, возможные явления, причинные связи и зависимости поведения от параметров.

Настоящая книга – о моделях первого и отчасти второго типов, что интересно и важно само по себе и что является необходимым для моделирования нестандартных задач и объектов третьего типа. Вместе с тем это не охватывает полностью проблему математического моделирования, к которому, безусловно, относится еще построение математических моделей на основе наблюдаемых опытных и экспериментальных данных путем их математической обработки. Одним из основных методов построения таких моделей является метод наименьших квадратов, представленный в трудах Гаусса и Лежандра около двух столетий назад. Его рассмотрение было осуществлено отдельно в [32].

Первые издания предлагаемой читателю книги вышли в 2003 году на английском языке в издательстве Springer и в 2004 году на русском в издательстве Нижегородского университета. Эта книга на общероссийском конкурсе учебных изданий для высших учебных заведений «Университетская книга – 2006» награждена дипломом в номинации «Лучшее учебное издание по точным наукам». В настоящем втором издании исправлены опечатки, внесены некоторые коррективы в текст и сделаны небольшие дополнения.

Вначале предполагалось пополнить книгу рассказом о приспособленном базисе [26], позволяющем кратко описывать значительную информацию о непростых объектах; рассказать о роли динамических негрубых систем, о преобразова-

телях, усилителях и генераторах стохастичности и явлении синхронизации в эволюции мира и др. В конечном счете я ограничился некоторым расширением главы, привлекающей внимание к важнейшей проблеме настоящего времени: эволюции человеческого общества на земном шаре, и реабилитацией плодотворной идеи о возможности инвариантного управления, высказанной Г.В. Щипановым в 1939 году [49]. В связи с этим книга получила новое название, более точно отражающее ее содержание.

В подготовке рукописи мне существенную помощь оказали сотрудники и друзья: В.П. Савельев, З.С. Баталова, П.С. Ланда, И.С. Емельянова, И.В. Котельников, Л.Г. Теклина и И.С. Гельфер, которым приношу искреннюю благодарность.

ВВЕДЕНИЕ

В этой книге вы познакомитесь с разнообразными математическими моделями механических, электрических, физических, астрономических, химических, биологических, экологических, кибернетических и других систем и процессов. Кроме того, этот курс поможет вам сформировать естественно-научное миропонимание, понимание науки и ваше отношение к специальности «прикладная математика», к изучаемым общим и специальным математическим предметам.

Вы поступили в университет, чтобы получить высшее математическое образование и специальность прикладного математика, что позволит вам иметь интересную работу, приносящую удовлетворение и средства к существованию. Возможно, вы об этом и не думали, а пришли сюда в силу природной любознательности и любви к математике и достижения успехов в ней.

Вам хочется постичь и изучить окружающий мир, к чему зовет вас ваша генетическая наследственность. Почему же вместо этого вас учат абстрактной, трудно понимаемой математике, оторванной от живого мира, – числам, векторам, матрицам, функциям, действиям над ними, дифференцированию, интегрированию и другому?

Я думаю, что вы знаете общий ответ на этот вопрос: потому что математика является основой естественно-научного познания. Но почему так велика роль в познании мира именно абстрактной математики, такой далекой от нашего конкретного, осязаемого, красочного мира, а, например, не астрологии, которая по расположению светил предсказывает будущее?

Каков метод использования математики в познании мира и предсказаний будущего?

Возможно, и на этот последний вопрос вы знаете ответ: мы познаем мир, создавая и изучая его математические модели.

Вот и мы с вами будем строить математические модели и будем их изучать, тем самым постигая тайны мира, в котором живем.

Что такое модель? Слово «модель» пришло из итальянского языка и означает «копия, образец, прототип», изучив который мы познаем оригинал. Значит, математическая модель – это прототип какой-то части мира – объекта, системы, устройства, машины, процесса, прибора, изучая который мы постигаем эту часть мира.

Чем же является математика и какова ее роль в создании этой вспомогательной модели, и почему мы эту математическую модель можем изучить, и почему изучив ее мы изучаем мир, и что, собственно, собой представляет математика? Я думаю, что на эти вопросы вы ответов не знаете.

К ответу на них подойдем издалека. Как-то давно, в позапрошлом веке, на одном из ученых заседаний бурно обсуждался вопрос о роли языков и математики в обучении студентов. Дебаты длились долго. Одни отстаивали роль языков, другие – значение математики. Среди обсуждающих был великий ученый Виллард Гиббс. Тот самый, который утверждал, что целое проще своей части, что изучить скопище миллиардов молекул проще, чем одну молекулу, и не только утверждал, но

и доказал своими замечательными научными открытиями. Он был всеми уважаем и необычайно молчалив. Но тут, к общему удивлению, он попросил слова и сказал: «Математика – тоже язык».

Что же имел в виду Гиббс? Что общего между русским, английским, греческим, французским и другими языками и математикой? А вот что: язык людей описывает их действия, чувства, желания, воспоминания, пожелания, мнения, приказания, а математика – это язык описания природы, язык научного естествознания. Если вы хотите поговорить с французом, вам надо изучить французский язык, с англичанином – английский язык, с природой – математический язык. Только на математическом языке природа открывает нам свои тайны, и если вы хотите их постичь, вы должны изучить математический язык, изучить математику. С природой и техникой – тоже частью природы, только созданной человеком, – нужно разговаривать на математическом языке. То, что книга природы написана на математическом языке, сказал еще Галилей.

Языки человека – это конкретные языки, в них каждое слово имеет конкретный смысл, напротив, математический язык – язык абстрактный. Но оба они – языки, и это различие не так уж существенно. Допустим, вы на заборе увидели надпись: «Глокая югора икает справдо».

Сначала вам показалось, что это бессмыслица, но, вдумавшись, вы поняли, что говорится о некой «югоре», которая «глокая» и которая «икает» и делает это «справдо». А теперь сравните эту фразу со следующими математическими фразами:

$$(a + b)^2 = a^2 + 2ab + b^2,$$

$$0 < \frac{d}{dt} x(t) < \frac{d}{dt} y(t),$$

$$\int u dv = uv - \int v du.$$

Во второй математической фразе говорится, что некоторая величина $x(t)$ возрастает с ростом t медленнее, чем величина $y(t)$. Что такое x и y , при этом остается неизвестным, как и то, чему равны a и b в первой фразе (формуле).

Языковые описания нашей жизни – рассказы, повести, романы – это ее языковые модели. Описание естественных явлений на математическом языке – это их математические модели, т.е. математика – это язык точных наук, а математические модели – это описание объектов и процессов природы или техники на математическом языке.

Как для того чтобы написать хорошее сочинение, нужно хорошо знать то, о чем пишешь, так и для того чтобы составить математическую модель, нужно хорошо понимать то, что хочешь описать, отличать главное от второстепенного, понимать существо и закономерности происходящих процессов, явлений и взаимоотношений, наконец, нужно уметь все это описать на математическом языке.

Написать сочинение не просто, но прочитать его и понять обычно не очень трудно. Хотя простота может быть обманчивой. Составить модель тоже достаточно трудно, но и понять ее, оказывается, не всегда просто, а подчас и очень трудно. Для того чтобы понять математическую модель, понять то, что она описывает, необходимо ее изучить, и это изучение может быть очень трудным.

Теперь, я думаю, вам понятно в общих чертах содержание этой книги.

Но на один вопрос мы не получили ответа: почему изучая математическую модель реального объекта или процесса мы получаем возможность изучения и их самих? Ведь они такие разные и, казалось бы, не имеют ничего общего. Это очень непростой вопрос. Один из известных математиков назвал эту чудесную возможность «непостижимой». Но, пожалуй, это не совсем так: кое-что понять можно.

Все основано на изоморфизме. Об этом – много позднее, а пока заметим только: то, что реальный объект и модель такие разные, не так уж страшно – как работает телевизор, можно понять по его радиосхеме, совершенно не похожей на телевизор, можно разобраться в улицах города по его карте.

Теперь я перейду к рассказу об одной из основных моделей естествознания и техники – динамической системе.

1. ДИНАМИЧЕСКАЯ СИСТЕМА

Архимед, когда постиг закон рычага, опьяненный открывшейся ему тайной природы, сказал: «Дайте мне точку опоры, и я переверну мир».

Много веков спустя, в XVIII веке, великий математик и астроном Лаплас, восхищенный открывшейся ему научной картиной мироздания и причинных связей всех явлений, выразил свое прозрение словами: «Я предскажу во всех деталях все будущее, все, что будет на все века вперед, от мелочей до самого крупного, если вы мне скажете или полностью опишите, в каком состоянии мир находится сейчас».

Это был апофеоз детерминизма, полной причинной связи будущего с настоящим.

Ясно, что никто не мог уличить Лапласа во лжи или ошибке: кто же может описать мироздание полностью? Хорошо бы это сделать хотя бы в какой-нибудь его части – и не всего, а только кое-чего. Ну и что тогда? Уже предсказывать нельзя? Нет, оказывается, иногда можно. Постараемся разобраться, когда можно и когда нельзя, и от чего это зависит, и заодно поймем, почему Лаплас так сказал.

Сказано это было в связи с успехами астрономии, в которой наука добилась полного понимания причин движения планет и научилась их рассчитывать на много лет вперед, и одновременно с воцарением механистической картины мира, в которой все происходящее объяснялось механическим движением мельчайших материальных частиц – молекул. При этом само собой разумелось, что молекулы (атомы), хотя и очень маленькие, но обладают такими же свойствами, как и окружающие нас тела: могут взаимодействовать между собой, ударяться друг о друга, соединяться и т.д. Таковы уж капризы истории, что величайшие прозрения пришли в науку «с небес» – благодаря успехам астрономии. Движение планет – блуждающих звезд – давно привлекало внимание человечества. Сейчас нам до них нет дела, а тогда было совсем не так. Широко известна птолемеяевская система мира. Она достаточно хорошо описывала и с точки зрения этого описания объясняла видимые причудливые движения планет (Меркурия, Венеры, Марса, Луны, Юпитера, Сатурна) и Солнца, еще проще это делала система Коперника, но только Ньютон их объяснил исходя из открытых им законов механического движения и всемирного тяготения. В отношении механического движения материальной точки он установил, что оно подчиняется дифференциальному уравнению второго порядка и однозначно определяется начальным положением и скоростью. Отсюда следовало, что движение всех планет и Солнца может быть точно вычислено, если известны их массы, начальные положения и скорости. Само вычисление Ньютон сумел выполнить лишь для двух притягивающихся свободных тел, но это уже принципиально. Вскоре научились вычислять приближенно, но с громадной точностью, движения всех планет, а сейчас это легко делают ЭВМ. Итак, для планет существует описание, по которому может быть предсказано будущее, т.е. то, как это описание меняется. Такое описание было названо состоянием механической системы. В какой-то мере понятие состояния может быть обобщено и распространено не только на механические системы. Именно по этому пути пошло дальней-

шее развитие точного естествознания в изучении эволюционных процессов и предсказании будущего.

Идея детерминизма, вдохновившая Лапласа, идея временной причинности и однозначной обусловленности событий нашего мира может рассматриваться как наличие связей между описаниями прошлого и настоящего.

При этом бессмысленно говорить о связи всего прошлого со всем настоящим: возможность фактического изучения появляется тогда, когда выбирается что-то как из описания, так и из окружающего мира. Этот выбор предмета изучения Ньютон описал как сосредоточение внимания на отдельных привлекательных красивых камушках на необозримом берегу моря. Тот «камушек», который мы выберем, назовем системой, а то, что привлечет к нему наше внимание, – ее описанием. Это описание можно обозначить буквой x , оно зависит от времени и совершенно не обязательно, что для него будет справедлив принцип детерминизма, т.е. что прошлое описание $x(t_1)$ определяет будущее описание $x(t_2)$ ($t_2 > t_1$). Это только тогда, когда нам крупно повезло как в выборе камушка, так и в выборе того, что нас в нем привлекает.

Теперь я приведу вам некоторые примеры выбора «камушков» и их описаний. Но прежде я хочу сказать еще несколько слов о детерминизме Лапласа. Если можно так выразиться, оградить его от естественных нападок, что он не видел и игнорировал повсеместно наблюдаемую случайность. Нет, он ее видел. Более того, он ввел в науку распределение вероятностей, получившее впоследствии название нормального (или гауссовского) распределения. Но так же как А. Эйнштейн, он не верил, что Бог играет в кости. Он не знал, как объяснить наблюдаемую случайность, но это не помешало ему принять детерминистическую картину мира. И сделал правильно. И случайные процессы допускают детерминистическое описание, но об этом много позднее.

Вернемся к прерванному разговору. Вот у меня в руках кусочек мела. Я могу заинтересоваться, как он пишет на доске и почему он это делает. Тогда меня будут интересовать его свойства и его описание с точки зрения его качества для письма на доске. Но я могу воспользоваться им для того, чтобы бросить или просто отпустить и посмотреть, как он будет падать. В этом последнем случае я могу быть удачлив и установить законы падения тяжелых тел на землю, обнаружив торжество детерминизма.

Я могу обратить внимание на стакан горячего чая и задуматься о том, почему он имеет такой вкус и запах, и что это такое, а могу поинтересоваться, когда же он, наконец, остынет и его можно будет пить. В последнем случае меня интересует только температура чая, описываемая в градусах Цельсия, и если я буду достаточно упорен, то найду закон ее уменьшения, подтверждающий общую идею детерминизма.

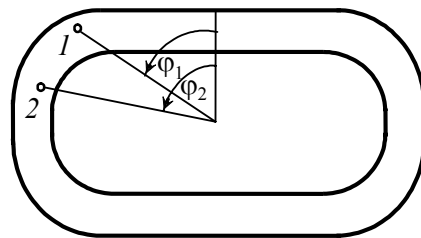
Эти примеры можно продолжать до бесконечности: горящая свеча, привлекающая внимание Фарадея; движение Земли вокруг Солнца, изученное Иоганном Кеплером и Исааком Ньютоном; электрический контур из емкости и индуктивности, тайна колебаний заряда которого была разгадана Томсоном; модель атома Нильса Бора; земная атмосфера, от которой зависит погода, которую мы никак не научимся предсказывать; живой организм, и посейчас полный тайн, и совсем таинственный мозг.

После этих беглых примеров рассмотрим несколько примеров попроще и более подробно.

Понаблюдаем спринтерский заезд двух велосипедистов на велодроме. Они приняли старт в одно и то же время, и выиграет тот, кто придет первым к финишу. Сколько времени они будут ехать, не имеет значения, и это приводит к тому, что ни один из них не рвется вперед, а маневрирует, желая обмануть противника и прийти первым. Просто рвануться без оглядки вперед нельзя, потому что «севший на хвост» соперник поедет за ним, затрачивая значительно меньше сил, и вблизи финиша легко придет первым.

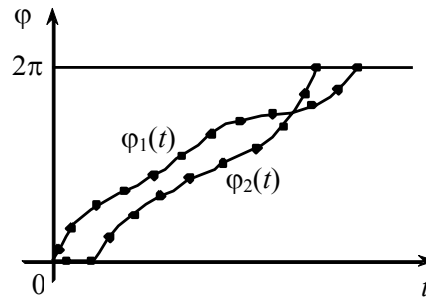
В качестве описания примем углы φ_1 и φ_2 , проходимые первым и вторым велосипедистами (рис. 1.1).

Рис. 1.1. Вид сверху велодрома и спринтерского заезда велосипедистов 1 и 2



Графики функций $\varphi_1(t)$ и $\varphi_2(t)$ дают представление о том, как протекало соревнование (рис. 1.2). При этом мы рисуем две кривые. Можно изобразить ход соревнования одной кривой на плоскости φ_1, φ_2 . (Старт отвечает углу $\varphi = 0$, финиш – $\varphi = 2\pi$.)

Рис. 1.2. Временные графики $\varphi_1(t)$ и $\varphi_2(t)$ одного из заездов велосипедистов



С течением времени точка M с координатами φ_1 и φ_2 опишет некоторую кривую (рис. 1.3).

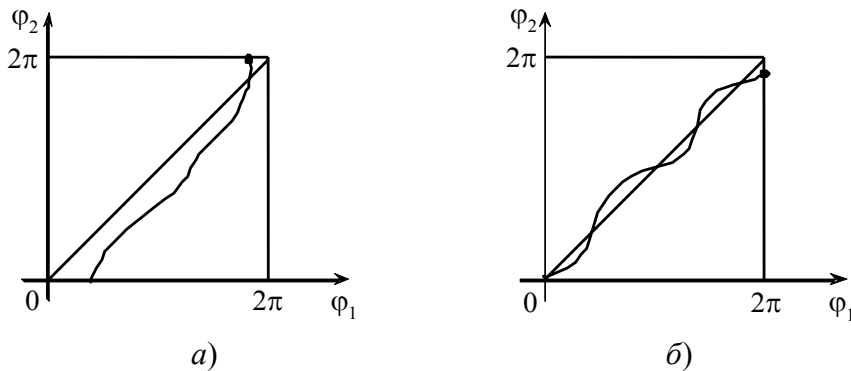


Рис. 1.3. Изображение на плоскости φ_1, φ_2 разных заездов велосипедистов

Ее вид также показывает, как протекало соревнование. Так, согласно рис. 1.3,а, первым уехал со старта первый велосипедист, затем второй и был почти до самого конца позади первого, а перед самым финишем его обогнал и пришел первым. На рис. 1.3,б изображен другой вариант соревнований, где много раз велосипедисты менялись местами и победил первый.

Нетрудно видеть, что принятое описание с помощью углов φ_1 и φ_2 до того, как один из них становится равен 2π , не допускает однозначного прогноза исхода соревнований, хотя, возможно, опытный глаз специалиста мог бы высказать более или менее вероятные предположения.

Рассмотрим еще пример с описанием течения болезни графиком температуры больного (рис. 1.4). Этот график опытному врачу может сказать о многом в отношении течения болезни, но все же его начальной части недостаточно для достоверного прогноза всей кривой или только финала болезни.

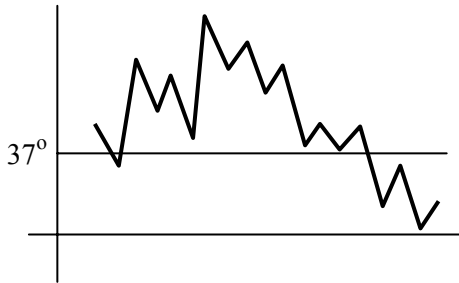


Рис. 1.4. График утренних и вечерних температур больного

Следующий пример – это свободное вертикальное падение тела, находящегося на высоте h_0 от земли и имеющего скорость v_0 (вертикальную).

Согласно закону свободного падения, имеем

$$h = h_0 + v_0 t + gt^2/2,$$

где t – время падения, а h – положение, отсчитываемое по вертикали вниз. На рис. 1.5 изображены допустимые виды графиков зависимости $h(t)$.

Эти графики показывают, как происходит падение тела с заданным h_0 , если оно брошено вверх, просто отпущено или брошено вниз. Так что величина h неплохо описывает падение тела. Но возможно ли с помощью одного h дать полное описание дальнейшего падения тела? Очевидно, нет, нужно знать еще $\dot{h}(t)$. Учитывая это, примем в качестве описания два графика – $h(t)$ и $\dot{h}(t)$ (рис. 1.6). По значениям $h(t)$ и $\dot{h}(t) = v$ в любой момент времени t можно предсказать их значения в любой следующий момент времени $\bar{t} > t$ по формулам:

$$\begin{aligned} h(\bar{t}) &= h(t) + \dot{h}(t)(\bar{t} - t) + g(\bar{t} - t)^2 / 2, \\ \dot{h}(\bar{t}) &= \dot{h}(t) + g(\bar{t} - t), \end{aligned} \tag{1.1}$$

хорошо известным вам из школьной физики.

Вместо двух графиков – $h(t)$ и $\dot{h}(t)$ – можно ограничиться одной кривой, пробегаемой при изменении времени t на плоскости (h, v) точкой M с координатами h и v

$$h = h(t), \quad v = \dot{h}(t).$$

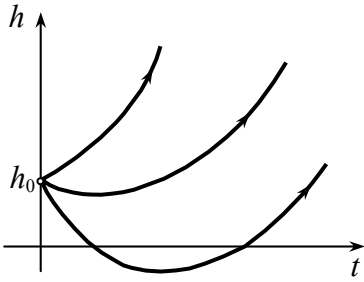


Рис. 1.5. Графики падения тела с одной и той же высоты, но с разными начальными скоростями

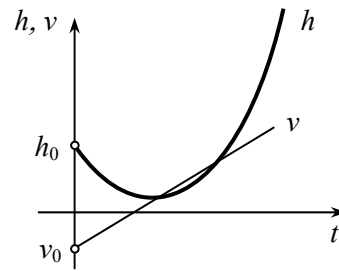


Рис. 1.6. Временные графики изменения высоты h и скорости v брошенного вертикально вверх тела

Уравнения этой кривой можно найти следующим образом.
Из очевидных соотношений

$$v = \dot{h}(t), \quad \dot{v} = g, \quad (1.2)$$

выражающих, что скорость — это производная от пройденного пути, а ускорение свободного падения равно g , следует, что

$$\frac{dh}{dv} = \frac{v}{g}$$

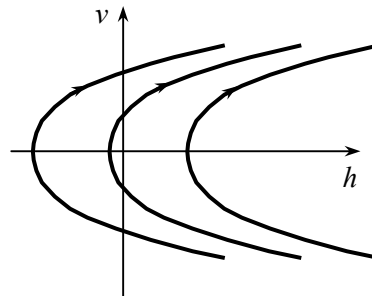
или

$$h = v^2/2g + C,$$

где C — произвольная постоянная.

В силу этого, траектории, описываемые точкой $M(h, v)$ на плоскости h, v при различных C , имеют вид, показанный на рис. 1.7. При возрастании времени точка M пробегает эти кривые в направлениях, показанных стрелками (это следует из того, что при $v > 0$ величина h возрастает, а при $v < 0$ — убывает). Принятое нами описание падения тела двумя величинами h и v обладает замечательным свойством самодостаточности и позволяет осуществлять однозначный прогноз. Действительно, задание любых h и v однозначно определяет единственную параболу, на которой лежит точка $M(h, v)$. Тем самым однозначно определено дальнейшее изменение величин h и v .

Рис. 1.7. Фазовый портрет падения тела



Плоскость переменных h, v имеет особое название — фазовая плоскость, соответственно переменные h и v — фазовые переменные, а нарисованные на этой фазовой плоскости траектории движения фазовых точек $M(h, v)$ — фазовые траектории.

Движение фазовой точки по траекториям замечательно тем, что знание ее положения в любой момент времени t позволяет найти, куда она переместится в любой последующий момент времени $\bar{t} > t$, а именно, согласно (1.1) или (1.2):

$$v(\bar{t}) = v(t) + g(\bar{t} - t),$$

$$h(\bar{t}) = h(t) + v(t)(\bar{t} - t) + \frac{1}{2}g(\bar{t} - t)^2.$$

Это свойство самодостаточности переменных h, v , состоящее в том, что для их нахождения через любой промежуток времени Δt достаточно знать их же в начальный момент, придает им особое значение, благодаря которому такое описание получило название состояния, а описывающие его переменные – фазовых переменных. Это же свойство состояния легло в основу определения математической модели, получившей наименование динамической системы, модели, описывающей детерминированные эволюционные процессы, одной из основных – если не основной – моделей современного научного естествознания и техники.

Еще раз взглянем на описание соревнования велосипедистов на плоскости φ_1, φ_2 и свободного падения тела на плоскости h, v . В первом случае описания различных заездов представляют собой всевозможные кривые в квадрате $0 \leq \varphi_1 \leq 2\pi, 0 \leq \varphi_2 \leq 2\pi$. Естественно только предположить, что во всех этих случаях φ_1 и φ_2 – неубывающие функции времени. Через каждую точку этого квадрата может проходить сколько угодно кривых. Именно поэтому задание значений φ_1 и φ_2 в некоторый момент времени не определяет дальнейшего хода проходящей через нее кривой. Напротив, на плоскости h, v кривые, изображающие различные случаи падения, не пересекаются между собой. Именно в силу того, что они не могут пересекаться и через каждую точку проходит единственная кривая, возможен однозначный прогноз дальнейшего изменения величин h и v .

Таким образом, при описании переменными φ_1, φ_2 невозможен однозначный прогноз, и это находит свое отражение в том, что через каждую точку (φ_1, φ_2) могут проходить несколько кривых, и, напротив, при описании свободного падения переменными h и v имеет место детерминизм, потому что через каждую точку (h, v) проходит единственная кривая. Рассмотренное математическое описание свободного падения тела в поле силы тяжести является примером математической модели, называемой динамической системой, в то время как математическое описание соревнования велосипедистов с помощью углов φ_1 и φ_2 не является математической моделью, называемой динамической системой.

Дадим теперь общее абстрактное определение математической модели, называемой динамической системой, в геометрической трактовке великого математика А. Пуанкаре.

Динамическая система определяется пространством X и заданным на нем однозначным оператором $T(\Delta t)$, зависящим от параметра $\Delta t \geq 0$ так, что каждой точке $x \in X$ оператор T ставит в соответствие единственную точку \bar{x} , т.е

$$\bar{x} = T(\Delta t)x.$$

При этом предполагается, что оператор $T(\Delta t)$ при любых допустимых $\Delta t_1 \geq 0$ и $\Delta t_2 \geq 0$ удовлетворяет соотношению

$$T(\Delta t_2)T(\Delta t_1) = T(\Delta t_1 + \Delta t_2).$$

Смысловое содержание пространства X и оператора $T(\Delta t)$ следующее: X – это пространство всевозможных состояний рассматриваемой системы. (Обычно x – это многомерный вектор с компонентами x_1, x_2, \dots, x_n .) Описание x называется состоянием или фазовой точкой, а пространство состояний X – фазовым пространством. Оператор $T(\Delta t)$ по состоянию x в начальный момент позволяет найти состояние \bar{x} спустя время Δt . При этом ясно, что переход из состояния x в \bar{x} , совершаемый сначала за время $\Delta t_1 \geq 0$, а затем еще за время $\Delta t_2 \geq 0$, должен быть таким же, как переход за время $\Delta t_1 + \Delta t_2$. В этом смысл написанного выше требования к оператору $T(\Delta t)$.

Описание x замечательно тем, что по нему можно найти такое же описание спустя любое время $\Delta t \geq 0$. Именно поэтому оно названо состоянием, а точка x – фазовой.

Рассмотренный выше пример свободного падения тела под действием силы тяжести полностью соответствует приведенному определению. В нем состояние x – это двумерный вектор с компонентами h и v . Пространство состояний или фазовое пространство представляет собой двумерную плоскость переменных h и v . Значения высоты h и скорости v за время Δt изменяются, и, согласно предыдущему, их новые значения равны \bar{h} и \bar{v} , где

$$\bar{h} = h + v\Delta t + \frac{1}{2}g(\Delta t)^2, \quad \bar{v} = v + g\Delta t. \quad (1.3)$$

Эти последние формулы определяют оператор $T(\Delta t)$. Можно непосредственно проверить, что для этого оператора $T(\Delta t)$ выполняется требование, налагаемое на него определением динамической системы.

Из соотношений, определяющих изменение переменных h и v за время Δt , следует, что

$$\frac{\bar{v} - v}{\Delta t} = g, \quad \frac{\bar{h} - h}{\Delta t} = v + \frac{1}{2}g\Delta t.$$

Переходя в них к пределу при $\Delta t \rightarrow 0$, находим, что функции $h(t)$ и $v(t)$ удовлетворяют дифференциальным уравнениям

$$\frac{dh}{dt} = v, \quad \frac{dv}{dt} = g.$$

Этот вывод имеет место и в общем случае, если предположить, что существует предел

$$\lim_{\Delta t \rightarrow 0} \frac{T(\Delta t)}{\Delta t} = L,$$

и тогда вектор состояния x , как функция времени t , удовлетворяет дифференциальному уравнению

$$\frac{dx}{dt} = Lx,$$

которое в компонентах записывается в виде

$$\frac{dx_i}{dt} = L_i(x_1, \dots, x_n), \quad i = \overline{1, n}.$$

Действительно, запишем соотношение

$$\bar{x} = T(\Delta t)x$$

в виде

$$x(t + \Delta t) = T(\Delta t)x(t),$$

далее в виде

$$\frac{x(t + \Delta t) - x(t)}{\Delta t} = \frac{T(\Delta t)x(t) - x(t)}{\Delta t}$$

или, учитывая, что

$$x(t) = T(0)x(t),$$

и переходя к пределу при $\Delta t \rightarrow 0$, в виде

$$\frac{dx(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{T(\Delta t) - T(0)}{\Delta t} x(t) = Lx(t),$$

что и требовалось.

Таким образом, изменение состояния динамической системы удовлетворяет некоторой системе дифференциальных уравнений первого порядка, которые для компонент вектора x могут быть записаны в виде:

$$\begin{aligned} \frac{dx_1}{dt} &= l_1(x_1, x_2, \dots, x_n), \\ &\dots\dots\dots \\ \frac{dx_n}{dt} &= l_n(x_1, x_2, \dots, x_n). \end{aligned}$$

При этом, конечно, предполагается, что рассматриваемые дифференциальные уравнения однозначно разрешимы, и тогда их решение $x_i(t)$ ($i = \overline{1, n}$) в любой момент времени $t \geq t_0$ определяется начальными условиями, т.е. значениями $x_i(t_0)$ ($i = \overline{1, n}$).

Напротив, интегрирование этих уравнений позволяет найти оператор преобразования состояния динамической системы $T(\Delta t)$. Тем самым задание дифференциальных уравнений, которым удовлетворяет состояние как функция времени t , определяет оператор динамической системы. Обратное имеет место лишь в случае дифференцируемости оператора динамической системы, что часто может не выполняться. Поэтому данное выше определение динамической системы несколько более общее, чем то, которое можно получить считая, что изменение состояния динамической системы подчиняется некоторым дифференциальным уравнениям. Заметим еще, что оператор $T(\Delta t)$ может быть определен не при всех $\Delta t \geq 0$, а только для некоторого множества значений $\Delta t \geq 0$.

Центральным наглядным геометрическим образом динамической системы является ее фазовый портрет, изображающий все возможные ее движения, т.е. всевозможные случаи эволюции во времени ее состояния (описания).

Знание фазового портрета дает полное представление о динамике (возможных изменениях) динамической системы, это портрет ее динамики.

С фазовым портретом мы уже сталкивались при изображении падений тела на двумерной плоскости h, v . Каждое отдельное падение на этой плоскости изображалось параболой

$$h = \frac{v^2}{2g} + C.$$

Вся совокупность возможных падений описывается параболой при всевозможных значениях C . Это привело к рис. 1.7. Отдельные кривые этого фазового портрета называются фазовыми траекториями, вся совокупность возможных фазовых траекторий образует фазовый портрет.

Опираясь на этот наглядный пример, нетрудно дать общие определения фазовой траектории и фазового портрета, обнаружить важнейшее свойство фазовых траекторий, состоящее в том, что они не могут раздваиваться, т.е. из каждой фазовой точки не может выходить более одной фазовой траектории и одна обязательно выходит.

Действительно, пусть x – произвольная точка фазового пространства X . За время Δt эта точка перейдет в точку \bar{x} . При изменении Δt от 0 до ∞ точка \bar{x} опишет некоторую «кривую», выходящую из точки x . Это и есть фазовая траектория, выходящая из фазовой точки x . Эта фазовая траектория единственная в силу однозначности оператора $T(\Delta t)$. Совокупность всех фазовых траекторий образует фазовый портрет.

Проиллюстрируем понятие фазового портрета на примерах описания катания на санках или лыжах, рассмотрев два вида профиля: ямку с всевозвышающимися краями и ямку на горизонтальной поверхности, изображенных на рис. 1.8. Будем считать, что сани симметричные и могут скользить без трения как вперед, так и назад.

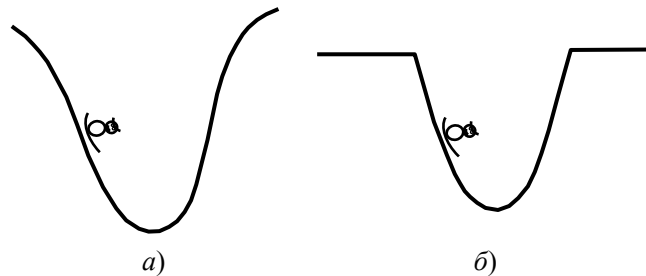


Рис. 1.8. Схема езды по снегу при разных профилях местности

Описание нужно выбрать так, чтобы оно было состоянием. Исходя из опыта примера со свободным падением тела, естественно выбрать в качестве описания положение s и скорость v саней. Переменную s выберем вдоль горизонтали, начало отсчета – от наиболее глубокого места ямки. За скорость v примем скорость движения саней вдоль профиля местности. Не будем писать никаких формул и уравнений. Попробуем сообразить без них, какой вид имеет фазовый портрет, т.е. фазовые траектории на плоскости s, v . В случае ямки с всевозвышающимися склонами всякое движение саней будет представлять собою либо неподвижное

положение внизу ямки, либо периодическое катание со склона на склон. Соответствующий фазовый портрет изображен на рис. 1.9,а. Он состоит из овальных, замкнутых, вложенных друг в друга фазовых траекторий, внутри которых находится фазовая траектория, изображаемая одной точкой $s=0, v=0$, отвечающая равновесному положению саней на дне ямки. В точках A, B, C и D фазовой траектории z сани последовательно находятся, соответственно, в крайнем левом верхнем положении, на дне ямки, в крайнем правом верхнем положении, опять на дне ямки и, наконец, снова в крайнем левом верхнем положении. Такое движение неограниченно периодически повторяется.

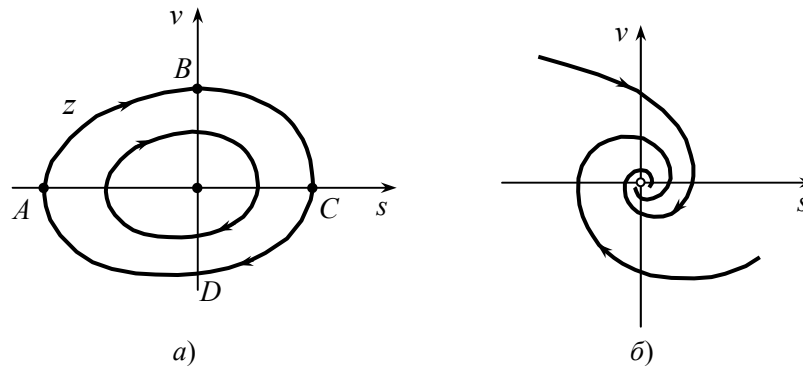


Рис. 1.9. Фазовый портрет движения саней: а) по идеально гладкому снегу; б) с трением при профиле местности рис. 1.8,а

Второй случай фазового портрета более разнообразен, поскольку вдоль горизонтальных участков сани движутся с постоянной скоростью и каждая из точек горизонтальных участков при $v=0$ является состоянием равновесия. Это приводит к фазовому портрету рис. 1.10,а.

Давайте сообразим, как изменятся фазовые портреты рис. 1.9,а и рис. 1.10,а, если учесть реально всегда существующее трение. Теперь колебания внутри ямки всегда затухают и переходят в равновесие на дне ямки. Движение по горизонтальной части также затухает. Если это учесть, то фазовые портреты рис. 1.9,а и рис. 1.10,а преобразуются к портретам, изображенным на рис. 1.9,б и рис. 1.10,б.

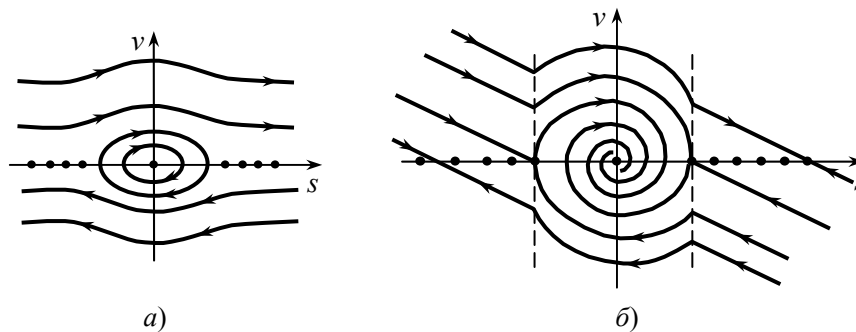


Рис.1.10. Фазовый портрет движения саней: а) по идеально гладкому снегу; б) с трением при профиле местности рис. 1.8,б

Теперь, я думаю, вы готовы к тому, чтобы найти уже достаточно сложный фазовый портрет плоского маятника. Мы опять не будем прибегать к уравнениям и формулам, а попробуем непосредственно сообразить, какой он имеет вид.

В свое время, позднее, мы напишем формулы и уточним его количественно, а сейчас ограничимся качественным видом. Пожалуй, самое трудное – это понять, что представляет собой фазовое пространство маятника. Если угол отклонения маятника от вертикали φ , то состоянием будет этот угол φ и угловая скорость вращения маятника $\omega = \dot{\varphi}$. Фазовое пространство – это пространство, точками которого являются всевозможные точки $M(\varphi, \omega)$. Угол φ меняется в пределах от $-\pi$ до π , причем значениям $\varphi = -\pi$ и $\varphi = \pi$ отвечает одно и то же положение маятника. Следовательно, фазовым пространством будет бесконечная полоса $-\pi \leq \varphi \leq \pi$, $-\infty < \omega < \infty$, где точки $\varphi = -\pi$, ω и $\varphi = \pi$, ω изображают одно и то же состояние маятника, поэтому им должна отвечать в фазовом пространстве только одна фазовая точка, а не две.

Можно просто условиться, что точки $(-\pi, \omega)$ и (π, ω) – это одна и та же точка. А можно реально осуществить их слияние, склеив в цилиндр полосу по линиям $\varphi = -\pi$ и $\varphi = \pi$, но, конечно, так, чтобы склеились друг с другом точки с одинаковыми ω . В результате приходим к тому, что фазовое пространство маятника – это двумерный цилиндр. Он изображен на рис. 1.11. Разрезав его вдоль линии $\varphi = \pm \pi$, приходим к полосе рис. 1.12 с отождествленными сторонами. Рисовать фазовые траектории на цилиндре затруднительно, и поэтому будем это делать на полосе, имея все время в виду, что это разрезанный и развернутый на плоскость цилиндр.

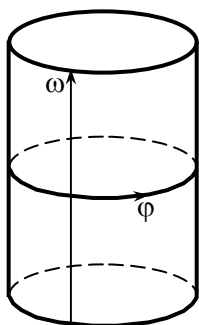


Рис. 1.11. Фазовый цилиндр

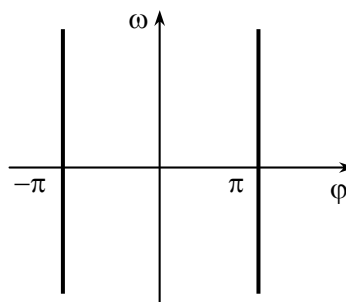


Рис. 1.12. Развертка фазового цилиндра на плоскости

У маятника есть два равновесия: одно нижнее устойчивое, отвечающее точке $\varphi = 0$, $\omega = 0$, и другое верхнее неустойчивое, отвечающее $\varphi = \pm \pi$, $\omega = 0$. Каждая из этих точек – целая фазовая траектория, ее начало и конец и вся она в одной точке.

Если неподвижно висящий маятник, изображаемый равновесием $\varphi = 0$, $\omega = 0$, немного толкнуть, то он начнет качаться с маленькой амплитудой возле этого состояния равновесия. Такое движение изобразится небольшим овалом, охватывающим точку нижнего равновесия. С усилением толчка маятник будет качаться все с большим и большим размахом (амплитудой), и наконец, его качания перейдут во вращательное движение либо по часовой стрелке, либо против часовой стрелки. Сказанное изображено на рис. 1.13,а. На этом фазовом портрете k_0 – нижнее устойчивое равновесие, k_∞ – верхнее неустойчивое, k_1, k_2 – периодические колебательные движения около нижнего равновесия; B_1, B_2 – все более и более быстрые вращательные движения против часовой стрелки; \bar{B}_1, \bar{B}_2 – убыстряю-

щиеся вращения по часовой стрелке. Угол φ отклонения маятника отсчитывается от нижнего его положения против часовой стрелки; P_1 и \bar{P}_1 – движения, приближающиеся к точке k_∞ при $t \rightarrow \infty$ и $t \rightarrow -\infty$. Эти последние фазовые траектории P_1 и \bar{P}_1 разделяют колебательные и вращательные движения.

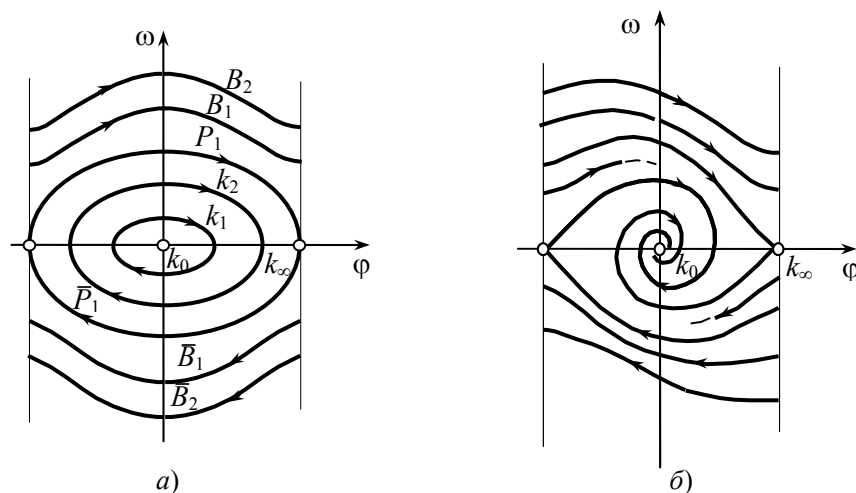


Рис. 1.13. Фазовый портрет маятника: а) в отсутствие трения; б) при наличии трения

Вы можете считать себя на этом первоначальном этапе усвоившими представления о фазовом портрете, если сумеете сами, не заглядывая в рис. 1.13,б, нарисовать, как изменится портрет рис. 1.13,а с появлением затухания, например из-за сопротивления воздуха, и объяснить, какие движения изображаются его отдельными фазовыми траекториями.

Все математические модели, которые далее будут возникать и изучаться, являются конкретизациями и частными случаями описанной выше общей математической модели динамической системы. Но сколь же они будут не похожи друг на друга, сколь разнообразны как по характеру протекающих в них процессов, так и по их физической природе! Всякий процесс, явление или система любой природы, описываемая дифференциальными уравнениями, – это динамическая система. В силу этого великие законы природы: законы механического движения твердых тел, жидкостей и упругих сред, изменения электромагнитного поля и законы электродинамики и квантовой физики – описываются математическими моделями, представляющими собою динамические системы. Но дальнейшее изложение начнется не с них. Мы будем рассматривать и изучать очень простые модели, и лишь затем они будут несколько усложняться, оставаясь все же конкретными и очень простыми. Несмотря на их простоту, а скорее именно благодаря их простоте, они наиболее подходят для обучения и в наибольшей мере способствуют приобретению столь важных общих и интуитивных представлений, подкрепленных конкретными, допускающими исчерпывающее осмысление примерами. Иногда это будут очень простые математические модели простых физических систем и иногда – простые модели очень сложных систем. Но даже в очень сложных случаях модели будут оставаться простыми, обнаруживая свое значение и познавательную мощь.

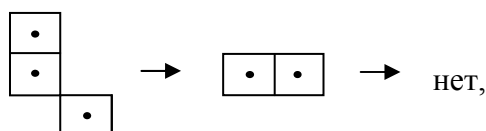
Как отмечалось, всякая система или процесс, описываемые дифференциальными уравнениями, всегда представляют собой динамическую систему. Это следует из теоремы существования и единственности решения дифференциальных уравнений при заданных начальных условиях. Ее состояние является совокупностью начальных условий, однозначно определяющих решение, а оператор определяется ее решением на промежутке времени Δt .

Вместе с тем мне не хотелось бы, чтобы вы подумали, что данное вам общее определение динамической системы – просто другая трактовка дифференциальных уравнений. Чтобы этого не произошло, я приведу пример, где очевидно, что от дифференциальных уравнений далеко и они ни при чем. Этот пример – игра «жизнь» по Конуэлю.

В этой игре, разыгрываемой на шахматной доске, состояние определяется расположением на ней фишек. Пространство состояний – это множество всевозможных расположений фишек. Состояние можно задать матрицей 8×8 , элементы которой – единицы и нули, в зависимости от того, есть фишка или нет на соответствующей клетке. Фазовое пространство состоит из всевозможных таких матриц. С каждым тактом $\Delta t = 1$ расположение фишек меняется в соответствии с оператором, определяемым тремя правилами выживания, гибели и рождения:

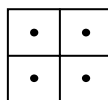
- 1) фишка остается, если рядом с ней находятся две или три другие фишки;
- 2) фишка убирается, если рядом с ней находится более трех или менее двух фишек;
- 3) на пустой клетке появляется новая фишка, если рядом с ней имелось три фишки.

В зависимости от начального расположения фишек возможны весьма многообразные их изменения. Так, например, три фишки, расположенные следующим образом



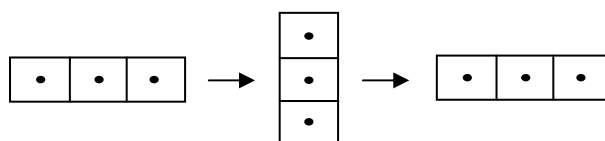
превращаются, как показано, сначала в две и затем исчезают.

Расположение четырех фишек в виде квадрата



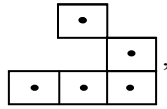
не влечет никаких изменений.

Три фишки в ряд



осциллируют (как показано) с периодом два такта.

Пять фишек, образующие фигуру «планер»



повторяются каждые четыре такта, смещаясь вправо и вниз на одну клетку.

Существуют расположения фишек, которые, осциллируя каждый период, рожают «планер».

Есть расположения фишек, пожирающие «планеры», и т.д. Это разнообразие возможностей, отдаленно и упрощенно имитирующее реальную жизнь, порождаемую комбинированием химических молекул, как уже говорилось, представляется динамической системой, состояние которой x – это матрица из нулей и единиц. В ней столько же строк и столбцов, сколько на доске, на которой разыгрывается эта игра. Элемент a_{ij} этой матрицы равен 1 или 0, в зависимости от того, есть фишка или нет на пересечении строки i и столбца j доски. Множество всевозможных таких матриц с единичными и нулевыми элементами составляет ее пространство состояний. Оператор $T(\Delta t)$ этой динамической системы определен только для дискретных значений Δt , равных $0, 1, 2, \dots$

При этом, если n целое, то

$$T(n) = T^n(1),$$

а $T(0)$ – оператор, никак не меняющий расположение фишек.

В заключение знакомства с математической моделью динамической системы и ее фазовым портретом найдем фазовый портрет физического маятника на вращающемся основании (рис. 1.14).

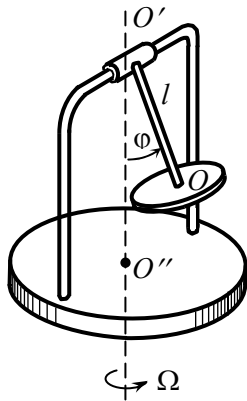


Рис. 1.14. Физический маятник на вращающемся основании

Фазовый портрет маятника при неподвижном основании нашли, не выписывая его уравнений движения: мы достаточно хорошо знаем маятник и нам не понадобились дифференциальные уравнения его движения. С маятником на вращающемся основании так поступить нельзя: его динамика для нас загадка и разгадать ее можно исследуя его математическую модель, что мы и сделаем. Для этого нам потребуются некоторые сведения из механики. Примите их на веру.

В качестве состояния маятника на вращающемся вокруг вертикальной оси с угловой скоростью Ω основании, как и маятника на неподвижном основании,

примем его угол отклонения от вертикали φ и его угловую скорость $\omega = \dot{\varphi}$. Нам нужно составить его дифференциальные уравнения движения. В этом может помочь теоретическая механика. Ее рецепт состоит в следующем: надо найти кинетическую T и потенциальную V энергии маятника и составить функцию Лагранжа $L = T - V$, выразив ее через переменные состояния φ и ω , после чего уравнения движения записываются в таком красивом виде:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \omega} \right) - \frac{\partial L}{\partial \varphi} = 0.$$

После сейчас непонятных вам вычислений находится функция L в виде

$$L = \frac{A}{2}(\omega^2 + \Omega^2 \sin^2 \varphi) + \frac{C}{2}\Omega^2 \cos^2 \varphi + Mgl \cos \varphi,$$

где A и C – моменты инерции маятника, l и M – его длина и масса.

После дифференцирования приходим к искомому дифференциальному уравнению второго порядка:

$$A\ddot{\varphi} + \frac{C-A}{2}\Omega^2 \sin 2\varphi + Mgl \sin \varphi = 0. \quad (1.4)$$

Примите на веру то, как составлено это уравнение. Нужно лишь его понимать, в частности видеть, что при $\Omega = 0$ оно переходит в уравнения движения обычного физического маятника на неподвижном основании. Еще следует представлять физический смысл моментов инерции A и C . Маятник имеет вид тела вращения относительно своей линии подвеса (линии, соединяющей точку подвеса с центром масс). Его центр масс находится на расстоянии l от оси подвеса, сам маятник имеет массу M , C – момент инерции относительно линии подвеса, A – момент инерции относительно осей, перпендикулярных линии подвеса и проходящих через точку подвеса. Для тела, вытянутого вдоль оси подвеса, $C < A$. Наоборот, для маятника в виде диска достаточно большого радиуса по сравнению с его длиной $C > A$.

Решить уравнение (1.4) в элементарных функциях нельзя. Но можно его один раз проинтегрировать. Для этого умножим его на $\dot{\varphi}$ и сделаем следующие вычисления:

$$\begin{aligned} & A\dot{\varphi}\ddot{\varphi} + \frac{C-A}{2}\Omega^2 \dot{\varphi} \sin 2\varphi + Mgl \dot{\varphi} \sin \varphi = \\ & = \frac{d}{dt} \left(\frac{A}{2} \dot{\varphi}^2 - \frac{C-A}{4} \Omega^2 \cos 2\varphi - Mgl \cos \varphi \right) = 0, \end{aligned}$$

откуда

$$\frac{A}{2} \dot{\varphi}^2 - \frac{C-A}{4} \Omega^2 \cos 2\varphi - Mgl \cos \varphi = h,$$

где h – постоянная интегрирования.

Из полученного интеграла находим, что

$$\sqrt{\frac{A}{2}} \dot{\varphi} = \pm \sqrt{h + \frac{C-A}{4} \Omega^2 \cos 2\varphi + Mgl \cos \varphi}. \quad (1.5)$$

Это и есть уравнение фазовых траекторий на цилиндрическом фазовом пространстве переменных состояния φ и $\dot{\varphi}$. Эти фазовые траектории надо построить, каждому h – своя фазовая траектория. Но вам сделать это не очень просто. Поэтому, сначала, исходя из этого уравнения (1.5), построим уже известный нам фазовый портрет маятника на неподвижном основании ($\Omega = 0$). В этом более простом случае

$$\sqrt{\frac{A}{2}} \dot{\varphi} = \pm \sqrt{h + Mgl \cos \varphi}. \quad (1.6)$$

Рассматривая подкоренное выражение как функцию φ , построим график постоянной h и график функции $-Mgl \cos \varphi$ (рис. 1.15,а). Первый график – прямая, параллельная оси φ на расстоянии h от нее, второй график – один период косинусоиды с амплитудой Mgl . Стрелками на этом рис. 1.15,а показаны значения подкоренной функции при разных φ . Если длину этой стрелки с учетом ее направления обозначить через ρ , то уравнение фазовой траектории (1.6) запишется в очень простом виде:

$$\dot{\varphi} = \pm \sqrt{2\rho/A}. \quad (1.7)$$

Для приведенного на рис. 1.15,а значения h возможные значения φ заключены между φ_1 и $-\varphi_1$ (вне их $\rho < 0$ и $\dot{\varphi}$ мнимое). Меняя φ от φ_1 до $-\varphi_1$, легко находим из рис. 1.15,а вид соответствующей фазовой траектории. Это будет овал, нарисованный ниже используемых нами графиков на плоскости $\varphi, \dot{\varphi}$ (точнее, полосу $-\pi \leq \varphi \leq \pi, -\infty < \dot{\varphi} < \infty$). Меняя h от $-Mgl$ до ∞ , найдем все фазовые траектории. Этот фазовый портрет нам уже известен и еще раз изображен на рис. 1.15,а.

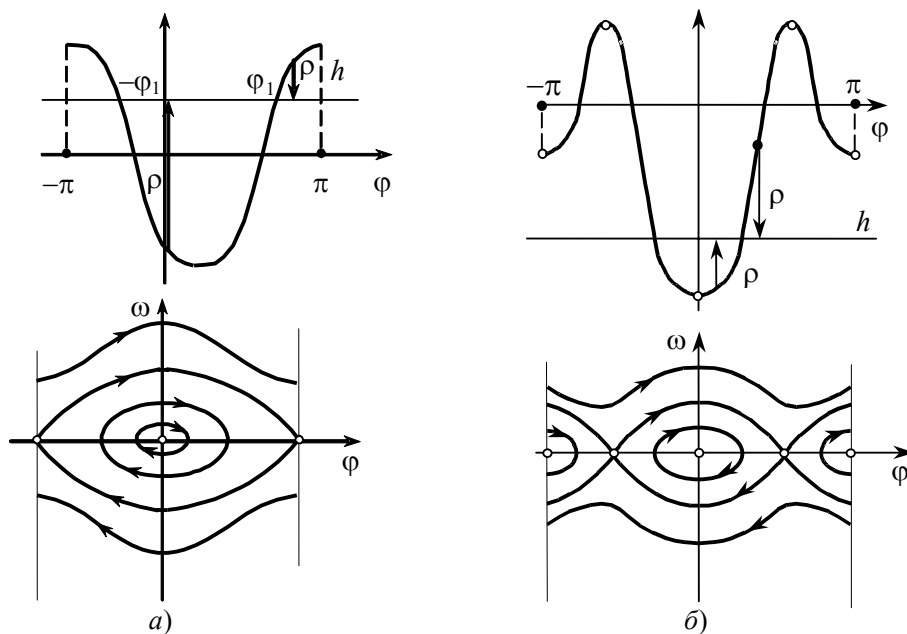


Рис. 1.15. Построение фазового портрета маятника на вращающемся основании: а) при $\Omega = 0$; б) при достаточно большом значении Ω

Колебания маятника отвечают $-Mgl < h < Mgl$, нижнему, устойчивому равновесию $-h = -Mgl$, верхнему, неустойчивому, $-h = Mgl$, вращениям $-h > Mgl$.

Теперь аналогичным образом построим значительно более сложный фазовый портрет по уравнению (1.5) при $\Omega \neq 0$.

При $(C - A)\Omega^2 > 4Mgl$ рисунок 1.15,а изменится и примет вид рис. 1.15,б. На нем построен график зависимости

$$F(\varphi) = -Mgl \cos \varphi - \frac{C - A}{4} \Omega^2 \cos 2\varphi$$

и проведена прямая, отвечающая значению h .

Как и ранее, стрелками показаны значения $\rho(\varphi)$, и по-прежнему фазовая траектория описывается уравнением (1.7). Меняя h , при некотором усердии находим фазовый портрет, который изображен на рис. 1.15,б. Этот фазовый портрет наделяет маятник совершенно удивительными свойствами: он может устойчиво висеть как вниз, так и вверх. Это поразительная способность появилась в силу сделанного предположения, что

$$(C - A)\Omega^2 > 4Mgl,$$

т.е. при $C > A$ и при достаточно большой скорости вращения Ω основания маятника. Интересно, что, как можно обнаружить аналогичным образом, при $C < A$ увеличение скорости вращения Ω сохраняет неустойчивость верхнего положения и делает неустойчивым и нижнее положение равновесия. При $C < A$ и $\Omega^2 > Mgl(A - C)^{-1}$ соответствующий фазовый портрет изображен на рис. 1.16.

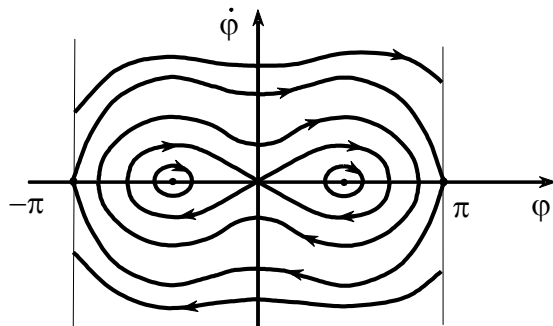


Рис. 1.16. Фазовый портрет маятника на вращающемся основании при достаточно большом значении Ω

Почему нижнее положение равновесия становится неустойчивым с ростом угловой скорости Ω , понять можно: устойчивость нарушается центробежной силой. Почему же это не имеет место при $C > A$ и почему при этом и верхнее равновесие становится устойчивым? А дело в том, что маятник не математический, его масса не сосредоточена в одной точке и суммарная центробежная сила при отклонении маятника от нижнего и верхнего равновесий не обязательно направлена от оси вращения, она при $C > A$ может быть направлена к оси вращения. Чтобы это обнаружить, следует всмотреться в уравнение (1.4), которое можно трактовать как уравнение движения маятника во вращающейся вместе с маятником системе координат под действием силы тяжести и центробежных сил. При этом на маятник действует суммарный момент сил, равный:

$$-\frac{C - A}{2} \Omega^2 \sin 2\varphi - Mgl \sin \varphi,$$

который при φ , близком к нулю, возвращает к равновесию при любом Ω и возвращает к равновесию $\varphi = \pi$ при достаточно большом Ω . Точная математическая модель поправляет нашу примитивную интуицию, порождая новую, более совершенную.

Рассмотрим еще один пример, где интуиция нас подводит. Пусть вам задают вопрос о том, как лучшей вооруженностью войск компенсировать их меньшую численность. Естественный ответ: если армия в два раза малочисленнее, то она должна быть в два раза лучше вооружена, подразумевая при этом, что она должна быть в два раза более эффективной, т.е. что каждый солдат должен в два раза быстрее выводить из строя солдат противника, чем это делают более многочисленные армии.

А теперь посмотрим, что дает соответствующая математическая модель. Пусть $x(t)$ и $y(t)$ – численности армий в момент времени t . При $t = 0$, когда началось сражение, они были $x(0)$ и $y(0)$. Пусть a и b – скорости, с которыми солдаты противостоящих армий выведут из строя солдат противника. Тогда

$$\begin{aligned}\dot{x} &= -by, \\ \dot{y} &= -ax\end{aligned}$$

или $axdx - bydy = 0$, $ax^2 - by^2 = ax^2(0) - by^2(0) = \text{const}$.

Согласно полученному интегралу, дающему уравнения фазовых траекторий, фазовый портрет имеет вид, представленный на рис.1.17.

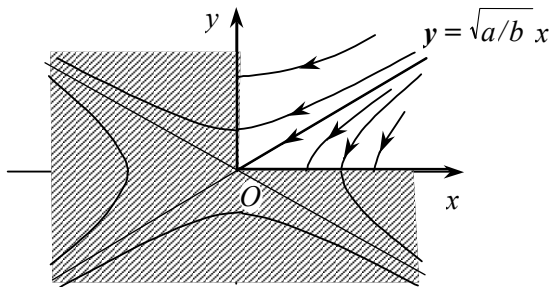


Рис. 1.17. Фазовый портрет динамики численностей сражающихся армий

Все фазовые траектории кончаются либо на оси OX , либо на OY , либо в точке O . Если фазовая траектория кончается на оси OX , то побеждает армия численности x , если на оси OY , то армия численности y .

В первом случае для начальной точки x_0, y_0 должно выполняться неравенство:

$$y_0 < \sqrt{\frac{a}{b}} x_0,$$

во втором случае – обратное неравенство. Таким образом, для победы армии численности x_0 над армией численности y_0 должно иметь место неравенство:

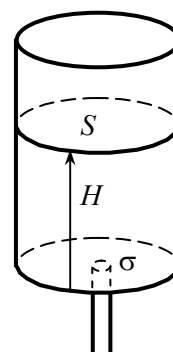
$$\frac{a}{b} > \frac{y_0^2}{x_0^2},$$

т.е. при численности y_0 , в два раза большей x_0 , необходимо, чтобы эффективность армии x была в четыре раза большей. Это следует из математической модели, а в чем ошибка интуиции, попробуйте сообразить сами.

2. ВЫТЕКАНИЕ ЖИДКОСТИ ИЗ СОСУДА

Рассмотрим очень простое явление вытекания воды из цилиндрического сосуда с маленькой дырочкой в его дне (рис. 2.1).

Рис. 2.1. Цилиндр с отверстием в дне, через которое выливается вода



Пусть площадь сечения сосуда S , площадь дырочки σ , а высота уровня воды в сосуде H . Как будет меняться высота уровня жидкости H , если она вытекает и в начальный момент вытекания воды ее уровень $H = H_0$? Ясно, что для того чтобы ответить на этот вопрос, достаточно знать скорость v , с которой вода вытекает через дырочку. Действительно, за время dt через дырочку вытекает объем воды, равный $\sigma v dt$, и поэтому скорость опускания уровня воды в сосуде равна

$$\dot{H} = -\frac{\sigma}{S} v.$$

Если известно v , то полученное соотношение является дифференциальным уравнением, из которого можно найти H как функцию времени t , т.е. $H(t)$. Первым догадался, чему равно v , очень давно, три века назад, Торричелли. Он сказал: «Вода будет вытекать с той же скоростью, с какой она упала бы с высоты H ». Ведь сверху вода исчезает, а снизу вытекает, и при этом она как бы падает с высоты H . Тело, падающее с высоты H , приобретает скорость

$$v = \sqrt{2gH}.$$

Это и есть знаменитая в свое время формула Торричелли. Нельзя не признать остроумия рассуждений Торричелли. Но ведь сверху опускается одна вода, а через дырочку вытекает другая. Не все здесь понятно, но формула верная – это показал эксперимент. Если поверить этой формуле, то приходим к дифференциальному уравнению

$$\dot{H} = -\frac{\sigma}{S} \sqrt{2gH}, \quad (2.1)$$

из которого следует, что

$$\frac{dH}{\sqrt{H}} = -\frac{\sigma}{S} \sqrt{2g} dt,$$

и поэтому

$$2\sqrt{H} = -\frac{\sigma}{S}\sqrt{2gt} + C,$$

где C – пока неизвестная постоянная.

Из начальных условий

$$H|_{t=0} = H_0$$

находим, что

$$C = 2\sqrt{H_0},$$

и окончательно

$$H = \left(\sqrt{H_0} - \frac{\sigma}{S} \sqrt{\frac{g}{2}} t \right)^2. \quad (2.2)$$

Теперь можно нарисовать график убывания H (рис. 2.2) и найти время полного вытекания воды

$$t_{\text{выт}} = \frac{S}{\sigma} \sqrt{\frac{2H_0}{g}}. \quad (2.3)$$

Из графика видно, что сначала вода вытекает быстрее, затем скорость ее вытекания уменьшается до нуля и это происходит за конечное время $t_{\text{выт}}$, пропорциональное корню квадратному из начальной высоты воды H_0 .

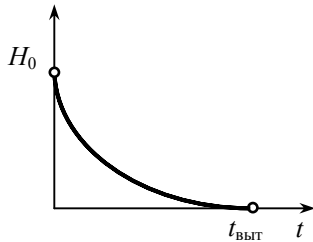


Рис. 2.2. Временной график изменения уровня воды при ее вытекании из цилиндра

Насколько эти выводы совпадают с экспериментом? Формула Торричелли практикой подтверждается: действительно, вода во всех случаях выливается с той же скоростью, как если бы она свободно падала с верхнего уровня воды. А вот время вытекания $t_{\text{выт}}$, получаемое по формуле (2.3), оказывается примерно в два раза меньшим. Соответственно, в два раза меньше расчетной и скорость опускания уровня воды H в сосуде. Откуда взялось такое громадное расхождение, если скорость вытекания воды v из дырочки определена правильно? Это возможно только за счет неправильности формулы

$$\dot{H} = -\frac{\sigma}{S} v,$$

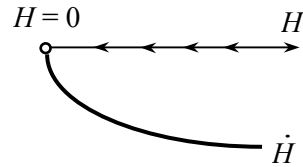
означающей, что сколько воды вытекло, на столько ее и убыло.

Отчего же она может быть неправильной? Сформулируем ее точный смысл: отношение скоростей v/\dot{H} равно отношению сечения цилиндра к сечению вытекающей струи. Мы приняли, не задумываясь, что сечение струи совпадает с размером дырки, и вот это совсем не так. Сечение струи на самом деле меньше сечения дыры, и вызвано это тем, что притекающая к дыре с разных сторон жидкость вызывает ее сжатие, благодаря чему сечение струи оказывается меньше сечения

отверстия. Сжатия струи можно избежать, снабдив отверстие постепенно сужающейся воронкой. Сложный гидродинамический расчет, подтвержденный экспериментом, дает, что сужение струи, вытекающей через отверстие в горизонтальном дне, равно примерно 2.

Построенная нами математическая модель вытекания жидкости является динамической системой. Ее фазовым пространством является полупрямая $H \geq 0$. Ее единственная фазовая траектория представляет эту полупрямую, пробегаемую от $H = \infty$ до $H = 0$ (рис. 2.3).

Рис. 2.3. Построение одномерного фазового портрета вытекания воды



Сосуд, из которого вытекает вода, в древние времена использовался для отсчета времени. Это так называемые водяные часы, клепсидры. Если выбрать S , σ и H_0 так, чтобы $t_{\text{выт}}$ равнялось суткам, то, налив сосуд водой, можно определять время в течение последующих суток по уровню воды H . Для того чтобы указать уровни, которым отвечают разные часы суток, нужно разделить отрезок оси t графика рис. 2.2 на 24 равные части и, согласно графику этого рисунка, найти соответствующие H . Нетрудно видеть, что получающаяся при этом шкала времени, выражаемая уровнями воды, будет неравномерной: сначала часу будет отвечать сильное опускание воды, а дальше все меньшее и меньшее. Попробуем сообразить, как нужно изменить водяные часы, чтобы их шкала была более удобной, равномерной. Еще раз взглянем на формулу (2.1): шкала будет равномерной, если равным dH вне зависимости от H будут отвечать равные dt , т.е. нужно сделать так, чтобы

$$\frac{\sigma}{S} \sqrt{2gH} = \text{const}.$$

На первый взгляд, это невозможно, но если подумать, то можно догадаться, что для этого достаточно цилиндрический сосуд, для которого S постоянно и не зависит от H , заменить на сосуд, у которого

$$S = a\sqrt{H}.$$

Что за форма у такого сосуда? Пусть он имеет вид тела вращения, тогда его горизонтальное сечение на высоте H – круг радиуса r и площади $S = \pi r^2$. Далее следует, что

$$\pi r^2 = a\sqrt{H}$$

или

$$H = \left(\frac{\pi}{a}\right)^2 r^4,$$

что соответствует форме сосуда, изображенной на рис. 2.4, форме, похожей на котелок.

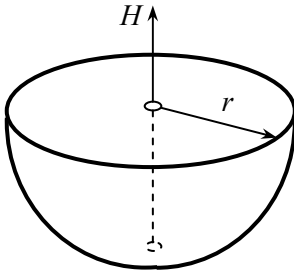


Рис. 2.4. Форма сосуда, из которого вода вытекает с постоянной скоростью убывания ее уровня

Итак, задача решена, закон вытекания найден и может быть использован для построения водяных часов с удобной равномерной шкалой, правда, сейчас это уже никому не нужно.

В основе решения задачи о вытекании лежит закон Торричелли, вывод которого несколько туманен. Попробуем его прояснить. Для этого прибегнем к тому, чего Торричелли не знал, к закону сохранения энергии. При вытекании воды потенциальная энергия тонкого слоя воды массы dm в связи с ее опусканием переходит в кинетическую энергию вытекающей воды, так что

$$dm g H = dm \frac{v^2}{2}$$

или

$$v = \sqrt{2gH},$$

что и есть формула Торричелли. Так что все обосновано.

Но давайте испытаем на прочность полученный выше результат (2.2)

$$H(t) = \left(\frac{\sigma}{S} \sqrt{\frac{g}{2}} t - \sqrt{H_0} \right)^2$$

тем, что посмотрим, что он дает при $\sigma = S$, когда вода не вытекает, а просто падает из бездонного цилиндра. Раз она свободно падает, то, очевидно,

$$H = H_0 - \frac{gt^2}{2},$$

что никак не согласуется с тем, что дает принятый нами закон Торричелли. На рис. 2.5 для сравнения приведены вместе оба графика изменения $H(t)$. Первый из них отвечает закону Торричелли при $\sigma = S$, второй – свободному падению воды. Они не совпадают, хотя вода выливается из сосуда в обоих случаях за одно и то же время $t_{\text{выт}}$.

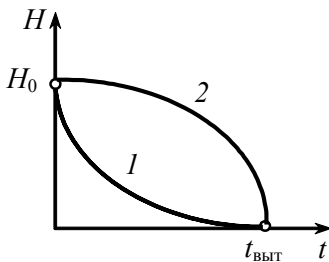


Рис. 2.5. Временные графики 1 и 2 «вытекания» воды из цилиндра без дна, следующие из уравнения (2.1) и, соответственно, закона свободного падения

В чем же дело? Где ошибка? Неверен закон Торричелли? В законе падения твердого тела усомниться трудно. Посмотрим, в чем разница между графиками 1 и 2 рис. 2.5. Согласно графику 1, вода сначала вытекает быстро, а затем скорость ее вытекания замедляется, а согласно кривой 2, наоборот, сначала медленно, а затем все быстрее и быстрее. При дырке во весь сосуд, пожалуй, вернее второе. Где же ошибка? Что, неверен закон сохранения энергии? Едва ли. Он неверно применен? Это скорее похоже на правду. Но где?

Ведь все так просто и ясно. Вернемся к основной посылке: энергия сохраняется (трением жидкости можно пренебречь) и поэтому потенциальная энергия верхнего слоя переходит в кинетическую энергию вытекающей жидкости. Но ведь это так, причем приближенно, только при $\sigma \ll S$. Действительно, ведь и верхний слой жидкости опускается и имеет некоторую кинетическую энергию, поэтому правильное написать

$$dm gH + dm \frac{\dot{H}^2}{2} = dm \frac{v^2}{2}$$

и соответственно, учитывая, что $\sigma v = -S\dot{H}$,

$$\dot{H} = -\frac{\sqrt{2gH}}{\sqrt{(S/\sigma)^2 - 1}} = -\frac{\sigma}{S} \sqrt{\frac{2gH}{1 - (\sigma/S)^2}}. \quad (2.4)$$

Эта новая формула при $\sigma \ll S$ переходит в формулу Торричелли. Так что требуемое уточнение найдено. Но не будем спешить: примем опять $\sigma = S$. Еще хуже: теперь $\dot{H} = \infty$. Опять ошибка. Вернемся снова к исходному рассуждению, сразу имея в виду $\sigma = S$. При падении жидкости ее потенциальная энергия, равная действительно $dm gH$, переходит в кинетическую, но не только жидкости, «вылезающей» из сосуда, но и находящейся в нем, точнее, в прибавку ее кинетической энергии, поскольку движение носит ускоренный характер. Так что ошибка в том, что нужно еще учитывать прибавку в кинетической энергии жидкости, находящейся в сосуде, которая действительно мала при $\sigma \ll S$. Но при σ , сравнимом с S , это совсем не так. Еще раз вернемся к исходному рассмотрению и сделаем необходимые уточнения.

Итак, пусть V – потенциальная энергия жидкости в сосуде, T – ее кинетическая энергия, а T_1 – кинетическая энергия вытекшей из сосуда за время dt жидкости, тогда, согласно закону сохранения энергии,

$$\frac{d}{dt}(T + V)dt + T_1 = 0.$$

Выражение для T_1 нам уже известно. Это

$$T_1 = dm \frac{v^2}{2} = -\rho S \dot{H} dt \frac{v^2}{2},$$

где ρ – объемная плотность жидкости.

Потенциальная энергия может быть сосчитана

$$V = \int gH dm = \int_0^H \rho S H g dH = \frac{\rho g S H^2}{2},$$

кинетическая энергия равна

$$T = \int \frac{u^2}{2} dm,$$

где интеграл берется по всему объему жидкости в сосуде, а u – скорость течения элемента массы dm этого объема жидкости. В разных точках сосуда скорость u своя, найти ее – это значит найти, как жидкость вытекает из сосуда, а это очень сложно. Как же быть? Как быть, ясно в двух крайних случаях: $\sigma \ll S$, и тогда можно пренебречь величиной порядка u^2 , и $\sigma = S$, когда $u = \dot{H}$. Последнее приближение применимо и в случае, когда $H \gg r$, где r – радиус цилиндрического сосуда с водой. Примем его, и тогда

$$T = \rho SH \frac{\dot{H}^2}{2}.$$

После этого закон сохранения энергии, согласно предыдущему, примет вид:

$$\frac{d}{dt} \left(\rho SH \frac{\dot{H}^2}{2} + \frac{\rho g SH^2}{2} \right) - \rho S \dot{H} \left(\frac{S}{\sigma} \right)^2 \frac{\dot{H}^2}{2} = 0$$

или

$$\rho S \frac{\dot{H}^3}{2} + \rho S H \dot{H} \ddot{H} - \rho S \left(\frac{S}{\sigma} \right)^2 \frac{\dot{H}^3}{2} + \rho g S H \dot{H} = 0,$$

откуда

$$\ddot{H} = -\frac{1}{H} \left\{ gH - \left[\left(\frac{S}{\sigma} \right)^2 - 1 \right] \frac{\dot{H}^2}{2} \right\}. \quad (2.5)$$

Как и можно было ожидать, теперь полученное дифференциальное уравнение уже второго порядка и успешно проходит проверку при $\sigma = S$, поскольку из нее следует, что, как требуется,

$$\ddot{H} = -g.$$

Остается понять, как из него при $\sigma \ll S$ следует ранее полученное уравнение, основанное на законе Торричелли или его уточнении. Обнаружить это не очень просто. Для облегчения рассуждения введем обозначение $u = -\dot{H}$, запишем исследуемое уравнение в виде

$$\dot{u} = \frac{1}{H} \left\{ gH - \left[\left(\frac{S}{\sigma} \right)^2 - 1 \right] \frac{u^2}{2} \right\} \equiv f(u) \quad (2.6)$$

и посмотрим, как меняется \dot{u} . При $f(u) < 0$ скорость u убывает, а при $f(u) > 0$, напротив, возрастает, т.е. существует такое значение

$$u^* = \sqrt{2gH \left[\left(\frac{S}{\sigma} \right)^2 - 1 \right]^{-1}},$$

что при $u < u^*$ u возрастает, а при $u > u^*$ u убывает. В результате этого u оказывается близким к u^* . Для наглядности изобразим полуось значений $u \geq 0$ и нарисуем график $f(u)$ как функции u . В соответствии с этим графиком u меняется, как показано на рис. 2.6. При этом, в силу того, что $S/\sigma \gg 1$, приближение u к u^* происходит быстро. Заметим, что при приближении u к u^* точка u^* несколько смещается, но это смещение точки u^* происходит значительно медленнее, чем приближение u к u^* , поскольку

$$\dot{u}^* = \frac{d}{dt} \sqrt{\frac{2gH}{(S/\sigma)^2 - 1}} = -\frac{1}{2} \frac{\sigma u}{S} \sqrt{\frac{2g}{H[1 - (\sigma/S)^2]}}. \quad (2.7)$$

В силу этого, приближенно и тем точнее, чем меньше σ/S , имеет место

$$u = \frac{\sigma}{S} \sqrt{2gH}, \quad (2.8)$$

что полностью совпадает с тем, что написано ранее (формулы (2.1) и (2.4)).

Какие же выводы следуют из того, что уравнение

$$\dot{H} = -\frac{\sigma}{S} \sqrt{2gH}$$

заменилось уравнением

$$\ddot{H} = \frac{1}{H} \left\{ -gH + \left[\left(\frac{S}{\sigma} \right)^2 - 1 \right] \frac{\dot{H}^2}{2} \right\} ?$$

Прежде всего то, что теперь в начальный момент $t = 0$, $\dot{H} = 0$. Затем происходит быстрое приближение \dot{H} к величине \dot{H} , определяемой первым уравнением, и далее процесс изменения \dot{H} происходит в соответствии с первым. Сказанное изображено на рис. 2.7, где сплошной линией изображен график изменения \dot{H} согласно первому уравнению, а пунктиром – второму. Часть графика, отвечающая промежутку времени от $t = 0$ до $t_{\text{раз}}$, соответствует разгону вытекающей струи жидкости, а остальная часть от $t_{\text{раз}}$ до $t_{\text{выт}}$ – ее длительному вытеканию с постепенным замедлением.

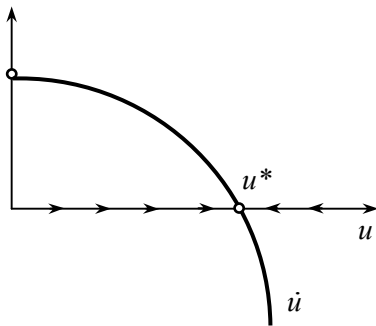


Рис. 2.6. Фазовый портрет дифференциального уравнения (2.6) при $H = \text{const}$

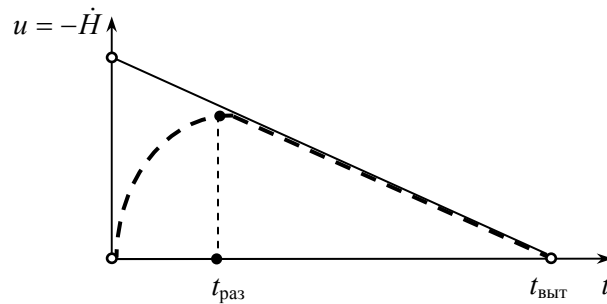


Рис. 2.7. Временные графики падения уровня воды после открытия дырочки: сплошная линия – без учета фазы разгона, пунктирная – с учетом ее

Для того чтобы правильно представить этот график изменения скорости, следует определить время разгона $t_{\text{раз}}$, а время вытекания уже определено

$$t_{\text{выт}} = \frac{S}{\sigma} \sqrt{\frac{2H_0}{g}}.$$

После разгона величина скорости вытекания жидкости достигает величины $\sqrt{2gH}$, причем скорость разгона \dot{i} , во всяком случае в начале, как следует из (2.6), равна

$$\dot{i} = g.$$

Поэтому приблизительно по порядку величины время разгона равно

$$t_{\text{раз}} = \frac{\sigma}{S} \sqrt{\frac{2H_0}{g}}.$$

Отсюда следует, что примерно

$$\frac{t_{\text{раз}}}{t_{\text{выт}}} = \left(\frac{\sigma}{S}\right)^2.$$

В частности, если отношение радиуса цилиндра сосуда к радиусу отверстия в дне равно 100, то время разгона меньше времени вытекания примерно в 10^8 раз, т.е. если время вытекания – около часа, то время разгона – десятитысячные доли секунды. Даже при относительно большей дыре, когда отношение радиусов ≈ 10 , отношение времен – порядка 10^4 . Таким образом, при $\sigma/S \ll 1$ разгон происходит практически мгновенно, а далее вытекание согласуется с законом Торричелли.

На этом краткий рассказ о том, как вытекает вода из сосуда через дырочку в его дне, можно было бы и закончить. Но я хочу еще вам показать, как можно использовать понятие фазового портрета для исследования движений уточненной модели (2.5).

Запишем ее дифференциальные уравнения в переменных H и v уровня жидкости и скорости вытекающей струи. Совокупность этих переменных образует состояние, и их изменения, согласно предыдущему, удовлетворяют дифференциальным уравнениям:

$$\dot{H} = -\frac{\sigma}{S} v, \quad \dot{v} = \frac{S}{\sigma} \left\{ g - \frac{1}{H} \left[1 - \left(\frac{\sigma}{S} \right)^2 \right] \frac{v^2}{2} \right\}. \quad (2.9)$$

В пределе при $\sigma/S \rightarrow 0$ дифференциальные уравнения принимают вид:

$$\dot{H} = 0, \quad \dot{v} = \begin{cases} \infty & \text{при } v^2 < 2gH, \\ 0 & \text{при } v^2 = 2gH, \\ -\infty & \text{при } v^2 > 2gH, \end{cases}$$

согласно которому фазовый портрет состоит из вертикальных фазовых траекторий, по которым фазовые точки бесконечно быстро перемещаются на кривую состояний равновесия $v^2 = 2gH$ (рис. 2.8).

При $\sigma/S \ll 1$, т.е. очень маленьком, мгновенные движения перейдут в быстрые, замедляющиеся по мере приближения к линии $v^2 = 2gH$, и одновременно каждая из фазовых точек, включая и бывшие состояния равновесия, будет смещаться влево с очень малой скоростью. В результате этого фазовый портрет изменится и примет вид, показанный на рис. 2.9. На нем всякое движение фазовой точки после быстрого изменения переходит в медленное движение вблизи и вдоль фазовой кривой $v^2 = 2gH$, где выполняется закон Торричелли.

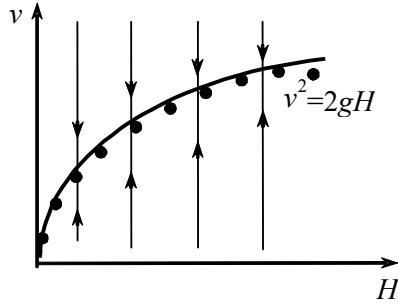


Рис. 2.8. Фазовый портрет предельного дифференциального уравнения (2.9).
Линия с черными точками – состояния равновесия

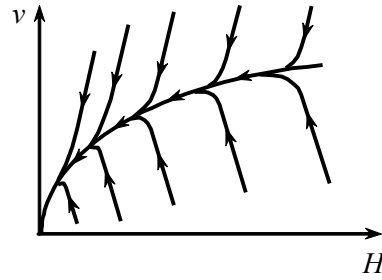


Рис. 2.9. Фазовый портрет дифференциального уравнения (2.9) при $\sigma/S \ll 1$

В частности, если начальной является фазовая точка $H = H_0, v = 0$, то сначала она очень быстро приблизится к кривой $v^2 = 2gH$ и затем будет двигаться вдоль нее налево, оставаясь вблизи нее. Соответствующий этому движению фазовой точки график изменения скорости v (точнее $-\dot{H} = (\sigma/S)v$) уже был представлен на рис. 2.7. Там же мы очень грубо оценили время приближения фазовой точки из начального положения $H = H_0, v = 0$ к малой окрестности кривой $v^2 = 2gH$. Сейчас мы можем уточнить эту оценку. Действительно, пусть нас интересует время приближения рассматриваемой начальной фазовой точки к ε -окрестности кривой $v^2 = 2gH$, определяемой неравенствами

$$-2gH\varepsilon < v^2 - 2gH < 2gH\varepsilon,$$

где $\varepsilon > 0$ мало. Из второго дифференциального уравнения (2.9) следует, что

$$\mu dv = \left[g - \frac{v^2(1-\mu^2)}{2H} \right] dt,$$

где введено обозначение $\mu = \sigma/S$. Пренебрегая малым изменением величины H , которая в начальный момент равна H_0 , и величиной μ^2 , запишем его в виде

$$\frac{\mu dv}{g(1 - v^2/(2gH_0))} = dt.$$

Интегрируя левую часть от значения $v = 0$ до значения $v = \sqrt{2gH(1-\varepsilon)}$, отвечающего достижению границы ε -окрестности кривой $v^2 = 2gH$, а правую часть от 0 до τ , находим, что искомое время движения τ фазовой точки приближенно равно

$$\tau = \frac{\mu}{g} \int_0^{\sqrt{2gH(1-\varepsilon)}} \frac{dv}{1-v^2/(2gH_0)} = \frac{1}{2} \frac{\sigma}{S} \sqrt{\frac{2H_0}{g}} \ln \frac{1+\sqrt{1-\varepsilon}}{1-\sqrt{1-\varepsilon}}.$$

Это по порядку величины близко к ранее грубо оцененному значению τ , равному $\sigma/S \sqrt{2H_0/g}$.

Таким образом, при небольшой дырочке ($\sigma/S \ll 1$) вытекание жидкости из сосуда четко разбивается на две фазы: быстрая фаза разгона до скорости, близкой к $\sqrt{2gH_0}$, и фаза длительного вытекания за конечное время с убывающей до нуля скоростью v . Вместе с тем из полученных оценок следует, что время разгона с уменьшением сечения дырочки σ неограниченно уменьшается. Это вызывает сомнения, так как соответствует неограниченному росту ускорения. Этот парадокс, по-видимому, – следствие допущенных приближений в подсчете кинетической энергии жидкости в сосуде, когда принято, что скорость жидкости во всех ее точках равна \dot{H} . Ясно, что вблизи дырочки это значительно не так, и тем более не так, чем больше S/σ . Вместе с тем, при достаточно малой дырочке в силу капиллярного поверхностного натяжения жидкость может не начать течь, если радиус отверстия меньше $2\mu/\rho g H_0$, где μ – коэффициент поверхностного натяжения. Для воды и $H_0 \approx 50$ см это отверстие много менее 1 мм. В силу этого же эффекта может прекратиться вытекание с уменьшением уровня H . Вообще, при маленькой дырочке возможно не только прекращение вытекания, но и возникновение прерывистого вытекания. Однако оставим эти явления в стороне и попытаемся уточнить вычисление кинетической энергии жидкости в сосуде на основе гидродинамической модели вытекания идеальной жидкости. Это уточнение вычисления наиболее существенно вблизи дырочки, где скорость много больше \dot{H} .

Пусть \dot{H} достаточно мало, так что поле скоростей жидкости внутри сосуда вблизи дырочки меняется медленно и его можно считать квазистационарным.

Для возможности сравнения полученного выражения для кинетической энергии T запишем его в виде (рис. 2.10):

$$T = \int_S d\sigma \int_0^H \frac{\rho u^2}{2} \frac{d\bar{\sigma}}{d\sigma} dx = \frac{1}{2} \rho v^2 \sigma K = \frac{1}{2} \rho K \sigma \left(\frac{S}{\sigma} \right)^2 \dot{H}^2 = \left(\rho S H \frac{\dot{H}^2}{2} \right) \frac{KS}{\sigma H}.$$

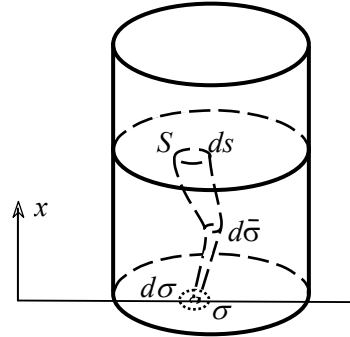
Выделенный скобками множитель – это прежнее выражение кинетической энергии. Отличие нового и прежнего – в множителе $KS(\sigma H)^{-1}$, где K – среднее значение интеграла по трубкам тока,

$$K = \frac{1}{\sigma} \int_S d\sigma \int_0^H \frac{d\sigma}{d\bar{\sigma}} dx.$$

Величина $d\sigma/d\bar{\sigma}$ убывает с ростом x от 1 до $d\sigma/dS$, где dS – значение $d\bar{\sigma}$ на поверхности жидкости. Поэтому K – величина, меньшая H , причем тем более меньшая, чем быстрее убывает вдоль трубки тока с ростом x $d\sigma/d\bar{\sigma}$. Итак, множитель отличия первоначальной и уточненной величин кинетической энергии меньше S/σ и его величина растет с уменьшением σ . Отсюда видно, что совершенная нами ошибка в первоначальном подсчете кинетической энергии может

быть значительна, но уменьшается с ростом H и убыстрением убывания $d\sigma/d\bar{\sigma}$ с ростом x .

Рис. 2.10. Трубка тока, по которой осуществляется подсчет кинетической энергии T



Найдем теперь, как изменится дифференциальное уравнение (2.5). Имеем

$$\frac{d}{dt} \left(\rho S^2 K \sigma^{-1} \frac{\dot{H}^2}{2} + \frac{\rho g S H^2}{2} \right) - \rho S \dot{H} \left(\frac{S}{\sigma} \right)^2 \frac{\dot{H}^2}{2} = 0,$$

откуда

$$\ddot{H} = -\frac{\sigma}{SK} \left\{ gH - \left(\frac{S}{\sigma} \right)^2 \frac{\dot{H}^2}{2} \right\}, \tag{2.10}$$

как и раньше, приходим к оценке времени разгона вида

$$t_{\text{раз}} \approx K \sqrt{\frac{2H_0}{g}}, \tag{2.11}$$

согласно которой нет стремления $t_{\text{раз}}$ к нулю вместе с уменьшением σ . Заметим, что $\sqrt{2H_0/g}$ – это время свободного падения с высоты H_0 , а $K < 1$. Время падения с высоты 50 см – это примерно 0,3 с, т.е. разгон остается быстрым.

3. РАВНОВЕСИЕ И АВТОКОЛЕБАНИЯ УРОВНЯ ЖИДКОСТИ В СОСУДЕ ПРИ ОДНОВРЕМЕННОМ ПРИТОКЕ И ОТТОКЕ

Пусть теперь одновременно с оттоком воды из цилиндрического сосуда с площадью сечения S через его дырочку в дне с эффективным сечением σ вода притекает снаружи с интенсивностью Q . В этом случае уравнение баланса жидкости в сосуде запишется в виде:

$$S\dot{H} = -\sigma v + Q, \quad (3.1)$$

где v – скорость вытекания воды через отверстие с сечением струи σ , а H , как и ранее, – высота уровня жидкости в сосуде, при этом жидкость считается несжимаемой. Если еще пренебречь ее вязкостью, то при $\sigma/S \ll 1$ и $v = \sqrt{2gH}$ дифференциальное уравнение (3.1) можно записать в виде:

$$\dot{H} = -\frac{\sigma}{S}\sqrt{2gH} + \frac{Q}{S}. \quad (3.2)$$

Оно легко интегрируется, но прибегнем к рассмотрению его фазового портрета, нарисовав график скорости \dot{H} как функции H (рис. 3.1). Слева от точки H^*

$$H^* = Q^2 / 2\sigma^2 g, \quad (3.3)$$

в которой $\dot{H} = 0$, фазовые точки по фазовой полупрямой $H \geq 0$ движутся вправо, H возрастает; справа от точки H^* , наоборот, H убывает. Это означает, что точка $H = H^*$ – устойчивое состояние равновесия. К нему приходят с течением времени все точки фазовой полупрямой $H \geq 0$. Из формулы (3.3) следует, что высота равновесного уровня прямо пропорциональна квадрату интенсивности Q притока воды и обратно пропорциональна квадрату сечения вытекающей струи.

Вот, собственно, и вся нехитрая динамика этой системы: она всегда приходит к устойчивому динамическому равновесию $H = H^*$ притока и оттока жидкости.

Допустим теперь, что отток жидкости происходит не через отверстие в дне сосуда, а через так называемый сифон, трубку сечения σ , изогнутую, как показано на рис. 3.2.

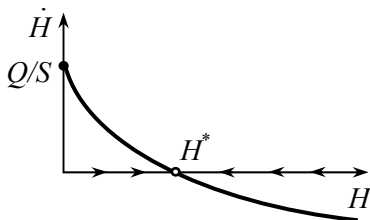


Рис. 3.1. Фазовый портрет изменения уровня воды H в цилиндре при притоке воды Q извне и оттоке ее через дырочку в дне

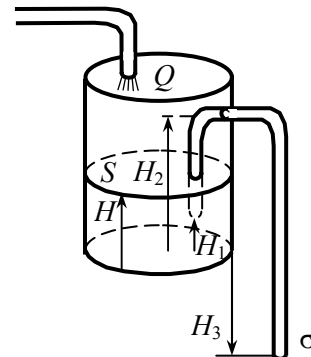


Рис. 3.2. Цилиндр с притоком воды Q и оттоком через сифон

Трубка не доходит до дна на расстояние, равное H_1 , вверху она поворачивает на высоте H_2 от дна и кончается ниже дна на расстоянии H_3 (рис. 3.2). Сифон – это замечательное устройство, позволяющее выливать воду из неполной бочки через ее край, т.е. при $H < H_2$. Но это возможно лишь тогда, когда сифон наполнен водой, если же он пуст, то при уровне воды $H < H_2$ через него вода не течет. Можно поэтому думать, что наша система – бочка с сифоном и притоком – будет иметь своим состоянием описание, достаточное для прогноза: высоту уровня воды H и значение переменной ξ , равное 1, если сифон заполнен водой, и равное 0, если он пуст. Итак, в качестве описания, претендующего на состояние, примем H и ξ . По-видимому, следует различать случаи:

$$H < H_1, \quad H_1 < H < H_2 \quad \text{и} \quad H > H_2.$$

При $H < H_1$ $\xi = 0$ и $S\dot{H} = Q$. При $H > H_2$ сифон наполняется (если не был наполнен), поэтому $\xi = 1$ и

$$S\dot{H} = Q - \sigma\sqrt{2g(H + H_3)}.$$

В промежуточном случае $H_1 < H < H_2$ возможны значения $\xi = 0$ и $\xi = 1$. При $\xi = 0$

$$S\dot{H} = Q.$$

При $\xi = 1$

$$S\dot{H} = Q - \sigma\sqrt{2g(H + H_3)}.$$

В случае $H_1 < H < H_2$ $\xi = 1$, если непосредственно перед этим $\xi = 1$ или $H > H_2$, и $\xi = 0$, если перед этим $\xi = 0$ или $H < H_1$.

Описанную зависимость ξ от значения и изменения H можно представить диаграммой рис. 3.3, на которой указаны условия смены значения ξ с 0 на 1 и обратно: ξ меняется с 0 на 1, если H , возрастая, обращается в H_2 , и ξ испытывает переход от 1 к 0, если H , убывая, проходит значение H_1 . Сделанное описание является математической моделью сифона. При этом ξ не является функцией H , хотя бы и двузначной. Это функционал от предшествующего изменения $H(\tau)$ с $\tau \leq t$, описываемый диаграммой рис. 3.3.

Уравнения изменения H , согласно предыдущему, записываются в виде

$$\dot{H} = \begin{cases} Q/S & \text{при } H < H_1, \\ Q/S & \text{при } H_1 \leq H \leq H_2 \text{ и } \xi = 0, \\ Q/S - \sigma/S\sqrt{2g(H + H_3)} & \text{при } H_1 \leq H \leq H_2 \text{ и } \xi = 1, \\ Q/S - \sigma/S\sqrt{2g(H + H_3)} & \text{при } H > H_2. \end{cases} \quad (3.4)$$

Фазовое пространство рассматриваемой системы состоит из двух частей: отрезка $0 \leq H \leq H_2$ и полупрямой $H_1 \leq H < \infty$, с переходами с одного участка на другой, показанными на диаграмме рис. 3.3.

Для того чтобы на этом фазовом пространстве построить фазовый портрет, нужно показать стрелкой направление движения каждой из его точек в соответствии с дифференциальными уравнениями (3.4). Знак \dot{H} непосредственно указывается уравнениями (3.4). При $H < H_2$ и $\xi = 0$ $\dot{H} > 0$. На части фазового пространства, отвечающей $\xi = 1$, знак \dot{H} зависит от местоположения точки $H=H^*$, в которой

$$Q - \sigma\sqrt{2g(H + H_3)} = 0.$$

Возможны три разных случая: а) $H^* < H_1$, б) $H_1 < H^* < H_2$ и в) $H^* > H_2$. Они представлены на рис. 3.4.

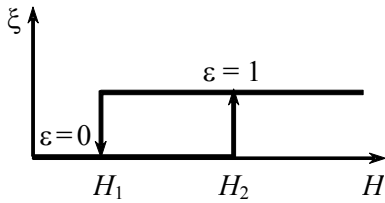


Рис. 3.3. Диаграмма смен состояния ξ сифона при изменении уровня воды H

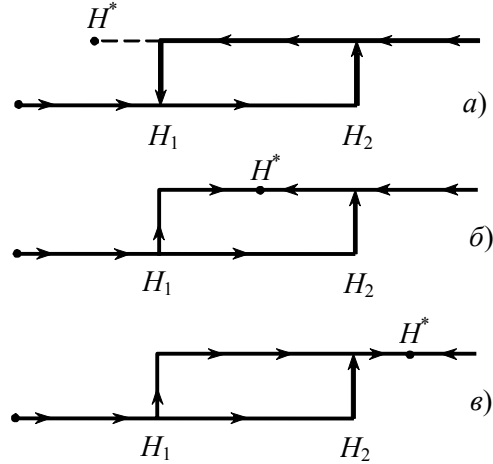


Рис. 3.4. Фазовое пространство уравнения (3.4)

Для того чтобы на этом фазовом пространстве построить фазовый портрет, нужно показать стрелкой направление движения каждой из его точек в соответствии с дифференциальными уравнениями (3.4). Знак \dot{H} непосредственно указывается уравнениями (3.4). При $H < H_2$ и $\xi = 0$ $\dot{H} > 0$. На части фазового пространства, отвечающей $\xi = 1$, знак \dot{H} зависит от местоположения точки $H = H^*$, в которой

$$Q - \sigma\sqrt{2g(H + H_3)} = 0.$$

Возможны три разных случая: а) $H^* < H_1$, б) $H_1 < H^* < H_2$ и в) $H^* > H_2$. Они представлены на рис. 3.4.

Первый случай (а) отвечает возникновению периодических колебаний уровня воды от H_1 к H_2 и обратно. Второй и третий – устойчивым равновесным уровням H^* между H_1 и H_2 и выше H_2 .

Пусть все параметры изучаемой системы фиксированы, а H_3 может меняться. То, какой из случаев будет иметь место, зависит от значения корня H^* уравнения

$$Q - \sigma\sqrt{2g(H + H_3)} = 0,$$

равного

$$H^* = \frac{Q^2}{2g\sigma^2} - H_3.$$

Согласно рис. 3.4, при $H^* > H_1$ имеет место устойчивое равновесие уровня жидкости в сосуде, при $H^* < H_1$ – устойчивые периодические его колебания, называемые автоколебаниями.

Скачкообразное изменение динамики уровня жидкости происходит при значении H_3 , определенном из условия

$$\frac{Q^2}{2g\sigma^2} - H_3 = H_1$$

или при

$$H_3 = H_3^* = \frac{Q^2}{2g\sigma^2} - H_1,$$

т.е. при $H_3 < H_3^*$ имеет место устойчивое равновесие, а при $H_3 > H_3^*$ будут наблюдаться автоколебания.

4. ПЕРЕХОДНЫЙ ПРОЦЕСС, СОСТОЯНИЕ РАВНОВЕСИЯ И АВТОКОЛЕБАНИЯ

В только что рассмотренных математических моделях мы познакомились с тремя типами движений динамических систем: переходный процесс, состояние равновесия и автоколебания.

Переходный процесс имел место при выливании воды из сосуда: система из одного состояния, отвечающего начальному уровню воды, переходит к другому, при котором сосуд пуст. Переходные процессы имели место и при установлении в сосуде равновесного уровня или периодических колебаний уровня. Последнее имело место в случае слива воды с помощью сифона.

Равновесие носило динамический характер в результате устойчивого выравнивания притока и оттока воды. Устойчивость возникла в силу того, что увеличение уровня вызывало превышение оттока над притоком, а его уменьшение, наоборот, превышение притока над оттоком.

Наиболее неожидан и интересен случай, когда возникают устойчивые колебания. Ведь они возникают без какой-либо видимой побуждающей их причины, сами по себе, почему и названы автоколебаниями (самоколебаниями). Еще совсем недавно возникновение таких колебаний связывали с каким-нибудь периодическим воздействием и с трудом приняли, что периодические колебания могут возникнуть сами по себе, без него.

В науке и инженерном деле признание возможности автоколебаний носило характер революционной ломки сознания, которая последовательно совершалась в электричестве, механике, химии, биологии и экономике. Казалось, что возможность автоколебаний противоречит природе вещей: всякие механические колебания, например колебания маятника, затухают, затухают и колебания электрического контура. А в химической реакции откуда возьмутся колебания: всякая реакция стремится к своему окончанию; некоторому динамическому равновесию. Но все оказалось значительно сложнее. Автоколебания возможны в самых разнообразных физических, химических и биологических системах. Автоколебания могли быть вредными, а могли быть полезными, и вскоре, после упорных и бурных обсуждений, их стали видеть всюду. Сегодня это уже банальность. Но еще в тридцатых–пятидесятых годах прошлого века это было совсем не так. В технике многие автоколебательные явления, такие как шимми передних колес автомобиля, состоящие в возникновении их колебаний; флаттер крыла самолета, при котором возбуждаются колебания крыла самолета, ведущие к его поломке или потере управляемости самолета; колебания режущего инструмента металлообрабатывающих станков и многие другие, длительное время пытались объяснить исходя из предположения о наличии вынуждающей силы и резонанса. Явление шимми предполагалось возникающим от неровностей дороги или дисбаланса колес, причем особенно опасным, если имел место резонанс – близость периода этих воздействий и собственной частоты колебаний колес. Считалось, что флаттер крыла

вызывается вихрями в атмосфере и опять же опасен, если имеет место резонанс. Режущий инструмент вибрировал под влиянием срывов снимаемой стружки.

Эти неправильные представления приводили к бессилию теории в борьбе с подчас очень вредными и опасными явлениями.

Несколько ранее, чем в механике, признание автоколебаний произошло в радиотехнике, поскольку именно они лежали в основе зарождающейся передачи радиосигналов и затем телевидения.

Особенно упорным сопротивлением идее автоколебаний было в химии и биологии, где консервативность проявлялась сильнее и длительнее. Даже когда Белоусов нашел реакцию, которая носила явный автоколебательный характер, видимый по периодическому изменению цвета раствора, химики не сдавались и утверждали, что это невозможно, потому что невозможно никогда.

Обычно принято думать, что великие открытия и перевороты в науке требуют сложных новых методов, на самом деле они требуют скорее простых, но принципиально новых.

Таким принципиально новым подходом, позволяющим понять естественность возникновения автоколебаний и заставляющим нас их принять в противовес ошибочной интуиции, является фазовый портрет динамической системы. Именно он легко убедил в том, что сифон может вызвать автоколебания. Но, пожалуй, общие соображения о том, каков образ автоколебаний на фазовом портрете и как он может у него возникнуть, еще действеннее, чем отдельные примеры. Этот простой геометрический образ – замкнутая фазовая траектория – открыл А. Пуанкаре, а его научное математическое открытие с практикой и теорией колебаний связал А.А. Андронов. Его знаменитая работа так и называлась: «Предельные циклы А. Пуанкаре и автоколебания».

Как же сегодня можно объяснить возникновение автоколебаний в динамических системах, естественность и обычность этого явления?

Пусть динамическая система описывается дифференциальными уравнениями

$$\dot{x} = X(x) \quad (x - \text{вектор}).$$

Его фазовые траектории являются огибающими векторов поля этой системы, т.е. скоростей X , которые в каждой точке x фазового пространства определяют вектор \dot{x} , т.е. векторное поле. Если поле скоростей (т.е. вектор-функция $X(x)$) таково, что фазовая траектория замыкается, то возникают автоколебания. Для этого нужно лишь подобрать подходящим образом вектор-функцию $X(x)$, а это сделать совсем нетрудно. Нужно лишь, чтобы пучок фазовых траекторий «зашел себе в хвост» и «сжался», как это показано на рис. 4.1. Сжатие фазовых траекторий обеспечивает не только существование замкнутой фазовой траектории, но и ее устойчивость. Так бывает очень часто: именно в силу такой ситуации поет скрипичная струна, гудит орган, переливается флейта, поет соловей, мы говорим и поем, работает наше сердце, мы ходим и бегаем, идут часы на нашей руке, передаются радиосигналы, стрекочут кузнечики, противно скрипят тормоза и еще неприятнее скрипит ноготь, которым проводят по стеклу.

В двумерном случае, т.е. при двумерном фазовом пространстве, нарисовать общее поведение фазовых траекторий, ведущих к наличию автоколебания – устойчивой замкнутой фазовой траектории, к которой приближаются соседние тра-

ектории, еще нагляднее. Пусть фазовая траектория γ_1 скручивается и расположенная внутри другая фазовая траектория γ_2 раскручивается, как показано на рис. 4.2.

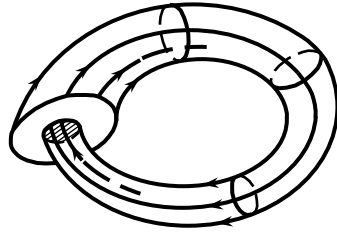


Рис. 4.1. Иллюстрация к возникновению автоколебаний

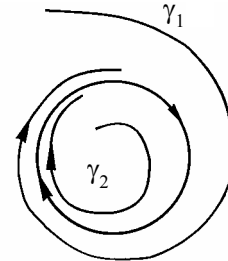


Рис. 4.2. Иллюстрация к возникновению автоколебаний на фазовой плоскости

Наглядно ясно, что между ними есть замкнутая фазовая траектория γ , на которую они наматываются (она изображена пунктиром). В дальнейшем будет приведен конкретный, т.е. описывающий реальный объект, пример подобного поведения фазовых траекторий. А сейчас ограничимся примером дифференциального уравнения [11]. Пусть φ , r – полярные координаты, удовлетворяющие дифференциальным уравнениям

$$\dot{\varphi} = 1, \quad \dot{r} = r - r^2.$$

Из этих уравнений видно, что фазовая траектория, начинающаяся при $\varphi = 0$ и $r = 1/2$, раскручивается, а начинающаяся в точке $\varphi = 0$ и $r = 2$ – скручивается, и их разделяет окружность $r = 1$, соответствующая устойчивому периодическому вращению – автоколебанию, что означает, что касательная в каждой ее точке совпадает с вектором \dot{x} , вектором скорости и перемещения этой точки в фазовом пространстве.

5. ДИНАМИКА УРОВНЯ ЗЕРКАЛА ВОДОХРАНИЛИЩА С ГИДРОСТАНЦИЕЙ

В предыдущих главах мы строили и изучали математические модели бочки с водой при ее оттоке через отверстие в дне или сифон и, возможно, еще притоке. Бочка – объект, заслуживающий внимания: самая известная для современников работа И. Кеплера, помимо его астрологических работ, – о том, как определить объем наполненной бочки, не опорожняя ее. Но это было давно, и с тех пор бочка потеряла свою «актуальность». В соответствии с этим и мы перейдем от бочки к математическим моделям более значимых объектов: водохранилища с гидростанцией и бесценного жидкостного насоса, помещающегося у нас в груди, именуемого сердцем. Эти объекты очень сложные, но их модели останутся простыми до примитивности, и все же кое о чем важном, неочевидном и интересном они нам поведают.

На рис. 5.1 схематически изображено водохранилище с уровнем воды H , подпиранием его плотиной и гидравлической турбиной с электрогенератором.

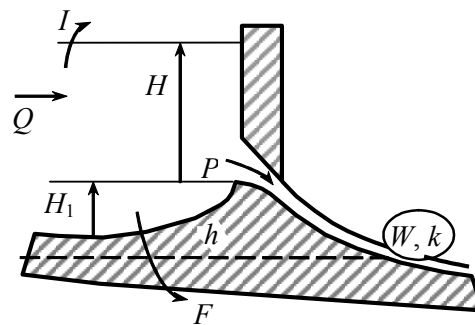


Рис. 5.1. Схематическое изображение водоема с плотиной и гидростанцией

Вода из водоема по трубопроводу поступает в турбину и вращает ее. Соединенный с ней генератор вырабатывает электрический ток требуемой мощности W . В водоем притекает вода, интенсивность общего притока которой равна Q . Величина Q меняется со временем, но мы примем ее постоянной. Вода в водоеме расходуется на вращение турбины, но, кроме того, следует учесть испарение и фильтрацию воды, которые растут с повышением уровня воды в водоеме. Нас будет интересовать изменение уровня воды H в водоеме. Расход воды, необходимый для выработки электроэнергии, зависит от интенсивности использования воды гидротурбиной и ее напора, определяемого уровнем H воды в водоеме. Учитывая все сказанное, можно записать

$$S \frac{dH}{dt} = Q - I - F - \frac{W}{k\rho g(H+h)}, \quad (5.1)$$

где $S(H)$ – площадь поверхности зеркала воды в водоеме, I и F – интенсивности испарения и фильтрации, а последний член выражает интенсивность расхода воды, необходимого для выработки электроэнергии мощности W при коэффициенте полезного действия гидроэлектростанции k и объемной плотности воды ρ . Выражение последнего члена расхода воды в формуле (5.1) находится следующим образом.

Пусть искомый объемный расход равен P . Тогда, если перепад уровня воды перед гидротурбиной $H+h$, то при перетекании воды из водохранилища в гидротурбину освобождается за время dt энергия, равная

$$\rho g P(H+h)dt.$$

Часть $k\rho g P(H+h)$ этой энергии превратится в электроэнергию (k – коэффициент полезного действия гидротурбины вместе с электрогенератором) требуемой мощности W . Поэтому

$$k\rho g P(H+h) = W,$$

откуда находим искомую интенсивность объемного расхода воды гидротурбиной

$$P = \frac{W}{k\rho g(H+h)}.$$

Перейдем к исследованию полученного дифференциального уравнения (5.1). Для этого нам необходимо нарисовать график правой части этого уравнения как функции H . Правая часть состоит из постоянного слагаемого Q , величин I и F , возрастающих с ростом уровня H , и последнего члена, который, напротив, увеличивается с убыванием H . При $H < 0$ последний член исчезает, так как при этом прекращается подача воды в гидротурбину. В соответствии с этим более полное уравнение (5.1) запишется в виде:

$$S\dot{H} = \begin{cases} Q - I - F - \frac{W}{k\rho g(H+h)} & \text{при } H > 0, \\ Q - I - F & \text{при } H < 0, \end{cases} \quad (5.2)$$

а график правой части уравнения (5.1) качественно имеет вид, представленный на рис. 5.2.

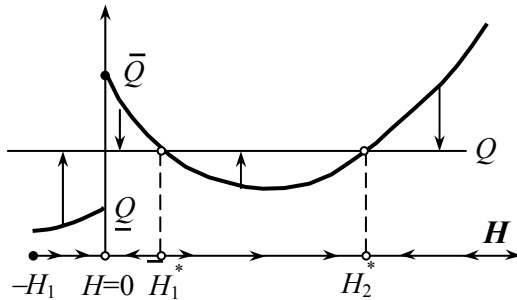


Рис. 5.2. Фазовый портрет системы «водоем – гидростанция с плотиной» при описании изменения уровня воды в водоеме дифференциальным уравнением (5.2)

На нем величина правой части есть разность между Q и остальными членами, которая изображается стрелками. Согласно этому графику, на фазовой полупрямой $H \geq -H_1$ имеются три состояния равновесия: $O_2 (H = H_2^*)$ – устойчивое равновесие, при котором обеспечена полноценная работа гидростанции с выдачей требуемой мощности W ; равновесие O с $H = 0$ также устойчивое, но оно обеспечивает получение не полной мощности W , а только ее части, равной

$$k\rho g(Q - I - F)|_{H=0} < W,$$

поскольку при $H = 0$, как видно из графика рис. 5.2,

$$Q - I - F - \frac{W}{kgph} < 0;$$

и, наконец, неустойчивое состояние равновесия O_1 ($H = H_1^*$). Оно интересно тем, что при $H > H_1^*$ уровень водоема возрастает до величины $H = H_2^*$ и при этом обеспечивается требуемая отдача электроэнергии. Напротив, при $0 < H < H_1^*$ уровень H уменьшается до нуля и полноценная работа гидроэлектростанции не обеспечивается. Заметим, что она не обеспечивается не за счет того, что принципиально не хватает требуемой воды, а за счет того, что режим работы системы выбран неправильно. Ведь при том же притоке Q в равновесии O_2 мощность W гидроэлектростанцией обеспечивается.

Из сказанного следует, что нельзя допускать падения уровня воды в водоеме ниже $H = H_1^*$, H_1^* – это критический уровень, ниже которого необходимо принимать неординарные меры: уменьшать временно потребление электроэнергии и мощность гидроэлектростанции и лишь после того, как уровень H превзойдет H_1^* , снова вернуться к нормальной эксплуатации.

Теперь посмотрим, какие сюрпризы нас могут ожидать при временных уменьшениях интенсивности притока Q или временной необходимости увеличения мощности W . При уменьшении Q состояния равновесия O_2 и O_1 сближаются, при $Q = Q_{кр}$ – сливаются и при $Q < Q_{кр}$ – исчезают, после чего уровень H приближается к $H = 0$. При $Q < Q_{кр}$ равновесный уровень меньше нуля. Аналогичный результат возможен и при временном увеличении W . В обоих случаях для того чтобы вернуться к прежнему нормальному режиму работы гидроэлектростанции, необходимо временно уменьшить мощность гидроэлектростанции и дождаться превышения уровнем H критического значения H_1^* .

Для придания наглядности сказанному нарисуем диаграмму зависимости равновесных режимов от величин Q . При одном значении Q , изображенном на рис. 5.2, эти равновесные состояния – O_1 (H_1^*), O_2 (H_2^*) и O ($H = 0$). Нанесем их при разных Q на диаграмму, как показано на рис. 5.3. Кружочками отмечены устойчивые равновесия и крестиком – неустойчивые.

Будем теперь менять Q и посмотрим по диаграмме рис. 5.3, как меняется равновесие. При уменьшении Q состояния равновесия O_1 и O_2 будут сближаться (H_2^* убывает, а H_1^* возрастает). Для состояния равновесия O величина H по-прежнему равна нулю. При некотором критическом значении $Q = Q_{кр}$ равновесия O_1 и O_2 сольются и при дальнейшем уменьшении Q исчезнут. После того как они исчезнут, в связи с уменьшением Q , останется одно устойчивое равновесие O . В этом равновесии $H = 0$, гидроэлектростанция не может вырабатывать полную мощность W , так как при этом расход воды P , находимый из соотношения

$$Q - I - F - P = 0,$$

меньше, чем расход $W/kgph$, требуемый для получения мощности W . При обратном возрастании Q гидроэлектростанция будет находиться в неполноценном режиме $H = 0$ работы вплоть до $Q = \bar{Q}$, где

$$\bar{Q} = I(0) + F(0) + \frac{W}{kgph}.$$

После возрастания Q сверх \bar{Q} режим O перестанет быть равновесным и гидроэлектростанция перейдет к режиму O_2 , в котором будет отдавать требуемую мощность W в выгодном режиме работы.

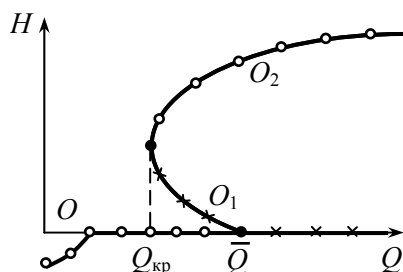


Рис. 5.3. Бифуркационная диаграмма равновесных уровней в водоеме с гидроэлектростанцией

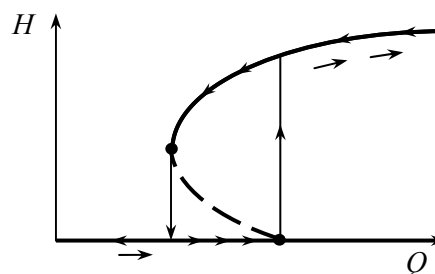


Рис. 5.4. Графическая иллюстрация явления гистерезиса зависимости уровня воды H в водоеме от ее притока Q

На рис. 5.4, повторяющем рис. 5.3, показано, как будет меняться режим работы гидроэлектростанции при медленном, описанном выше, сначала убывании, а потом возрастании притока воды Q . Характерной особенностью этого изменения является то, что при одном и том же притоке Q при $Q_{кр} < Q < \bar{Q}$ возможны два существенно разных режима работы гидроэлектростанции в соответствии с двумя устойчивыми равновесиями O_2 и O . С точки зрения работы гидроэлектростанции хорошим является первый и плохим – второй. То, какой из режимов на самом деле будет иметь место, зависит от предыстории изменения Q . Нетрудно заметить, что характер этой зависимости похож на тот, что имел место для сифона.

С помощью диаграммы рис. 5.3 и рис. 5.4 легко прослеживается зависимость возможных равновесных режимов от притока Q и смена их при медленном изменении Q . Ее называют бифуркационной диаграммой состояний равновесия по параметру Q – диаграммой изменения равновесий с изменением параметра Q .

Как следует из сказанного, стратегия управления гидроэлектростанцией, при которой она все время вне зависимости от уровня H и притока Q должна выдавать мощность W , не только не всегда реализуема, но и плоха, так как может приводить к ситуации, когда гидроэлектростанция не выдает требуемую мощность W , хотя это было бы вполне возможным при другой стратегии управления ею.

Выше предполагалось, что уровень устойчивого равновесия не выше допустимого H_{max} . Это может быть не так, и тогда необходимо использовать неустойчивое равновесие, трансформировав его в устойчивое средствами управления, о чем будет идти речь далее. Если пренебречь потерями воды на испарение и фильтрацию по сравнению с расходом ее на выработку электроэнергии, то станет ясно, что наилучшим режимом работы гидроэлектростанции является ее работа на максимально возможном H_{max} (определяемом высотой плотины), потому что в этом случае расход воды на единицу получаемой мощности минимален. Этот режим работы не только самый экономный с точки зрения расхода воды, но и дающий максимально возможную мощность при заданном притоке Q . Это, если можно так выразиться, идеальный режим работы гидроэлектростанции. Но он реализуется при заданной H_{max} при вполне определенном притоке Q , равном

$$I(H_{\max}) + F(H_{\max}) + \frac{W}{kg\rho(h + H_{\max})}.$$

При уменьшении Q этот оптимальный режим нарушается и при сохранении мощности W переходит в менее выгодный режим с меньшей, чем H_{\max} , высотой H . Нетрудно видеть, что при этом происходят потери возможной суммарной энергии. Чтобы этого не происходило, можно снизить мощность W так, чтобы уровень воды оставался H_{\max} . Как видите, перед нами возникает задача об оптимальном управлении гидростанцией. Теоретическим изучением таких задач занимается специальная сравнительно недавно возникшая и еще развивающаяся наука – теория оптимального управления, являющаяся самостоятельной частью более общей науки, называемой теорией операций.

6. ЭНЕРГЕТИЧЕСКАЯ МОДЕЛЬ СЕРДЦА

Мы построим и исследуем очень простую энергетическую модель сердца. Будем изучать орган очень сложный с помощью модели очень простой. Оказывается, это возможно. Оказывается, и очень простая модель может быть полезной и сказать что-то очень важное и интересное.

Функционально сердце – четырехкамерный насос, снабжающий кровью весь организм. Одна его половина прокачивает кровь по так называемому малому кругу, через легкие, обогащая ее кислородом. Синяя венозная кровь превращается в красную, ее эритроциты, поглотив кислород, меняют цвет. Вторая половина сердца снабжает артериальной кровью, наполненной кислородом и многими другими веществами, все органы и ткани человека. Это так называемый большой круг кровообращения. Сердце не может остановиться, оно должно непрерывно работать днем и ночью, всю нашу жизнь.

Но сердце – не просто насос. Это управляемый насос, подчиняющийся командам вегетативной и центральной нервных систем. При своей работе оно должно согласовываться с физическими и психическими нагрузками своего обладателя, поддерживать баланс между большим и малым кругами, менять свой ритм и интенсивность работы в соответствии с командами нервной системы. Его нервная регуляция многоступенчата и достаточно сложна. Сердце совершает механическую работу за счет запасенной им химической энергии, которая все время неустанно пополняется поступающей в него кровью за счет его же работы. Оно как бы питает само себя, работает, потому что себя питает.

Сердце – это сложный управляемый насос, он должен прогонять кровь по малому и большому кругам, принимая ее предсердиями и прокачивая желудочками. Для этого нужна система клапанов: митральный, аортальный и другие, нужны сложнейшие сети артериальных и венозных сосудов вплоть до тончайших капиллярных сосудов, доставляющих кровь клеткам тканей. Процесс превращения химической энергии АТФ в механическую работу очень сложен, очень сложна и система управления работой сердца, она должна в достаточной мере снабжать организм кровью, не излишествуя и увеличивая это снабжение по мере необходимости при физических, умственных и стрессовых нагрузках. Она должна обеспечить согласованность прокачки крови по большому и малому кругам, не вызывая переполнения кровью легких, органов и тканей организма. В этом ей помогают эластичные сокращения и расслабления сосудов. Система управления должна обеспечить и синхронное сокращение отдельных мышечных волокон сердца, и возможность его автономной работы. Все это достаточно сложно. Современной медицине очень нужны математические модели сердца, они нужны, чтобы понять, как можно ему помочь, особенно в так называемых критических состояниях: при реанимации, операциях на сердце и тех или иных его заболеваниях.

В настоящее время существуют модели сердца, отражающие лишь отдельные стороны его очень сложного функционирования. Я вам расскажу об очень простой модели, где отражено только то, что им управляет нервная система и некоторые поступающие в него химические вещества, и то, что оно живет только

потому, что, работая, себя питает. В соответствии с этим будем описывать его только двумя переменными: командой управления u , которую сердце беспрерывно выполняет, и его текущим энергетическим запасом Q , который расходуется для выполнения работы и пополняется протекающей через него кровью. Энергетический запас сердца Q расходуется и тогда, когда сердце не работает. Он расходуется для поддержания жизнедеятельности его тканей. В соответствии с этим можно записать дифференциальное уравнение

$$\frac{dQ}{dt} = -a - f(u, Q) + g(u, Q), \tag{6.1}$$

где a – интенсивность расходования Q на поддержание жизненного состояния неработающего сердца, $f(u, Q)$ – интенсивность расхода энергетического запаса на совершение работы по перекачке крови и $g(u, Q)$ – интенсивность пополнения запаса Q за счет поступающей в сердце крови. Величины u и Q могут меняться в каких-то пределах $0 \leq u \leq u_{\max}$, $0 \leq Q \leq Q_{\max}$ (рис. 6.1).

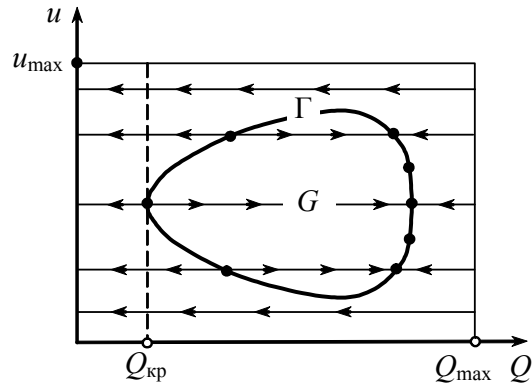


Рис. 6.1. Геометрическая интерпретация энергетической модели сердца, описываемой дифференциальным уравнением (6.1), на плоскости переменных запаса энергии сердца Q и управления u

Для того чтобы написанная модель приобрела конкретность, нужно знание функций $f(u, Q)$ и $g(u, Q)$, хотя бы качественное. Ясно, что как при $u = 0$, так и при $Q = 0$ эти функции обращаются в нуль, т.е.

$$f(0, Q) = f(u, 0) = g(0, Q) = g(u, 0) = 0.$$

Далее $f(u, Q)$ – неубывающая функция обоих аргументов.

Так как Q не может быть больше Q_{\max} , то при $Q = Q_{\max}$

$$\frac{dQ}{dt} \leq 0$$

и поэтому

$$-a - f(u, Q_{\max}) + g(u, Q_{\max}) \leq 0.$$

Естественно предположить, что при $u = u_{\max}$, т.е. максимальном требовании к работе сердца, оно быстро истощается и при этом $dQ/dt < 0$. Таким образом, на сторонах $Q = 0$, $u = 0$ и $u = u_{\max}$ прямоугольника рис. 6.1 правая часть дифференциального уравнения (6.1) отрицательна. На четвертой стороне $Q = Q_{\max}$ она не положительна. Вместе с тем мы длительно живем, и поэтому внутри квадрата рис. 6.1 должна быть область, где правая часть уравнения (6.1) положительна. Это приводит к картинке рис. 6.1, где стрелками указаны направления перемещения

точки $M(u, Q)$ при фиксированном u . Конечно, сколько-нибудь точный вид границы области, где внутренний запас Q энергии сердца растет, остался неизвестным, но для наших целей он и не нужен. Мы будем основывать дальнейшие выводы только на общих представлениях о виде области G , в которой Q возрастает вплоть до ее границы. Вне области G точка $M(u, Q)$ при своем движении при фиксированной команде управления u либо приходит на отрезок, где $Q = 0$, либо приходит на часть Γ границы области G .

Наблюдая возможные движения точки $M(u, Q)$ при изменениях u , можно заметить, что до тех пор пока точка $M(u, Q)$ не оказалась левее прямой $Q = Q_{кр}$, она может побывать в любом месте и быть приведена к любому месту с $Q > Q_{кр}$. Напротив, при $Q < Q_{кр}$ ее судьба предопределена и независимо от изменения u точка $M(u, Q)$ приходит на границу $Q = 0$, это приводит к гибели сердца. К неотвратимой, если не будут приняты соответствующие реанимационные меры. Отсюда ясно, что пребывание вне области G в той ее части, где точка $M(u, Q)$ приходит к отрезку, на котором $Q = 0$, допустимо только до тех пор, пока точка $M(u, Q)$ лежит правее прямой $Q = Q_{кр}$. Причем меры, с помощью которых можно предотвратить это, могущее стать роковым, движение точки $M(u, Q)$, различны для случаев, когда точка $M(u, Q)$ лежит выше области G и когда – ниже нее. В первом случае необходимо вовремя прекратить дальнейшее требование высокой нагрузки на работу сердца, т.е. нужно уменьшить u . Во втором случае, напротив, необходимо стимулировать работу сердца и увеличить команду u . Нетрудно видеть, что первый случай отвечает длительным нагрузкам, которые следует прекратить, а второй – недостаточной работе сердца, недостаточной ее стимуляции и требует увеличения команды, в том числе с помощью возбуждающих медикаментозных средств.

Таким образом, с точки зрения этой модели, возможны два кардинально разных кризисных состояния сердечной деятельности: одно – называемое длительной нагрузкой, другое – длительным чрезмерным ослаблением его деятельности. Заметим, что длительная интенсивная работа сердца может быть вызвана не только реальной физической нагрузкой, но и стрессовым состоянием и перевозбуждением нервной системы.

Теперь посмотрим, как меняется область G при сужении сосудов, питающих сердце, при уменьшении его коэффициента полезного действия, вызываемого, в частности, детренировкой, интоксикацией или общим истощением организма.

В первом случае уменьшается величина функции питания $g(u, Q)$, во втором – увеличивается значение функции $f(u, Q)$.

Оба случая имеют один и тот же эффект – уменьшение размеров области G , что влечет сужение жизненных возможностей организма. В частности, при этом увеличивается $Q_{кр}$, уменьшаются нагрузочные возможности и одновременно возможности сердцу «отдохнуть». Как это ни парадоксально, но сердце вынуждено работать интенсивнее и в суженных пределах изменения нагрузки. Оно лишено отдыха и имеет очень малый запас мощности и суженный резерв. Некоторого улучшения этой ситуации можно добиться путем искусственного лекарственного расширения коронарных сосудов и, возможно, еще умеренной тренировкой, призванной увеличивать коэффициент полезного действия сердечной мышцы и сердца в целом.

Таковы выводы, которые можно сделать из принятой примитивной энергетической модели сердца. В модели учтено лишь то, что интенсивность работы сердца подчиняется управлению *и* и что оно поддерживает свою жизнь, все время работая. Можно заметить, что эти же особенности есть и у каждого живого существа, предоставленного самому себе в окружении безжалостной природы: оно должно все время убегать от врагов, искать и добывать себе пищу, и интенсивность этого поиска и убегания диктуется им самим и обстоятельствами. Разница между бедным животным и нашим сердцем только в жесткости внешних требований, которые человек сумел в большей или меньшей мере смягчить. Пожалуй, в некоторой мере эта модель относится и к сообществу животных и человека.

7. ЗАСОЛЕНИЕ ВОДОЕМА С ЗАЛИВОМ И ЗАГАДКИ КАСПИЙСКОГО МОРЯ

Следующая математическая модель, которую мы рассмотрим, это засоление или растворимое загрязнение водоема стекающими в него водами. Помимо экологического аспекта, эта модель имеет отношение к загадке существенно разных соленостей Черного и Каспийского морей, имеющих общее происхождение. Если говорить о засолении замкнутого ограниченного водоема, то ясно, что по мере стекания в него вод, несущих растворимые вещества, он будет все более и более засаливаться до тех пор, пока не наступит насыщение и выпадение осадка.

Пусть Q – интенсивность втекающего потока, а ν – концентрация загрязняющих веществ. Далее, пусть v – объем водоема, I – интенсивность испарения с его поверхности. Интенсивность испарения зависит от многих факторов: влажности воздуха, его температуры, ветра. Но ее среднее значение определяется в основном размером испаряющей поверхности, которая определяется объемом v . Аналогично интенсивность втекающего потока Q меняется и зависит от очень многих причин, но среднее значение Q достаточно постоянно.

Запишем уравнение баланса притекающей и испаряющейся воды

$$\frac{dv}{dt} = Q - I(v)$$

и уравнение накопления в водоеме общего количества загрязнений

$$\frac{dM}{dt} = \nu Q,$$

где M – масса растворимого загрязнения в водоеме.

Так как $I(v)$ – возрастающая функция v , то из первого уравнения следует установление устойчивого равновесного уровня v^* , определяемого из уравнения

$$Q - I(v) = 0,$$

что поясняется графиком рис. 7.1.

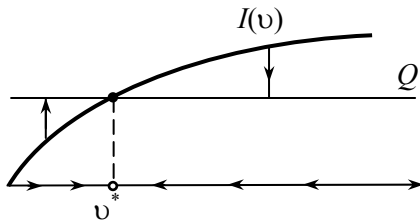


Рис. 7.1. Фазовый портрет динамики уровня воды в водоеме с испарением, зависящим от количества воды в нем

Из второго уравнения следует, что масса загрязнения неуклонно растет. Сначала это влечет за собой увеличение концентрации загрязнения, равной M/v , а затем, с наступлением насыщения, ее рост прекращается и начинается образование донных отложений.

Как видно, в рамках этой модели причина различия в солености Черного и Каспийского морей установлена быть не может. Необходимо учесть еще какие-то

обстоятельства. Но что именно? Черное море не совсем замкнутое, оно соединено узким проливом со Средиземным морем. Но через Босфор вода в основном вытекает, так обстоит дело сейчас, а ранее оно было замкнутым. Каспийское море все время изолировано от океана и, несмотря на это, его соленость много меньше, чем Черного моря.

В чем же может быть существенное различие? Это различие не бросается в глаза. Оно – в наличии у Каспийского моря залива Кара-Богаз-Гол. На первый взгляд кажется, что это не может играть роли, ведь все равно Каспий остается замкнутым водоемом, но это не так. Не так, потому что между водами Кара-Богаз-Гола и Каспия нет перемешивания, вода из Каспия все время течет в залив. Пролит, их соединяющий, узкий и длинный, а снижение уровня воды в нем около 4,5 метров. Учтем все это и составим более полную математическую модель.

Пусть v и v_1 – объемы воды в водоеме и заливе; Q и v – имеют прежний смысл; I и I_1 – интенсивности испарения в водоеме и заливе; q – интенсивность перетекания воды из водоема в залив (рис. 7.2). В этих обозначениях непосредственно имеем:

$$\begin{aligned} \frac{dv}{dt} &= Q - I - q, \\ \frac{dv_1}{dt} &= q - I_1. \end{aligned} \tag{7.1}$$

Кроме того,

$$\begin{aligned} \frac{dM}{dt} &= vQ - \mu q, \\ \frac{dM_1}{dt} &= \mu q, \end{aligned} \tag{7.2}$$

где M и M_1 – общие массы растворимого загрязнения в водоеме и заливе, а μ – концентрация загрязнения в водоеме. Считая, что загрязнение в водоеме не достигает насыщенности, имеем $\mu = M/v$. Напротив, насыщение в заливе имеет место уже давно и в нем образовались гигантские донные отложения солей натрия, магния и других.

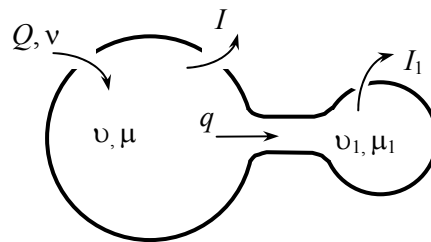


Рис. 7.2. Схематическое изображение Каспийского моря с заливом Кара-Богаз-Гол

Полученная система дифференциальных уравнений (7.1) и (7.2) описывает водный баланс и баланс растворимого загрязнения. Первые два уравнения водного баланса не зависят от оставшихся двух, так что их можно рассматривать сами по себе. Однако это рассмотрение осложнено тем, что I и I_1 зависят от v и v_1 соответственно, а q – функция v и v_1 , причем выписать конкретный вид этих зависимостей весьма затруднительно. Вместе с тем представляется очевидным на-

личие у них устойчивого равновесного режима v^* , v_1^* . Для того чтобы в этом убедиться, надлежит установить наличие решения у системы нелинейных уравнений

$$\begin{aligned} I(v) + q(v, v_1) &= Q, \\ I_1(v_1) - q(v, v_1) &= 0 \end{aligned} \quad (7.3)$$

и обнаружить устойчивость отвечающего ему равновесия (v^*, v_1^*) . Этим займемся позднее, а сейчас, предполагая наличие такого устойчивого равновесия, перейдем к анализу дифференциальных уравнений (7.2) математической модели водоема с заливом. Последнее уравнение уже хорошо известно и не несет ничего нового. Масса примесей в заливе растет со временем, это влечет за собой рост концентрации примесей, который прекращается с достижением насыщения и образованием осадка.

Первое дифференциальное уравнение, напротив, говорит о коренном изменении ситуации и подмене неограниченного роста количества загрязняющего вещества равновесной концентрацией μ^* , находимой из условия обращения правой части рассматриваемого дифференциального уравнения в нуль:

$$\mu q(v, v_1) - vQ = 0.$$

Учитывая, что $v \rightarrow v^*$, $v_1 \rightarrow v_1^*$, находим равновесную концентрацию

$$\mu^* = \frac{vQ}{q(v^*, v_1^*)}.$$

Следует отметить, что процесс установления водного равновесия значительно более быстрый, чем установление равновесной концентрации, и поэтому общий процесс установления можно разбить на два этапа: относительно быстрое установление равновесных объемов v^* , v_1^* и сравнительно медленное установление равновесной концентрации μ^* . После установления равновесных объемов v^* , v_1^* процесс установления равновесной концентрации подчиняется, согласно (7.2), дифференциальному уравнению

$$v^* \frac{d\mu}{dt} = vQ - q(v^*, v_1^*)\mu,$$

решение которого легко находится и при нулевом начальном условии запишется в виде:

$$\mu = \frac{vQ}{q(v^*, v_1^*)} \left[1 - \exp\left(-\frac{q(v^*, v_1^*)}{v^*} t\right) \right] = \mu^* \left[1 - \exp\left(-\frac{q(v^*, v_1^*)}{v^*} t\right) \right]. \quad (7.4)$$

Заметим, что концентрация μ , как следует из формулы (7.4), составляет $1 - e^{-1}$ часть равновесной μ^* спустя время v^*/q^* , равное времени, в течение которого в равновесном режиме объем v^* перетекает из Каспийского моря в залив Кара-Богаз-Гол. При удвоении этого времени μ уже равно $(1 - e^{-2})\mu^*$, так что в качестве оценки времени установления равновесной концентрации μ^* можно принять время $\tau = v^*/q^*$.

Вернемся к равновесной концентрации μ^* . Она равна

$$\mu^* = \frac{vQ}{q(v^*, v_1^*)}$$

и, казалось бы, требует знания v , Q и q . Однако, пользуясь уравнениями (7.3), определяющими v^* и v_1^* , легко находим, что

$$\mu^* = \frac{vQ}{q(v^*, v_1^*)} = v \frac{q^* + I^*}{q^*} = v \left(1 + \frac{I^*}{q^*} \right) = v \left(1 + \frac{I^*}{I_1^*} \right), \quad (7.5)$$

где звездочки обозначают, что соответствующие величины взяты при равновесных значениях $v = v^*$ и $v_1 = v_1^*$.

Отношение I^*/I_1^* , входящее в (7.5), грубо приближенно равно отношению площадей водных поверхностей водоема и залива, т.е. S/S_1 . Эта величина для Каспийского моря и Кара-Богаз-Гола равна примерно 37, и поэтому равновесная концентрация в заливе примерно только в 40 раз больше средней концентрации солей в потоках пресной воды, поступающей в Каспийское море, благодаря втекающим рекам и дождям.

Таким образом, залив Кара-Богаз-Гол играет для Каспийского моря роль отстойника, отсасывающего растворимые загрязнения. Эту же роль в заселяемых стоками воды водоемах могут сыграть искусственные заливы. При этом необходимо обеспечить наличие неисчезающего перепада уровня воды от водоема к заливу, так как обратный поток из залива в водоем все испортит.

Поэтому меньшую, чем у Черного моря, соленость Каспийского моря можно объяснить наличием залива Кара-Богаз-Гол. Но на самом деле наше предположение о равновесности объемов v^* и v_1^* не совсем соблюдается. Уровень Каспийского моря меняется причудливым образом: то колеблется возле одной величины, то возле другой, и различие этих уровней значительно больше колебаний воды возле каждого из них.

Эти изменения не удается объяснить ни изменениями погоды, ни тектоническими изменениями ложа Каспийского моря. Так что это еще одна загадка Каспийского моря.

Относительно небольшие колебания уровня Каспийского моря не нарушают данного выше объяснения малой его солености и формулу (7.5), но они значительны для эксплуатации Каспийского моря и интересны сами по себе. Для того чтобы перейти к разгадке этого удивительного явления, рассмотрим, казалось бы, очевидное предположение о наличии устойчивых равновесных объемов v^* и v_1^* вод моря и залива и зададим вопрос: единственно ли равновесие и нет ли автоколебаний?

Вернемся к уравнениям (7.1) водного баланса в Каспийском море и заливе. Фазовым пространством этой системы дифференциальных уравнений является первый октант плоскости v, v_1 , для которого $v \geq 0$ и $v_1 \geq 0$. Прежде всего установим, что все фазовые траектории приходят в конечную его часть вида:

$$v + v_1 = C > 0.$$

Это следует из того, что при достаточно больших v и v_1 (большом C)

$$\frac{d}{dt}(v + v_1) = Q - I(v) - I_1(v_1) < 0.$$

На рис. 7.3, изображающем поперечное сечение Каспия, залива и соединяющего их протока, виден наиболее высокий уровень соединяющего их протока.



Рис. 7.3. Профиль дна водоема, залива и соединяющего их узкого пролива

Из рисунка видно, что фазовое пространство, первый октант $v > 0$, $v_1 > 0$ плоскости v , v_1 , разбивается на части, когда $q = 0$, $q > 0$ и $q < 0$. Ясно, что $q = 0$, когда в Каспии и заливе уровень ниже указанного на рис. 7.3. И пусть это имеет место при $v < \underline{v}$ и $v_1 < \underline{v}_1$, $q = 0$. Это имеет место еще и тогда, когда уровни Каспия и залива равны, но $v > \underline{v}$ и $v_1 > \underline{v}_1$. Множество точек фазовой плоскости, отвечающее $q = 0$, состоит из прямоугольника $v \leq \underline{v}$ и $v_1 \leq \underline{v}_1$ и некоторой кривой γ , изображенных на рис. 7.4.

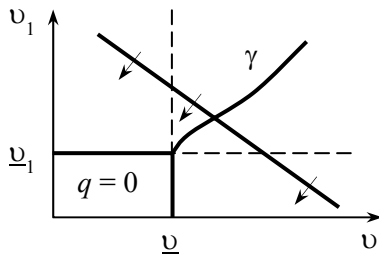


Рис. 7.4. Графическая иллюстрация к поиску области, в которую приходят все траектории фазового пространства $v \geq 0$, $v_1 \geq 0$

На этом же рис. 7.4 изображена прямая $v + v_1 = C$, пересекаемая фазовыми траекториями снаружи внутрь, что показано стрелками. Эта прямая вместе с осями $v = 0$ и $v_1 = 0$ ограничивает область, в которую входят все фазовые траектории, т.е. спустя достаточное время любая фазовая точка (v, v_1) находится в ней.

После этих предварительных соображений перейдем к выяснению установившихся движений системы (7.1). Ясно, что они зависят от величины притока Q , функций $I(v)$, $I(v_1)$ и функции $q(v, v_1)$.

При $Q < I(\underline{v})$ все фазовые траектории входят в четырехугольник, в котором $q = 0$, и далее приближаются к состоянию равновесия $v = v^* < \underline{v}$ и $v_1 = 0$, где v^* – корень уравнения

$$Q - I(v) = 0,$$

единственный, если $I'(v) > 0$. Так как при составлении второго уравнения (7.1) мы учли приток q воды только из Каспия, то получается, что залив высыхает. Отметим, что именно это произошло после временного перекрытия плотиной протока из Каспия в залив.

При $Q > I(\underline{v})$ все фазовые траектории входят в треугольник, ограничиваемый прямыми $v = \underline{v}$, $v_1 = 0$ и $v + v_1 = C$, и поэтому в нем есть состояния равновесия. В одной части этого треугольника $q > 0$, в другой – $q < 0$. Их разделяет кривая γ , на которой $q = 0$. Выше кривой и на ней $q \leq 0$ и $\dot{v}_1 < 0$, так что все фазовые траектории, начинающиеся в криволинейном треугольнике, ограниченном прямыми $v = \underline{v}$, $v + v_1 = C$ и кривой γ , выходят из него и, следовательно, в нем не может быть равновесия.

Теперь покажем, что это равновесие единственно и устойчиво. Далее мы покажем, что замкнутых фазовых траекторий нет, и поэтому это единственное равновесие глобально устойчиво. Конечно, это основывается на определенных свойствах функций $I(v)$, $I_1(v_1)$ и $q(v, v_1)$.

Мы примем, помимо $Q > I(v)$, что

$$\frac{dI(v)}{dv} > 0, \quad \frac{dI_1(v_1)}{dv_1} > 0, \quad \frac{\partial q}{\partial v} > 0, \quad \frac{\partial q}{\partial v_1} \leq 0. \quad (7.6)$$

Преобразуем (7.3) к виду:

$$\begin{aligned} I(v) + I_1(v_1) &= Q, \\ I(v) + q(v, v_1) &= Q. \end{aligned} \quad (7.7)$$

Дифференцируя первое уравнение (7.7) по v , получим

$$\frac{dI}{dv} + \frac{dI_1}{dv_1} \frac{dv_1}{dv} = 0,$$

и поэтому, в силу (7.6), $dv_1/dv < 0$.

Аналогично из второго уравнения (7.7) следует

$$\frac{dI}{dv} + \frac{\partial q}{\partial v} + \frac{\partial q}{\partial v_1} \frac{dv_1}{dv} = 0,$$

и тогда $dv_1/dv > 0$, причем может обращаться в бесконечность.

Первое соотношение (7.7) определяет v_1 как убывающую функцию v ; второе соотношение (7.7) – как возрастающую функцию, могущую иметь вертикальные отрезки или вообще быть вертикальной прямой. Пересечение этих кривых единственно и определяет единственное равновесие v^* , v_1^* .

Таким образом, при естественных предположениях равновесие существует и пусть имеет место при $v = v^*$ и $v_1 = v_1^*$.

Покажем, что оно устойчиво. Для этого пусть ξ и η – малые отклонения от равновесия, так что $v = v^* + \xi$, $v_1 = v_1^* + \eta$.

Отклонения ξ и η удовлетворяют дифференциальным уравнениям

$$\begin{aligned} \frac{d(v^* + \xi)}{dt} &= Q - I(v^* + \xi) - q(v^* + \xi, v_1^* + \eta), \\ \frac{d(v_1^* + \eta)}{dt} &= q(v^* + \xi, v_1^* + \eta) - I_1(v_1^* + \eta) \end{aligned}$$

или после линеаризации

$$\begin{aligned} \dot{\xi} &= \left(-\frac{\partial q}{\partial v} - \frac{dI}{dv} \right) \xi - \frac{\partial q}{\partial v_1} \eta, \\ \dot{\eta} &= \frac{\partial q}{\partial v} \xi + \left(\frac{\partial q}{\partial v_1} - \frac{dI_1}{dv_1} \right) \eta, \end{aligned}$$

где производные вычисляются при $v = v^*$, $v_1 = v_1^*$.

Устойчивость или неустойчивость определяется корнями характеристического уравнения, которое в рассматриваемом случае второй степени, вида

$$\lambda^2 + A\lambda + B = 0.$$

Устойчивость имеет место, если корни этого уравнения лежат слева от мнимой оси. Для этого необходимо и достаточно, чтобы

$$A > 0, \quad B > 0.$$

Вычисляя коэффициенты A и B непосредственно, в силу отмеченных выше неравенств, находим, что

$$A = \frac{\partial q}{\partial v} + \frac{dI}{dv} - \frac{\partial q}{\partial v_1} + \frac{dI_1}{dv_1} > 0,$$

$$B = -\frac{\partial I}{\partial v} \frac{\partial q}{\partial v_1} + \frac{\partial q}{\partial v} \frac{\partial I_1}{\partial v_1} + \frac{\partial I}{\partial v} \frac{\partial I_1}{\partial v_1} > 0.$$

Таким образом, состояние равновесия имеется, оно устойчивое и единственное.

Интуитивно ясно, что это исследование было излишним, а с другой стороны, из него еще нельзя заключить, что кроме устойчивого равновесия нет других установившихся движений, которыми могут быть только устойчивые периодические движения. Но их все же нет, как это следует из широко известного критерия Бендиксона, согласно которому справедливость неравенства

$$\frac{\partial P}{\partial v} + \frac{\partial Q}{\partial v_1} > 0$$

или обратного, где P и Q – правые части дифференциальных уравнений (7.1), достаточно для отсутствия предельных циклов в области, где это неравенство выполняется. В рассматриваемом случае, в силу сделанных предположений,

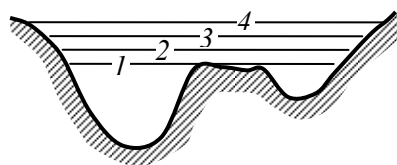
$$\frac{\partial P}{\partial v} + \frac{\partial Q}{\partial v_1} = -\frac{dI}{dv} - \frac{\partial q}{\partial v} + \frac{\partial q}{\partial v_1} - \frac{dI_1}{dv_1} < 0,$$

поэтому состояние равновесия (v^*, v_1^*) является глобально устойчивым и к нему сходятся все движения, т.е. при любых начальных условиях $v(t) \rightarrow v^*$, $v_1(t) \rightarrow v_1^*$ при $t \rightarrow \infty$.

Итак, при естественных предположениях не только интуитивно кажется, но и доказано, что равновесие одно и оно глобально устойчиво. Откуда же наблюдаются переходы равновесного уровня Каспия от одного уровня к другому? Это действительно загадка. Ведь сделанные предположения так естественны. Но податься некуда, и надлежит думать, какое же предположение все же неверно.

Напомним, что сделанные предположения – это $Q > I(v)$ и неравенства (7.6). Конечно, загадка была бы полностью устранена, если бы усредненный по нескольким годам водный приток значительно менялся или менялось ложе Каспийского моря и почему-то все время одинаково. Но этого не наблюдалось. Что же может быть еще не так? Посмотрим на принятые неравенства (7.6), пожалуй, самое слабое место – это $dI/dv > 0$. Действительно, можно придумать, когда это может быть не так. Пусть дно Каспия имеет вид, показанный на рис. 7.5.

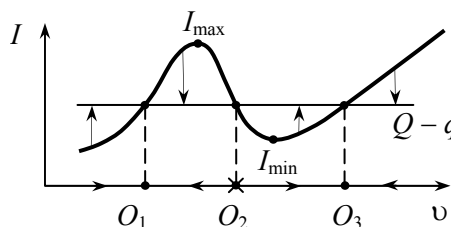
Рис. 7.5. Гипотетический профиль дна Каспийского моря и различные возможные в нем уровни: 1, 2, 3 и 4



Горб в центре рисунка – это возможные мели. Теперь посмотрим, как меняется интенсивность испарения при жаркой погоде при переходе от нижнего уровня, обозначенного на рис. 7.5 цифрой 1, к верхнему уровню 4.

При переходе от 1 к 2 происходит сильное возрастание испарения. При переходе от 2 к 3, несмотря на общее увеличение площади моря, испарение может уменьшаться, так как вода над мелью стала более глубокой и значительно холоднее, а следовательно, и менее испаряющей. При дальнейшем переходе от 3 к 4 снова происходит увеличение испарения. Таким образом, наличие мелей может привести к тому, что график $I(v)$ примет вид, показанный на рис. 7.6.

Рис. 7.6. Фазовый портрет изменения объема вод Каспийского моря при значении притока Q , когда пересечение происходит в трех точках, которым отвечают два устойчивых равновесия O_1 и O_3 и одно неустойчивое O_2



И теперь равновесий может быть три, как это видно из того же рисунка. Пренебрегая ради простоты величиной $q \ll Q$, рассмотрим, что происходит с этими равновесиями при временных изменениях Q , иногда выходящих за рамки I_{\max} и I_{\min} , указанных на рис. 7.6.

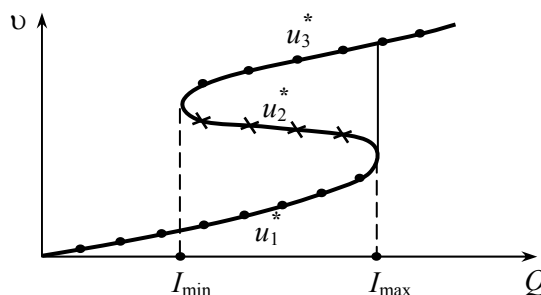
Теперь фазовое пространство – это полупрямая $v \geq 0$, а фазовый портрет определяется уравнением

$$\frac{dv}{dt} = Q - I(v).$$

Точки пересечения графика $I(v)$ и прямой Q на рис.(7.6) определяют равновесия, а стрелки – величину и знак разности $Q - I(v)$, т.е. dv/dt . В случае, показанном на рис. 7.6, состояний равновесия три: O_1 , O_2 и O_3 , причем O_1 и O_3 – устойчивые, а O_2 – неустойчивое.

Теперь мы можем нарисовать графики зависимости O_1 , O_2 и O_3 от Q . Координаты состояний равновесия O_1 , O_2 и O_3 обозначим соответственно u_1^* , u_2^* и u_3^* и придем к зависимостям u_1^* , u_2^* и u_3^* от Q , изображенным на рис. 7.7.

Рис. 7.7. Бифуркационная диаграмма равновесных уровней объемов воды в Каспийском море (кружочки – устойчивые равновесия, крестики – неустойчивые)



Кружочкам на этом графике, называемом бифуркационной диаграммой, соответствуют устойчивые равновесия, а крестикам – неустойчивые. При $I_{\min} < I < I_{\max}$ имеется два устойчивых равновесия u_1^* и u_3^* . При убывании Q сверх I_{\min} оба они переходят в u_1^* , причем равновесие u_3^* – скачком. При последующем возрастании Q u_1^* меняется также скачком на u_3^* при возрастании Q сверх I_{\max} .

При последующем убывании Q равновесие u_3^* сохраняется, перескакивая на u_1^* при уменьшении Q сверх I_{\min} .

Сказанное может объяснить имеющиеся место на Каспийском море изменения его уровня. Эти изменения происходят скачками с низкого уровня на высокий, когда Q при своем текущем изменении превосходит I_{\max} . И наоборот, высокий уровень моря снижается скачком на низкий при снижении Q ниже I_{\min} . Здесь идет речь о скачкообразном изменении устойчивых равновесий с изменением притока Q . Уровень же Каспийского моря меняется вовсе не скачком, а достаточно медленно в соответствии со скоростью перехода к новому состоянию равновесия, причем разные уровни u_3^* и u_1^* имеют место при одинаковом притоке Q , а их смена происходит сравнительно редко, через несколько десятилетий, когда $Q > I_{\max}$ и, соответственно, $Q < I_{\min}$.

8. ЭКСПОНЕНЦИАЛЬНЫЕ ПРОЦЕССЫ

От рассказа об отдельных математических моделях перейдем к описанию достаточно широкого класса динамических процессов, так называемых экспоненциальных процессов, и некоторых других (логистических, взрывных).

Экспоненциальные процессы могут быть двух видов: экспоненциального возрастания и экспоненциального убывания. Они оба описываются очень простым дифференциальным уравнением

$$\frac{dx}{dt} = \lambda x. \quad (8.1)$$

При $\lambda > 0$ – это экспоненциально возрастающий процесс, при $\lambda < 0$ – экспоненциально убывающий. Экспоненциальное возрастание или убывание может быть как по времени t , так и по какой-либо другой, отличной от времени физической переменной.

Решение дифференциального уравнения (8.1), обращающееся в x_0 при $t = 0$, имеет вид:

$$x = x_0 e^{\lambda t} \quad (8.2)$$

и обладает следующим замечательным характеристическим свойством: спустя промежуток времени $\tau = |\lambda|^{-1} \ln 2$ от любого начального момента t ,

$$\frac{x(t + \tau)}{x(t)} = 2, \quad (8.3)$$

если $\lambda > 0$ и процесс возрастающий, и

$$\frac{x(t + \tau)}{x(t)} = \frac{1}{2}, \quad (8.4)$$

если $\lambda < 0$ и процесс убывающий. Это говорит о том, что арифметической прогрессии возрастающих значений $t + n\tau$ ($n = 0, 1, 2, \dots$) соответствует геометрическая прогрессия со знаменателем 2 или 1/2. На рис. 8.1 представлены эти свойства возрастающего и убывающего экспоненциальных процессов.

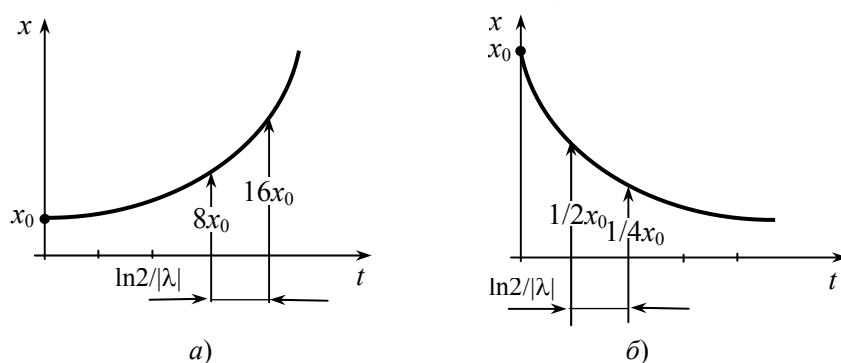


Рис. 8.1. Графическая иллюстрация экспоненциального возрастания (а) и убывания (б) и времени удвоения и уменьшения вдвое

Основное, что поражает в экспоненциальных процессах, это их фантастически быстрое, не вполне осмысливаемое нами, все убыстряющееся возрастание

или, наоборот, замедляющееся убывание. Хорошим примером может служить притча о царе и изобретателе шахмат, которого царь пожелал вознаградить и сказал: «Проси у меня, что хочешь, хоть полцарства, хоть царскую дочь». Но изобретатель шахмат оказался очень скромным человеком. Он сказал: «Не надо мне ничего, дай мне самую малость. На первую клетку шахматной доски положи пшеничное зернышко, на вторую – два, на третью – четыре и т.д.». Царь сказал: «Что за пустяк ты просишь, но пусть будет по-твоему». Когда дошли, до середины доски, то закрома царя уже были пусты. А чтобы заполнить все 64 клетки шахматной доски, требуется зерна много больше, чем на всем земном шаре.

Столь же фантастически ускоряясь, происходит беспрепятственное размножение клеток, микробов и вирусов. Клетки и микробы делятся каждые час-два, и поэтому через 3–6 суток их число того же порядка, что и число пшеничных зерен на всех клетках шахматной доски. У вирусов время удвоения много меньше часа, так что за время, меньшее суток, возможна катастрофа, если им не мешать размножаться. Так же, все ускоряясь, истощая ресурс, протекают цепные ядерные и химические реакции, лавинообразующие и лавинообразные процессы. В них время удвоения может измеряться тысячными долями секунды, так что процесс может носить взрывной характер, достигая фантастических размеров за ничтожные доли секунды.

Аналогичны примеры распространения эпидемий и слухов. Они тоже носят в своей начальной стадии экспоненциальный характер. Ярким примером плохой нашей ориентировки в скорости роста экспоненциального процесса является недобдуманный ответ на вопрос: «Сколько рукопожатий жителя маленького городка нашей страны связывает с президентом США?». Обычно отвечают, что достаточно много, в то время как на самом деле – заведомо меньше десяти. Быстро оценить это можно, если учесть, что каждый человек, как правило, здоровается за свою жизнь не менее, чем со 100 людьми. Поэтому в цепочке из 5 человек, если не учитывать повторных рукопожатий, насчитывается 10^{10} человек, больше чем их на земном шаре. Это, в частности, указывает на стремительность распространения эпидемий с коротким продромальным периодом и еще большую стремительность распространения слухов.

Во всех приведенных примерах скорость нарастания некоторой величины тем больше, чем больше сама эта величина. До тех пор, пока не истощается материал, из которого создается эта величина, и не нарушаются условия ее роста, такой процесс носит нарастающий экспоненциальный характер и подчиняется дифференциальному уравнению

$$\frac{dN}{dt} = \lambda N,$$

где N – число размножившихся клеток или вирусов, число заболевших людей во время эпидемии, число размножающихся нейтронов или молекул при цепной ядерной или химической реакциях, число людей, до которых дошел слух о сенсационном событии или новости, число жителей страны или всей нашей планеты и т.д. В какой-то мере такая же закономерность относится к приобретению знаний, изучению языков, поскольку уже приобретенные знания облегчают и ускоряют их дальнейшее постижение. Это относится как к накоплению знаний отдельным че-

ловеком, так и всем человечеством. Это относится к развитию науки, производства, нарастанию экологического загрязнения и разрушения.

Быстрота нарастания экспоненциального процесса приводит к явлениям внезапного возникновения кризиса и исчезновения, «схлопывания» ресурсов. Представьте себе, что вы летом устроились в палатках на берегу чудесного озера, в котором предполагали купаться и плавать. Пусть это озеро имеет площадь 16384 м^2 . Большое, хорошее озеро метров 200 длиной и метров 100 шириной. В день вашего приезда в укромном уголке озера были зеленые водоросли, занимающие площадь в 1 м^2 . Вы их и не заметили. Пусть время их удвоения при размножении равно одним суткам. В течение 13 дней они вам особенно не мешали, хотя в последние 3–4 дня вы уже их заметили. Но вот наступил 14 день, и все озеро заросло зелеными водорослями и купаться и поплавать уже негде. Катастрофа произошла за одну ночь, еще вчера ничто не предвещало беды и половина озера была без водорослей. Это 8192 м^2 , более чем достаточно. Чистая вода озера «схлопнулась» за сутки, а 13 дней до этого все было очень хорошо и не предвещало никаких неприятностей.

Экспоненциальный рост знаний, науки, промышленности, богатства приводит еще к одной связанной с ними особенности. В качестве примера возьмем индивидуальное накопление знаний, знания языков в частности. Для каждого человека в период его активной жизни и при наличии благоприятствующих обстоятельств накопление знаний носит экспоненциально нарастающий характер, поскольку приобретение новых знаний облегчается имеющимися, т.е. грубо приближенно

$$\frac{dZ}{dt} = \lambda Z.$$

В силу этого, приобретение знаний носит пороговый характер, т.е. как только они достигают некоторого порога Z^* , их дальнейшее накопление приобретает необычайно быстрый характер и человек за свою короткую жизнь может достичь гигантских высот. Но для этого он должен в своих знаниях превзойти некоторый трудно достижимый порог. Это удается немногим. Но кто этого достиг, далее устремляется к вершинам. Это приводит к тому, что среди всех людей резко выделяется по объему и качеству познаний лишь весьма ограниченное число, ничтожные доли процента, и между ними и остальными имеет место колоссальный разрыв. В этом можно видеть объяснение явления исключительности.

Следующий пример – пример внезапного кризиса – возможно, уже из вашего опыта. Вы заразились неделю тому назад. Шесть дней длился ничем не обнаруживающий себя, так называемый, продромальный период заболевания. На седьмой день – озноб, высокая температура, плохое или даже тяжелое состояние. Конечно, микробы или вирусы размножаются в вашем организме не беспрепятственно. Им противостоит мощная иммунная система, но, возможно, еще не мобилизованная и недостаточная, чтобы подавить инфекцию. Для этого организму нужно время. В силу этого, экспоненциальная катастрофа в значительной мере смазана, но все же она проявляется, и чем расслабленной и неподготовленной иммунная система, чем позже ее реакция, тем четче и быстрее растет численность

возбудителей болезни и интоксикация достигает критических значений, и после первых слабых признаков заболевания наступает состояние, угрожающее жизни.

Об экспоненциальном росте населения земли впервые сказал Мальтус в конце XVIII века. У него был предшественник в XVII веке – известный математик Фибоначчи. Он поставил и разрешил вопрос: «Сколько пар кроликов в год от одной пары рождается».

Экспоненциальное размножение предполагает отсутствие всяких ограничений. На самом деле они всегда есть (например, ограничение ресурса), и простейший их учет состоит во введении отрицательного квадратичного члена, так что соответствующее дифференциальное уравнение

$$\frac{dx}{dt} = \lambda x - \varepsilon x^2 \quad (8.5)$$

имеет решение вида, так называемой, логистической кривой, изображенной на рис. 8.2. При этом первоначальное возрастание, близкое к экспоненциальному, замедляется и ограничено предельно достижимой величиной $x_\infty = \lambda/\varepsilon$ (см рис. 8.2).

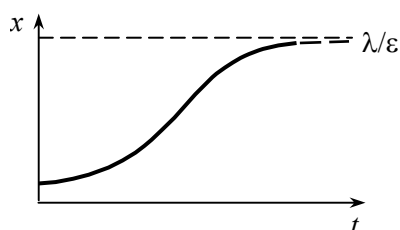


Рис. 8.2. График логистической зависимости (кривой)

Многие процессы, экспоненциальные на начальном участке, затем испытывают замедление и близки к логистической кривой. Таковы рост народонаселения, рост численности животных, развитие промышленности, науки, образования и многое другое.

Возможны процессы нарастающие быстрее, чем экспоненциальные. Таково, например, размножение насекомых, где численность потомства определяется частотой встреч, которая на ограниченном участке пропорциональна квадрату числа его населения, т.е.

$$\frac{dN}{dt} = aN^2. \quad (8.6)$$

Решение этого уравнения

$$N = \frac{1}{N_0^{-1} - at} \quad (8.7)$$

при начальном условии

$$N|_{t=0} = N_0$$

неограниченно возрастает при приближении t к конечному значению $1/a N_0$, что соответствует взрывному росту численности, хорошо известному, например, у саранчи.

Экспоненциальный характер нарастания, во всяком случае на начальном участке, характерен для неравновесных уходов от неустойчивых состояний равновесия. Так, пусть система описывается дифференциальным уравнением

$$\dot{x} = f(x), \quad (8.8)$$

$x = x^*$ – ее неустойчивое состояние равновесия. На рис. 8.3 показана горка, вершинная точка которой отвечает неустойчивому равновесию. Скатываясь с нее, точка все быстрее и быстрее набирает скорость.

Введем новую переменную ξ , положив

$$x = x^* + \xi,$$

и подставим ее в уравнение (8.8), найдем, что

$$\dot{\xi} = f(x^* + \xi) = f(x^*) + f'(x^*)\xi + \dots \quad (f(x^*) = 0),$$

где многоточия означают члены второго и выше порядков малости относительно ξ . На начальном участке ухода от неустойчивого равновесия, когда ξ еще мало, можно приближенно принять, что

$$\dot{\xi} = \lambda \xi,$$

где $\lambda = f'(x^*) > 0$, т.е. возрастание ξ , во всяком случае на начальном участке, действительно происходит по экспоненте.

До сих пор описывались экспоненциально возрастающие процессы. Перейдем к экспоненциально убывающим. Их тоже очень много, и они весьма разнообразны. Это остывание чашки горячего чая, разряд электрически заряженного конденсатора, и в частности электронной трубки телевизора после его выключения, убывание интенсивности излучения при прохождении его через частично прозрачную среду, убывание массы разгоняющейся ракеты, торможение падения после раскрытия парашюта, гибель популяции, попавшей в неблагоприятные условия, например, резкого недостатка пищи, приход к равновесному состоянию и многое, многое другое.

Экспоненциальное убывание – это все время замедляющееся убывание, настолько быстрое, что весь процесс замедления всегда ограничен. Именно ввиду этой конечности времени быстроногий Ахилл не мог догнать черепаху, потому что он пробежал сначала метр, потом полметра, затем четверть метра, ... и всего не более двух метров. Ахилл не догоняет черепаху, но вместе с тем приближается к ней он очень быстро, в том смысле, что расстояние до нее сокращается сначала в два раза, потом еще в два раза и всего в четыре, потом еще в два раза, и всего в восемь и т.д. Из сказанного следует, что экспоненциальное убывание можно трактовать как все убыстряющееся приближение к финальному значению.

Перейдем к конкретным примерам. На рис. 8.4 представлена схема конденсатора емкости C , разряжающегося на сопротивление R .

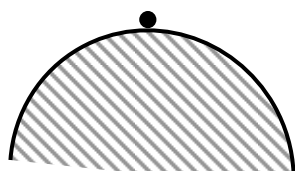


Рис. 8.3. Иллюстрация неустойчивого равновесия в виде материальной точки на вершине выпуклой кривой

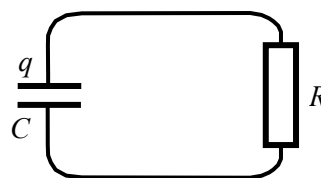


Рис. 8.4. Электрический замкнутый контур, состоящий из сопротивления и емкости

В соответствии с законами Кирхгофа, дифференциальное уравнение разряда конденсатора пишется в виде:

$$R\dot{q} = -\frac{q}{C}$$

или

$$\frac{dq}{dt} = -\frac{1}{RC}q$$

и, следовательно, описывается экспоненциальным убывающим процессом с временем уменьшения вдвое, равным

$$\tau = RC \ln 2.$$

Это время τ пропорционально емкости C и сопротивлению R . При начальном заряде $q = q_0$

$$q(t) = q_0 e^{-t/RC}.$$

Следующий пример – торможение парашюта. Пусть по достижении скорости падения v_0 парашют раскрылся и тормозит падение пропорционально его скорости. Согласно закону Ньютона,

$$m\dot{v} = -mg - hv.$$

Решение этого уравнения следующее:

$$v = \frac{mg}{h} + \left(v_0 - \frac{mg}{h} \right) e^{-(h/m)t}.$$

Из него следует, что начальная скорость v_0 ($v_0 > 0$) экспоненциально замедляется до постоянной скорости спуска, равной mg/h . График этого процесса изображен на рис. 8.5.

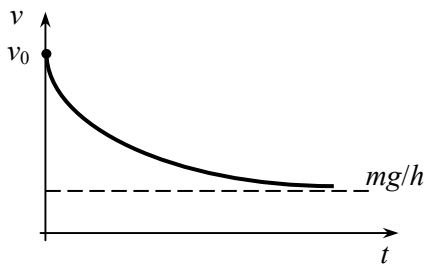


Рис. 8.5. Временной график падения тела при вязком сопротивлении

До сих пор рассматриваемые нами экспоненциальные процессы носили временной характер, т.е. они менялись экспоненциально с изменением времени. Рассмотрим теперь также экспоненциальные процессы, но не от временной переменной.

Сейчас уже мало кто читает Жюль Верна, а когда-то им зачитывались. Он в своих фантастических романах предвидел очень многое, но иногда ошибался. Одной из таких явных ошибок был описанный им полет на Луну из пушки. Полет он предсказал, но из пушки – это ошибка. Жюлю Верну казалось, что чем больше пушка, тем быстрее она выбрасывает ядро. Но это не совсем так; есть предельная скорость выброса и ее превзойти нельзя. Пусть используется порох (как у Жюль Верна) или другое взрывчатое вещество, и пусть при взрыве его массы m выделя-

ется энергия $E_{\text{хим}}$ на единицу массы. Эта энергия может сообщить массе m (без учета массы ядра) скорость v . В лучшем случае

$$\frac{mv^2}{2} = mE_{\text{хим}},$$

т.е. скорость выбрасываемого ядра не может быть больше

$$v_{\text{max}} = \sqrt{2E_{\text{хим}}}.$$

Вот тот абсолютный теоретический предел скорости, который можно достигнуть стрельбой из пушки. Он много меньше требуемой первой космической скорости, равной 8 км/с.

Циолковский был первым, кто понял, что необходим другой путь использования того же химического вещества, при котором теоретически может быть достигнута любая, сколь угодно большая скорость. Это путь использования ракетного двигателя. Ракетный двигатель развивает тягу за счет выбрасываемой им струи газа. Пусть ракета массы m имеет скорость v и из нее выбрасывается в обратном к скорости направлении масса со скоростью c (рис. 8.6). Эта выбрасываемая масса $-dm$ приобретает скорость c за счет химической энергии сгорающего в ракете топлива, так что

$$\frac{dmc^2}{2} = dE_{\text{хим}}, \quad (8.9)$$

где $dE_{\text{хим}}$ – затрата химической энергии, запасенной в ракете, переходящая в кинетическую энергию движения газа. То, что ракета при выбрасывании из себя назад массы $-dm$ получает ускорение движения вперед, можно обнаружить и вычислить величину этого прироста скорости, исходя из закона сохранения энергии, т.е. приравнивания энергии до и после выброса массы.

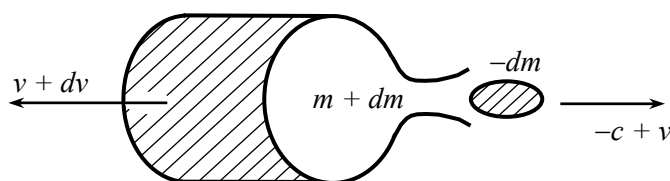


Рис. 8.6. Схема, поясняющая вывод уравнения разгона ракеты при выбрасывании из нее через сопло газа

До выброса массы ракета обладала энергией, равной

$$E_{\text{хим}} + \frac{mv^2}{2}.$$

Она складывалась из химической энергии $E_{\text{хим}}$ содержащегося в ней топлива и ее кинетической энергии движения. После выброса массы общая энергия складывается из уменьшенной химической энергии, измененной кинетической энергии ракеты и кинетической энергии выброшенной массы газа $-dm$ и в соответствии с этим равна

$$E_{\text{хим}} + dE_{\text{хим}} + \frac{(m + dm)(v + dv)^2}{2} - \frac{dm(v - c)^2}{2}.$$

Приравнивая общие энергии до и после выбрасывания массы, найдем, что

$$E_{\text{хим}} + dE_{\text{хим}} + \frac{(m + dm)(v + dv)^2}{2} - \frac{dm(v - c)^2}{2} = E_{\text{хим}} + \frac{mv^2}{2}.$$

Отсюда, учитывая (8.9) и, естественно, отбрасывая члены, содержащие $dm dv$, $(dv)^2$ более высокого порядка малости, приходим к простому соотношению

$$mv dv + cv dm = 0$$

или

$$\frac{dm}{dv} = -\frac{1}{c} m.$$

Таким образом, масса m ракеты убывает как экспоненциальная функция скорости v :

$$m = m_0 e^{-v/c}. \quad (8.10)$$

Так как нас интересует возрастание скорости с убыванием массы в связи с ее отбрасыванием назад, то естественно последнее соотношение записать в виде формулы Циолковского

$$v = c \ln \frac{m_0}{m}.$$

Из выражения скорости v через отношение начальной массы m_0 к массе m следует, что, выбрасывая все большую и большую часть массы, ракета может достичь все больших и больших скоростей, причем каждому уменьшению массы в два раза отвечает увеличение скорости ракеты на величину $c \ln 2$. Эта закономерность в принципе позволяет достичь любых сколь угодно больших скоростей, но для этого, так как конечная масса ракеты не может быть очень маленькой, необходимо, соответственно, значительное увеличение начальной массы m_0 . Так что для полета на Луну нужна не большая пушка, как думал Жюль Верн, а большая ракета. Не большая, а громадная. Из соотношения (8.10) следует, что начальная масса m_0 – экспоненциально возрастающая функция скорости v :

$$m_0 = m e^{v/c},$$

и удвоение начальной массы при заданной конечной происходит с каждым возрастанием скорости на величину $c \ln 2$. Поэтому, чтобы снизить начальный вес ракеты, необходимо предельно уменьшить ее конечный вес. Существенного снижения конечного веса ракеты достигают путем ее многоступенчатой конструкции, когда уже в полете ненужные части ракеты отбрасываются и остается минимально необходимая ступень ракеты.

Другой пример. Динамика численности популяции N , т.е. величина производной \dot{N} определяется соотношением скоростей рождаемости P и гибели Γ , так что

$$\dot{N} = P - \Gamma,$$

P и Γ являются функциями численности N и внешних условий. При неизменности внешних условий и естественной простейшей зависимости от N ($P = aN$, $\Gamma = bN$) с учетом знака величины $\lambda = a - b$ приходим к процессам экспоненциального размножения или гибели.

Процессу гибели популяции аналогичен радиоактивный распад, влекущий экспоненциальное убывание нераспавшегося вещества. Скорость радиоактивного распада характеризуют временем полураспада. Чем меньше это время, тем быстрее исчезает опасное для человека радиоактивное излучение.

Как процесс только гибели, можно представить поглощение излучения не вполне прозрачной средой. При этом очень удобно представить себе излучение как поток большого числа N фотонов, а среду – обладающей некоторыми независимыми вероятностями поглощения каждого из фотонов потока на единицу пробегаемого им пути. В этой интерпретации при очень большом N

$$\frac{dN}{dt} = -\lambda N,$$

что приводит к экспоненциально убывающему процессу с пробегом, уменьшающим интенсивность излучения вдвое, равным $\lambda^{-1} \ln 2$.

Экспоненциальным характером убывания интенсивности излучения объясняется факт крошечной тьмы в глубинах океана при достаточной освещенности в его верхних слоях. Действительно, пусть освещенность падает вдвое на каждые 20 метров. Это падение освещенности в ясный светлый день совершенно незаметно. Но уже на глубине километра будет тьма, поскольку интенсивность излучения будет в $2^{50} \approx 10^{12}$ раза меньшей.

Эта же модель потока частиц может привести к экспоненциально растущему лавинообразному процессу, если при столкновении быстро движущейся частицы с молекулой газа может образоваться более одной так же быстро движущихся частиц. Такие лавинные процессы наблюдаются при вхождении в атмосферу из космоса частиц со скоростями, близкими к скорости света, и огромными энергиями.

Охлаждение не очень сильно нагретого тела – стакана чая или кофе – это также экспоненциально убывающий процесс. Действительно, если температуру окружающей среды принять за отсчетную, то изменение температуры T нагретого тела подчиняется дифференциальному уравнению

$$\frac{dT}{dt} = -\lambda T$$

в силу того, что отдача тепла нагретым телом окружающей среде приблизительно пропорциональна разности температур тела и среды.

Наконец, отметим, что все процессы прихода к равновесному состоянию, а именно так можно трактовать разряд заряженного конденсатора, торможение движущегося тела, гибель популяции, охлаждение нагретого тела, как правило, имеют экспоненциально убывающий характер. Это обосновывается тем, что линейаризованное в окрестности равновесия уравнение движения в общем случае совпадает с дифференциальным уравнением экспоненциально убывающего процесса.

Так, для одномерного процесса

$$\frac{dx}{dt} = f(x),$$

имеющего состояние равновесия x^* , для величины отклонения ξ от x^* имеем:

$$\frac{d\xi}{dt} \approx f'(x^*)\xi,$$

где в силу устойчивости $f'(x^*) < 0$.

Для многомерного случая все несколько сложнее и процесс убывания описывается выражением вида:

$$\sum C_s e^{-\lambda_s t} + \sum D_k e^{-\mu_k t} \cos(\omega_k t + \varphi_k),$$

где λ_s и μ_k – положительные, т.е. суперпозицией экспоненциально убывающих и экспоненциально убывающих осциллирующих составляющих. Ясно, что финальный приход к равновесию будет описываться членом с наименьшим показателем λ_s или μ_k , и в этом смысле тоже может характеризоваться как экспоненциально убывающий или осциллирующий экспоненциально убывающий процесс.

9. ДИНАМИКА СОСУЩЕСТВОВАНИЯ ПОПУЛЯЦИЙ

Выше рассказывалось о беспрепятственном размножении популяции. Оно происходило по экспоненте. Однако в реальных условиях популяция находится в весьма разнообразных взаимоотношениях с другими популяциями и внутри себя. Эти взаимоотношения могут быть типа хищник и жертва (антагонизм), взаимопомощи (симбиоз), взаимной конкуренции. В простейшем случае взаимодействующих популяции две, так что именно это взаимодействие является решающим и наиболее важным, а остальные носят второстепенный характер.

Первая экологическая модель взаимодействующих популяций – это модель хищник–жертва, написанная Лоткой и Вольтерра [14–16]. Она учитывает, что популяция жертвы может существовать сама по себе, а популяция хищника – только питаясь жертвой. Поэтому в отсутствие хищника жертва размножается экспоненциально согласно уравнению

$$\dot{x} = ax,$$

а хищник в отсутствие жертвы вымирает, так что

$$\dot{y} = -cy.$$

Хищник съедает тем больше жертвы, чем ее больше и чем более многочислен он сам, поэтому при наличии хищника

$$\dot{x} = ax - bxy. \quad (9.1)$$

Съеденное количество жертвы способствует размножению хищника и поэтому

$$\dot{y} = -cy + dxy. \quad (9.2)$$

Дифференциальные уравнения (9.1) и (9.2) и есть знаменитые и широко известные дифференциальные уравнения Вольтерра–Лотки, уравнения, с которых началось развитие теории взаимодействующих популяций.

Состоянием этой динамической системы, описываемой уравнениями (9.1) и (9.2), являются численности x и y жертвы и хищника, фазовым пространством – октант $x \geq 0, y \geq 0$. При $y = 0$ x возрастает по экспоненте, при $x = 0$ y убывает по экспоненте. Точка $x = y = 0$ является состоянием неустойчивого равновесия типа седла. Это следует из того, что вблизи точки равновесия $x = y = 0$ членами, содержащими произведение малых величин x и y , можно пренебречь и приближенно записать:

$$\dot{x} = ax, \quad \dot{y} = -cy.$$

Кроме равновесия $x = y = 0$, возможно еще равновесие $x = x^*, y = y^*$, где x^* и y^* – отличные от нуля решения уравнений

$$ax - bxy = 0, \quad -cy + dxy = 0,$$

то есть

$$x^* = c/d, \quad y^* = a/b.$$

Тип равновесия (x^*, y^*) можно найти, как обычно, составляя его характеристическое уравнение

$$\begin{vmatrix} a - by^* - \lambda & -bx^* \\ dy^* & -c + dx^* - \lambda \end{vmatrix} = 0$$

или

$$\lambda^2 + ac = 0.$$

Корни характеристического уравнения чисто мнимые $\lambda_{1,2} = \pm i\sqrt{ac}$ и отвечают в линейном приближении состоянию равновесия типа центр, которое при учете отбрасываемых при линеаризации членов может оказаться как устойчивым, так и неустойчивым фокусом. В данном случае более полный анализ обнаруживает, что нелинейные члены сохраняют вид состояния равновесия, т.е. оно остается типа центра. Более детальный анализ основывается на возможности интегрирования дифференциальных уравнений Вольтерра–Лотки. Уравнение Вольтерра–Лотки можно записать в виде:

$$\frac{dx}{ax - bxy} = -\frac{dy}{-cy + dxy}.$$

Откуда, разделяя переменные x и y , получим

$$\frac{(-c + dx)dx}{x} = \frac{(a - by)dy}{y}$$

и после интегрирования

$$-a \ln y - c \ln x + dx + by = h,$$

где h – произвольная константа. Последнее уравнение – это уравнение фазовых траекторий. Каждой траектории соответствует свое значение константы интегрирования h . Качественно построить фазовые траектории на плоскости x, y можно, заметив, что каждая из фазовых траекторий получается проектированием на плоскость x, y линии пересечения поверхности

$$z = dx + by - c \ln x - a \ln y \equiv f(x, y) \quad (9.3)$$

плоскостью $z = h$. Вид этой поверхности, показанный на рис. 9.1, можно найти из соображений, что на осях x и y $f(x, y) = \infty$ и $f(x, y) \rightarrow \infty$ при возрастании хотя бы одной из переменных x или y . Точке минимума этой поверхности $z = f(x, y)$ отвечает состояние равновесия (x^*, y^*) . Все остальные фазовые траектории представляют собою овалы, заключенные друг в друге и охватывающие равновесие (x^*, y^*) , так что фазовый портрет оказывается вида, изображенного на рис. 9.2.

Несмотря на свою простоту, модель Вольтерра–Лотки качественно правильно отражает колебательный характер численностей x и y в сосуществовании хищника и жертвы. Он был обнаружен, например, при многолетних наблюдениях численности зайцев и рысей в Канаде и многих других случаях.

Согласно фазовому портрету, колебания численности популяции имеют периодический характер, но могут происходить с разными размахами: от нулевых до бесконечных, но, как заметил Вольтерра, временные средние значения числен-

ности x и y остаются неизменными, равными координатам x^* и y^* состояния равновесия. Это непосредственно следует из усреднения по периоду соотношений

$$\frac{\dot{x}}{x} = a - by, \quad \frac{\dot{y}}{y} = -c + dx,$$

поскольку средние значения их левых частей равны нулю.

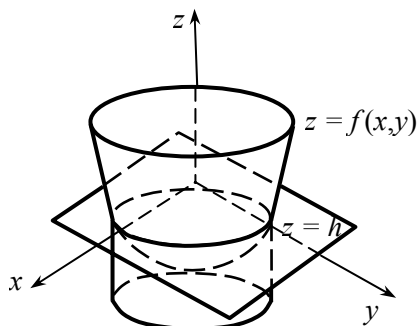


Рис. 9.1. Геометрическое построение фазового портрета модели Вольтерра – Лотки

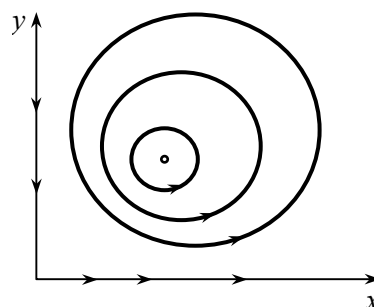


Рис. 9.2. Фазовый портрет модели Вольтерра – Лотки

Согласно этой модели, ни изменения параметров a , b , c и d , ни возмущения, приводящие к изменению начальных условий, до тех пор, пока y не обращается в нуль, не могут нарушить печальную судьбу жертвы быть пищей для хищника. Пожалуй, самое благоприятное для нее – находиться вблизи равновесного состояния, поскольку при больших размахах популяция жертвы, впрочем, как и хищника, проходит периоды очень малой численности.

Если распространить, что допустимо лишь с большими оговорками, эту модель на сосуществование человека (жертва) с микробом или вирусом (хищник), то можно сделать любопытные выводы о том, как следует и как не следует принимать лекарства, уничтожающие хищника. Прежде всего, если это возможно, то хищника нужно добивать полностью: оставляя даже очень малое его количество, мы никакой цели не достигнем; спустя некоторое время он достигнет, возможно, еще большей численности. Если же полное истребление хищника невозможно, то уменьшать его число целесообразно в некоторой мере и только в верхнем над равновесием положении фазовой точки, т.е. вблизи максимальной его численности. Частичное уничтожение при малой численности приведет только к большему последующему обострению заболевания.

Таковы выводы, которые можно сделать из очень простой модели сосуществования хищника и жертвы Вольтерра–Лотки. В дальнейшем эта модель уточнялась и усложнялась. С чисто математической стороны модель Вольтерра–Лотки неудовлетворительна в силу возможности значительных качественных изменений фазового портрета под влиянием сколь угодно малых изменений правых частей ее дифференциальных уравнений. Это означает, что неучет даже незначительных факторов может существенно его изменить, и поэтому нет никаких оснований считать модель правильной, так как игнорирование второстепенных факторов неизбежно.

Такое нежелательное поведение модели Вольтерра–Лотки объясняется тем, что его равновесие типа центр, которому отвечают чисто мнимые корни характе-

ристического уравнения, может при самых незначительных добавках $\delta(x, y)$ и $\varepsilon(x, y)$ к правым частям уравнений (9.1), (9.2) смениться на устойчивый или неустойчивый фокус, а вокруг него могут появиться замкнутые устойчивые периодические движения, что существенно изменяет фазовый портрет и вытекающее из него поведение сосуществующих жертвы и хищника.

Модель, лишенная этого недостатка, учитывающая насыщение хищника и конкуренцию жертвы за источники питания и жизни, имеет вид [5]:

$$\begin{aligned}\dot{x} &= ax - \frac{bxy}{1 + Ax} - Bx^2, \\ \dot{y} &= -cy + \frac{dxy}{1 + Ax}.\end{aligned}\tag{9.4}$$

Она переходит в модель Вольтерра–Лотки при $A = 0$ и $B = 0$. Параметр A учитывает насыщение хищника, что даже при очень большом числе x жертв он не может съесть жертвы больше, чем некоторое количество bA^{-1} , в то время как при $A = 0$ он мог съесть неограниченно много при $x \rightarrow \infty$. Параметр $B > 0$ учитывает ограниченность ресурсов для размножения жертвы.

Исследование дифференциальных уравнений (9.4) уже не так просто, как модели Вольтерра–Лотки. Его целесообразно начать с уменьшения числа параметров. Сейчас их шесть. Оказывается, что можно привести к двум. Замена переменных $t = \tau/\alpha$, $x = c/d u$, $y = a/b v$ приводит дифференциальные уравнения (9.4) к виду:

$$\begin{aligned}\dot{u} &= u - \frac{uv}{1 + \alpha u} - \varepsilon u^2, \\ \dot{v} &= -\gamma v \left(1 - \frac{u}{1 + \alpha u} \right).\end{aligned}\tag{9.5}$$

Как показывает исследование, качественный вид их фазового портрета зависит только от положительных параметров α и ε , параметр γ на него не влияет ($\alpha = Ac/d$, $\varepsilon = cB/ad$).

Отыскание двумерного (переменные $u \geq 0$, $v \geq 0$) фазового портрета этих уравнений и бифуркационного портрета распадается на ряд этапов. Сначала находятся и исследуются состояния равновесия. Это исследование подсказывает возможность автоколебаний. Далее находятся области параметров, отвечающие различным качественным структурам фазового портрета. В результате приходим к бифуркационному портрету (рис. 9.3).

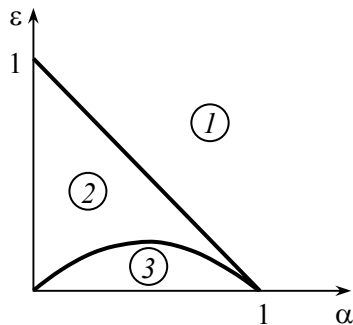


Рис. 9.3. Параметрический портрет уточненной модели «хищник–жертва» Вольтерра – Лотки

Он содержит три области 1, 2 и 3. Каждой из них отвечает свой вид фазового портрета, показанный на рис. 9.4 с указанием номера области бифуркационного портрета.

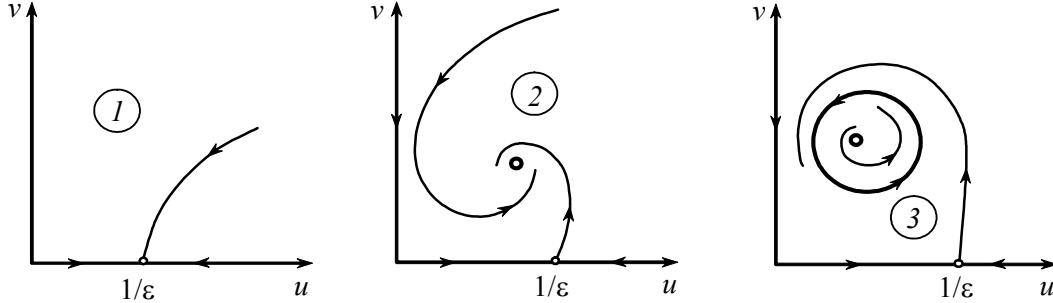


Рис. 9.4. Фазовые портреты уточненной модели «хищник–жертва» для значений параметров из областей 1–3 параметрического портрета

При параметрах из области 1 хищник гибнет, остается одна жертва в количестве $u^* = 1/\varepsilon$, отвечающем устойчивому равновесию фазового портрета 1 рис. 9.4.

В области 2 изолированное существование жертвы становится неустойчивым и появляется равновесное взаимодействие хищника и жертвы.

С переходом в область 3 нарушается равновесное сосуществование жертвы и хищника и оно приобретает автоколебательные характер.

После этого краткого описания перейдем к изложению исследования.

Состояния равновесия находятся из уравнений:

$$u - \frac{uv}{1 + \alpha u} - \varepsilon u^2 = 0, \quad \gamma v \left(1 - \frac{u}{1 + \alpha u} \right) = 0. \quad (9.6)$$

Решений три: 1) $u = v = 0$, 2) $u = 1/\varepsilon, v = 0$, 3) $u = 1/(1 - \alpha), v = 1 - \alpha - \varepsilon/(1 - \alpha)^2$.

Первые два равновесия имеют место при любых положительных α и ε , последнее – только в области 2 и 3, где $1 - \alpha - \varepsilon > 0$. На границе областей 1 и 2 третье и второе равновесия сливаются.

Для исследования устойчивости найденных равновесий составим характеристическое уравнение:

$$X(\lambda) = \begin{vmatrix} 1 - 2\varepsilon u - \frac{v}{1 + \alpha u} + \frac{\alpha uv}{(1 + \alpha u)^2} - \lambda & -\frac{u}{1 + \alpha u} \\ -\gamma v \left[\frac{\alpha u}{(1 + \alpha u)^2} - \frac{1}{1 + \alpha u} \right] & -\gamma \left(1 - \frac{u}{1 + \alpha u} \right) - \lambda \end{vmatrix} = 0,$$

в которое последовательно подставляются вместо u и v значения координат равновесий, исследуемых на тип и устойчивость. Для состояний равновесия 1, 2 и 3 последовательно находим корни этого характеристического уравнения:

- 1) $\lambda_1 = 1, \lambda_2 = -\gamma$;
- 2) $\lambda_1 = -1, \lambda_2 = -\gamma(\alpha + \varepsilon - 1)$;
- 3) $\lambda_{1,2} = -\sigma \pm i\omega$,

где σ обращается в нуль на граничной кривой $\varepsilon = \alpha(1 - \alpha)/(1 + \alpha)$ областей 2 и 3 (в области 2 – значение σ отрицательно, а в области 3 – положительно). В соот-

ветствии с теорией бифуркаций при переходе из области 2 в 3 должно либо родиться устойчивое периодическое движение от третьего состояния равновесия, либо неустойчивое периодическое движение должно исчезать, вливая в него. Выяснить, какой из случаев имеет место, можно, подсчитывая знак так называемой ляпуновской величины. Эта процедура приводит к тому, что имеет место рождение устойчивого периодического движения – автоколебания – и в соответствии с этим имеет место фазовый портрет 3 рис. 9.4.

Справедливости ради укажем, что при строгом подходе нужно еще выяснить, не могут ли возникнуть сложные предельные циклы. Легче всего в этом убедиться с помощью компьютера, с дисплея которого и были срисованы все приведенные картинки фазовых портретов.

Мы рассмотрели математические модели взаимодействующих популяций типа хищник–жертва. Аналогичные модели могут быть построены для двух конкурирующих популяций и помогающих друг другу. Дифференциальные уравнения двух конкурирующих популяций при учете еще и внутренней конкуренции в каждой из них можно записать в виде:

$$\dot{x}_1 = a_1 x_1 - a_{11} x_1^2 - a_{12} x_1 x_2, \quad \dot{x}_2 = a_2 x_2 - a_{22} x_2^2 - a_{21} x_2 x_1$$

или, после замены переменных

$$t = \frac{\tau}{a_1}, \quad x_1 = \frac{a_1}{a_{11}} u_1, \quad x_2 = \frac{a_2}{a_{22}} u_2,$$

в эквивалентном виде:

$$\begin{aligned} \dot{u}_1 &= u_1(1 - u_1 - \varepsilon_1 u_2), \\ \dot{u}_2 &= \gamma u_2(1 - u_2 - \varepsilon_2 u_1). \end{aligned} \tag{9.7}$$

Как и в предыдущей модели, приведенной к такому виду, зависимость от γ несущественная, не отражающаяся на качественном виде фазового портрета. По существенным параметрам ε_1 и ε_2 бифуркационная диаграмма содержит четыре области 1, 2, 3 и 4 (рис. 9.5), каждой из которых отвечает свой вид фазового портрета, показанный на рис. 9.6 под тем же номером, что и соответствующая область бифуркационного портрета.

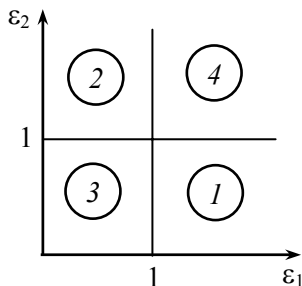


Рис. 9.5. Параметрический портрет модели конкурирующих видов

Общий смысл их состоит в том, что в конкурентной борьбе выживает та популяция, которая в параметрах ε_1 и ε_2 оказывает на другую большее конкурентное давление. Эта ситуация имеет место в областях 1 и 2, где один из параметров ε_1 и ε_2 больше единицы, а другой меньше. В области 3 имеет место равновесное сосуществование. В области 4 выживает одна из популяций в зависимости от начальных условий.

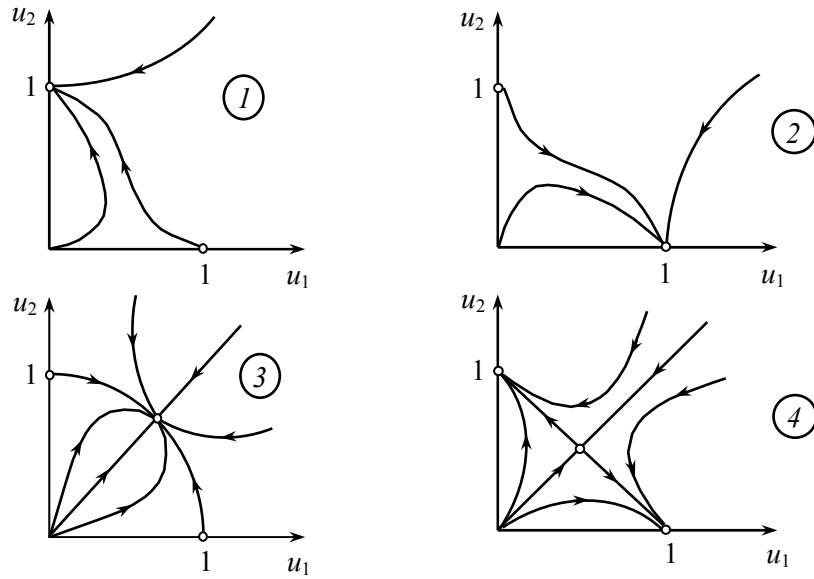


Рис. 9.6. Фазовые портреты, соответствующие параметрическому портрету (рис.9.5) модели конкурирующих видов

Для осмысливания этих выводов нелишне знание связи исходных параметров модели $a_1, a_2, a_{11}, a_{12}, a_{21}, a_{22}$ с редуцированными ε_1 и ε_2 :

$$\varepsilon_1 = \frac{a_2 a_{12}}{a_1 a_{22}}, \quad \varepsilon_2 = \frac{a_1 a_{21}}{a_2 a_{11}}.$$

Следующие две модели соответствуют взаимопомощи популяций, известной в биологии как явление симбиоза. Первая модель относится к случаю, когда взаимопомощь – необходимое условие существования: каждая из популяций сама по себе гибнет. В ней учитываются еще внутренние конкуренции в каждой из популяций. В исходных параметрах эта модель имеет вид:

$$\begin{aligned} \dot{x}_1 &= -a_1 x_1 + \frac{b x_1 x_2}{1 + A_1 x_2} - c_1 x_1^2, \\ \dot{x}_2 &= -a_2 x_2 + \frac{b x_1 x_2}{1 + A_2 x_1} - c_2 x_2^2. \end{aligned} \tag{9.8}$$

Как и ранее, все параметры неотрицательные. Здесь возможно только два вида фазового портрета: один, когда при любых начальных условиях численности обеих популяций x_1 и x_2 стремятся к нулю и популяции вымирают, и второй, отвечающий рис. 9.7.

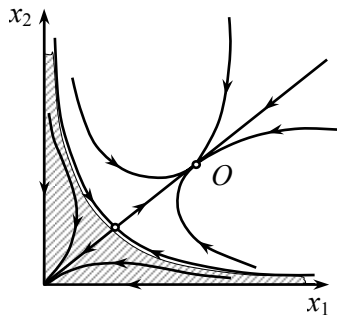


Рис. 9.7. Фазовый портрет модели (9.8) симбиоза двух видов

В заштрихованной части фазового пространства все фазовые траектории стремятся к началу координат, что означает гибель обеих популяций. При больших значениях численности x_1 и x_2 популяций, отвечающих незаштрихованной области, устанавливается равновесное взаимовыгодное сосуществование. Это равновесное состояние отмечено на фазовом портрете буквой O .

Значительное разнообразие динамики симбиоза двух популяций наблюдается в модели, где отдельное существование каждой из популяций возможно, но требует, чтобы ее численность превзошла некоторый порог: малочисленная популяция вымирает. Такая особенность биологического поведения изолированной популяции может быть отображена моделью вида:

$$\dot{x} = -ax(x - L_1)(x - L_2) \quad (0 < L_1 < L_2), \quad (9.9)$$

как это видно из рис. 9.8, на котором изображен график правой части этого уравнения и соответствующий ему фазовый портрет.

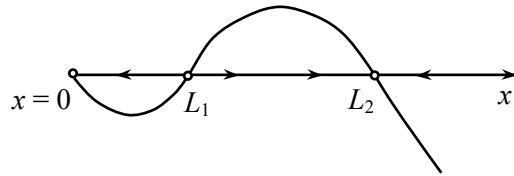


Рис. 9.8. Фазовый одномерный портрет существования одного вида модели (9.9)

В соответствии с этим математическая модель симбиоза двух популяций записывается в виде:

$$\begin{aligned} \dot{x}_1 &= a_1 x_1 (x_1 - L_1)(L_2 - x_1) + b_1 x_1 x_2, \\ \dot{x}_2 &= a_2 x_2 (x_2 - K_1)(K_2 - x_2) + b_2 x_1 x_2. \end{aligned} \quad (9.10)$$

Один из возможных здесь фазовых портретов, который, как и другие, можно наблюдать на дисплее персональной ЭВМ, изображен на рис. 9.9. В соответствии с поведением фазовых траекторий на рис. 9.9 в описываемом симбиозе может произойти вымирание обеих популяций, любая из популяций может выжить, а другая погибнуть, и, наконец, они могут сосуществовать при больших численностях, чем каждая отдельно.

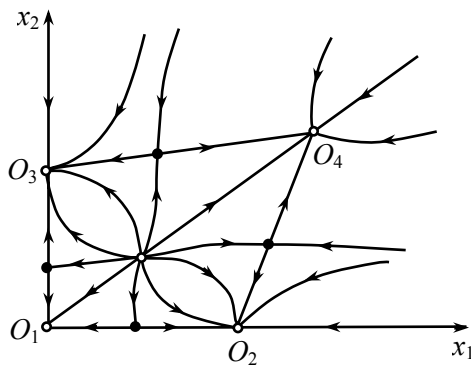


Рис. 9.9. Фазовый портрет симбиоза двух видов для модели (9.10)

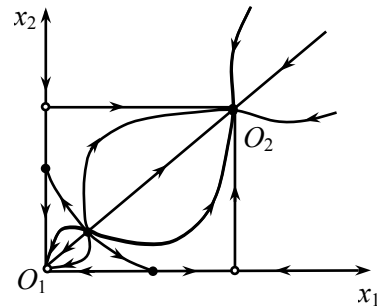


Рис. 9.10. Другой фазовый портрет симбиоза двух видов для модели (9.10)

Другой возможный случай изображен на рис. 9.10. В нем, как и в предыдущей модели симбиоза (9.8), возможно только совместное существование обеих популяций или их обоюдная гибель.

Мы рассмотрели несколько разных математических моделей сосуществования двух популяций. Соображения, положенные в их основу, естественны и просты, но выводы неожиданны и поучительны.

Можно ли было ожидать, что сосуществование хищника и жертвы ведет к автоколебаниям? Можно ли было думать, что конкуренция жертв способствует ее избавлению от хищника, что при сосуществовании конкурирующих популяций та же внутренняя конкуренция, которая еще более затрудняет жизнь, способствует выживанию? Менее неожидан, но также интересен вывод о том, что при симбиозе популяции выживают только при достаточно больших численностях.

10. ПРОТОЧНЫЙ БИОЛОГИЧЕСКИЙ РЕАКТОР

Проточный биологический реактор – хемостат – представляет собою емкость с питательной средой, в которой размножаются микробы [18]. Его особенность состоит в том, что среда, в которой происходит непрерывное размножение, все время пополняется свежим интенсивным питающим раствором и одновременно с той же скоростью отбирается жидкость с микроорганизмами и субстратом. Среда хемостата интенсивно размешивается, так что во всем его объеме концентрации интенсивного субстрата и микробов одинаковы. Пополнение и отбор происходят с одинаковыми объемными скоростями. В хемостате поддерживается постоянная температура.

Наша задача состоит в том, чтобы, построив математическую модель динамики хемостата, изучить ее и оптимизировать его работу по управляемым параметрам. Размножение микробов в питательной среде хемостата подобно размножению хищника с учетом его насыщения. В соответствии с этим скорость размножения примем пропорциональной массе микробов x с коэффициентом пропорциональности вида

$$\mu = \frac{\mu_m S}{K + S}, \quad (10.1)$$

где S – концентрация субстрата в хемостате, а μ_m и K – параметры. Пусть поступающий в хемостат питательный субстрат концентрации S_0 и скорость объемного поступления его в хемостат равна D . Ясно, что убыль интенсивного вещества пропорциональна скорости размножения микробов. Поступление интенсивного раствора и отбор содержимого хемостата происходят с постоянной объемной скоростью D .

Согласно сказанному, используя (10.1), найдем, что

$$\begin{aligned} \dot{x} &= \frac{\mu_m S}{K + S} x - \frac{D}{V} x, \\ V \dot{S} &= D(S_0 - S) - \frac{l \mu_m S}{K + S} x, \end{aligned}$$

где V – объем жидкости в хемостате. Введем новые параметры: D вместо D/V и l вместо l/V , запишем эти уравнения в виде:

$$\begin{aligned} \dot{x} &= \frac{\mu_m S}{K + S} x - D x, \\ \dot{S} &= D(S_0 - S) - \frac{l \mu_m S}{K + S} x \end{aligned} \quad (10.2)$$

и приступим к их изучению.

Фазовым пространством хемостата, точнее, его математической модели (10.2), является первый октант $x \geq 0$, $S \geq 0$ плоскости x , S . Помимо тривиального равновесия $x = 0$, $S = S_0$, имеется еще одно равновесие, определяемое из уравнений:

$$\frac{\mu_m S}{K + S} - D = 0, \quad D(S_0 - S) - \frac{l \mu_m S}{K + S} x = 0.$$

Непосредственно находим, что его координаты x^* и S^* равны

$$S^* = \frac{KD}{\mu_m - D}, \quad x^* = [(\mu_m - D)S_0 - KD]l^{-1}(\mu_m - D)^{-1}. \quad (10.3)$$

Этим значениям S^* и x^* будет отвечать равновесие, если они не отрицательные, т.е. при выполнении неравенства

$$\mu_m > D + \frac{KD}{S_0}, \quad (10.4)$$

т.е. если точка (S_0, D) лежит в области, изображенной на рис. 10.1.

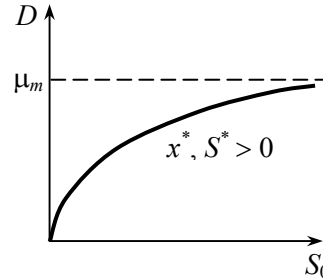


Рис. 10.1. Область существования равновесия (x^*, S^*) хемостата по параметрам S_0 и D

Устойчивость равновесий определяется корнями характеристического уравнения:

$$X(\lambda) = \begin{vmatrix} \mu - D - \lambda & \frac{\mu_m K}{(K + S)^2} x \\ -l\mu & -D - \frac{\mu_m Kl}{(K + S)^2} x - \lambda \end{vmatrix} = 0, \quad (10.5)$$

где использовано обозначение (10.1), а S и x – значения координат рассматриваемого равновесия. Для равновесия $x = 0, S = S_0$ корни действительные разных знаков, и для равновесия $x = x^*, S = S^*$, определяемого (10.3), коэффициенты квадратного уравнения относительно λ (10.5) положительны и, следовательно, оно устойчиво. Соответствующий фазовый портрет представлен на рис. 10.2. Он очень прост: любая фазовая точка (x, S) ($x > 0$) асимптотически приближается к равновесию с координатами S^*, x^* .

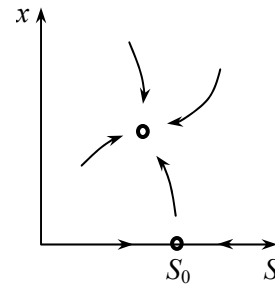


Рис. 10.2. Фазовый портрет хемостата при значениях параметров S_0 и μ_m из области существования равновесия, показанной на рис. 10.1

Параметрами хемостата, согласно (10.2), являются μ_m, K, D, S_0, l , причем величины параметров D и S_0 можно менять, они управляемые. Менять их следует так, чтобы не нарушить устойчивую работу хемостата, т.е. соблюдая условия (10.4), и желательно так, чтобы максимизировать Dx^* . Из (10.3) следует, что

$$Dx^* = D[(\mu_m - D)S_0 - KD]l^{-1}(\mu_m - D)^{-1} = \frac{DS_0}{l} - \frac{KD^2}{l(\mu_m - D)}. \quad (10.6)$$

При выбранном S_0 максимум по D достигается при некотором $D(S_0) < \mu_m$, причем $D(S_0)$ – возрастающая функция S_0 . Таким образом, S_0 желательно выбирать побольше и при выбранном S_0 существует оптимальная скорость отбора D . На рис. 10.3 изображен поясняющий этот факт график зависимости скорости производства микроорганизмов Dx^* от скорости отбора D .

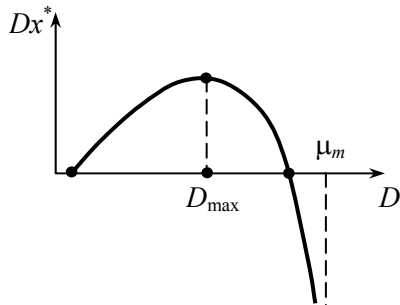


Рис. 10.3. График производительности хемостата от скорости отбора D . Четко видно наличие максимума

Точка пересечения этого графика с осью абсцисс имеет координату $\mu_m S_0(K + S_0)^{-1}$ и поэтому после этой точки график теряет физический смысл, поскольку нарушается неравенство (10.4), обеспечивающее функционирование хемостата.

11. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ИММУННОГО ОТВЕТА ОРГАНИЗМА НА ВТОРЖЕНИЕ ИНФЕКЦИИ

Иммунные системы животного и человека очень сложны. Сложно и многообразно их устройство, сложны и многовариантны их реакции на инфекцию, сложны и многолики средства противодействия. Сегодня описание иммунной системы занимает пару томов, и при этом еще многое остается незатронутым и неизвестным. Вместе с тем, иммунная система, сколь бы сложна она ни была, предназначена оказывать противодействие и уничтожать вторгшуюся в организм чужеродную и вредную для него инфекцию, и в общих понятиях это противодействие инфекции и ее вредное влияние на организм можно описать, опуская детали того, как это фактически происходит.

Есть инфекция, которая стремится размножиться и воздействует на организм, подавляя его жизнеспособность. Есть организм, который пытается ее истребить, организуя, кроме изначального противодействия, еще и дополнительное, как только ее обнаружит. Быстрота и эффективность ответа организма зависят от наследственных факторов и от приобретенного «жизненного опыта», а также от состояния организма в данный момент и резервов, которыми он располагает.

Таким образом, несмотря на всю сложность иммунного ответа организма на инфекцию, на первый план выступают три фактора: инфекция; ее размножение и влияние на организм; противодействие организма и возможности его реализации, зависящие от его состояния. Эти факторы можно количественно охарактеризовать тремя величинами: x , y и z – численностью инфекции, величиной противодействия организма и потенциалом организма. Инфекция, помимо своей численности x , характеризуется быстротой размножения в среде организма и подавляющим действием на организм. Противодействие характеризуется своим количеством y и его эффективностью, быстротой и мощностью пополнения w , зависящими от организма и его потенциала z .

Все сказанное позволяет описать динамику иммунного ответа – изменение величин x , y , w и z – следующими четырьмя дифференциальными уравнениями:

$$\begin{aligned} \dot{x} &= \lambda x - \frac{\alpha xy}{1 + \alpha x} - \bar{\epsilon} x^2; \\ \dot{y} &= \begin{cases} -\frac{\bar{b}xy}{1 + \alpha x} + w = K & \text{при } y > 0 \text{ или } y = 0 \text{ и } K \geq 0, \\ 0 & \text{при } K < 0; \end{cases} \\ \tau \dot{w} + w &= \begin{cases} 0 & \text{при } x \leq \bar{x}_0, \\ \bar{B}z(2z_0 - z)(x + \bar{\beta} x^2) & \text{при } x > \bar{x}_0; \end{cases} \\ \dot{z} &= \begin{cases} \frac{c(z_0 - z)}{1 + \gamma x} - dy - e = F & \text{при } z > 0 \text{ или } z = 0 \text{ и } F \geq 0, \\ 0 & \text{при } z = 0 \text{ и } F < 0. \end{cases} \end{aligned} \quad (11.1)$$

В этих уравнениях \bar{x}_0 – порог чувствительности организма к инфекции, τ – временная задержка дополнительного иммунного ответа, величина которого зависит от объема инфекции x и потенциала организма z . При $\bar{\beta} = 0$ зависимость от x линейная, при $\bar{\beta} > 0$ она форсирована, а при $\bar{\beta} < 0$, напротив, уменьшена. Последнее уравнение системы (11.1) описывает изменение потенциала z , при этом $0 \leq z \leq z_0$. Первый член последнего уравнения (11.1) – это скорость пополнения потенциала z до его предельного значения z_0 , второй и третий – его расход на иммунный ответ организма y и на потребности самого организма. Обращение z в нуль трактуется как полное истощение организма, ведущее к его гибели.

У четырех уравнений (11.1) 13 параметров. Параметр λ может иметь любой знак: $\lambda < 0$ означает, что инфекция, попав в среду организма, гибнет; при $\lambda > 0$ она все же размножается, но ей мешает в этом противодействие организма y , которого может и не быть. В дальнейшем будем считать $\lambda > 0$. Все остальные параметры, кроме $\bar{\beta}$, положительны. Процесс нарастания и динамики противодействия y инфекции x описывается последними тремя уравнениями.

Хотя сказанного достаточно для понимания предлагаемой модели (11.1), все же дадим более подробное описание каждого из ее уравнений.

В первом уравнении член λx соответствует размножению инфекции, второй – ее уничтожению фактором y противодействия организма. Этот член, естественно, равен нулю при $x = 0$, максимален при $x = \infty$ и линейно зависит от y . Последний член отражает возможность конкуренции за условия размножения.

Второе уравнение описывает изменение противодействия инфекции y . Первый член во втором уравнении при $y > 0$ соответствует тому, что на каждую единицу уничтоженной инфекции x происходит определенное уменьшение y . Второй член w в этом же уравнении – это скорость пополнения организмом противодействия y .

Третье уравнение указывает, что иммунный ответ на инфекцию имеет место только при $x > x_0$, происходит с временной задержкой и определяется величиной, зависящей от потенциала организма z , причем пропорционально x или с некоторой форсировкой при $\bar{\beta} > 0$, или, наоборот, уменьшением с ростом x при $\bar{\beta} < 0$.

Последнее уравнение описывает изменение потенциала z организма. В этом уравнении $0 \leq z < z_0$. Первый член правой части увеличивает z при $z < z_0$. Скорость этого роста потенциала убывает с увеличением вредного влияния инфекции x . Второй и третий члены – затраты на поддержание противодействия y и всего организма.

Модель (11.1) для исследования достаточно сложна, но ее можно значительно упростить, ограничившись переменными x и y . Это тем более возможно, что роль остальных уравнений довольно очевидна и может быть мысленно учтена и понята. Упрощенная модель имеет вид:

$$\dot{x} = \lambda x - \frac{axy}{1 + \alpha x} - \bar{\epsilon}x^2,$$

$$\dot{y} = \begin{cases} -\frac{bxy}{1+\alpha x} & \text{при } x \leq \bar{x}_0 \text{ и } y \geq 0, \\ -\frac{bxy}{1+\alpha x} + \bar{B}(x + \beta x^2) = K & \text{при } x > \bar{x}_0 \text{ и } K \geq 0, \\ 0 & \text{при } x > \bar{x}_0 \text{ и } K < 0. \end{cases} \quad (11.2)$$

Отличие уравнений (11.2) от (11.1) в том, что величина \bar{B} теперь постоянна, в то время как в уравнениях (11.1) она меняется согласно двум последним уравнениям. Формально перейти от системы (11.1) к (11.2) можно, положив $\tau = 0$ и приняв z постоянным. Величина $\tau > 0$ определяет быстроту возрастания w от нуля до предельного значения $\bar{B} z (2z_0 - z)(x + \beta x^2)$, а \bar{B} в уравнениях (11.2) равно $\bar{B} z (2z_0 - z)$ при принимаемом постоянном z . Так что роль последних двух уравнений в (11.1) может быть учтена, как изменяемость \bar{B} в уравнениях (11.2).

Число параметров в упрощенных уравнениях (11.2) может быть уменьшено путем изменения масштабов по времени t и переменным x и y , после чего они принимают вид:

$$\dot{x} = x - \frac{xy}{1+x} - \varepsilon x^2,$$

$$\dot{y} = \begin{cases} -\frac{bxy}{1+x} & \text{при } x \leq \bar{x}_0 \text{ и } y > 0, \\ -\frac{bxy}{1+x} + B(x + \beta x^2) = K & \text{при } x > \bar{x}_0 \text{ и } y > 0 \text{ или } y = 0 \text{ и } K \geq 0, \\ 0 & \text{при } x > \bar{x}_0 \text{ и } K < 0. \end{cases} \quad (11.3)$$

Поймем, что следует из упрощенной модели (11.3), сначала приняв $\varepsilon = \beta = 0$. При этом остается три параметра: x_0 , b и B . Как оказывается, возможны два существенно разных случая: $b > B$ и $b < B$. Второй случай, $b < B$, распадается еще на два: $b < 1$ и $b > 1$. Так что имеют место четыре разных вида фазовых портретов, представленных на рис. 11.1.

На рис. 11.1 показана A – область начальных условий, для которых имеет место выздоровление. Вне области A инфекция сохраняется на уровне чувствительности x_0 и имеет место бациллоносительство.

Как видно из рис. 11.1 в случае $b > B$ организм избегает гибели только при начальных условиях в области A , т.е. при достаточно высоком начальном уровне y . При недостаточном начальном противодействии y организм гибнет (происходит неограниченное возрастание x , ведущее, согласно последнему уравнению (11.1), к занулению потенциала z). Это происходит даже при самой незначительной начальной инфицируемости. Напротив, при $b < B$ организм никогда не гибнет. Однако исход встречи с инфекцией может кончиться либо полным ее уничтожением (область A и приход к одному из равновесий $x = 0, y > y^*$), либо сохранением инфекции на пороге чувствительности (весь положительный октант, кро-

го нами уменьшения потенциала z , влекущего уменьшение величины B в уравнениях (11.3). И еще необходимо отметить, что на самом деле, согласно полной модели (11.1), вначале, в момент заражения, $B = 0$, и лишь затем достигает своего, рассматриваемого нами значения, причем рост B в течение времени порядка τ происходит только после того, как x превзойдет x_0 .

Приведенные рисунки фазовых портретов указывают на желательность привнесенного извне увеличения y , стимуляции иммунной системы. Это вообще увыстряет выздоровление и позволяет избежать перехода в хроническую инфицированность. Это будет видно и из дальнейшего, когда выявится роль β . Но, пожалуй, самый существенный вывод, кроме крайней желательности обеспечения $b < B$, тот, что целесообразно поддерживать изначальное значение y больше, чем y^* (см. рис. 11.1).

Действительно, если в начальный момент, до заражения, иммунная система находится в равновесии, $x = 0, y \geq 0$, то при $y < y^*$, в силу неустойчивости равновесия, даже незначительное увеличение x приводит к заболеванию, в то время как при $y > y^*$ происходит уничтожение инфекции и наступление нового близкого равновесного состояния. Уменьшение y может произойти в силу отсутствия асимптотической устойчивости состояний равновесий $x = 0, y > y^*$. Напротив, при сожительстве с инфекцией состояние равновесия организма O асимптотически устойчиво, и в этом смысле оно оберегает организм от вторичного заражения. Конечно, это только при стабильности параметров b и B , при их изменении эта инфекция, наоборот, может быть причиной повторного заболевания.

Теперь посмотрим, что нового вносят параметры β и ε , не учитываемые ранее ($\varepsilon = \beta = 0$). Ограничимся иллюстрацией ряда фазовых портретов, срисованных с дисплея ЭВМ. Фазовые портреты, представленные на рис. 11.2, 11.3 и 11.4, демонстрируют роль форсированного иммунного ответа $\beta > 0$ на весьма неблагоприятном случае $b > B$.

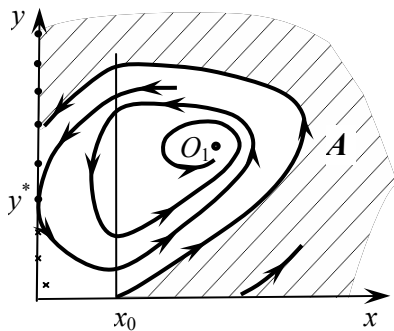


Рис. 11.2. Вид фазового портрета упрощенной модели (11.3) при $x_0=1, b=1,1, B=1,0, \varepsilon=0$ и $\beta=0,2$. Состояние равновесия O_1 устойчиво и соответствует хроническому заболеванию. В области A – выздоровление

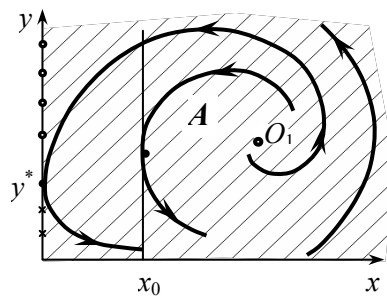


Рис. 11.3. Вид фазового портрета упрощенной модели (11.3) при $x_0=1, b=0,8, B=0,5, \varepsilon=0$ и $\beta=0,2$. Состояние равновесия O_1 неустойчиво и при любых начальных условиях – выздоровление

При этом все кончается хроническим заболеванием, бациллоносительством, или выздоровлением. Конечно, длительное хроническое заболевание может навлечь существенное уменьшение потенциала z и следующее за ним уменьшение величины B , что может привести к фатальным последствиям.

На рис. 11.5 представлен другой вариант форсирования иммунного ответа ($\beta > 0$), при котором возможно течение болезни с периодически повторяющимися рецидивами (автоколебание Γ).

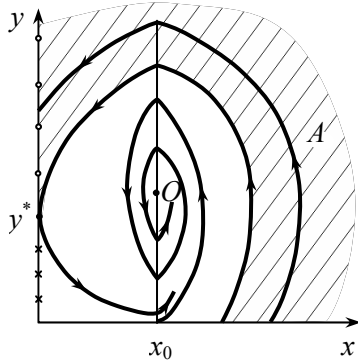


Рис. 11.4. Вид фазового портрета упрощенной модели (11.3) при $x_0=1$, $b=1$, $B=0,9$, $\varepsilon=0,01$ и $\beta=0,05$. В области A – выздоровление, вне A – бациллоносительство

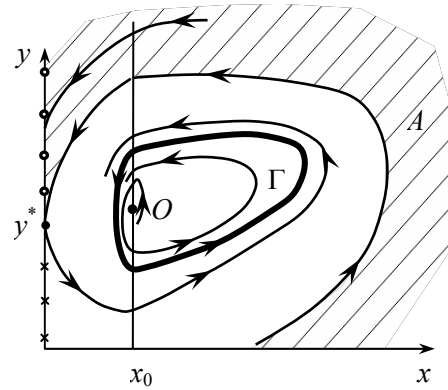


Рис. 11.5. Вид фазового портрета упрощенной модели (11.3) при $x_0=1$, $b=0,9$, $B=0,9$, $\varepsilon=0,01$ и $\beta=0,05$. В области A – выздоровление, вне A – периодическое течение болезни (автоколебание Γ)

Что может произойти при ослаблении иммунного ответа ($\beta < 0$), демонстрируют рис. 11.6 и 11.7: несмотря на то, что $B > b$, возможен летальный исход. К ослаблению иммунного ответа может привести не только $\beta < 0$, но и уменьшение потенциала z согласно последнему, не учитываемому упрощенной системой (11.3), уравнению системы (11.1).

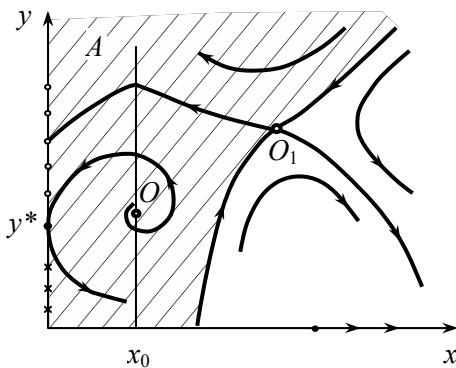


Рис. 11.6. Вид фазового портрета упрощенной модели (11.3) при $x_0=1$, $b=0,5$, $B=2$, $\varepsilon=0$ и $\beta=-0,1$. В области A – выздоровление, вне A – гибель

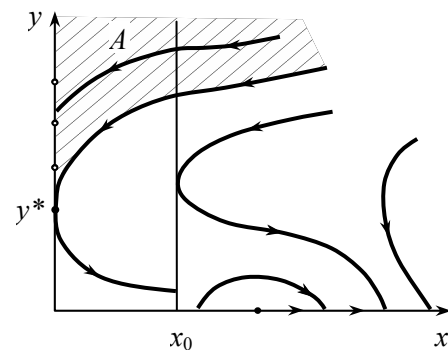


Рис. 11.7. Вид фазового портрета упрощенной модели (11.3) при $x_0=1$, $b=0,8$, $B=1$, $\varepsilon=0$ и $\beta=-0,2$. В области A – выздоровление, вне A – гибель

12. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СООБЩЕСТВА «ПРОИЗВОДИТЕЛИ – ПРОДУКТ – УПРАВЛЕНЦЫ»

Человеческое общество необычайно сложно, каждый человек – целый мир с многообразным непредсказуемым поведением, а людей – тысячи и миллионы, и они находятся в сложнейших взаимоотношениях друг с другом. Кроме того, есть государство, системы управления и распределения, финансы, промышленность, бытовые учреждения, транспорт, связь, учебные заведения, медицинские учреждения и т.д. Кажется, что это все невозможно описать математически, да это так и есть. Но можно попытаться описать только некоторые стороны этого неопишуемого сложного человеческого сообщества, опираясь на какие-то присущие ему особенности. Например, для того чтобы жить, необходимо добывать пищу, чтобы ее эффективнее добывать, нужна организованность; да и не хлебом единым жив человек: ему нужно жилище, семья, орудия производства, зрелища и удовольствия. При реализации своих естественных потребностей он вступает в конфликт с другими членами сообщества, имеющими свои потребности, и вместе с тем члены сообщества объединяются, чтобы эффективнее добывать то, что они хотят потреблять. Выделим во всем этом три основных компонента: производители, продукт, управленцы. Производители – это те, которые непосредственно производят продукт, управленцы не производят продукт, но способствуют его производству; продукт – это все то, что нужно для жизни человека, что он потребляет и чем он пользуется.

Придадим производителям, управленцам и продукту количественную определенность в виде величин x , y и z , означающих численности производителей, управленцев и наличное количество используемого сообществом продукта. Производители и управленцы находятся в конкурентных взаимоотношениях как между собой, так и друг с другом, и вместе с тем, производители совместно производят продукт, а управленцы могут способствовать этому производству. Управленцы оказывают влияние на жизнь и производство продуктов, управление сообществом, но непосредственными их кормильцами, как и себя самих, являются производители. Итак, есть конкурентные взаимоотношения, есть естественное размножение и есть совместное производство продукта, эффективность которого определяется уровнем технологии сообщества.

На рис. 12.1 указаны связи и взаимодействия между производителями x , управленцами y и продуктом z .

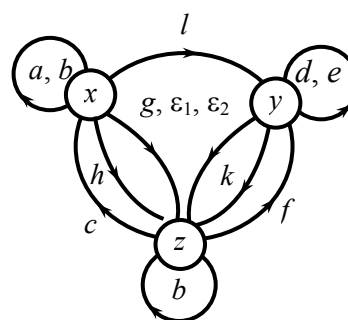


Рис. 12.1 Схема влияний в модели
«производители – продукт – управленцы»

Каждая стрелка этой схемы отвечает воздействию переменной, от которой она идет, на изменение той, к которой она ведет. Так, стрелка от переменной x к переменной z означает, что скорость изменения количества продукта z зависит от переменной x – числа производителей. Таких стрелок две: одна отвечает тому, что x производит продукт z , а такая же другая – что он его потребляет. Стрелка от z к x означает, что скорость изменения x зависит от наличного продукта z . Все стрелки, идущие от x , y и z к z , указывают, что скорость изменения накопленного продукта z зависит от x , y и z .

Эта схема не раскрывает конкретного характера зависимостей и взаимных связей, она лишь указывает на их наличие и то, что мы хотим в нашей модели их учесть. Ясно, что сколько-нибудь точный учет их очень труден, да и при агрегированности переменных x , y и z невозможен. Ограничимся грубо приближенным, качественным их учетом, описываемым следующей системой трех дифференциальных уравнений:

$$\begin{aligned} \dot{x} &= (a - bx - ly + cz)x, \\ \dot{y} &= (-d - ey + fz)y, \\ \dot{z} &= \begin{cases} F = g \frac{1 + \varepsilon_1 y}{1 + \varepsilon_2 y} \frac{x}{1 + \delta z} - hx - ky & \text{при } z > 0 \text{ или } z = 0 \text{ и } F > 0, \\ 0 & \text{при } z = 0 \text{ и } F \leq 0, \end{cases} \end{aligned} \quad (12.1)$$

которая является предлагаемой математической моделью сообщества «производители – продукт – управленцы». В этой модели очень многое не учтено. Посмотрим, к какой динамике модели это приводит. Опишем еще раз, более детально, общие соображения, лежащие в ее основе. Принимается, что скорость изменения x в целом пропорциональна x с коэффициентом пропорциональности $a - bx - ly + cz$, убывающим с ростом x и y и возрастающим с ростом z . Это призвано отразить внутреннюю конкуренцию среди производителей x , давление на них управленцев y и способствование производителям накопленного продукта z . Без учета этих влияний x , y и z на изменение x первое уравнение системы (12.1) приняло бы вид

$$\dot{x} = ax,$$

отвечающий беспрепятственному размножению производителей.

Соображения, положенные в основу второго уравнения системы (12.1), такие же, но принято, что давление производителей на управленцев отсутствует или им можно пренебречь. Возможно, это не всегда так и не представляет труда это учесть. Если производители еще как-то, возможно, могут существовать сами по себе при $z = 0$, то в отношении управленцев это исключено и поэтому знак перед коэффициентом d отрицательный, в соответствии с чем «изолированные» управленцы подчиняются дифференциальному управлению

$$\dot{y} = -dy,$$

отвечающему их гибели.

Третье уравнение отражает процесс накопления продукта в результате его производства и потребления. Потребляют продукт производители и управленцы. Он амортизируется и разрушается сам по себе и производится совместными уси-

лиями производителей и управленцев. Характер участия производителей и управленцев в производстве продукта разный: производители непосредственно его производят, а управленцы способствуют (или тормозят) его производству. В соответствии с этим, правая часть дифференциального уравнения для z содержит три слагаемых: первое отражает скорость производства продукта с учетом увеличивающейся трудности его производства, с ростом его объема и учетом его «старения» (амортизации), второе и третье – его потребление производителями и, соответственно, управленцами. Потребления продукта производителями и управленцами принимаются пропорциональными с разными коэффициентами их численности x и y . С производством продукта сложнее. Принято, что скорость его производства пропорциональна численности производителей x с коэффициентом пропорциональности, в который входят сомножителями уровень технологии сообщества g , величина $(1 + \varepsilon_1 y)(1 + \varepsilon_2 z)^{-1}$, описывающая влияние на производство управленцев y , и множитель $(1 + \delta z)^{-1}$, убывающий с ростом z , призванный учесть амортизацию накопленного продукта и увеличение трудности его накопления с возрастанием его объема. При $\varepsilon_1/\varepsilon_2 > 1$ управленцы способствуют в пределах от 1 до $\varepsilon_1/\varepsilon_2$ производству, напротив, при $\varepsilon_1/\varepsilon_2 < 1$ они его тормозят в тех же пределах.

В принятой модели понятия продукта, производителей и управленцев агрегированные, не учитывающие их, возможно, очень сложной и многообразной дифференциации. Для первобытного общества продукт – это прежде всего продукты питания и, возможно, жилище; для более развитого общества – еще и орудия охоты, скотоводства и земледелия, включая сам скот и землю; для современного общества – это вся совокупность материальных и интеллектуальных богатств общества. Трактовка переменных x и y усредненная, обобщенная, ассоциированная с производственной и общественной активностью, отражаемой в производственных и конкурентных взаимоотношениях.

Такие обобщенные, усредненные и расплывчатые трактовки переменных x , y и z , естественно, вызывают чувство неудовлетворенности. Много еще хочется учесть, уточнить, дифференцировать, но вместе с уточнениями модель неизбежно потеряет не только свою простоту, но и общность. Поэтому не будем излишне требовательны. Ведь все же что-то, и, возможно, самое главное, она учитывает. Посмотрим, что из нее следует, и тогда составим свое мнение.

Модель определяется большим числом параметров, эти параметры вовсе не постоянные и не всегда меняются достаточно медленно, иногда даже очень быстро. Они сложно зависят от текущего и прошлого состояния сообщества. Это в реальности, а в модели они принимаются постоянными. Анализируя модель, можно выяснить, что происходит при тех или иных фиксированных значениях параметров и как поведение модели меняется с их изменением.

То, как текущие значения x , y и z влияют на изменение параметров модели, зависит от того, понимают ли ее современники закономерности динамики сообщества, в котором живут, и каковы их фактические представления о ней, какова управленческая структура общества, каковы нравы и воспитание его членов, каковы верования, мораль и этика? Ясно, что на характер зависимости параметров могут существенно повлиять и сведения о сообществе, полученные в результате изучения ее математической модели. Все это учесть очень трудно, поэтому огра-

начимся изучением модели при различных постоянных значениях ее параметров, а остальное предоставим нашему интеллекту.

Изучение модели, несмотря на ее простоту, – непростая задача: она трехмерная и в ней много параметров. Качественное изучение ее динамики сводится к построению фазовых и бифуркационных портретов. Фазовый портрет трехмерен и может быть нарисован. С бифуркационным портретом дело сложнее и можно лишь получить некоторые общие о нем представления. Даже после обычного сокращения числа параметров их остается не меньше девяти, так что удобнее сохранить тринадцать исходных.

Выяснение возможных видов фазовых портретов и их зависимости от параметров в основном осуществлено старомодным трудоемким аналитическим исследованием. Некоторая поддержка получена наблюдением фазовых траекторий на дисплее персональной ЭВМ.

Обнаружить удалось следующее. Одним из основных параметров, определяющих структуру сообщества, является его уровень технологии g . В зависимости от его величины сообщества можно подразделить на три типа: с низким уровнем технологии, средним и высоким. Низкий уровень – это когда $g < h$, средний – при $h < g < h(1 + \delta d/f)$, и высокий – когда $g > h(1 + \delta d/f)$.

Фазовые портреты сообществ с низким и средним уровнями технологии зависят еще от величин ε_1 и ε_2 , определяющих влияние управленцев на эффективность производства. При $\varepsilon_1 < K(\varepsilon_2)$, где $K(\varepsilon_2)$ зависит, конечно, и от остальных параметров, имеют место фазовые портреты рис. 12.2 и 12.3. Первый из них отвечает сообществу с низким уровнем производства, когда в нем остаются одни производители без накопленного продукта и управленцев: что добыл, то и съел. Второй – сообществу со средним уровнем производства, в котором устанавливается общество производителей, имеющих накопленный продукт. Состояния, при которых имеются управленцы, меняются так, что они со временем исчезают, а для сообщества с низким уровнем исчезает и накопленный продукт.

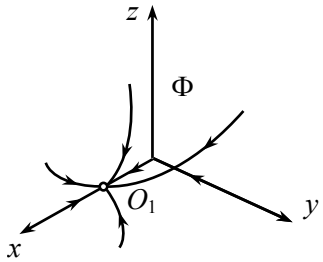


Рис. 12.2. Фазовый портрет модели «производители–продукт–управленцы» при $g < h$, когда в глобально устойчивом равновесии O_1 нет продукта и управленцев

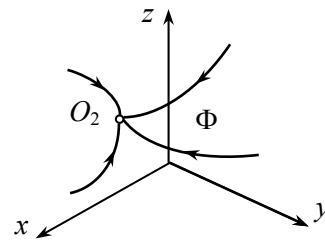


Рис. 12.3. Фазовый портрет модели «производители–продукт–управленцы» при $h < g < h(1 + \delta d/f)$, когда в глобально устойчивом равновесии O_2 есть производители и продукт, но нет управленцев

Если управленцы могут значительно повысить эффективность производства, так что $\varepsilon_1 > K(\varepsilon_2)$, то это приводит к измененным фазовым портретам, показанным на рис. 12.4,а, 12.5,а. В этом случае, наряду с сохранением прежних устойчивых сообществ «производители» или «производители–продукт», возможно воз-

никновение сообществ с накопленным продуктом и управленцами, т.е. сообщества «производители–продукт–управленцы».

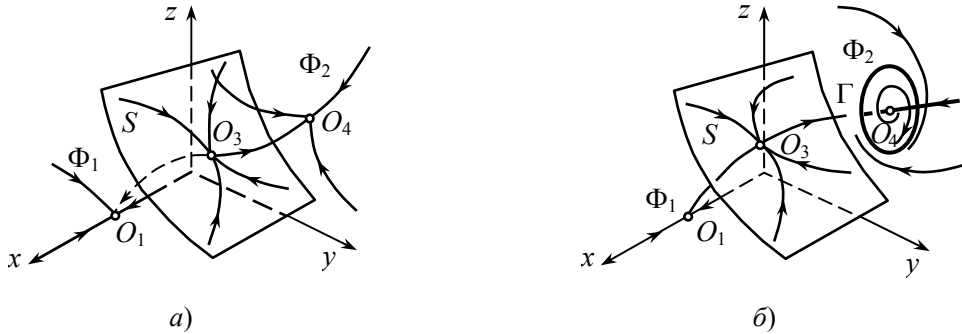


Рис.12.4. Фазовые портреты модели «производители–продукт–управленцы»:

- а) два устойчивых равновесия O_1 и O_4 , области притяжения Φ_1 и Φ_2 которых разделены сепаратрисной поверхностью S седлового равновесия O_3 ;
- б) одно равновесие O_1 устойчиво, второе – O_4 , став неустойчивым, породило устойчивое периодическое движение Γ , области их (O_1 и Γ) притяжения Φ_1 и Φ_2 разделены сепаратрисной поверхностью S седлового равновесия O_3

Сообщества «производители – продукт – управленцы» могут быть стабильными и нестабильными. Все это приводит к четырем видам фазовых портретов, изображенных на рис. 12.4 и рис. 12.5. Стабильность имеет место при $\Delta > 0$, а нестабильность – при $\Delta < 0$ (см. (12.9)).

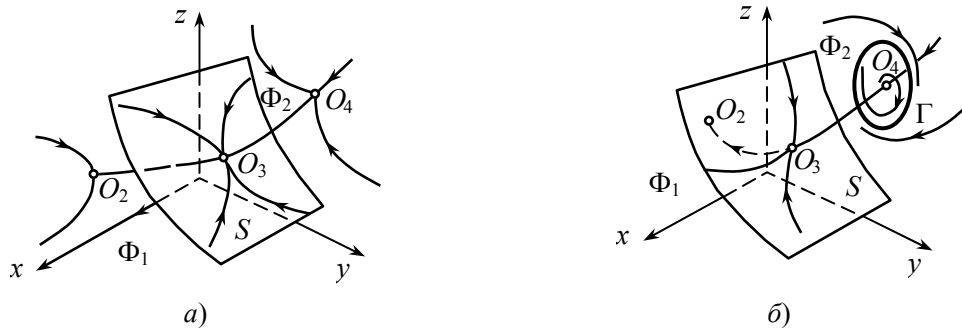


Рис. 12.5. Фазовые портреты такие же, как на рис. 12.4, но устойчивое равновесие O_1 , в котором есть только производители, заменилось устойчивым равновесием O_2 , где нет только управленцев

Нестабильное сообщество можно представить себе возникшим из соответствующего стабильного, благодаря потере устойчивости состоянием равновесия O , отвечающего сообществу «производители – продукт – управленцы», в результате рождения от него устойчивого периодического движения Γ . Периодическое движение Γ может располагаться целиком выше плоскости $z = 0$ полного истощения накопленного продукта, а может включать в себя участок, расположенный на плоскости $z = 0$.

Модель сообщества с высоким уровнем технологии $g > h(1 + \delta df)$ всегда имеет два состояния равновесия O_2 и O_4 . Ранее устойчивое равновесие O_2 , отве-

чающее сообществу «производители–продукт», становится неустойчивым. Появляется новое состояние равновесия O_4 . В зависимости от знака величины Δ оно устойчиво или неустойчиво: первое имеет место при $\Delta > 0$ и второе – при $\Delta < 0$. При его локальной устойчивости оно глобально устойчиво и соответствует сообществу «производители–продукт–управленцы». При смене знака величиной Δ и потере устойчивости состоянием равновесия из него рождается устойчивое периодическое движение Γ . Его устойчивость также глобальная, и оно соответствует сообществу «производители–продукт–управленцы». Как и ранее, это периодическое решение, отвечающее неустойчивому автоколебательному сообществу «производители–продукт–управленцы», располагается выше плоскости $z = 0$, но может содержать части, лежащие на плоскости $z = 0$. Соответствующие сделанному описанию фазовые портреты изображены на рис. 12.6.

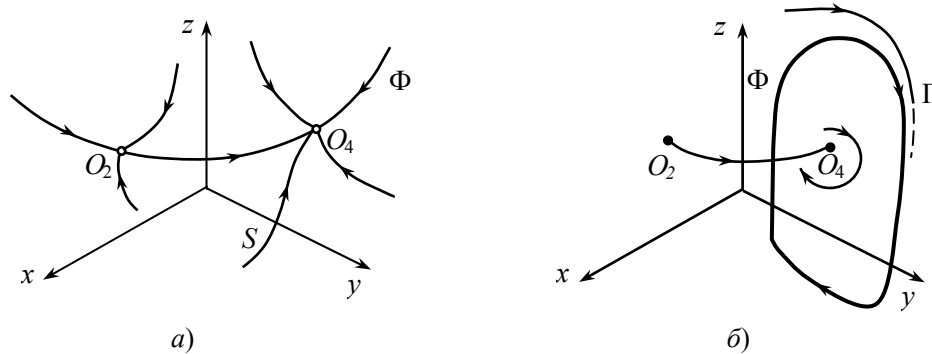


Рис. 12.6. Фазовые портреты при высоком уровне технологии и стабильности (а) и существенной неустойчивости (б) общества, где есть производители и управленцы

Кроме этих качественных различий, можно отметить, что устойчивое или неустойчивое равновесное сообщество, точнее, отвечающее ему равновесие $x^* \neq 0$, $y^* \neq 0$, $z^* \neq 0$, может по-разному вести себя при неограниченном росте параметра g : x^* , y^* , z^* неограниченно растут, либо остаются ограниченными. Первое имеет место при $ce - lf > 0$, второе – при обратном неравенстве.

Описанные выше состояния равновесия и периодические движения лежат целиком в параллелепипеде, определяемом неравенствами:

$$0 \leq x \leq A, \quad 0 \leq y \leq B, \quad 0 \leq z \leq C,$$

где конечные A , B и C определяются неравенствами:

$$\max(1, \varepsilon_1/\varepsilon_2) - (1 + \delta C)h < 0, \quad -d - eB + fC < 0, \quad a - bA + cC < 0,$$

поскольку через его стороны $x = A$, $y = B$ и $z = C$ фазовые траектории входят внутрь него.

Определенный интерес представляют количественные значения координат x^* , y^* и z^* для устойчивых сообществ, отвечающие устойчивым равновесиям. Они легко находятся для обществ производителей и производителей и продукта:

$$x^* = \frac{a}{b}, \quad y^* = z^* = 0 \quad (12.2)$$

и, соответственно,

$$x^* = \frac{a}{b} + \frac{c(g-h)}{\delta bh}, \quad y^* = 0, \quad z^* = \frac{g-h}{\delta h}. \quad (12.3)$$

В случае сообщества «производители – продукт – управленцы» отыскание x^* , y^* , z^* сложнее и определяется следующим образом: находим максимальный положительный корень y^* уравнения

$$g \frac{1 + \varepsilon_1 y}{1 + \varepsilon_2 y} \frac{a_1 + a_2 y}{1 + \delta b_1 + \delta b_2 y} - ha_1 - (ha_2 + k)y = 0, \quad (12.4)$$

где

$$a_1 = \frac{a}{b} + \frac{cd}{bf}, \quad a_2 = \frac{ce}{bf} - \frac{l}{b}, \quad b_1 = \frac{d}{f}, \quad b_2 = \frac{e}{f},$$

и затем x^* , z^* определяются формулами:

$$x^* = a_1 + a_2 y^*, \quad z^* = b_1 + b_2 y^*. \quad (12.5)$$

При условии

$$g < h \left(1 + \frac{\delta d}{f} \right) \quad (12.6)$$

уравнение (12.4) имеет либо два положительных корня, либо ни одного. Слияние корней происходит при

$$\varepsilon_1 = K(\varepsilon_2), \quad (12.7)$$

и они отсутствуют при $\varepsilon_1 < K(\varepsilon_2)$.

При условии

$$g > h \left(1 + \frac{\delta d}{f} \right) \quad (12.8)$$

уравнение (12.4) имеет один и только один положительный корень.

Устойчивость равновесия (x^*, y^*, z^*) со всеми координатами, отличными от нуля и где y^* – максимальный корень уравнения (12.4), имеет место при выполнении неравенства:

$$\Delta = (ey^* + bx^*)(bBx^* - Afy^* + eBy^* + bex^* y^*) - (bkfx^* y^* + lkfx^* y^{*2} - cekx^{*2} y^{*2}) > 0, \quad (12.9)$$

где

$$A = g \frac{\varepsilon_1 - \varepsilon_2}{(1 + \varepsilon_2 y^*)^2} \frac{x^*}{1 + \delta z^*}, \quad B = \frac{\delta(hx^* + ky^*)}{1 + \delta z^*}.$$

При убывании Δ в момент $\Delta = 0$ из состояния равновесия (x^*, y^*, z^*) рождается устойчивый предельный цикл Г.

Подведем итоги сказанному. Они представлены на диаграмме рис. 12.7. На ней уже указаны номера рисунков соответствующих фазовых портретов. Прокомментируем диаграмму. При $g < h$ в зависимости от остальных параметров возможны три вида фазового портрета, представленные на рис. 12.2 и рис. 12.4,а и б. В первом случае имеет место глобально устойчивое равновесие $O_1(x^* > 0, y^* = z^* = 0)$, отвечающее сообществу производителей. Во втором случае все фазовое про-

странство разбивается на две части: часть Φ_1 – это область притяжения устойчивого равновесия $O_1 (x^* > 0, y^* = z^* = 0)$ и часть Φ_2 , являющуюся областью притяжения либо устойчивого равновесия $O_4 (x^* > 0, y^* > 0, z^* > 0)$ (рис. 12.4,а), либо устойчивого периодического движения – автоколебания – Γ (рис. 12.4,б). Возникновение второй области Φ_2 и ее притягивающего установившегося движения при малом g объясняется его фактическим увеличением за счет воздействия управленцев, возникновение которых стало возможным, благодаря их значительному положительному воздействию на производительность общества. С исчезновением этой положительной роли управленцы, а с ними и накопленный продукт, исчезают.

При среднем уровне технологии производства, когда $h < g < h(1 + \delta df)$, также в зависимости от ε_1 , возможны три разных вида фазовых портретов, изображенных на рис. 12.3–12.5. В случае, изображенном на рис. 12.3, имеет место сообщество «производители и продукт». Управленцы в нем изживаются. Это если $\varepsilon_1 < K$. Напротив, при $\varepsilon_1 > K$ возникает сообщество, в котором, помимо производителей и накопленного продукта, есть еще и управленцы. В зависимости от знака величины Δ сообщество стабильное или нестабильное, а его установившийся режим – состояние равновесия или автоколебание. В третьем случае – высокого уровня технологии – появление управленцев не зависит от того, положительно или отрицательно их воздействие на общество: они возникают, потому что производительность общества позволяет им это сделать. Возникающие варианты фазового портрета зависят от знака величины Δ и соответствуют стабильному равновесному сообществу и нестабильному автоколебательному. Все фазовое пространство в обоих случаях является областью притяжения равновесия или автоколебания, т.е. они глобально устойчивы.

Отметим, что наряду с сообществами, в которых устанавливающееся состояние определяется только значениями параметров, возможны случаи, когда устанавливающееся движение зависит еще и от начальных условий.

Разделение стабильных обществ на общества неограниченного и ограниченно-го развития не имеет качественного характера и относится к характеру изменения координат глобально устойчивого равновесия x^*, y^*, z^* при $g \rightarrow \infty$. Отметим, что такова же роль знака величины $ce - lf$ на характер изменения средних значений периодических функций $x^*(t), y^*(t), z^*(t)$, описывающих автоколебания, как это следует из соотношений для средних значений этих функций, которые можно получить из дифференциальных уравнений модели:

$$\begin{aligned} a - bx^*(t) - ly^*(t) + cz^*(t) &= 0, \\ -d - ey^*(t) + fz^*(t) &= 0, \end{aligned}$$

поскольку из них следует, что

$$\tilde{x}^* = \frac{a}{b} + \frac{cd}{bf} + \frac{1}{bf}(ce - lf)\tilde{y}^*.$$

В заключение обратим внимание на рис. 12.7. На нем указаны условия, отвечающие различным видам фазовых портретов, и показано, как они переходят друг в друга с изменением параметров. Параметр g , и отчасти ε_1 , определяют общую структуру сообщества. При этом можно отметить, что если при высоком уровне

технологии управленцы возникают вне зависимости от их полезности или вредности для сообщества, то при более низком уровне технологии их возникновение имеет место только при их достаточно высокой полезности. От знака параметра Δ зависит стабильность сообщества, т.е. является ли его предельное установившееся движение устойчивым равновесием или автоколебанием. Наконец, важен еще параметр $ce - lf$: он определяет количественный характер развития сообщества с ростом уровня технологии.

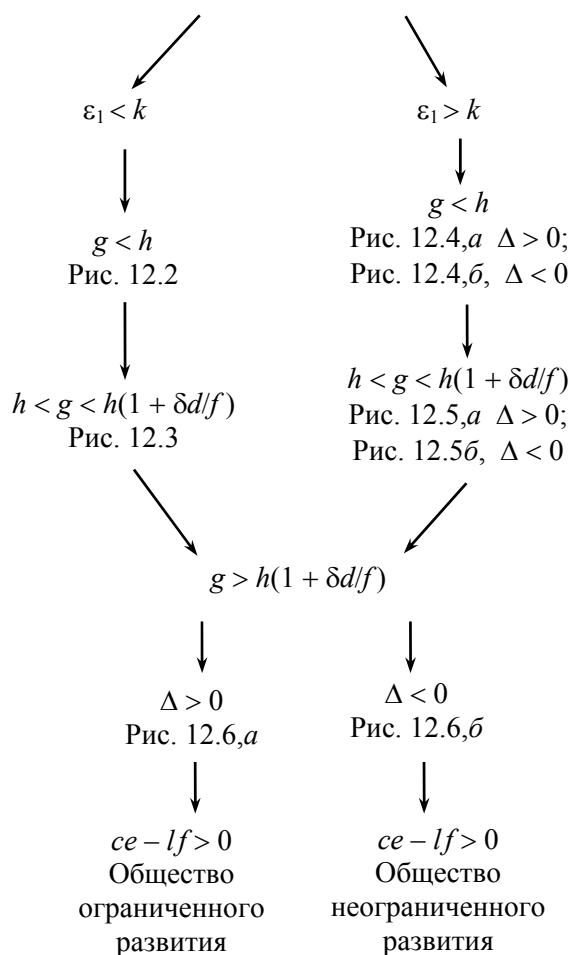


Рис. 12.7. Условия, отвечающие различным видам фазовых портретов

Можно смотреть на диаграмму и фазовые картинки как на игру математического ума и воображения, а можно и задуматься о том, какое отношение они имеют к многовековой истории человечества и нашей судьбе.

13. ЛИНЕЙНЫЙ ОСЦИЛЛЯТОР

Линейный осциллятор – очень простая математическая модель, замечательная как по разнообразию и широте конкретных интерпретаций, так и многообразию описываемых ею явлений. В механике осциллятор всюду, где есть взаимодействующие масса и упругость, в электродинамике – там, где есть емкости и индуктивности. В схематическом виде они представлены на рис. 13.1 и 13.2.

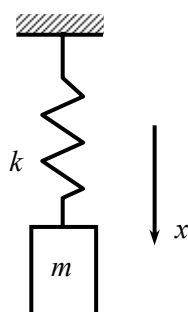


Рис. 13.1. Грузик, удерживаемый пружиной

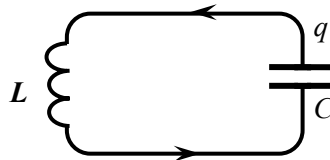


Рис. 13.2. Электрический контур с емкостью и индуктивностью

Линейный осциллятор описывает периодические гармонические колебания, затухающие и нарастающие колебания разных частот, различные равновесия: устойчивые и неустойчивые, типа узла, фокуса и седла. Математической моделью линейного осциллятора является линейное дифференциальное уравнение второго порядка вида

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = 0. \quad (13.1)$$

В нем всего лишь два параметра δ и ω^2 , его фазовое пространство двумерно.

Простейшими физическими объектами, приводящими к уравнению (13.1), являются масса на пружинке и электрический контур из индуктивности и емкости, изображенные на рис. 13.1 и 13.2.

Масса m может перемещаться только вдоль оси x , а сила упругости пружины при ее растяжении на x от нерастянутого состояния $x = 0$ равна $-kx$. Минус потому, что сила противоположна смещению массы m . Согласно уравнению Ньютона,

$$m\ddot{x} = -kx$$

или

$$\ddot{x} + \frac{k}{m} x = 0,$$

что при $\delta = 0$ и $k/m = \omega^2$ совпадает с (13.1).

В замкнутом контуре, включающем в себя емкость C и индуктивность L (см. рис. 13.2), q – заряд конденсатора и, следовательно, $I = \dot{q}$ – сила тока в контуре. На обкладках конденсатора возникает напряжение q/C , а в индуктивности L – электродвижущая сила (ЭДС), равная $-LdI/dt$. Опять же минус, потому что ЭДС

самоиндукции противоположна нарастанию тока. В силу того, что конденсатор замкнут на индуктивность, напряжения q/C и $-Ldl/dt$ равны друг другу, т.е.

$$-L \frac{dl}{dt} = \frac{q}{C},$$

или, учитывая, что $I = \dot{q}$,

$$\ddot{q} + \frac{q}{LC} = 0.$$

Это совпадает с (13.1) при $\delta = 0$ и $\omega^2 = 1/LC$.

В любом реальном электрическом контуре имеется омическое сопротивление (если только мы не имеем дело с исключительным случаем сверхпроводимости). Схема соответствующего контура, включающего сопротивление, представлена на рис. 13.3.

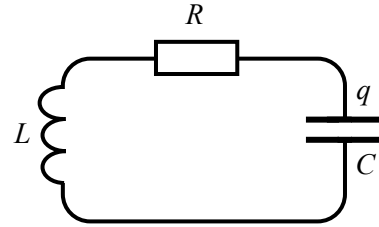


Рис. 13.3. Электрический контур с конденсатором, индуктивностью и сопротивлением

В этом случае ЭДС конденсатора q/C противостоит ЭДС последовательно соединенных индуктивности с ЭДС $-Ldl/dt$ и сопротивления с ЭДС $-RI$. В силу этого имеем:

$$-L \frac{dl}{dt} - RI = \frac{q}{C}$$

или

$$\ddot{q} + \frac{R}{L} \dot{q} + \frac{1}{LC} q = 0,$$

что совпадает с (13.1) при $2\delta = R/L$ и $\omega^2 = 1/LC$. Аналогично и в случае массы на пружинке может присутствовать препятствующее движению сопротивление, пропорциональное скорости \dot{x} с коэффициентом h , и тогда

$$m\ddot{x} = -kx - h\dot{x}$$

или

$$\ddot{x} + \frac{h}{m} \dot{x} + \frac{k}{m} x = 0,$$

что опять приводит к уравнению линейного осциллятора (13.1) с $2\delta = h/m$ и $\omega^2 = k/m$.

Дифференциальное уравнение (13.1) имеет частные решения вида $x = e^{\lambda t}$, где λ – любой из корней так называемого характеристического уравнения

$$\lambda^2 + 2\delta\lambda + \omega^2 = 0,$$

получаемого прямой подстановкой этого предполагаемого решения в уравнение (13.1). Вообще, таких корней два, и общее решение, получаемое согласно принципу суперпозиции линейных динамических систем, имеет вид:

$$x = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}. \quad (13.2)$$

Напомним, что принцип суперпозиции применительно к осциллятору (13.1) означает, что если $x_1(t)$ и $x_2(t)$ – произвольные решения дифференциального уравнения (13.1), то и их любая линейная комбинация $C_1 x_1(t) + C_2 x_2(t)$ – также решение.

Общее решение уравнения (13.1) при $\lambda_1 = \lambda_2$, т.е. при $\delta^2 = \omega^2$, также можно найти, опираясь на принцип суперпозиции. Действительно, пусть λ_1 близко к λ_2 и имеется два независимых решения $e^{\lambda_1 t}$ и $e^{\lambda_2 t}$. Однако при $\lambda_2 \rightarrow \lambda_1$ они сливаются и остается только одно. Но можно рассмотреть решение

$$x = (\lambda_1 - \lambda_2)^{-1} (e^{\lambda_1 t} - e^{\lambda_2 t}),$$

которое при $\lambda_2 \rightarrow \lambda_1$ не исчезает и не совпадает с $e^{\lambda_1 t}$: оно равно $t e^{\lambda_1 t}$. Таким образом, при $\lambda_1 = \lambda_2 = \lambda$ общее решение может быть записано в виде:

$$x = C_1 e^{\lambda t} + C_2 t e^{\lambda t}. \quad (13.3)$$

Это простое рассуждение говорит о том, что мы можем считать $\lambda_1 \neq \lambda_2$ и, если имеет место обратное, то в конечном результате следует перейти к пределу при $\lambda_2 \rightarrow \lambda_1$.

Корни λ_1 и λ_2 характеристического уравнения могут быть действительными или комплексными. В любом случае при подходящих C_1 и C_2 общее решение (13.3) может удовлетворить любым начальным условиям вида:

$$x|_{t=0} = x_0, \quad \dot{x}|_{t=0} = x_1. \quad (13.4)$$

Действительно, расписывая эти условия, придем к двум линейным уравнениям относительно C_1 и C_2 :

$$C_1 + C_2 = x_0, \quad \lambda_1 C_1 + \lambda_2 C_2 = x_1,$$

откуда легко находим, что

$$C_1 = \frac{\lambda_2 x_0 - x_1}{\lambda_2 - \lambda_1}, \quad C_2 = \frac{\lambda_1 x_0 - x_1}{\lambda_1 - \lambda_2},$$

и поэтому соответствующее решение принимает вид:

$$x = (\lambda_2 - \lambda_1)^{-1} \{(\lambda_2 x_0 - x_1) e^{\lambda_1 t} - (\lambda_1 x_0 - x_1) e^{\lambda_2 t}\}. \quad (13.5)$$

При комплексных корнях, когда

$$\lambda_{1,2} = -\delta \pm i\Omega, \quad \Omega = \sqrt{\omega^2 - \delta^2}, \quad (13.6)$$

из (13.5) находим, что

$$\begin{aligned} x &= -\frac{1}{2} i \Omega \{ [(-\delta - i\Omega)x_0 - x_1] e^{-\delta t} (\cos \Omega t + i \sin \Omega t) - \\ &\quad - [(-\delta + i\Omega)x_0 - x_1] e^{-\delta t} (\cos \Omega t - i \sin \Omega t) \} = \\ &= e^{-\delta t} \{ x_0 \cos \Omega t + \Omega^{-1} (\delta x_0 + x_1) \sin \Omega t \}. \end{aligned} \quad (13.7)$$

К этому же результату (13.7) и общему виду действительного решения при комплексных корнях характеристического уравнения можно прийти и другим путем. Опять же в силу принципа суперпозиции, наряду с $e^{\lambda_1 t}$ и $e^{\lambda_2 t}$, решениями будут и

$$\frac{1}{2}(e^{\lambda_1 t} + e^{\lambda_2 t}) = e^{-\delta t} \cos \Omega t,$$

и

$$\frac{1}{2i}(e^{\lambda_1 t} - e^{\lambda_2 t}) = e^{-\delta t} \sin \Omega t,$$

которые уже действительные. Суперпозиция этих действительных решений и дает его общий вид.

Что же собою представляют найденные нами общие решения

$$x = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} \tag{13.8}$$

в случае действительных λ_1 и λ_2 и общие решения

$$x = e^{-\delta t} (C_1 \cos \Omega t + C_2 \sin \Omega t) \tag{13.9}$$

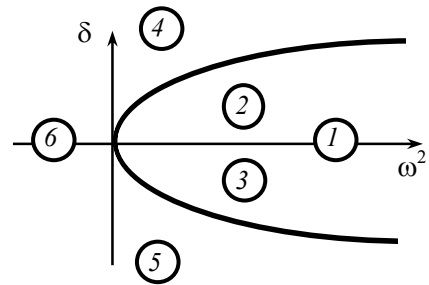
при комплексных λ_1 и λ_2 ?

Нас интересуют графики $x(t)$ и фазовые портреты, отвечающие случаям (13.8) и (13.9). И то и другое зависит от параметров δ и ω^2 . Прежде всего, от этих параметров зависит, какой из случаев (13.8) или (13.9) имеет место.

Для наглядности введем в рассмотрение плоскость параметров ω^2 и δ (рис. 13.4). Подчеркнем, что оба параметра δ и ω^2 могут быть любого знака.

Граничному случаю, разделяющему (13.8) и (13.9), отвечает равенство $\delta^2 = \omega^2$, которому на плоскости рис. 13.4. соответствует нанесенная на нем парабола. Внутри этой параболы $\delta^2 < \omega^2$ и имеет место (13.9), вне ее $\delta^2 > \omega^2$ и имеет место (13.8).

Рис. 13.4. Разбиение плоскости параметров ω^2, δ линейного осциллятора на области с различными корнями λ_1 и λ_2 характеристического уравнения: 1 – оба корня чисто мнимые, 2 – комплексные с отрицательными действительными частями, 3 – комплексные с положительными действительными частями, 4 – действительные отрицательные, 5 – действительные положительные, 6 – действительные разных знаков



Рассмотрим сначала осцилляторы, отвечающие точкам внутри параболы. Точкам на оси $\delta = 0$, помеченным на рис. 13.4 цифрой 1, внутри параболы отвечают гармонические осцилляторы

$$\ddot{x} + \omega^2 x = 0, \tag{13.10}$$

движения которых – хорошо известные гармонические колебания

$$x = C_1 \cos \omega t + C_2 \sin \omega t = A \cos(\omega t + \varphi), \tag{13.11}$$

где A – амплитуда колебаний, ω – частота колебаний, а φ – фаза. С постоянными C_1 и C_2 амплитуда A и фаза φ связаны соотношениями:

$$A \sin \varphi = C_1, \quad A \cos \varphi = C_2,$$

из которых следует, что

$$A = \sqrt{C_1^2 + C_2^2}, \quad \operatorname{tg} \varphi = \frac{C_1}{C_2}.$$

График гармонических колебаний представлен на рис. 13.5. На нем же показаны амплитуда A , фаза φ и период $T = 2\pi/\omega$.

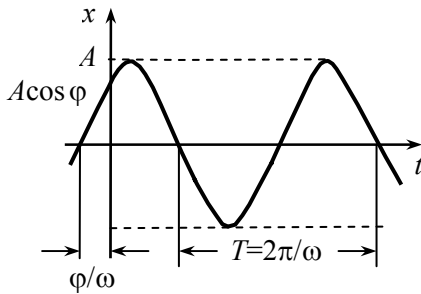


Рис. 13.5. Временной график гармонического колебания

Из (13.11) находим параметрические уравнения фазовых траекторий в виде:

$$x = A \cos(\omega t + \varphi),$$

$$\dot{x} = -\omega A \sin(\omega t + \varphi).$$

При изменении t точка (x, \dot{x}) пробегает по часовой стрелке эллипс

$$\frac{x^2}{A^2} + \frac{\dot{x}^2}{\omega^2 A^2} = 1.$$

Различным A отвечают разные эллипсы, что в итоге приводит к фазовому портрету рис. 13.6.

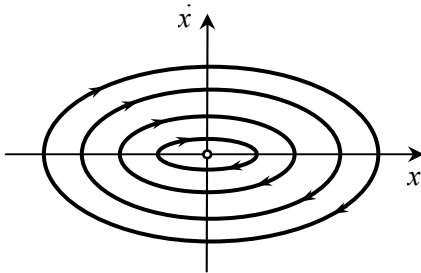


Рис. 13.6. Фазовый портрет гармонического осциллятора (область I рис. 13.4)

Заметим, что найти фазовые траектории можно было и непосредственно, интегрируя дифференциальное уравнение (13.10) гармонического осциллятора. Именно, умножая (13.10) на \dot{x} и интегрируя, находим, что

$$\frac{\dot{x}^2}{2} + \frac{\omega^2 x^2}{2} = h,$$

где постоянная интегрирования h , имеющая смысл приведенной энергии осциллятора, связана с амплитудой колебаний осциллятора соотношением

$$h = \frac{1}{2} \omega^2 A^2.$$

Сместимся теперь в область плоскости параметров ω^2, δ выше оси абсцисс, оставаясь внутри параболы. В этом случае придем к тому, что согласно (13.9),

$$x = Ae^{-\delta t} \cos(\omega t + \varphi). \quad (13.12)$$

Это соответствует затухающим гармоническим колебаниям, изображенным на рис. 13.7. Их период колебаний $2\pi/\Omega$ несколько больше, чем период колебаний соответствующего гармонического осциллятора, равного $2\pi/\omega$ ($\Omega = \sqrt{\omega^2 - \delta^2}$). При $\delta < 0$, т.е. при смещении вниз от оси абсцисс по плоскости параметров ω^2, δ рис. 13.4, также имеет место (13.12). Однако теперь соответствующие колебания нарастающие, как показано на рис. 13.8. Фазовые портреты в рассмотренных случаях состоят из скручивающихся ($\delta > 0$) или раскручивающихся ($\delta < 0$) спиралей, представленных на рис. 13.9, а и б.

Рис. 13.7. Экспоненциально затухающее гармоническое колебание (область 3 рис.13.4)

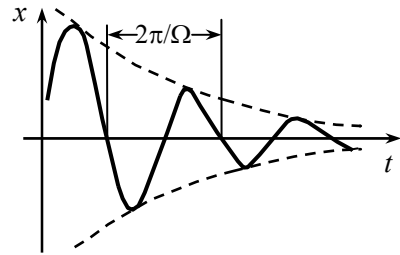


Рис. 13.8. Экспоненциально нарастающие гармонические колебания

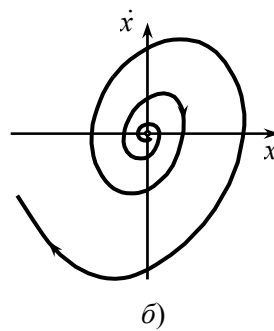
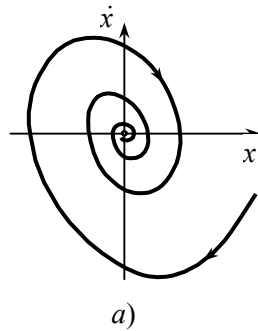
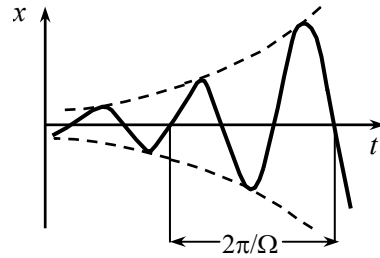


Рис. 13.9. Фазовые портреты (область 2 рис.13.4): а) для $\delta > 0$; б) для $\delta < 0$

Итак, мы нашли графики движения и фазовые портреты в областях 1 ($\delta = 0, \omega^2 > 0$), 2 ($\delta > 0, \delta^2 < \omega^2$) и 3 ($\delta < 0, \delta^2 < \omega^2$), показанных на рис. 13.4.

Найдем еще графики изменения x и фазовые портреты для осцилляторов из областей 4 ($\delta > 0, \delta^2 > \omega^2 > 0$), 5 ($\delta < 0, \delta^2 > \omega^2 > 0$) и 6 ($\omega^2 < 0$) того же рис. 13.4. Если предшествующие случаи отвечали комплексным корням характеристического уравнения, то теперь эти корни действительные, и общее решение определяет-

ся формулой (13.8). Это решение является суперпозицией двух решений $x = e^{\lambda_1 t}$ и $x = e^{\lambda_2 t}$. Графики этих решений – возрастающие или убывающие экспоненты в зависимости от знаков λ_1 и λ_2 , что позволяет представить себе, каковы графики их суперпозиций с любыми коэффициентами C_1 и C_2 .

Для построения фазовых портретов построим сначала фазовые траектории для частных решений $x = e^{\lambda_1 t}$ и $x = e^{\lambda_2 t}$. Пусть λ – это λ_1 либо λ_2 , тогда параметрическое уравнение фазовой траектории записывается в обоих случаях в виде

$$x = e^{\lambda t}, \quad \dot{x} = \lambda e^{\lambda t} \quad (-\infty < t < \infty), \quad (13.13)$$

и фазовые траектории представляют собой лучи

$$\dot{x} = \lambda x \quad (x > 0 \text{ или } x < 0),$$

пробегаемые фазовой точкой, согласно (13.13), от начала координат до бесконечности при $\lambda > 0$ или, наоборот, из бесконечности до нуля при $\lambda < 0$. Нетрудно понять векторное равенство

$$(x, \dot{x}) = C_1(1, \lambda_1)e^{\lambda_1 t} + C_2(1, \lambda_2)e^{\lambda_2 t}, \quad (13.14)$$

непосредственно следующее из (13.13). Соответствующая ему геометрическая интерпретация представлена на рис. 13.10. Она позволяет, меняя знаки и величины C_1 и C_2 , построить фазовый портрет.

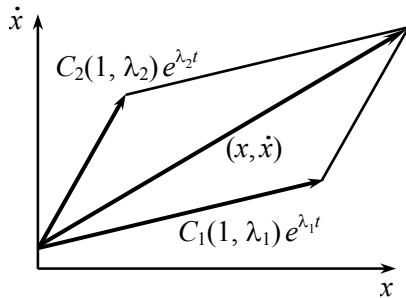


Рис. 13.10. Вспомогательное построение для получения фазового портрета узлового состояния равновесия

В области 4 рис. 13.4 λ_1 и λ_2 отрицательны и соответствующий фазовый портрет представлен на рис. 13.11. В области 5 λ_1 и λ_2 положительны, и фазовый портрет имеет вид, показанный на рис. 13.12. В области 6 λ_1 и λ_2 – разных знаков, а фазовый портрет вида, показанного на рис. 13.13.

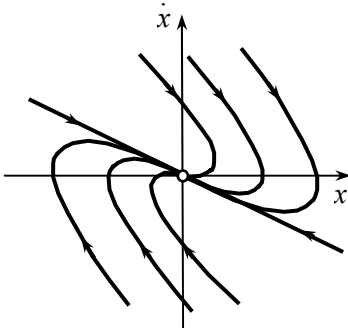


Рис. 13.11. Фазовый портрет устойчивого узла (для области 4 рис. 3.4)

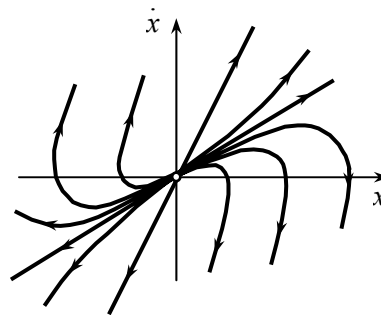


Рис. 13.12. Фазовый портрет неустойчивого узла (для области 5 рис. 3.4)

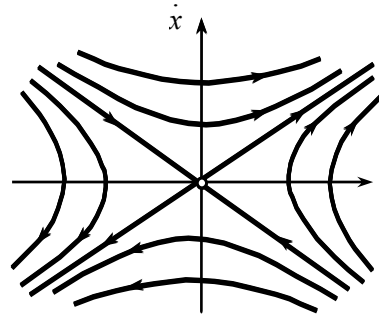


Рис. 13.13. Фазовый портрет седла (для области 6 рис. 13.4)

Таким образом, мы нашли все возможные виды фазовых портретов линейного осциллятора. На всех них есть выделенная фазовая траектория, сводящаяся к одной точке – состоянию равновесия $x = 0, \dot{x} = 0$. Остальные фазовые траектории как бы расположены вокруг нее и поэтому можно говорить о различных типах состояний равновесия. Они соответствуют областям 1, 2, 3, 4, 5 и 6 рис. 13.4 и называются центром, устойчивым и неустойчивым фокусом, устойчивым и неустойчивым узлом и седлом. В этом перечислении выделена граница 1 между областями 2 и 3. Остальные границы оставлены без внимания. Это вызвано ее особой ролью. Эта граница между экспоненциальной устойчивостью и неустойчивостью, затухающими и нарастающими колебаниями, и этот особый граничный случай является общим для консервативных и, в частности, гамильтоновых систем.

Фазовые портреты гармонических осцилляторов, принадлежащих этому граничному случаю, обладают очень важной для статистической механики особенностью. Она проявляется, когда вместо одного осциллятора рассматривается ансамбль идентичных осцилляторов, каждый из которых представляется своей движущейся фазовой точкой. При этом фазовый портрет можно интерпретировать как поток частиц, отображающих осцилляторы, или как течение фазовой жидкости. Фазовые частицы с течением времени перемещаются и как-то деформируются. Если задана область фазового пространства G_0 , то, спустя время t , она перейдет в область G_t согласно преобразованию:

$$x = e^{-\delta t} \left\{ \left(\cos \Omega t + \frac{\delta}{\Omega} \sin \Omega t \right) x_0 + \frac{\dot{x}_0}{\Omega} \sin \Omega t \right\},$$

$$\dot{x} = e^{-\delta t} \left\{ -x_0 \left(\Omega + \frac{\delta^2}{\Omega} \right) \sin \Omega t + \left(-\frac{\delta}{\Omega} \sin \Omega t + \cos \Omega t \right) \dot{x}_0 \right\},$$

следующему из формулы (13.7), определяющей решение $x(t)$ через начальные условия x_0 и \dot{x}_0 . Якобиан этого преобразования

$$J = \begin{vmatrix} \frac{\partial x}{\partial x_0} & \frac{\partial x}{\partial \dot{x}_0} \\ \frac{\partial \dot{x}}{\partial x_0} & \frac{\partial \dot{x}}{\partial \dot{x}_0} \end{vmatrix}$$

находится после несложных вычислений и оказывается равным $e^{-2\delta t}$. Этот результат означает, что при своем течении фазовая жидкость сжимается при $\delta > 0$, расширяется при $\delta < 0$, и при $\delta = 0$ ведет себя как несжимаемая. Точнее, любой элемент фазового объема $d\sigma_0$ преобразуется за время t в элемент $d\sigma_t$ и

$$d\sigma_t = e^{-2\delta t} d\sigma_0,$$

так что $e^{-2\delta t}$ является коэффициентом сжатия ($\delta > 0$) или коэффициентом растяжения ($\delta < 0$) текущей фазовой жидкости. При $\delta = 0$ этот коэффициент растяжения равен единице, что соответствует течению несжимаемой жидкости.

Перечисленные выше виды фазового портрета линейного осциллятора – центр, фокус, узел и седло – имеют место не только у равновесий линейного осциллятора, но и у равновесий $O(x^*, y^*)$ любых двумерных динамических систем, описываемых дифференциальными уравнениями второго порядка

$$\dot{x} = X(x, y), \quad \dot{y} = Y(x, y). \quad (13.15)$$

Мы с этим уже сталкивались, рассматривая фазовые портреты взаимодействующих популяций, причем эти фазовые портреты имели место не на всей фазовой плоскости, а только в некоторой окрестности состояния равновесия. Как и для осциллятора, вид фазового портрета в окрестности состояния равновесия x^*, y^* определяется корнями λ_1 и λ_2 характеристического уравнения, которое для уравнения (13.15), как известно, принимает вид:

$$\begin{aligned} X(\lambda) &= \begin{vmatrix} \left(\frac{\partial X}{\partial x}\right)^* - \lambda & \left(\frac{\partial X}{\partial y}\right)^* \\ \left(\frac{\partial Y}{\partial x}\right)^* & \left(\frac{\partial Y}{\partial y}\right)^* - \lambda \end{vmatrix} = \\ &= \lambda^2 - \left(\left(\frac{\partial X}{\partial x}\right)^* + \left(\frac{\partial Y}{\partial y}\right)^* \right) \lambda + \left(\frac{\partial X}{\partial x}\right)^* \left(\frac{\partial Y}{\partial y}\right)^* - \left(\frac{\partial X}{\partial y}\right)^* \left(\frac{\partial Y}{\partial x}\right)^* = 0, \end{aligned}$$

где звездочки означают, что производные вычисляются в точке равновесия x^*, y^* .

В окрестности состояния равновесия линеаризованные уравнения (13.15) имеют вид:

$$\begin{aligned} \dot{\xi} &= \left(\frac{\partial X}{\partial x}\right)^* \xi + \left(\frac{\partial X}{\partial y}\right)^* \eta, \\ \dot{\eta} &= \left(\frac{\partial Y}{\partial x}\right)^* \xi + \left(\frac{\partial Y}{\partial y}\right)^* \eta, \end{aligned}$$

где ξ и η – отклонения от равновесия x^*, y^* , т.е. $\xi = x - x^*$ и $\eta = y - y^*$.

Исключение из этих дифференциальных уравнений ξ или η приводит к уравнению линейного осциллятора для ξ :

$$\ddot{\xi} - \left[\left(\frac{\partial X}{\partial x}\right)^* + \left(\frac{\partial Y}{\partial y}\right)^* \right] \dot{\xi} + \left[\left(\frac{\partial X}{\partial x}\right)^* \left(\frac{\partial Y}{\partial y}\right)^* - \left(\frac{\partial X}{\partial y}\right)^* \left(\frac{\partial Y}{\partial x}\right)^* \right] \xi = 0$$

или точно такому же уравнению для η .

Ниже будут приведены конкретные примеры линейных осцилляторов из всех областей 1–6 бифуркационного портрета рис. 13.4. При этом модель линейного осциллятора (13.1), как правило, является тем или иным приближением реальной

системы, очень часто приемлемой только при малых или не очень больших отклонениях от равновесия. Так, в линейном приближении нижнее состояние равновесия маятника описывается линейным гармоническим осциллятором с состоянием равновесия типа центр; движение вблизи верхнего равновесия – осциллятором с неустойчивым седловым равновесием. Действительно, круговой маятник

$$\ddot{\varphi} + \frac{g}{l} \sin \varphi = 0 \quad (13.16)$$

имеет два равновесия: нижнее $\varphi^* = 0$ и верхнее $\varphi^* = \pi$. Полагая $\varphi = \varphi^* + x$, имеем

$$\ddot{x} + \frac{g}{l} \sin(\varphi^* + x) = 0,$$

или

$$\ddot{x} + \frac{g}{l} \sin \varphi^* + \frac{g}{l} x \cos \varphi^* + \dots = 0, \quad (13.17)$$

где многоточие означает величины третьего и большего порядка малости относительно предполагаемого малым x . Из (13.17) при $\varphi^* = 0$ следует

$$\ddot{x} + \omega^2 x = 0, \quad \omega^2 = \frac{g}{l},$$

а при $\varphi^* = \pi$ следует

$$\ddot{x} - \omega^2 x = 0,$$

как и утверждалось.

Маятник, отвечающий математической модели (13.16), мы представляем в виде тела, подвешенного в поле силы тяжести. Но его можно представить и в виде диполя в электрическом поле. Если I – момент инерции диполя длины $2l$ относительно центра масс, e и $-e$ – заряды на его концах и E – напряженность однородного постоянного электрического поля, то для угла φ отклонения диполя от направления электрического поля имеет место уравнение

$$I\ddot{\varphi} = -2le \sin \varphi,$$

т.е. опять же классическое уравнение маятника, а в линейном приближении при малых отклонениях от равновесных значений $\varphi = 0$ и $\varphi = \pi$ – линейный осциллятор.

В хорошем приближении линейным осциллятором описываются малые колебания лодки или другого плавающего на воде тела. Для лодки или корабля различают бортовые и килевые колебания. Каждое из них имеет свою частоту. Можно приближенно линейным осциллятором описывать колебания струны, здания, имея в виду его вертикальные колебания на фундаменте и боковые колебания. Стоэтажный небоскреб колеблется так, что боковые смещения его верхней части порядка метра.

Осциллятором можно описать упругие колебания газа в сосуде в виде резонатора Гельмгольца, колебания молекул в кристалле, электромагнитных волн в проводящей полости и многое, многое другое: колебания автомобиля, поезда, пузырька воздуха в воде, разнообразных электрических контуров в радиосхемах; колебание воды в сообщающихся сосудах; колебания жидкости в нашем вестибулярном аппарате; крутильные колебания маховика на упругом валу.

Во всех приведенных примерах осцилляторов имеет место рассеяние энергии и колебания затухают, при этом $\delta > 0$, или $\delta = 0$, если этим затуханием можно пренебречь. Ниже будут приведены примеры с $\delta < 0$, с так называемым отрицательным трением, когда, напротив, происходит подкачка энергии от некоторого источника. Кроме того, будут рассмотрены некоторые нелинейные осцилляторы, для которых линейное описание недостаточно, а поведение существенно другое.

В заключение, как на итог всего сказанного, посмотрите на сводный рис. 13.14. На нем изображено, как ведут себя осциллограммы изменения координаты $x(t)$ и какой вид имеют фазовые портреты в зависимости от параметров ω^2 и δ , из разных областей бифуркационной диаграммы.

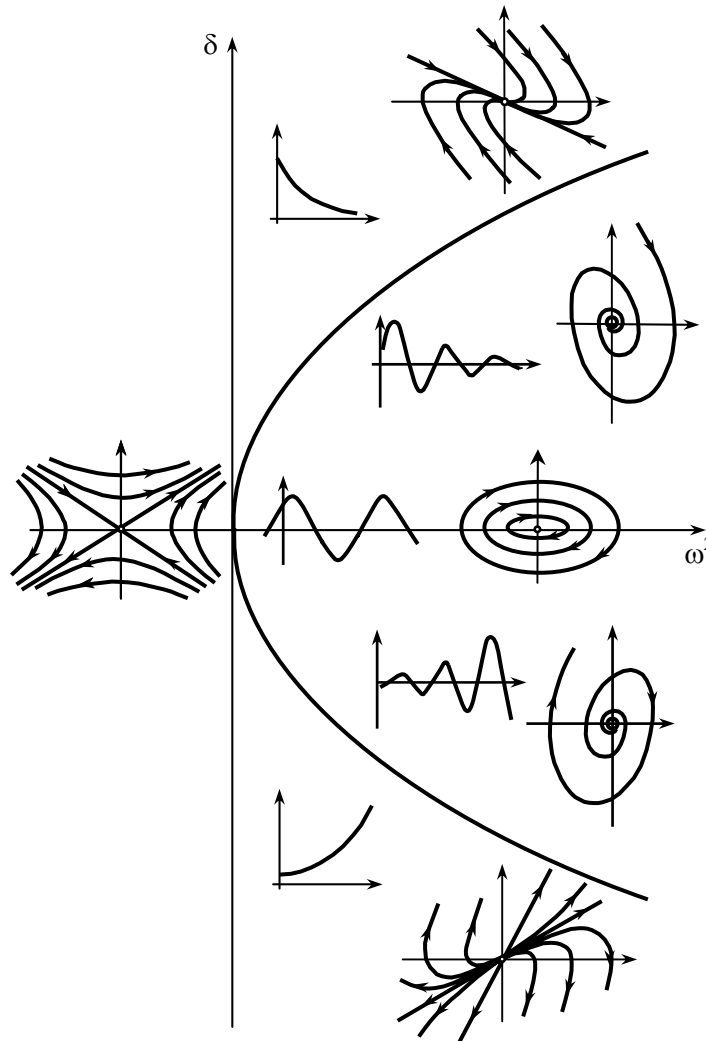


Рис. 13.14. Итоговый параметрический портрет линейного осциллятора. В каждой из его областей изображены временные графики и фазовые портреты

14. ЭЛЕКТРОМЕХАНИЧЕСКИЕ АНАЛОГИИ. УРАВНЕНИЯ ЛАГРАНЖА – МАКСВЕЛЛА

Вернемся к простейшим примерам механического и электрического линейных осцилляторов. Первая физическая модель описывается дифференциальным уравнением

$$m\ddot{x} + h\dot{x} + kx = 0, \quad (14.1)$$

а вторая – дифференциальным уравнением

$$L\ddot{q} + R\dot{q} + (1/C)q = 0. \quad (14.2)$$

Модели (14.1) и (14.2) переходят друг в друга при заменах переменной x на q (\dot{x} на \dot{q} , \ddot{x} на \ddot{q}) и параметров m на L , h на R , k на $1/C$. Поэтому если $m = L$, $h = R$, $k = 1/C$ и если в некоторый начальный момент времени $x = q$ и $\dot{x} = \dot{q}$, то равенства $x = q$ и $\dot{x} = \dot{q}$ будут иметь место после этого все время; x – это координата положения массы, а q – электрический заряд конденсатора, и они при перечисленных условиях будут меняться одинаково. Аналогично все время будут совпадать величина скорости \dot{x} перемещения массы и электрический ток \dot{q} в контуре. Поэтому можно говорить об аналогии массы на пружине с вязким трением и электрического контура, состоящего из индуктивности, емкости и сопротивления. Аналогии в том смысле, что процессы изменения совершенно разных физических величин x и q одинаковы, если только $m = L$, $h = R$ и $k = 1/C$. А аналогичными являются: механические смещения и электрический заряд, скорость и ток, масса и индуктивность, коэффициент вязкого трения и омическое сопротивление, коэффициент жесткости пружины и величина, обратная емкости конденсатора.

Продолжим эту аналогию. Кинетическая энергия массы равна $m\dot{x}^2/2$. Аналогом ее будет величина $L\dot{q}^2/2$. Это, как известно, энергия магнитного поля катушки индуктивности. Потенциальная энергия пружины – это $kx^2/2$, ее аналог – $q^2/2C$ и это, как мы видели ранее, энергия электрического поля конденсатора. Произведение kx – это сила натяжения пружины, ей отвечает величина электрического напряжения q/C конденсатора.

Мы установили аналогичность величин смещения, скорости, силы, кинетической и потенциальной энергий соответственно заряду, току, напряжению, энергии магнитного поля индуктивности и электрического поля конденсатора. Одновременно масса m , вязкое сопротивление h и жесткость k аналогичны индуктивности L , омическому сопротивлению R и величине, обратной емкости $1/C$.

Эта аналогия относится не только к массе на пружине и электрическому контуру. Она значительно шире. Чтобы в этом убедиться, представим механическую систему и электрическую, состоящими из элементов: масс, упругостей, вязкого сопротивления и, соответственно, индуктивности, емкостей и омического сопротивления. Каждый из этих простейших элементов имеет свою математическую модель. Укажем их. Масса m связывает ускорение \ddot{x} с силой F соотношением

$m\ddot{x} = F$. Аналогично индуктивность L связывает скорость изменения тока \dot{q} с напряжением E , так что $L\dot{q} = E$. Таковы же аналогии между упругостью и емкостью, механическим вязким и электрическим сопротивлениями. Выпишем их все:

$$\begin{aligned} m\ddot{x} = F &\leftrightarrow L\ddot{q} = E, \\ h\dot{x} = F &\leftrightarrow R\dot{q} = E, \\ kx = F &\leftrightarrow q/C = E. \end{aligned} \quad (14.3)$$

При этом в каждой из аналогий (14.3) сила F и напряжение E имеют свой смысл и величину. Слева – математические модели массы, вязкого сопротивления (демпфера), упругой пружины. Справа – индуктивности, омического сопротивления, емкости.

После этого ясно, что если имеются две системы: механическая и электрическая, составленные из перечисленных элементов и так, что между аналогичными элементами возникают одинаковые связи, то и системы в целом будут аналогичными. На этом основано электромеханическое аналоговое моделирование, позволяющее механическую систему изучить с помощью электрической и наоборот.

Не будем пока развивать дальше эту тему и ограничимся сказанным. Продумайте только, что чему соответствует в этой замечательной аналогии. Еще раз выпишем ее

$$\begin{aligned} x \leftrightarrow q, \quad mv^2/2 \leftrightarrow LI^2/2, \quad m \leftrightarrow L, \\ v = \dot{x} \leftrightarrow I = \dot{q}, \quad kx^2/2 \leftrightarrow q^2/2C, \quad k \leftrightarrow 1/C, \\ F \leftrightarrow E, \quad h \leftrightarrow R. \end{aligned} \quad (14.4)$$

Перейдем теперь к трудному вопросу составления математических моделей механических, электрических и электромеханических систем.

Для механических систем Лагранж придумал общий простой и удобный универсальный способ составления их динамических математических моделей в виде знаменитых дифференциальных уравнений Лагранжа. Вообще, это большая и красивая теория. Вы о ней узнаете из соответствующих курсов. А сейчас, не вдаваясь в детали, я постараюсь вам описать алгоритм составления этих дифференциальных уравнений для произвольной механической системы, положение всех материальных точек которой определяется конечным числом скалярных величин q_1, q_2, \dots, q_n , получивших наименование обобщенных координат.

Пусть сначала все силы, действующие в системе, потенциальные, т.е. для них есть потенциальная энергия $V(q_1, q_2, \dots, q_n, t)$, зависящая от обобщенных координат, и возможно, времени t . Это означает, что работа этих сил на любом виртуальном перемещении $\delta q_1, \delta q_2, \dots, \delta q_n$ равна уменьшению потенциальной энергии:

$$\delta A = -\delta V = -\sum_{\alpha} \frac{\partial V}{\partial q_{\alpha}} \delta q_{\alpha}.$$

Далее ясно, что кинетическую энергию T системы в любой момент времени можно выразить через обобщенные координаты q_1, q_2, \dots, q_n и обобщенные скорости $\dot{q}_1, \dot{q}_2, \dots, \dot{q}_n$. Ведь кинетическая энергия – это сумма кинетических энергий всех материальных точек системы, положение которых однозначно зависит от q_1, q_2, \dots, q_n и t .

Согласно Лагранжу, нужно составить функцию

$$L = T - V, \quad (14.5)$$

с ее помощью искомые уравнения движения записываются в виде:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} - \frac{\partial L}{\partial q_\alpha} = 0 \quad (\alpha = 1, 2, \dots, n). \quad (14.6)$$

Этот очень красивый общий способ составления уравнений движения, в силу изложенной аналогии, можно перенести на электрические системы. В свое время это сделал Максвелл. Это можно сделать не только для электрических систем, но и для электромеханических. Нужно лишь, в силу изложенной выше аналогии, под кинетической энергией T понимать сумму механической кинетической энергии и энергии магнитных полей индуктивностей, а под потенциальной энергией V – сумму механической потенциальной энергии и энергий электрических полей емкостей. Под обобщенными координатами q_1, q_2, \dots, q_n при этом понимаются любые скалярные величины, которые определяют не только механическое положение, но и все заряды, производные которых определяют все электрические токи. Подчеркнем, что при этом в системе нет ни механических, ни электрических сопротивлений.

Поясним на очень простом примере, как возникает запись дифференциальных уравнений математической модели движения в виде уравнения Лагранжа (14.6).

Рассмотрим уравнение Ньютона движения линейного механического осциллятора

$$m\ddot{x} + kx = 0 \quad (14.7)$$

и выполним с ним следующие преобразования:

$$m\ddot{x} = \frac{d}{dt} \frac{\partial}{\partial \dot{x}} \frac{m\dot{x}^2}{2}, \quad kx = \frac{\partial}{\partial x} \frac{kx^2}{2},$$

следовательно,

$$m\ddot{x} + kx = \frac{d}{dt} \frac{\partial}{\partial \dot{x}} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) - \frac{\partial}{\partial x} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) = 0, \quad (14.8)$$

что по форме совпадает с уравнением Лагранжа (14.6) и при этом x – обобщенная координата, $m\dot{x}^2/2$ – кинетическая энергия и $kx^2/2$ – потенциальная энергия.

Аналогично к лагранжеву виду приводится и уравнение электрического осциллятора:

$$L\ddot{q} + \frac{q}{C} = \frac{d}{dt} \frac{\partial}{\partial \dot{q}} \left(\frac{L\dot{q}^2}{2} - \frac{q^2}{2C} \right) - \frac{\partial}{\partial q} \left(\frac{L\dot{q}^2}{2} - \frac{q^2}{2C} \right) = 0, \quad (14.9)$$

где обобщенной координатой является заряд q конденсатора, $L\dot{q}^2/2$ – магнитная энергия катушки индуктивности, а $q^2/2C$ – электрическая энергия конденсатора.

Эти выкладки представляются искусственными и ничего не доказывающими. Это так и есть. Для того чтобы придать им смысл, нужно проникнуть в тайну лагранжевой записи уравнений движения. А тайна их в том, что они соответствуют некоторой вариационной задаче минимума функционала. Необходимое условие

минимума функционала, как известно, состоит в обращении в нуль первой его вариации. Условия обращения в нуль первой вариации оказываются уравнениями Лагранжа. Такова тайна удивительной формы и универсальности уравнений Лагранжа. Но давайте не будем спешить, а сначала познакомимся еще с удивительным свойством инвариантности лагранжевой записи уравнений движения, независимости их вида от выбранных обобщенных координат, в частности, уравнений (14.8) или (14.9). Оказывается, при любой замене переменных вида $x = f(y)$ ($q = f(p)$) вид их сохраняется, чего нет для исходных уравнений Ньютона или Кирхгофа.

Действительно, замена переменных в дифференциальном уравнении Ньютона (14.7) приводит его к виду:

$$m \left(\frac{\partial^2 f}{\partial y^2} \dot{y}^2 + \frac{\partial f}{\partial y} \ddot{y} \right) + k f(y) = 0, \quad (14.10)$$

отличному по форме от исходной записи (14.7). А теперь сделаем такую же замену переменных в лагранжевой записи (14.8). Нам надо убедиться, что из уравнения

$$\frac{d}{dt} \frac{\partial}{\partial \dot{x}} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) - \frac{\partial}{\partial x} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) = 0 \quad (14.11)$$

следует

$$\frac{d}{dt} \frac{\partial}{\partial \dot{y}} \left(\frac{m(\partial f / \partial y \cdot \dot{y})^2}{2} - \frac{k f^2(y)}{2} \right) - \frac{\partial}{\partial y} \left(\frac{m(\partial f / \partial y \cdot \dot{y})^2}{2} - \frac{k f^2(y)}{2} \right) = 0 \quad (14.12)$$

и обратно, т.е. замена переменных в уравнении Лагранжа сводится к замене переменных в функции Лагранжа (функция Лагранжа должна быть функцией принимаемых нами обобщенных координат и скоростей). Для этого достаточно показать, что (14.12) совпадает с (14.10), поскольку (14.10) совпадает с (14.7), а (14.7) – с (14.11). Действительно,

$$\frac{d}{dt} \frac{\partial}{\partial \dot{y}} \left(\frac{m f'^2 \dot{y}^2}{2} - \frac{k f^2}{2} \right) - \frac{\partial}{\partial y} \left(\frac{m f'^2 \dot{y}^2}{2} - \frac{k f^2}{2} \right) = [m(f' \ddot{y} + f'' \dot{y}^2) + k f] \frac{\partial f}{\partial y} = 0$$

и поскольку $\partial f / \partial y \neq 0$, то приходим к уравнению (14.10).

Итак, лагранжева форма записи обладает удивительным свойством инвариантности формы записи по отношению к замене переменных. Эта форма записи не зависит от выбранных переменных, она инвариантна по отношению к ним. Интересно, откуда эта инвариантность? Разгадать это не просто. Но в науке очень часто важно найти зацепку, дальше уже следует «ковырять», пока не поймешь. А что нужно «ковырять», пожалуй, ясно: существует, по-видимому, какой-то надкоординатный подход отделения истинных движений системы среди всех мыслимо возможных. Ведь именно это, в конечном счете, делают дифференциальные уравнения движения.

Не зависящая от координат вариационная формулировка законов механического движения исторически назревала давно и имела своим истоком теологические представления о целесообразности божественного промысла. Об этом много

написано. Наверное, и сейчас к этому трудно прийти чисто логически. Так что не будем лукавить и прямо скажем, в чем дело.

Возьмем интеграл по времени от функции Лагранжа от момента времени t_0 до момента t_1 :

$$\int_{t_0}^{t_1} L dt$$

и назовем эту величину действием. Обозначим ее через W . Распишем детальнее, от чего зависит действие W :

$$W = \int_{t_0}^{t_1} L(q_1(t), \dots, q_n(t); \dot{q}_1(t), \dots, \dot{q}_n; t) dt .$$

Оно определяется функциями $q_\alpha(t)$ ($\alpha = 1, 2, \dots, n$), зависящими от времени t на промежутке от t_0 до t_1 , т.е. является функционалом. Рассмотрим, как меняется этот функционал при малых изменениях функции $q_\alpha(t)$. Изменим функции $q_\alpha(t)$, заменив $q_\alpha(t)$ на $q_\alpha(t) + \delta q_\alpha(t)$, и найдем изменение действия W с точностью до величин первого порядка малости относительно малых изменений $\delta q_\alpha(t)$ и $\delta \dot{q}_\alpha$. Получим:

$$\begin{aligned} \delta W &= \delta \int_{t_0}^{t_1} L dt = \int_{t_0}^{t_1} \left(\sum \frac{\partial L}{\partial q_\alpha} \delta q_\alpha + \frac{\partial L}{\partial \dot{q}_\alpha} \delta \dot{q}_\alpha \right) dt = \int_{t_0}^{t_1} \left(\sum_\alpha \frac{\partial L}{\partial q_\alpha} \delta q_\alpha + \frac{\partial L}{\partial \dot{q}_\alpha} \frac{d}{dt} \delta q_\alpha \right) dt = \\ &= \sum_\alpha \frac{\partial L}{\partial \dot{q}_\alpha} \delta q_\alpha \Big|_{t_0}^{t_1} + \int_{t_0}^{t_1} \sum_\alpha \left(\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} \right) \delta q_\alpha dt . \end{aligned}$$

Последнее действие в этих вычислениях состоит в интегрировании по частям. В результате мы с удивлением обнаруживаем, что под интегралом стоят выражения

$$\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} , \tag{14.13}$$

обращение которых в нуль как раз и есть уравнения Лагранжа. Что же нужно потребовать, чтобы они обратились в нуль? Какое предположение в отношении действия W приведет к такому выводу? Догадаться не очень трудно. Потребуем, чтобы

$$\delta W = 0$$

при условии, что изменений функций $q_\alpha(t)$ на концах не происходит и поэтому $\delta q_\alpha = 0$ при $t = t_0$ и $t = t_1$. Действительно, из этих предположений следует, что

$$\int_{t_0}^{t_1} \left[\sum_\alpha \left(\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} \right) \delta q_\alpha \right] dt = 0 \tag{14.14}$$

для произвольных δq_α ($\alpha = 1, \dots, n$), обращающихся в нуль при $t = t_0$ и $t = t_1$. Это влечет, как нетрудно понять, обращение в нуль всех выражений (14.13) и имеет место только благодаря произвольности δq_α , которые в силу этого можно, в частности, выбрать равными

$$\delta q_\alpha = \left(\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} \right) \varepsilon^2(t),$$

где $\varepsilon^2(t)$ обращается в нуль только на концах замкнутого интервала $[t_0, t_1]$. Для таких вариаций функций $q_\alpha(t)$ из (14.14) следует, что

$$\int_{t_0}^{t_1} \left[\sum_\alpha \left(\frac{\partial L}{\partial q_\alpha} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} \right)^2 \right] \varepsilon^2(t) dt = 0, \quad (14.15)$$

и если бы при каком-либо значении t внутри промежутка $[t_0, t_1]$ одно из выражений (14.13) было отлично от нуля, то и интеграл (14.13) был бы положительным, отличным от нуля.

Итак, уравнения Лагранжа эквивалентны требованию

$$\delta W = 0 \quad (14.16)$$

при любых вариациях функций $q_\alpha(t)$, при которых не меняются их значения при $t = t_0$ и $t = t_1$.

Это и есть желанная вариационная формулировка принципа отбора истинного движения от всех мыслимо возможных. Истинное движение за время от t_0 до t_1 от начальной точки $q_\alpha = q_\alpha^0$ до конечной $q_\alpha = q_\alpha^1$ выделяется из всех мыслимых тем, что для него вариация действия равна нулю.

Можно показать, что при этом на истинном движении не только достигается равенство нулю первой вариации, но действие на нем достигает минимума. Но нам достаточно и того, что мы узнали: мы пришли к формулировке закона движения, не зависящей от выбираемых переменных, от того, как конкретно мы его описываем. Из этой независимости и следует инвариантность лагранжевой записи уравнений движения.

Вместе с тем мы обнаружили, что в качестве математической модели могут выступать не только дифференциальные уравнения движения, но и, пожалуй, более общие вариационные принципы. Более общие, потому что они не зависят от координатного описания. Они как бы над ним. Повторим еще раз, в чем же состоит эта вариационная математическая модель.

В основе ее уже не лежит пространство состояний, ее основа – пространство положений, конфигураций, пространство переменных q_1, \dots, q_n . Для механической системы – это положение всех ее материальных точек. Для электрических, тех, о которых речь шла выше, – это электрические заряды, производные по времени от которых – электрические токи.

Движение в вариационном принципе определяется не начальным состоянием, а исходным и конечным положениями. Во всяком случае, они предполагаются как бы заданными или фиксированными. А сам рассматриваемый нами вариационный принцип состоит в том, что действительное движение от заданной начальной точки к конечной выделяется тем, что для него вариация действия равна нулю.

Можно смотреть на это, как на математический прием новой формы записи дифференциальных уравнений движения, можно видеть в этом новый принцип. Он так и называется принципом наименьшего действия, открытие которого при-

надлежит Гамильтону. Это принцип некоей целесообразности, по идее иной, чем принцип детерминизма. Вообще говоря, из принципа целесообразности не следует принцип детерминизма, но конкретно, применительно к законам природы, пока это всегда так. Это и делает обе точки зрения совместимыми, дополняющими друг друга.

Теперь я хочу привести вам несколько примеров применения уравнений Лагранжа – Максвелла. Они убедят вас в их эффективности, в том числе и в случаях, когда вы не знаете даже, как подступиться к составлению математической модели.

В качестве первого примера возьмем обычный круговой маятник. Для него кинетическая энергия равна $J\dot{\varphi}^2/2$, а потенциальная равна $mg l \cos \varphi$, где φ – угол отклонения маятника от вертикали, J и m – момент инерции и масса маятника, l – расстояние от точки подвеса до центра масс. Поэтому функция Лагранжа равна

$$L = \frac{J\dot{\varphi}^2}{2} + mg l \cos \varphi.$$

После чего легко находим, что

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\varphi}} - \frac{\partial L}{\partial \varphi} = J\ddot{\varphi} + mg l \sin \varphi = 0.$$

Теперь пусть маятник не круговой, а параболический, т.е. масса m может двигаться по расположенной в вертикальной плоскости параболе $z = ax^2$ (рис. 14.1). В этом случае

$$T = \frac{mv^2}{2} = \frac{m(\dot{x}^2 + \dot{y}^2)}{2} = \frac{m}{2}(1 + 4a^2x^2)\dot{x}^2,$$

$$V = mgz = mga^2x^2$$

и

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = m(1 + 4a^2x^2)\ddot{x} + 2mgax = 0.$$

Наконец, пусть материальная точка массы m движется в вертикальной плоскости Oxz (рис. 14.2) по произвольной кривой, уравнение которой запишем в параметрическом виде:

$$x = x(s), \quad z = z(s),$$

где параметр s – длина дуги кривой, отсчитываемая от некоторой точки O .

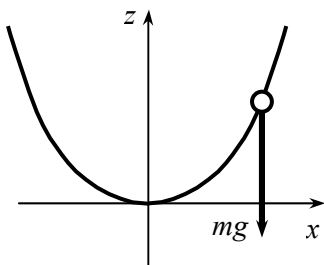


Рис. 14.1. Параболический математический маятник

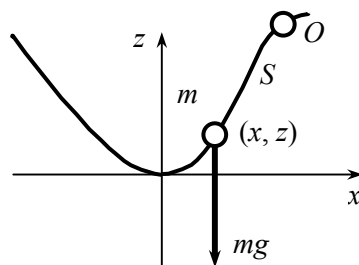


Рис. 14.2. Иллюстрация к движению материальной точки по произвольной гладкой плоской кривой в поле силы тяжести

Непосредственно находим, что

$$L = \frac{m\dot{s}^2}{2} + mgz(s),$$

и поэтому

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{s}} - \frac{\partial L}{\partial s} = m\ddot{s} + mgz'(s) = 0.$$

Возьмем электрический контур, состоящий из индуктивности L и емкости C . При этом

$$L = \frac{L\dot{q}^2}{2} - \frac{q^2}{2C},$$

и поэтому

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = L\ddot{q} + \frac{q}{C} = 0,$$

что нам хорошо известно.

Усложним задачу, предположив, что внутри катушки индуктивности может качаться подвешенный на пружине железный стержень массы m , так что ее индуктивность L зависит от отклонения x стержня от его равновесного положения $x = 0$ (рис. 14.3).

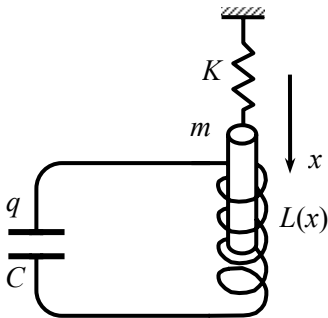


Рис. 14.3. Взаимодействующие между собой механический и электрический осцилляторы

Эта система довольно сложна. Она состоит из двух осцилляторов (замкнутого контура, состоящего из индуктивности и емкости, и массы m , подвешенной на пружине), которые взаимодействуют друг с другом. Движение стержня меняет магнитное поле катушки индуктивности, вызывая в ней ЭДС индукции, а меняющееся магнитное поле катушки индуктивности воздействует на железный сердечник. Для того чтобы написать уравнения движений этих осцилляторов, нужно найти индуцируемую в катушке индуктивности движением стержня ЭДС и меняющуюся силу воздействия магнитного поля катушки индуктивности на находящийся в ней железный стержень. Это для математика не так уж просто. А написать функцию Лагранжа – Максвелла очень легко.

Помимо координаты x , введем еще заряд конденсатора q , тогда

$$L = \frac{L(x)\dot{q}^2}{2} + \frac{m\dot{x}^2}{2} - \frac{q^2}{2C} - \frac{kx^2}{2}$$

и приходим к двум уравнениям для координат x и q как функций времени:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = m\ddot{x} - \frac{1}{2} L'(x) \dot{q}^2 + kx = 0,$$

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = L(x) \ddot{q} + L'(x) \dot{x} \dot{q} + \frac{q}{C} = 0.$$

Из этих уравнений, как и следовало ожидать, вытекает закон сохранения энергии в виде:

$$\frac{L(x) \dot{q}^2}{2} + \frac{m \dot{x}^2}{2} + \frac{q^2}{2C} + \frac{kx^2}{2} = h = \text{const}.$$

Получается это соотношение умножением первого уравнения на \dot{x} , второго на \dot{q} , затем их сложения и интегрирования.

Перепишем найденные выше уравнения Лагранжа – Максвелла в виде:

$$m\ddot{x} + kx = \frac{1}{2} L' \dot{q}^2,$$

$$L\ddot{q} + \frac{q}{C} = -L' \dot{x} \dot{q}.$$

При $L' = 0$ они распадаются на два независимых осциллятора – механический и электрический. Их взаимодействие тем сильнее, чем больше L' . При $L' \neq 0$ электрический осциллятор действует на массу механического с силой $L' \dot{q}^2 / 2$, а механический осциллятор наводит в электрическом контуре ЭДС индукции, равную $-L' \dot{x} \dot{q}$.

Вы, вероятно, уже заметили, что во всех приведенных примерах отсутствуют механические и электрические сопротивления. Это потому, что уравнения Лагранжа предполагают потенциальность действующих сил. При независимости потенциальной функции от времени из этого следует, как и в рассмотренном примере, закон сохранения энергии.

При непотенциальных силах и, в частности, при наличии вязких трений и омических сопротивлений, энергия не сохраняется и уравнения Лагранжа не применимы. Но их можно обобщить. Это несложно. Нужно только исходить не из потенциальной энергии действующих сил и энергии электрического поля, а из выражения элементарной работы δA . При наличии потенциальной энергии V элементарная работа δA запишется:

$$\delta A = -\delta V = -\sum \frac{\partial V}{\partial q_\alpha} \delta q_\alpha. \quad (14.17)$$

В общем же случае (точнее, более общем)

$$\delta A = \sum Q_\alpha \delta q_\alpha, \quad (14.18)$$

т.е. производные $-\partial V / \partial q_\alpha$ заменяются на величины Q_α , называемые обобщенными силами. К этой формальной замене, как оказывается, сводится обобщение уравнений Лагранжа. Запишем уравнения Лагранжа в виде:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_\alpha} - \frac{\partial T}{\partial q_\alpha} = -\frac{\partial V}{\partial q_\alpha}.$$

Их обобщение будет

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_\alpha} - \frac{\partial T}{\partial q_\alpha} = Q_\alpha. \quad (14.19)$$

Заметим еще, что, разделяя потенциальные и непотенциальные силы, можно уравнение Лагранжа записать в виде:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_\alpha} - \frac{\partial T}{\partial q_\alpha} = -\frac{\partial V}{\partial q_\alpha} + Q_\alpha$$

или

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} - \frac{\partial L}{\partial q_\alpha} = Q_\alpha, \quad (14.20)$$

где по-прежнему $L = T - V$, но V относится только к потенциальным силам, а непотенциальные силы представлены обобщенными силами Q_α .

Проиллюстрируем сначала уравнения (14.19) и (14.20) очень простым примером массы на пружине при наличии вязкого трения. В этом случае кинетическая энергия T равна $m\dot{x}^2/2$. Сил, совершающих работу при виртуальном изменении x , две – сила упругости $-kx$ и сила трения $-h\dot{x}$. Их работа на перемещении δx равна

$$\delta A = -kx \delta x - h\dot{x} \delta x,$$

поэтому соответствующая координате x обобщенная сила (в данном случае просто сила) равна

$$Q = kx - h\dot{x}.$$

Уравнение Лагранжа (14.19) принимает вид:

$$m\ddot{x} = -kx - h\dot{x}.$$

Составляя уравнения движения в форме (14.20), потенциальную силу жесткости относим к функции Лагранжа L , равной

$$L = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2},$$

а обобщенная сила вязкого трения Q будет $-h\dot{x}$. В силу этого уравнение (14.20) принимает вид:

$$\frac{d}{dt} \frac{\partial}{\partial \dot{x}} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) - \frac{\partial}{\partial x} \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2} \right) = -h\dot{x}$$

или

$$m\ddot{x} + kx = -h\dot{x}.$$

Для электрического контура из индуктивности, емкости и омического сопротивления все аналогично, так как работа диссипативных сил омического сопротивления при переносе заряда δq равна $-R\dot{q} \delta q$.

15. ЧАСЫ ГАЛИЛЕЯ – ГЮЙГЕНСА

Исчисление времени – древнейшая проблема человечества. Миллионы лет, столетия, годы, месяцы, сутки, часы, минуты, секунды и кратчайшие ее доли – наносекунды – это те разнообразные масштабы, в которых для разных целей научился измерять время человек. Первоначальное измерение времени основывалось на обращении Земли вокруг Солнца и ее собственном вращении. Знаки Зодиака указывали времена года, а Солнце днем и звезды ночью – время суток.

Вызванная мореплаванием потребность определения местоположения корабля потребовала точного измерения времени суток. Широта определялась только по астрономическим наблюдениям. Для определения долготы требовалось еще и точное знание времени. Определение местоположения по долготе с точностью в один километр соответствует ошибке в отсчете времени примерно в три секунды. А плавать приходилось, не имея возможности откорректировать показания корабельных часов, месяц и более.

В XVII веке английское адмиралтейство объявило конкурс на создание точных часов. Такие часы были сконструированы и изготовлены Х. Гюйгенсом. Это были первые маятниковые часы – часы, в которых для отсчета времени использовались колебания маятника. Галилей изучил колебания маятника, установил неизменность периода их колебаний (при малых отклонениях маятника изменения частоты его колебаний Галилей заметить не мог). Гюйгенс создал часы на этой основе. Поэтому эти новые часы можно назвать часами Галилея – Гюйгенса. До Галилея и Гюйгенса были часы, использующие колебания балансира, однако этот балансир не имел собственной частоты колебаний, его частота колебаний зависела и определялась прилагаемыми к нему усилиями часового механизма.

Использование в часах устройства с собственной частотой колебаний оказалось необычайно плодотворным и привело к серии разнообразных конструкций все более и более точных часов, использующих этот принцип: часы с балансиром на пружинке с цилиндрическим, анкерным и хронометровым ходами; кварцевые часы, использующие собственные упругие колебания кварцевой пластинки; часы, использующие электрический колебательный контур с собственной частотой и, наконец, феноменально точные молекулярные и атомные часы, позволяющие обнаружить и измерить неравномерность вращения Земли [11].

Для того чтобы колеблющийся маятник сделать часами, нужно считать его колебания и показывать их число, например, стрелкой на циферблате в соответствующем масштабе. Кроме того, нужно поддерживать его колебания, которые сами по себе затухают. Необходимо затухающие колебания маятника превратить в незатухающие автоколебания.

Обе эти задачи успешно решает так называемый часовой ход, простейший вариант которого схематически изображен на рис. 15.1. После каждого колебания маятника туда и обратно храповик под влиянием заведенной пружины или опускающегося груза поворачивается на один зубец и одновременно сообщает маятнику подталкивающий импульс. Тем самым скорость вращения храпового колеса

определяется частотой колебаний маятника, а его зубцы в момент его поворота подталкивают маятник, поддерживая его колебания.

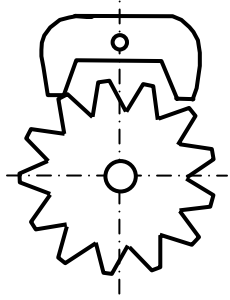


Рис. 15.1. Схематическое изображение ходового механизма часов

Точность хода часов с таким устройством определяется стабильностью колебаний маятника. Период колебаний маятника зависит от многих причин, и каждая из них может повлечь его изменение. Основными факторами, определяющими период колебаний маятника, являются его длина, масса, момент инерции, сопротивление среды и трение в точке подвеса, величина ускорения силы тяжести, удары храпового колеса, а также амплитуда его колебаний.

Рассмотрим сначала зависимость периода колебаний маятника от его амплитуды. Запишем уравнение колебаний маятника в виде

$$J\ddot{\varphi} + mgl \sin \varphi = 0, \quad (15.1)$$

где J – момент инерции маятника, m и l – масса и длина (расстояние от оси подвеса до центра масс). Введя обозначение $\omega^2 = mgl/J$, умножая уравнение (15.1) на $\dot{\varphi}$ и интегрируя, найдем, что

$$\frac{\dot{\varphi}^2}{2} - \omega^2 \cos \varphi = h, \quad (15.2)$$

где h – константа интегрирования. Полагая угловую амплитуду маятника равной α , имеем, что при $\varphi = \alpha$ $\dot{\varphi} = 0$. Выражаем, согласно (15.2), константу интегрирования через α и (15.2) записываем в виде:

$$\frac{\dot{\varphi}^2}{2} - \omega^2 \cos \varphi = -\omega^2 \cos \alpha$$

или

$$\dot{\varphi} = \omega \sqrt{2(\cos \varphi - \cos \alpha)}. \quad (15.3)$$

Из уравнения (15.3) следует, что

$$\frac{d\varphi}{\omega \sqrt{2(\cos \varphi - \cos \alpha)}} = dt. \quad (15.4)$$

Пусть период колебания маятника при амплитуде α равен T . Тогда при изменении угла отклонения маятника от нуля до α пройдет четверть периода и, согласно (15.4), найдем, что

$$\frac{T}{4} = \int_0^{\alpha} \frac{d\varphi}{\omega \sqrt{2(\cos \varphi - \cos \alpha)}}.$$

Этот интеграл не вычисляется в элементарных функциях. Он может быть найден численно. При малых α , что нас устраивает, можно найти его разложение в ряд по параметру α . Если ограничиться точностью до α^2 , то это представление находится следующим образом:

$$\begin{aligned}
 [2(\cos \varphi - \cos \alpha)]^{-1/2} &= \left\{ 2 \left[\left(1 - \frac{\varphi^2}{2} + \frac{\varphi^4}{24} - \dots \right) - \left(1 - \frac{\alpha^2}{2} + \frac{\alpha^4}{24} - \dots \right) \right] \right\}^{-1/2} = \\
 &= (\alpha^2 - \varphi^2)^{-1/2} \left(1 - \frac{\varphi^2 + \alpha^2}{12} + \dots \right)^{-1/2} = (\alpha^2 - \varphi^2)^{-1/2} \left(1 + \frac{\varphi^2 + \alpha^2}{24} + \dots \right),
 \end{aligned}$$

и поэтому

$$T = \frac{4}{\omega} \int_0^\alpha \left(1 + \frac{\varphi^2 + \alpha^2}{24} + \dots \right) \frac{d\varphi}{\sqrt{\alpha^2 - \varphi^2}}$$

или после подстановки $\varphi = \alpha \sin \theta$

$$T = \frac{4}{\omega} \int_0^{\pi/2} \left(1 + \frac{\alpha^2(1 + \sin^2 \theta)}{24} + \dots \right) d\theta = \frac{2\pi}{\omega} \left(1 + \frac{\alpha^2}{16} + \dots \right). \tag{15.5}$$

О точности этой формулы можно судить по графику на рис. 15.2, где сплошной кривой изображена численно определенная величина, стоящая в скобках правой части формулы (15.5), т.е. $\omega T/2\pi$, и пунктиром – график первых двух членов этой скобки, т.е. $1 + \alpha^2/16$. Из них видно, что при $\alpha < \pi/2$ эти графики на рисунке неразличимы.

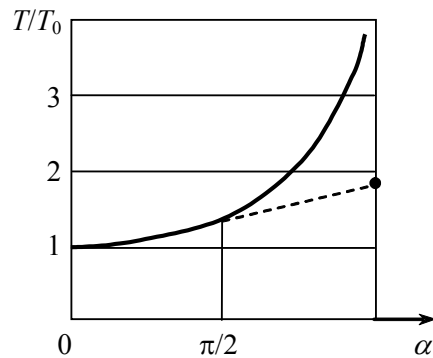


Рис. 15.2. График зависимости периода T колебаний маятника от его амплитуды

Из самой формулы (15.5) видно, что при малых амплитудах α изменение периода колебаний маятника невелико. Так, при $\alpha \leq 0,1$ изменение периода не более чем 0,7%. А изменение периода T при изменении амплитуды α на $\Delta\alpha$ таково, что

$$\frac{dT}{T} \approx \frac{\alpha \Delta\alpha}{8}.$$

При $\alpha = 0,1$ и требуемой точности $dT/T = 10^{-5}$ $\Delta\alpha$ не должно превосходить $4 \cdot 10^{-4}$, т.е. требуется очень маленькая амплитуда колебаний маятника. В связи с этим в свое время задумывались, нельзя ли сделать маятник, у которого период колебаний не зависел бы от амплитуды. Такой маятник был придуман. Его идея состояла в том, чтобы маятник качался не по окружности, а по какой-нибудь дру-

гой кривой, чего можно достичь, как это показано на рис. 15.3, где П – гибкая пластинка, а Н – ограничивающие ее движение направляющие. При этом линия L, направленная вдоль подвеса маятника, все время касается направляющей Н.

Пусть искомая кривая, по которой должен качаться маятник (материальная точка массы m), имеет уравнение

$$x = x(s), \quad y = y(s), \tag{15.6}$$

где s – длина дуги, отсчитываемая от нижней точки кривой (рис. 15.4).

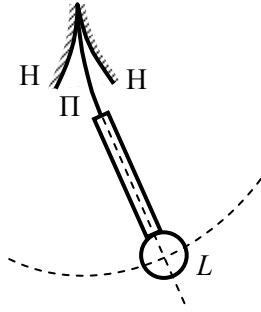


Рис. 15.3. Ограничивающее препятствие Н, позволяющее изменить траекторию колебаний маятника

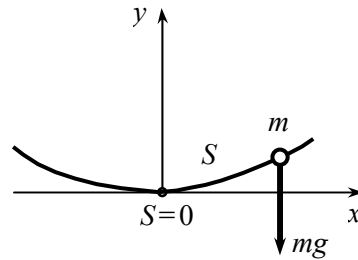


Рис. 15.4. Иллюстрация к поиску кривой, по которой должен качаться маятник, чтобы его период колебаний не зависел от амплитуды

Составим уравнения движения массы m . Находим функцию Лагранжа

$$L = \frac{m}{2} \dot{s}^2 - mgy(s).$$

Откуда

$$m\ddot{s} + mgy'(s) = 0. \tag{15.7}$$

Решение этого уравнения будет периодическим с периодом, не зависящим от начальных условий, а следовательно, и амплитуды, если оно совпадает с уравнением линейного гармонического осциллятора, т.е. когда

$$gy'(s) = \omega^2 s \tag{15.8}$$

при произвольном ω . Решая это уравнение, находим, что

$$y = \frac{\omega^2}{2g} s^2, \tag{15.9}$$

где произвольная постоянная отсутствует, так как, согласно рис. 15.3, при $s = 0$ $y = 0$.

Осталось найти функцию $x(s)$. Это легко сделать, исходя из соотношения

$$dx^2 + dy^2 = ds^2$$

или

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1. \tag{15.10}$$

Из (15.9) и (15.10) находим, что

$$\frac{dx}{ds} = \sqrt{1 - \left(\frac{\omega^2}{2g}\right)^2 s^2} \quad (15.11)$$

или

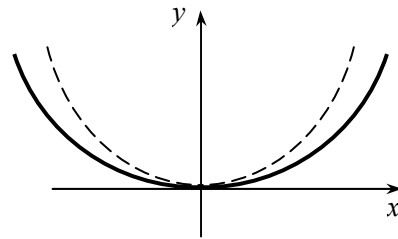
$$x = \int \sqrt{1 - \left(\frac{\omega^2}{2g}\right)^2 s^2} ds = \int \sqrt{1 - a^2 s^2} ds. \quad (15.12)$$

Используя замену $s = a^{-1} \sin \varphi$, вычисляем интеграл (15.12) и, учитывая (15.9), приходим к параметрическим уравнениям искомой кривой:

$$x = \frac{g}{2\omega^2} \varphi + \frac{g}{4\omega^2} \sin 2\varphi, \quad y = \frac{2g}{\omega^2} \sin^2 \varphi. \quad (15.13)$$

Это хорошо известная циклоида, ее вид показан на рис. 15.5. Она отличается от окружности тем, что поднимается вверх от точки $x = y = 0$ все быстрее и быстрее, чем окружность, тем самым ускоряется скатывание по ней маятника при увеличении его отклонения. Это отличие, естественно, при малых φ невелико, как невелико и отличие уравнения колебаний кругового маятника от уравнения линейного гармонического осциллятора, т.е. как отличие $\sin \varphi$ от φ .

Рис. 15.5. Сопоставление траекторий движения кругового и циклоидального маятника (пунктирная кривая)



Мы рассмотрели зависимость периода колебаний маятника от его амплитуды колебаний, оценили влияние изменения амплитуды на точность хода часов и обнаружили, что от этой зависимости и возможной причины уменьшения точности хода часов можно избавиться, если круговой маятник заменить циклоидным. Однако изменение амплитуды далеко не единственная причина, от которой зависит точность хода часов. У маятника может меняться длина, при значительных перемещениях часов может меняться ускорение силы тяжести. Мы никак не учли наличие трения, вязкого и сухого. Нестабильность периода может внести и механизм подталкивания маятника, поддерживающий его колебания.

Таким образом, нестабильность и потеря точности хода часов могут возникнуть от многих причин, и если мы хотим создать точные часы, то нам надлежит изучить все эти факторы нестабильности, оценить их величину и найти пути устранения.

Все эти факторы нестабильности в силу их малости можно изучить отдельно, независимо друг от друга, учитывая не все вместе, а каждое отдельно. Это значительное упрощение можно обосновать следующим образом. Период колебаний T является функцией многих параметров часового механизма u_1, u_2, \dots, u_m , так что

$$T = T(u_1, u_2, \dots, u_m).$$

Пусть $u_1^*, u_2^*, \dots, u_m^*$ – номинальные значения параметров, а их возможные изменения – $\delta u_1, \delta u_2, \dots, \delta u_m$. Тогда с точностью до этих малых изменений

$$\delta T = \left(\frac{\partial T}{\partial u_1} \right)^* \delta u_1 + \left(\frac{\partial T}{\partial u_2} \right)^* \delta u_2 + \dots + \left(\frac{\partial T}{\partial u_m} \right)^* \delta u_m,$$

где производные от функции T вычисляются при номинальных значениях параметров, т.е. производная $\partial T / \partial u_s$ вычисляется в предположении отсутствия зависимости от остальных параметров $u_1^*, u_2^*, \dots, u_m^*$, кроме u_s^* .

В силу этого с принимаемой точностью зависимость от длины, вязкого трения и ускорения силы тяжести можно вычислить, исходя из упрощенной модели колебаний маятника, не учитывающей зависимость его периода от амплитуды, т.е. из уравнения линейного осциллятора

$$\ddot{\varphi} + 2\delta\dot{\varphi} + \omega^2\varphi = 0 \quad \left(\omega^2 = \frac{mgl}{J} \right), \quad (15.14)$$

период колебаний которого равен

$$T = \frac{2\pi}{\sqrt{\omega^2 - \delta}}. \quad (15.15)$$

Ради простоты примем, что основная масса маятника сосредоточена в одном месте. Тогда $J = ml^2$ и (15.15) принимает вид:

$$T = \frac{2\pi}{\sqrt{g/l - \delta^2}} = 2\pi \sqrt{\frac{l}{g}} \left(1 + \frac{1}{2} \frac{l\delta^2}{g} + \dots \right),$$

из которого находим оценку

$$\frac{dT}{T} = \frac{1}{2} \frac{dl}{l} - \frac{1}{2} \frac{dg}{g} + \frac{l\delta d\delta}{g}. \quad (15.16)$$

Эта оценка указывает на сильную зависимость точности хода от изменений длины и силы тяжести, а также от вязкого трения, если оно не очень мало.

Длина маятника меняется в основном под влиянием изменения температуры. Поэтому этот факт потери точности хода часов можно устранить, поддерживая окружающую температуру постоянной. Но есть и другой способ поддержания постоянства l , несмотря на изменение температуры. Он состоит в особой конструкции маятника, показанной на рис. 15.6. При увеличении температуры длины l_1 и l_2 увеличиваются на δl_1 и δl_2 , а длина маятника увеличивается при этом на $\delta l_1 - \delta l_2$. Стержни длин l_1 и l_2 изготовлены из разных материалов с коэффициентами теплового расширения α_1 и α_2 , подобранными так, чтобы величина удлинения, равная

$$\delta l = \delta l_1 - \delta l_2 = (\alpha_1 l_1 - \alpha_2 l_2) \delta T,$$

где δT – изменение температуры, равнялась нулю, т.е. так, чтобы $\alpha_1 l_1 - \alpha_2 l_2 = 0$.

С зависимостью от ускорения силы тяжести, если часы значительно меняют свое местонахождение, например, находясь на корабле, вроде бы ничего сделать нельзя. Можно лишь скомпенсировать ее соответствующими вычислительными

поправками или заменить маятник часов балансиrom на пружинке, что спустя некоторое время и было сделано: часы Гюйгенса сменил корабельный хронометр.

Оценим еще влияние сухого трения. Для этого, согласно сказанному, можно ограничиться рассмотрением модели

$$\ddot{\varphi} + \omega^2 \varphi = -f \operatorname{sign} \dot{\varphi}, \tag{15.17}$$

где f – параметр, отражающий наличие сухого кулоновского трения. Как мы сейчас увидим, сухое трение приводит к затуханиям колебаний, но их период не меняется. Действительно, на каждом полупериоде колебаний $\dot{\varphi}$ не меняет знака и поэтому, например, при $\dot{\varphi} > 0$ (15.17) записывается в виде

$$\ddot{\varphi} + \omega^2 \varphi = -f$$

или, после замены φ на $\varphi - f/\omega^2$, в виде

$$\ddot{\varphi} + \omega^2 \varphi = 0,$$

где явная зависимость от трения f исчезла, т.е. ее нет.

Перейдем к рассмотрению механизма поддержания колебаний маятника и его влияния на стабильность хода часов.

Часовые ходы весьма различны по конструкции (цилиндрический ход, анкерный, хронометровый и др.), и их точное описание довольно сложно. Примем идеализированную схему часового хода, считая, что он осуществляет подталкивание при угле $\varphi = \alpha$ ($\dot{\varphi} > 0$), сообщая маятнику импульс p , а вне этого подталкивающего импульса движение маятника подчиняется уравнению

$$\ddot{\varphi} + 2\delta\dot{\varphi} + \omega^2 \varphi = 0. \tag{15.18}$$

Толчок происходит при $\varphi = \alpha$ и $\dot{\varphi} > 0$, и в результате его скорость $\dot{\varphi}$ мгновенно возрастает на величину p . Так, что

$$\dot{\varphi}_+ = \dot{\varphi}_- + p, \tag{15.19}$$

где $\dot{\varphi}_-$ – доударная угловая скорость, а $\dot{\varphi}_+$ – послеударная. Угол φ при этом измениться не успеваеt. Нарисуем фазовую траекторию, отвечающую одному колебанию от точки $\varphi = \alpha$ снова к $\varphi = \alpha$. Пусть начальная точка M_0 , а M_1 – точка повторного значения $\varphi = \alpha$ при условии, что движение происходит в соответствии с дифференциальным уравнением (15.18). В момент прихода в точку M_1 происходит передача импульса p и точка M_1 перемещается вверх вдоль оси $\dot{\varphi}$ на расстояние p (рис. 15.7) в точку M_2 .

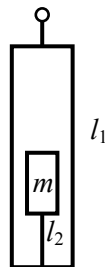


Рис. 15.6. Схема устройства подвеса маятника, делающего его длину не зависящей от температуры

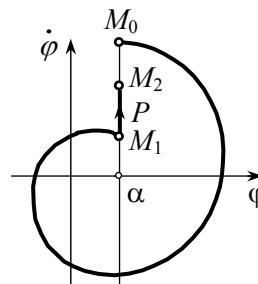


Рис. 15.7. Фазовая траектория часов и способ ее изучения путем сведения к точечному отображению точки M_0 в точку M_2

После этого весь описанный цикл повторяется (за исключением случая, когда выходящая из точки M_2 фазовая траектория больше не пересекает линии $\varphi = \alpha$).

Найдем аналитические зависимости описанного движения от фазовой точки M_0 к M_1 и затем к M_2 . Пусть точки M_0 , M_1 и M_2 имеют координаты $\varphi_0 = \alpha$ и $\dot{\varphi}_0$, $\varphi_1 = \alpha$ и $\dot{\varphi}_1$, $\varphi_2 = \alpha$ и $\dot{\varphi}_2$. Точки M_0 и M_1 связаны, согласно (13.7), зависимостями:

$$\begin{aligned} \varphi_1 &= e^{-\delta\tau} \left(\varphi_0 \cos \Omega\tau + \frac{\dot{\varphi}_0 + \delta\varphi_0}{\Omega} \sin \Omega\tau \right), \\ \dot{\varphi}_1 &= e^{-\delta\tau} \left[-\delta \left(\varphi_0 \cos \Omega\tau + \frac{\dot{\varphi}_0 + \delta\varphi_0}{\Omega} \sin \Omega\tau \right) - \Omega\varphi_0 \sin \Omega\tau + (\dot{\varphi}_0 + \delta\varphi_0) \cos \Omega\tau \right], \end{aligned} \quad (15.20)$$

где τ – время движения от точки M_0 до M_1 , а $\Omega = \sqrt{\omega^2 - \delta^2}$. Для точек M_1 и M_2 , согласно (15.19),

$$\varphi_2 = \varphi_1, \quad \dot{\varphi}_2 = \dot{\varphi}_1 + p. \quad (15.21)$$

При этом следует еще иметь в виду, что

$$\varphi_0 = \varphi_1 = \varphi_2 = \alpha. \quad (15.22)$$

Из (15.20)–(15.22), вводя обозначения $\dot{\varphi}_0 = u$ и $\dot{\varphi}_2 = \bar{u}$, находим, что

$$\begin{aligned} \alpha &= e^{-\delta\tau} \left(\alpha \cos \Omega\tau + \frac{u + \delta\alpha}{\Omega} \sin \Omega\tau \right), \\ \bar{u} &= p + e^{-\delta\tau} \left(u \cos \Omega\tau + \frac{u + \delta\alpha - \Omega^2\alpha}{\Omega} \sin \Omega\tau \right). \end{aligned} \quad (15.23)$$

Первое соотношение (15.23) определяет время τ в зависимости от u , а второе позволяет найти \bar{u} .

Анализ этих уравнений достаточно сложен, но при $\alpha = 0$ он существенно упрощается, так как в этом случае формулы (15.23) принимают вид:

$$\sin \Omega\tau = 0 \quad (u \neq 0), \quad \bar{u} = e^{-\delta\tau} u \cos \Omega\tau + p,$$

из которого следует, что $\tau = 2\pi/\Omega$ (взяв второй по величине положительный корень первого уравнения, так как для первого корня соответствующее $\dot{\varphi}_1 < 0$, это же следует из рис. 15.7) и что

$$\bar{u} = e^{-2\pi\delta/\Omega} u + p. \quad (15.24)$$

Прежде чем проанализировать этот необычайно простой результат, покажем, что именно он в первую очередь достоин внимания. Дело в том, что именно при $\alpha = 0$ исчезает зависимость периода колебаний τ от скорости u . Он постоянен и равен $2\pi/\omega$. Напротив, при $\alpha \neq 0$ такая зависимость есть и следует из уравнений (15.23). Действительно, дифференцируя по u первое уравнение (15.23), находим, что

$$0 = -e^{-\delta\tau} \frac{d\tau}{du} \left(\alpha \cos \Omega\tau + \frac{u + \delta\alpha}{\Omega} \sin \Omega\tau \right) +$$

$$+ e^{-\delta\tau} \left[-\alpha\Omega \sin\Omega\tau \frac{d\tau}{du} + \frac{1}{\Omega} \sin\Omega\tau + (u + \delta\alpha) \cos\Omega\tau \frac{d\tau}{du} \right],$$

откуда

$$\frac{d\tau}{du} = \frac{\sin\Omega\tau}{\Omega[(\alpha - u - \delta\alpha) \cos\Omega\tau + ((u + \delta\alpha)/\Omega + \alpha\Omega) \sin\Omega\tau]}$$

и, следовательно, $d\tau/du$ не обращается в нуль, поскольку τ отлично от $2\pi/\omega$.

Этот же факт можно обнаружить и непосредственно из вида фазовой траектории (рис. 15.7), причем из него видно, что $\tau > 2\pi/\omega$, так как на нем $\alpha > 0$. Напротив, при $\alpha < 0$ видно, что $\tau < 2\pi/\omega$. Сказанное становится очевидным, если заметить, что любой луч $\dot{\phi} = a\varphi$ любая фазовая траектория повторно пересекает через одно и то же время $2\pi/\omega$.

Вернемся к найденному ранее соотношению (15.24), определяющему связь между последовательными угловыми скоростями, с которыми после подталкивания маятник проходит через вертикальное равновесное положение. Этому соотношению, связи между u и \bar{u} , можно придать очень простую, полезную во многих случаях геометрическую интерпретацию. Возьмем прямоугольную систему координат, откладывая на ее осях u и \bar{u} , и построим в ней график зависимости (15.24) (рис. 15.8). Дополним его биссектрисой и найдем по u \bar{u} и затем перенесем его снова на ось u , как показано на том же рисунке.

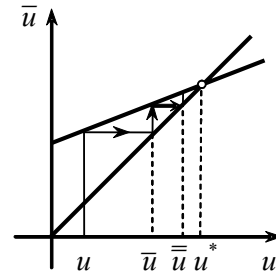


Рис.15.8. Диаграмма Ламеря для часов Галилея – Гюйгенса: u^* – устойчивая неподвижная точка, соответствующая автоколебаниям часов; $u, \bar{u}, \bar{\bar{u}}, \dots$ – последовательные приближения к точке u^*

Повторяя это построение, найдем последовательно $\bar{u}, \bar{\bar{u}}$ и т.д. Из него видно, что последовательные точки $u, \bar{u}, \bar{\bar{u}}, \dots$ стремятся к точке u^* , отвечающей пересечению кривой (15.24) с биссектрисой. Само же значение u^* переходит в себя, поскольку $\bar{u}^* = u^*$, и, следовательно, согласно (15.24),

$$u^* = e^{-2\pi\delta/\Omega} u^* + p$$

и

$$u^* = \frac{p}{1 - e^{-2\pi\delta/\Omega}}. \tag{15.25}$$

Таким образом, при подталкивающих ударах, сообщающих маятнику постоянный импульс при прохождении им состояния равновесия, его колебания асимптотически приближаются к периодическим с периодом $2\pi/\omega$ и амплитудой по скорости, равной u^* . Тем самым затухающие колебания маятника превращаются в автоколебания. Достигается это с помощью механизма – часового хода, осуществляющего обратную связь, полнее – управления движением маятника с помощью обратной связи. Общая идея управления и обратной связи имеют в науке и техни-

ке огромное значение. Мы еще не раз с ней столкнемся, а сейчас ограничимся объяснением, в чем же состоит управление и почему оно реализуется с помощью обратной связи.

Управление состоит в том, что мы действуем на маятник импульсами, имеющими сторонний источник энергии. Эти импульсы сообщаются маятнику тогда, когда он проходит через свое состояние равновесия, т.е. влияние на маятник осуществляется в зависимости от движения самого маятника. Маятник как бы действует сам на себя, осуществляя обратную связь и используя при этом стороннюю энергию пружинного завода часов или опускающегося груза. В результате этого не толчки определяют движение маятника, а сам маятник поддерживает свои колебания, управляя толчками, и часы Гюйгенса оказываются автоколебательной автономной (не зависящей явно от времени) динамической системой.

Вообще управление используется человеком и существует в природе для реализации каких-то целей. Во многих случаях оно реализуется как корректировка того, что происходит с объектом управления, т.е. с помощью обратной связи и некоторого необходимого ей источника энергии. Но возможны и случаи, когда управление осуществляет намеченный заранее план или подчиняется каким-то командам извне.

В чем же секрет точности часов Гюйгенса? Ведь и до него были часы, использующие часовой ход, колеблющий балансир. Но Гюйгенс придумал балансир собственной частоты колебаний и этим колоссально увеличил точность хода часов. При этом частоту колебаний маятника уже определяли не толчки «туда-обратно» часового хода, а собственная частота его колебаний.

Маятник – конкретный частный вид осциллятора с собственной частотой – стал определяющим фактором отсчета времени, и стабильность его собственных колебаний определяла точность часов. А стабильность колебаний осциллятора может быть очень высокой, и тем большей, чем незначительней подталкивающие его импульсы, т.е. чем меньше его затухание, и поэтому еще более точные часы – это часы, в которых вместо маятника используется упруго колеблющаяся кварцевая пластинка или, наконец, молекулы и атомы. Молекулярные и атомные часы уже настолько точны, что позволяют регистрировать неравномерность вращения Земли.

В заключение несколько слов о соотношении (15.24) с диаграммой рис. 15.8.

Способ, с помощью которого мы пришли к соотношению (15.24), носит общий характер. Его придумал А. Пуанкаре, и он широко используется в теории колебаний и теории динамических систем. В общей формулировке он состоит в сведении описания и исследования движений динамической системы к составлению и исследованию точечного отображения. С этой общей точки зрения соотношение (15.24) – это точечное отображение прямой в себя, преобразующее любую ее точку с координатой u в другую точку с координатой \bar{u} . Именно это точечное отображение строится изложенным выше способом на рис. 15.8. Диаграмма рис. 15.8 впервые использовалась Ламереем, а точечное отображение прямой в прямую исследовалось Кенигсом. В связи с этим диаграмма рис. 15.8 получила наименование диаграммы Кенигса – Ламерея. Это название ей дал А.А. Андронов, который усовершенствовал и широко использовал ее для решения ряда актуальных задач теории колебаний и автоматического регулирования [2, 3, 25].

Вот, пожалуй, для первого знакомства с методом Пуанкаре и точечными отображениями и все. Дальнейшее вы сможете узнать из соответствующих курсов теории управления и теории динамических систем [36].

16. ГЕНЕРАТОР ЭЛЕКТРИЧЕСКИХ КОЛЕБАНИЙ

Электрическим аналогом маятника является колебательный контур, составленный из емкости и самоиндукции. Так же как в часах Галилея–Гюйгенса, колебания маятника поддерживаются подталкиваниями часового механизма, можно поддерживать колебания электрического контура с помощью специальных устройств – электронной лампы или полупроводникового триода. Эти устройства должны осуществлять подкачку энергии для того, чтобы компенсировать ее рассеяние на неизбежном омическом сопротивлении. Подталкивающее воздействие на маятник часов должно согласовываться с его фазой и в этом смысле имеет управляемый характер. Аналогично этому управляемый характер имеет и поддержка колебаний электрического контура – генератора электрических колебаний, осуществляемая электронной лампой или полупроводниковым триодом через взаимную индукцию M . Электрическая схема одного из вариантов такой управляемой подкачки энергии изображена на рис. 16.1. Входящая в нее электронная лампа имеет катод, анод и управляющую сетку. Напряжением на сетке можно, практически не затрачивая энергии, менять сопротивление лампы от катода к аноду. Поэтому, если между катодом и сеткой приложить некоторое напряжение, то текущий от катода к аноду ток (поток электронов) I_a будет определяться напряжением на сетке E_g . Эта зависимость имеет вид, представленный на графике рис. 16.2. Со школы вы, наверное, помните, почему график зависимости $I_a(E_g)$ имеет именно такой вид: раскаленный катод излучает электроны, и напряжение E_g на сетке, помещенной между катодом и анодом, помогает или препятствует их попаданию на анод, т.е. увеличивает или уменьшает анодный ток.

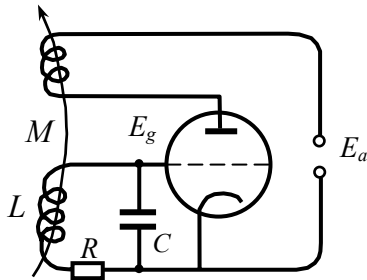


Рис. 16.1. Схема лампового генератора

Напишем математическую модель для электрической схемы рис. 16.1. Пусть I_a – анодный ток, E_g – напряжение на сетке (относительно катода), а q – заряд конденсатора. Пусть еще C , R и L – соответственно емкость конденсатора, величина сопротивления (резистора) и самоиндукция катушки, а M – ее коэффициент взаимной индукции с катушкой в анодном контуре. Для колебательного контура, состоящего из самоиндукции L , емкости C и сопротивления R , имеем:

$$I = \dot{q}, \quad E_g = \frac{q}{C},$$

$$L \frac{dI}{dt} + RI + \frac{q}{C} = M \frac{dI_a}{dt}. \quad (16.1)$$

Для того чтобы продвинуться дальше, необходимо конкретизировать вид зависимости $I_a(E_g)$. Примем, что

$$I_a = \alpha + \beta E_g - \gamma E_g^3. \quad (16.2)$$

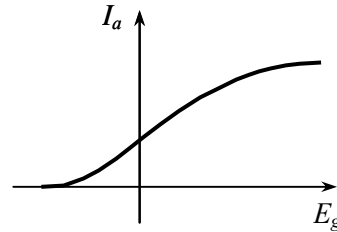
При этом мы пренебрегаем самоиндукцией в анодном контуре и принимаем, что анодный ток определяется формулой (16.2), аппроксимирующей зависимость рис. 16.2:

$$L\ddot{q} + R\dot{q} + \frac{q}{C} = M \left(\frac{\beta}{C} - \frac{3\gamma}{C^3} q^2 \right) \dot{q}$$

или

$$\ddot{q} + \left(\frac{R}{L} - \frac{M\beta}{CL} \right) \dot{q} + \frac{3\gamma M}{LC^3} \dot{q} q^2 + \frac{q}{CL} = 0. \quad (16.3)$$

Рис. 16.2. Вид зависимости анодного тока от напряжения на сетке трехэлектродной лампы



Уравнение (16.3) приводится к виду широко известного уравнения Ван-дер-Поля [2]

$$\ddot{u} - 2\delta(1 - au^2)\dot{u} + \omega^2 u = 0, \quad (16.4)$$

где

$$2\delta = \frac{M\beta}{CL} - \frac{R}{L}, \quad a = \left(\frac{M\beta}{CL} - \frac{R}{L} \right)^{-1} \frac{3\gamma}{LC^3}, \quad \omega^2 = \frac{1}{LC}.$$

При этом предполагается, что $\delta > 0$, $a > 0$ и $\omega^2 > 0$. При $a = 0$ или очень малых u уравнение Ван-дер-Поля превращается в линейный осциллятор с отрицательным трением $\delta > 0$. При $u^2 > 1/a$ коэффициент при \dot{u} становится положительным и мы вправе предположить, что при этом колебания затухают. Таким образом, состояние равновесия осциллятора Ван-дер-Поля неустойчиво и малые колебания нарастают, а очень большие колебания затухают и, следовательно, между ними должно быть устойчивое периодическое движение Γ (рис. 16.3).

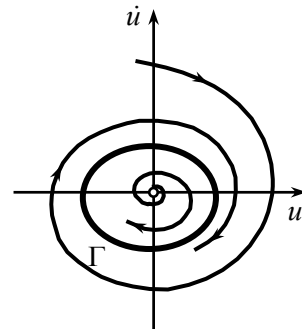


Рис. 16.3. Фазовый портрет уравнения Ван-дер-Поля: Γ – автоколебание – устойчивое периодическое движение

В этом правдоподобном утверждении можно непосредственно убедиться, наблюдая на дисплее ЭВМ фазовый портрет уравнения (16.3). Аналитическое обоснова-

ние этого факта сравнительно просто при достаточно малом δ , при этом мы не только аналитически обоснуем автоколебательность рассматриваемой нами системы при $\delta > 0$ (т.е. при достаточно большом M), но и найдем приближенное значение амплитуды автоколебаний.

Запишем уравнение Ван-дер-Поля (16.4) в виде:

$$\ddot{u} + \omega^2 u = -\varepsilon u^2 \dot{u} + 2\delta \dot{u}, \quad (16.5)$$

где величина

$$\varepsilon = \frac{3\gamma M}{LC^3} > 0,$$

как и δ , принимается малой. Это позволяет записать решение уравнения (16.5) в виде:

$$u = A \cos(\omega t + \varphi), \quad (16.6)$$

где в силу малости ε и δ A и φ – медленно меняющиеся функции времени. Возможность такого представления не вызывает особых сомнений, но вместе с тем требование его конкретного вида может вызвать затруднения. Оно, прежде всего, связано с тем, что такое представление неоднозначно. Но именно пользуясь этой неоднозначностью, можно к условию (16.6) добавить еще одно, например,

$$\dot{u} = -\omega A \sin(\omega t + \varphi), \quad (16.7)$$

после чего A и φ выражаются через u и \dot{u} . Далее, можно непосредственно убедиться, что производные \dot{A} и $\dot{\varphi}$ в силу уравнения (16.5) малы ввиду малости δ и ε .

Рассмотрим, как меняется со временем энергия осциллятора, отвечающего уравнению (16.5) при $\varepsilon = \delta = 0$. Энергия такого осциллятора равна

$$E = \frac{1}{2}(\dot{u}^2 + \omega^2 u^2), \quad (16.8)$$

а ее производная по времени при изменении величины u , согласно (16.5), имеет вид:

$$\frac{dE}{dt} = \dot{u}\ddot{u} + \omega^2 u\dot{u} = (2\delta - \varepsilon u^2)\dot{u}^2$$

или, учитывая (16.6) и (16.7),

$$\frac{dE}{dt} = [2\delta - \varepsilon A^2 \cos^2(\omega t + \varphi)] A^2 \omega^2 \sin^2(\omega t + \varphi). \quad (16.9)$$

Интегрируя по периоду $2\pi/\omega$ полученное соотношение (16.9) при тех же предположениях $\dot{A} = \dot{\varphi} = 0$, найдем, что изменение энергии за это время

$$\Delta E = 2\pi\delta\omega A^2 - \frac{\pi\varepsilon\omega}{2} A^4.$$

С другой стороны, согласно (16.7) и (16.6), $E = \omega^2 A^2/2$. Поэтому

$$\Delta E = \frac{\omega^2}{2} \Delta A^2 = 2\pi\delta\omega A^2 - \pi\varepsilon\omega \frac{A^4}{2}$$

или

$$\Delta(A^2) = \frac{4\pi\delta}{\omega} A^2 - \frac{\pi\varepsilon}{\omega} A^4. \quad (16.10)$$

Обозначая последовательные значения амплитуды A , отстоящие друг от друга на промежутке времени $2\pi/\omega$, через A_n и A_{n+1} , приходим, согласно (16.10), к связывающему их соотношению

$$A_{n+1}^2 = \left(1 + \frac{4\pi\delta}{\omega}\right) A_n^2 - \frac{\pi\varepsilon}{\omega} A_n^2, \quad (16.11)$$

которое уже известным нам образом может быть рассмотрено с помощью диаграммы Кенигса–Ламерея.

Из вида этих диаграмм при $\delta < 0$ и $\delta > 0$ и постоянном $\varepsilon > 0$ следует, что при $\delta < 0$ (рис. 16.4,а) амплитуда A движения уменьшается до нуля, т.е. состояние равновесия глобально устойчиво. Напротив, в интересующем нас случае, когда $\delta > 0$ (рис. 16.4,б), из вида соответствующей диаграммы следует неустойчивость равновесия и наличие глобально устойчивого автоколебания (периодического движения). Амплитуда этого периодического движения A^* находится из условия $A_{n+1} = A_n$ и равна

$$A^* = 2\sqrt{\frac{\delta}{\varepsilon}}. \quad (16.12)$$

Период соответствующего автоколебания при сделанных предположениях малости δ и ε приближенно равен $2\pi/\omega$.

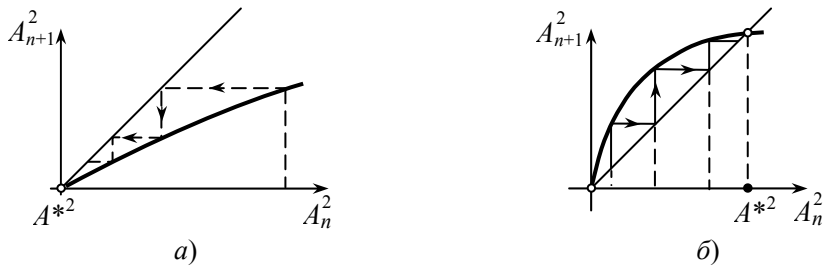
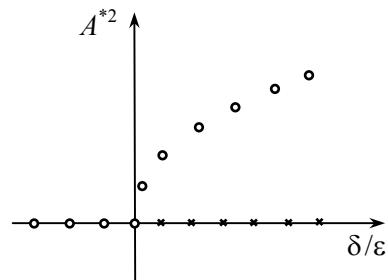


Рис. 16.4. Диаграмма Ламерея в случае мягкого возбуждения автоколебаний с ростом параметра δ : а) при $\delta < 0$, б) при $\delta > 0$.

При $\delta = 0$ происходит рождение из теряющего устойчивость равновесия устойчивой неподвижной точки A^* , соответствующий автоколебанию

Построим еще для наглядности бифуркационную диаграмму равновесия и возникающего из него при смене знака δ автоколебания. Она представлена на рис. 16.5. По оси абсцисс отложена величина δ/ε .

Рис. 16.5. Бифуркационная диаграмма мягкого возбуждения колебаний. С ростом параметра δ/ε при переходе его через нуль равновесие теряет устойчивость и становится неустойчивым. Одновременно с этим рождается автоколебание, амплитуда которого растет с последующим возрастанием величины δ/ε



Точки оси абсцисс отвечают состояниям равновесия, устойчивым при $\delta < 0$ и неустойчивым при $\delta > 0$. Потеря устойчивости происходит при $\delta = 0$ и вместе с этим при возрастании δ возникает автоколебание, амплитуда которого растет от нуля, т.е. оно как бы рождается из сменяющего устойчивость равновесия.

Все сделанные выводы мы получили из приближенного вида точечного отображения (16.11), считая его точным. Не может ли это привести к ошибке? Вообще это возможно, но только при достаточно больших $\dot{\phi}$ или \dot{A} . При малых $\dot{\phi}$ и \dot{A} точечные отображения рис. 16.4 меняются мало, и общие выводы остаются прежними. Они остаются прежними при достаточной малости $\dot{\phi}$ и \dot{A} , т.е. параметров δ и ε . А вот изменение вида анодно-сеточной характеристики $I_a = I_a(E_g)$ может повлечь значительные изменения. Мы их рассмотрим, так как это позволяет обнаружить разные пути возникновения автоколебаний при потере устойчивости состоянием равновесия и обнаружить и дать математические описания явлениям мягкого и жесткого возбуждения автоколебаний и так называемого явления гистерезиса.

Мягкое возбуждение колебаний – плавное от нуля возникновение автоколебаний из теряющего устойчивость состояния равновесия в силу изменения параметров системы – это то, которое описано выше. Подчеркнем, что оно возникает в результате изменения параметров системы.

При обратном изменении параметров вся картина возникновения автоколебаний повторяется в обратном порядке. Адекватной математической моделью мягкого возбуждения являются последовательности фазовых портретов или точечных отображений, отвечающих бифуркационной диаграмме рис. 16.5.

Эта последовательность точечных отображений и фазовых портретов воспроизведена в серии из трех рис. 16.6. Первый рисунок отвечает устойчивости равновесия, второй – состоянию непосредственно после возникновения неустойчивости, третий – при дальнейшем изменении параметра, приводящего к росту амплитуды (размаха) автоколебаний.

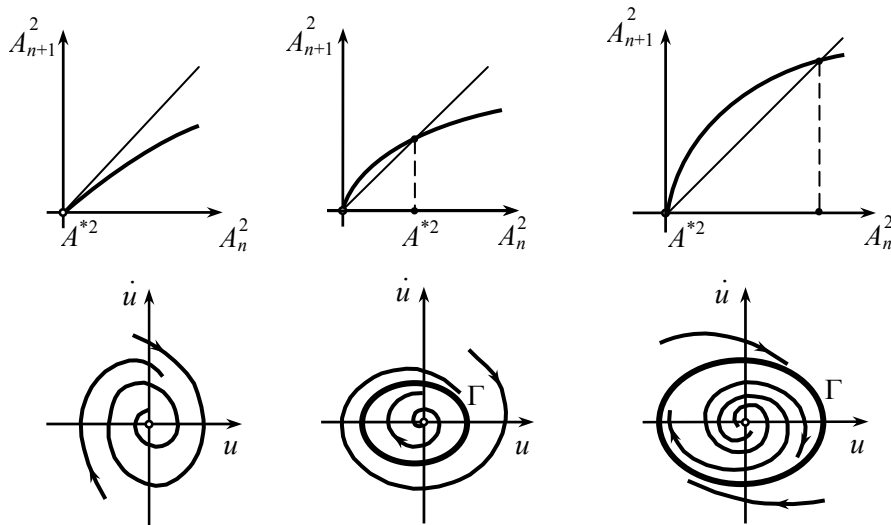


Рис. 16.6. Последовательные виды диаграмм Ламерея и соответствующих фазовых портретов при мягком возбуждении автоколебаний с ростом δ/ε

Такое мягкое возбуждение колебаний в результате изменения параметров наблюдается в реальных системах, но наряду с этим можно наблюдать и совсем другой вариант возникновения автоколебаний, он назван жестким. При жестком режиме возбуждения колебаний все не так: автоколебания возникают скачком, приобретая сразу конечное и, возможно, большое значение. При обратном изменении картина не повторяется в обратном порядке: она другая.

17. МЯГКИЙ И ЖЕСТКИЙ РЕЖИМЫ ВОЗБУЖДЕНИЯ АВТОКОЛЕБАНИЙ

На сыгравшем историческую роль в начале прошлого века примере лампового генератора электрических колебаний и описывающего его динамику дифференциального уравнения Ван-дер-Поля мы познакомились с явлением мягкого возбуждения автоколебаний. Это явление носит общий характер и сопряжено с бифуркацией рождения в фазовом пространстве из теряющего устойчивость состояния равновесия устойчивого периодического движения – автоколебания. Впервые эта бифуркация динамической системы была рассмотрена А.А. Андроном и носит его имя. На соответствующем рисунке были приведены последовательные стадии – кадры киноленты – мягкого возбуждения автоколебаний в двумерном фазовом пространстве. Эти кадры являются геометризованной математической моделью этого явления, а уравнение Ван-дер-Поля можно считать простейшей конкретной его аналитической моделью.

Но возбуждение автоколебаний при изменении параметров может носить не только мягкий, но и так называемый, жесткий характер. Его можно наблюдать на том же ламповом генераторе с другого вида анодной характеристикой $I_a = I_a(E_g)$, например при

$$I_a = \beta E_g + \gamma E_g^3 + \mu E_g^5, \quad (17.1)$$

в котором отсутствие четных степеней несущественно, так как они никак себя не проявляют при нашем приближенном методе исследования.

Повторим предшествующие выкладки для этого более общего случая, когда в характеристику (17.1) добавлен еще один член, и мы придем к точечному отображению вида:

$$A_{n+1}^2 = \left(1 + \frac{4\pi\delta}{\omega}\right) A_n^2 + \varepsilon_1 A_n^4 + \varepsilon_2 A_n^6, \quad (17.2)$$

первые два члена которого совпадают с тем, что было раньше, а третий член новый. Коэффициенты ε_1 и ε_2 пропорциональны γ и μ и, как и δ , по предположению, малы. Кроме того, описание анодной характеристики полиномом (17.1) приемлемо только в некоторой конечной области изменения E_g , что влечет ограниченность амплитуды A_n в (17.2).

Рассмотрим точечное отображение (17.2) с учетом описанных ограничений. Если различать все возможные случаи по знакам параметров δ , ε_1 и ε_2 , то это приводит к восьми различным вариантам, для каждого из которых на рис. 17.1 изображены соответствующие графики точечного отображения – диаграммы Кенигса – Ламерея. Они разбиты на четыре пары $a)$, $b)$, $v)$ и $z)$, каждая из которых отвечает $\delta < 0$ и $\delta > 0$ при одинаковых ε_1 и ε_2 .

Случай $z)$ соответствует уже известному нам мягкому возбуждению колебаний при непрерывном переходе от $\delta < 0$ к $\delta > 0$, а случай $v)$ – интересующему нас новому жесткому возбуждению. Случаи $a)$ и $b)$ скорее умозрительного характера, так как не отвечают реальному виду характеристики электронной лампы.

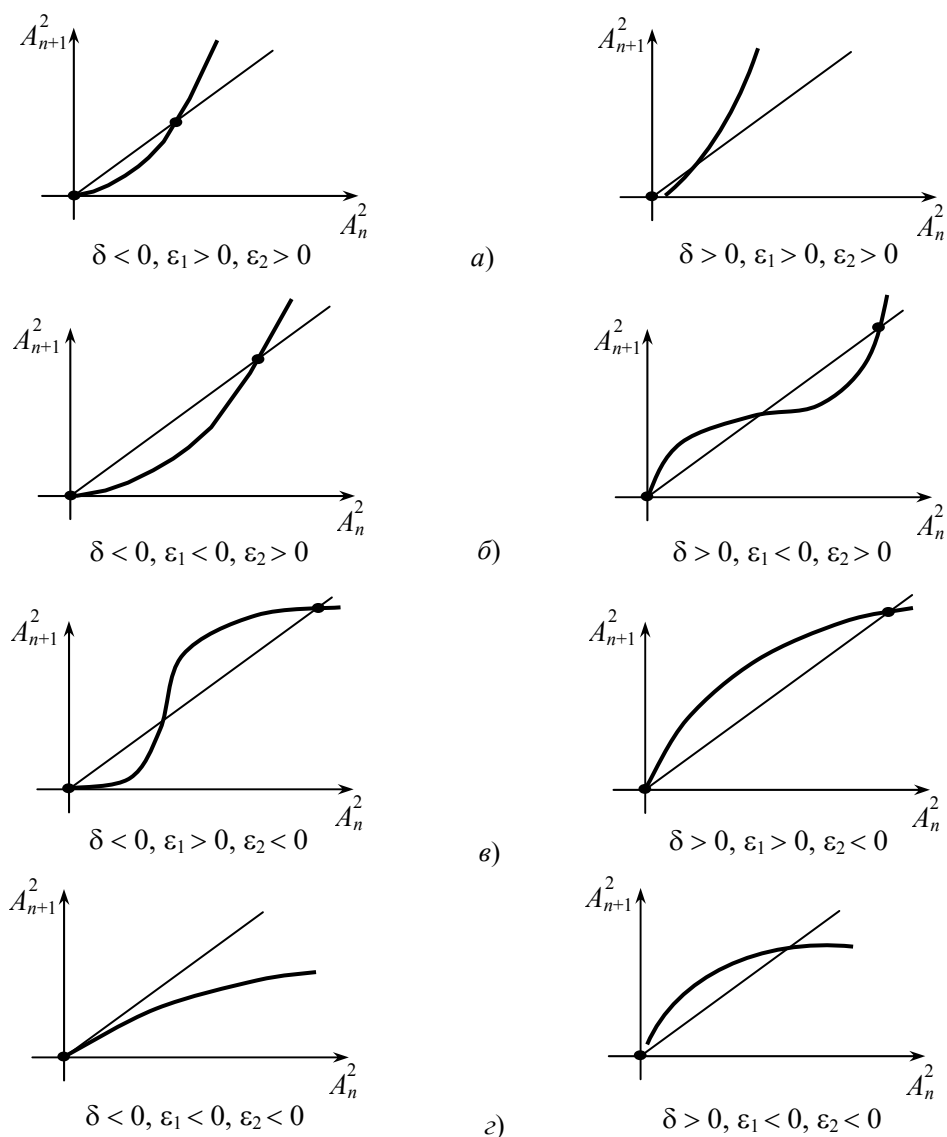


Рис. 17.1. Возможные виды диаграммы Ламерея при полиномиальной характеристике лампы (17.1) в соответствии с видом точечного отображения (17.2)

Сосредоточим внимание на случае *в*), изобразив его подробнее в виде пяти последовательных графиков точечного отображения (рис. 17.2) для возрастающих значений величины δ .

Из этих графиков видно, что сначала имеет место устойчивое состояние равновесия O , затем появляется полуустойчивое периодическое движение M_{12} (точнее, полуустойчивое периодическое движение, отвечающее неподвижной точке M_{12}), затем распадающееся на устойчивое M_2 и неустойчивое M_1 периодические движения. При дальнейшем возрастании δ амплитуда неустойчивого периодического движения M_1 уменьшается, а устойчивого M_2 – увеличивается, в результате чего M_1 сливается с состоянием равновесия O , которое из устойчивого становится неустойчивым.

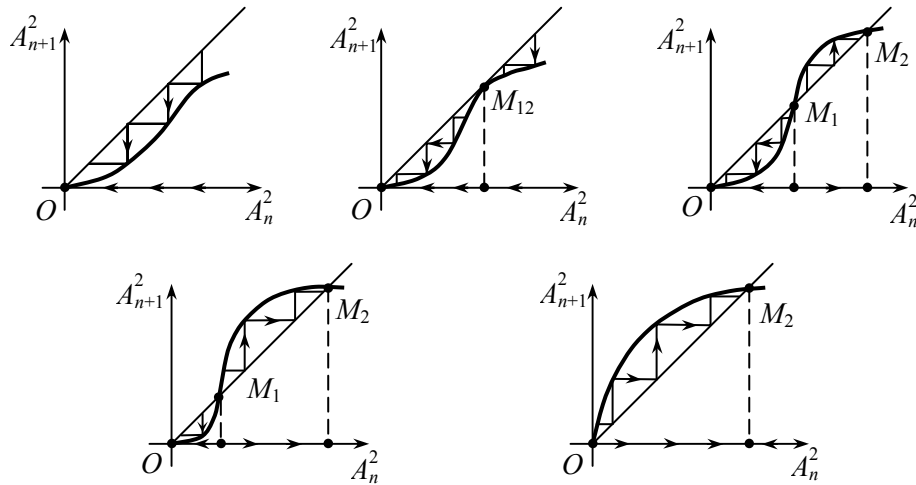


Рис. 17.2. Последовательность диаграмм Ламерея жесткого возбуждения колебаний: сначала равновесие O глобально устойчиво, затем рождается полуустойчивая неподвижная точка M_{12} , которая распадается на две: M_1 и M_2 ; затем точка M_1 сливается с равновесием O и делает его неустойчивым, в результате чего глобально устойчивой оказывается неподвижная точка M_2

Фазовые портреты, отвечающие этим превращениям, представлены на рис. 17.3. Найдем, основываясь на этих рисунках, бифуркационную диаграмму. Она изображена на рис. 17.4. Как видно, она отличается от бифуркационной диаграммы, имеющей место в случае мягкого возбуждения (см. рис. 16.5). Что же следует из этого различия? Прежде всего, если раньше при переходе δ через нуль, возрастая, состояние равновесия теряло устойчивость, отрождая от себя устойчивое периодическое движение, то теперь при потере устойчивости с ним сливается неустойчивое периодическое движение и система, находясь ранее в устойчивом равновесии, покидает его и переходит сразу к периодическому движению конечной амплитуды M_2 . Это и есть жесткий режим возбуждения колебаний, когда сразу скачком возникают автоколебания конечной амплитуды. Конечно, они устанавливаются не мгновенно, для этого нужно некоторое время и размах колебаний растет непрерывно, а скачок с изменением параметра имеет место для установившегося движения.

Таким образом, первое отличие жесткого режима возникновения автоколебаний как физического явления, состоит не в плавном, а скачкообразном изменении характера движения с изменением параметра. В математическом описании это различие отражается в том, что из состояния равновесия не рождается устойчивое периодическое движение, а с ним сливается неустойчивое периодическое движение и точки, близкие к ставшему неустойчивым состоянию равновесия, устремляются к ранее существующему устойчивому периодическому движению.

Второе существенное отличие относится к необратимому характеру жесткого перехода от равновесия к установившемуся движению: при возрастании δ и переходе через значение $\delta = 0$ жестко возникают автоколебания, но при обратном уменьшении δ через нуль автоколебания не исчезают, они сохраняются, непрерывно меняя амплитуду. Для того чтобы от жестко возникающих автоколебаний перейти к устойчивому состоянию равновесия, нужно уменьшить параметр до не-

которого отрицательного значения $\delta = \delta^*$ (рис. 17.4), после чего автоколебания скачком исчезают и происходит переход к устойчивому равновесию. Исчезновение устойчивого автоколебания сопровождается его слиянием с неустойчивым периодическим движением и их совместным исчезновением (при $\delta = \delta^*$ устойчивое периодическое движение M_2 сливается с неустойчивым M_1 и при $\delta < \delta^*$ они исчезают). Таким образом, возникновение и исчезновение автоколебаний при жестком их возбуждении имеет место при разных значениях параметра δ , как это показано на рис. 17.4. Это явление подобно гистерезису ферромагнитных тел.

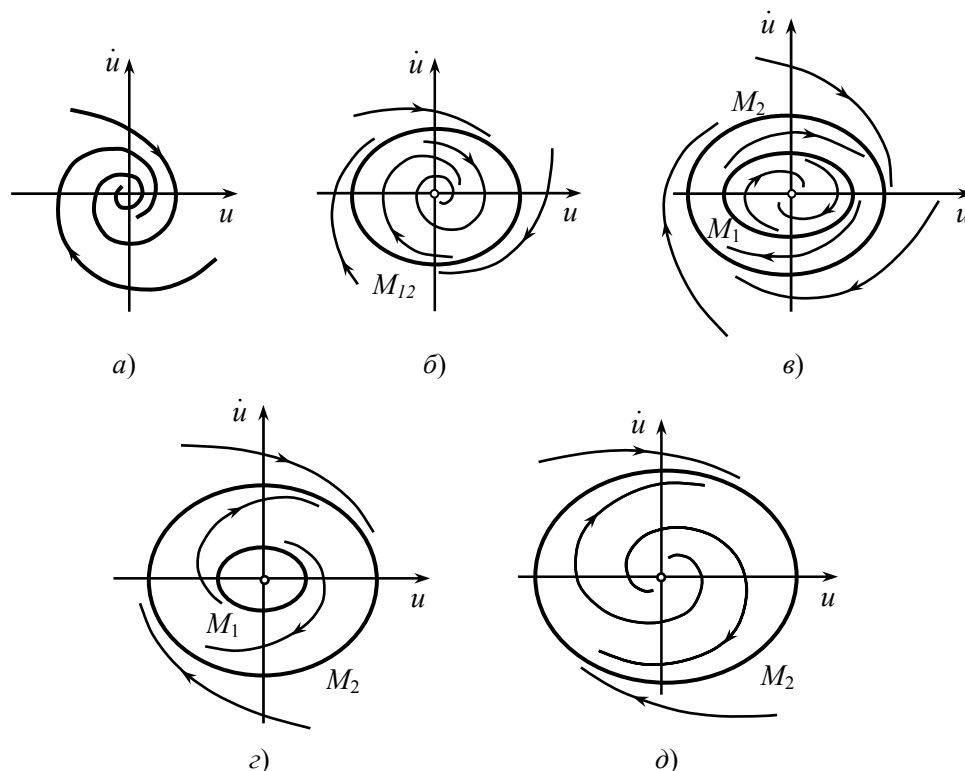
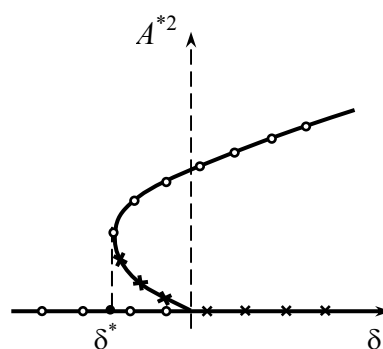


Рис. 17.3. Изменения фазового портрета, отвечающие изменениям диаграммы Ламерея на рис. 17.2

Рис. 17.4. Бифуркационная диаграмма жесткого возбуждения колебаний. Кружочки и крестики – соответственно устойчивые и неустойчивые равновесия и периодические колебания



Закончим повествование о мягком и жестком возбуждениях автоколебаний мыслимым экспериментом. Пусть перед вами ручка, вращая которую вы можете увеличивать и уменьшать параметр динамической системы, наблюдая одновременно на дисплее ЭВМ изменение состояния системы (одну из его компонент).

Тогда при мягком режиме при медленном вращении ручки сначала не будет колебаний, затем они возникнут, и их амплитуда будет медленно нарастать, тем медленнее, чем медленнее вы вращаете ручку (рис. 17.5,*а*). При обратном движении все виденное вами повторится в обратном порядке.

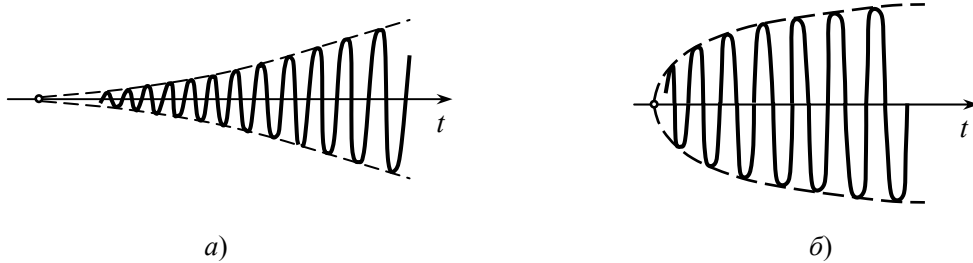


Рис. 17.5. Осциллограммы мягкого (*а*) и жесткого (*б*) возбуждения колебаний при плавном изменении параметра

При жестком возбуждении вначале также не будет колебаний, но затем они возникнут сразу, причем быстрота этого перехода не уменьшается с замедлением вращения ручки (рис. 17.5,*б*). При обратном изменении картина не будет меняться в обратном порядке. Исчезновение автоколебаний, и притом тоже скачком, произойдет при более длительном вращении ручки в обратную сторону.

18. СТОХАСТИЧЕСКИЙ ОСЦИЛЛЯТОР («ЧАСЫ НАОБОРОТ»)

В предыдущих лекциях мы рассмотрели часы Галилея – Гюйгенса. Основой этих часов является маятник или любой другой осциллятор, колебания которого поддерживаются за счет какого-нибудь источника энергии и период которых достаточно стабилен. Стабильность периода колебаний – это точность часов. Мы изучили возможные причины нестабильности и пути ее уменьшения. Очень важным моментом стабильности была малость затухания, одновременно и малая зависимость от затухания периода, достаточность очень малых толчков и малая, вызываемая толчками, нестабильность.

Однако осциллятор может быть не только высокоточным измерителем времени с весьма стабильным периодом, но может обладать и беспорядочными, значительно меняющимися временами колебаний, настолько разными, что говорить о периоде колебаний становится бессмысленным. Более того, эти колебания приобретают свойства случайности, становятся непредсказуемыми и допускают вероятностное описание. Эта удивительная возможность стала научной сенсацией недавних лет. Она изменила общепринятые представления о случайности. Оказалось, что случайность может порождаться динамической системой, описываемой дифференциальными уравнениями. И это несмотря на то, что остается верной теорема о единственности решения, о единственности его определения начальными условиями (начальным состоянием). Как это может быть? Представляется очевидным, что это совершенно невозможно, и все-таки это может быть и реализуется вокруг нас не менее часто, чем детерминированные процессы, описываемые дифференциальными уравнениями. Оказывается, что реальные решения дифференциальных уравнений могут быть детерминированными и предсказуемыми, а могут быть и случайными непредсказуемыми.

Целью дальнейшего рассказа является объяснение того, как это может быть, на очень простом примере стохастического осциллятора или «часов наоборот».

Часы – это затухающий осциллятор, время от времени подталкиваемый. «Часы наоборот» – это самораскачивающийся осциллятор, колебания которого время от времени одерживаются.

Математическая модель часов имела вид:

$$\begin{aligned} \ddot{x} + 2\delta\dot{x} + \omega^2 x &= 0 \text{ для } x \neq 0 \text{ или } x = 0 \text{ и } \dot{x} < 0, \\ \dot{x}_+ &= \dot{x}_- + p \text{ для } x = 0 \text{ и } \dot{x} > 0. \end{aligned} \quad (18.1)$$

Математическая модель «часов наоборот» ($p \leq a$):

$$\begin{aligned} \ddot{x} - 2\delta\dot{x} + \omega^2 x &= 0 \text{ для } x \neq 0 \text{ или } x = 0 \text{ и } \dot{x} < a, \\ \dot{x}_+ &= \dot{x}_- - p \text{ для } x = 0 \text{ и } \dot{x} \geq a. \end{aligned} \quad (18.2)$$

Уравнения (18.2) получаются из (18.1) заменой $-\delta$ на δ и p на $-p$. Кроме того, еще ноль заменяется числом $a > 0$.

Нам надлежит исследовать модель (18.2). Прежде всего представим себе вид ее фазовой траектории (рис. 18.1). Затем сведем исследование динамики модели

(18.2) к точечному отображению, к диаграмме Кенигса – Ламерея. Для этого рассмотрим последовательные уходы фазовой точки с полуоси $x = 0, \dot{x} > 0$.

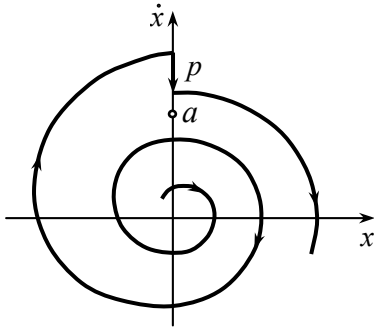


Рис. 18.1. Фазовый портрет «часов наоборот» (стохастического осциллятора)

Пусть после первого пересечения $\dot{x} = u$, а после второго $\dot{x} = \bar{u}$. Найдем связь между u и \bar{u} . Пусть траектория выходит из точки $M(0, u)$. Она приходит вновь на луч $x = 0, \bar{u} > 0$ в некоторой точке $M_1(0, u_1)$, где

$$u_1 = ue^{2\pi\delta/\Omega},$$

и не одерживается, если $u_1 < a$ (рис. 18.2,а), и тогда $\bar{u} = u_1$, либо одерживается, если $u_1 \geq a$, и тогда $\bar{u} = u_1 - p$ (рис. 18.2,б). Итак, имеем:

$$\bar{u} = \begin{cases} qu & \text{для } u < a/q, \\ qu - p & \text{для } u \geq a/q, \end{cases} \quad (18.3)$$

где для сокращения записи введено обозначение $q = \exp(2\pi\delta/\Omega) > 1$.

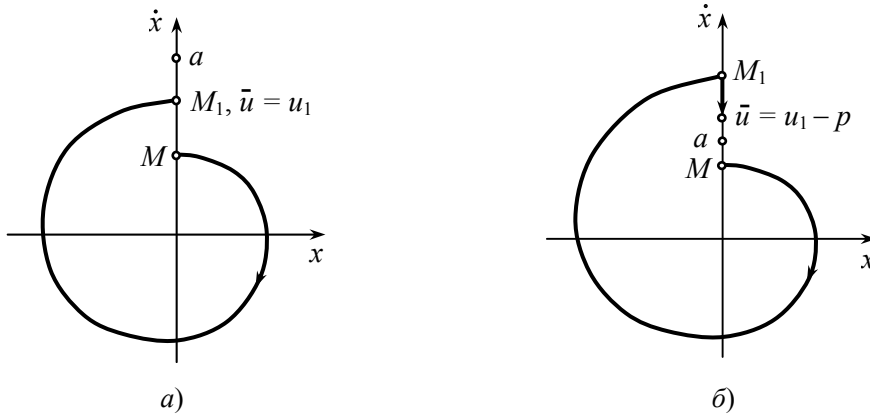


Рис. 18.2. Различные возможные виды фазовых траекторий «часов наоборот»:

а) при u_1 точки M_1 меньшем a ; б) при $u_1 > a$.

За цикл колебания точка M с координатой u переходит в точку с координатой \bar{u} , причем при $u_1 > a$ этот переход совершается со скачком от u_1 до \bar{u}

Нарисуем для точечного отображения (18.3) диаграмму Кенигса – Ламерея. Для этого построим график зависимости \bar{u} от u (рис. 18.3) и проведем биссектрису. Изучение диаграмм, несмотря на их простоту, достаточно сложно. Но оно стоит того, чтобы потрудиться. Даже потрудившись, мы далеко не все сможем обосновать, кое-что придется принять на веру.

Получаемые таким образом диаграммы точечного отображения могут иметь три различных вида. Они показаны на рис.18.3,а, б и в.

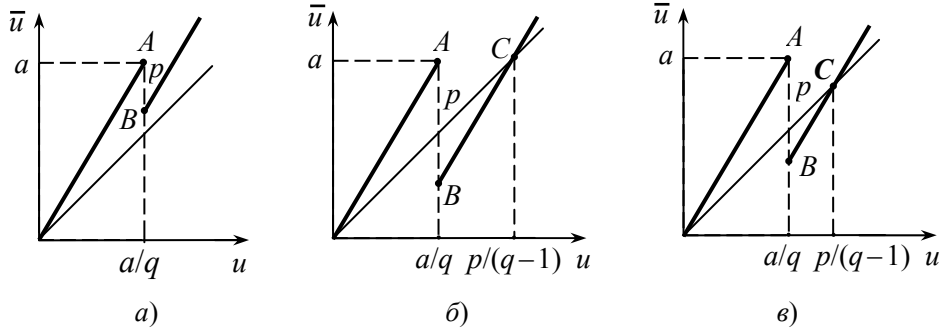
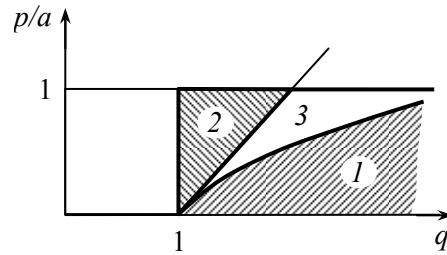


Рис. 18.3. Возможные виды диаграммы Ламерея для «часов наоборот»

Первый случай (рис. 18.3,а): точки A и B обе лежат выше биссектрисы. При этом $a - p > a/q$ или $p/a < 1 - 1/q$. Второй случай (рис. 18.3,б): точка B лежит ниже биссектрисы, а точка C выше точки A . При этом $a - p < a/q$ и $a < p/(q - 1)$ или $p/a > 1 - 1/q$ и $p/a > q - 1$. Третий случай (рис. 18.3,в): точка B лежит ниже биссектрисы и точка C ниже точки A . При этом $a - p < a/q$ и $a > p/(q - 1)$ или $p/a > 1 - 1/q$ и $p/a < q - 1$. Как видно, от величин q и p/a зависит то, какой из случаев имеет место. При этом $q > 1, 0 < p/a \leq 1$.

Каждому из перечисленных случаев на плоскости величин q и p/a отвечает своя область. Эти области, обозначенные соответственно номерами 1, 2 и 3, показаны на рис.18.4.

Рис. 18.4. Параметрический портрет, соответствующий изображенным на рис. 18.3 различным видам диаграммы Ламерея



В случаях 1 и 3 почти все последовательные преобразования уходят в бесконечность. Это соответствует неограниченному раскачиванию осциллятора. В случае 2 в зависимости от начального значения величины u возможны два разных типа поведения. При $u < p/(q - 1)$ последовательные преобразования u все время ограничены. Напротив, при $u > p/(q - 1)$ они неограниченно возрастают.

Наибольший интерес для нас представляет случай 2, когда значения преобразований все время ограничены. Для любого начального $u_0 \in (0, p/(q - 1))$ последовательные преобразования u_1, u_2, \dots с некоторого номера попадают на отрезок $[a - p, a]$ и на нем остаются. Любая такая фазовая траектория u_0, u_1, u_2, \dots неустойчива, и близкие к ней траектории разбегаются. Они, в силу наличия у отображения (18.3) разрыва, могут сблизиться, но потом снова разбегаются и т.д.

Проследим, как это происходит, сопоставляя две траектории u_0, u_1, \dots и v_0, v_1, \dots с близкими начальными точками u_0 и v_0 . Если между точками u_0 и v_0 не лежит точка разрыва a/q отображения (18.3), то, согласно (18.3),

$$u_1 - v_1 = q(u_0 - v_0) \tag{18.4}$$

и, следовательно, расстояние между ними увеличивается в $q > 1$ раз. Если же между точками u_0 и v_0 лежит точка разрыва a/q отображения (18.3), то

$$u_1 - v_1 = q(u_0 - v_0) \pm p. \quad (18.5)$$

Такая же альтернатива имеет место и на всех последующих преобразованиях. Поэтому в любом случае после некоторого числа преобразований, сколь бы близки ни были начальные точки, они расходятся на конечное расстояние, не меньшее, чем $p/2q$. Причем это заведомо происходит за число шагов, не более чем n , для которого

$$q^n |u_0 - v_0| > \frac{p}{2q}. \quad (18.6)$$

Поэтому, сколь бы точно, но все же с некоторой погрешностью, мы ни вычисляли последовательные преобразования отображения (18.3), их ошибки после некоторого их числа превзойдут величину $p/2q$. Исключение составляет лишь случай абсолютно точного вычисления, который вообще нереализуем. А так как эти малые ошибки, как и малые возмущения, при физической реализации непредсказуемы и не поддаются учету, то фактически мы сталкиваемся с непредсказуемостью фазовой траектории движения «часов наоборот». Движения «часов наоборот» непредсказуемы, и эта непредсказуемость явно проявляется очень быстро. Действительно, неравенства (18.6) при $q = 1,76$, $p = 4$ и очень маленькой ошибке в начальных значениях порядка 10^{-8} выполняются для $n \geq 32$. Для ошибки порядка размеров атома оно выполняется при $n \geq 48$.

Движение «часов наоборот» непредсказуемо еще и потому, что мы не можем задать или знать их начальные условия абсолютно точно. Любая же погрешность в начальном значении через некоторое время ведет к большим различиям, ведет к непредсказуемости.

Идея о том, что случайность может порождаться исчезающе малыми, не известными нам неопределенностями с последующим их экспоненциальным усилением из-за неустойчивости, высказывалась еще А. Пуанкаре и позднее в более развернутом виде Н.С. Крыловым и М. Борном. Применительно к долгосрочному прогнозу погоды эффект необычайно чувствительной зависимости от малых возмущений отмечался Э. Лоренцем как «эффект бабочки», которая своим полетом может изменить состояние атмосферы и будущую погоду так же, как она повлияла на события, развертывающиеся в одном из рассказов Брэдбери.

Осмысливание сказанного, естественно, приводит к пониманию того, что неустойчивость порождает непредсказуемость, но это не значит, что случайность. Для случайности необходимо наличие еще вероятностного описания. Откуда ему взяться? Мне представляется, что сегодня еще трудно дать исчерпывающий однозначный ответ. Этот вопрос требует дальнейшей глубокой проработки. Но кое-что сказать можно.

Оказывается, что динамическая система, описываемая отображением (18.3), обладает удивительными свойствами, которые позволяют ее непредсказуемости приписать вероятность, определяемую только ее динамикой, а не тем, какие случайные возмущения вызывают непредсказуемость. Нужно лишь наличие этих ничтожно малых случайностей. Достаточно тех, которые порождаются тепловым движением молекул или даже квантовой вероятностной природой микромира.

Поясним сказанное на примере «часов наоборот» с параметрами из области 2 рис. 18.4, взяв в ней точку, где $p = a$ и $q = 2$. Диаграмма точечного отображения (18.3), которое обозначим через T , принимает вид, показанный на рис. 18.5, а само оно записывается так:

$$\bar{u} = \begin{cases} 2u & \text{при } u < a/2, \\ 2u - a & \text{при } u \geq a/2. \end{cases} \quad (18.7)$$

Отображение, обратное (18.7), двузначное. Пусть T_1^{-1} и T_2^{-1} – его однозначные ветви. Согласно (18.7),

$$T_1^{-1} : u = \frac{\bar{u}}{2}; \quad T_2^{-1} : u = \frac{\bar{u} + a}{2}. \quad (18.8)$$

Зададим начальную точку u_0 , предполагая ее случайной с некоторой плотностью вероятности $p_0(u)$. Малая неопределенность в задании u_0 означает, что функция $p_0(u)$ плотности вероятности похожа на δ -функцию и имеет вид, изображенный на рис. 18.6. Чем уже и выше пик этой функции, тем точнее задается начальное значение u_0 .

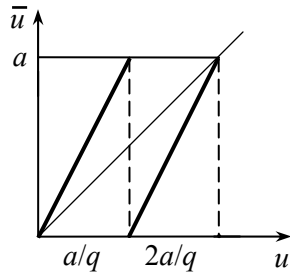


Рис. 18.5. Частный случай диаграммы Ламерея для «часов наоборот», когда $p = a$ и $q = 2$

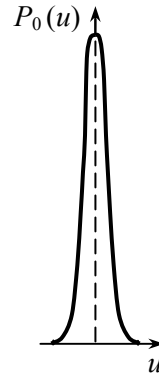


Рис. 18.6. Вид плотности вероятности задания начального значения u

Произведем преобразование T , при этом каждое u_0 перейдет в новое значение u_1 . Если u_0 случайные, то и u_1 тоже случайные. По плотности вероятности $p_0(u)$ случайной точки u_0 можно найти плотность вероятности преобразованной точки u_1 . Пусть это $p_1(u)$. Продолжая преобразования T , можно найти $p_2(u)$, $p_3(u)$ и так далее. Для того чтобы действительно их вычислить, найдем связь между $p_s(u)$ и $p_{s+1}(u)$.

Точки $u_{s+1} \in (u, u + du)$, если $u_s \in (T_1^{-1}(u), T_1^{-1}(u + du))$ или $u_s \in (T_2^{-1}(u), T_2^{-1}(u + du))$. Поэтому

$$\begin{aligned} P(u_{s+1} \in (u, u + du)) &= \\ &= P(u_s \in (T_1^{-1}(u), T_1^{-1}(u + du))) + P(u_s \in (T_2^{-1}(u), T_2^{-1}(u + du))), \end{aligned} \quad (18.9)$$

где P означает вероятность события, стоящего за P в скобках. Так что соотношение (18.9) означает, что вероятность того, что точка u_{s+1} , являющаяся отображением точки u_s , принадлежит интервалу $(u, u + du)$, равна сумме вероятностей того, что

u_s принадлежит соответственно интервалу $(T_1^{-1}(u), T_1^{-1}(u + du))$ или интервалу $(T_2^{-1}, T_2^{-1}(u + du))$.

Учитывая (18.8), перепишем (18.9) в виде:

$$P(u_{s+1} \in (u, u + du)) = P\left(u_s \in \left(\frac{u}{2}, \frac{u + du}{2}\right)\right) + P\left(u_s \in \left(\frac{u + a}{2}, \frac{u + a + du}{2}\right)\right)$$

или

$$p_{s+1}(u)du = p_s\left(\frac{u}{2}\right)\frac{du}{2} + p_s\left(\frac{u + a}{2}\right)\frac{du}{2},$$

или, сокращая на du ,

$$p_{s+1}(u) = \frac{1}{2}\left[p_s\left(\frac{u}{2}\right) + p_s\left(\frac{u + a}{2}\right)\right]. \quad (18.10)$$

Это искомое соотношение (18.10) позволяет, начиная с известной плотности вероятностей $p_0(u)$, найти последовательно $p_1(u)$, $p_2(u)$ и так далее. Естественен вопрос: есть ли предельная вероятность и какова она? Установить сходимость и найти предел можно следующим образом.

Пусть начальная плотность вероятностей $p_0(u)$ имеет производную. Тогда, согласно (18.10), ее имеют и все последующие вероятности $p_s(u)$. Продифференцируем (18.10) и найдем, что

$$p'_{s+1}(u) = \frac{1}{4}\left\{p'_s\left(\frac{u}{2}\right) + p'_s\left(\frac{u + a}{2}\right)\right\}. \quad (18.11)$$

Из (18.11) с очевидностью следует, что

$$\lim_{s \rightarrow \infty} p'_s(u) = 0,$$

и поэтому предельная функция $p_\infty(u)$ постоянна. Учитывая условие нормировки, непосредственно находим, что

$$p_\infty(u) = \frac{1}{a}.$$

Таким образом, спустя некоторое время, величины послеударных скоростей u_s в моменты прохождения маятником «часов наоборот» равновесия, являются равномерно распределенными случайными величинами с интервалом значений от нуля до a . Эти величины не независимы. Напротив, близкие по номеру величины тесно связаны друг с другом, но по мере удаления друг от друга эта связь слабеет и очень отдаленные значения практически независимы.

Мы обосновали наличие вероятности лишь на конкретном примере, да еще только в частном случае $p = a$ и $q = 2$. Обоснование может быть и более широким. В том, что изложено, рассуждение о связи плотностей вероятностей $p_s(u)$ и $p_{s+1}(u)$, имеет общий характер. Трудности возникают в обосновании существования предела $p_s(u)$ при $s \rightarrow \infty$. Именно поэтому мы ограничимся частным случаем $p = a$ и $q = 2$, хотя существующие исследования позволяют сделать значительно более общие утверждения. Они позволяют утверждать, что наличие предела не есть что-то исключительное, что это одна из общих возможностей.

Таким образом, ситуация, обнаруженная в рассмотренном очень простом примере «часов наоборот», имеет общий характер. Она имеет место в широком круге явлений турбулентности жидкостей и газов. Именно ею вызваны трудности долгосрочного прогноза погоды. Стохастичность наблюдается во многих электрических и механических системах, в движениях молекул газа. Можно сказать, что устойчивости и неустойчивости отвечают два принципиально разных основных типа поведения динамических систем. В первом случае имеет место предсказуемость и детерминизм, а во втором – непредсказуемость и стохастичность.

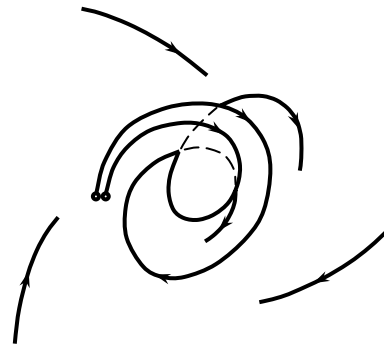
Простейшим примером первого типа динамических систем являются часы Галилея – Гюйгенса, типичным примером второго типа – «часы наоборот» или простейший стохастический осциллятор [40, 41].

Часы служат для измерения времени, в регулярности и стабильности их хода их достоинство, стохастический осциллятор порождает из микрослучайности макрослучайность. Часы Галилея – Гюйгенса генерируют равномерный ряд моментов времени, «часы наоборот» генерируют случайную последовательность.

В динамических системах первого типа фазовые траектории сближаются, стремясь к состояниям равновесия или периодическим движениям, в системах второго типа они все время разбегаются, образуя сложный запутанный клубок.

Ранее, при обсуждении автоколебаний, была выявлена естественность их возникновения в динамических системах и указан геометрический образ автоколебаний – замкнутая фазовая кривая, к которой приближаются все соседние траектории [2]. С непредсказуемыми хаотическими движениями это сделать много сложнее. Они значительно более многолики [22–24]. Общее их свойство – в нерегулярности и непредсказуемости, которые могут вызываться самыми ничтожными возмущениями. Можно выделить притягивающие множества неустойчивых фазовых траекторий. Это когда каждая отдельно взятая фазовая траектория неустойчива и соседние от нее убегают, а все в целом они притягивающие, что схематически представлено на рис. 18.7.

Рис. 18.7. Схематичное представление притягивающего множества неустойчивого непредсказуемого движения



Этого достаточно для непредсказуемости. Для стохастичности, по крайней мере, следует еще потребовать наличия предельных плотностей вероятностей, о которых шла речь выше, а, в общем, еще все впереди и ждет своих исследователей и исследований, которые могут видоизменить и обогатить нашу феноменологическую аксиоматически построенную теорию вероятностей.

На этом можно было бы закончить, но мне хочется познакомить вас еще с замечательной системой дифференциальных уравнений Лоренца:

$$\begin{aligned}\dot{x} &= -\sigma x + \sigma y, \\ \dot{y} &= -rx - y - xz, \\ \dot{z} &= -bz + xy,\end{aligned}$$

дающих приближенное описание тепловой конвекции между нагретой нижней и более холодной верхней горизонтальными пластинками. При $b = 0$ это уравнение описывает конвективную циркуляцию жидкости в замкнутой круговой трубке, которую снизу нагревают. В отсутствие нагрева снизу эта система имеет устойчивое равновесие. По мере нагрева оно становится неустойчивым и рождаются две устойчивые циркуляции жидкости по часовой стрелке и против часовой стрелки. Затем по мере роста нагрева и эти циркуляции становятся неустойчивыми и наступает хаос: равновесие нарушается – возникает циркуляция, она нарушается и возникает другая циркуляция, она, в свою очередь, прекращается и возникает снова прежняя, и т.д. без какой-либо регулярности, без намека на периодичность, с единственно закономерной сменой циркуляции в одну сторону циркуляцией в другую, сменой непредсказуемой, совершаемой через разные случайные промежутки времени. Но что замечательно: оказывается, исследование этих смен циркуляций может быть описано точечным отображением, похожим на то, которое имело место в рассмотренном нами примере «часов наоборот». Вот на этом уже поставим точку.

19. НЕУСТОЙЧИВОСТЬ И АВТОКОЛЕБАНИЯ, ВЫЗЫВАЕМЫЕ ТРЕНИЕМ

Последующий рассказ – о механизме и простейших математических моделях возникновения неустойчивости и автоколебаний благодаря трению.

В том, что трение вызывает неустойчивость и автоколебания, есть что-то парадоксальное: мы привыкли к тому, что трение тормозит и прекращает движение. Но, оказывается, это не всегда так, и парадокс разрешается тем, что трение выступает в роли распределителя энергии некоторого другого, не связанного с трением, источника энергии. Само же по себе оно действительно может только тормозить движение.

Схематическая упрощенная модель, с которой мы начнем исследование систем с парадоксальной ролью сухого трения, состоит из массы, которую через пружину тянут с постоянной скоростью по горизонтальной плоскости. Массивное тело при этом движется под влиянием силы растягиваемой пружины (рис. 19.1). В этой модели можно считать, что движется с постоянной скоростью плоскость, а конец пружины, прикрепленной к телу, неподвижен (рис. 19.2).

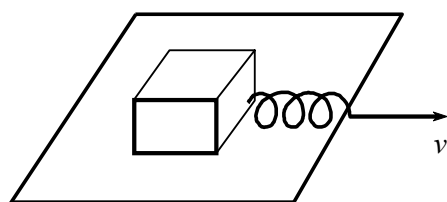


Рис. 19.1. Движение тела под влиянием силы растягиваемой пружины

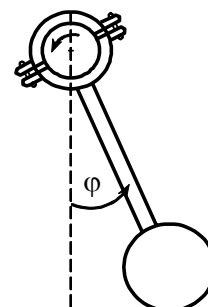


Рис. 19.3. Маятник Фроуда

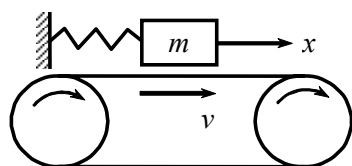


Рис. 19.2. Другой вид конкретной реализации математической модели с трением

Остановимся на этом последнем варианте модели (рис. 19.2). В несколько другом конкретном физическом виде она предстает в так называемом маятнике Фроуда, изображенном на рис. 19.3, и представляющем собою обычный маятник, насаженный с помощью муфты на вращающуюся с постоянной скоростью ось [74]. Ось вращается в муфте с некоторым не очень большим трением, чтобы маятник мог качаться на ней, но не вращался вместе с нею. Реальными прототипами описанной схематической модели могут служить тормозные колодки автомобиля, трамвая, железнодорожного вагона, а также других разнообразных механических

тормозных устройств. С некоторой натяжкой сюда же можно отнести скрипичную струну, по которой с трением скользит смычок.

Вернемся к принятой схематической модели. Обозначим через x величину смещения массы m от ее положения, при котором пружина не растянута, а через F – силу трения массы о движущуюся плоскость опоры, тогда уравнение Ньютона запишется в виде:

$$m\ddot{x} = -kx + F. \quad (19.1)$$

Природа силы сухого трения очень сложна. Первая математическая модель силы сухого трения была предложена еще в XVIII веке Кулоном. Она так и называется – кулоновское трение. Об этой модели вам рассказывали в школе в разделе о законах сухого трения, согласно которым сила трения не зависит от взаимной скорости трущихся поверхностей, действие ее направлено против скорости и по величине пропорционально нормальному давлению, так что

$$F = F_0 \text{sign}(v - \dot{x}), \quad (19.2)$$

где $\text{sign } \xi$ – знак величины ξ . Это при $\dot{x} \neq v$. При $\dot{x} = v$, т.е. когда масса движется вместе с плоскостью, как бы скреплена с ней, сила трения – сила трения покоя – может быть любой в пределах от $-F_0$ до F_0 . Последнее означает, что любая сила, по величине меньшая F_0 , не может сдвинуть покоящуюся относительно движущейся плоскости массу. Движение начинается, когда действующая сила по величине превазойдет F_0 .

Эти последние утверждения о силе трения покоя можно включить в формулу (19.2), приняв, что $\text{sign } 0$ означает любую величину между -1 и $+1$. После чего соотношение (19.2) становится математической моделью кулоновского трения и уравнение движения запишется в виде:

$$m\ddot{x} + kx = F_0 \text{sign}(v - \dot{x}). \quad (19.3)$$

Дифференциальное уравнение (19.3) – это математическая модель рассматриваемой нами системы. Выяснение ее динамики сводится к рассмотрению фазового портрета динамической системы (19.3).

Фазовым пространством динамической системы, определяемой дифференциальным уравнением (19.3), является плоскость фазовых переменных x и $\dot{x} = y$. В полуплоскостях $y < v$ и $y > v$ движения фазовых точек подчиняются, согласно (19.3), уравнениям

$$m\ddot{x} + kx = \mp F_0, \quad (19.4)$$

каждое из которых является гармоническим осциллятором с состоянием равновесия соответственно в точке $x = -F_0/k$, $\dot{x} = y = 0$ или в точке $x = F_0/k$, $\dot{x} = y = 0$. Это позволяет непосредственно нарисовать, как показано на рис. 19.4, фазовые траектории отдельно в каждой из полуплоскостей $y > v$ и $y < v$.

Движение фазовой точки на этом рисунке определено всюду, кроме точек линии $\dot{x} = y = v$. Эта линия разбивается на три части: луч $(-\infty, O_1)$, по отношению к которому фазовые траектории приходят снизу и уходят сверху; отрезок $[O_1, O_2]$, на который фазовые траектории приходят снизу и сверху, и луч (O_2, ∞) , где они приходят сверху и удаляются снизу.

На лучах $(-\infty, O_1)$ и (O_2, ∞) естественно предположить, в силу непрерывности изменения скорости $y = \dot{x}$ и координаты x (фазовая точка не может совершать скачков), что при своем движении фазовая точка пересекает их соответственно снизу вверх и сверху вниз; а на отрезке $[O_1, O_2]$, придя к нему снизу или сверху, остается на нем, двигаясь вдоль отрезка согласно тому, что на нем $y = \dot{x} = v$, и, достигнув в результате этого движения точки O_2 , ее покинет, переходя в нижнюю полуплоскость.

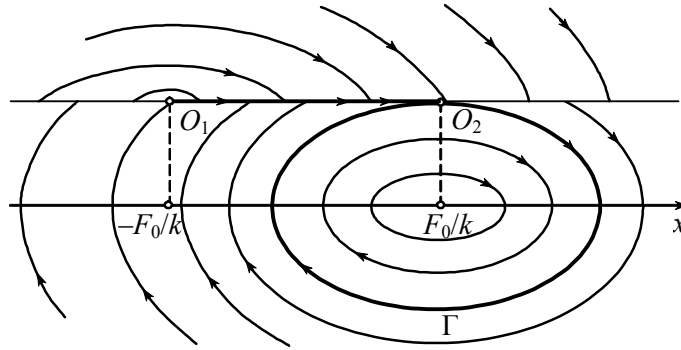


Рис. 19.4. Фазовый портрет систем типа маятника Фруда при кулоновской идеализации трения

Сделанное описание полностью определяет движение фазовых точек на всей фазовой плоскости. Однако, возможно, оно кажется не совсем обоснованным или, во всяком случае, его желательно детальнее осмыслить.

Движение массы m может быть с проскальзыванием по опорной плоскости и без него, т.е. вместе с плоскостью. В полуплоскостях $y < v$ и $y > v$ движения всюду с проскальзыванием. На лучах $(-\infty, O_1)$ и (O_2, ∞) имеют место мгновенные остановки относительно плоскости, вызванные сменой направления проскальзывания, совершаемого со скоростью $v - \dot{x}$. На отрезке $[O_1, O_2]$ проскальзывания нет, масса движется вместе с плоскостью со скоростью v . На этом отрезке $\dot{x} = y = v$. На отрезке $[O_1, O_2]$ сила трения – это сила трения покоя. На нем в течение конечного времени $y = v$ и $\dot{y} = 0$ и поэтому, в силу уравнения (19.3),

$$kx = F_0 \text{sign}(0). \tag{19.5}$$

Так как на отрезке $[O_1, O_2]$ $-F_0/k < x < F_0/k$, то сила натяжения пружины kx не превосходит величины трения покоя F_0 и проскальзывания не возникает. Напротив, на лучах $(-\infty, O_1)$ и (O_2, ∞) остановка имеет мгновенный характер, так как на них сила натяжения пружины kx превосходит предельную величину силы трения покоя F_0 .

Перейдем к общему описанию полученного фазового портрета (см. рис. 19.4) и возможных движений массы m .

Имеется равновесие $O(x = F_0/k, y = 0)$. Оно устойчивое типа центра. Возле него возможно бесконечное множество гармонических колебаний до некоторой предельной амплитуды, отвечающей кривой Γ . Эти гармонические колебания не затухают, несмотря на наличие трения. Трение как бы исчезает. Все остальные движения после нескольких колебаний с уменьшающимся размахом приводят к временному движению массы m вместе с движущейся плоскостью (к движению

фазовой точки по отрезку $[O_1, O_2]$), после чего возникают гармонические колебания, отвечающие замкнутой фазовой кривой Γ .

Мы провели теоретическое исследование движений массы на пружине, лежащей на движущейся плоскости, исходя из кулоновской модели сухого трения, и обнаружили не только устойчивое равновесие, что довольно естественно, но и, к удивлению, множество различных периодических колебаний. Однако эксперимент не подтверждает эти теоретические результаты: на небольших конечных временах теоретические и экспериментальные данные близки друг другу, однако со временем расхождения нарастают. Это значительное нарастание различий имеет место только в области фазового пространства, ограниченного кривой Γ , и состоит в том, что равновесие может быть неустойчивым и вместо бесконечного числа периодических движений либо нет ни одного, либо чаще всего одно. Это имеет место одновременно с соблюдением законов Кулона с большой точностью. Податься некуда, по-видимому, и ничтожные отличия от закона Кулона могут привести к таким существенным различиям. Это действительно так, и вы уже с этим сталкивались при рассмотрении модели «хищник – жертва» Вольтерра – Лотки. Там причина была в некорректности математической модели, и вызвана она была наличием состояния равновесия типа центр. В рассматриваемом сейчас случае центр тоже есть. Так что, возможно, именно в нем все дело и нужно уточнить математическую модель сухого трения таким образом, чтобы математическая модель рассматриваемой нами системы стала корректной. Желательно это сделать так, чтобы учесть реальные отличия характеристики трения от идеальной кулоновской и одновременно не слишком усложнить задачу исследования получающейся уточненной математической модели.

Закону Кулона отвечает график зависимости силы трения F от взаимной скорости u трущихся поверхностей, представленный на рис. 19.5.

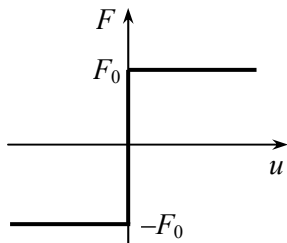


Рис. 19.5. Кулоновская идеализация сухого трения

Эксперимент дает несколько другие графики. Они в основном трех следующих типов: сила трения возрастает с увеличением скорости, убывает со скоростью, сначала убывает, а затем возрастает. Возможны и другие варианты, но оставим их без внимания. На рис. 19.6. изображены перечисленные три типа. Для первых двух изображенных случаев естественна кусочно-линейная аппроксимация, при которой сила трения равна

$$F = F_0 \operatorname{sign} u + hu, \quad (19.6)$$

где u – взаимная скорость трущихся поверхностей, в нашем случае $u = v - \dot{x}$. В третьем случае аппроксимируем силу трения полиномом третьего порядка

$$F = F_0 \operatorname{sign} u + \beta u + \gamma u^3. \quad (19.7)$$

Эти аппроксимации, естественно, могут быть хорошими только при ограниченных значениях u .

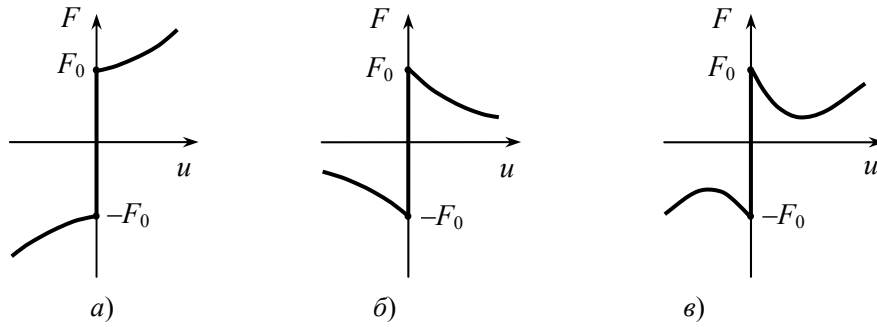


Рис. 19.6. Возможные отклонения от кулоновской характеристики трения:
 а) сила трения возрастает с возрастанием скорости; б) сила трения убывает;
 в) сила трения сначала убывает, а затем возрастает

В случае (19.6) математическая модель рассматриваемой нами системы примет вид:

$$m\ddot{x} + h\dot{x} + kx = F_0 \text{sign}(v - \dot{x}) + hu, \tag{19.8}$$

а в случае (19.7) – вид:

$$m\ddot{x} + h\dot{x} + kx = F_0 \text{sign}(v - \dot{x}) + \beta u + \gamma u^3, \tag{19.9}$$

где h, β и γ малы.

В общем же случае уравнение уточненной математической модели можно записать в виде:

$$m\ddot{x} + kx = F(u), \quad u = v - \dot{x}, \tag{19.10}$$

где $F(u)$ – кусочно-гладкая функция, имеющая разрыв непрерывности при $u = 0$. Ясно, что математическая модель (19.10) содержит модели (19.8) и (19.9) как частные случаи.

Исследование мы начнем с самой общей модели, но оно коснется только ее состояния равновесия $x = F(v)/k, \dot{x} = 0$, когда сила трения уравновешивается силой натяжения пружины. Интересовать нас будет устойчивость этого равновесия.

Для исследования устойчивости нам следует рассмотреть движения, близкие к равновесию $x = F(v)/k$. Введем новую переменную

$$\xi = x - \frac{F(v)}{k} \tag{19.11}$$

и запишем уравнение (19.10) в виде:

$$m\ddot{\xi} + k\xi = F(u) - F(v) = -\left. \frac{\partial F}{\partial u} \right|_{u=v} \dot{\xi} + \dots$$

Пренебрегая в дальнейшем членами не ниже второго порядка малости относительно малой величины $\dot{\xi}$, обозначенными в этом уравнении многоточием, рассмотрим линейризованное уравнение

$$m\ddot{\xi} + \left. \frac{\partial F}{\partial u} \right|_{u=v} \dot{\xi} + k\xi = 0. \tag{19.12}$$

Это хорошо известный нам осциллятор. Его равновесие устойчиво при

$$\left. \frac{\partial F}{\partial u} \right|_{u=v} > 0 \quad (19.13)$$

и неустойчиво при противоположном неравенстве.

Таким образом, если скорости v отвечает восходящий участок характеристики трения, то равновесие устойчиво, если нисходящий, то оно неустойчиво. Это общее утверждение, полученное из общей модели, верно, конечно, и для моделей (19.6) и (19.7). Характеристика трения модели (19.6) качественно изображена на рис. 19.6. Для нее

$$\left. \frac{\partial F}{\partial u} \right|_{u=v} = h. \quad (19.14)$$

Для модели (19.7)

$$\left. \frac{\partial F}{\partial u} \right|_{u=v} = \beta + 3\gamma v^2. \quad (19.15)$$

Поэтому состояние равновесия $x = F(v)/k$, $\dot{x} = 0$ этих моделей устойчиво соответственно при положительных h и $\beta + 3\gamma v^2$ и неустойчиво при отрицательных.

Сказать еще что-нибудь о модели (19.10), не уточняя вида характеристики трения, едва ли возможно, так что обратимся к более конкретизированным моделям (19.8) и (19.9).

Уравнение (19.8) перепишем в виде:

$$m\ddot{x} + h\dot{x} + kx = F_0 \text{sign}(v - \dot{x}) + hu.$$

В полуплоскостях $\dot{x} > v$ и $\dot{x} < v$ своей фазовой плоскости x , $\dot{x} = y$ оно принимает вид:

$$m\ddot{x} + h\dot{x} + kx = \mp F_0 + hu, \quad (19.16)$$

т.е. в каждой из них является линейным осциллятором со смещенным в точку

$$x = \frac{h}{k} v \mp \frac{F_0}{k}; \quad \dot{x} = 0$$

состоянием равновесия (в отличие от привычной записи, когда равновесие находится в точке $x = \dot{x} = 0$). Величина h , по предположению, мала, поэтому фазовые портреты – устойчивый фокус при $h > 0$ и неустойчивый при $h < 0$. При этом «скручивание» и «раскручивание» фазовых траекторий происходит в силу малости h медленно, так что существенное изменение фазовых траекторий уравнений (19.16) по отношению к фазовым траекториям ранее рассмотренных уравнений (19.4) произойдет только с замкнутыми фазовыми траекториями, охватывающими точку $(F_0/k, 0)$ (см. рис. 19.4) (точнее, близкую к ней в силу малости h точку $(F_0/k + hv/k, 0)$). Эти изменения при сколь угодно малых $h > 0$ и $h < 0$ разные и представлены на рис. 19.7,а и б. Ранее рассмотренный случай относится к $h = 0$. В первом случае при $h > 0$ всякое движение оканчивается устойчивым состоянием равновесия O с координатами F_0/k и 0 . Во втором, когда $h < 0$, это равновесие неустойчиво и всякое движение оканчивается устойчивым периодическим движением – автоколебанием, изображаемым на рис. 19.7,б замкнутой кривой Γ (на рисунке она выделена жирной линией).

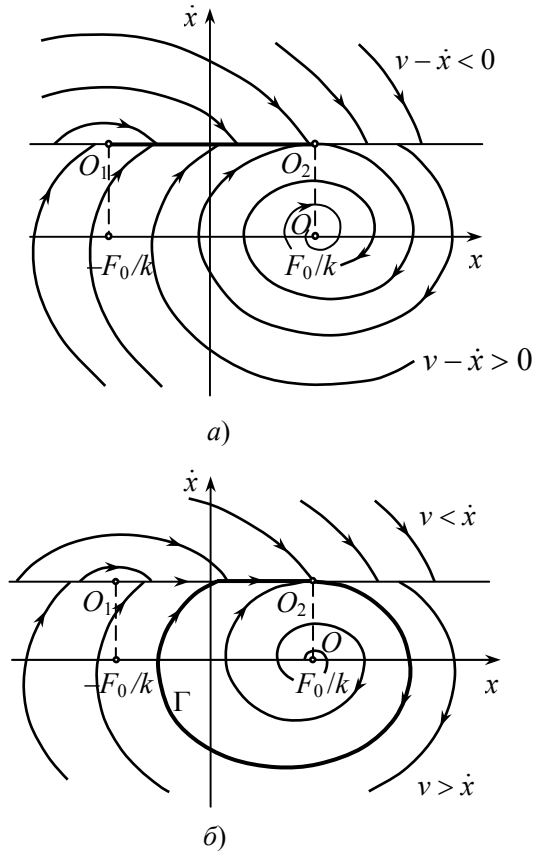


Рис. 19.7. Фазовые портреты систем типа маятника Фруда:
 а) при характеристике трения вида, показанного на рис.19.6,а;
 б) при характеристике трения вида, показанного на рис. 19.6,б

Эти качественные изменения имеют место при сколь угодно малом h . Это подтверждает и обосновывает сделанное ранее утверждение о некорректности математической модели (19.3), совпадающей с рассматриваемой моделью при $h = 0$. При $h \neq 0$ математическая модель (19.8) корректна. Это следует из общей теории грубости – структурной устойчивости – динамических систем второго порядка. Поэтому можно надеяться, что полученный результат имеет общий характер, т.е. справедлив для любых монотонных не очень крутых характеристик трения, возрастающих и убывающих. Конечно, не лишнее, прежде чем делать такое предположение, посмотреть на дисплей компьютера. Кроме того, следует иметь в виду, что сделанные аппроксимации силы трения предполагают не очень большие величины \dot{x} и v .

Перейдем к модели (19.9). Ее характеристика трения состоит из падающего и последующего восходящего участков (см. рис. 19.6,б). Возьмем за образец проведенное ранее исследование динамики лампового генератора, описываемого уравнением Ван-дер-Поля.

Пусть $v > 0$ и \dot{x} мало. Запишем уравнение (19.9) в виде:

$$m\ddot{x} + kx = F_0 + \beta v + \gamma v^3 + (-\beta - 3\gamma v^2)\dot{x} + 3\gamma v\dot{x}^2 - \gamma\dot{x}^3 = a_0 + a_1\dot{x} + a_2\dot{x}^2 + a_3\dot{x}^3.$$

Это дифференциальное уравнение после замены переменных $\xi = x - a_0/k$ приводится к виду:

$$\ddot{\xi} - 2\delta(1 - a\dot{\xi} - b\xi^2)\dot{\xi} + \omega^2\xi = 0, \quad (19.17)$$

где $2\delta = a_1/m$, $a = a_2/ma_1$, $b = a_3/ma_1$ и $\omega^2 = k/m$. При $a = 0$ оно очень похоже на уравнение Ван-дер-Поля. Член $a\dot{\xi}$, как будет видно из дальнейшего, несущественен. Уравнение (19.17) с $a = 0$ и уравнение Ван-дер-Поля являются простейшими математическими моделями автоколебательных систем.

Приступим к исследованию уравнения (19.17). Как и ранее, примем, что

$$\xi = A \cos(\omega t + \varphi), \quad (19.18)$$

и найдем производную по времени от энергии $E = (\dot{\xi}^2 + \omega^2\xi^2)/2$ соответствующего линейного осциллятора:

$$\begin{aligned} \frac{dE}{dt} &= \frac{d}{dt} \frac{\omega^2 A^2}{2} = \dot{\xi}\ddot{\xi} + \omega^2\xi\dot{\xi} = [2\delta(1 - a\dot{\xi} - b\xi^2)\dot{\xi}]\dot{\xi} = \\ &= 2\delta[1 + a\omega A \sin(\omega t + \varphi) - b\omega^2 A^2 \sin^2(\omega t + \varphi)]\omega^2 A^2 \sin^2(\omega t + \varphi). \end{aligned} \quad (19.19)$$

Напомним, что эти последние вычисления выполняются в пренебрежении членами с \dot{A} и $\dot{\varphi}$, т.е. при $\dot{A} = \dot{\varphi} = 0$.

Интегрируя (19.19) по периоду $2\pi/\omega$, находим что

$$\Delta \frac{\omega^2 A^2}{2} = \left(\delta \omega^2 A^2 - \frac{1}{2} \delta b \omega^4 A^4 \right) \frac{2\pi}{\omega}$$

или

$$\Delta(A^2) = \frac{4\pi\delta A^2}{\omega} - 2\pi\delta b \omega A^4. \quad (19.20)$$

Из (19.20) следует рекуррентное соотношение для последовательных амплитуд A_n и A_{n+1} :

$$A_{n+1}^2 = \left(1 + \frac{4\pi\delta}{\omega} \right) A_n^2 - 2\pi\delta b \omega A_n^4, \quad (19.21)$$

которое можно рассматривать как точечное отображение полупрямой в себя. Его неподвижной точке $A = 0$ соответствует состояние равновесия, устойчивое при $\delta < 0$ и неустойчивое при $\delta > 0$. При $b > 0$ имеется еще одна неподвижная точка

$$A^* = \frac{1}{\omega} \sqrt{\frac{2}{b}}, \quad (19.22)$$

которая устойчива при $\delta > 0$, т.е. при неустойчивости равновесия.

Для характеристики трения вида, изображенного на рис. 19.6,в, и приближенно аппроксимируемой полиномом (19.7), если скорость v лежит на ее падающем участке, $\delta > 0$ и $b > 0$ и, следовательно, состояние равновесия неустойчиво и возникают автоколебания с амплитудой, приближенно определяемой формулой (19.22) (рис. 19.8).

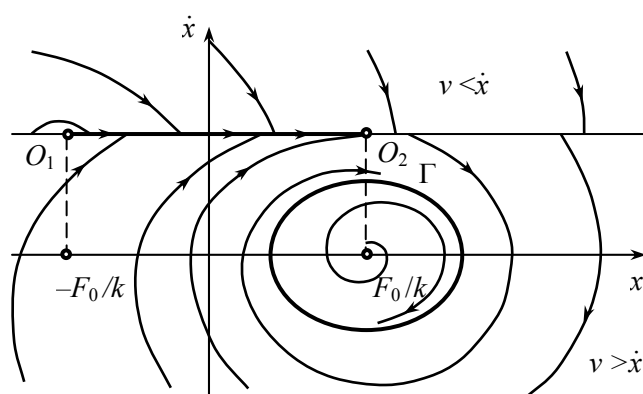


Рис. 19.8. Фазовый портрет системы типа маятника Фруда при наличии падающего участка характеристики трения, показанного на рис. 19.6,в

С некоторой долей оптимизма можно надеяться, что так же будет для любой характеристики вида, изображенного на рис. 19.6,в (но при δ малом).

20. ВЫНУЖДЕННЫЕ КОЛЕБАНИЯ ЛИНЕЙНОГО ОСЦИЛЛЯТОРА

Мы познакомились с простейшими физическими реализациями линейного осциллятора: масса на пружине при наличии вязкого трения и электрический контур, состоящий из индуктивности, емкости и сопротивления. Теперь нам предстоит рассмотреть, каков результат воздействия на эти объекты гармонической силы и, соответственно, напряжения (рис. 20.1). Математическая модель в обоих случаях одна и та же и имеет вид:

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = a \cos vt. \quad (20.1)$$

Мы значительно облегчим себе выкладки в этом и подобных случаях, если прибегнем к так называемой комплексной форме записи, заменив уравнение (20.1) уравнением

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = ae^{ivt}. \quad (20.2)$$

Понимать его можно как то, что его действительная часть – это и есть рассматриваемое уравнение (20.1) и, в соответствии с этим, $a \exp(ivt)$ – это комплексная запись действительного воздействия $a \cos vt$, а в (20.1) x есть действительная часть x , входящего в (20.2).

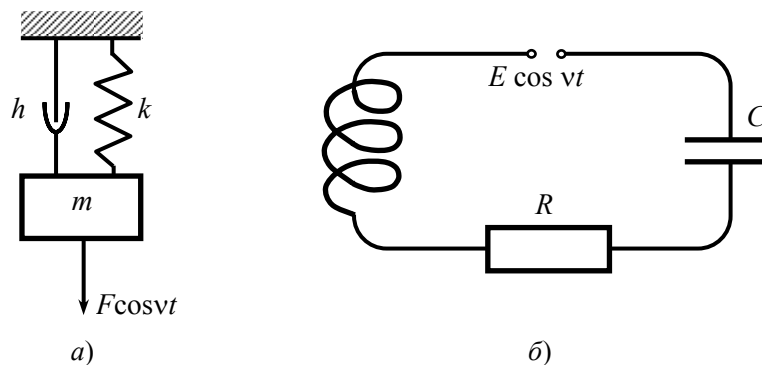


Рис. 20.1. Условное изображение конкретных реализаций механического (а) и электрического (б) осцилляторов с действующими на них внешними гармоническими воздействиями: силой $F \cos vt$ и напряжением $E \cos vt$

Решение неоднородного уравнения (20.2) состоит из общего решения однородного, которое хорошо нам известно, и какого-нибудь решения неоднородного уравнения, которое мы постараемся найти. Будем его искать в виде $x = A \exp(ivt)$, т.е. как колебания той же частоты, что и внешнее воздействие $a \exp(ivt)$. Прямая подстановка дает

$$(-v^2 + 2\delta vi + \omega^2)A = a,$$

откуда

$$A = \frac{a}{\omega^2 - v^2 + 2\delta vi}. \quad (20.3)$$

Если предположить, что общее решение однородного уравнения при $t \rightarrow \infty$ исчезает – это имеет место при $\delta > 0$ и $\omega^2 > 0$, то остающееся при неограниченном возрастании времени колебание осциллятора (20.2) совпадает с найденным нами частным решением неоднородного уравнения

$$x = \frac{a}{\omega^2 - \nu^2 + 2\delta\nu i} e^{i\nu t}. \quad (20.4)$$

Выясним, что оно собою представляет. Начнем с общего рассмотрения. Пусть

$$x = K(i\nu)e^{i\nu t}, \quad (20.5)$$

где $K(i\nu)$ – комплексная функция частоты ν . Реальный смысл имеет действительная часть (20.5). Найдем ее, сохранив за ней обозначение x :

$$x = \operatorname{Re} (K(i\nu)e^{i\nu t}) = \operatorname{Re} [|K(i\nu)| e^{i(\nu t + \arg K(i\nu))}] = |K(i\nu)| \cos(\nu t + \varphi),$$

где $\varphi = \arg K(i\nu)$.

Таким образом, найденное нами частное решение (20.4), которое устанавливается с возрастанием t при асимптотической устойчивости осциллятора, называемое поэтому вынужденным колебанием, является гармоническим колебанием с частотой внешнего воздействия ν , амплитудой $|K(i\nu)|$ и фазой $\varphi = \arg K(i\nu)$.

Обратите внимание, что характеристики вынужденного гармонического колебания – амплитуда и фаза – находятся по одному и тому же комплексному числу $K(i\nu)$, которое определяет эти величины как функции частоты внешнего воздействия ν . Благодаря этому зависимость $K(i\nu)$ получила наименование амплитудно-фазовой частотной характеристики, т.е. характеристики, объединяющей частотные зависимости амплитуды и фазы одновременно.

Частотной амплитудно-фазовой характеристике можно придать геометрически наглядную форму. Для этого на комплексной плоскости W нарисуем годограф (кривую), пробегаемую комплексной точкой $W = K(i\nu)$ при изменении ν от 0 до ∞ (рис. 20.2). Длина радиуса-вектора точки W этой кривой, отвечающего частоте ν , дает амплитуду, а угол с действительной осью – фазу соответствующего вынужденного колебания.

Кривая-годограф, отвечающая формуле (20.3), легко строится по трем своим характерным точкам, для которых $\nu = 0$, $\nu = \omega$ и $\nu = \infty$ (рис. 20.3). При малом $\delta > 0$ годограф сильно вытягивается вдоль отрицательного направления мнимой оси, принимая при $\nu = \omega$ значение $-ia/2\delta\omega$. Начинается он в точке $\nu = 0$ и кончается в точке $\nu = \infty$.

По амплитудно-фазовой частотной характеристике рис. 20.3 легко построить амплитудную зависимость – резонансную кривую – и фазовую зависимость от частоты внешнего воздействия ν . Эти зависимости при разных $\delta > 0$ качественно представлены на рис. 20.4. Из них следует широко известное явление амплитудного резонанса – большое возрастание амплитуды вблизи $\nu = \omega$ при $\delta \ll 1$ – и менее известное, но не менее замечательное явление противофазы вынужденных колебаний по отношению к внешнему воздействию при $\nu > \omega$. При $\delta = 0$ фаза колебаний зависит от частоты ν разрывно: при $\nu < \omega$ она равна нулю, а при $\nu > \omega$ она равна $-\pi$. При небольшом $\delta > 0$ эта зависимость уже непрерывная, но с быстрым

изменением фазы φ вблизи $\nu = \omega$ (рис. 20.5). Резонанс тем ярче выражен, чем меньше затухание δ . При малом δ максимальная амплитуда достигается примерно при $\nu = \omega$, а точнее, при $\nu = \sqrt{\omega^2 - 2\delta^2}$, ибо именно при этом значении частоты внешнего воздействия достигается максимум модуля величины A , определяемой формулой (20.3) (см. рис. 20.4).

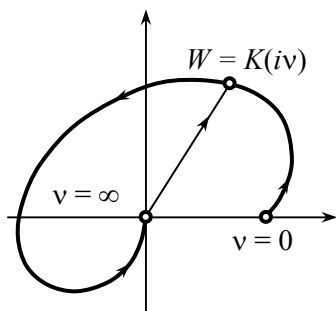


Рис. 20.2. Амплитудно-фазовая частотная характеристика $W = K(iv)$

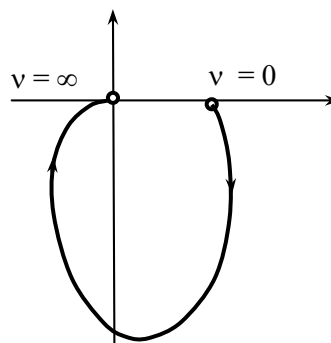


Рис. 20.3. Амплитудно-фазовая частотная характеристика линейного осциллятора

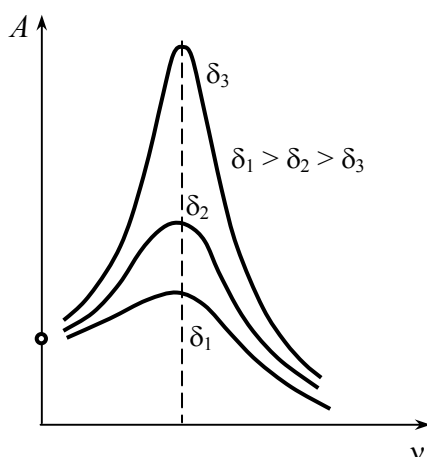


Рис. 20.4. Резонансные кривые для трех различных значений $\delta > 0$

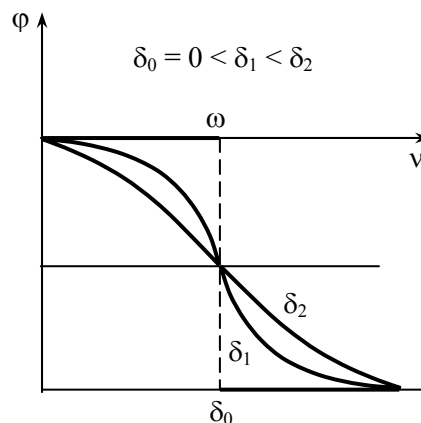


Рис. 20.5. Фазовая частотная характеристика для трех убывающих значений коэффициента δ

Оба эти явления ярко иллюстрируются бортовой и килевой качками корабля на бегущих волнах. В обоих случаях приближенной моделью качки корабля является осциллятор, на который действует внешняя гармоническая сила, на что в свое время обратил внимание А.Н. Крылов. Особый интерес представляет килевая качка корабля, когда при достаточно большой амплитуде воздействия волн на корабль их частота зависит от скорости и курса корабля. При неудачных курсе и скорости набегающие волны могут иметь частоту, несколько большую собственной частоты килевой качки, и тогда явление резонанса будет еще сопровождаться тем, что корабль не взбегает носом на приближающуюся волну, а опускается вниз и зарывается в нее носом.

Найдем условия возникновения такой нежелательной и опасной качки. Пусть θ – угол между направлениями движения волны со скоростью u и корабля, движущегося со скоростью ν навстречу волне. При этом скорость набегающих волн на

корабль равна $u + v \cos \theta$ (рис. 20.6), а частота их набегания $\nu = 2\pi\lambda^{-1}(u + v \cos \theta)$, где λ – длина бегущих волн. Пусть ω – частота собственных свободных колебаний корабля. Тогда особенно опасна ситуация, когда при близости частот ν и ω $\nu > \omega$, так как именно в этом случае имеет место близость к резонансу, а колебания носа корабля находятся в противофазе с набегающей волной.

Следующий пример вынужденных колебаний системы несколько более сложный, чем осциллятор. Основной его частью является машина на упругом основании под действием возникающей при ее работе гармонической силы. Ее математическая модель совпадает с (20.1). Периодическое воздействие может иметь своей причиной, например, наличие эксцентрических вращающихся деталей или деталей, поступательно движущихся туда и обратно с некоторой частотой ν . Под влиянием этой силы машина начинает вибрировать, что может быть весьма нежелательным.

Как подавить или хотя бы уменьшить эти вибрации? Одним из эффективных средств их подавления является динамический демпфер, который представляет собой массу с пружиной, подсоединяемой к основной большой массе вибрирующей машины (рис. 20.7).

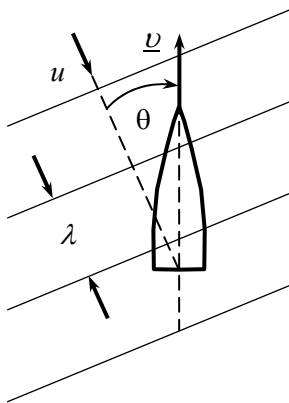


Рис. 20.6. Курс корабля, движущегося со скоростью v под углом θ к набегающим со скоростью u волнам

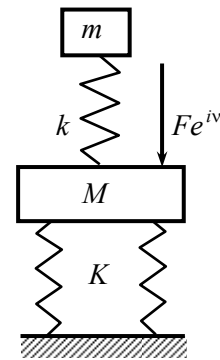


Рис. 20.7. Схема механического резонансного демпфера вынужденных колебаний массы M гармонической силой $F e^{i\nu t}$

Пусть величины перемещений машины и демпфера от их равновесных положений соответственно x и y ; M и m – массы машины и демпфера, а K и k – жесткости основания (крепления) машины и пружины демпфера. Математическую модель машины на упругом основании с динамическим демпфером (см. рис. 20.7) можно записать в виде:

$$\begin{aligned} M\ddot{x} &= -Kx - k(x - y) + F e^{i\nu t}, \\ m\ddot{y} &= k(x - y). \end{aligned} \tag{20.6}$$

В этой модели ради простоты трение не учитывается. Для реального расчета динамического демпфера этот учет, конечно, необходим, но для наших целей демонстрационного характера он несущественен и только усложняет рассмотрение.

Уравнение (20.6) можно написать, исходя непосредственно из законов Ньютона или на основе функции Лагранжа, которая в данном случае равна

$$L = \frac{M\dot{x}^2}{2} + \frac{m\dot{y}^2}{2} - \frac{Kx^2}{2} - \frac{k(x-y)^2}{2},$$

и, кроме того, есть силы

$$Q_x = Fe^{ivt}, \quad Q_y = 0.$$

Найдем вынужденные колебания модели (20.6). В соответствии с предыдущим примем, что

$$x = Ae^{ivt}, \quad y = Be^{ivt}, \quad (20.7)$$

и подставим их в (20.6). Имеем

$$\begin{aligned} -Mv^2 A &= -KA - k(A - B) + F, \\ -mv^2 B &= k(A - B), \end{aligned} \quad (20.8)$$

откуда

$$\begin{aligned} A &= \frac{F(k - mv^2)}{(K + k - Mv^2)(k - mv^2) - k^2}, \\ B &= \frac{kF}{(K + k - Mv^2)(k - mv^2) - k^2}. \end{aligned} \quad (20.9)$$

Из этих формул следует, что вынужденные колебания машины на основании имеют амплитуду A , обращаясь в нуль при

$$k - mv^2 = 0. \quad (20.10)$$

Масса демпфера при этом колеблется с амплитудой $B = -F/k$, так что пружины демпфера действуют на массу машины с силой $kBe^{ivt} = -Fe^{ivt}$, в точности компенсирующей воздействие Fe^{ivt} . Так как, согласно (20.10), $k = mv^2$, а масса m мала, то амплитуда колебаний демпфера может быть большой и именно такой, чтобы нейтрализовать воздействие на машину гармонической силы.

Это замечательный и удивительный результат: демпфер колеблется, а вибраций машины нет. Однако как же заколебался демпфер? По-видимому, ответ может быть только один: машина вначале колебалась и возбудила колебания демпфера, колебания демпфера подавили колебания машины и они исчезли, остались одни колебания демпфера. Это довольно правдоподобно, но не совсем так, поскольку о вынужденных колебаниях, как об установившихся, можно говорить только при затухании собственных, т.е. при наличии трения. Так что это может быть так только при наличии исчезающе малых трений. Они, конечно, всегда есть и все в конечном счете верно. Только при наличии очень малых трений вибрации машины не отсутствуют, а очень малы.

Мы рассмотрели действие гармонической силы на линейный осциллятор, нашли вызываемые им колебания и обнаружили явления синхронизма, резонанса и сдвига фаз.

Синхронизм выражается в том, что гармоническое воздействие частоты ν вызывает гармонические вынужденные колебания осциллятора той же частоты. Явления резонанса проявляются тем сильнее, чем меньше параметр трения $\delta > 0$, так что в резонансе амплитуда вынужденных колебаний неограниченно растет при $\delta \rightarrow 0$.

Явление сдвига фазы – это ее зависимость от частоты воздействия ν . Она такова, что при маленьких частотах сдвиг фаз близок к нулю, при больших – к $-\pi$, т.е. к противофазе. При $\delta \rightarrow 0$ эта зависимость приобретает скачкообразный характер: при $\nu < \omega$ сдвиг фаз равен 0, при $\nu > \omega$ он равен $-\pi$.

Мы ввели еще понятие частотной амплитудно-фазовой характеристики $K(i\nu)$, так что в комплексной записи связь между гармоническим воздействием $ae^{i\nu t}$ и вынужденными колебаниями $Ae^{i\nu t}$ записывается в очень простом виде:

$$A = K(i\nu)a, \quad (20.11)$$

напоминающем закон рычага. Конечно, из этой записи следуют явления синхронизма, резонанса и сдвига фаз. Таковы успехи хорошей теории: сложное становится простым.

К этому краткому итогу наших исследований мне хочется сделать одно важное добавление. Формула (20.11) относится к гармоническому воздействию. Нельзя ли ее обобщить на произвольное периодическое воздействие? Оказывается, это возможно. Возможно в силу применимости принципа суперпозиции и представимости любого периодического воздействия в виде суммы гармонических. С представимостью вас познакомили в курсе математического анализа, где она называется рядом Фурье.

Согласно теории рядов Фурье, периодическую функцию $f(t)$ с периодом $T = 2\pi/\nu$ при выполнении для нас неограничительных условий можно записать в виде:

$$f(t) = \sum a_s \cos s\nu t + b_s \sin s\nu t, \quad (20.12)$$

где

$$a_0 = \frac{\nu}{2\pi} \int_0^T f(t) dt, \quad a_s = \frac{\nu}{\pi} \int_0^T f(t) \cos s\nu t dt,$$

$$b_s = \frac{\nu}{\pi} \int_0^T f(t) \sin s\nu t dt \quad (s = 1, 2, \dots),$$

или в виде

$$f(t) = \operatorname{Re} \sum_{s=0}^{\infty} (a_s - ib_s) e^{is\nu t}. \quad (20.13)$$

Последняя формула (20.13) означает, что периодическое воздействие является суммой гармонических воздействий (в комплексном представлении)

$$f(t) = \operatorname{Re} \sum_{s=0}^{\infty} c_s e^{is\nu t}, \quad c_s = a_s - ib_s. \quad (20.14)$$

Такое представление функции $f(t)$ называется спектральным, частоты входящих в него гармонических составляющих – его спектр. С нулевой частотой входит постоянное воздействие величины $c_0 = a_0$. Гармоническая составляющая с частотой ν называется основной, остальные воздействия, называемые гармониками, имеют частоты $2\nu, 3\nu, \dots$. Амплитуды этих гармонических составляющих, основной и гармоник, равны соответственно $|c_1|, |c_2|, \dots$

Теперь, чтобы найти вынужденные колебания осциллятора, отвечающие произвольному периодическому воздействию $f(t)$, достаточно воспользоваться принципом суперпозиции. Отдельному гармоническому воздействию, входящему в (20.14), отвечает вынужденное колебание

$$c_s K(isv) e^{isvt}, \quad (20.15)$$

поэтому сумме воздействий (20.14), т.е. воздействию $f(t)$, отвечает сумма вынужденных гармонических колебаний (20.15)

$$\sum_{s=0}^{\infty} c_s K(isv) e^{isvt}. \quad (20.16)$$

Соответствующие (20.16) гармонические колебания имеют частоты sv , амплитуды $|c_s K(isv)|$ и фазы $\arg [c_s K(isv)]$. Явление синхронизма сохраняется в том смысле, что ответ осциллятора имеет тот же период, что и воздействие, явление сдвига фаз при периодическом воздействии (20.14) теряет смысл, а явление резонанса сохраняется, однако теперь резонанс возможен не только при $v = \sqrt{\omega^2 - 2\delta^2}$, но и при $sv = \sqrt{\omega^2 - 2\delta^2}$ с любым целым s , т.е. не только на основной частоте, но и на ее гармониках. Но в связи с тем, что $c_s \rightarrow 0$ при $s \rightarrow \infty$, явление резонанса может проявляться только на не очень далеких гармониках. Это при $\delta > 0$; при $\delta = 0$ резонанс возможен на любой гармонике.

Сказанное о периодическом внешнем воздействии на осциллятор может быть обобщено на произвольное воздействие $f(t)$. Это дальнейшее обобщение также основывается на принципе суперпозиции, но исходит из более общих способов представления воздействия $f(t)$ в виде суперпозиции более простых. При этом под более простыми понимаются такие воздействия, для которых отклик осциллятора может быть легко найден. Как было показано ранее, таким простым воздействием было гармоническое воздействие e^{ivt} . Отклик на него также гармонический и равен $K(iv)e^{ivt}$, где $K(iv)$ – амплитудно-фазовая частотная характеристика осциллятора. Далее, мы воспользовались тем, что периодическое воздействие $f(t)$ периода T может быть представлено в виде ряда Фурье, т.е. в виде суперпозиции гармонических колебаний частот $2\pi n/T$ с $n = 1, 2, 3, \dots$ и некоторого постоянного воздействия. Но, кроме ряда Фурье, существует еще интеграл Фурье, который любую функцию $f(t)$, достаточно быстро убывающую при $t \rightarrow \pm\infty$, представляет в виде интеграла

$$f(t) = \int_{-\infty}^{\infty} g(v) e^{ivt} dv, \quad (20.17)$$

т.е. тоже в виде суперпозиции гармонических воздействий $g(v)e^{ivt} dv$ частоты v , амплитуды $|g(v) dv|$ и фазы $\arg [g(v) dv]$. При этом, как известно, функция $g(v)$, называемая преобразованием Фурье функции $f(t)$, находится по формуле

$$g(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-ivt} dt. \quad (20.18)$$

На каждую гармоническую составляющую $g(v)e^{ivt} dv$ интегральной суммы (20.17) откликом осциллятора является $K(iv)g(iv)e^{ivt} dv$, а отклик на все воздействие $f(t)$ будет

$$x(t) = \int_{-\infty}^{\infty} K(iv)g(v)e^{ivt} dv. \quad (20.19)$$

Наряду со спектральным представлением внешнего воздействия $f(t)$ возможно и импульсное представление, т.е. представление $f(t)$ как последовательности мгновенных воздействий – импульсов. В момент t за время dt величина этого импульса равна $f(t) dt$, где длительность dt исчезающе мала. Задать такой импульс с помощью обычных функций невозможно и это достигается с помощью обобщенной функции, получившей название дельта-функции. Мгновенный единичный импульс в момент времени τ записывается в виде $\delta(t - \tau)$. Составить представление об этом описании, об этой математической модели мгновенного единичного импульса можно, представив его как предел пикообразной функции $\varphi(t - \tau, \varepsilon)$, изображенной на рис. 20.8 при $\varepsilon \rightarrow 0$.

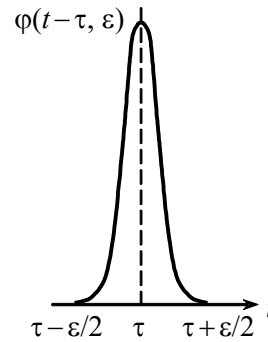


Рис. 20.8. График импульсного воздействия

При этом предполагается, что несмотря на неограниченное сужение области ширины ε возле точки τ , в которой функция $\varphi(t - \tau, \varepsilon)$ отлична от нуля, интеграл

$$\int_{-\infty}^{\infty} \varphi(t - \tau, \varepsilon) dt = \int_{\tau - \varepsilon/2}^{\tau + \varepsilon/2} \varphi(t - \tau, \varepsilon) dt \quad (20.20)$$

все время равен единице. Ясно, что такое требование ведет к неограниченному росту функции $\varphi(t - \tau, \varepsilon)$ при $\varepsilon \rightarrow 0$ и отсутствию предельной функции в обычном понимании. Обойдем эту трудность, назвав предельную функцию обобщенной и записав в виде $\delta(t - \tau)$. С помощью этой обобщенной функции любое воздействие $f(t)$ можно записать в виде

$$f(t) = \int_{-\infty}^{\infty} f(\tau) \delta(t - \tau) d\tau. \quad (20.21)$$

Действительно, согласно теореме о среднем,

$$\int_{-\infty}^{\infty} f(\tau) \delta(t - \tau) d\tau = \lim_{\varepsilon \rightarrow 0} \int_{\tau - \varepsilon/2}^{\tau + \varepsilon/2} f(\tau) \varphi(t - \tau, \varepsilon) d\tau =$$

$$= \lim_{\varepsilon \rightarrow \infty} f(\tilde{\tau}) \int_{\tau-\varepsilon/2}^{\tau+\varepsilon/2} \varphi(t-\tau, \varepsilon) d\tau = f(t).$$

Напомним, что $\tilde{\tau}$ – это некоторое значение, лежащее в суживающемся интервале $(t - \varepsilon/2, t + \varepsilon/2)$.

Формула (20.21) – это представление внешнего воздействия в виде суперпозиции импульсных воздействий, описываемых обобщенной функцией $\delta(t - \tau)$. Обозначим отклик осциллятора на единичный импульс через $\psi(t - \tau)$. С его помощью, согласно принципу суперпозиции, отклик осциллятора $x(t)$ на произвольное воздействие $f(t)$ можно записать в виде

$$x(t) = \int_{-\infty}^{\infty} f(\tau) \psi(t - \tau) d\tau. \quad (20.22)$$

Осталось найти отклик $\psi(t - \tau)$ осциллятора на единичный импульс $\delta(t - \tau)$, т.е. на воздействие $f(t) = \delta(t - \tau)$. Но прежде чем приступить к поиску этого отклика $\psi(t - \tau)$, уточним условия его действия, как и то, что мы понимаем под откликом осциллятора на воздействие $f(t)$.

Дело в том, что движение осциллятора определяется его начальным состоянием и последующим внешним на него воздействием. Поэтому естественно под откликом на внешнее воздействие понимать только ту его часть, которая отвечает этому воздействию, а не начальному возмущению осциллятора, его начальным, отличным от нуля условиям. Итак, под $\psi(t - \tau)$ мы понимаем результат воздействия импульса $\delta(t - \tau)$ при нулевых начальных условиях, т.е. мы хотим решить задачу: найти решение дифференциального уравнения

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = \delta(t - \tau) \quad (20.23)$$

при условиях

$$x(\tau - \xi) = 0, \quad \dot{x}(\tau - \xi) = 0 \quad (\xi > 0). \quad (20.24)$$

Позволим себе найти его не формально. Затем можно проверить прямой постановкой верность полученного результата.

Мгновенный импульс вызывает скачок скорости \dot{x} , а координата x , ввиду бесконечной малости времени импульса и конечности скорости \dot{x} , не меняется. Поэтому интегрируя (20.23) в пределах от $\tau - \varepsilon/2$ до $\tau + \varepsilon/2$ и устремляя ε к нулю, имеем

$$\Delta\dot{x} = \dot{x}(\tau + 0) - \dot{x}(\tau - 0) = 1,$$

откуда $\dot{x}(\tau + 0) = 1$, так как $\dot{x}(\tau - 0) = 0$. К этому добавим, что $x(\tau + 0) = x(\tau - 0) = x(\tau) = 0$.

Таким образом, откликом на воздействие $\delta(t - \tau)$ является решение уравнения:

$$\begin{aligned} \ddot{x} + 2\delta\dot{x} + \omega^2 x &= 0, \\ x(\tau) &= 0, \quad \dot{x}(\tau) = 1, \end{aligned}$$

при $t < \tau$ равное нулю, а при $t \geq \tau$ определяемое начальными условиями

$$x(\tau) = 0, \quad \dot{x}(\tau) = 1.$$

Это решение, согласно (1.7), при $\omega^2 - \delta^2 > 0$ записывается в виде:

$$x(t) = \psi(t - \tau) = \begin{cases} 0 & \text{при } t \leq \tau, \\ e^{-\delta(t-\tau)} \frac{1}{\Omega} \sin \Omega(t - \tau) & \text{при } t \geq \tau. \end{cases} \quad (20.25)$$

В заключение я хотел бы отметить, что полученная нами формула отклика (20.22) (с учетом (20.25)) является прямым следствием использования новой, непривычной для вас обобщенной функции $\delta(t)$, обладающей свойством

$$\delta(t) = \begin{cases} 0 & \text{при } t \neq 0, \\ \infty & \text{при } t = 0, \end{cases} \quad \int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (20.26)$$

Несмотря на необычность функции $\delta(t)$, она всего лишь естественная математическая модель мгновенного импульсного воздействия. Одновременно она же – математическая модель плотности единичной точечной массы на прямой или точечного единичного заряда. Впервые эту функцию в науке ввел Хэвисайд в связи с созданным им операционным исчислением, позднее – физик Дирак в связи с разрабатываемой им квантовой физикой. Физики ее называют функцией Дирака, математики – обобщенной δ -функцией или просто δ -функцией.

21. ПАРАМЕТРИЧЕСКОЕ ВОЗБУЖДЕНИЕ И СТАБИЛИЗАЦИЯ

В предыдущих лекциях мы рассмотрели действие внешней гармонической силы на линейный осциллятор, в результате которого его дифференциальное уравнение принимает вид:

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = a \cos vt. \quad (21.1)$$

Но воздействие, зависящее от времени, может вызывать и другие изменения уравнения линейного осциллятора. Например, оно может состоять в том, что параметр частоты ω^2 меняется гармонически со временем и тогда уравнение осциллятора принимает вид:

$$\ddot{x} + 2\delta\dot{x} + (\omega^2 + \varepsilon \cos vt)x = 0. \quad (21.2)$$

Непосредственной физической моделью для уравнения (21.2) является масса на пружине с переменной гармонически меняющейся жесткостью при наличии вязкого сопротивления. С приемлемым приближением уравнением (21.2) описываются малые колебания маятника с вертикально гармонически колеблющейся точкой подвеса. Если точка подвеса колеблется горизонтально, то с аналогичным приближением приходим к уравнению (21.1).

Рассмотрим вместе оба этих случая, предполагая, что точка подвеса маятника смещается по горизонтали и вертикали согласно законам $u(t)$ и $v(t)$ (рис. 21.1).

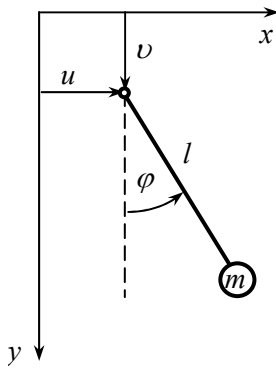


Рис. 21.1. Маятник с перемещаемой точкой подвеса (u и v – величины перемещений в горизонтальном и вертикальном направлениях)

Для составления уравнений движения мы прибегнем к уравнению Лагранжа. Выразим координаты x и y массы m маятника через смещения $u(t)$ и $v(t)$ и угол φ отклонения его от вертикали (см. рис. 21.1):

$$x = u + l \sin \varphi, \quad y = v + l \cos \varphi$$

и запишем функцию Лагранжа

$$L = \frac{m}{2} \{(\dot{u} + l\dot{\varphi} \cos \varphi)^2 + (\dot{v} - l\dot{\varphi} \sin \varphi)^2\} + mg(l \cos \varphi + v).$$

После несложных вычислений находим искомое уравнение Лагранжа в виде:

$$\ddot{\varphi} + \frac{g}{l} \sin \varphi = -\frac{1}{l} \ddot{u} \cos \varphi + \frac{1}{l} \ddot{v} \sin \varphi$$

или, учитывая еще вязкое трение и заменяя $\sin \varphi$ на φ , а $\cos \varphi$ на 1, в виде:

$$\ddot{\varphi} + 2\delta\dot{\varphi} + \omega^2\varphi = -\frac{1}{l}\ddot{u} + \frac{1}{l}\ddot{v}\varphi. \quad (21.3)$$

Уравнение (21.3) переходит в (21.1) и (21.2) соответственно при гармонических горизонтальных ($u = alv^{-2}\cos vt$, $v = 0$) и вертикальных ($u = 0$, $v = \varepsilon lv^{-2}\cos vt$) колебаниях точки подвеса маятника.

Таким образом, если гармонически горизонтально перемещать точку подвеса маятника, то это равнозначно действию на маятник гармонической силы. Если же точку подвеса гармонически перемещать вертикально, то это равнозначно гармоническому изменению частоты маятника. Это позволяет нам наглядно наблюдать, что происходит с маятником в первом и втором случаях, описываемых дифференциальными уравнениями (21.1) и (21.2). Воспользуемся этим, чтобы еще раз продемонстрировать явления синхронизма, сдвига фаз и резонанса, о которых была речь в предыдущей лекции, а в заключение этой лекции продемонстрируем вывод и исследование дифференциального уравнения (21.2). Для демонстрации я воспользуюсь одной из сумочек присутствующих здесь студенток. Я держу ее за ремешок на весу. Затем начинаю перемещать руку периодически вправо и влево. Сумка начинает качаться в такт с рукой, демонстрируя явление синхронизма. Резонанс наступает тогда, когда сумка сильно качается, а перемещения руки практически незаметны; сумка колеблется влево, а рука перемещается вправо, когда частота движения руки больше собственной частоты колебаний сумки, при меньшей частоте движения руки рука и сумка движутся одновременно вправо и влево.

Вернемся к уравнению (21.2). Путем замены переменных $x = ye^{-\delta t}$ в нем можно убрать член вязкого трения и привести уравнение к виду:

$$\ddot{y} + (\Omega^2 + \varepsilon \cos vt)y = 0,$$

где $\Omega^2 = \omega^2 - \delta^2$. В свою очередь, замена масштаба времени $\tau = vt$ позволяет оставить в нем только два параметра, приведя его к виду широко известного уравнения Маттье:

$$\ddot{y} + (v^{-2}\Omega^2 + \varepsilon v^{-2}\cos \tau)y = 0,$$

которое запишем, возвращаясь к исходным обозначениям, в виде:

$$\ddot{x} + (\omega^2 + \varepsilon \cos t)x = 0. \quad (21.4)$$

Исследование уравнения Маттье (21.4) не очень просто и может служить примером, когда хорошая идея порождает метод исследования, само исследование ведет к вычислениям, а вычисления приводят к четким и неожиданным результатам.

Идея идет от великого математика А. Пуанкаре и состоит в сведении изучения решений уравнения Маттье к точечному отображению. Метод и вычисления состоят в построении, изучении и численном исследовании этого точечного отображения.

Пусть x и $\dot{x} = y$ – начальные значения при $t = 0$, а \bar{x} и \bar{y} – значения этих же переменных спустя время 2π . Значения \bar{x} и \bar{y} однозначно находятся по x и y , так что

$$(\bar{x}, \bar{y}) = T(x, y), \quad (21.5)$$

где T – некоторый не известный нам однозначный оператор, переводящий x и y в \bar{x} и \bar{y} .

Соотношения (21.5) можно трактовать и как точечное отображение точки (x, y) в точку (\bar{x}, \bar{y}) . С такой трактовкой мы уже сталкивались при рассмотрении динамики часов Галилея–Гюйгенса и «часов наоборот» (стохастического осциллятора). Это были диаграммы Кенигса–Ламерея, которые геометрически представляли одномерные точечные отображения. Теперь перед нами – двумерное точечное отображение. К сожалению, столь же удобное геометрическое представление для него неизвестно. Но и без него использование точечного отображения оказывается очень полезным. Полезность эта состоит в том, что продолжение решения уравнения Матье (21.2) с начальными условиями x, y на времени $2\pi, 4\pi, 6\pi, \dots$ сводится к однократному, двукратному, трехкратному, и т.д. преобразованиям (21.5), т.е. при любом целом положительном n и любом решении $x(t), y(t) = \dot{x}(t)$ уравнения Матье (21.2) имеет место

$$(x(2n\pi), y(2n\pi)) = T^n(x(0), y(0)). \quad (21.6)$$

Этот очевидный факт сводит исследование поведения решений уравнения Матье (21.2) при неограниченном возрастании времени t к рассмотрению итераций (последовательных преобразований) точки (x, y) .

Итак, идея состоит в том, чтобы вместо решений уравнения Матье (21.2) и его трехмерного фазового портрета изучать отображение T плоскости переменных x, y в себя. Найдем отображение T . Для этого надлежит проинтегрировать дифференциальное уравнение Матье (21.4). Пусть мы это сделали для начальных условий $x = 1, y = 0$ и $x = 0, y = 1$. Обозначим эти решения $\varphi_1(t)$ и $\varphi_2(t)$. Тогда при произвольных начальных условиях x, y решение $x(t)$ в силу линейности уравнения Матье запишется в виде:

$$x(t) = x\varphi_1(t) + y\varphi_2(t)$$

и

$$y(t) = \dot{x}(t) = x\dot{\varphi}_1(t) + y\dot{\varphi}_2(t).$$

Поэтому

$$\begin{aligned} \bar{x} &= x(2\pi) = \varphi_1(2\pi)x + \varphi_2(2\pi)y, \\ \bar{y} &= \dot{x}(2\pi) = \dot{\varphi}_1(2\pi)x + \dot{\varphi}_2(2\pi)y, \end{aligned} \quad (21.7)$$

т.е. мы нашли отображение T , и оно оказалось линейным. Для его отыскания нам потребовалось знание только двух частных решений $\varphi_1(t)$ и $\varphi_2(t)$. Даже не самих решений, а только их значений и значений их производных при $t = 2\pi$.

Линейное отображение мы исследовать умеем, и как оно устроено нам хорошо известно. Применим эти знания. Самым важным являются собственные значения матрицы линейного отображения T , т.е. корни квадратного уравнения

$$\begin{vmatrix} \varphi_1(2\pi) - \lambda & \varphi_2(2\pi) \\ \dot{\varphi}_1(2\pi) & \dot{\varphi}_2(2\pi) - \lambda \end{vmatrix} = \quad (21.8)$$

$$= \lambda^2 - (\varphi_1(2\pi) + \dot{\varphi}_2(2\pi))\lambda + \varphi_1(2\pi)\dot{\varphi}_2(2\pi) - \dot{\varphi}_1(2\pi)\varphi_2(2\pi) = 0.$$

Это уравнение не общего вида: его свободный член равен единице. Действительно, пусть

$$\varphi_1(t)\dot{\varphi}_2(t) - \dot{\varphi}_1(t)\varphi_2(t) = W(t).$$

При $t = 0$ $W(t) = 1$. Далее обнаруживаем, что $\dot{W}(t) = 0$ и поэтому $W(2\pi) = 1$, т.е. свободный член тоже равен единице. Итак, характеристическое уравнение (21.8) имеет вид:

$$\lambda^2 - 2A\lambda + 1 = 0, \tag{21.9}$$

где

$$2A = \varphi_1(2\pi) + \dot{\varphi}_2(2\pi).$$

Корни λ_1 и λ_2 уравнения (21.9)

$$\lambda_{1,2} = A \pm \sqrt{A^2 - 1}$$

при $|A| > 1$ – действительные различные, один по модулю больше единицы, другой – меньше (следует иметь в виду, что произведение корней равно единице); при $|A| < 1$ – корни комплексные сопряженные, по модулю равные единице. Переход от первого случая ко второму происходит при $|A| = 1$: при $A = 1$ – через двойной корень, равный единице, и при $A = -1$ – через двойной корень, равный минус единице.

Как известно, при различности корней λ_1 и λ_2 , т.е. $|A| \neq 1$, линейное отображение T после однородной линейной замены переменных (возможно, комплексной) представимо в диагональном виде:

$$\bar{u} = \lambda_1 u, \quad \bar{v} = \lambda_2 v. \tag{21.10}$$

При $|A| > 1$ корни λ_1 и λ_2 – действительные, замена переменных также действительная и, согласно (21.10), отображение T в одном направлении сжимающее, а в другом – растягивающее. При $A < -1$ растягивание и сжатие сопровождаются еще симметричным отображением относительно преобразуемой в себя точки $x = y = 0$, совпадающей с точкой $u = v = 0$. Ясно, что при этом почти все точки x, y при последовательных преобразованиях удаляются в бесконечность. Исключение составляют точки линии $u = 0$ или $v = 0$, вдоль которой имеет место сжатие (соответствующее λ по модулю меньше единицы).

Из этого следует, что при $|A| > 1$ состояние равновесия $x = y = 0$ уравнения Матье неустойчиво и почти все его решения со временем неограниченно возрастают. Имеет место так называемое параметрическое возбуждение колебаний осциллятора, при котором амплитуда колебаний неограниченно растет.

При $|A| < 1$ ситуация совсем другая. В этом случае λ_1 и λ_2 , как и переменные u и v , – комплексные сопряженные. Перейти к действительным переменным можно с помощью замены

$$\xi = \frac{u + v}{2}, \quad \eta = \frac{u - v}{2i}.$$

В новых переменных ξ и η преобразование (21.10) и исходное для него (21.7) (если учесть, что $|\lambda_1| = |\lambda_2| = 1$, и поэтому при некотором φ $\lambda_1 = \cos \varphi - i \sin \varphi$, а $\lambda_2 = \cos \varphi + i \sin \varphi$) принимают вид:

$$\bar{\xi} = \xi \cos \varphi - \eta \sin \varphi, \quad \bar{\eta} = \xi \sin \varphi + \eta \cos \varphi,$$

т.е. в этих переменных являются вращением вокруг точки $x = y = 0$ на угол φ . Это означает, что при неограниченном возрастании времени любое решение $x(t)$ $v(t) = \dot{x}(t)$ уравнения Матье остается ограниченным, а его равновесие $x = y = 0$ является устойчивым.

Итак, при $|A| > 1$ имеет место неустойчивость с экспоненциальным возрастанием, а при $|A| < 1$ – устойчивость, не сопровождающаяся приближением к равновесию. Граничный случай $|A| = 1$ позволим себе оставить без внимания.

Значение величины A определяется параметрами ω^2 и ε уравнения Матье. Эту зависимость не удастся найти аналитически, но для любых ω^2 и ε можно вычислить соответствующее значение A . Это вычисление легко выполняется с помощью ЭВМ. Представим себе, что мы ею воспользовались и нашли области на плоскости параметров ω^2 и ε , где $|A| > 1$ и где $|A| < 1$. При этом мы придем к картинке, показанной на рис. 21.2.

На этой картинке есть детали, которые не могут быть выяснены вычислительно – это поведение граничных кривых вблизи оси ε . Доходят ли они до нее или оканчиваются очень близко к ней и как именно ведут себя около нее? Но именно этот вопрос может быть исследован аналитически, поскольку при этом можно ограничиться малыми ε . Это аналитическое исследование было проделано. Приведенный рис. 21.2 соответствует его результату: «носы» областей неустойчивости упираются своими заострениями в ось ω в точках, для которых $\omega^2 = n^2/4$ ($n = 1, 2, \dots$).

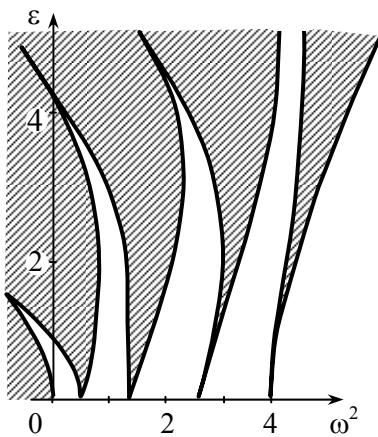


Рис. 21.2. Области по параметрам ω^2 и ε параметрического резонанса (заштрихованы) и устойчивости параметрически возбуждаемого осциллятора, описываемого уравнением Матье (21.4)

Что же из этого следует в отношении движений осциллятора, описываемого уравнением Матье? Отсюда следует не только удивительное явление параметрического возбуждения под влиянием гармонически меняющегося параметра, но и возможность этого возбуждения при сколь угодно малых амплитудах изменения параметра, если только частота исходного приведенного невозмущаемого осциллятора ω достаточно близка к одному из значений $n/2$ (это, конечно, при $\omega^2 > 0$).

Параметрическое возбуждение удивительно потому, что оно может происходить и тогда, когда при каждом фиксированном значении меняющегося параметра осциллятор устойчив. Это еще более удивительно потому, что может происходить

при сколь угодно малых изменениях параметра, однако только при вполне определенных условиях на частоту осциллятора, точнее, на соотношение частоты ω исходного осциллятора (2.9) и частоты ν изменения параметра ε . Это последнее явление такого же типа, как явление резонанса при вынужденных колебаниях осциллятора, и потому получило наименование параметрического резонанса; но если обычный резонанс (при $\delta = 0$) происходит при равенстве частот осциллятора и внешнего воздействия, то параметрический резонанс имеет место при счетной серии соотношений частот, когда

$$\omega = \frac{n}{2} \quad (n = 1, 2, \dots). \quad (21.11)$$

Наиболее ярко проявляется параметрический резонанс, отвечающий $n = 1$, когда отношение частот осциллятора и параметрического воздействия ω и ν равно $1/2$. Именно благодаря этому резонансу мы раскачиваемся на качелях, приседая и вставая в два раза чаще, чем качания качелей. Приседания и вставания приводят к изменению параметра длины маятника, что также с некоторым приближением описывается уравнением Матье. «Понять» этот основной параметрический резонанс можно с энергетической точки зрения, так как периодическое изменение жесткости или изменение длины маятника требует затраты работы, которая переходит в колебательную энергию качания маятника. Для подкачки энергии нужно вставать, когда вас сильнее давит вниз, и приседать, когда это давление меньше: внизу вставать, в верхних положениях приседать. Ясно, что при этом мы совершаем работу и, значит, раскачиваем качели.

Параметрическое возбуждение в основном резонансе можно трактовать, как преобразование колебаний частоты ν в колебания с частотой $\nu/2$. При обычном резонансе деления частоты не происходит: вынужденные колебания происходят с той же частотой, что и внешнее воздействие. Это преобразование частоты широко используется в технике, в так называемых параметрических преобразованиях частоты. В ряде случаев параметрический резонанс, как и обычный, приводит к крайне нежелательным и даже опасным колебаниям, например, при возбуждении колебаний спаренных колес электровоза, в силу того, что жесткость их соединения зависит от угла поворота колес.

В трактовке результатов исследования уравнения Матье, предполагалось, что отсутствует диссипация и $\delta = 0$. При обычном резонансе с появлением диссипации амплитуда колебаний становится конечной вне зависимости от величины воздействия. Для параметрического резонанса это не так. Амплитуда колебаний и при наличии диссипации может расти неограниченно. Нужно только, чтобы амплитуда изменения ε превосходила некоторый предел. То есть граница неустойчивости располагается несколько выше указанной на рис. 21.2 линии. Параметрическое возбуждение имеет место выше этой линии. Естественно, что эта линия с ростом δ смещается все выше и выше.

Внимательно посмотрим на рис. 21.2. На нем области устойчивости и неустойчивости изображены как при $\omega^2 > 0$, так и при $\omega^2 < 0$. Штриховка относится к областям неустойчивости. Штриховка при $\omega^2 > 0$ означает параметрическое возбуждение. При $\omega^2 < 0$ осциллятор неустойчив и при $\varepsilon = 0$, и поэтому при $\omega^2 < 0$ область неустойчивости примыкает к оси абсцисс. При $\omega^2 > 0$ «нос» областей

неустойчивости указывают на явление параметрического резонанса, а наличие областей устойчивости при $\omega^2 < 0$, напротив, – на явление параметрической стабилизации статически неустойчивого осциллятора – перевернутого маятника.

На этом рассказ об осцилляторе с параметрическим гармоническим воздействием можно было бы закончить. Однако отчего же при обычном резонансе даже малая диссипация делает амплитуду вынужденных колебаний конечной, а при параметрическом резонансе она продолжает неограниченно расти?

Рассмотрим процесс нарастания амплитуды при обычном резонансе в отсутствие диссипации. Ему соответствует решение дифференциального уравнения

$$\ddot{x} + \omega^2 x = a \cos \omega t \quad (21.12)$$

при нулевых начальных условиях $x = \dot{x} = 0$ при $t = 0$. Это решение легко находится и имеет вид:

$$x = At \sin \omega t, \quad (21.13)$$

где после прямой подстановки выясняется, что $A = a/2\omega$.

Таким образом, рост амплитуды при обычном резонансе пропорционален времени. Убывание же амплитуды, вызываемое $\delta > 0$, экспоненциальное, так что рост амплитуды обязательно приостановится. При параметрическом резонансе все иначе. Нарастание амплитуды при неустойчивости ($A > 1$) имеет экспоненциальный характер, и показатель этой экспоненты может быть больше по величине показателя – δ экспоненциального затухания. Следовательно, обычный резонанс с помощью диссипации можно подавить лишь частично, уменьшив амплитуду. Для параметрического резонанса это подавление полное, поскольку превращает неустойчивое равновесие в устойчивое. Однако если это подавление недостаточное, то амплитуда продолжает экспоненциально расти.

В заключение я, как обещал, продемонстрирую вам параметрическое возбуждение, но только для главного резонанса, когда $\nu = 2\omega$, с помощью той же знакомой вам сумочки на ремешке. Как видите, я ее поднимаю и опускаю в два раза чаще, чем она может качаться, и она начинает раскачиваться. Затем это раскачивание нарастает, но, конечно, не до бесконечности, так как с ростом ее колебаний становится неприемлемой наша модель.

Более сложные нелинейные модели вынужденных и параметрических колебаний рассмотрены сотрудниками кафедры теории управления и динамики машин в работах [7–10] и др.

22. НОРМАЛЬНЫЕ КОЛЕБАНИЯ И БИЕНИЯ

Ранее мы познакомились с замечательной математической моделью – линейным осциллятором. Его обобщением являются системы взаимодействующих, связанных осцилляторов и линейные динамические системы. Последующий рассказ ознакомит вас с некоторыми новыми явлениями и фактами, относящимися к этим обобщениям линейного осциллятора. Изложение не будет носить общего характера, оно ограничится простым примером, осмыслив который, вы все же получите некоторое представление о том, что нового может быть в многомерных линейных динамических системах.

Пример: две массы, перемещения которых вдоль прямой сдерживаются тремя пружинами (рис. 22.1).

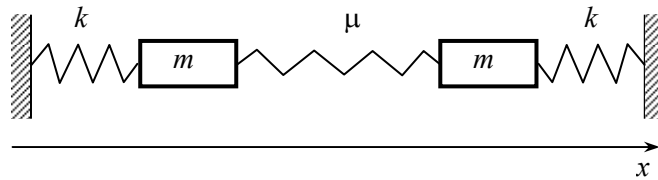


Рис. 22.1. Два гармонических осциллятора, взаимодействующих друг с другом посредством упругой связи жесткости μ

Массы одинаковые и равные m , крайние пружины равных жесткостей k . Между массами – пружина жесткости μ . Если x_1 и x_2 – смещения масс от их равновесных положений, то дифференциальные уравнения Ньютона непосредственно записываются в виде:

$$m\ddot{x}_1 = -kx_1 - \mu(x_1 - x_2),$$

$$m\ddot{x}_2 = -kx_2 + \mu(x_1 - x_2)$$

или

$$m\ddot{x}_1 + (k + \mu)x_1 = \mu x_2, \tag{22.1}$$

$$m\ddot{x}_2 + (k + \mu)x_2 = \mu x_1.$$

Если в правых частях этих уравнений положить $\mu = 0$, то мы придем к двум, никак не связанным между собой линейным осцилляторам. Несущественным, но упрощающим последующие выкладки обстоятельством является то, что осцилляторы, получаемые фиксированием одного из них (x_1 или x_2 равно нулю), имеют одинаковые частоты $\omega = \sqrt{(k + \mu)/m}$. Это естественно, и в соответствии с (22.1) наталкивает на мысль рассматривать систему (см. рис. 22.1) как два взаимодействующих осциллятора: первый действует на второй с силой μx_1 , а второй на первый – с силой μx_2 . Каков же будет результат этого взаимодействия? Для того чтобы ответить на этот вопрос, необходимо найти решение системы (22.1). Это делается так же, как и для одного линейного осциллятора, путем поиска решений вида

$$x_1 = a_1 e^{\lambda t}, \quad x_2 = a_2 e^{\lambda t}, \tag{22.2}$$

когда обе переменные x_1 и x_2 меняются пропорционально $e^{\lambda t}$. Такое решение, точнее, отвечающие ему действительные решения, называются нормальными колебаниями, т.е. колебаниями, когда все части системы колеблются с одной и той же частотой, но, возможно, с разными амплитудами. После подстановки (22.2) в (22.1) приходим к двум уравнениям:

$$(m\lambda^2 + k + \mu)a_1 - \mu a_2 = 0, \quad \mu a_1 + (m\lambda^2 + k + \mu)a_2 = 0. \quad (22.3)$$

Это два уравнения относительно трех неизвестных a_1 , a_2 и λ . Ясно, что найти их однозначно не удастся. Это и сразу было понятно, потому что наряду с (22.2) решением является и любое другое, получаемое умножением (22.2) на один и тот же множитель. Исключая переменные a_1 и a_2 или пользуясь общей теоремой линейной алгебры о наличии у линейной однородной системы ненулевого решения, приходим к характеристическому уравнению для возможных значений λ :

$$\chi(\lambda) = \begin{vmatrix} m\lambda^2 + k + \mu & -\mu \\ -\mu & m\lambda^2 + k + \mu \end{vmatrix} = 0. \quad (22.4)$$

Это уравнение легко решается (так и был подобран пример) и оказывается, что

$$\lambda_{1,2} = \pm i \sqrt{\frac{k}{m}} = \pm i \omega_1, \quad \lambda_{3,4} = \pm i \sqrt{\frac{k+2\mu}{m}} = \pm i \omega_2. \quad (22.5)$$

Теперь, подставляя последовательно в (22.3) полученные значения λ (22.5), находим постоянные a_1 и a_2 с точностью до множителя. При этом следует подставлять найденные значения λ только в одно из уравнений (22.3), так как они линейно зависимы.

При $\lambda = \lambda_{1,2}$ имеем

$$\mu a_1 - \mu a_2 = 0$$

или

$$a_2 = a_1.$$

При $\lambda = \lambda_{3,4}$

$$-\mu a_1 - \mu a_2 = 0$$

или

$$a_2 = -a_1.$$

Таким образом, мы нашли частные решения вида

$$x_1 = a e^{i\omega_1 t}, \quad x_2 = a e^{i\omega_1 t}$$

и вида

$$x_1 = a e^{i\omega_2 t}, \quad x_2 = -a e^{i\omega_2 t}.$$

Им отвечают следующие действительные решения:

$$x_1 = a \frac{\cos}{\sin} \omega_1 t, \quad x_2 = a \frac{\cos}{\sin} \omega_1 t \quad (22.6)$$

и

$$x_1 = a \frac{\cos}{\sin} \omega_2 t, \quad x_2 = -a \frac{\cos}{\sin} \omega_2 t. \quad (22.7)$$

Эти решения, при которых обе массы колеблются с одинаковой частотой, называются нормальными колебаниями. Их частоты разные, ω_1 и соответственно ω_2 , причем $\omega_2 > \omega_1$. Для первого нормального колебания амплитуды колебаний обеих масс одинаковы и они колеблются в фазе. Во втором – амплитуды тоже одинаковы, но колебания происходили в противофазе. Это означает, что при первом нормальном колебании обе массы колеблются вместе вправо и влево. При втором, напротив, навстречу друг другу с большей частотой. Любое другое движение рассматриваемой системы двух масс или связанных осцилляторов – это суперпозиция нормальных колебаний (22.6) и (22.7). Складывая колебания (22.6) и (22.7), можно получить много различных графиков, но, пожалуй, самый интересный возникает при ω_1 , близком к ω_2 , так как при этом наблюдается неожиданное явление последовательной перекачки энергии от одного осциллятора к другому, когда графики изменений x_1 и x_2 имеют вид, показанный на рис. 22.2.

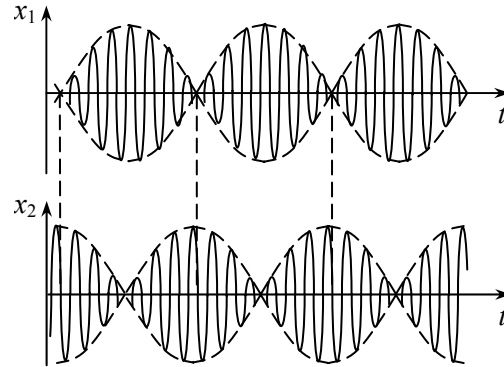


Рис. 22.2. Явление биений в колебаниях двух связанных осцилляторов

Частоты ω_1 и ω_2 близки друг к другу, согласно (22.5), когда μ мало, что и будем предполагать. При этом, пользуясь комплексной записью колебаний, имеем

$$x_1 = C_1 e^{i\omega_1 t} + C_2 e^{i\omega_2 t},$$

$$x_2 = C_1 e^{i\omega_1 t} - C_2 e^{i\omega_2 t}$$

или, несколько преобразовывая эту запись в предположении, что $|C_1| \geq |C_2|$,

$$x_1 = C_1 e^{i\omega_1 t} \left(1 + \frac{C_2}{C_1} e^{i(\omega_2 - \omega_1)t} \right), \tag{22.8}$$

$$x_2 = C_1 e^{i\omega_1 t} \left(1 - \frac{C_2}{C_1} e^{i(\omega_2 - \omega_1)t} \right).$$

Каждое из этих колебаний в силу записи (22.8) и условия $\omega_1 \approx \omega_2$ можно трактовать, как колебание с частотой ω_1 и медленно меняющимися амплитудами, равными величинам

$$\left| 1 \pm \frac{C_2}{C_1} e^{i(\omega_2 - \omega_1)t} \right|,$$

находящимся в пределах от $1 - |C_2/C_1|$ до $1 + |C_2/C_1|$.

Частота изменения амплитуды мала и равна $\omega_2 - \omega_1$. Мы предположили, что $|C_1| \geq |C_2|$; при $|C_1| \leq |C_2|$ аналогично пришли бы к колебаниям частоты ω_2

с медленно меняющейся амплитудой в пределах от $1 - |C_2/C_1|$ до $1 + |C_2/C_1|$. Это подтверждает график на рис. 22.2. При этом полнота перекачки энергии от одного осциллятора к другому зависит от отношения $|C_2/C_1|$. При $|C_1| = |C_2|$ перекачка полная, при $C_2 = 0$ или $C_1 = 0$ она отсутствует, но это имеет место только при нормальном колебании с частотой ω_1 или ω_2 , т.е. в тех случаях, когда массы колеблются вместе либо навстречу друг другу.

Этот простой пример следует рассматривать как введение в разнообразный, подчиненный принципу суперпозиции мир колебаний связанных осцилляторов и линейных систем, в частности, образующих цепочки, кристаллические решетки или отдельные молекулы, с их затейливыми спектрами нормальных колебаний и многообразием их форм и суперпозиций. Качественно новыми, пожалуй, будут нормальные колебания в виде стоячих и бегущих волн, ярко проявляющихся при большом числе осцилляторов. Эти волновые явления перебрасывают мостик от скромного линейного осциллятора к колебаниям распределенных сред, соответствующих неограниченно большому их числу.

Перейдем к описанию наглядной демонстрации обнаруженного нами явления биений или перекачки энергии от одного осциллятора к другому и обратно. И это несмотря на малость связей осцилляторов между собой, когда жесткость μ мала, но при близости частот каждого из осцилляторов при $\mu = 0$. То есть перекачка энергии от одного осциллятора к другому и обратно происходит при близости их частот в условиях резонанса. Становится понятным, почему малая связь влечет за собой значительные последствия – полную перекачку энергии с одного осциллятора на другой.

Эту перекачку можно наглядно показать с помощью двух одинаковых маятников с общей осью, соединенных между собой слабой пружинкой или резинкой (рис. 22.3). Если качнуть один из маятников, то он начнет раскачивать другой до тех пор, пока сам не остановится, после чего уже его начнет раскачивать раскачавшийся маятник и т.д. Эта взаимная раскачка обусловлена малым силовым, но резонансным их взаимодействием.

Аналогичное явление резонансной перекачки энергии можно наблюдать, когда одна из связей между осцилляторами силовая, а другая – параметрическая. Представьте себе грузик, висящий на длинной пружинке. Он может качаться вверх и вниз, а может еще колебаться, как маятник (рис. 22.4).

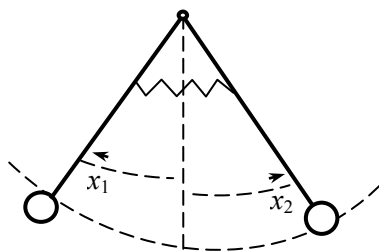


Рис. 22.3. Упруго связанные осцилляторы в виде маятников

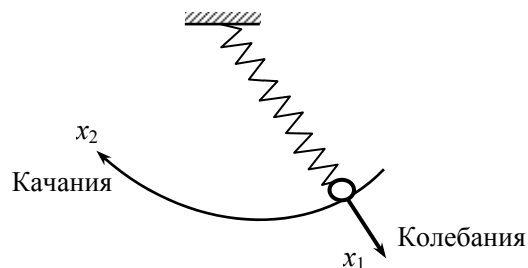


Рис. 22.4. Маятник с упругим подвесом как два связанных осциллятора: от x_1 к x_2 связь параметрическая, от x_2 к x_1 – силовая

Частоты этих колебаний разные, и пусть все подобрано так, что частота колебаний в два раза больше частоты качаний. Тогда при колебаниях грузика периодически меняется длина маятника и происходит параметрическое возбуждение его качаний за счет, естественно, его колебаний. Далее, качания маятника вызывают колебания грузика, так как сила инерции имеет частоту, удвоенную по отношению к качаниям маятника. Все это приводит к очень интересной и красивой последовательности переходов от колебаний к качаниям и обратно – от качаний к колебаниям.

В заключение мне хотелось бы подчеркнуть, что формальная математическая сторона нашей сегодняшней лекционной беседы в основном вам хорошо известна из курса дифференциальных уравнений: вас там научили, как решаются системы линейных дифференциальных уравнений с постоянными коэффициентами. В этом плане ничего нового вы не услышали. Но мы придали некоторую образность тому, что делали, введя понятие нормального колебания, его частот и форм и представляя любое движение как суперпозицию нормальных колебаний. Чтобы ощутить полезность этих общих наглядных представлений, попробуйте выяснить, каковы возможные колебания трех равных масс, насаженных без трения на круговое кольцо при условии, что сила притяжения каждой пары масс друг к другу за счет сближающих их пружин пропорциональна углу между ними. По общим правилам эту задачу можно решать долго и нудно, а с использованием образного представления нормального колебания решение потребует не более строчки.

23. СТАБИЛИЗАЦИЯ ПЕРЕВЕРНУТОГО МАЯТНИКА

Я хочу рассказать вам об универсальном методе, который позволяет кардинально менять свойства разнообразных объектов, на простом примере превращения неустойчивой, всегда падающей, вертикально стоящей палочки в устойчивую, саму по себе стоящую. Этот универсальный метод, широко используемый живой природой и человеком, называется управлением. Вертикально поставленная палочка стоять не может. Она обязательно упадет. Заставить ее не падать могут в цирке. Это делают жонглеры едва заметными перемещениями ее точки опоры. Они меняют положение ее точки опоры так, что палочка или другой предмет стоит вертикально и не падает. Как они это делают? Попробуем это понять, построив математическую модель стабилизации вертикального маятника.

Итак, перед нами плоский маятник с перемещаемой точкой опоры (рис. 23.1). Угол отклонения маятника от вертикали пусть φ , горизонтальное смещение точки опоры в плоскости качания маятника u , длина маятника l , масса m . Найдем функцию Лагранжа и составим с ее помощью уравнения движения.

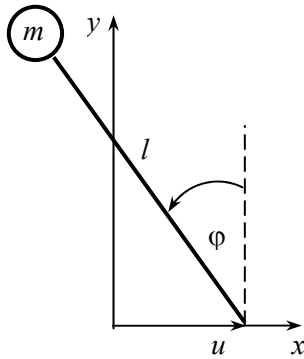


Рис. 23.1. Перевернутый маятник с углом φ отклонения от верхнего неустойчивого положения

Непосредственно находим координаты x, y массы m маятника

$$x = u - l \sin \varphi, \quad y = l \cos \varphi.$$

Далее находим кинетическую и потенциальную энергии

$$T = \frac{m}{2} [(\dot{u} - l\dot{\varphi} \cos \varphi)^2 + l^2 \dot{\varphi}^2 \sin^2 \varphi] = \frac{m}{2} (\dot{u}^2 - 2l\dot{u}\dot{\varphi} \cos \varphi + l^2 \dot{\varphi}^2),$$

$$V = mgy = mgl \cos \varphi$$

и функцию Лагранжа

$$L = T - V = \frac{m}{2} (\dot{u}^2 - 2l\dot{u}\dot{\varphi} \cos \varphi + l^2 \dot{\varphi}^2) - mgl \cos \varphi. \quad (23.1)$$

Составляем уравнение Лагранжа

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\varphi}} - \frac{\partial L}{\partial \varphi} = 0. \quad (23.2)$$

В рассматриваемом случае

$$\frac{\partial L}{\partial \dot{\varphi}} = -mli \cos \varphi + ml^2 \dot{\varphi}, \quad \frac{\partial L}{\partial \varphi} = mli \dot{\varphi} \sin \varphi + mgl \sin \varphi,$$

и поэтому уравнение (23.2) принимает вид:

$$-m\ddot{u} \cos \varphi + mli \dot{\varphi} \sin \varphi + ml^2 \ddot{\varphi} - mli \dot{\varphi} \sin \varphi - mgl \sin \varphi = 0$$

или

$$\ddot{\varphi} - \frac{g}{l} \sin \varphi = \frac{\ddot{u}}{l} \cos \varphi.$$

Ограничимся малыми углами φ и упростим это уравнение, записав его в виде:

$$\ddot{\varphi} - \frac{g}{l} \varphi = \frac{1}{l} \ddot{u}. \quad (23.3)$$

При $u = 0$, т.е. при неподвижной точке опоры маятника, уравнение (23.3) переходит в хорошо нам известное уравнение маятника, линеаризованное вблизи верхнего неустойчивого положения равновесия. Это неустойчивое равновесие типа седла. А мы хотим, чтобы оно стало устойчивым равновесием типа узла или фокуса. Как этого достигнуть? Воспользуемся возможностью выбора смещения u точки опоры, возможностью управлять ее положением. Как это нужно делать, мы не знаем, но хотим, чтобы уравнение (23.3) превратилось в уравнение

$$\ddot{\varphi} + 2\delta \dot{\varphi} + \omega^2 \varphi = 0 \quad (23.4)$$

осциллятора с устойчивым положением равновесия $\varphi = 0$, т.е. чтобы маятник, стоя кверху, вел себя так же, как если бы он висел вниз и колебания его затухали ($\delta > 0$).

Сравнивая (23.3) и (23.4), находим, что для этого нужно, чтобы

$$\ddot{u} = -(g + l\omega^2)\varphi - 2l\delta \dot{\varphi}. \quad (23.5)$$

Реализовать такое u можно, наблюдая за маятником и измеряя его отклонение φ и скорость отклонения $\dot{\varphi}$ и сообщая точке опоры ускорение \ddot{u} согласно формуле (23.5).

Способ управления, задаваемый формулой (23.5), называют стратегией управления. Как уже говорилось, для ее реализации нужно измерять φ и $\dot{\varphi}$, и, естественно, прикладывать к точке опоры силу, которая бы сообщала точке опоры требуемое ускорение \ddot{u} . Таким образом, для реализации управления нужна измерительная и исполнительная системы. В цирке обе эти функции с успехом выполняет жонглер.

Описанный процесс управления можно изобразить схемой, показанной на рис. 23.2. Схема включает в себя объект управления – перевернутый маятник, на который можно воздействовать, перемещая его опору; систему измерения угла φ и угловой скорости $\dot{\varphi}$; систему управления, реализующую стратегию управления и силовое воздействие, передвигающее опору перевернутого маятника и сообщаящее ей требуемое стратегией управления ускорение.

Что же собою представляет стратегия управления (23.5)? Она линейна по φ и $\dot{\varphi}$, так что $\ddot{u} = -a\varphi - b\dot{\varphi}$, причем требуется для устойчивости, чтобы $a > g$, а

$b > 0$. При $b = 0$ перевернутый маятник также устойчив, но его колебания возле равновесия не затухают, так как они описываются осциллятором (23.4) с $\delta = 0$.

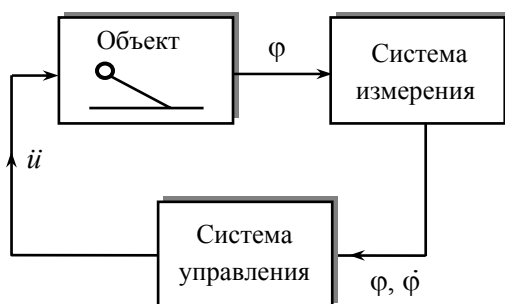


Рис. 23.2. Принципиальная схема стабилизации стоящего вертикально вверх маятника, отображающая основной принцип управления – обратную связь

Первый член стратегии управления $a\varphi$ – это управление по отклонению маятника φ . Согласно этому члену, основание маятника нужно перемещать ускоренно в ту же сторону, что и наклон маятника, т.е. при $\varphi > 0$ $\ddot{i} < 0$, а при $\varphi < 0$ $\ddot{i} > 0$.

Следующий член корректирует величину ускорения \ddot{i} , учитывая угловую скорость вращения маятника. Если маятник движется в ту же сторону, что и его отклонение, то это приводит к увеличению необходимого ускорения \ddot{i} по сравнению с членом $a\varphi$, напротив, если он движется в сторону, обратную по отношению к отклонению, то это ускорение уменьшается по величине, становясь меньше $a\varphi$.

Перейдем теперь от жонглирования палочкой к управлению лодкой (судном или кораблем). Целью управления является поддержание определенного курса ее движения, а средством воздействия – руль. Если рулем не управлять, то лодка может двигаться в любую сторону или вообще все время будет менять курс.

Построим математическую модель движущейся со скоростью v лодки и модель ее управления, т.е. модель движения управляемой лодки.

Пусть лодка движется с постоянной скоростью v , так что сила сопротивления ее движению компенсируется силой тяги весел или гребного винта. Если набегающий на лодку поток направлен точно вдоль ее корпуса, то никакой момент сил не возникает. Момент сил (пусть относительно ее центра масс) возникает только тогда, когда набегающий поток отклоняется от продольной оси лодки, от направления ее киля. В грубом приближении угол расхождения скорости движения центра масс лодки и направления ее продольной оси определяется скоростью ее вращения $\dot{\varphi}$ и вызывает момент, равный $-h\dot{\varphi}$, где в зависимости от конструкции лодки коэффициент h может быть положительным или отрицательным. При этом изменение угла φ поворота лодки относительно ее центра масс, без учета воздействия руля, описывается уравнением вида:

$$J\ddot{\varphi} = -h\dot{\varphi}, \quad (23.6)$$

где J – момент инерции лодки. Как видим, лодка, как объект управления, представляет собою осциллятор с нулевой жесткостью.

Интегрируя уравнение (23.6), находим, что

$$J\dot{\varphi} + h\varphi = \text{const}. \quad (23.7)$$

Наличие первого интеграла (23.7) позволяет непосредственно указать фазовые портреты лодки, движущейся без руля, изображенные на рис. 23.3, а, б. Рисунок 23.3, а относится к $h > 0$, а рис. 23.3, б – к $h < 0$. В обоих случаях лодка имеет

множество равновесий, отвечающих фазовым точкам оси $\dot{\phi} = 0$, при которых лодка движется с постоянной скоростью под некоторым постоянным углом ϕ .

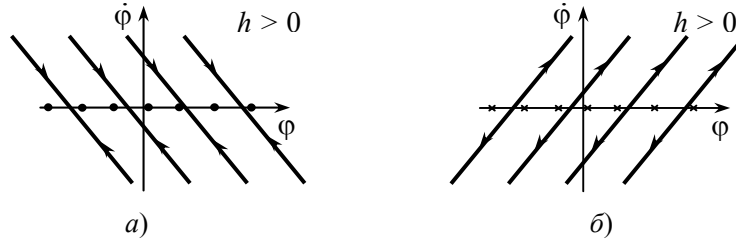


Рис. 23.3. Фазовые портреты неуправляемого объекта управления (корабля или лодки): а) устойчивого и б) неустойчивого

При $h > 0$ эти равновесия устойчивые и любое движение лодки переходит в движение в каком-то постоянном направлении. Напротив, при $h < 0$ состояния равновесия неустойчивы и лодка стремится быстрее и быстрее вращаться (рис. 23.3,б). Так как наше линеаризованное упрощенное уравнение (23.6) справедливо только для не очень больших $\dot{\phi}$, то сказанное нужно понимать, как то, что лодка не может придерживаться определенного курса и обязательно будет вращаться.

При $h > 0$ состояния равновесия устойчивые, но эта устойчивость не асимптотическая и малейшее возмущение движения лодки вообще приводит к неустраняемому изменению направления движения лодки.

Примем, что $\phi = 0$ отвечает требуемому курсу движения лодки. Из предыдущего следует, что ни при $h > 0$, ни тем более при $h < 0$, лодка не может устойчиво придерживаться заданного курса $\phi = 0$.

Для того чтобы лодка его придерживалась, надо ею управлять, например, с помощью руля. Пусть ψ – угол отклонения руля (рис. 23.4).

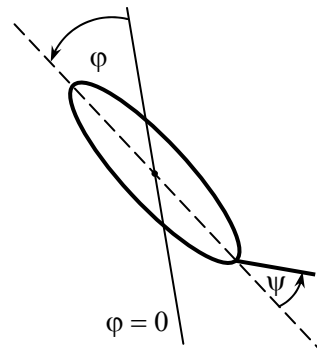


Рис. 23.4. Лодка (корабль), уклонившаяся на угол ϕ , с рулем, повернутым на угол ψ

При $\psi \neq 0$ давление текущей воды на руль создает вращающий момент и уравнение (23.6) заменится уравнением

$$J\ddot{\phi} + h\dot{\phi} = -k\psi. \tag{23.8}$$

Такое уравнение приемлемо, конечно, только при не очень больших ψ . Уравнение (23.8) предполагает $\dot{\phi}$ и ψ малыми, а ϕ может быть любым. Кроме того, следует иметь в виду, что теперь коэффициент h относится не к лодке без руля, а к лодке с закрепленным рулем ($\psi = 0$). Отличие может быть значительным при

большой площади руля и приводит к увеличению параметра h . При отрицательном, но небольшом $|h|$ это может сделать его положительным, что кардинально изменит управляемость лодки. Для придания лодке нового свойства асимптотической устойчивости движения по требуемому курсу $\varphi = 0$ используем ту же стратегию, что и для жонглирования палочкой, а именно, примем, что

$$\psi = a\varphi + b\dot{\varphi}. \quad (23.9)$$

Подставляя (23.9) в (23.8), находим, что

$$J\ddot{\varphi} + (h + b)\dot{\varphi} + ka\varphi = 0. \quad (23.10)$$

Состояние равновесия $\varphi = 0$ этого дифференциального уравнения – линейно-го осциллятора – будет устойчивым, если все коэффициенты его характеристического уравнения

$$J\lambda^2 + (h + b)\lambda + ka = 0 \quad (23.11)$$

положительные, т.е. при выполнении условий $h + b > 0$, $ka > 0$, которые нетрудно выполнить, выбирая соответствующие a и b ,

$$a > 0, \quad b > -h. \quad (23.12)$$

Итак, лодка, управляемая стратегией (23.9) при условии (23.12), асимптотически устойчива на курсе. Смысл стратегии при условии (23.12) прост. Он такой же, как и в случае жонглирования палочкой, но если ранее требовалось достаточно большое значение параметра a , то теперь при $h < 0$ это требование налагается на параметр b . Из этого следует, что лодкой с $h > 0$ управлять сравнительно просто. Напротив, при $h < 0$ это становится достаточно сложным, ибо требуется достаточно большое и правильное реагирование на скорость вращения лодки.

Для этой правильной реакции есть даже свое название, нужно вовремя энергично «одерживать» уходящую с курса лодку. Если лодка сама по себе устойчивая ($h > 0$), то управление ею выглядит так: лодка влево – руль вправо, лодка вправо – руль влево. При $h < 0$ стратегия управления значительно сложнее, главным в ней является идея «одерживания», «одерживания» даже при $\varphi = 0$, если $\dot{\varphi} \neq 0$.

Вернемся к жонглированию палочкой. Но прежде чем продолжить исследование, отметим, что цирковое жонглирование имеет много общего со стабилизацией вертикального взлета и разгона ракеты. Как и палочка, ракета неустойчиво торчит вверх и подпирается снизу реактивными двигателями. Конечно, для нее нужны более точные и полные модели, но в принципе, все так же, как и для палочки. А более точные модели требуются хотя бы потому, что нужно учесть упругость корпуса ракеты, влияние ее колебаний на показания измерительных приборов и процесс стабилизации, запаздывания реализации наблюдения и управления и, возможно, многое другое.

Все же снова о стабилизации перевернутого маятника. Стратегия (23.5) поддерживает вертикальность перевернутого маятника, но при этом его точка опоры может значительно смещаться, не имея тенденции к возврату. Спрашивается, а можно ли стабилизировать верхнее положение маятника, сделав одновременно асимптотически устойчивым и исходное положение точки опоры?

Попробуем применить к этой более требовательной задаче стратегию такого же вида, как и ранее, т.е. линейную по фазовым переменным объекта. Итак, мы хотим стабилизировать маятник не только по φ , но и по координате точки опоры u . Теперь фазовыми переменными маятника с перемещающейся точкой опоры будут переменные φ , $\dot{\varphi}$ и u , \dot{u} , для которых примем уравнения

$$\begin{aligned} \ddot{\varphi} - \frac{g}{l}\varphi &= \frac{\ddot{u}}{l}, \\ \ddot{u} &= a\varphi + b\dot{\varphi} + cu + d\dot{u}, \end{aligned} \quad (23.13)$$

где стратегия управления, как и ранее, – линейная функция всех фазовых переменных.

Задача существенно усложнилась. Характеристическое уравнение, от корней которого зависит устойчивость, не второго, а четвертого порядка. Найдем его. Из (23.13) непосредственно находим характеристическое уравнение в виде определителя

$$\begin{vmatrix} \lambda^2 - g/l & -\lambda^2/l \\ a + b\lambda & c + d\lambda - \lambda^2 \end{vmatrix} = 0$$

или, раскрывая определитель:

$$\lambda^4 + \left(-d + \frac{b}{l}\right)\lambda^3 + \left(-c - \frac{g}{l} + \frac{a}{l}\right)\lambda^2 + d\frac{g}{l}\lambda + c\frac{g}{l} = 0. \quad (23.14)$$

Корни этого уравнения за счет выбора параметров a , b , c и d можно сделать любыми. Это следует из того, что свободный член – любой за счет выбора c , коэффициент при λ – за счет выбора d , коэффициент при λ^2 – за счет выбора a и, наконец, коэффициент при λ^3 – за счет выбора b .

Таким образом, подходящая стратегия вида

$$\ddot{u} = a\varphi + b\dot{\varphi} + cu + d\dot{u},$$

при которой все корни уравнения (23.14) лежат слева от мнимой оси, обеспечивает требуемую стабилизацию перевернутого маятника. Подчеркнем, что при $c = d = 0$, т.е. при стратегии управления $\ddot{u} = -a\varphi - b\dot{\varphi}$, характеристическое уравнение имеет два нулевых корня и равновесие $\varphi = 0$, $u = 0$ неустойчиво, что приводит к отклонению u от нуля.

Рассмотрим еще вопрос о так называемой левитации, т.е. бесконтактном устойчивом «подвешивании» тела в некотором месте. Согласно легенде, так висит в какой-то пещере гроб Магомета. Сразу же приходит мысль сделать это с помощью магнитов, т.е. вывесить железный или содержащий железо предмет магнитами. Однако есть общая теорема, которая это запрещает, но ферромагнитный шарик можно устойчиво подвесить с помощью управления, используя такую же стратегию управления, как и ранее. Это очень соблазнительно, так как бесконтактная подвеска позволяет создавать многие новые высокоточные измерительные приборы и другие устройства, избавленные от неточностей и диссипаций, вызываемых трением, не говоря уж об отсутствии подчас нежелательных контактов с другими телами.

В детстве, возможно, вы безрезультатно пытались «вывесить» металлическое перышко (или какой-либо небольшой железный предмет) магнитом: при этом перышко неизменно либо падало, либо прилипало к магниту. Управлять можно положением магнита, но значительно проще заменить магнит электромагнитом и управлять его током I . При этом, как и ранее, нужна измерительная система положения и скорости вывешиваемого тела, а стратегия управления может состоять в изменении тока как линейной функции этих измерений. При подходящем подборе параметров тело «свободно» и асимптотически устойчиво висит. Висящее тело можно потрогать пальцем и почувствовать, что оно упирается, не желая менять своего места. Можно его завертеть и оно будет долго вращаться, так как трение о воздух незначительно. Можно трение уменьшить, откачав воздух, и тогда оно будет вращаться еще дольше.

Мы рассмотрели жонглирование (стабилизацию) одного перевернутого маятника, но нельзя ли таким же путем жонглировать с несколькими поставленными друг на друга маятниками, управляя положением точки опоры только самого нижнего? Кажется, что это нельзя сделать: верхние маятники упадут. Но посмотрим, что дает математическая модель. Ограничимся двумя маятниками с равными длинами и массами, которые без потери общности примем единичными (рис. 23.5).

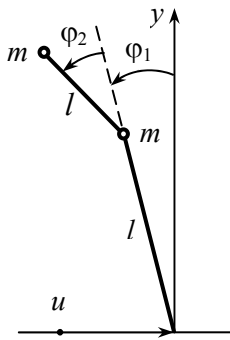


Рис. 23.5. Двойной маятник с отклонениями от вертикали на углы φ_1 и $\varphi_1 + \varphi_2$ соответственно

В обозначениях рис. 23.5, полагая без ограничения общности $m = l = 1$, непосредственно находим, что функция Лагранжа L с учетом последующей линеаризации, т.е. сохранения в дифференциальных уравнениях Лагранжа только членов, линейных относительно $\varphi_1, \varphi_2, \dot{\varphi}_1, \dot{\varphi}_2$, равна

$$L = \frac{1}{2}(2\dot{u}^2 - 4\dot{u}\dot{\varphi}_1 - 2\dot{u}\dot{\varphi}_2 + 2\dot{\varphi}_1^2 + \dot{\varphi}_2^2 + 2\dot{\varphi}_1\dot{\varphi}_2) + g\varphi_1^2 + \frac{g}{2}\varphi_2^2.$$

Отсюда непосредственно находим, что имеют место уравнения:

$$2\ddot{\varphi}_1 + \ddot{\varphi}_2 - 2\ddot{u} - 2g\varphi_1 = 0,$$

$$\ddot{\varphi}_1 + \ddot{\varphi}_2 - \ddot{u} - g\varphi_2 = 0$$

или

$$\begin{aligned} -\ddot{\varphi}_2 - 2g\varphi_1 + 2g\varphi_2 &= 0, \\ \ddot{\varphi}_1 + \ddot{\varphi}_2 - \ddot{u} - g\varphi_2 &= 0. \end{aligned} \tag{23.15}$$

Как и ранее, примем стратегию управления в виде линейной функции всех фазовых переменных, т.е.

$$\ddot{u} = a_1\varphi_1 + b_1\dot{\varphi}_1 + a_2\varphi_2 + b_2\dot{\varphi}_2. \quad (23.16)$$

Составим характеристическое уравнение системы (23.15) с учетом (23.16)

$$\begin{vmatrix} -2g & -\lambda^2 + 2g \\ \lambda^2 - a_1 - b_1\lambda & \lambda^2 - g - a_2 - b_2\lambda \end{vmatrix} = 0$$

или

$$\lambda^4 - b_1\lambda^3 - (a_1 + 4g)\lambda^2 + 2g(b_1 + b_2)\lambda + 2g(g - a_1 + a_2) = 0. \quad (23.17)$$

Из вида характеристического уравнения (23.17) следует, что коэффициенты при λ^3 , λ^2 , λ и λ^0 могут быть любыми при соответствующем выборе параметров стратегии управления b_1 , a_1 , b_2 и a_2 .

Таким образом, стабилизация двух поставленных друг на друга палочек возможна, и для этого достаточно надлежащим образом перемещать нижнюю точку опоры в соответствии со стратегией управления (23.16). Естественно предположить, что аналогично возможно жонглировать и тремя стоящими друг на друге палочками, и, может быть, даже любым их числом, и в соответствии с этим даже заставить кусок веревки из руки стоять вертикально вверх. В юности я видел фильм, где великолепный актер Игорь Ильинский перебрался через высокий забор по вертикально поставленной веревке, которую он принес с собой в мешочке. Это было заведомо не управление, она стояла сама по себе.

Вместе с тем, следует иметь в виду, что эта и предшествующие модели составлены в предположении идеальности как измерительной, так и исполнительской систем управления, т.е. что измерительная система достаточно точно и своевременно находит значения фазовых переменных, а исполнительная система быстро и точно реализует требуемое управление (ускорение точки опоры, угла поворота руля). При этом из рассмотренных нами математических моделей не ясно, в какой мере можно уклониться от этих идеальных требований, как и то, каким образом их можно осуществить.

Едва ли в общем курсе целесообразно сколько-нибудь подробно рассматривать эти сами по себе важные вопросы – это удел специальных курсов. Поэтому ограничимся сказанным и примером учета неизбежных задержек (запаздываний в измерении φ и $\dot{\varphi}$ и реализации управляющего ускорения \ddot{u}) при стабилизации вертикального положения маятника. В отсутствие этих задержек уравнение движения системы управления имело вид (уравнения (23.3) и (23.5)):

$$\ddot{\varphi} - \frac{g}{l}\varphi = -a\varphi - b\dot{\varphi}.$$

Учитывая временную задержку τ измерений φ и $\dot{\varphi}$ и реализации управления, заменим их уравнением

$$\ddot{\varphi}(t) - \frac{g}{l}\varphi(t) = -a\varphi(t - \tau) - b\dot{\varphi}(t - \tau). \quad (23.18)$$

Его характеристическое уравнение, составляемое обычным образом, полагая $\varphi = e^{\lambda t}$, имеет вид:

$$\lambda^2 - \frac{g}{l} + ae^{-\tau\lambda} + b\lambda e^{-\tau\lambda} = 0. \quad (23.19)$$

Для того чтобы система управления, описываемая линейными дифференциальными уравнениями с запаздываниями (23.18), имела устойчивое состояние равновесия $\varphi = 0$, как и в случае обыкновенного дифференциального уравнения, нужно, чтобы корни его характеристического уравнения (23.19) лежали слева от мнимой оси, т.е. удовлетворяли условию $\operatorname{Re} \lambda < 0$. Расположение корней характеристического уравнения (23.19) зависит от параметров стратегии управления a и b . При отсутствии запаздывания, когда $\tau = 0$, устойчивость имеет место при

$$a > g/l, \quad b > 0. \quad (23.20)$$

Как обнаруживает соответствующее исследование, с методикой которого вас познакомят в курсе теории управления, при τ малых, меньших $\sqrt{2l/g}$, область устойчивости по параметрам управления a и b существует и имеет вид, показанный на рис. 23.6. При $\tau \rightarrow 0$ эта область устойчивости стремится к области D_0 , определяемой неравенствами (23.20). При увеличении τ область устойчивости по параметрам a и b на рис. 23.6 уменьшается (D_1) и при $\tau = \tau^* = \sqrt{2l/g}$ исчезает. При $\tau > \tau^*$ ее нет.

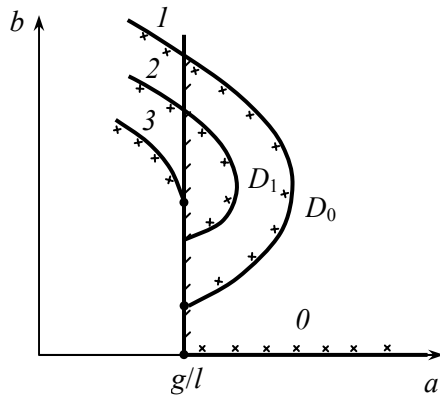


Рис. 23.6. Область стабилизируемости перевернутого маятника и ее зависимость от запаздывания τ реализации линейной стратегии управления. Граничные кривые 0, 1, 2 и 3 соответствуют возрастанию запаздывания. При $\tau = 0$ область устойчивости $a > g/l, b > 0$; при τ , отвечающем граничной кривой 3, область устойчивости исчезает

Из сказанного следует: чтобы жонглирование было успешным, нужно не только правильно выбрать параметры a и b стратегии управления, но и реализовать измерения и управление достаточно быстро. Эта быстрота должна быть тем большей, чем меньше длина палочки. Пытаясь жонглировать очень короткими палочками, вы неизбежно потерпите поражение и по этой наименьшей, доступной вам длине успешного жонглирования, можете судить о быстроте выполнения вами требуемого управления и, конечно, одновременно и об умении подобрать требуемую стратегию управления.

Для палочки (шариковой ручки) длиной 16 см мне это никак не удавалось. Осуществление жонглирования представлялось безнадежным. Она падала раньше, чем я успеваю что-либо сделать для предотвращения ее падения. Критическое значение запаздывания для такой палочки равно примерно $\tau \approx 0,17$ с. Для $\tau^* = 0,3$ с соответствующая длина равна $l = 45$ см. Такой палочкой жонглировать уже удастся. Я имею в виду себя. А как у вас? При желании вы можете сопоставить τ^* , получающиеся для вас из экспериментов жонглирования, с прямым измерением времени вашей реакции, которое вы можете провести с кем-нибудь из своих знакомых, располагая обыкновенной линейкой, причем интересно, что по такой же

формуле $\tau = \sqrt{2l/g}$, где l – минимальная длина падающей между вашими большим и указательным пальцами линейки, которую после отпущения на ваших глазах, вам еще удастся зажать и не дать упасть. У меня вышло примерно 20 см, что отвечает $\tau \approx 0,2$ с.

Мы познакомились на простых примерах с величайшей придумкой природы и человека – управлением, без которого жизнь была бы невозможна. Человек управляет своим поведением, с помощью управления через энзимы он осуществляет биохимические процессы в своем организме; зажигая электрический свет в темной комнате, он нажатием выключателя управляет электрическим током; поворачивая руль, управляет велосипедом, а нажимая на педали, заставляет его ехать; руками и ногами управляет автомобилем; произнося команды, через других людей управляет производственными процессами; посредством нервной системы управляет сокращением и расслаблением мышц и, тем самым, движением своего тела, биением сердца, дыханием, сокращениями и перистальтикой кишечника, поступлением в желудок и двенадцатиперстную кишку необходимых для пищеварения веществ. Посмотрите, как красиво бежит или скачет лошадь, грациозно перемещается и движется кошка – это тоже благодаря управлению. А мимика нашего лица, отражающая внутренние психологические процессы, обогащающая наше общение, да и сама наша речь возможна только благодаря управлению голосовыми связками, горлом и артикуляцией рта.

Человек создал и разработал новую науку – теорию управления, и она подсказала ему, как построить автопилот, авторулевой, как управлять ядерным реактором, полетом самолета, работой станков и механизмов, как создать системы связи и передачи информации и как управлять ракетой и своей собственной жизнью на Земле, что наиболее трудно и еще далеко от совершенства. Да и сама наша жизнь возможна лишь благодаря ферментам, управляющим биохимическими реакциями в нашем организме, и нервной системе, управляющей сердцем и нашими органами, движениями и действиями.

24. УПРАВЛЯЕМЫЙ МАЯТНИК И ДВУНОГАЯ ХОДЬБА

В предыдущих разделах я вам рассказывал об управлении на примерах жонглирования – стабилизации одной и двух вертикально стоящих палочек – и управления курсом лодки. С этим вы встречаетесь, но сравнительно редко. Сейчас я вам расскажу тоже о конкретном феномене управления, но с которым вы в содружестве все время, пользуетесь им каждый день. Хотя жонглированием вы занимаетесь тоже каждый день, поддерживая свое вертикальное положение, но модели этого жонглирования не те, что мы рассмотрели.

Простейший вид жонглирования своим телом – вы стоите, опираясь на обе ноги. Стоите вы вполне уверенно, но все же немного, еле заметно, покачиваясь. Прислушавшись к себе, можно заметить, как меняются усилия в опорных ногах. Это подсказывает следующую модель управления в виде вертикального маятника, в неподвижной точке опоры которого имеется управляющий момент M . В каждой из двух перпендикулярных плоскостей имеют место линеаризованные уравнения вида:

$$\ddot{\varphi} - \frac{g}{l} \varphi = \frac{1}{ml^2} M.$$

Согласно предыдущему ясно, что стратегия, обеспечивающая ваше вертикальное положение, может быть очень простого вида:

$$M = a\varphi + b\dot{\varphi}.$$

Но стоять вертикально вы можете не только опираясь на подошвы обеих ног, но и встав на носки, т.е. опираясь на маленькую площадку, не позволяющую вам создать момент M . Вы это делаете, двигая руками и корпусом, а если стоите на одной ноге, то и другой свободной ногой. В этом случае оказывается, что для стабилизации вертикального положения достаточно одного управления корпусом. В качестве упрощенной модели такого жонглирования примем двойной маятник с закрепленной точкой опоры и управляемым моментом M в их сочленении. Для написания математической модели в этом случае можно воспользоваться полученными ранее формулами для управляемого двойного маятника, положив в них $u = 0$ и добавив к правым частям моменты M и $-M$ соответственно. В упрощенном линеаризованном виде, согласно формулам (23.14), (23.15), имеем:

$$\begin{aligned} 2\ddot{\varphi}_1 + \dot{\varphi}_2 - 2g\varphi_1 &= M, \\ \ddot{\varphi}_1 + \ddot{\varphi}_2 - g\varphi_2 &= -M. \end{aligned}$$

Стратегию управления в соответствии с предыдущим возьмем зависящей от всех фазовых переменных, т.е. от $\varphi_1, \dot{\varphi}_1, \varphi_2, \dot{\varphi}_2$, так что

$$M = a_1\varphi_1 + b_1\dot{\varphi}_1 + a_2\varphi_2 + b_2\dot{\varphi}_2.$$

Это приводит к характеристическому уравнению

$$\begin{vmatrix} 2\lambda^2 - 2g - a_1 - b_1\lambda & \lambda^2 - a_2 - b_2\lambda \\ \lambda^2 + a_1 + b_1\lambda & \lambda^2 - g + a_2 + b_2\lambda \end{vmatrix} = 0$$

или

$$\lambda^4 + (3b_2 - 2b_1)\lambda^3 + (3a_2 - 2a_1 - 4g)\lambda^2 + (b_1 - 2b_2)g\lambda + g(2g + a_1 - 2a_2) = 0.$$

Коэффициенты полученного характеристического уравнения за счет выбора a_1, b_1, a_2, b_2 могут быть любыми, и поэтому при надлежащем их выборе асимптотическая устойчивость равновесия $\varphi_1 = \varphi_2 = 0$ обеспечивается.

Итак, жонглирование двумя стоящими друг на друге перевернутыми маятниками при закрепленной опоре возможно с помощью линейной стратегии управления моментом M в шарнирном соединении маятников. Природа тоже нашла такие стратегии, они вырабатываются у нас после рождения в процессе обучения и, возможно, что-то заложено генетически.

Таким образом, стоять мы умеем. Но мы умеем еще и ходить. Природа ходьбы кажется очень сложной: в ней участвуют многие десятки мышц нашего тела. Вместе с тем феномен ходьбы в своем простейшем виде – это не более чем автоколебания, при которых все периодически повторяется. Добиться генерации колебаний, как это было видно на примере часов Галилея–Гюйгенса и генератора электрических колебаний, рассмотренных ранее, можно с помощью обратной связи, управления, осуществляемого обратной связью. Выше указывались и многочисленные другие примеры применения управления и обратной связи для достижения требуемых свойств различных объектов. Пожалуй, управление и обратная связь по достоинству должны быть произведены в некий принцип обратной связи. Именно так сделал Н. Винер и провозгласил новую, основанную на этом принципе, науку кибернетику.

Но вернемся к двуногой ходьбе. Рассказ об этом феномене начнем издалека. Мы рассмотрели и поняли, как жонглер заставляет не падать вертикально стоящую палочку: он ею управляет. Управляет, перемещая ее точку опоры в зависимости от ее движения. Ясно, что это не единственный способ управления перевернутым маятником. Маятником можно управлять и иначе. Например, подсоединить к концу маятника еще одно звено и поддерживать его направление постоянным. Не задумывайтесь над тем, зачем это нужно, лучше сначала разберемся, что из этого последует.

Итак, перед нами двойной маятник (рис. 24.1). Его звенья имеют длины l_1 и l_2 ; массы, которые ради простоты сосредоточим на концах палочек, m_1 и m_2 ; углы звеньев с вертикалью φ_1 и φ_2 . В сочленении звеньев маятника находится управляющее устройство, которое с помощью момента M между подсоединенным звеном и маятником поддерживает угол φ_2 постоянным, равным α . Момент M нам неизвестен, но известно, что система уравнений справляется со своей задачей, так что угол φ_2 постоянно равен α .

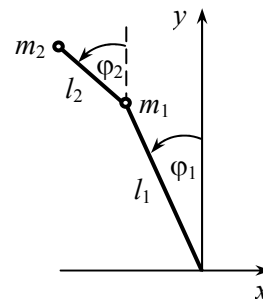


Рис. 24.1. Двойной маятник

Составим математическую модель этого причудливо управляемого перевернутого маятника.

Маятник с подсоединенным звеном – это двойной маятник. С ним вы уже встречались, когда рассматривали жонглирование двумя стоящими одна на другой вертикальными палочками. Сейчас нас интересует другое: как будет двигаться маятник под воздействием описанного управления.

Пусть x_1, y_1 и x_2, y_2 – координаты точечных масс m_1 и m_2 . Непосредственно из рис. 24.1 находим, что

$$\begin{aligned}x_1 &= -l_1 \sin \varphi_1, & y_1 &= l_1 \cos \varphi_1, \\x_2 &= -l_1 \sin \varphi_1 - l_2 \sin \varphi_2, & y_2 &= l_1 \cos \varphi_1 + l_2 \cos \varphi_2.\end{aligned}$$

Далее находим кинетическую энергию T и потенциальную функцию V сил тяжести:

$$\begin{aligned}T &= \frac{1}{2}(m_1 + m_2)l_1^2 \dot{\varphi}_1^2 + m_2 l_1 l_2 \dot{\varphi}_1 \dot{\varphi}_2 \cos(\varphi_1 - \varphi_2) + \frac{1}{2}m_2 l_2^2 \dot{\varphi}_2^2, \\V &= (m_1 + m_2)gl_1 \cos \varphi_1 + m_2 gl_2 \cos \varphi_2.\end{aligned}$$

Обобщенные силы Q_1 и Q_2 , отвечающие моменту M управляющего устройства, равны: $Q_1 = M$, $Q_2 = -M$.

Теперь мы знаем функцию Лагранжа и дополнительные обобщенные силы и можем написать уравнения Лагранжа:

$$\begin{aligned}(m_1 + m_2)l_1^2 \ddot{\varphi}_1 + m_2 l_1 l_2 \ddot{\varphi}_2 \cos(\varphi_1 - \varphi_2) - m_2 l_1 l_2 \dot{\varphi}_2 (\dot{\varphi}_1 - \dot{\varphi}_2) \sin(\varphi_1 - \varphi_2) + \\+ m_2 l_1 l_2 \dot{\varphi}_1 \dot{\varphi}_2 \sin(\varphi_1 - \varphi_2) - (m_1 + m_2)l_1 g \sin \varphi_1 = M, \\m_2 l_2^2 \ddot{\varphi}_2 + m_2 l_1 l_2 \ddot{\varphi}_1 \cos(\varphi_1 - \varphi_2) - m_2 l_1 l_2 (\dot{\varphi}_1 - \dot{\varphi}_2) \dot{\varphi}_1 \sin(\varphi_1 - \varphi_2) - \\- m_2 l_1 l_2 \dot{\varphi}_1 \dot{\varphi}_2 \sin(\varphi_1 - \varphi_2) - m_2 l_2 g \sin \varphi_2 = -M\end{aligned}\tag{24.1}$$

или, учитывая, что $\varphi_2 = \alpha$ ($\dot{\varphi}_2 = \ddot{\varphi}_2 = 0$)

$$\begin{aligned}(m_1 + m_2)l_1^2 \ddot{\varphi}_1 - (m_1 + m_2)l_1 g \sin \varphi_1 = M, \\m_2 l_1 l_2 \dot{\varphi}_1 \cos(\varphi_1 - \alpha) - m_2 l_1 l_2 \dot{\varphi}_1^2 \sin(\varphi_1 - \alpha) - m_2 l_2 g \sin \alpha = -M.\end{aligned}\tag{24.2}$$

Эта система уравнений определяет изменение угла φ_1 маятника с вертикалью и управляющий момент M . Исключая из этих уравнений неизвестный момент M , приходим к уравнению для угла φ_1 . Опустив у угла φ_1 индекс, запишем получившееся дифференциальное уравнение второго порядка в виде:

$$\begin{aligned}[(m_1 + m_2)l_1^2 + m_2 l_1 l_2 \cos(\varphi - \alpha)]\ddot{\varphi} - \\- m_2 l_1 l_2 \dot{\varphi}^2 \sin(\varphi - \alpha) - (m_1 + m_2)gl_1 \sin \varphi - m_2 gl_2 \sin \alpha = 0.\end{aligned}\tag{24.3}$$

Для любого решения уравнения (24.3) каждое из уравнений (24.2) можно использовать для нахождения момента M .

Приступим к изучению решений – фазового портрета – уравнения (24.3). Прежде всего, приведем его к виду

$$[1 + \kappa \cos(\varphi - \alpha)]\ddot{\varphi} - \kappa \dot{\varphi}^2 \sin(\varphi - \alpha) - \omega^2 \sin \varphi - \kappa \omega^2 \sin \alpha = 0,\tag{24.4}$$

содержащему только два параметра κ и ω :

$$\kappa = \frac{m_2}{m_1 + m_2} \frac{l_2}{l_1}, \quad \omega^2 = \frac{g}{l_1}.$$

При $\kappa = 0$ уравнение (24.4) превращается в уравнение маятника, который при $\varphi = 0$ стоит вертикально вверх. Оказывается, что уравнение (24.4) можно проинтегрировать. Для этого умножим его на $[1 + \kappa \cos(\varphi - \alpha)] \dot{\varphi}$:

$$[1 + \kappa \cos(\varphi - \alpha)]^2 \dot{\varphi} \ddot{\varphi} - \kappa [1 + \kappa \cos(\varphi - \alpha)] \dot{\varphi}^3 \sin(\varphi - \alpha) - \omega^2 [1 + \kappa \cos(\varphi - \alpha)] \dot{\varphi} \sin \varphi - \kappa \omega^2 [1 + \kappa \cos(\varphi - \alpha)] \dot{\varphi} \sin \alpha = 0$$

и преобразуем к виду

$$\frac{1}{2} \frac{d}{dt} \{ [1 + \kappa \cos(\varphi - \alpha)]^2 \dot{\varphi}^2 \} - [1 + \kappa \cos(\varphi - \alpha)] \omega^2 (\sin \varphi + \kappa \sin \alpha) \dot{\varphi} = 0.$$

Интегрируя, находим, что

$$[1 + \kappa \cos(\varphi - \alpha)]^2 \dot{\varphi}^2 + F(\varphi) = h, \quad (24.5)$$

где h – постоянная интегрирования и

$$\begin{aligned} F(\varphi) &= -2\omega^2 \int [1 + \kappa \cos(\varphi - \alpha)] (\sin \varphi + \kappa \sin \alpha) d\varphi = \\ &= 2\omega^2 \left[\cos \varphi - \frac{3\kappa}{2} \varphi \sin \alpha + \frac{\kappa}{4} \cos(2\varphi - \alpha) - \kappa^2 \sin \alpha \sin(\varphi - \alpha) \right]. \end{aligned}$$

Разрешая (24.5) относительно $\dot{\varphi}$, имеем

$$\dot{\varphi} = \pm \frac{1}{1 + \kappa \cos(\varphi - \alpha)} \sqrt{h - F(\varphi)}. \quad (24.6)$$

Как вы уже знаете, построение фазового портрета, согласно (24.6), сводится к нахождению графика функции $F(\varphi)$. Однако при $\kappa \geq 1$ возможно неограниченное возрастание угловой скорости при приближении угла φ к значению, при котором

$$1 + \kappa \cos(\varphi - \alpha) = 0. \quad (24.7)$$

Этот неожиданный феномен проистекает оттого, что при реализации требуемого управления управляющий момент M неограниченно растет, так как из уравнений (24.2), исключая $\ddot{\varphi}$, находим, что

$$M = \frac{m_2 l_1 l_2}{1 + \kappa \cos(\varphi - \alpha)} \left[\dot{\varphi}^2 \sin(\varphi - \alpha) + \omega^2 \sin \alpha - \omega^2 \sin \varphi \cos(\varphi - \alpha) \right], \quad (24.8)$$

где φ_1 заменено на φ .

Реализация неограниченного момента невозможна, и поэтому невозможно соблюдение условия $\varphi_2 = \alpha$ и мы вынуждены, если хотим изучить этот случай, внести коррективы в рассматриваемую математическую модель. Эта корректировка может, например, состоять в том, что при достижении $|M|$ некоей предельной величины принимается $M = M^* \text{sign } M$ до тех пор, пока M , определяемое из (24.8), не окажется по модулю меньше M^* . При этом углы φ_1 и φ_2 определяются из уравнений (24.1). Однако ограничимся случаем, когда описанное явление не имеет места. Отметим только, что это явление подобно удару и приводит к очень быстрым изменениям скоростей $\dot{\varphi}_1$ и $\dot{\varphi}_2$.

Как уже отмечалось, для построения фазового портрета достаточно располагать графиком функции $F(\varphi)$, определяемой формулой (24.5). Вид графика функции $F(\varphi)$ зависит от двух параметров κ и α . Достаточно очевиден его вид при малых κ и α . На рис. 24.2 приведен его вид при $\alpha > 0$. Согласно этому графику, построены фазовые траектории, определяемые уравнением (24.6) при разных h , совокупность которых дает фазовый портрет на цилиндрическом фазовом пространстве, изображенном на рис. 24.3.

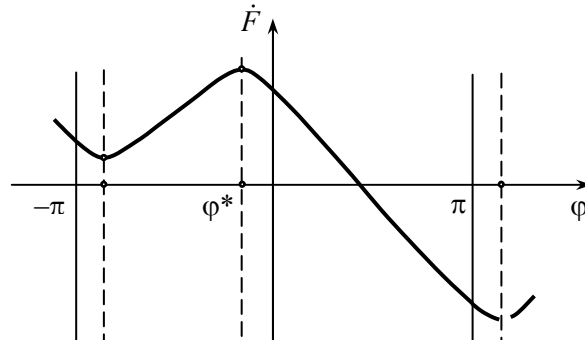


Рис. 24.2. График функции $F(\varphi)$ в формуле (24.6)

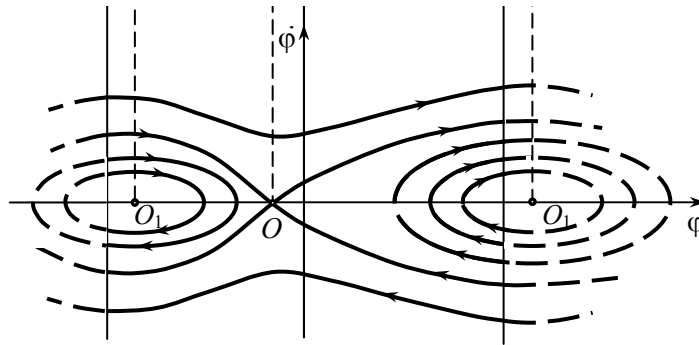


Рис. 24.3. Фазовый портрет на развертке цилиндра управляемого двухзвенного маятника, получаемый с помощью графика функции $F(\varphi)$

Физически то, что состояние рассматриваемого маятника имеет своим состоянием φ и $\dot{\varphi}$, причем $-\pi \leq \varphi \leq \pi$ и точки $\varphi = \pi, \dot{\varphi}$ и $\varphi = -\pi, \dot{\varphi}$ соответствуют одному и тому же состоянию, очевидно. Но вместе с тем уравнение фазовых траекторий (24.6) содержит φ , значение которого находится в пределах от $-\infty$ до $+\infty$, причем никакой периодичности по этому φ нет и изображать кривые линии, определяемые уравнением (24.6), надо на бесконечной плоскости $-\infty < \varphi < +\infty$, $-\infty < \dot{\varphi} < +\infty$. Как же согласовать это несоответствие и как уместить эти кривые на линии на цилиндре так, чтобы они были фазовыми траекториями исследуемого объекта.

Плоскость $(\varphi, \dot{\varphi})$, на которой можно поместить все кривые, определяемые уравнением (24.6), можно разбить на бесконечное число полос $(m\pi, (m+2)\pi)$, где m — любое нечетное целое число. На любой полосе располагаются кривые, определяемые уравнением вида

$$\dot{\varphi} = \pm \frac{2\omega^2}{1 + k \cos(\varphi - \alpha)} \left[h - \frac{3k}{2} \varphi \sin \alpha + G(\varphi) \right],$$

где $G(\varphi)$ – периодическая функция с периодом 2π . Эти кривые в m -й полосе можно перенести в полосу $(-\pi, \pi)$, отвечающую $m = -1$, заменив φ на $\varphi - (m + 1)\pi$, что эквивалентно замене h на $h - 3k/2 (m + 1)\pi \sin \alpha$. Таким образом, все кривые, описываемые уравнением (24.6), оказываются на цилиндре и составляют фазовый цилиндрический портрет рассматриваемого управляемого двухзвенного маятника.

На этом фазовом портрете два состояния равновесия O и O_1 . Равновесие $O(\varphi = \varphi^*)$, где φ^* находится из условия $\sin \varphi^* = -k \sin \alpha$, неустойчиво и отвечает положению маятника, показанному на рис. 24.4, т.е. перевернутому маятнику. Равновесие O_1 , отвечающее тому же условию, соответствует маятнику, висящему вниз. Это равновесие устойчиво. Возле этого устойчивого равновесия возможны колебания. Они соответствуют замкнутым фазовым траекториям, изображенным на рис. 24.3 области, содержащей равновесие O_1 .

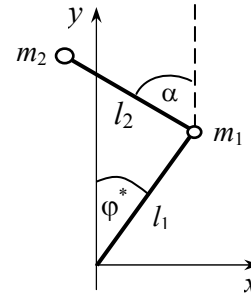


Рис. 24.4. Положение двухзвенного маятника, отвечающее неустойчивому равновесию (точка O на фазовом портрете рис. 24.3)

Все остальные фазовые траектории, кроме граничных траекторий, сепаратрис седла O , отвечают движениям с неограниченно возрастающей скоростью вращения маятника ($\dot{\varphi} \rightarrow \infty$). Это неожиданный эффект управления вторым звеном маятника: стремление поддерживать второе звено под одним и тем же углом α к вертикали приводит к ускоренному вращению основного звена маятника.

Мы привыкли, что маятник можно завращать, подталкивая его в одну и ту же сторону, сильно крутанув рукой непосредственно или с помощью ременной или зубчатой передачи. В этих случаях воздействие на маятник носит внешний характер. Но, оказывается, это можно сделать и иначе, с помощью подсоединения к нему управляемого маятника. Это приводит к тому, что основной маятник все быстрее и быстрее крутится.

Конечно, такое неограниченное раскручивание предполагает отсутствие диссипации (вязкого и сухого трения). При учете диссипации, например, небольшим вязким трением, возникает устойчивое вращение Γ с некоторой скоростью, как это видно из рис. 24.5. В области, охватываемой кривой, составленной из сепаратрис седла O , все фазовые траектории неограниченно приближаются к устойчивому равновесию O_1 . Равновесие O – по-прежнему неустойчивое седло. С дальнейшим увеличением коэффициента трения все движения сходятся к устойчивому равновесию O_1 .

Нас будет интересовать случай, когда имеется устойчивое вращение Γ , что соответствует фазовому портрету рис. 24.5.

Таким образом, в реальных условиях наличия потерь энергии описанное управление приводит к вращению маятника. Мы уже отмечали необычность этого способа порождения и поддержки вращения. В дополнение к этому отметим, что этот эффект неожиданным образом используется человеком в организации его двуногой ходьбы. Едва ли вы задумывались, как вы ходите. Но уже самое поверхностное наблюдение обнаруживает, что при ходьбе вовсе не обязательно подталкивать себя вперед ногами. Скорее при ходьбе ощущается, что какая-то «сила» тянет вас, и вы только выставляете поочередно ноги вперед, чтобы не упасть. Эта влекущая вас вперед сила ощущается не только при спуске, но и при ходьбе на ровной дороге и даже чуть-чуть в гору. Эту влекущую вперед «силу» можно увеличить, наклонив еще немного корпус вперед.

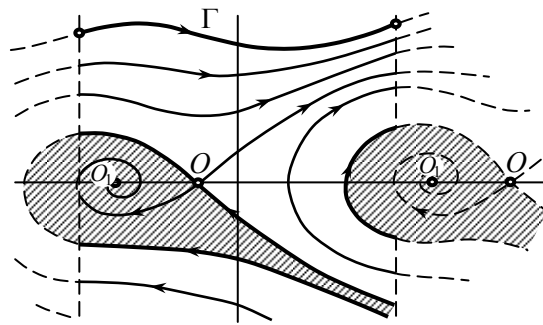


Рис. 24.5. Фазовый портрет перевернутого управляемого двойного маятника с учетом диссипации

Понять, откуда берется эта тяга вперед, можно, обратившись к фазовому портрету рис. 24.5, трактуя нашу ногу, на которую мы опираемся, как звено перевернутого маятника, а подсоединенный к ней наш корпус – как подсоединенное звено, которое мы поддерживаем под некоторым постоянным углом к вертикали: при стоянии корпус вертикален, при ходьбе несколько наклонен вперед. Наша опорная нога не может совершать полных вращений около своей точки опоры, она лишь поворачивается на некоторый угол от $-\beta$ до β , после чего нога H_1 сменяется и другая нога H_2 поворачивается от $-\beta$ до β . И так все время, пока вы не меняете длину шага (рис. 24.6). При этом мы несколько упрощаем характер движения опорной ноги; она не опирается одной точкой, опора осуществляется всей ступней. В целом это движение со сменой ног уподобляется качению колеса со спицами без обода.

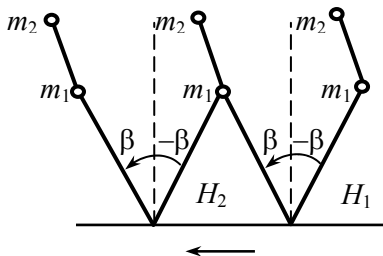


Рис. 24.6. Схема двуногой ходьбы как циклически повторяемые движения управляемого двойного маятника (автоколебательная ходьба)

В соответствии со сказанным проведем на фазовом портрете рис. 24.5 две вертикальные линии $\varphi = -\beta$ и $\varphi = \beta$ (рис. 24.7).

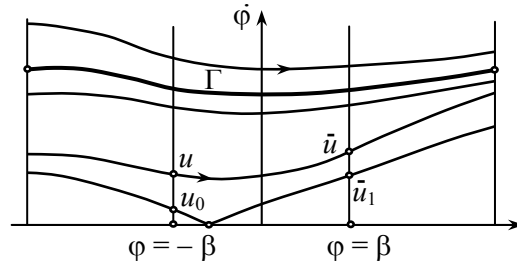


Рис. 24.7. Фрагмент фазового портрета двойного маятника с учетом диссипации

Фазовые траектории, покидая точки отрезка $\phi = -\beta$, могут пересекать при своем продолжении отрезок $\phi = \beta$. Пусть на начальном отрезке $\dot{\phi} = u$, а на конечном $\dot{\phi} = \bar{u}$. Отображение u в \bar{u} рис. 24.7 имеет диаграмму, показанную на рис. 24.8. Устойчивая неподвижная точка этой диаграммы отвечает равномерной ходьбе, у которой каждый новый шаг начинается со скоростью u , равной координате u^* этой неподвижной точки. При этом если вы использовали рис. 24.3, то диссипация полностью отсутствует, а если рис. 24.5, то она есть в течение всего движения от $\phi = -\beta$ до $\phi = \beta$.

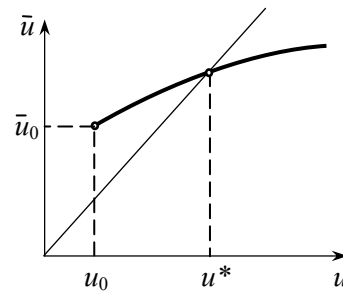


Рис. 24.8. Диаграмма Ламерея, обосновывающая автоколебательную ходьбу

При реальной ходьбе диссипация есть при перекачивании на опорной ноге и особенно при смене ног, так как при смене опорной ноги происходит «толчок» ее об опору, сопровождаемый потерей энергии. Тонкости механизма потери энергии не влияют на качественно правильный вид диаграммы рис. 24.8, неподвижная точка которой соответствует равномерной периодической ходьбе.

Резюмируя сказанное, отметим, что, как следует из предыдущего, не только человек, но и природа придумали свое колесо и не только колесо, а и своеобразные способы его вращения; у человека способ вращения основан на управлении положением корпуса и своевременной сменой ног, предотвращающей падение. Можно еще отметить, что для динозавров, у которых $k > 1$, этот способ требует дополнительных опор, что, по-видимому, и делал динозавр, опираясь на громадный хвост. Но это взгляд «сквозь магический кристалл»: детального анализа случая $k > 1$ мы не провели.

25. ДИНАМИЧЕСКИЕ МОДЕЛИ ИГР, ОБУЧЕНИЯ И ЦЕЛЕСООБРАЗНОГО ПОВЕДЕНИЯ

В последующих лекциях я расскажу вам о простейших динамических моделях целесообразного поведения, игр и обучения. Разработка этих вопросов только началась и в своем начале пересекалась с популярной дискуссией о том, «может ли машина мыслить?» Дискуссия была бурной и, пожалуй, формально безрезультатной: каждая из сторон осталась при своем мнении. Одни утверждали, что машина может мыслить, так как умеет решать сложные математические задачи, подчас не доступные человеку, умеет доказывать геометрические теоремы, находить выход из лабиринта, играть в шахматы, управлять ядерным реактором или ракетой, «читать» чертеж и по нему управлять изготовлением соответствующей детали и т.д. Другие возражали и говорили, что машина не мыслит, а лишь выполняет действия, начертанные ей человеком.

Ясно, что каждая из сторон по-своему была права, что дискуссия не имеет смысла, пока не договорились о том, что значит «мыслить», пока каждая из сторон понимает это по-своему.

Где же этот водораздел между бездумным выполнением команд и предписаний и разумным поиском решения и его осуществлением? Чем «мыслить» отличается от выполнения некоторой программы действий? Ведь всякий разумный подход тоже происходит по некоторому неведомому нам алгоритму. Возможно, этот алгоритм не детерминированный, а включает в себя элементы случайности. Может, в этой случайности и последующей фильтрации и есть суть мышления? А, может быть, в ассоциациях и умении увидеть аналогии и использовать имеющиеся знания или, наконец, в умении извлечь выводы из неудачных и удачных попыток? Но даже если все это так, то и это происходит в соответствии с некоторыми правилами.

Едва ли сегодня можно дать исчерпывающий ответ на эти вопросы: точно определить, что означает «мыслить». Да и как это можно сделать, когда мы фактически ничего не знаем о том, как работает наш мозг, которому мы приписываем возможность мышления?

Но мы знаем, как работают создаваемые нами ЭВМ, и применительно к ним можем попытаться установить, где наступает различие между «бездумным» выполнением команд, предписанных человеком, а где начинается «мышление» машины. В признании способности ЭВМ мыслить нас не должно смущать, что этой способностью машину наделил человек, ведь и человека ею тоже наделила природа, миллиарднолетний эволюционный процесс, как мы сейчас думаем, сводящийся к жесткому отбору случайных изменений.

Даже в этой специализированной суженной постановке вопроса едва ли сейчас целесообразно стремиться к однозначным четким ответам, лучше лишь обсудить их с вами, дав некоторую заправку для размышлений. Эта заправка состоит в рассказе о простейших моделях, в которых обнаруживаются элементы того, что, пожалуй, уже воспринимается как мышление.

Начнем с описания автоматных моделей целесообразного поведения и обучения. При этом субъект, принимающий решение, будет «мыслить» как некий автомат с внутренним состоянием, от которого зависит принимаемое решение, а само состояние определяется успешностью предшествующих решений. Идея и исследование таких моделей ведет начало от М.Л. Цетлина [47, 48]. Об использовании автоматов для оптимизации и управления см. [43].

Пусть у автомата возможных действий только два, тогда все состояния делятся на два класса: в одном принимается одно решение, в другом – другое. Обозначим эти решения «+1» и «-1». Примером такой ситуации может быть решение вопроса о том, в правой или левой руке находится монета, которая служит призом при угадывании. Можно представить и ситуацию с выбором одной из двух кормушек *A* или *B*, как изображено на рис. 25.1. С помощью кормушек изучалась способность крыс к обучению. Крысу время от времени выпускали через дверцу в помещение, из которого она могла пойти к одной из кормушек: направо *A* или налево *B*. Пища помещалась время от времени то в одну, то в другую кормушку и крыса в зависимости от того, куда она пошла, могла утолить свой голод или остаться голодной до следующего раза.

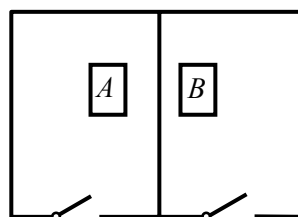


Рис. 25.1. Схема расположения двух кормушек для проведения эксперимента по обучаемости крыс или других животных или птиц

Первая ситуация соответствует игре в отгадывание, вторая напоминает жизненную ситуацию о том, куда отправиться за пищей.

Начнем с простейших автоматных описаний различных типов игроков: простака, упряма, хитреца и мистика. Будем считать, что игрок может находиться в двух состояниях, и в одном он принимает решение -1, а в другом +1 (рис. 25.2).

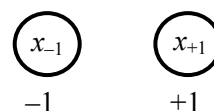


Рис. 25.2. Два разных состояния автомата-игрока

Решения -1 и +1 означают соответственно, что игрок выбирает левую или правую руку. Отгадывающий – отгадывая, загадывающий – кладя монету. То, какое состояние он имеет и как их меняет в зависимости от результатов предшествующих игр, зависит от того, кто он: простака, хитрец, упрямец или мистик. Простака считает, что если он выиграл, то не должен менять своего мнения – состояния о том, в какой руке монета. Напротив, если он проиграл, то состояние нужно сменить. Пусть величины $\xi = +1$ и $\xi = -1$ означают, выиграл или, соответственно, проиграл игрок предшествующее отгадывание. Тогда при $\xi = +1$ он не меняет своего состояния и при $\xi = -1$, напротив, его меняет, как это показано на рис. 25.3.

Хитрец же считает, что если он выиграл, то его партнер сменит положение монеты и поэтому он должен, выиграв, сменить свое состояние. При проигрыше он, напротив, не меняет состояние, так как думает, что его партнер сам его сменит. Стратегия игры хитреца изображена на рис. 25.4.

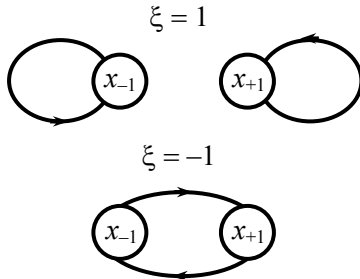


Рис. 25.3. Граф смены состояний автомата-игрока «простака» в зависимости от выигрыша ($\xi = 1$) или проигрыша ($\xi = -1$) в предыдущей партии

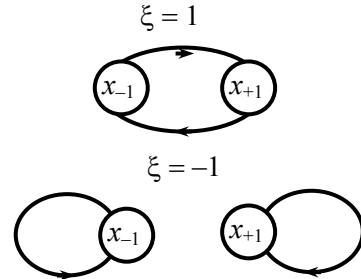


Рис. 25.4. Граф смены состояний автомата-игрока «хитрец» в зависимости от выигрыша ($\xi = 1$) или проигрыша ($\xi = -1$) в предыдущей партии

Упрямец не меняет своего решения и находится все время в состоянии x_{-1} или x_{+1} вне зависимости от результатов игры. Граф смен его состояний не зависит от результата ξ предшествующей игры (рис. 25.5).

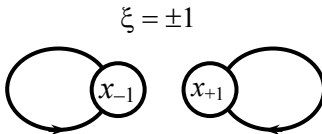


Рис. 25.5. Граф смены состояний автомата-игрока «упрямец» в зависимости от выигрыша ($\xi = 1$) или проигрыша ($\xi = -1$) в предыдущей партии

Мистик доверяет свое решение случаю, бросая некий, не обязательно симметричный, кубик и принимая в соответствии с этим независимо состояния x_{-1} или x_{+1} с вероятностями p и $q = 1 - p$. При этом он не обращает внимания на свои выигрыши и проигрыши, полностью полагаясь на свой талисман – кубик. Такое поведение также можно описать (моделировать) автоматом с двумя состояниями, но теперь его состояния меняются не в строгой зависимости от проигрыша или выигрыша, а случайно в соответствии со стрелками и указанными на них вероятностями графа смен состояний (рис. 25.6).

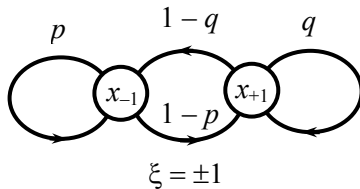


Рис. 25.6. Граф вероятностей перехода «мистика» ($p + q = 1$)

Мы описали стратегию отгадывающего простака, хитреца, упрянца и мистика. Аналогичные стратегии могут быть у загадывающего простака, хитреца, упрянца и мистика. Такие же стратегии могут быть и у крысы, ищущей пищу. Однако в случае крысы едва ли целесообразны стратегии хитреца, мистика или упрянца. Хотя все зависит от того, каковы правила размещения пищи в кормушках.

В описании возможных стратегий игры в отгадывание мы ограничились двумя состояниями и предполагали, кроме мистика, детерминированную их смену

после каждой партии в зависимости от ее результата. Число состояний можно увеличить и их смене придать стохастический характер. Так можно ввести простака с памятью, который не сразу меняет свое мнение, а после первой неудачи только начинает сомневаться в нем и лишь при повторном проигрыше его меняет. То, как происходит смена состояний у такого памятьливого простака, изображено на рис. 25.7.

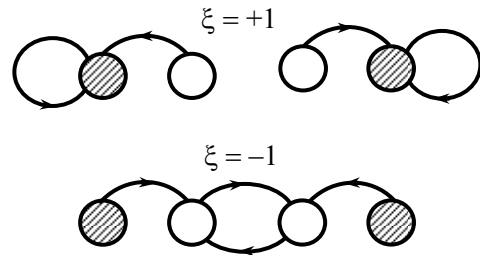


Рис. 25.7. Граф смен состояний для «простака» с памятью (так называемая линейная тактика игры)

На нем четыре состояния: крайние правое и левое соответствуют уверенности, какое надо принять решение; средние, напротив, тому, когда решение принимается, но с некоторым сомнением. Ясно, что состояний может быть не четыре, а любое число. Каждому из решений могут соответствовать не обязательно равные числа состояний, как в предшествующем примере, где их по два. Можно придумать и другие автоматные модели, например, с графом смены состояний (рис. 25.8).

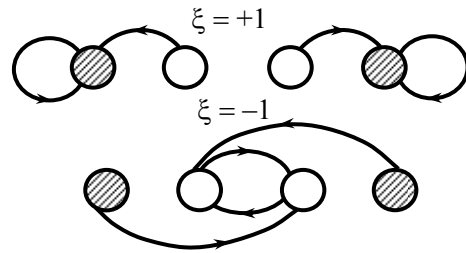


Рис. 25.8. Автомат-игрок с памятью, но другим графом смен состояний, чем на рис. 25.7

На рис. 25.9 изображен граф смены состояний для принимающего решения автомата с двумя состояниями, но в случае, когда их смена происходит случайно с вероятностями $1 - p$ при выигрыше, когда $\xi = +1$, и, соответственно, с вероятностями $1 - q$ при проигрыше, когда $\xi = -1$. Игрок со стохастической стратегией, изображенной на графе рис. 25.9: при $p = 1, q = 0$ – простак; при $p = 0, q = 1$ – хитрец; при $p = 1, q = 1$ – упрямец.

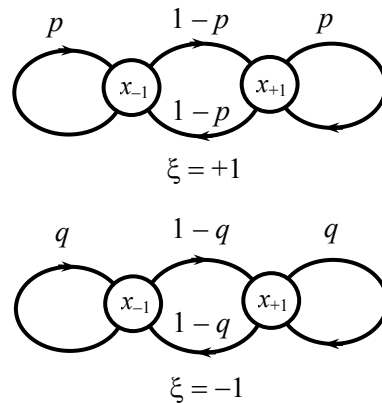


Рис. 25.9. Граф вероятностей перехода для автомата-игрока «стохастик»

Мы описали простейшие стратегии для игры в отгадывание у другого игрока или у «природы». Природу также можно рассматривать как игрока, безразличного к проигрышу или выигрышу.

В качестве простейших моделей стратегии стохастической природы можно принять последовательность независимых решений с вероятностями p и q ($p + q = 1$), как у мистика, или марковскую последовательность, определяемую некоторой матрицей смен состояний. Например, имеется два состояния x_{-1} и x_{+1} , в них принимаются решения -1 и соответственно $+1$. Смена состояний x_{-1} и x_{+1} происходит случайно, с некоторыми неизменными вероятностями, например, δ (рис. 25.10). Случайные независимые смены состояний x_{-1} на x_{+1} и наоборот могут происходить и с разными вероятностями δ_1 и δ_2 .

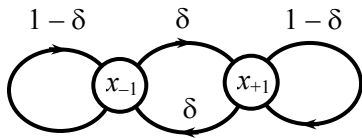


Рис. 25.10. Граф вероятностей перехода «стохастика» при безразличии к выигрышу или проигрышу

После того как мы познакомились с простейшими автоматными моделями игроков, рассмотрим игру в отгадывание между двумя игроками: простаком и хитрецом. Посмотрим, кто кого обыграет. Состояние одного игрока – простака – обозначим x_{-1} и x_{+1} , состояние второго – хитреца – y_{-1} и y_{+1} . У них вместе будет четыре состояния (x_{-1}, y_{-1}) , (x_{+1}, y_{-1}) , (x_{-1}, y_{+1}) , и (x_{+1}, y_{+1}) . В состояниях x_{-1} и y_{-1} отгадывающий выбирает левую руку, а загадывающий в нее берет монету, а в состояниях x_{+1} и y_{+1} отгадывающий называет правую руку, а загадывающий в ней держит монету. В состояниях (x_{-1}, y_{-1}) и (x_{+1}, y_{+1}) выигрывает отгадывающий, а в состояниях (x_{-1}, y_{+1}) и (x_{+1}, y_{-1}) выигрывает загадывающий. Теперь можно нарисовать граф смены состояний динамической системы загадывающий – отгадывающий (рис. 25.11). Результат несколько неожиданный. Если выиграл первую партию отгадывающий простака, то он будет выигрывать и все остальные партии. Если же выиграл первую партию загадывающий хитрец, то он будет выигрывать и дальше.

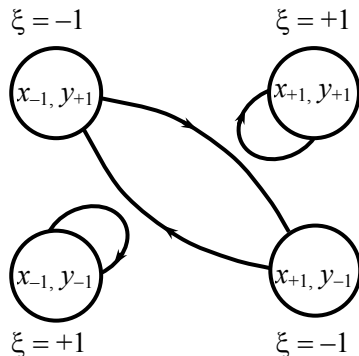


Рис. 25.11. Фазовый портрет игры «простака» с «хитрецом»

Аналогичный граф для двух простаков изображен на рис. 25.12. Их игра протекает периодически с периодом в четыре игры: в двух партиях выигрывает загадывающий простака, а в двух других – отгадывающий. Так что общий результат ничейный.

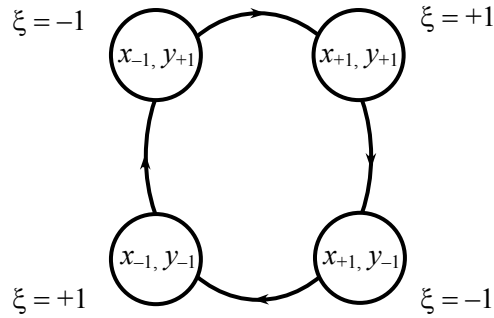


Рис. 25.12. Фазовый портрет игры двух «простаков»

Рассмотрим игру двух игроков со стохастическими стратегиями. Это рассмотрение включает в себя как частные случаи все варианты игр хитреца и простака, а также мистика или природы (см. рис. 25.9). У одного пусть соответствующие вероятности p, q , а у другого r, s . Игроки могут находиться в четырех комбинациях своих состояний $(x_{-1}, y_{-1}), (x_{+1}, y_{-1}), (x_{-1}, y_{+1})$ и (x_{+1}, y_{+1}) . В первом и последнем состояниях выигрывает отгадывающий, в двух средних – загадывающий. Смены состояний обоих игроков могут происходить в любом случае: при выигрыше с вероятностями $1 - p$ и $1 - r$, а при проигрыше – с вероятностями $1 - q$ и $1 - s$ соответственно. Пусть играют последовательно несколько партий и $p_n(x_{-1}, y_{-1}), p_n(x_{+1}, y_{-1}), p_n(x_{-1}, y_{+1})$ и $p_n(x_{+1}, y_{+1})$ – вероятности соответствующих комбинаций x_i и y_j в n -й партии. В начальный момент может иметь место определенное сочетание состояний x_i, y_j , так что только одна из вероятностей $p(x_i, y_j)$ равна единице, а остальные – нулю.

Ясно, что теперь комбинации (x_i, y_j) не являются состояниями, так как знание x_i и y_j не позволяет предсказать, какие будут x_i и y_j на следующем такте. Но можно предсказать, каковы их вероятности. Это наводит на мысль, что за состояние следует принять вектор \bar{p} , компонентами которого являются вероятности $p(x_i, y_j)$.

Допустим, что на n -й игре вектор состояний известен и равен \bar{p}^n . Найдем вектор \bar{p}^{n+1} , т.е. найдем его компоненты

$$p^{n+1}(x_{-1}, y_{-1}), p^{n+1}(x_{+1}, y_{-1}), p^{n+1}(x_{-1}, y_{+1}) \text{ и } p^{n+1}(x_{+1}, y_{+1})$$

через

$$p^n(x_{-1}, y_{-1}), p^n(x_{+1}, y_{-1}), p^n(x_{-1}, y_{+1}) \text{ и } p^n(x_{+1}, y_{+1}).$$

Оказывается, что

$$\bar{p}^{n+1} = \bar{p}^n P, \tag{25.1}$$

где P – матрица вероятностей перехода,

$$P = \begin{pmatrix} ps & (1-p)s & p(1-s) & (1-s)(1-p) \\ (1-q)r & qr & (1-q)(1-r) & q(1-r) \\ q(1-r) & (1-q)(1-r) & qr & (1-q)r \\ (1-p)(1-s) & (1-s)p & s(1-p) & ps \end{pmatrix}, \tag{25.2}$$

а \bar{p}^n и \bar{p}^{n+1} – векторы строки.

Если перенумеровать комбинации «состояний» (x_{-1}, y_{-1}) , (x_{-1}, y_{+1}) , (x_{+1}, y_{-1}) , (x_{+1}, y_{+1}) номерами 1, 2, 3, 4, а элементы матрицы P обозначить p_{ij} , где i – номер строки, а j – номер столбца, то p_{ij} – это вероятность перехода из комбинации «состояний» с номером i в комбинацию «состояний» с номером j . Так, элемент p_{12} – это вероятность перехода (x_{-1}, y_{-1}) в (x_{-1}, y_{+1}) , равная произведению вероятностей смены состояния выигравшего отгадывающего, т.е. $1 - p$, на вероятность сохранения состояния проигравшего загадывающего, т.е. s . Аналогично находятся все остальные элементы матрицы P .

Справедливость формулы (25.1) следует из соотношения

$$p_i^{n+1} = p_1^n p_{1i} + p_2^n p_{2i} + p_3^n p_{3i} + p_4^n p_{4i},$$

означающего, что i -я комбинация состояний может возникнуть из первой с вероятностью p_{1i} , из второй – p_{2i} , из третьей – p_{3i} и из четвертой – с вероятностью p_{4i} , а вероятности первой, второй, третьей и четвертой комбинаций, соответственно, p_1^n , p_2^n , p_3^n и p_4^n .

Итак, знание вектора вероятностей \bar{p}^n позволяет найти его следующее значение \bar{p}^{n+1} после еще одной сыгранной партии. Это позволяет нам игру двух стохастических автоматов рассматривать как динамическую систему с состоянием, описываемым вектором \bar{p} .

Фазовым пространством этой динамической системы являются всевозможные векторы $\bar{p}(p_1, p_2, p_3, p_4)$, у которых все p_i не отрицательны и их сумма равна единице, так как в одном из несовместных «состояний» 1, 2, 3 или 4 рассматриваемая система обязательно находится. Это фазовое пространство имеет размерность три и представляет собою трехмерный симплекс. Вершины этого симплекса Σ_3 , изображенного на рис. 25.13, отвечают единичным векторам $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ и $(0, 0, 0, 1)$. На четырех гранях соответственно $p_1 = 0$, $p_2 = 0$, $p_3 = 0$ и $p_4 = 0$.

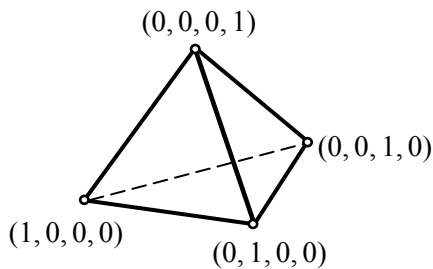


Рис. 25.13. Фазовое пространство в виде трехмерного симплекса в случае, когда хотя бы один автомат-игрок меняет свои состояния случайно с какими-то вероятностями, могущими зависеть от результата предшествующей игры

Динамическая система с фазовым пространством в виде симплекса Σ_3 (а в общем случае с симплексом любой размерности) и оператором, определяемым матрицей P , согласно формуле (25.1), называется дискретной марковской системой.

Фазовый портрет такой системы в общем случае очень прост: все его фазовые траектории стремятся к единственному равновесию \bar{p}^* , определяемому из условия

$$\bar{p}^{n+1} = \bar{p}^n = \bar{p}^*$$

или, согласно (25.1),

$$\bar{p}^* = \bar{p}^* P. \quad (25.3)$$

Уравнение (25.3) в общем случае имеет единственное ненулевое нормированное решение \bar{p}^* . Марковская система определяется матрицей P , т.е. ее элементами $p_{ij} \geq 0$. В силу этого под общим случаем понимается тот, когда ни одно из p_{ij} не обращается в нуль. В нашем конкретном случае, когда матрица P вида (25.2), общий случай имеет место при p, q, r и s , отличных от нуля и единицы. При выполнении этого общего условия уравнение (25.3) имеет единственное решение $\bar{p}^*(p_1^*, p_2^*, p_3^*, p_4^*)$, для которого $p_1^* + p_2^* + p_3^* + p_4^* = 1$.

Это условие существования единственного глобально устойчивого равновесия, к которому асимптотически приближаются все точки фазового пространства, является достаточным. Необходимое и достаточное условие состоит в том, что некоторая степень матрицы P , т.е. P^k , удовлетворяет этому достаточному условию положительности всех своих элементов.

В ситуациях (x_{-1}, y_{-1}) и (x_{+1}, y_{+1}) , т.е. 1 и 4, выигрывает угадывающий, в остальных (2 и 3) он столько же (пятак) проигрывает, так что математическое ожидание его выигрыша f при длительной игре близко к его значению в равновесном состоянии

$$Mf = 5(p_1^* + p_4^*) - 5(p_2^* + p_3^*). \quad (25.4)$$

Полученное выражение (25.2) матрицы переходных вероятностей используем, чтобы рассмотреть игры простака, хитреца и мистика со стохастиком с вероятностями $0 < r < 1$ и $0 < s < 1$.

Для простака $p = 1$ и $q = 0$, поэтому, согласно (25.1), матрица P записывается в виде:

$$P = \begin{pmatrix} s & 0 & 1-s & 0 \\ r & 0 & 1-r & 0 \\ 0 & 1-r & 0 & r \\ 0 & 1-s & 0 & s \end{pmatrix}, \quad (25.5)$$

что приводит к уравнениям, определяющим равновесие:

$$\begin{aligned} p_1^* &= sp_1^* + rp_2^*, \\ p_2^* &= (1-r)p_3^* + (1-s)p_4^*, \\ p_3^* &= (1-s)p_1^* + (1-r)p_2^*, \\ p_4^* &= rp_3^* + sp_4^*, \end{aligned} \quad (25.6)$$

к которым следует присоединить условие нормировки

$$p_1^* + p_2^* + p_3^* + p_4^* = 1. \quad (25.7)$$

Решение этих уравнений (25.6) и (25.7) легко находится, если сначала найти искомые p_1^*, p_2^*, p_3^* , и p_4^* и с точностью до общего множителя из (25.6):

$$p_2^* = p_3^* = \frac{1-s}{r} p_1^*, \quad p_4^* = p_1^*, \quad (25.8)$$

а затем – этот множитель из (25.7). В результате найдем, что

$$p_1^* = p_4^* = \frac{r}{2(1+r-s)}, \quad p_2^* = p_3^* = \frac{1-s}{2(1+r-s)}, \quad (25.9)$$

и, следовательно, математическое ожидание (25.4) выигрыша отгадывающего простака равно

$$Mf = 5 \frac{-1+r+s}{1+r-s}. \quad (25.10)$$

Теперь легко исследовать, как зависит выигрыш в глобально устойчивом равновесии от вероятностей r и s . Полагая $Mf = 5\mu$, согласно (25.10) легко приходим к тому, что на линии

$$s + \frac{r(1-\mu)}{1+\mu} = 1 \quad (-1 \leq \mu \leq 1) \quad (25.11)$$

внутри квадрата $0 < r < 1, 0 < s < 1$ математическое ожидание (25.10) равно 5μ (рис. 5.14).

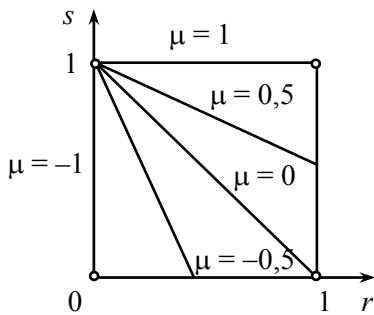


Рис. 5.14. Геометрическое представление зависимости выигрыша μ от вероятностей r и s смен состояний в игре «простака» со «стохастиком»

Таким образом, простака в зависимости от вероятностей r и s может как выигрывать, так и проигрывать, а также иметь любой промежуточный результат. Близок простака к максимально возможному выигрышу при $\mu \approx +1$ и, напротив, к максимальному проигрышу при $\mu \approx -1$. При $\mu \approx -1$ стохастик близок к хитрецу с $r = 0$ и $s = 1$, но хитрец у простака не всегда выигрывает, а сколь угодно близкий стохастик выигрывает. Этот скачок вызван тем, что при $r = 0$ и $s = 1$ нарушаются условия существования глобально устойчивого равновесия.

Аналогично почти максимально проигрывающий стохастик близок к упрямцу с $r = 1$ и $s = 1$, который у простака не выигрывает. Это несоответствие (скачок) опять же объясняется нарушением условия глобальной устойчивости равновесия \bar{p}^* при $r = s = 1$.

Для игры хитреца со стохастиком ($p = 0, q = 1$ в формуле (25.2)) матрица P запишется в виде

$$P = \begin{pmatrix} 0 & s & 0 & 1-s \\ 0 & r & 0 & 1-r \\ 1-r & 0 & r & 0 \\ 1-s & 0 & s & 0 \end{pmatrix}, \quad (25.12)$$

и поэтому равновесие находится из уравнений:

$$\begin{aligned} p_1^* &= (1-r)p_3^* + (1-s)p_4^*, \\ p_2^* &= sp_1^* + rp_2^*, \\ p_3^* &= rp_3^* + sp_4^*, \\ p_4^* &= (1-s)p_1^* + (1-r)p_2^* \end{aligned} \quad (25.13)$$

и условия нормировки (25.7). Из (25.13) и (25.7) находим, что

$$p_1^* = p_4^* = \frac{1-r}{2(1+s-r)}, \quad p_2^* = p_3^* = \frac{s}{2(1+s-r)}. \quad (25.14)$$

Далее оказывается, что математическое ожидание выигрыша хитреца в глобально устойчивом равновесном режиме равно

$$Mf = 5 \frac{1-r-s}{1+s-r}. \quad (25.15)$$

Найденное Mf равно $-Mf$ формулы (25.10), если в ней поменять местами s и r , поэтому имеет место тот же, что и раньше, график (см. рис. 25.13) с заменой μ на $-\mu$ и оси r на s и s на r .

Теперь рассмотрим игру мистика, для которого $p = q = \alpha$, со стохастиком. В этом случае матрица переходных вероятностей имеет, согласно (25.2), вид:

$$P = \begin{vmatrix} \alpha s & (1-\alpha)s & \alpha(1-s) & (1-\alpha)(1-s) \\ (1-\alpha)r & \alpha r & (1-\alpha)(1-r) & \alpha(1-r) \\ \alpha(1-r) & (1-\alpha)(1-r) & \alpha r & (1-\alpha)r \\ (1-\alpha)(1-s) & \alpha(1-s) & (1-\alpha)s & \alpha s \end{vmatrix}. \quad (25.16)$$

В соответствии с (25.16) уравнения устойчивого равновесия запишутся в виде:

$$\begin{aligned} (\alpha s - 1)p_1^* + (1-\alpha)rp_2^* + \alpha(1-r)p_3^* + (1-\alpha)(1-s)p_4^* &= 0, \\ (1-\alpha)sp_1^* + (\alpha r - 1)p_2^* + (1-\alpha)(1-r)p_3^* + \alpha(1-s)p_4^* &= 0, \\ \alpha(1-s)p_1^* + (1-\alpha)(1-r)p_2^* + (\alpha r - 1)p_3^* + (1-\alpha)sp_4^* &= 0, \\ (1-\alpha)(1-s)p_1^* + \alpha(1-r)p_2^* + (1-\alpha)rp_3^* + (\alpha s - 1)p_4^* &= 0. \end{aligned} \quad (25.17)$$

Эти уравнения линейно зависимы и поэтому допускают ненулевое решение $p_1^*, p_2^*, p_3^*, p_4^*$, которое затем следует нормировать на единицу. Манипулируя с этой системой уравнений (25.17) (складывая первое и третье уравнения, и третье и четвертое), приведем ее к виду (отбрасывая четвертое):

$$\begin{aligned} p_1^* - p_2^* + p_3^* - p_4^* &= 0, \\ (1-s)p_1^* + (1-r)p_2^* + (r-1)p_3^* + (s-1)p_4^* &= 0, \\ \alpha(1-s)p_1^* + (1-\alpha)(1-r)p_2^* + (\alpha s - 1)p_3^* + (1-\alpha)sp_4^* &= 0 \end{aligned}$$

и непосредственно найдем, что требуемое нормированное на единицу решение будет следующим:

$$\begin{aligned}
 p_1^* &= p_4^* = \Delta^{-1}(2-s-r)(r+\alpha-2\alpha r), \\
 p_2^* &= p_3^* = \Delta^{-1}(2-\alpha-s)(\alpha+s-2\alpha s), \\
 \Delta &= 2(\alpha-s-r)[2\alpha+s+r-2\alpha(r+s)],
 \end{aligned}$$

а математическое ожидание выигрыша в равновесном состоянии будет

$$Mf^* = 5(p_1^* + p_4^* - p_2^* - p_3^*) = \frac{5(r-s)(1-2\alpha)}{2\alpha+r+s-2\alpha(r+s)}. \quad (25.18)$$

Из выражения (25.18) следует, что при $\alpha = 0,5$ мистик не проигрывает и не выигрывает у любого автомата (при любых r и s). Напротив, при $\alpha \neq 0,5$ он может как проигрывать, так и выигрывать.

При $\alpha < 0,5$ он выигрывает при $r > s$ и проигрывает при $r < s$. При этом его максимальный выигрыш у простака, когда $r = 1, s = 0$; максимальный проигрыш хитрецу, когда $r = 0, s = 1$.

При $\alpha > 0,5$ все наоборот: он выигрывает при $r < s$ максимально у хитреца, а проигрывает при $r > s$ максимально у простака. (Эти выводы о максимальном выигрыше и проигрыше предполагают α отличным от нуля и единицы, что опять связано с нарушением условий глобальной устойчивости равновесия.)

Итак, мы рассмотрели игры простака и хитреца между собой, а также игры простака, хитреца и мистика со стохастиком. Мы нашли вероятности различных ситуаций в игре в установившемся равновесном состоянии и выигрыши и проигрыши игроков в нем. Для полноты представления об этих играх напомним соответствующие им графы смены ситуаций для игр детерминированных игроков простака с хитрецом и простака с простаком, изображенные на рис. 25.11 и 25.12. В этом случае графы смен ситуаций одновременно являются и фазовыми портретами, потому что ситуация является одновременно и состоянием. При наличии в игре стохастичности ситуация не является состоянием и граф смены ситуаций не является фазовым портретом. Фазовый портрет – это всевозможные фазовые траектории точки $\bar{p} (p_1, p_2, p_3, p_4)$ трехмерного симплекса

$$p_s \geq 0, \quad p_1 + p_2 + p_3 + p_4 = 1.$$

Этот фазовый портрет в общем случае положительности всех вероятностей перехода p_{ij} прост и состоит из фазовых траекторий, стягивающихся к единственному устойчивому равновесию $\bar{p}^* (p_1^*, p_2^*, p_3^*, p_4^*)$ (рис. 25.15).

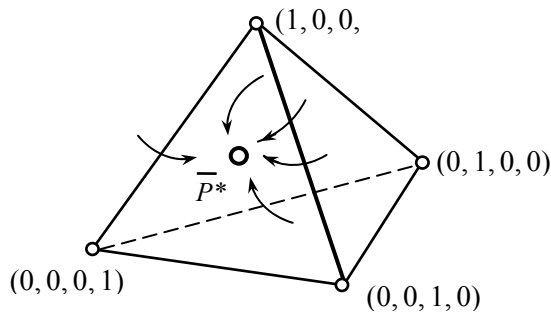


Рис. 25.15. Иллюстрация к наличию глобально-устойчивого равновесия в игре со случайными сменами состояний

Граф смены ситуаций может быть непосредственно нарисован по матрице переходных вероятностей P (25.2). Матрицы смен состояний игроков при играх простака и хитреца со стохастиком и стохастического мистика со стохастиком являются частными случаями этой матрицы P . Поэтому можно ограничиться одним рис. 25.16 графа переходов матрицы P вида (25.2).

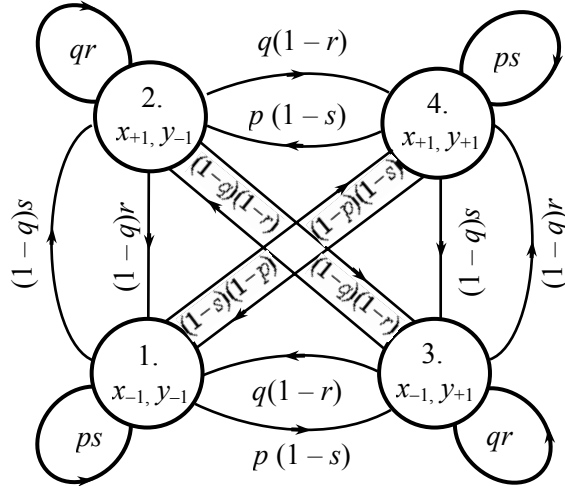


Рис. 25.16. Граф вероятностей перехода в игре двух «стохастиков»

Мы рассмотрели очень простые примеры игр в отгадывание. А где же в них обучение? Оно состоит в том, что игроки «простак», «хитрец», «стохастик» действовали не по заранее фиксированному плану, а на основе результатов предшествующих игр, поскольку они выбирали принимаемое в следующей партии решение в зависимости от результатов игры. Упрямец не обучался, он действовал по заранее определенному плану, он мог бы быть и другим, реализующим некоторую произвольную, заранее заданную последовательность действий (ходов). Мистик также не обучался, хотя и не имел заранее заданного плана, но его действия не зависели от хода игры, они определялись независимым от него случаем. Стратегии упряמצа и мистика не предполагают обучение. Напротив, стратегии простака, хитреца и стохастика таковы, что имеет место обучение.

Простейшие стратегии игры в отгадывание, которые мы рассмотрели, применимы и к принятию крысой решения о том, к какой из кормушек пойти, чтобы поесть, куда отправиться за пищей. Этот случай соответствует игре отгадывающего с природой, поскольку принимается, что то, где окажется еда, определяется причинами, не зависящими от действий крысы. Есть еще одно отличие в стратегии природы от стратегии загадывающего игрока: не обязательно одна кормушка с едой, другая без нее, возможны и случаи, когда обе кормушки с едой или обе без нее.

В интерпретации игры в отгадывание, как и крысы, ищущей еду и улучшающей свой поиск в результате учета результатов своих предшествующих действий, элемент обучения очевиден. Этот элемент обучения заложен в играющего, он очень прост до примитивности и поэтому, возможно, еще не ощущаются его принципиальное отличие и совершенно новые возможности по сравнению с жестко заданными, не использующими предшествующего опыта, алгоритмами поведения. Хотелось бы подчеркнуть, что новое в обучении не только в иных алго-

ритмах действий, а и в новой информации, используемой алгоритмом: информации о результатах предшествующих действий.

В рассмотренных выше примерах эта информация использовалась для принятия решения перехода автомата в то или иное состояние. Предполагается и более общее и глубокое использование этой информации для изменения самого автомата, в частности, вероятностей p и q стохастического автомата.

Описанная стратегия поведения крысы с обучением представляется нам целесообразной. Вместе с тем, не так уж очевидно, почему это так: неясно, с чем ее сравнивать и по отношению к чему она лучше. Это очень непростой вопрос. Его обсуждение увело бы нас очень далеко от нашей темы. Поэтому ограничимся сравнением со случайным поведением, и то, что лучше его, будем считать целесообразным. Случайное поведение с независимыми выборами имеет равные вероятности выигрыша и проигрыша. Целесообразная стратегия – это та, для которой выигрыш более 0,5. Такое понимание, по-видимому, согласуется с нашей интуицией. Более полный анализ этого вопроса вас, возможно, ждет в специальных курсах.

В заключение хотелось бы остановиться на двух общих вопросах, не имеющих прямого отношения к теме нашего разговора – это объяснение того, почему стохастик с $p = q = 0,5$ ни у кого не выигрывает и никому не проигрывает и почему марковская система с матрицей P , у которой все элементы p_{ij} положительны, имеет единственное устойчивое равновесие \bar{p}^* . То, что он не проигрывает и не выигрывает ни у простака, ни у хитреца, ни у стохастика, мы получили из конкретного рассмотрения математических моделей их игр. Но естественен вопрос: не может ли все же существовать такая стратегия игры против стохастика, равновероятно и независимо выбирающего -1 и $+1$, при которой его можно обыграть, как об этом мечтают фанатики игорных домов? Простое рассуждение показывает, что это невозможно. Невозможно в силу независимости реализаций у стохастика -1 и $+1$ ни от его предшествующих ходов, ни от действий партнера. В силу этой независимости условные вероятности его выигрыша и проигрыша всегда одни и те же, равные 0,5.

В свое время мне и Д.М. Черток захотелось придумать более сложные, чем описано выше, обучающиеся стратегии, и с их помощью обыграть человека. Человек – не равновероятный стохастик: человек невольно учитывает результаты предыдущих игр, и если этот его учет разгадать – проигрыш человека неизбежен. Для того чтобы реализовать стратегию игры с человеком, основанную на разгадывании его стратегии, нужна математическая модель человека-игрока. Достаточно эффективным оказалось предположение, что стратегия человека-игрока имеет случайный характер, но вероятности выбора -1 и $+1$ зависят от конечного, не очень большого числа предшествующих партий, и то, что человек в результате этого выбирает -1 и $+1$, не равновероятно и не независимо.

Вместе с тем машина, таким способом обыгрывающая человека, ставит себя в рискованное положение и, следовательно, может проиграть человеку, если он придумает еще более совершенную, чем у машины, стратегию, основанную на разгадке ее стратегии. Этот процесс взаимного совершенствования стратегий может длиться долго, как войны человечества. Это не совсем прямая аналогия, потому что в войнах совершенствуются не только стратегии ее ведения, но и средст-

ва, которыми она ведется, что, впрочем, тоже можно считать стратегией. Насущная жизненно необходимая проблема человечества – обрыв этой цепочки эволюционного совершенствования, заложенного в инстинктах человека как представителя мира живых существ. Это, по-видимому, возможно только через осознание им гибельности этого «естественного» пути, гибельности в новых невиданных еще условиях его существования, в осознании необходимости коренного изменения стратегий коллективного поведения перед лицом экологической катастрофы. В связи со сказанным особый интерес представляет построение не только моделей сосуществования на основе дарвиновской теории борьбы за существование, но и моделей сосуществования современного человечества в новых условиях его существования. Скорее всего, это переход к новым эволюционным стратегиям, основанным не на открытой борьбе за существование, а на стратегиях компромиссов и соглашений, которые должны быть взаимовыгодными, как в кратковременных, так и долгосрочных планах, в понимании, что кроме примитивных инстинктивных стратегий, возможны другие, много более целесообразные и выгодные стратегии.

Наличие таких стратегий ниже иллюстрируется примером Ю.Б. Гермейера, хорошо объясняющим возможность более выгодных стратегий, чем прямая максимизация собственного выигрыша. Речь идет об игре n лиц, в которой выигрыш s -го игрока равен

$$f_s = 1 - x_s + \mu \sum_{i \neq s} x_i \quad (s = 1, 2, \dots, n). \quad (25.19)$$

Каждый игрок волен распоряжаться своей, выбираемой им переменной $0 \leq x_s \leq 1$.

Ясно, что непосредственная максимизация s -м игроком своего выигрыша f_s приводит к выбору $x_s = 0$. Если так поступят все игроки, то выигрыш каждого из них равен 1. Вместе с тем, если они договорятся о том, что не будут стремиться к непосредственной выгоде, а все выберут $x_s = 1$, то выигрыши каждого из них будут не единица, а, возможно, много большие $\mu(n-1)$. Такая договоренность не имеет устойчивого характера: каждому из игроков выгодно ее нарушить и получить еще больший выигрыш $\mu(n-1) + 1$. Однако если так поступят m игроков, то их выигрыш будет $\mu(n-m-1) + 1$, а остальных – $\mu(n-m-1)$. Ясно, что если $\mu(n-m-1)$ окажется меньше единицы, то для них общая договоренность станет невыгодной и «союз» распадется к общей невыгоде всех. Договоренности можно придать устойчивый характер, если прибегнуть к общей кассе, т.е. договориться, что выигрыши всех поступают в общую кассу и затем, например, поровну раздаются всем. В этом случае общий выигрыш равен

$$\begin{aligned} F &= \sum f_s = n - x_1 - \dots - x_n + \mu(n-1)(x_1 + \dots + x_n) = \\ &= n + (\mu(n-1) - 1)(x_1 + x_2 + \dots + x_n). \end{aligned}$$

Пусть $\mu(n-1) - 1 > 1$, тогда не только всем вместе, но и каждому из игроков выгодно выбирать $x_s = 1$.

Второй общий вопрос, по-видимому, был вам обоснован в лекциях по теории вероятностей. Удивительность и важность этого общего факта делает желательным эвристически наглядное его понимание, которое и приводится ниже. Я только хочу напомнить вам, что условия $p_{ij} > 0$ достаточные, а не необходимые. Необ-

ходимые и достаточные условия тоже известны. Я не буду вам о них рассказывать, хотя они очень красивые и непосредственно проверяются по виду графа возможных смен ситуаций. Итак, пусть все $p_{ij} > 0$.

При этом линейном преобразовании с матрицей P симплекс Σ_n переходит опять-таки в симплекс $\bar{\Sigma}_n$, который лежит строго внутри симплекса Σ_n . Строго внутри потому, что все вершины симплекса $\Sigma_n(1, 0, \dots, 0)$, $(0, 1, \dots)$, \dots , $(0, 0, \dots, 1)$ переводятся в силу $p_{ij} > 0$ строго внутрь Σ_n . Это означает, что симплекс Σ_n преобразуется внутрь геометрически подобного симплекса $\tilde{\Sigma}_n$, который в $q < 1$ раз меньше симплекса Σ_n . При следующем преобразовании аналогично симплекс $\tilde{\Sigma}_n$ преобразуется в q раз меньший, подобный ему симплекс $\tilde{\tilde{\Sigma}}_n$ и так далее. Последовательность подобных неограниченно уменьшающихся с числом преобразований симплексов стягивается к некоторой внутренней точке p^* .

26. ПЕРСЕПТРОН И РАСПОЗНАВАНИЕ ОБРАЗОВ

Перейдем теперь к модели, которая появилась сравнительно недавно, в середине XX века, и получила широкую известность уже в шестидесятые годы. Модели безусловно интеллектуальной деятельности человека, которая реализуема на ЭВМ и тем самым доказывает, что машина может мыслить. Модели, которая потрясла современников и послужила началом нового раздела кибернетики – распознавания образов [26]. Сначала появилась не математическая модель, а удивительное устройство, которое придумал Ф. Розенблат в 1957 году и которому дал звучное название «персептрон». Придуманно оно было, исходя из тогдашних представлений о конструкции мозга, сетчатки глаза и обработки ее сигналов в слоях нервных клеток. По профессии Розенблат был физиолог. Его устройство могло воспринимать, различать и узнавать зрительные образы, отличать треугольники от овалов, опознавать буквы и другие изображения. Самое удивительное было в том, что эти способности не были заложены в устройстве, а возникали в результате обучения. Обучение позволяло различать новые геометрические фигуры и зрительные образы. Это умение ранее считалось прерогативой человеческого интеллекта, недоступной машине. Удивление подкреплялось еще и тем, что было совершенно непонятно, как персептрон это делает: доска с фотоэлементами, имитирующая сетчатку глаза, случайные их связи с *A*-слоем, *B*-слой, сумматор, формальный нейрон, многочисленные обратные связи – все это как-то связано между собой и как-то функционирует согласно непостижимым принципам нейродинамики. Возможно, если бы не реальное устройство, а просто его описание и те же рассуждения о его работе, никто бы на них не обратил внимание. Но было устройство и оно работало, распознавало образы, обучалось этому распознаванию.

Розенблат написал толстую книгу «Принципы нейродинамики» [44], она была переведена на русский язык. Скорее всего, мало кто ее читал, едва ли она была понятна кому-нибудь, кроме автора. Но человеческая мысль работала. Принципы были поняты. Они оказались в потрясающе простой, далекой от нейрофизиологии, математической модели пространства признаков и образа, как области в нем, модели, основанной на абстрактном фантастическом понятии многомерного пространства, придуманном необузданным воображением математиков. Таинственность, запутанность, непостижимость исчезли. Все стало очень просто, даже тривиально, и эта мощь простоты привела к бурному развитию нового раздела кибернетики – «распознаванию образов» – и его разнообразным, многочисленным применениям и использованиям. Сейчас робот может играть на фортепиано с нот, поставленных перед ним, станок – изготавливать детали по их чертежам, компьютер – печатать статью или книгу с рукописи или голоса, ставить диагноз заболевания, обнаруживать неисправности машин, находить месторождения полезных ископаемых и многое, многое другое.

Вернемся к персептрону. Я не буду вас затруднять детальным описанием его устройства, хотя, может быть, это было бы полезно. Я попытаюсь вам рассказать о его функционировании с точки зрения сегодняшнего его понимания.

С помощью табло 100×100 фотоэлементов перцептрон воспринимал зрительный образ, как 10^4 сигналов $x_1, x_2, \dots, x_{10000}$, образующих вектор признаков или описание x . На каждый входной сигнал x перцептрон отвечал: да или нет, $+1$ или -1 . То, как он отвечает на данный входной сигнал x в каждом конкретном случае, зависит от его внутреннего состояния, характеризуемого величинами параметров $\gamma_1, \gamma_2, \dots, \gamma_m$, составляющих вектор γ . Эти параметры γ могут меняться под воздействием учителя, который в каждом случае поначалу указывает перцептрону, когда он отвечает верно и когда неверно. Предполагается, что учитель все показываемые перцептрону объекты делит на два класса A и B и хочет этому разделению научить перцептрон, так чтобы на объект образа (класса) A он отвечал $+1$, а на объект образа B отвечал -1 . Этого учитель хочет достигнуть, обучая перцептрон на примерах показа разных образов и вызывая соответствующие изменения его внутреннего состояния γ . Учитель надеялся, что после обучения перцептрон будет верно (так же, как учитель) отвечать на новые показы, которых, возможно, не было при обучении.

Согласно этому описанию, внутреннее устройство перцептрона, каким бы оно ни было, может быть представлено моделью

$$y = f(x, \gamma), \quad (26.1)$$

в которой $f(x, \gamma)$ – функция двух векторов x и γ , могущая принимать только два значения $+1$ и -1 . В простейшем варианте перцептрона Розенблата эта функция имеет вид:

$$f(x, \gamma) = \text{sign} \sum_{s=1}^m \gamma_s \varphi_s(x_1, \dots, x_n) = \text{sign} \gamma \varphi(x). \quad (26.2)$$

Это только модель ответа перцептрона на сигнал x , нужно еще описать, как происходит процесс обучения перцептрона.

В общем виде обучение можно понимать, как использование информации учителя для определения вектора γ состояния перцептрона, т.е.

$$\gamma = g(x^1, z^1; x^2, z^2; \dots, x^p, z^p), \quad (26.3)$$

где x^i, z^i – описание i -го показываемого объекта и указание учителя, к какому образу он принадлежит: к образу A , если $z = +1$, и B , если $z = -1$. Последовательность $x^1, z^1; \dots; x^p, z^p$ – это обучающая информация учителя. Перцептрон ее использует, определяя по ней свое состояние γ . При новом показе ответ перцептрона (26.1) на показ объекта с описанием x определяется x и состоянием γ . Ответ z учителя определяется какой-то нам неизвестной функцией $f_{\text{учит}}(x) = z$. Так что процесс обучения можно видеть в том, чтобы приблизить функцию перцептрона (26.1) к неизвестной функции учителя $f_{\text{учит}}(x)$ по обучающей информации $I(x^1, z^1; \dots; x^p, z^p)$. После этого пояснения перейдем к описанию алгоритма обучения (26.3) для простейшего варианта перцептрона.

Пусть X – множество всевозможных объектов x . Их конечное или бесконечное число. Каждый объект $x \in X$ учитель относит к одному из классов A или B . Эту способность учителя можно представить некоторой функцией $f_{\text{учит}}(x)$, принимающей для любого x только одно из двух значений: $+1$ и -1 . Задача обучения перцептрона в идеале состоит в том, чтобы ученик-перцептрон делал то же, что и учитель, т.е. чтобы при любом $x \in X$

$$y = f(x, \gamma) = \text{sign } \gamma\varphi(x) = f_{\text{учит}}(x). \quad (26.4)$$

Ясно, что для принципиальной возможности обучения необходимо существование такого числа $\gamma = \gamma^*$, для которого имеет место (26.4) при всех $x \in X$. Далее возникает вопрос, каким должно быть значение γ из (26.3), чтобы найти такое γ^* при условии, что оно существует? Напомним, что в рассматриваемом нами случае функция $f(x, \gamma)$ определяется формулой (26.2).

Розенблат придумал следующий алгоритм подбора требуемого γ^* в процессе последовательных показов объектов x^1, x^2, \dots персептрону и указаний учителя $f_{\text{учит}}(x^1), \dots, f_{\text{учит}}(x^2), \dots$, обнаруживающих, верно или неверно ответил персептрон.

Возьмем произвольное γ . Если персептрон ответил верно, то это подтверждает учитель, и персептрон, удовлетворенный, ничего не меняет, т.е. оставляет γ без изменения. Если же ответ персептрона не совпал с ответом учителя, т.е. если он неверен, то происходит изменение γ , приближающее персептрон к верному ответу. А именно, пусть учитель ответил +1, а персептрон -1, тогда

$$f(x, \gamma) = \text{sign } \gamma\varphi(x) = -1,$$

и для того чтобы приблизить персептрон к верному ответу +1, можно к вектору γ добавить вектор $\varepsilon\varphi(x)$ ($\varepsilon > 0$), т.е. заменить γ на $\bar{\gamma}$ согласно формуле

$$\bar{\gamma} = \gamma + \varepsilon\varphi(x).$$

При этом

$$\bar{\gamma}\varphi(x) = \gamma\varphi(x) + \varepsilon\varphi^2(x) \quad (\varepsilon > 0),$$

следовательно, $f(x, \bar{\gamma}) > f(x, \gamma)$ и тем самым персептрон приблизился к верному ответу.

Пусть теперь, наоборот, учитель ответил -1, а персептрон +1. Тогда величину $\gamma\varphi(x)$ следует уменьшить, т.е. γ заменить на $\bar{\gamma}$ согласно формуле

$$\bar{\gamma} = \gamma - \varepsilon\varphi(x),$$

поскольку при этом $f(x, \gamma) = \gamma\varphi(x)$ уменьшается на $-\varepsilon\varphi^2(x)$.

Пусть ответ учителя $f_{\text{учит}}(x)$ и ответ персептрона не совпадают, тогда, согласно сказанному, γ меняется в соответствии с формулой

$$\bar{\gamma} = \gamma + \varepsilon f_{\text{учит}}(x)\varphi(x). \quad (26.5)$$

Это изменение происходит только при ошибке персептрона. В случае же, когда персептрон отвечает правильно, вместо (26.5) имеет место

$$\bar{\gamma} = \gamma. \quad (26.6)$$

Не представляет труда выражения (26.5) и (26.6) записать в виде одной формулы

$$\bar{\gamma} = \begin{cases} \gamma & \text{при } f_{\text{учит}}(x) = f, \\ \gamma + \varepsilon f_{\text{учит}}(x)\varphi(x) & \text{при } f_{\text{учит}}(x) \neq f \end{cases} \quad (26.7)$$

или

$$\bar{\gamma} = \gamma - 0,5\varepsilon(f - f_{\text{учит}})\varphi(x),$$

где f – ответ персептрона, определяемый формулой (26.2), а $f_{\text{учит}}$ – ответ учителя на показ одного и того же объекта $x \in X$.

Прежде чем перейти к изучению алгоритма обучения (26.7), кратко и схематично опишем, как сказанное было реализовано в персептроне Розенблата.

Как уже говорилось, Розенблат, конструируя персептрон, исходил из представлений о структуре зрительного рецептора и последующей обработке восприятия зрительного образа сетчаткой глаза и мозгом. Эту обработку он предполагал послойной: слой рецепторов, слой формирования функций $\varphi_s(x_1, \dots, x_n)$, слой умножения функций φ_s на коэффициенты γ_s и, наконец, нейрон, реализующий пороговую функцию sign . В соответствии с этим его схема имела вид, показанный на рис. 26.1.

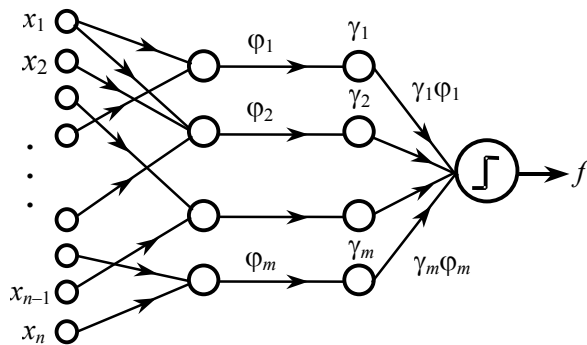


Рис. 26.1. Схема трехслойного персептрона с формальным нейроном (пороговым элементом) на выходе. Обратные связи изменения фазовых переменных $\gamma_1, \dots, \gamma_m$ в третьем слое не показаны

Механизм обучения был реализован путем сопоставления ответа нейрона с ответом учителя, по результатам которого происходила корректировка с помощью обратных связей каждого из множителей γ_s согласно формуле (26.7). Таково схематическое описание работы персептрона.

Помимо функционального и схематического описаний возможно геометрическое описание. Поступление образа x в персептрон можно интерпретировать как возникновение точки x в многомерном пространстве признаков R_n . В этом пространстве функцию (26.2), реализуемую персептроном, можно представить геометрически как то, с какой стороны от поверхности

$$\gamma\varphi(x) = 0 \tag{26.8}$$

находится точка x . С одной стороны этой поверхности $\gamma\varphi(x) > 0$, и это соответствует ответу +1, с другой ее стороны $\gamma\varphi(x) < 0$, что отвечает ответу -1. Множество точек X делится на точки класса A и точки класса B , которые должны лежать по разные стороны поверхности (26.8).

В соответствии со сказанным множество точек x , отвечающих образу A , можно трактовать как модель образа A ; аналогично множество точек образа B – как модель образа B . Обучение при этом трактуется как построение поверхности (26.8), разделяющей множества точек образов A и B , согласно формуле

$$x \in \begin{cases} A, & \text{если } \gamma\varphi(x) > 0, \\ B, & \text{если } \gamma\varphi(x) < 0. \end{cases}$$

Итак, интеллектуальная задача распознавания образов получила геометрическую интерпретацию в построении поверхности, разделяющей множества точек – моделей – образов A и B .

Такая геометрическая интерпретация распознавания зрительных образов возымела огромные последствия. Образы начали воспринимать не только зритель-

ные, а любые, описываемые какими-то числовыми признаками x_1, x_2, \dots, x_n . А это сделало возможным распознавание почерков, различных звуковых сигналов, нефтяных месторождений, чтение текстов, узнавание голоса человека и произносимых им слов, медицинской и технической диагностики, и многого, многого другого.

Наконец, немаловажно и интересно, что персептрон – это динамическая система, что процесс его обучения, как и само узнавание, – динамический процесс.

Состояние персептрона – это вектор γ размерности m . Оператор этой динамической системы дается формулой (26.7), которая определяет изменение его состояния γ . Ответ персептрона на внешнее воздействие x определяется формулой (26.2). Обученность персептрона – это достижение им равновесного состояния, в котором состояние γ уже не меняется при любом показе x .

Перейдем к обоснованию возможности его обучения согласно алгоритму (26.7). При этом мы будем трактовать персептрон как динамическую систему. Вектор состояния γ будем интерпретировать как точку в многомерном фазовом пространстве Γ_m . Формула (26.7) определяет закон изменения состояния и может трактоваться как оператор динамической системы. Состояния γ^* , при которых персептрон обучен и всегда отвечает одинаково с учителем, согласно (26.7), не меняются и поэтому являются его равновесными состояниями.

Нетрудно видеть, что множество состояний равновесия, если они существуют, образуют выпуклый конус. Действительно, равновесие γ^* определяется требованием

$$\text{sign } \gamma^* \varphi(x) = f_{\text{учит}}(x) \quad (26.9)$$

для всех $x \in X$. Но если (26.9) выполнено для некоторого γ^* , то оно выполнено и для $\lambda \gamma^*$ при любом $\lambda > 0$. Аналогично, если γ_1^* и γ_2^* – равновесия, то и $\lambda_1 \gamma_1^* + \lambda_2 \gamma_2^*$ – тоже равновесие при λ_1 и λ_2 положительных (в частности, при $\lambda_2 = 1 - \lambda_1$ и $0 \leq \lambda_1 \leq 1$).

Теперь докажем, что если персептрон имеет конус равновесий, содержащий внутреннюю точку γ^* , то он может совершить не более некоторого конечного числа ошибок, после чего уже их не совершает. Эта ошеломляющая теорема Новикова доказывается сравнительно просто. Смысл ее состоит в том, что любая дискретная фазовая траектория персептрона при любой заданной последовательности показов не может содержать более некоторого конечного числа точек. Из этого еще не следует, что она обязательно окончится равновесием, просто после некоторого конечного числа ошибок персептрон на любой заданной последовательности обязательно перестает ошибаться. Но если мы будем несколько раз показывать последовательно все точки множества X , то фазовая траектория обязательно закончится состоянием равновесия и персептрон научится правильно различать все точки множества X .

Доказательство этой замечательной теоремы Новикова, как уже говорилось, несложно.

Пусть для $x \in X$ $0 < r \leq \|\varphi(x)\| \leq R < \infty$ и пусть γ^* – внутренняя точка множества равновесий, которая соответствует обученности персептрона, тогда

$$|\gamma^* \varphi(x)| \geq \rho \|\gamma^*\| \|\varphi(x)\|, \quad \rho > 0, \quad (26.10)$$

вне зависимости от длины вектора γ^* , которую можно выбрать произвольно.

Действительно, если γ^* – внутренняя точка конуса равновесий, то при некотором $\rho > 0$ для любого вектора a , для которого $\|a\| \leq \rho$, для всех $x \in X$ имеет место неравенство

$$|(\gamma^* + \alpha \|\gamma^*\|)\varphi(x)| > 0$$

или, принимая $\alpha = v\varphi(x)/\|\varphi(x)\|$, неравенство

$$|\gamma^*\varphi(x) + v\|\gamma^*\|\|\varphi(x)\| > 0$$

при любом $|v| \leq \rho$, из чего следует (26.10). Пусть еще γ^0 – произвольное начальное состояние персептрона; $\gamma^1, \gamma^2, \gamma^3, \dots$ – последовательность состояний, возникающая в результате его обучения показами некоторых произвольных точек x^1, x^2, x^3, \dots множества X . Рассмотрим, как меняется расстояние $V(\gamma^s) = (\gamma^s - \gamma^*)^2$ точки γ^s от точки γ^* с ростом s . Непосредственно находим, что

$$\begin{aligned} V(\gamma^{s+1}) - V(\gamma^s) &= (\gamma^s + \varepsilon f_{\text{учит}}\varphi(x) - \gamma^*)^2 - (\gamma^s - \gamma^*)^2 = \\ &= 2\varepsilon f_{\text{учит}}\gamma^s\varphi(x) - 2\varepsilon f_{\text{учит}}\gamma^*\varphi(x) + \varepsilon^2 f_{\text{учит}}^2\varphi^2(x). \end{aligned}$$

Так как

$$f_{\text{учит}}\gamma^s\varphi(x) < 0, \quad f_{\text{учит}}\gamma^*\varphi(x) > 0,$$

то, учитывая (26.10),

$$V(\gamma^{s+1}) - V(\gamma^s) < -2\varepsilon|\gamma^*\varphi(x)| + \varepsilon^2\varphi^2(x) \leq -2\varepsilon\rho\|\gamma^*\|\|\varphi(x)\| + \varepsilon^2\varphi^2(x).$$

Пользуясь произвольностью выбора длины вектора γ^* , выберем ее так, чтобы

$$-2\varepsilon\rho\|\gamma^*\|\|\varphi(x)\| + \varepsilon^2\varphi^2(x) \leq -q < 0.$$

Тогда с каждым переходом от γ^s к γ^{s+1} расстояние до γ^* будет уменьшаться не менее чем на $q > 0$, и поэтому таких переходов не может быть больше, чем $N = V(\gamma)/q$. Утверждение доказано.

Найдем явное выражение оценки N верхней границы числа ошибок персептрона. При $\|\gamma^*\| = \varepsilon R/\rho$ и $\gamma^0 = 0$,

$$N = \frac{\varepsilon^2 R^2}{\rho^2} \max \left[2\varepsilon\rho \frac{\varepsilon R}{\rho} \|\varphi(x)\| - \varepsilon^2 \varphi^2(x) \right]^{-1} = \frac{R^2}{\rho^2} \max_{r \leq \xi \leq R} [2R\xi - \xi^2]^{-1} = \frac{1}{\rho^2},$$

что весьма естественно, так как вынужденное уменьшение ρ означает все большую и большую трудность разделения подмножеств A и B множества X .

Таким образом, персептрон совершает не более конечного числа ошибок, если он в принципе способен к различению образов A и B . Это так даже при бесконечности множества X . Если множество X конечно, то при повторном показе персептрон в конце концов обязательно научится правильному распознаванию заведомо за число повторных показов, меньшее N . Напротив, при бесконечности множества X никакое сколь угодно длительное обучение не может гарантировать его обученности. С точки зрения ускорения обучаемости желательно, чтобы персептрон при обучении совершал возможно большее число ошибок и исчерпал допустимое их число. Если это число исчерпано, то персептрон правильно классифицирует все

остальные точки, несмотря на то, что учитель ему их не показывал. В этой экстраполяционной прогностической возможности состоит величайшая его ценность и практическое значение. Она, в частности, позволяет по показу нескольких больных с той или иной болезнью затем ставить диагноз этого заболевания для новых больных. По нескольким показам буквы A затем узнавать ее при других написаниях и т.д.

Выше фактически одновременно с описанием модели персептрона была описана и общая модель обучения распознаванию двух классов объектов (то, что только двух классов, не является существенным ограничением, так как путем дихотомии можно осуществить разделение на любое число классов). Эта модель состоит из решающего правила $f(x, \gamma)$ и алгоритма обучения, алгоритма изменения состояния γ , который записывается в виде (26.3). Сегодняшние знания не позволяют указать на все возможные, приводящие к цели, алгоритмы для любого заданного решающего правила $f(x, \gamma)$. Нельзя указать и лучшее из решающих правил. Но уже существует довольно много разных видов алгоритмов обучения. Мы укажем на один общий способ их построения. Он основан на методе минимизации стохастического функционала и получил название метода стохастической аппроксимации.

Пусть имеется некоторый функционал $G(x, \lambda)$, зависящий от случайного вектора $x \in X$ и вектора параметров λ . Случайный вектор имеет некоторую функцию распределения $p(x)$. Метод стохастической аппроксимации указывает процедуру последовательных приближений к значению λ^* , которое минимизирует математическое ожидание функционала

$$MG(x, \lambda) = \int_x G(x, \lambda) p(x) dx \quad (26.11)$$

по параметру λ . Последовательные приближения строятся согласно формуле

$$\lambda^{p+1} = \lambda^p - \varepsilon^p \nabla_{\lambda} G(x^{p+1}, \lambda^p), \quad (26.12)$$

где точки x^1, x^2, \dots возникают случайно в соответствии с плотностью вероятностей $p(x)$.

Не будем формулировать условий сходимости последовательных приближений $\lambda^1, \lambda^2, \dots$ к минимуму функционала (26.11). Это при желании вы узнаете из соответствующих спецкурсов или книг. Ограничимся тем, что покажем, как этим общим способом можно прийти к описанному выше алгоритму обучения и многим другим, как показал Я.З. Цыпкин.

Ясно, что цель обучения можно сформулировать как стремление к возможно меньшему числу ошибок, т.е. как отыскание такого λ^* , при котором решающее правило $f(x, \lambda)$ допускает минимум ошибок. Введем вместо числа ошибок функцию штрафа совершенной ошибки $G(x, \lambda)$ и потребуем, чтобы математическое ожидание этой функции штрафа было минимальным. Отыскание требуемого λ можно осуществить согласно (26.12).

Примем

$$f(x, \lambda) = \text{sign} \lambda \varphi(x),$$

$$G(x, \lambda) = (f(x, \lambda) - f_{\text{учит}}(x)) \lambda \varphi(x)$$

и согласно (26.12) найдем

$$\begin{aligned}\lambda^{p+1} &= \lambda^p - \varepsilon^p \nabla_{\lambda} (f(x, \lambda) - f_{\text{учит}}(x)) \lambda \varphi(x) = \\ &= \lambda^p - \varepsilon^p (f(x, \lambda) - f_{\text{учит}}(x)) \varphi(x),\end{aligned}$$

что, как нетрудно видеть, совпадает с (26.7) при $\varepsilon^p = 0,5\varepsilon$.

Функция штрафа $G(x, \lambda)$ была выбрана так, что при верном ответе ($f = f_{\text{учит}}$) она обращается в нуль, а при неверном ($f \neq f_{\text{учит}}$) оказывается равной $|\lambda \varphi(x)|$. Выбирая другие функции штрафа, можно найти другие алгоритмы обучения.

27. ЗАКОНЫ КЕПЛЕРА И ПРОБЛЕМА ДВУХ ТЕЛ, РЕШЕННАЯ НЬЮТОНОМ

Сегодня я начну рассказ о динамической системе, описание и математическое моделирование которой прошло через всю историю человечества. От египтян, вавилонян, арабов и древних греков к европейцам и всему современному человечеству протянулась извилистая нить этих исследований через двадцать пять веков, отражая изменения взглядов людей на окружающий мир и свое место в нем, на возможности использования их в своей жизни. Началась с примитивных описаний непосредственно созерцаемого, прошла через постижения величественного замысла богов и созданной ими гармонии мира, через описание этой гармонии совершенными идеальными математическими образами (кругами, сферами, правильными многогранниками, золотым сечением и др.) к современному лаконичному, сугубо математизированному, абстрактному моделированию и исследованию, к пониманию единства мира земного и небесного и их законов, выражаемых языком математики. Эта система – звездное небо с Солнцем, Луной и планетами, и Землей под ними.

День сменяет ночь и снова наступает день, так из поколения в поколение, испокон веков. Зиму сменяет весна, весну – лето, лето – осень, снова зима и так все время. Жизнь человека непосредственно связана с этой вечной сменой времен года, дня и ночи. Луна то появляется, то исчезает, меняются ее фазы, причудливо движутся по небесному своду бродяги-планеты. Иногда вдруг затмевается Луна или Солнце. Загадки этих метаморфоз не могли не волновать человека, ему казалось, что его жизнь целиком в их власти, во власти величественного божественного неба и его светил. Мы все со школьной скамьи знаем о геоцентрической системе мира древних греков и средневековья, об описательной кинематической модели Птолемея, о сменившей ее гелиоцентрической системе Коперника, о законах Кеплера и всемирном законе тяготения. Но как из этих законов Ньютона следуют три закона Кеплера, осталось тайной. Вот эту тайну мы развеем. Именно о ее раскрытии будет дальнейший рассказ.

Законы Кеплера несравнимо проще описывают движение небесных светил и позволяют предсказывать их положения на небе, чем это делала предшествующая очень сложная кинематическая модель Птолемея и более простая Коперника, но такая же в основном принципе представления движения планет в виде суперпозиции равномерных круговых вращений. Если системы Птолемея содержали 77 кругов, то система Коперника уменьшила их число до 34. Вместе с тем системы Птолемея и Коперника достигли очень большой точности предсказания порядка угловых минут, но вычисления были очень сложными. Кеплер все необычайно упростил, но описание по-прежнему оставалось кинематическим. Три закона Кеплера не были связаны между собой, не имели единого объяснения. Единое объяснение им дал Ньютон. Единое не только для них, но и для всех механических явлений на небе и на земле. Все стало следствием одних и тех же законов, законов механического движения.

Напомним законы Кеплера.

Первый закон говорит о том, что орбита каждой планеты – это эллипс, в одном из фокусов которого расположено Солнце.

Второй закон описывает, как, с какой скоростью движется каждая планета по своему эллипсу. Скорость ее движения переменная, но постоянна так называемая секторная скорость – скорость, с которой описывает площадь радиус-вектор, соединяющий Солнце и планету.

Третий закон, как думал Кеплер, – это общая гармония небесных сфер, выражаемая постоянством отношения квадрата периода обращения планеты к кубу большой полуоси ее эллипса.

В целом три закона Кеплера дают полное кинематическое описание закономерностей движения планет Солнечной системы. Но нет ли более общих обоснований этим закономерностям, не есть ли они следствие каких-то более общих законов природы? Чтобы это выяснить, Ньютон преодолел громадные трудности и титаническими усилиями человеческого гения заложил основы ряда новых направлений науки: механики, теории тяготения, дифференциального и интегрального исчисления. Более того, само вычисление, как говорил Ньютон, законов Кеплера было для того времени невообразимо сложной задачей, но Ньютон их «вычислил».

Мы сейчас в осовремененном виде повторим эти вычисления и предположим им выдержку из сочинений Ньютона, которая говорит о близости нам его методологии и о месте нашего рассказа в его общей программе постижения мира. «Вся трудность физики... состоит в том, чтобы по явлениям движения распознать силы природы, а затем по этим силам объяснить остальные явления».

В аннотации загадочного тяготения, действующего через пустоту, он писал: «... гипотез же я не изобретаю. Довольно того, что тяготение на самом деле существует и действует согласно изложенным нами законам и вполне достаточно для объяснения всех движений небесных тел и моря».

Может быть, следует напомнить, что в изложенном подходе великий Ньютон следовал великому Галилею.

Как и до сих пор, нам надлежит построить математическую модель и ее исследовать. Объектом моделирования являются две точечные массы, которые притягиваются друг к другу с силой, по величине равной $\gamma mM/\rho^2$, где m и M – массы притягивающихся тел, ρ – расстояние между ними и γ – экспериментально найденный коэффициент, константа тяготения. Сила притяжения направлена по линии, соединяющей два тела.

Пусть \vec{r} и \vec{R} – радиусы-векторы этих масс в некоторой инерциальной системе координат (рис. 27.1), тогда, согласно второму закону Ньютона и закону всемирного тяготения, имеем:

$$\begin{aligned} M\ddot{\vec{R}} &= \gamma \frac{Mm}{(\vec{R} - \vec{r})^2} \frac{\vec{r} - \vec{R}}{|\vec{r} - \vec{R}|}, \\ m\ddot{\vec{r}} &= -\gamma \frac{Mm}{(\vec{R} - \vec{r})^2} \frac{\vec{r} - \vec{R}}{|\vec{r} - \vec{R}|}. \end{aligned} \quad (27.1)$$

В этих дифференциальных уравнениях $(\bar{r} - \bar{R})|\bar{r} - \bar{R}|^{-1}$ – единичный вектор, направленный от тела M к телу m . Если их написать в скалярной форме, как это, естественно, делал Ньютон, то они составят систему шести уравнений второго порядка, т.е. систему двенадцатого порядка.

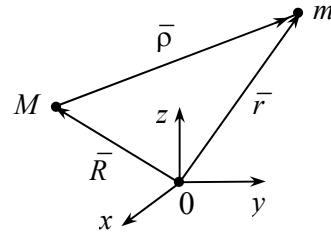


Рис. 27.1. Графическая иллюстрация к задаче двух тел – притягивающихся друг к другу масс m и M

Перейдем к вычислениям с этой системой. Они необычайно облегчаются векторной записью (27.1). Первое – сложим друг с другом эти уравнения и найдем, что

$$M\ddot{\bar{R}} + m\ddot{\bar{r}} = 0$$

или

$$\frac{d}{dt} \frac{M\dot{\bar{R}} + m\dot{\bar{r}}}{M + m} = \bar{V}_c = \text{const.} \tag{27.2}$$

Это означает, что центр масс двух тел движется равномерно и прямолинейно с постоянной скоростью \bar{V}_c .

Запишем уравнения (27.1) в виде:

$$\ddot{\bar{R}} = \gamma \frac{m(\bar{r} - \bar{R})}{|\bar{r} - \bar{R}|^3}, \quad \ddot{\bar{r}} = -\gamma \frac{M(\bar{r} - \bar{R})}{|\bar{r} - \bar{R}|^3}$$

и вычтем одно из другого, найдем, что

$$\ddot{\bar{\rho}} = -\gamma(M + m) \frac{\bar{\rho}}{|\bar{\rho}|^3}, \tag{27.3}$$

где $\bar{\rho} = \bar{r} - \bar{R}$ – радиус-вектор массы m (планеты) по отношению к массе M (Солнцу), т.е. это дифференциальные уравнения их относительного движения. Как движется их центр масс, мы уже знаем.

Умножим векторно обе части уравнения (27.3) на $\bar{\rho}$ и придем к тому, что

$$\ddot{\bar{\rho}} \times \bar{\rho} = 0$$

или

$$\begin{aligned} \frac{d}{dt} (\dot{\bar{\rho}} \times \bar{\rho}) &= 0, \\ \dot{\bar{\rho}} \times \bar{\rho} &= \bar{g} = \text{const.} \end{aligned} \tag{27.4}$$

Выясним смысл полученного соотношения (27.4). Из него следует, что обе массы M и m все время находятся в плоскости, проходящей через центр масс, перпендикулярно вектору \bar{g} . Действительно, радиусы-векторы $\bar{\rho}_m$ и $\bar{\rho}_M$ масс m и M (планеты и Солнца) по отношению к их центру масс равны соответственно:

$$\begin{aligned}\bar{\rho}_m &= \bar{r} - \frac{M\bar{R} + m\bar{r}}{M+m} = -\frac{m}{M+m}\bar{\rho}, \\ \bar{\rho}_M &= \bar{R} - \frac{M\bar{R} + m\bar{r}}{M+m} = \frac{M}{M+m}\bar{\rho},\end{aligned}$$

где вектор $\bar{\rho}$, согласно (27.4), все время перпендикулярен вектору \bar{g} .

Далее распишем (27.4) в скалярной форме, выбрав начало координат в начале вектора $\bar{\rho}$, т.е. в центре Солнца, и направим ось Oz вдоль вектора \bar{g} длины g . Имеем

$$\dot{\bar{\rho}} \times \bar{\rho} = \begin{vmatrix} \bar{i} & \bar{j} & \bar{k} \\ \dot{x} & \dot{y} & \dot{z} \\ x & y & z \end{vmatrix} = g \bar{k}, \quad (27.5)$$

где \bar{i} , \bar{j} , \bar{k} – орты осей Ox , Oy и Oz ; x , y , z – проекции вектора $\bar{\rho}$ на эти орты, причем векторы $\bar{\rho}$ и $\dot{\bar{\rho}}$ ортогональны орту \bar{k} согласно его выбору, и поэтому $z = \dot{z} = 0$. В силу этого из (27.5) следует только одно соотношение

$$\dot{x}y - y\dot{x} = g. \quad (27.6)$$

Перейдем к полярным координатам $x = \rho \cos \varphi$, $y = \rho \sin \varphi$ и приведем (27.6) к виду

$$\rho^2 \dot{\varphi} = -g. \quad (27.7)$$

Нетрудно заметить, что полученное соотношение (27.7) выражает второй закон Кеплера. Действительно, приращение площади $d\sigma$, описываемой вектором $\bar{\rho}$, равно

$$d\sigma = \frac{1}{2} \rho^2 d\varphi,$$

и поэтому

$$d\sigma = \frac{1}{2} \rho^2 \dot{\varphi} dt = -\frac{g}{2} dt$$

или

$$\frac{d\sigma}{dt} = -\frac{g}{2} = \text{const},$$

что и требовалось показать (рис. 27.2).

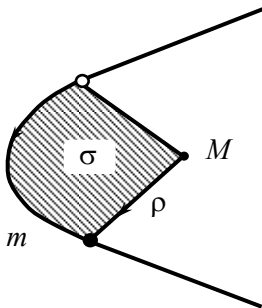


Рис. 27.2. Иллюстрация ко второму закону Кеплера: σ – площадь, описанная вектором $\bar{\rho}$, массой m , движущейся вокруг Солнца

Перейдем теперь к поиску орбиты планеты относительно Солнца.

В нашем распоряжении уже имеется одно уравнение (27.7), которое запишем в виде:

$$\frac{d\varphi}{dt} = -\frac{g}{\rho^2}. \quad (27.8)$$

Второе уравнение для $d\rho/dt$ можно получить из векторного уравнения (27.3), умножив его скалярно на вектор скорости $\dot{\vec{\rho}}$:

$$\ddot{\vec{\rho}} \dot{\vec{\rho}} = -\gamma(M+m) \frac{\vec{\rho} \dot{\vec{\rho}}}{\rho^3}$$

или

$$\begin{aligned} \frac{d}{dt} \left(\frac{\dot{\vec{\rho}}^2}{2} \right) &= -\gamma(M+m) \frac{1}{\rho^3} \frac{d}{dt} \frac{\vec{\rho}^2}{2} = -\gamma(M+m) \frac{1}{\rho^3} \frac{d}{dt} \left(\frac{\rho^2}{2} \right) = \\ &= -\gamma(M+m) \frac{\dot{\rho}}{\rho^2} = \gamma(M+m) \frac{d}{dt} \left(\frac{1}{\rho} \right), \end{aligned}$$

откуда, интегрируя, находим, что

$$\frac{\dot{\vec{\rho}}^2}{2} - \frac{\gamma(M+m)}{\rho} = h = \text{const}. \quad (27.9)$$

Выразим в (27.9) $\dot{\vec{\rho}}^2$ через полярные координаты ρ и φ :

$$\dot{\vec{\rho}}^2 = (\dot{x}^2 + \dot{y}^2) = (\dot{\rho} \cos \varphi - \rho \dot{\varphi} \sin \varphi)^2 + (\dot{\rho} \sin \varphi + \rho \dot{\varphi} \cos \varphi)^2 = \dot{\rho}^2 + \rho^2 \dot{\varphi}^2.$$

После подстановки в (27.9) имеем

$$\dot{\rho}^2 + \rho^2 \dot{\varphi}^2 - \frac{2\gamma(M+m)}{\rho} = 2h$$

или, выражая $\dot{\varphi}$ через ρ , согласно (27.8), получим

$$\dot{\rho}^2 + \frac{g^2}{\rho^2} - \frac{2\gamma(M+m)}{\rho} = 2h.$$

Разрешая последнее уравнение относительно $\dot{\rho}$, находим, что

$$\frac{d\rho}{dt} = \sqrt{2h + \frac{2\gamma(M+m)}{\rho} - \frac{g^2}{\rho^2}}. \quad (27.10)$$

Найденное уравнение (27.10) вместе с полученным ранее уравнением (27.8) представляют собою дифференциальные уравнения относительного движения планеты в плоскости OXY . Если из этих уравнений исключить время t , разделив почленно уравнения (27.10) на уравнение (27.8), то придем к одному уравнению орбиты планеты

$$\frac{d\rho}{d\varphi} = -\frac{\rho^2}{g} \sqrt{2h + \frac{2\gamma(M+m)}{\rho} - \frac{g^2}{\rho^2}}.$$

Проинтегрировать это дифференциальное уравнение первого порядка и найти уравнение орбиты можно следующим образом. Сделаем замену переменных, введя новую переменную $u = 1/\rho$, в результате чего придем к уравнению

$$\frac{du}{d\varphi} = \frac{1}{g} \sqrt{2h + 2\gamma(M + m)u - g^2 u^2} = \frac{1}{g} \sqrt{f}$$

и повысим его порядок, продифференцировав по φ :

$$\begin{aligned} \frac{d^2 u}{d\varphi^2} &= -\frac{1}{g} \frac{1}{2} \frac{1}{\sqrt{f}} \frac{df}{du} \frac{du}{d\varphi} = -\frac{1}{2g} \frac{1}{\sqrt{f}} \frac{df}{du} \frac{1}{g} \sqrt{f} = \\ &= -\frac{1}{2g^2} \frac{df}{du} = -\frac{1}{2g^2} (-2\gamma(M + m) + 2g^2 u) \end{aligned}$$

или

$$\frac{d^2 u}{d\varphi^2} + u = \frac{\mu}{g^2} \quad (\mu = \gamma(M + m)). \quad (27.11)$$

Мы пришли к хорошо нам известному неоднородному дифференциальному уравнению гармонического осциллятора, общее решение которого нам хорошо известно и пишется в виде:

$$u = \frac{\mu}{g^2} + c_1 \cos \varphi + c_2 \sin \varphi = \frac{\mu}{g^2} \{1 + e \cos(\varphi - \varphi_0)\},$$

где e и φ_0 – константы интегрирования. Постоянная φ_0 имеет смысл начала отсчета переменной φ на орбите и может быть выбрана равной 0 или π , так чтобы постоянная e была неотрицательной. После этого уравнение планетной орбиты примет вид:

$$\rho = \frac{g^2}{\mu(1 + e \cos \varphi)}, \quad e \geq 0. \quad (27.12)$$

Эта орбита при $e = 0$ – окружность, при $0 < e < 1$ – эллипс, при $e = 1$ – парабола и при $e > 1$ – гипербола (рис. 27.3). В точке $\rho = 0$ расположен фокус. Ближайшей к фокусу точкой орбиты является точка $\varphi = 0$, $\rho = g^2 / \mu(1 + e)$. Параметр e имеет смысл эксцентриситета¹.

Между величинами a и b большой и малой полуосей эллипса и расстоянием $c = \sqrt{a^2 - b^2}$ от центра эллипса до его фокусов F_1 и F_2 и параметрами формулы (27.12) имеют место соотношения:

$$e = \frac{c}{a}, \quad a - c = \frac{g^2}{\mu(1 + e)}, \quad a + c = \frac{g^2}{\mu(1 - e)}. \quad (27.13)$$

¹ Это следует непосредственно из определений перечисленных конических сечений. Так, для эллипса, согласно рис. 27.4 и его определению, $(2a - \rho)^2 = \rho^2 + 4c^2 + 2\rho c \cos \varphi$, откуда $\rho = (a^2 - c^2) / [a(1 + c/a \cos \varphi)]$, что при обозначениях $e = c/a$, $g^2/\mu = a^2 - c^2/a$ совпадает с (27.12).

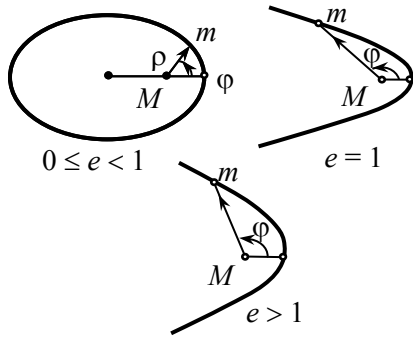


Рис. 27.3. Эллиптическая, параболическая и гиперболическая орбиты движения массы m относительно инерциальной системы с началом координат в центре масс

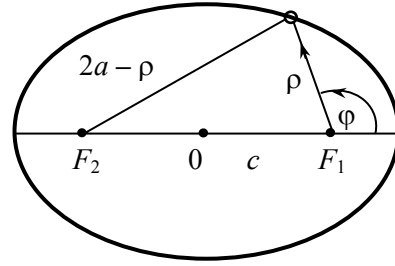


Рис. 27.4. Эллипс с фокусами F_1 и F_2

Перейдем к извлечению из найденных соотношений третьего закона Кеплера. Прежде всего, интегрируя формулу секторной скорости планеты

$$\frac{d\sigma}{dt} = -\frac{g}{2},$$

находим, что

$$\int_0^T \frac{d\sigma}{dt} dt = -\frac{g}{2} T.$$

Используя то, что площадь эллипса равна πab , находим

$$\pi ab = -gT/2.$$

или

$$T = -\frac{2\pi ab}{g}. \tag{27.14}$$

Соотношение (27.14) вместе с соотношениями (27.13) и $c = \sqrt{a^2 - b^2}$ составляет 5 уравнений с шестью величинами a, b, c, e, g, T , что позволяет, в частности, получить соотношение между a и T . Оно может быть получено следующим образом:

$$T = -\frac{2\pi ab}{g} = -\frac{2\pi a}{g} \sqrt{a^2 - c^2} = -\frac{2\pi a^2}{g} \sqrt{1 - e^2}. \tag{27.15}$$

Далее из соотношений (27.13) следует, что

$$a = \frac{g^2}{\mu(1 - e^2)}$$

или

$$\sqrt{1 - e^2} = \frac{-g}{\sqrt{\mu a}}.$$

Подставляя найденное выражение в (27.15), находим, что

$$T = -\frac{2\pi a^2}{g} \left(\frac{-g}{\sqrt{\mu a}} \right) = \frac{2\pi a^{3/2}}{\sqrt{\mu}}$$

или

$$\frac{T^2}{a^3} = \frac{4\pi^2}{\mu}, \quad (27.16)$$

что и является соотношением, выражающим третий закон Кеплера. Величина

$$\mu = \gamma(M + m) = \gamma M \left(1 + \frac{m}{M} \right)$$

несколько различна для разных планет, но эти различия не могли быть замечены во времена Кеплера, так как для солнечной системы m/M очень малая величина.

Применим теперь полученные результаты в отношении движения двух тел, притягивающихся друг к другу по закону всемирного тяготения, к движению тела, запущенного с горы на Земле в горизонтальном направлении со скоростью v , в отсутствие сопротивления воздуха. Прежде всего заметим, что шарообразная Земля при сферической однородности притягивает брошенное тело так же, как материальная точка, имеющая массу Земли и находящаяся в ее центре. Этот нетривиальный факт был установлен Ньютоном. Он же поставил и решил сформулированный выше вопрос о траектории движения запущенного с земной поверхности тела и обнаружил, что при достаточной скорости это тело станет спутником Земли подобно Луне.

Согласно предыдущему, движение будет происходить по кривой, определяемой уравнением (27.12):

$$\rho = \frac{g^2}{\mu(1 + e \cos \varphi)}, \quad (27.17)$$

и в одном из ее фокусов будет находиться центр Земли. При $\varphi = 0$ $\rho = R_3$, где R_3 – радиус Земли или, точнее, расстояние от центра Земли до места, с которого запускается тело, так что

$$R_3 = \frac{g^2}{\mu} \frac{1}{1 + e}. \quad (27.18)$$

Осталось еще учесть, что тело запускается со скоростью v . Из соотношения (27.8) следует, что

$$\dot{\varphi} = -\frac{g}{R_3^2},$$

где

$$\dot{\varphi} = \frac{v}{R_3}.$$

Так что (27.18) можно записать в виде:

$$R_3 = \frac{(vR_3)^2}{\mu} \frac{1}{1 + e},$$

откуда получим

$$1 + e = \frac{v^2 R_3}{\mu}$$

или, пренебрегая массой запущенного тела, запишем

$$e = \frac{v^2 R_3}{\gamma M_3} - 1. \quad (27.19)$$

Соотношение (27.19) позволяет по скорости v найти эксцентриситет e орбиты брошенного тела, что вместе с (27.18), определяющим g^2/μ , определяет параметры уравнения орбиты (27.17).

Формула (27.19) позволяет найти первую и вторую космические скорости v_1 и v_2 . Первая – это минимальная скорость, при которой брошенное тело не падает на Землю и становится ее спутником. Вторая – это минимальная скорость, при которой брошенное тело перестает быть спутником и покидает Землю. Первое имеет место при $e = 0$ и круговой орбите, второе – при $e = 1$, когда орбита становится параболической. Из (27.19) находим, что

$$v_1 = \sqrt{\frac{\gamma M_3}{R_3}} = \sqrt{\frac{\gamma M_3}{R_3^2} R_3} = \sqrt{9,8 \cdot 6 \cdot 10^6} \approx 8 \text{ км/с}, \quad (27.20)$$

$$v_2 = \sqrt{\frac{2\gamma M_3}{R_3}} \approx 11 \text{ км/с}.$$

Эти скорости очень велики; с такой скоростью можно облететь земной шар менее чем за два часа. Но человек их достиг и сделал это только в XX веке с помощью ракеты, о которой у нас была речь ранее. Этих скоростей могут достигать и молекулы газа. На Земле это удается очень редким молекулам, а на Луне, где вторая космическая скорость много меньше, их настолько много, что на ней нет атмосферы. Вторая космическая скорость на Солнце много больше, чем на Земле, но температура на нем настолько велика, что Солнце все время испускает поток заряженных частиц.

Помимо этих очевидных замечаний по поводу полученных нами формул (27.20), укажем еще на два их следствия. Первое относится к тому, когда «звезда» становится черной дырой, которую ничто и даже свет не может покинуть. Черная дыра уже не светится и ее удастся обнаружить лишь по влиянию ее тяготения на ближайшие к ней настоящие видимые звезды. Ответ на поставленный вопрос очень прост: «звезда» будет черной дырой, если ее вторая космическая скорость больше скорости света: быстрее света c ничто перемещаться не может. Таким образом, условие черной дыры имеет вид:

$$\sqrt{\frac{2\gamma M}{R}} > c, \quad (27.21)$$

где M – ее масса, а R – радиус. Чтобы почувствовать, что представляет собою это условие, укажем, что Солнце было бы черной дырой, если бы его размер был менее трех километров, т.е. при гигантской плотности нескольких миллионов тонн в кубическом сантиметре. Эта фантастическая плотность все же возможна, и чер-

ные дыры существуют. Вблизи черной дыры геометрия пространства неевклидова и часы идут медленнее в соответствии с теорией тяготения Эйнштейна, эффекты которой, еле уловимые в поле тяготения Солнца и планет, становятся значительными в гигантских гравитационных полях черных дыр.

Второй вопрос о том, сбегаются под влиянием тяготения небесные тела вселенной или, наоборот, разбегаются? Он уже не так прост, я бы даже сказал, настолько сложен, что ответ не столь уж очевиден и ясен.

Начнем издалека. Еще Ньютону было известно, что сила f притяжения единичной массы на расстоянии r от центра массивного однородного шара радиуса R и массы M определяется формулой

$$f = \begin{cases} \frac{\gamma M}{r^2} & \text{для } r \geq R, \\ \frac{\gamma M}{r^2} \frac{r^3}{R^3} & \text{для } r \leq R, \end{cases} \quad (27.22)$$

т.е. при $r \geq R$ шар можно заменить точечной массой M в его центре, а при $r < R$ наружная часть шара по отношению к притягиваемой единичной массе никакого притяжения не оказывает (все ее притяжения, направленные в разные стороны, уравновешиваются).

Сделанные выше выводы из формулы (27.22) обобщаются на неоднородный шар, у которого плотность ρ зависит только от радиуса удаления от центра. Теперь перенесем эти соображения на вселенную. Плотность ее в пределах досягаемости и точности измерений постоянна. Конечно, под плотностью понимается усредненная плотность по достаточно большому шару, но и не очень большому: много, много меньшему размеров обозримой нами вселенной. Можно думать, что где-то еще дальше она уменьшается и даже исчезает и что наша Солнечная система находится где-то вблизи центра вселенной, центра по отношению к этому очень далекому убыванию. Об этом говорит отсутствие притяжения вселенной нашей Солнечной системы, с точностью до локальных флуктуаций плотности.

Далее, чтобы ответить на вопрос, сжимается или разбегаются вселенная, нужно иметь некоторое представление о скоростях составляющих ее масс. Эти сведения дает закон Хаббла, согласно которому на хаотические скорости звезд накладывается систематическая радиальная составляющая. Звезды радиально разбегаются со скоростью

$$V = HR \text{ км/с}, \quad (27.23)$$

где H – постоянная Хаббла, примерно равная $75 \text{ км/с} \cdot \text{Мик}$, а R – расстояние до звезды, измеряемое в Мик (Мик – это единица расстояния, равная примерно пути, пробегаемому светом за три года).

Теперь выделим большой шар радиуса R с центром, где мы находимся, и спросим, могут ли звезды или другие тела его покинуть? Они смогут это сделать, если скорости на поверхности шара превосходят вторую космическую скорость выделенного нами шара, т.е. если

$$V = HR > \sqrt{\frac{2\gamma M}{R}}, \quad (27.24)$$

и не смогут это сделать при противоположном неравенстве. Выражая массу M выделенного шара радиуса R через среднюю плотность ρ , найдем, что

$$M = \frac{4}{3}\pi\rho R^3.$$

Подставляя значение M в (27.24), находим, что критическая, разделяющая случаи разбегания и сжатия, плотность $\rho_{\text{кр}}$, не зависит от R и равна

$$\rho_{\text{кр}} = \frac{3H^2}{8\pi\gamma} \approx 10^{-29} \text{ г/см}^3. \quad (27.25)$$

Таким образом, если плотность массы во вселенной больше $\rho_{\text{кр}}$, вселенная будет сжиматься, если меньше – расширяться. Сегодняшние подсчеты ρ для вселенной требуют учета межзвездного темного вещества и дают с его учетом цифру, близкую к $\rho_{\text{кр}}$, так что вопрос остается открытым.

Мы с достаточной скрупулезностью проделали все вычисления, с которыми впервые справился Ньютон. Вы понимаете, что в те времена это был громадный, титанический труд, полный неожиданных препятствий. Ведь до Ньютона не было ни дифференциального, ни интегрального исчисления, не было и дифференциальных уравнений.

В дальнейшем пытались обобщить результаты Ньютона на большое число притягивающихся тел, на так называемую проблему n тел. Но никому это не удалось. Спустя значительное время стало ясно, что это нельзя сделать: проблема n тел при $n > 2$ неинтегрируемая. Возникли сложные теории приближенного решения, методы так называемой теории возмущений. Эта теория достигла высокой степени совершенства и сыграла в науке громадную роль, не только в небесной механике. Путем вычислений, основанных на теории возмущений, удалось не только предсказать на много лет вперед положение планет, но и открыть новую планету Нептун. Но проинтегрировать уравнения не удалось даже в упрощенной проблеме трех тел, когда одно из тел имеет исчезающе малую массу и притяжением двух других тел можно пренебречь.

Теория возмущений основана на построении последовательных приближений, например, в виде степенного ряда по некоторому малому параметру. В первом приближении учитывается основное взаимодействие каждой планеты с Солнцем. В нем все планеты движутся по неизменным кеплеровым орбитам. Плоскости этих орбит близки к плоскости эклиптики. Положение каждой из орбит относительно плоскости эклиптики определяется какими-то параметрами. В первом приближении они постоянны. Затем строится второе приближение, учитывающее влияние всех или части остальных планет, при этом считается, что они движутся по неизменным эллиптическим орбитам. Затем эта процедура повторяется с учетом уточненных орбит планет и т.д. В результате возникает решение для изменений параметров орбит и движений по ним планет в виде ряда по малому параметру – отношения масс планет к массе Солнца. Построение этих рядов достаточно сложно и, с одной стороны, достигло высокой степени разработанности, а с другой, обнаружило ряд трудностей, среди которых самой стойкой и непреодоленной оказалась проблема малых знаменателей.

Качественная теория дифференциальных уравнений А. Пуанкаре – ответ на эти трудности. Начало ей положили знаменитые труды А. Пуанкаре «Новые методы небесной механики» [40, 41] и «О кривых, определяемых дифференциальными уравнениями». Но и эта новая замечательная теория оказалась бессильной не только перед проблемой n тел, но даже перед ограниченной (упрощенной) проблемой трех тел. Сегодня расчет движения планет и искусственных спутников, благодаря успехам в развитии методов вычислений и вычислительной техники, выполняют ЭВМ. Они это делают быстро и очень точно. Но не все им подвластно. И сегодня нет ответа, какой будет Солнечная система через миллиарды лет. Устойчива ли она или может существенно измениться? Прямые вычисления здесь помочь не могут: как бы ни были малы ошибки счета за миллиарды лет, они накапливаются и становятся очень большими и непредсказуемыми.

Это вызвано тем, что дифференциальные уравнения движения в задаче n тел – консервативные гамильтоновы, не допускающие экспоненциально устойчивых движений. Любое их решение либо экспоненциально неустойчиво, и тогда неизбежно экспоненциальное же нарастание ошибки при его приближенном вычислении, либо нейтральное, для которого ошибки счета так же, хотя и значительно медленнее, но все же нарастают. Более того, при сколь угодно малых независимых случайных ошибках спустя достаточное время Солнечная система, скорее всего, может прийти к любому состоянию, достижимому при ненарушении основных законов сохранения энергии, количества и момента количества движения.

Вместе с тем, у Солнечной системы наблюдаются удивительные структурные особенности и закономерности. Особое удивление вызвали удивительные синхронизмы между периодами обращения планет, орбитальными и собственными, тонкая структура колец Сатурна и другие. Все это ни теория, ни счет долго не объясняли. Они и не могли объяснить, пока господствовала только консервативная гамильтонова модель Солнечной системы.

Эти закономерности своим появлением обязаны неконсервативным силам, не только силам ньютоновского притяжения, но и, пусть и очень малым, диссипативным силам. Солнечная система существует миллиарды лет и именно эти, казалось бы, ничтожные силы, породили удивительную структуру и закономерности ее движения.

Изучение этих структурных особенностей долго эволюционирующей Солнечной системы невозможно и без более полного учета гравитационных взаимодействий между ее телами. Если однородные шарообразные тела взаимодействуют друг с другом как материальные точки, помещенные в их центрах, что знал еще Ньютон, то для тел нешарообразных и неоднородных это не так. Нужно учитывать не только нешарообразность тел Солнечной системы, но и их деформируемость, приводящую к приливному трению и силам, вызываемым внутренними деформациями. Этот учет кардинально качественно меняет фазовый портрет и движение Солнечной системы, т.к. консервативная гамильтонова система дифференциальных уравнений движения превращается в диссипативную, допускающую экспоненциально устойчивые движения. Их уже можно находить с помощью вычислений, а экспоненциально устойчивые установившиеся движения можно найти теоретически с помощью современной теории систем с быстровращающимися фазами.

Наличие малой диссипации упростило фазовый портрет Солнечной системы, в нем уже нет прежнего, возможно, бесконечного числа периодических движений, он упростился и стал доступен теоретическому и вычислительному изучению, и это сделал учет, казалось бы, совершенно ничтожных добавок, не обнаруживаемых в первых приближениях.

28. РАСПРЕДЕЛЕННЫЕ ДИНАМИЧЕСКИЕ МОДЕЛИ МЕХАНИКИ И ФИЗИКИ

До сих пор рассматриваемые нами математические модели – динамические системы – были в основном такими, что их состояние задавалось конечным числом скаляров, а его изменение подчинялось обыкновенным дифференциальным уравнениям

$$\frac{dx}{dt} = X(x), \quad (28.1)$$

где x – конечномерный вектор состояния, а $X(x)$ – вектор-функция x .

Сейчас мы перейдем к другому типу динамических систем, у которых состояние определяется функциями от каких-то переменных, а изменение состояния во времени, т.е. изменение описывающих его функций, подчиняется уравнению в частных производных вида

$$\frac{\partial u}{\partial t} = F\left(u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n}, \dots\right), \quad (28.2)$$

где $u(x_1, \dots, x_n; t)$ – функция или вектор-функция, описывающая состояние в момент времени t , а правая часть уравнения (28.2) зависит не только от функции u , но и ее частных производных по переменным x_1, \dots, x_n до какого-то порядка.

При этом данное ранее общее определение динамической системы полностью сохраняется. Изменился лишь способ описания ее состояния и в связи с этим и вид ее оператора. Как и ранее, изменение состояния во времени, т.е. производная $\partial u / \partial t$ определяется через состояние, т.е. функцию $u(x_1, \dots, x_n; t)$ и определяемые ею частные производные по x_1, \dots, x_n . Как и ранее, это позволяет приближенно найти состояние в следующий момент времени $t + \Delta t$ через состояние в предшествующий момент времени t :

$$u(x_1, \dots, x_n; t + \Delta t) \approx u(x_1, \dots, x_n; t) + \tilde{F}(x_1, \dots, x_n; t) \Delta t$$

и надеяться делать это сколь угодно точно с уменьшением Δt . В предположении существования гладкого решения ошибка на одном шаге порядка $(\Delta t)^2$, а на конечном промежутке времени T порядка $(\Delta t)^2 T / \Delta t = T \Delta t$ и стремится к нулю вместе с Δt .

Описание состояния функциями во многих случаях вызвано пространственной распределенностью описываемой реальной системы. Так, тепловое состояние среды описывается температурой $T(x, y, z, t)$, зависящей от пространственных переменных x, y, z и меняющейся со временем t . Давление в жидкости (газе) определяется его величиной $p(x, y, z, t)$ в каждой ее точке (x, y, z) . Аналогичны описания $\rho(x, y, z, t)$ плотности вещества или распределения в пространстве электрического заряда.

Это все примеры скалярных полей и описаний их одной функцией от пространственных переменных и времени.

Поля могут быть векторными. Таково поле скоростей $\bar{v}(x, y, z, t)$ течения жидкости или газа; поле смещений $\bar{r}(x, y, z, t)$ точек упругого деформируемого тела; электрические и магнитные поля, описываемые векторами электрического и магнитного полей.

Рассматриваемые нами ранее системы, состояние которых описывалось конечным числом скалярных переменных, называются дискретными. Системы, о которых сейчас речь и у которых состояние описывается функциями, называются распределенными.

Распределенную систему можно трактовать как дискретную с бесконечным числом переменных. Действительно, функцию, например, трех переменных $f(x, y, z)$ можно разложить в ряд по некоторым функциям $\varphi_1(x, y, z)$, $\varphi_2(x, y, z)$, ..., так что

$$f(x, y, z) = \sum_{j=1}^{\infty} c_j \varphi_j(x, y, z),$$

и тем самым функция $f(x, y, z)$ определяется бесконечным числом скалярных переменных c_1, c_2, \dots

При такой замене при независимости функций $\varphi_1, \varphi_2, \dots$ уравнение (28.2) повлечет за собой эквивалентную ему бесконечную систему обыкновенных дифференциальных уравнений

$$\frac{dc_j}{dt} = F(c_1, c_2, \dots) \quad (j = 1, 2, 3).$$

Сказанное достаточно очевидно, но требует обоснований, и они не так уж и просты, скорее даже очень сложны, если пытаться уменьшать ограничения на функции f и F . Но это уже не предмет этих лекций.

Мир распределенных систем, пожалуй, еще более многообразен и широк, чем дискретных. Более того, на дискретные системы можно смотреть как на частный конечномерный случай распределенных систем. Это не делается только потому, что существующие методы исследования уравнений в частных производных существенно отличаются от методов теории обыкновенных дифференциальных уравнений. Фактически это разные научные дисциплины, разные разделы математики, которыми занимаются разные люди.

Вместе с тем, те и другие дифференциальные уравнения при использовании общих понятий функционального анализа записываются одинаково в виде (28.1), где теперь x – элемент некоторого функционального пространства, а $X(x)$ не вектор-функция от x , а оператор X от x .

Дальнейший рассказ – не более чем знакомство, предисловие или введение в науку о распределенных динамических системах. Более основательные встречи с ними могут состояться в курсах математической физики, гидродинамики, теории упругости, электродинамики, статистической физики, теории вероятностей и случайных процессов и других. Поскольку это предисловие, то многое будет носить описательный, а не доказательный характер. Это в первую очередь относится к основным моделям классической математической физики и великим уравнениям в частных производных механики и физики: уравнениям Эйлера, Навье – Стокса и уравнениям Максвелла. Вам придется поверить на слово, не требуя их обоснова-

ния и вывода, и это не должно вам мешать их «понять» и пользоваться ими при рассмотрении конкретных примеров. Исключением из этого будет уравнение теплопроводности или диффузии, для того чтобы дать вам возможность увидеть, как возникают и откуда берутся уравнения в частных производных. Речь идет об уравнении

$$\frac{\partial u}{\partial t} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right), \quad (28.3)$$

которое одновременно описывает явления теплопроводности, диффузии и случайного блуждания. В первом случае $u(x, y, z, t)$ – температура в точке (x, y, z) в момент времени t , во втором – плотность диффундирующего вещества, в третьем – плотность вероятности того, что случайная блуждающая точка в момент времени t находится в точке (x, y, z) .

Начнем с распространения тепла в однородной одномерной среде. Поле температуры в этом случае описывается функцией $T(x, t)$. Ее изменение происходит благодаря естественному перетеканию тепла с более горячих мест в более холодные. Это перетекание во многих случаях можно принять пропорциональным градиенту температуры, так что

$$dQ = -k \frac{\partial T}{\partial x} dt, \quad (28.4)$$

где dQ – количество тепла, перетекшего в направлении оси Ox за время dt через сечение x (рис. 28.1). Знак минус указывает на то, что перетекание тепла происходит в направлении, обратном возрастанию температуры. Однородность среды выражается в постоянстве коэффициента теплопроводности k .

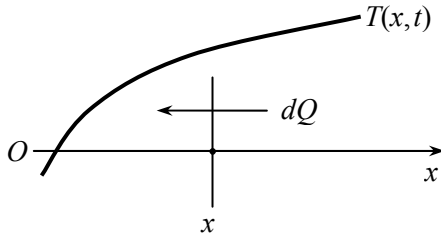


Рис. 28.1. Распространение тепла в однородной одномерной среде

Выделим произвольный участок $[x_1, x_2]$. Принимая, что теплоемкость среды постоянна и равна c , находим, что в момент времени t в нем содержится количество тепла, равное $\int_{x_1}^{x_2} cT(x, t) dx$.

Прирост этого количества тепла, если нет излучения, за время dt определяется его притоком через сечения $x = x_1$ и $x = x_2$ ($x_2 > x_1$), т.е.

$$\frac{\partial}{\partial t} \left(\int_{x_1}^{x_2} cT(x, t) dx \right) dt = -k \frac{\partial T}{\partial x} \Big|_{x=x_1} dt + k \frac{\partial T}{\partial x} \Big|_{x=x_2} dt,$$

сокращая на dt и дифференцируя по t , получим

$$c \int_{x_1}^{x_2} \frac{\partial T}{\partial t} dx = k \left(\frac{\partial T}{\partial x} \Big|_{x=x_2} - \frac{\partial T}{\partial x} \Big|_{x=x_1} \right). \quad (28.5)$$

Нам осталось перейти к пределу при неограниченном сближении сечений x_1 и x_2 к произвольному x . Воспользуемся теоремой о среднем для интеграла в соотношении (28.5) и запишем это соотношение в виде:

$$c(x_2 - x_1) \frac{\partial T}{\partial t} \Big|_{x=\xi} = k \left(\frac{\partial T}{\partial x} \Big|_{x=x_2} - \frac{\partial T}{\partial x} \Big|_{x=x_1} \right),$$

где ξ – некоторая промежуточная точка между x_1 и x_2 . Разделим полученное равенство на $x_2 - x_1$ и перейдем к пределу при стремлении этой разности к нулю в связи со сближением точек x_1 и x_2 к точке x . В результате придем к искомому уравнению теплопроводности

$$c \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2},$$

которое обычно записывают в виде

$$\frac{\partial T}{\partial t} = a^2 \frac{\partial^2 T}{\partial x^2} \quad \left(a^2 = \frac{k}{c} \right). \quad (28.6)$$

Таким же уравнением описывается и процесс диффузии одного вещества с концентрацией $\rho(x, t)$ в другом. Это следует из того, что при диффузии происходит перемещение вещества из мест с большей концентрацией в места с меньшей, причем для количества переместившегося вещества dm за время dt имеет место тот же закон, что и для перемещения тепла, т.е.

$$dm = -D \frac{\partial \rho}{\partial x} dt, \quad (28.7)$$

где D – коэффициент диффузии. Поэтому такие же рассуждения и выкладки приводят к уравнению (28.6), в котором $a^2 = D$.

К такому же уравнению приводит и задача о случайных блужданиях частицы по прямой Ox . Назовем эту задачу задачей об абсолютно пьяном, который не знает, куда ему идти, и делает шаг вперед или назад равновероятно. Пусть шагает он через время τ , а длина его шага a . Пусть еще $p(sa, n\tau)$ – вероятность того, что в момент $t = n\tau$ после n шагов он оказался в точке $x=sa$ (s – целое).

Нетрудно установить связь между вероятностями $p(sa, n\tau)$ и $p(sa, (n+1)\tau)$. Действительно, в точке sa он может оказаться, только шагнув перед этим вперед из точки $(s-1)a$, или шагнув назад из точки $(s+1)a$. То и другое он делает равновероятно с вероятностью $1/2$, поэтому

$$p(sa, (n+1)\tau) = \frac{1}{2} p((s-1)a, n\tau) + \frac{1}{2} p((s+1)a, n\tau). \quad (28.8)$$

Соотношение (28.8) позволяет, зная функцию $p(sa, n\tau)$ при $n = 0$, найти последовательно ее при любом n . Так, если при $n = 0$ абсолютно пьяный – блуждающая частица – находится в точке $x = 0$, то функция $p(sa, 0)$ известна и равна

$$p(sa,0) = \begin{cases} 0 & \text{при } s \neq 0, \\ 1 & \text{при } s = 0, \end{cases}$$

и по ней может быть найдена любая из вероятностей $p(sa, n\tau)$ при любых s и n .

Допустим теперь, что шаги пьяного все убыстряются, а их длина уменьшается. Пусть τ и a стремятся к нулю, так, что

$$\lim \frac{a^2}{2\tau} = D. \tag{28.9}$$

Запишем соотношение (28.8) в виде:

$$\frac{p(sa, (n+1)\tau) - p(sa, n\tau)}{\tau} = \frac{a^2}{2\tau} \frac{p((s+1)a, n\tau) - 2p(sa, n\tau) + p(sa, n\tau)}{a^2},$$

переходя в нем к пределу при $\tau \rightarrow 0$, $a \rightarrow \tau$ и учитывая условие (28.9), приходим к уравнению в частных производных относительно функции $p(x, t)$:

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}, \tag{28.10}$$

совпадающему по виду с уравнением теплопроводности (28.6) ($a^2 = D$).

Не следует думать, что в уравнении (28.10), как это было бы естественно, стоит предел дискретной функции $p(sa, n\tau)$. Этот предел, как нетрудно видеть, равен нулю, так как все вероятности $p(sa, n\tau)$ при $a \rightarrow 0$ стремятся к нулю. Конечно, нуль уравнению (28.10) удовлетворяет, но это нас не устраивает. Нам нужно ненулевое решение уравнения (28.10), удовлетворяющее условию нормировки

$$\int_{-\infty}^{+\infty} p(x, t) dx = 1,$$

которое интерпретируется для дискретных функций $p(sa, n\tau)$ в виде

$$\sum_{s=-\infty}^{+\infty} p(sa, n\tau) = 1.$$

Его можно записать в виде

$$\sum \frac{p(sa, n\tau)}{a} a = 1,$$

приводящем к мысли, что $p(x, t)$ следует считать пределом дискретной функции $p(sa, n\tau)/a$.

Приведенные выводы нетрудно повторить в более общих предположениях двумерности или трехмерности пространства. При этом мы во всех трех случаях приходим к уравнению вида

$$\frac{\partial T}{\partial t} = a^2 \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) \tag{28.11}$$

или, соответственно, вида

$$\frac{\partial T}{\partial t} = a^2 \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right). \tag{28.12}$$

Дифференциальное уравнение в частных производных (28.12) (и его частные случаи (28.6) и (28.11)) является одним из основных уравнений классической математической физики. Вторым основным ее уравнением является волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right). \quad (28.13)$$

Это уравнение описывает колебания упругих сред. Для звуковых колебаний газа функция $u(x, y, z, t)$ имеет смысл давления в точке (x, y, z) в момент времени t . В одномерном случае, когда

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2},$$

функция $u(x, t)$ может быть отклонением натянутой струны от ее равновесного положения. Это же уравнение описывает продольные упругие колебания однородного стержня. В этом случае $u(x, t)$ – упругое продольное смещение материала стержня в сечении x в момент t . В двумерном варианте уравнение (28.11) может описывать поперечные колебания мембраны.

Уравнения теплопроводности и волновое уравнение классической математической физики – линейные. Это значит, что к ним применим принцип суперпозиции, тот же самый, что и для линейного осциллятора или любой другой линейной системы, т.е. найдя какие-нибудь их решения, можно с их помощью строить новые решения или, может быть, даже любое нужное нам решение. Такими частными решениями являются так называемое фундаментальное решение уравнения теплопроводности и гармонические бегущие волны. С тем и другим ниже мы познакомимся.

Существенным отличием динамических систем, описываемых дифференциальными уравнениями в частных производных, от тех, которые описываются обыкновенными дифференциальными уравнениями, состоит в том, что равновесные состояния дискретных динамических систем находятся сравнительно просто из обращения в нуль их правых частей, т.е. из некоторых нелинейных уравнений, в то время как для распределенных динамических систем уравнения, определяющие равновесие, оказываются также дифференциальными. Так, полагая в (28.12) или (28.13) $T(x, y, z, t) = T^*(x, y, z)$ и $u(x, y, z, t) = u^*(x, y, z)$ равновесными состояниями, не меняющимися со временем, приходим к тому, что они удовлетворяют уравнениям в частных производных

$$\frac{\partial^2 T^*}{\partial x^2} + \frac{\partial^2 T^*}{\partial y^2} + \frac{\partial^2 T^*}{\partial z^2} = 0 \quad (28.14)$$

и, соответственно,

$$\frac{\partial^2 u^*}{\partial x^2} + \frac{\partial^2 u^*}{\partial y^2} + \frac{\partial^2 u^*}{\partial z^2} = 0. \quad (28.15)$$

Решение этих уравнений настолько сложно, что они также входят в список уравнений классической математической физики.

Продолжим знакомство с другими математическими моделями распределенных динамических систем.

Движение вязкой жидкости подчиняется нелинейным дифференциальным уравнениям в частных производных Навье–Стокса

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \nabla) \mathbf{v} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{v} \quad (\mathbf{v} = (v_x, v_y, v_z)). \quad (28.16)$$

Частным его случаем при $\nu = 0$ является уравнение Эйлера движения идеальной (невязкой) жидкости. В этом уравнении $\mathbf{v}(x, y, z, t)$ – вектор скорости жидкости в точке x, y, z в момент t ; $p(x, y, z, t)$ – давление; ρ – плотность; ν – параметр, так называемая кинематическая вязкость; ∇ и Δ – известные символы операторов градиента и Лапласа. Именно градиент

$$\nabla = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z},$$

где i, j, k – орты прямоугольной системы координат, а оператор Лапласа имеет вид:

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

В случае несжимаемой жидкости к уравнению Навье–Стокса или Эйлера требуется добавить условие несжимаемости, состоящее в том, что скорости жидких частиц должны быть такими, чтобы их объем не менялся. Это условие также записывается с помощью уравнения в частных производных, а именно:

$$\operatorname{div} \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = 0, \quad (28.17)$$

а для сжимаемой жидкости в виде условия

$$\operatorname{div} \rho \mathbf{v} = 0, \quad (28.18)$$

выражающего сохранение не объема, а массы.

Вас бегло познакомлю еще с великими уравнениями Максвелла. Это достаточно сложная система уравнений в частных производных. В нее входит много разных функций: вектор-функции $E(x, y, z, t)$, $D(x, y, z, t)$, $H(x, y, z, t)$, $B(x, y, z, t)$, описывающие электрическое и магнитное поля (поля вектора электрической напряженности и индукции, магнитной напряженности и индукции); скалярная функция распределенного электрического заряда $\rho(x, y, z, t)$ и векторная функция поля электрических токов $j(x, y, z, t)$. Уравнения Максвелла связывают эти функции между собой. В векторной форме и в сокращающих их запись общепринятых обозначениях они могут быть записаны в виде следующих основных уравнений¹:

¹ $\operatorname{rot} A$ – дифференциальная операция над вектором $A (A_x, A_y, A_z)$, записанная в виде определителя:

$$\operatorname{rot} A = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} = i \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + j \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + k \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right).$$

Определение $\operatorname{div} A$ содержится в уравнении (28.17).

$$\frac{\partial D}{\partial t} = -4\pi j + c \operatorname{rot} H, \quad \frac{\partial B}{\partial t} = -c \operatorname{rot} E, \quad (28.19)$$

к которым следует добавить еще два уравнения

$$\operatorname{div} B = 0, \quad \operatorname{div} D = -4\pi\rho, \quad (28.20)$$

и в простейшем, частном случае, еще три алгебраических соотношения

$$D = \varepsilon E, \quad B = \mu H, \quad j = \sigma E. \quad (28.21)$$

Это очень сложная система уравнений. Это и понятно, ведь из нее следуют все явления электростатики, магнитостатики, электродинамики и электромагнитного излучения и волн. Они при независимости ε и μ от полей E и H линейные и поэтому допускают применение принципа суперпозиции. В этом смысле более простые на вид уравнения гидродинамики куда хуже, они нелинейные и принцип суперпозиции для них не имеет места.

Вы увидели несколько замечательных, даже великих, распределенных математических моделей механики и физики. Возможно, они вам чужды и непонятны. Не смущайтесь этим, ведь это не более как математические описания, математические модели. Только это от вас сейчас требуется. Вы должны принять их в этом качестве, не смущаясь ни их сложностью, ни вашим сегодняшним незнанием, как их решать и применять.

Во всех приведенных моделях описание состояния функциями возникло потому, что речь шла о тех или иных физических пространственных полях: поле температур, плотности, скорости, смещений, давления, электрической и магнитной напряженностей и индукции, плотности вероятностей. Но возможны и совсем другие причины возникновения в описании состояния функций. Этот другой случай вам встретится позднее при рассмотрении математических моделей микромира. Состояние динамических систем микромира описывается функцией от координат всех входящих в систему частиц и их спинов, а уравнение, которому эта функция удовлетворяет, – это знаменитое уравнение Шредингера для функции состояния ψ :

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi, \quad (28.22)$$

которое пишется очень кратко, но в котором отражается весь окружающий мир. Уравнение это на самом деле очень сложное и решается очень трудно: функция ψ – комплексная и может зависеть от очень большого числа переменных, а под знаком оператора Гамильтона H спрятан сложный дифференциальный оператор, содержащий частные производные второго порядка. Этому не следует удивляться, ведь именно он, в конечном счете, определяет все происходящее в окружающем нас неживом и живом мире.

Ко всему сказанному о распределенных моделях динамических систем (классических уравнениях математической физики, уравнениях Эйлера и Навье–Стокса, Максвелла и Шредингера) необходимо добавить, что сами по себе они еще не дают полного описания изменения входящих в них функций. К этим уравнениям необходимо присоединить еще некоторые дополнительные условия. Они весьма разнообразны и определяются конкретным рассматриваемым объектом. Для уравнений математической физики при неограниченности пространства пе-

ременных x, y, z – это некоторые условия ограниченности на бесконечности, при ограниченности области переменных x, y, z – это некоторые условия на ее границе, называемые краевыми. То же имеет место и для уравнений гидродинамики, электродинамики и уравнений Шредингера. Только совместно с этими дополнительными условиями входящие в эти уравнения функции от пространственных переменных определяют состояние, т.е. будучи заданы при $t = t_0$, они могут быть найдены при любом $t > t_0$. С некоторыми конкретными краевыми условиями вы встретитесь в дальнейшем. А сейчас ограничимся простым примером натянутой струны, концы которой неподвижно закреплены.

В этом случае уравнение малых колебаний струны – это волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \quad (28.23)$$

где $u(x, t)$ – отклонение струны от равновесного натянутого положения в момент времени t . Это уравнение имеет место на всей струне, т.е. при $0 \leq x \leq l$, где $x = 0$ и $x = l$ – координаты закрепленных концов струны. На этих концах отклонение u отсутствует, и поэтому имеют место краевые условия

$$u|_{x=0} = 0, \quad u|_{x=l} = 0. \quad (28.24)$$

Уравнения в частных производных (28.23) совместно с краевыми условиями (28.24) допускают при заданном начальном положении и скоростях

$$u(x, t)|_{t=0} = \varphi(x), \quad \left. \frac{\partial u}{\partial t} \right|_{t=0} = \psi(x) \quad (28.25)$$

единственное решение $u(x, t)$. Состояние в этом случае определяется в начальный момент функциями $\varphi(x)$ и $\psi(x)$, удовлетворяющими краевым условиям $\varphi(0) = \varphi(l) = 0$, $\psi(0) = \psi(l) = 0$. При любом t состояние описывается удовлетворяющими краевым условиям функциями от x : $u(x, t)$ и $\partial u(x, t)/\partial t$. Вас не должно удивлять, что наряду с функцией $u(x, t)$ в описание состояния входит еще и функция $\partial u(x, t)/\partial t$. Это объясняется тем, что волновое уравнение (28.13) в виде (28.2) записывается как

$$\frac{\partial}{\partial t}(u, v) = (v, a^2 \Delta u), \quad (28.26)$$

и состоянием для него является вектор, составленный из двух функций $u(x, t)$ и $v(x, t)$, как функций x при фиксированном времени t .

29. ФУНДАМЕНТАЛЬНОЕ РЕШЕНИЕ УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ

Распространение тепла в неограниченной однородной одномерной теплопроводящей среде описывается дифференциальным уравнением в частных производных вида

$$\frac{\partial T}{\partial t} = a^2 \frac{\partial^2 T}{\partial x^2}, \quad T = T(x, t). \quad (29.1)$$

Такое же по виду уравнение описывает диффузию или случайное блуждание частицы.

Пусть в начальный момент времени $t = 0$ имеется некоторое произвольное поле температур $T_0(x)$, так что

$$T(x, t)|_{t=0} = T_0(x). \quad (29.2)$$

Уравнение теплопроводности (29.1) по начальному распределению температур $T_0(x)$ определяет дальнейшее поле температур $T(x, t)$ при любом $t > 0$.

Уравнение (29.1) и начальное условие (29.2) описывают не только изменение начального поля температур, но и диффузию вещества, и изменение со временем плотности вероятности случайно блуждающей частицы. При этом $T_0(x)$ – начальная плотность вещества или, соответственно, начальная плотность вероятности.

Не будем решать эту общую для всех перечисленных явлений задачу (29.1), (29.2) сразу для произвольного начального поля $T_0(x)$. Попытаемся найти ее решение для точечного начального поля, когда только в одной точке x_0 среда разогрета, или только в ней находится диффундирующее вещество, или в ней находится случайная частица.

Сразу же возникает вопрос: как, какой моделью описать точечное сосредоточение тепла, вещества или плотности вероятности? С физической точки зрения идеализации точечным сгущением массы или тепла обычны. Но как их описать математически? Ведь в этих случаях температура, плотность диффундирующего вещества или, соответственно, плотность вероятности становятся бесконечными, а всюду вне этой точки равными нулю. Не будем смущаться и запишем, что для описывающей функции $T_0(x)$ имеет место

$$T_0(x) = \begin{cases} \infty & \text{при } x = x_0, \\ 0 & \text{при } x \neq x_0, \end{cases} \quad (29.3)$$

и еще запишем, что в точке $x = x_0$ сосредоточено c (c – теплоемкость вещества) единиц тепла, единичная масса или вероятность, равная единице,

$$\int_{-\infty}^{\infty} T_0(x) dx = 1. \quad (29.4)$$

Безусловно, функция $T_0(x)$, определяемая (29.3), – не функция в обычном смысле, а интеграл (29.4) – не интеграл. Будем ее рассматривать как некий новый символ – δ -функцию или обобщенную функцию. Эту необычную δ -функцию ввел

в математику О. Хевисайд в связи с созданным им операционным исчислением, и позднее физик П. Дирак – в квантовой механике.

Обобщенная функция $\delta(x)$ определяется условиями:

$$\delta(x) = \begin{cases} \infty & \text{при } x = 0, \\ 0 & \text{при } x \neq 0, \end{cases} \quad (29.5)$$

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

Из (29.5) формально следует, что для любой непрерывной функции

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = f(x_0). \quad (29.6)$$

Действительно,

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0) dx = \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} f(x) \delta(x - x_0) dx$$

при любом сколь угодно малом $\varepsilon > 0$. По известной теореме о среднем этот последний интеграл равен

$$f(\xi) \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} \delta(x - x_0) dx = f(\xi),$$

где ξ лежит в промежутке от $x_0 - \varepsilon$ до $x_0 + \varepsilon$. Так как ε может быть сколь угодно малым и функция $f(x)$ – непрерывна, то $\xi = x_0$.

Пусть $x_0 = 0$ и начальному распределению температур $T_0(x) = \delta(x)$ отвечает меняющееся поле температур $\varphi(x, t)$. Функция $\varphi(x, t)$ показывает, как растекается c единиц тепла, сосредоточенных в точке $x = 0$, по всей прямой. Для задач диффузии и случайного блуждания функция $\varphi(x, t)$ описывает диффузию вещества, поначалу сосредоточенного в точке $x = 0$, а для случайного блуждания частицы – плотность вероятности нахождения и в точке x спустя время t , если в начальный момент она находилась в точке $x = 0$.

Найти путем логического рассуждения функцию $\varphi(x, t)$ не очень просто, догадаться можно, но трудно. Мы не будем делать ни того, ни другого, а просто ее напишем

$$\varphi(x, t) = \frac{1}{2a\sqrt{\pi t}} \exp\left(-\frac{x^2}{4a^2 t}\right) \quad (29.7)$$

и проверим, что эта функция $\varphi(x, t)$ удовлетворяет при любых x и $t > 0$ уравнению теплопроводности (29.1) и начальным условиям (29.2), когда $T_0(x) = \delta(x)$. Не нужно только последнее понимать буквально: если в (29.7) подставить $t = 0$, то нам дважды придется делить на 0. Поэтому будем не подставлять $t = 0$, а переходить к пределу при $t \rightarrow 0$, т.е. покажем, что

$$\lim_{t \rightarrow 0} \varphi(x, t) = \begin{cases} 0 & \text{при } x \neq 0, \\ \infty & \text{при } x = 0 \end{cases}$$

и что

$$\lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \varphi(x, t) dx = 1.$$

Первое очевидно, второе мы установим, показав, что при всех $t > 0$

$$\int_{-\infty}^{\infty} \varphi(x, t) dx = 1.$$

Действительно,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{2a\sqrt{\pi t}} \exp\left(-\frac{x^2}{4a^2t}\right) dx &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{4a^2t}\right) d\frac{x}{2a\sqrt{t}} = \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-\xi^2) d\xi = 1. \end{aligned} \tag{29.8}$$

Полученный несобственный определенный интеграл хорошо известен и равен $\sqrt{\pi}$. Убедиться в этом можно следующим образом. Обозначим этот интеграл через I . Тогда

$$I^2 = \int_{-\infty}^{\infty} \exp(-\xi^2) d\xi \int_{-\infty}^{\infty} \exp(-\eta^2) d\eta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-\xi^2 - \eta^2) d\xi d\eta,$$

переходя к полярным координатам r и φ , получим

$$I^2 = \int_0^{\infty} dr \int_{-\pi}^{\pi} \exp(-r^2) r d\varphi = \pi \int_0^{\infty} \exp(-r^2) dr^2 = \pi.$$

То, что функция $\varphi(x, t)$ удовлетворяет уравнению теплопроводности (29.1) при всех $t > 0$, можно убедиться непосредственно подстановкой. Итак, функция $\varphi(x, t)$ – та, которую мы искали. Напомним еще раз, что $\varphi(x, t)$, определяемая формулой (29.7), соответствует тому, что в начальный момент $t = 0$ в точке $x = 0$ помещено c единиц тепла. В случае если речь идет о диффузии, то в начальный момент $t = 0$ в точке $x = 0$ помещена единичная масса диффундирующего вещества. Наконец, при случайном блуждании в момент $t = 0$ блуждающая частица находится в точке $x = 0$. При $t \approx 0$ функция $\varphi(x, t)$ имеет вид высокого (высоты $1/2a\sqrt{\pi t}$) и тонкого пика. С ростом времени t этот пик расплывается, как показано на рис. 29.1. Площадь, ограниченная этой расплывающейся колоколообразной кривой, все время при всех t , согласно (29.8), равна единице, что отвечает сохранению тепла, вещества или, соответственно, нормировке плотности вероятности.

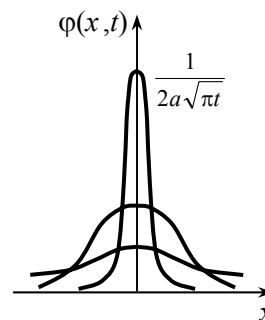


Рис. 29.1. Изменение со временем функции отклика $\varphi(x, t)$

Зная функцию $\varphi(x, t)$, описывающую расплывание тепла, сосредоточенного в одной точке, можно найти по ней аналогичные функции в двумерном и трехмерном случае. В трехмерном случае такой функцией будет

$$\varphi(x, y, z, t) = \varphi(x, t)\varphi(y, t)\varphi(z, t) = \frac{1}{8a^3(\sqrt{\pi t})^3} \exp\left(-\frac{x^2 + y^2 + z^2}{4a^2 t}\right). \quad (29.9)$$

Проверить это очень просто. Непосредственно находим, что

$$\lim_{t \rightarrow 0} \varphi(x, y, z, t) = \begin{cases} \infty & \text{при } x = y = z = 0, \\ 0 & \text{при } x^2 + y^2 + z^2 \neq 0. \end{cases} \quad (29.10)$$

Запишем

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y, z, t) dx dy dz = \int_{-\infty}^{\infty} \varphi(x, t) dx \int_{-\infty}^{\infty} \varphi(y, t) dy \int_{-\infty}^{\infty} \varphi(z, t) dz = 1.$$

Далее уравнение

$$\frac{\partial \varphi(x, y, z, t)}{\partial t} = \frac{\partial^2 \varphi(x, y, z, t)}{\partial x^2} + \frac{\partial^2 \varphi(x, y, z, t)}{\partial y^2} + \frac{\partial^2 \varphi(x, y, z, t)}{\partial z^2}$$

записывается в виде:

$$\begin{aligned} & \varphi(y, t)\varphi(z, t) \frac{\partial \varphi(x, t)}{\partial t} + \varphi(x, t)\varphi(z, t) \frac{\partial \varphi(y, t)}{\partial t} + \varphi(x, t)\varphi(y, t) \frac{\partial \varphi(z, t)}{\partial t} = \\ & = \varphi(y, t)\varphi(z, t) \frac{\partial^2 \varphi(x, t)}{\partial x^2} + \varphi(x, t)\varphi(z, t) \frac{\partial^2 \varphi(y, t)}{\partial y^2} + \varphi(x, t)\varphi(y, t) \frac{\partial^2 \varphi(z, t)}{\partial z^2}, \end{aligned}$$

из которого следует его выполнимость.

Функция $\varphi(x, t)$ описывает температурное поле, которое возникает растеканием s единиц тепла, помещенных в точке $x = 0$ в момент времени $t = 0$. Если то же количество тепла помещается в точку x_0 в момент времени τ , то последующее тепловое поле будет описываться функцией $\varphi(x, t)$ с заменой x на $x - x_0$ и t на $t - \tau$, т.е. функцией $\varphi(x - x_0, t - \tau)$.

Оказывается, что с помощью функции $\varphi(x, t)$ можно найти тепловое поле $T(x, t)$ для любого начального распределения температуры $T_0(x)$. В силу этого частное решение $\varphi(x, t)$ названо фундаментальным. Для того чтобы это установить, прибегнем к принципу суперпозиции. Однако прежде рассмотрим описываемое этой функцией случайное блуждание.

Плотность вероятности случайно блуждающей частицы – абсолютно пьяного – со временем меняется, как показано на рис. 29.1. Спрашивается, как далеко уйдет абсолютно пьяный от начальной точки $x = 0$ за время t ?

Средняя удаленность (математическое ожидание x) равна нулю, потому что возможные уходы влево и вправо уравновешивают друг друга, поэтому следует прибегнуть к математическому ожиданию $|x|$ или $\sqrt{Mx^2}$. Та и другая величины легко вычисляются и равны

$$M|x| = a\sqrt{\frac{t}{\pi}}, \quad \sqrt{Mx^2} = a\sqrt{2t}, \quad (29.11)$$

т.е. в обоих случаях «удаление» пропорционально \sqrt{t} , в то время как при равномерном движении удаление пропорционально t . Поэтому даже очень медленно движущаяся частица уходит дальше, чем блуждающая бесконечно быстро случайная. Вместе с тем, с очень маленькой вероятностью спустя любое время, в том числе и очень малое, она может оказаться сколь угодно далеко от начального положения.

Растекание тепла и диффузия вещества происходят так же, как и растекание плотности вероятностей случайно блуждающей частицы. Интересно отметить, что это растекание несколько необычно. За конечное и даже малое время температура повышается сколь угодно далеко. Это повышение ничтожно, но оно есть, т.е. тепло распространяется бесконечно быстро. То же самое относится и к диффузии. Эта несурзница, несущественная ввиду ничтожной малости удаленных значений фундаментального решения $\varphi(x, t)$, является все же некоторым принципиальным дефектом наших моделей. Сущность дефекта понятна из соотношения (28.9), согласно которому τ и a стремятся к нулю, а величина $a^2/2\tau = a/\tau \cdot a/2$ – к конечному числу D , и поэтому скорость перемещения случайной частицы a/τ неограниченно растет.

Вернемся к роли фундаментального решения $\varphi(x, t)$ и принципа суперпозиции в решении общей задачи о тепловом поле $T(x, t)$ при начальном его значении $T_0(x)$.

Из принципа суперпозиции следует, что если отдельно для начальных полей $T_{01}(x)$ и $T_{02}(x)$ температурные поля есть $T_1(x, t)$ и $T_2(x, t)$, то при начальном поле температур $c_1T_{01}(x) + c_2T_{02}(x)$ последующее температурное поле будет $c_1T_1(x, t) + c_2T_2(x, t)$.

Начальное поле температур $T_0(x)$ можно записать в виде:

$$T_0(x) = \int_{-\infty}^{\infty} T_0(\xi)\delta(x - \xi)d\xi \quad (29.12)$$

и воспринимать его как сумму полей $\delta(x - \xi)T_0(\xi)d\xi$ при всевозможных ξ , каждому из слагаемых которой спустя время t отвечает температурное поле $\varphi(x - \xi)T_0(\xi)d\xi$.

В соответствии с этим, согласно принципу суперпозиции,

$$T(x, t) = \int_{-\infty}^{\infty} T_0(\xi)\varphi(x - \xi, t)d\xi. \quad (29.13)$$

Это и есть решение задачи (29.1), (29.2). При желании, с пользой для себя, вы можете проверить это путем подстановки (29.13) в (29.1) и (29.2).

Напишем еще раз полученное решение (29.13), выписав конкретный вид фундаментального решения

$$T(x, t) = \frac{1}{2a\sqrt{\pi t}} \int_{-\infty}^{\infty} T_0(\xi) \exp\left(-\frac{(x - \xi)^2}{4a^2 t}\right) d\xi. \quad (29.14)$$

В двумерном и трехмерном случае искомые решения запишутся аналогично. Так, в трехмерном случае

$$T(x, y, z, t) = \int_{-\infty}^{\infty} T_0(\xi, \eta, \zeta) \varphi(x - \xi, t) \varphi(y - \eta, t) \varphi(z - \zeta, t) d\xi d\eta d\zeta =$$

$$= \frac{1}{8a^3 (\sqrt{\pi t})^3} \int_{-\infty}^{\infty} T_0(\xi, \eta, \zeta) \exp\left(-\frac{(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2}{4a^2 t}\right) d\xi d\eta d\zeta,$$

где, как и ранее, $T_0(x, y, z)$ – трехмерное начальное при $t = 0$ поле температур.

Этим возможности фундаментального решения не ограничиваются. С его помощью можно решить и более общую задачу, когда наряду с начальным полем температур имеется распределенное тепловыделение или поглощение тепла. Первое можно представить как горение или экзотермическую химическую реакцию в теплопроводящей среде, а второе – как таяние мелких льдинок или эндотермическую химическую реакцию в ней. Мы получим решение этой новой задачи, даже не выписывая уравнения в частных производных, которое ей соответствует.

Пусть выделяющееся тепло за время $d\tau$ в каждый момент времени τ создает некоторое дополнительное распределение тепла $cf(x, \tau)d\tau$. Ему, как нам уже известно, отвечает решение (29.13) или (29.14), в котором $T_0(\xi)$ следует заменить на $cf(\xi, \tau)d\tau$, а t на $t - \tau$, т.е.

$$\int_{-\infty}^{\infty} f(\xi, \tau) d\tau \varphi(x - \xi, t - \tau) d\xi.$$

Ко времени t все эти тепловые поля просуммируются, т.е.

$$T(x, t) = \int_0^t \int_{-\infty}^{\infty} f(\xi, \tau) \varphi(x - \xi, t - \tau) d\xi d\tau. \quad (29.15)$$

Это, конечно, при начальном поле температур $T_0(x) = 0$. Если $T_0(x) \neq 0$, то следует к решению (29.15) добавить еще (29.13), так что в итоге

$$T(x, t) = \int_{-\infty}^{\infty} T_0(\xi) \varphi(x - \xi, t) d\xi + \int_0^t \int_{-\infty}^{\infty} f(\xi, \tau) \varphi(x - \xi, t - \tau) d\xi d\tau. \quad (29.16)$$

Это есть решение уравнения (29.1), правая часть которого дополнена слагаемым $f(x, t)$, соответствующим непрерывному тепловыделению, при условии (29.2). Вот так изящно с помощью фундаментального решения находятся решения задач о теплопроводности в неограниченной теплопроводящей среде. Изящество прежде всего в том, что решения сложные, а находятся легко, без выкладок, только на основе принципа суперпозиции. Но надо было догадаться, что нужно найти решение $\varphi(x, t)$, отвечающее точечному распределению тепла.

Теперь рассмотрим задачу о прогреве полупространства, в решении которой нам опять поможет фундаментальное решение $\varphi(x, t)$. Эта задача формулируется так. С одной стороны пространства температура все время равна T_1 , а с другой – находится теплопроводящая среда, имеющая в начальный момент температуру T_0 . Спрашивается, как будет меняться температура этого последнего полупространства?

Выберем начало координат O на границе сред и направим ось Ox вглубь теплопроводящей среды (рис. 29.2). Тогда ее температурное поле будет зависеть

только от x и, следовательно, описывается функцией $T(x, t)$. Эта функция $T(x, t)$ удовлетворяет уравнению теплопроводности

$$\frac{\partial T}{\partial t} = a^2 \frac{\partial^2 T}{\partial x^2}, \tag{29.17}$$

начальным условиям

$$T(x, t) \Big|_{t=0} = T_0 \tag{29.18}$$

и, кроме того, граничному условию

$$T(x, t) \Big|_{x=0} = T_1, \tag{29.19}$$

так как по условиям задачи в полупространстве $x \leq 0$ поддерживается температура T_1 . Новым по сравнению с предыдущим является наличие границы области, в которой рассматривается тепловое поле, и в связи с этим появление нового дополнительного условия – граничного условия (29.19).

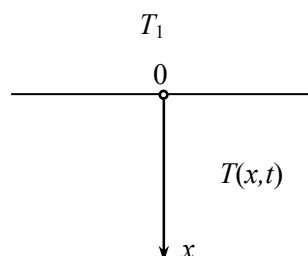


Рис. 29.2. Иллюстрация к задаче о прогреве бесконечно толстой стенки

Заметим, что температуру T_0 можно считать равной нулю. При этом T_1 заменится на $T_1 - T_0$. А можно этого и не делать, а просто искать тепловое поле в виде

$$T(x, t) = T_0 + \bar{T}(x, t),$$

где $\bar{T}(x, t)$ удовлетворяет условиям (29.17), (29.18) и (29.19), в которых T_0 и T_1 заменяются соответственно нулем и $T_1 - T_0$.

Решая эту задачу, можно, наверное, исходить только из чисто математических соображений. Но мне хочется показать вам на этом примере, как могут помочь и подсказать решение физические соображения.

Пусть ради определенности $T_1 > T_0$ и, следовательно, происходит прогрев полупространства $x > 0$. Этот прогрев происходит за счет того, что через граничную точку $x = 0$ (через единицу площади границы раздела $x = 0$) перетекает тепло с интенсивностью $q(t)$. Эта функция $q(t)$ нам неизвестна, но через нее с помощью фундаментального решения можно записать $T(x, t)$ в виде

$$T(x, t) = T_0 + \int_0^t q(\tau) \varphi(x, t - \tau) d\tau. \tag{29.20}$$

Функция $T(x, t)$, определяемая формулой (29.20), удовлетворяет (29.17) и (29.18) при любой функции $q(\tau)$. Остается удовлетворить еще только требованию (29.19), т.е.

$$T_1 = T_0 + \int_0^t q(\tau) \varphi(0, t - \tau) d\tau$$

или

$$T_1 = T_0 + \int_0^t \frac{q(\tau)}{2a\sqrt{\pi(t-\tau)}} d\tau. \quad (29.21)$$

Вводя новую переменную $u = \tau/t$, запишем (29.21) в виде

$$T_1 - T_0 = \frac{1}{2a\sqrt{\pi}} \int_0^t \frac{q(tu)\sqrt{t}}{\sqrt{1-u}} du. \quad (29.22)$$

Для выполнения этого последнего условия, во всяком случае, следует избавиться от зависимости подынтегральной функции от t . Для этого следует выбрать

$$q(tu) = \frac{c}{\sqrt{tu}}, \quad (29.23)$$

после чего (29.22) принимает вид:

$$T_1 - T_0 = \frac{1}{2a\sqrt{\pi}} \int_0^t \frac{du}{\sqrt{u(1-u)}} = \frac{\sqrt{\pi}}{2a} c.$$

Отсюда находим, что

$$c = \frac{2a(T_1 - T_0)}{\sqrt{\pi}}, \quad (29.24)$$

и искомое решение, согласно (29.22), (29.23) и (29.24), принимает вид:

$$T(x, t) = T_0 + \frac{T_1 - T_0}{\pi} \int_0^t \frac{d\tau}{\sqrt{\tau(1-\tau)}} \exp\left(-\frac{x^2}{4a^2(t-\tau)}\right). \quad (29.25)$$

Исследование этого решения обнаруживает, что оно является функцией t/x^2 . Это означает, в частности, что для того чтобы на глубине $2x$ установилась та же температура, какая уже есть на глубине x , необходимо не в два, а в четыре раза больше времени.

Устанавливается сделанное выше общее утверждение преобразованием после замены переменной τ на $\eta = \tau/x^2$ выражения (29.25) к виду:

$$T(x, t) = T_0 + \frac{T_1 - T_0}{\pi} \int_0^{t/x^2} \frac{d\eta}{\sqrt{\eta(t/x^2 - \eta)}} \exp\left(\frac{1}{4a^2(t/x^2 - \eta)}\right) = T_0 + F(t/x^2). \quad (29.26)$$

Вид графика этого решения показан на рис. 29.3.

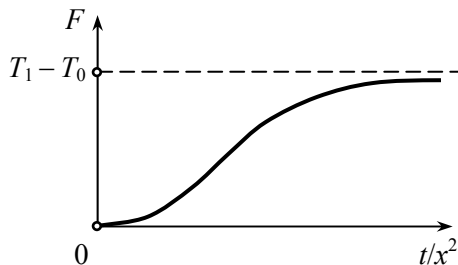


Рис. 29.3. График изменения температуры в прогреваемой толстой стене в идеализации полупространством

Такой его вид следует из монотонности возрастания F как функции аргумента t/x^2 и из того, что при $t/x^2 = 0$ $T(x, t) = T_0$, а при $t/x^2 = \infty$ $T(x, t) = T_1$.

30. БЕГУЩИЕ ВОЛНЫ И ДИСПЕРСИОННОЕ УРАВНЕНИЕ

Уравнение линейного осциллятора

$$\ddot{x} + 2\delta\dot{x} + \omega^2 x = 0$$

имеет частные решения вида $e^{\lambda\tau}$, где значения λ находятся из характеристического уравнения

$$\chi(\lambda) = \lambda^2 + 2\delta\lambda + \omega^2 = 0.$$

Аналогично для уравнения теплопроводности [19]

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (30.1)$$

и волнового уравнения

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (30.2)$$

имеются решения вида $e^{i(\omega t - kx)}$, где возможные значения ω и k находятся из дисперсионного уравнения

$$\chi(\omega, k) = 0. \quad (30.3)$$

Подставляя в (30.1) предполагаемое решение

$$u(x, t) = e^{i(\omega t - kx)}, \quad (30.4)$$

приходим к дисперсионному уравнению

$$i\omega + a^2 k^2 = 0. \quad (30.5)$$

Для уравнения (30.2) аналогичная подстановка дает

$$i\omega - a^2 k^2 = 0. \quad (30.6)$$

Выражая k через ω , согласно (30.5) и соответственно (30.6), находим, что уравнение (30.1) имеет всевозможные решения вида

$$u(x, t) = \exp i \left(\omega t \pm \frac{\sqrt{\omega}}{a} \frac{1-i}{\sqrt{2}} x \right), \quad (30.7)$$

а уравнение (30.2) – решения вида

$$u(x, t) = \exp i \left(\omega t \pm \frac{\omega}{a} x \right). \quad (30.8)$$

Выясним, что представляют собою решения (30.7) и (30.8), т.е. решения вида

$$u(x, t) = \exp i(\omega t + kx), \quad (30.9)$$

где ω действительно, а k может быть комплексным.

Пусть поначалу ω и k – действительные. Тогда

$$\exp i(\omega t + kx) = \cos(\omega t + kx) + i \sin(\omega t + kx)$$

и, следовательно, действительными решениями являются функции

$$\cos(\omega t + kx), \quad \sin(\omega t + kx). \quad (30.10)$$

При $t = 0$ этим решениям отвечают переменные, гармонически меняющиеся поля с длиной волны $\lambda = 2\pi/|k|$ (рис. 30.1). С ростом времени t это поле как целое перемещается вдоль оси x со скоростью $v = -\omega/k$. Так что (30.10) является, а (30.9) можно трактовать как бегущую со скоростью $v = -\omega/k$ гармоническую волну длины $\lambda = 2\pi/|k|$. При этом в каждой точке x , как это видно из (30.10), поле при изменении времени t меняется тоже по гармоническому закону с частотой ω и единичной амплитудой. В силу этого величины ω и k названы частотой и волновым числом бегущей гармонической волны (30.10) ((30.9)).

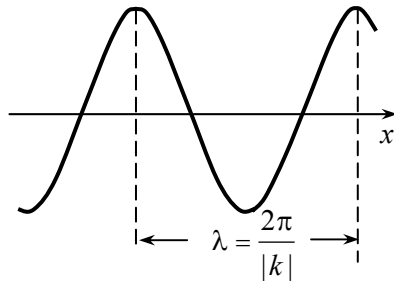


Рис. 30.1. Гармонически меняющееся одномерное скалярное поле

Пусть теперь волновое число k – комплексное, т.е. $k = k_1 + ik_2$. Тогда

$$\exp i(\omega t + kx) = \exp i(\omega t + k_1 x) e^{-k_2 x} = e^{-k_2 x} \cos(\omega t + k_1 x) + ie^{-k_2 x} \sin(\omega t + k_1 x),$$

и, следовательно, действительными решениями являются

$$e^{-k_2 x} \cos(\omega t + k_1 x), \quad e^{-k_2 x} \sin(\omega t + k_1 x).$$

Если бы $k_2 = 0$, то это была бы гармоническая бегущая волна. Наличие множителя $e^{-k_2 x}$ при $k_2 > 0$ приводит к ее экспоненциальному убыванию в направлении положительных x или, напротив, экспоненциальному возрастанию при $k_2 < 0$. Сама волна перемещается направо, если $k_1 < 0$, и налево, если $k_1 > 0$. Скорость этого перемещения v , как и в предыдущем случае, равна $-\omega/k_1$, так что при $k_1 < 0$ и $k_2 > 0$ – это убывающая бегущая направо волна (рис. 30.2); при $k_1 < 0$ и $k_2 < 0$ – это возрастающая бегущая направо волна (рис. 30.3). Аналогично при $k_1 > 0$ меем бегущую налево волну, убывающую при $k_2 < 0$ и возрастающую при $k_2 > 0$. Во всех случаях в каждой точке x поле меняется по гармоническому закону с частотой ω и амплитудой $e^{-k_2 x}$.

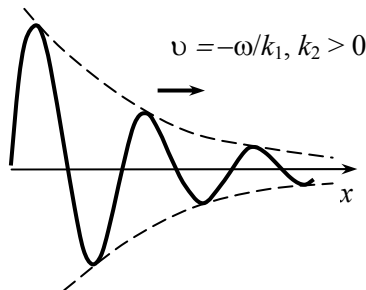


Рис. 30.2. Бегущая затухающая гармоническая волна

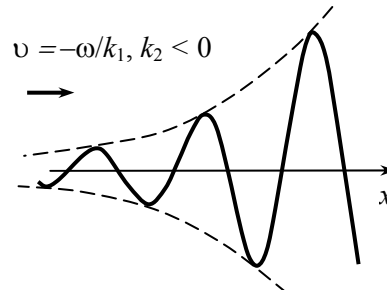


Рис. 30.3. Бегущая нарастающая гармоническая волна

Волновое уравнение (30.2), согласно (30.8), допускает всевозможные бегущие гармонические волны с всевозможными частотами и волновыми числами ω и $\pm\omega/a$ соответственно. Скорость распространения у всех этих волн одна и та же: $-a$ и, соответственно, $+a$. Длины волн $2\pi a/\omega$ разные, как и разные соответствующие им частоты.

В силу принципа суперпозиции, наряду с всевозможными бегущими гармоническими волнами (30.8), имеют место и всевозможные бегущие волны вида

$$\sum_s c_s \cos \left[\omega_s \left(t \pm \frac{x}{a} \right) + \varphi_s \right]$$

и вида

$$\int_0^{\infty} c(\omega) \cos \left[\omega \left(t \pm \frac{x}{a} \right) + \varphi(\omega) \right] d\omega.$$

Это наталкивает на мысль, что волновое уравнение допускает произвольные бегущие со скоростью $\pm a$ волны

$$u(x, t) = f \left(t \pm \frac{1}{a} x \right) \quad (30.11)$$

любой формы. Так это на самом деле и есть. Это легко проверить, непосредственно убеждаясь, что при любой функции f уравнение (30.11) удовлетворяет волновому уравнению (30.2).

Если все бегущие волны волнового уравнения распространяются в обе стороны с одной и той же скоростью a , не затухают и не возрастают, то волновые решения (30.7) уравнения теплопроводности имеют разные скорости распространения и экспоненциально уменьшаются. Согласно (30.7), действительными волновыми решениями являются переменные поля вида

$$u(x, t) = \exp \frac{\sqrt{\omega}}{a\sqrt{2}} x \cos \left(\omega t + \frac{\sqrt{\omega}}{a\sqrt{2}} x \right) \quad (30.12)$$

и вида

$$u(x, t) = \exp \left(-\frac{\sqrt{\omega}}{a\sqrt{2}} x \right) \cos \left(\omega t - \frac{\sqrt{\omega}}{a\sqrt{2}} x \right) \quad (30.13)$$

при разных неотрицательных ω .

Поле, описываемое формулой (30.12), – затухающая волна, бегущая налево, а поле (30.13) – бегущая направо затухающая волна. При этом как быстрота затухания, так и скорость движения волны зависят от ее частоты ω . Скорость ее перемещения равна $a\sqrt{2\omega}$ и тем больше, чем больше ее частота ω или меньше длина волны $\lambda = 2\sqrt{2}\pi a/\sqrt{\omega}$. Вид этих бегущих волн показан на рис. 30.4. На этом же рисунке указаны скорость распространения, длина волны и быстрота ее затухания.

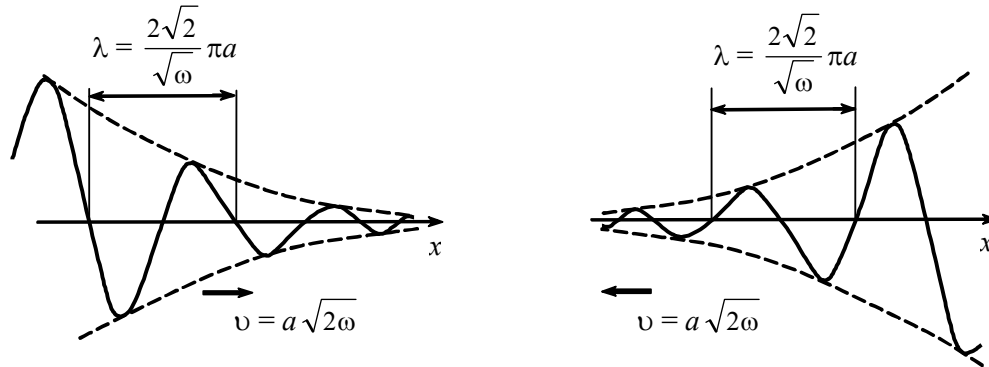


Рис. 30.4. Бегущие гармонические тепловые волны; ω – временная частота волны, λ – ее пространственная длина, v – скорость распространения

Применим теперь полученные нами сведения к следующим двум похожим задачам для волнового уравнения и уравнения теплопроводности. В качестве физической модели первой задачи возьмем полубесконечную веревку, конец которой двигают вверх и вниз по закону $A \cos \Omega t$. В качестве физической модели второй задачи – полупространство, на границе которого температура меняется по тому же закону.

Я думаю, вы, наверное, не раз в детстве забавлялись пусканием волн по веревке, быстро колебля ее конец в руке. Так же природа со сменой дня и ночи или зимы и лета меняет температуру над земной поверхностью. Мы придем к сформулированным выше задачам, если примем, что движение руки и температура меняются по гармоническому закону. Конечно, это грубое приближение. Но оно достаточно для понимания происходящих при этом явлений.

Построим математические модели в первом и втором случаях. Примем, что веревку начали колебать уже очень давно, как и очень давно меняют температуру над земной поверхностью.

В первом случае движение натянутой полубесконечной веревки описывается волновым уравнением

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (30.14)$$

при $-\infty < t < +\infty$ и $x \geq 0$. При $x = 0$ и всех t имеет место

$$u(x, t) \Big|_{x=0} = A \cos \Omega t. \quad (30.15)$$

Во втором примере тепловое поле подчиняется уравнению теплопроводности

$$\frac{\partial T}{\partial t} = a^2 \frac{\partial^2 T}{\partial x^2} \quad (30.16)$$

тоже при $-\infty < t < +\infty$ и $x \geq 0$, а при $x = 0$ и всех t

$$T(x, t) \Big|_{x=0} = A \cos \Omega t. \quad (30.17)$$

Как видите, постановки обеих задач очень похожи. В первом случае вы хорошо знаете и много раз видели, что будет. Это позволяет нам, сочетая виденное с тем, что мы узнали о гармонических волнах, сразу сообразить, какое будет решение. Это будет гармоническая волна частоты ω , бегущая от руки вдоль веревки, то есть

$$u(x, t) = A \cos \Omega \left(t - \frac{x}{a} \right). \quad (30.18)$$

Она удовлетворяет уравнению (30.14) и краевому условию (30.15).

По аналогии решение второй задачи запишем в виде тепловой волны, бегущей от поверхности Земли вглубь и имеющей частоту Ω , т.е. согласно (30.13),

$$T(x, t) = A \exp \left(-\frac{\sqrt{\Omega}}{a\sqrt{2}} x \right) \cos \left(\Omega t - \frac{\sqrt{\Omega}}{a\sqrt{2}} x \right). \quad (30.19)$$

Ясно, что функция $T(x, t)$, определяемая (30.19), удовлетворяет уравнению теплопроводности (30.16) и краевому условию (30.17).

Решение (30.18), я думаю, вам хорошо известно из наблюдения. Хотя, если бы веревка была достаточно длинная, то вы бы заметили, что бегущая волна затухает. Это уже дефект нашей идеализации: наличие трения в веревке при ее деформациях и трение ее о воздух мы не учли. При желании это можно сделать, считая, например, что сила трения пропорциональна скорости движения веревки и противоположно направлена. Это привело бы к уравнению вида

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} - h \frac{\partial u}{\partial t}. \quad (30.20)$$

Далее можно найти волновое решение вида (30.4) и, используя его, уточнить полученное нами решение (30.18). Действительно, подставляя (30.4) в (30.20), найдем, что

$$\omega^2 - a^2 k^2 - ih\omega = 0$$

или

$$k = \frac{\omega}{a} \sqrt{1 - i \frac{h}{\omega}}.$$

При $h/\omega \ll 1$ приближенно

$$k = \pm \frac{\omega}{a} \left(1 - i \frac{h}{2\omega} \right).$$

Нас интересует волна, бегущая от руки, поэтому в полученном значении волнового числа k берем знак минус. Соответствующее уточненное решение будет

$$u(x, t) = A \exp \left(-\frac{hx}{2a} \right) \cos \left(\Omega t - \frac{\Omega}{a} x \right). \quad (30.21)$$

Оно действительно затухает с ростом x и при $h = 0$ переходит, как и положено, в (30.18).

Бегущая вдоль веревки волна может привести к неожиданному эффекту, если веревка утончается. Это утончение вызывает рост скорости распространения волны a , так как $a^2 = T/\rho$, где T – натяжение веревки, а ρ – ее линейная плотность. При этом растет не только скорость, но и амплитуда колебаний, волна нарастает. В результате может возникнуть сильный хлопок, вызванный тем, что скорость движения к концу веревки превзойдет звуковую. Чтобы осмыслить это явление, учтем, что волна сохраняет энергию, а колеблющаяся масса уменьшается. Наблю-

дать это явление можно на пастбище, когда пастухи хлопают утончающимися к концу длинными кнутами.

Полученное нами решение (30.19) качественно описывает прогрев земной поверхности при суточных и годовых колебаниях температуры. При этом предполагалось, что наружная температура меняется согласно (30.17), колеблясь возле среднего значения, равного нулю. Если считать среднее значение температуры отличным от нуля и равным T_0 , то и к решению (30.19) добавится T_0 и оно примет вид:

$$T(x, t) = T_0 + A \exp\left(-\frac{\sqrt{\Omega}}{a\sqrt{2}}x\right) \cos\left(\Omega t - \frac{\sqrt{\Omega}}{a\sqrt{2}}x\right). \quad (30.22)$$

Проанализируем это решение. Оно состоит из постоянной составляющей T_0 и экспоненциально убывающей бегущей волны. Экспоненциальное убывание для суточных колебаний много больше, чем для годовых. Амплитуда колебаний температуры убывает в e раз на глубине $a\sqrt{2}/\sqrt{\Omega}$ и поэтому, если для годовых колебаний она порядка метра, для суточных она в $\sqrt{\Omega_{\text{год}}}/\sqrt{\Omega_{\text{сут}}} \approx 20$ раз меньше, т.е. порядка сантиметров. Отсюда следует, что на глубине нескольких метров температура практически постоянна и равна средней температуре. Это объясняет наличие вечной мерзлоты на глубине, если наружная средняя температура ниже нулевой.

Длина тепловой волны λ равна $a\pi\sqrt{2}/\sqrt{\Omega}$ и значительно больше для годовых колебаний, чем для суточных, примерно в двадцать раз. Минимальное и максимальное значения температуры на поверхности смещаются вглубь одно за другим со скоростью распространения волны v , равной $a\sqrt{2\Omega}$. Поэтому на глубине x максимум и минимум температур наступают позднее на время $x/a\sqrt{2\Omega}$, чем на поверхности. Это объясняет тот факт, что если в холодную зиму глубоко зарытые в землю водопроводные трубы замерзают, то происходит это не в самые сильные морозы, а значительно позднее, возможно, весной при значительном потеплении. Суточные волны температуры распространяются вглубь быстрее, чем годовые (примерно в 20 раз), а затухают они значительно быстрее, по экспоненте с показателем тоже примерно в 20 раз большим и, конечно, отрицательным.

Мы рассмотрели задачи о теплопроводности, которые можно трактовать, как прогрев земной поверхности или толстой стенки здания, а также как изменение температуры в глубине земной поверхности при суточных и годовых изменениях ее на поверхности.

Рассмотрим еще, как происходит намерзание льда над водой в озере или реке. Вода начинает замерзать, когда она охлаждается до нуля и продолжает охлаждаться со своей поверхности. На рис. 30.5 схематически представлена толща льда и под ним еще не замерзшая вода при нулевой температуре. На поверхности льда температура отрицательная T_1 , толщина льда ξ . Нас интересует, как растет со временем ξ . Пусть температура льда $T(x, t)$. Для нее имеет место уравнение теплопроводности

$$\frac{\partial T}{\partial t} = a^2 \frac{\partial^2 T}{\partial x^2} \quad (30.23)$$

и краевые условия

$$T(x, t) \Big|_{x=0} = T_1, \quad T(x, t) \Big|_{x=\xi} = 0, \quad (30.24)$$

означающие, что на верхней и нижней поверхностях льда его температура соответственно T_1 и 0.

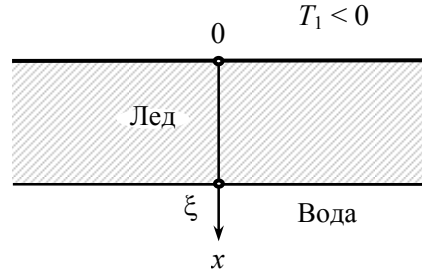


Рис. 30.5. Иллюстрация к намерзанию льда на поверхности воды

Поток тепла через нижнюю поверхность льда от более нагретой воды к более холодному льду равен

$$Q = k \frac{\partial T}{\partial x} \Big|_{x=\xi}. \quad (30.25)$$

В соответствии с этим оттоком тепла происходит образование нового льда, причем

$$k \frac{\partial T(x, t)}{\partial x} \Big|_{x=\xi} = \rho c \frac{d\xi}{dt}, \quad (30.26)$$

где c – тепло, выделяющееся при замерзании единицы массы льда, а ρ – его плотность. Такова наша исходная математическая модель. Она достаточно сложна. Сложность эта обусловлена тем, что в краевое условие (30.25) входит неизвестная переменная ξ , меняющаяся в соответствии с уравнением (30.26), которое само зависит от неизвестного ξ .

Учтем особенности задачи и упростим ее математическую модель. Образование льда происходит очень медленно, так что можно считать, что в толще льда температура квазистационарна, т.е. такая, которая устанавливается при постоянном ξ .

При постоянном ξ установившееся, не зависящее от времени температурное поле $T(x, t) = u^*(x)$ удовлетворяет уравнению

$$\frac{d^2 u^*}{dx^2} = 0, \quad (30.27)$$

следующему из (30.23), и краевым условиям

$$u^*_{x=0} = T_1, \quad u^*_{x=\xi} = 0, \quad (30.28)$$

следующим из (30.24). Из (30.27), (30.28) непосредственно находим, что

$$u^* = -\frac{T_1}{\xi} x + T_1,$$

и поэтому уравнение (30.25) принимает вид

$$\rho c \frac{d\xi}{dt} = -k \frac{T_1}{\xi}. \quad (30.29)$$

Откуда, зная, что при $t = 0$ $\xi = 0$, получим

$$\xi = \sqrt{-\frac{2kT_1}{\rho c}t}, \quad (30.30)$$

что означает увеличение толщи льда по закону \sqrt{t} , т.е. замедляясь со временем. Так, если один сантиметр льда возникает за сутки, то 6 сантиметров, когда по льду можно ходить, не за шесть, а за 36 суток. Зависимость от наружной температуры тоже замедленная, пропорциональная $\sqrt{-T_1}$. Поэтому при температуре в два раза более низкой скорость образования льда увеличивается только в $\sqrt{2}$ раз.

Конечно, реально все немного не так, поскольку температура T_1 не постоянна, а меняется со временем. Если эти изменения температуры не очень быстрые и в толще льда успевают все время устанавливаться тепловое поле, отвечающее температуре $T_1(t)$, то можно по-прежнему воспользоваться уравнением (30.29), но считая в нем T_1 функцией времени. Из этого уравнения (30.29) находим, что

$$\rho c \xi d\xi = -k T_1(t) dt$$

или

$$\frac{\rho c}{2} \xi^2 = -k \int_0^t T_1(\tau) d\tau,$$

откуда

$$\xi = \sqrt{\frac{2k}{\rho c} \left(-\int_0^t T_1(\tau) d\tau \right)^{1/2}}. \quad (30.31)$$

Из приближенной формулы (30.31) следует, что в пределах ее погрешностей толщина намерзшего льда за некоторое время t определяется той же формулой (30.30), где уже T_1 – средняя температура над поверхностью льда за время t . Так что толщина льда, намерзшего за зиму, пропорциональна корню квадратному от средней зимней температуры и корню квадратному из ее длительности.

31. ТЕОРИЯ ЭЛЕКТРОМАГНЕТИЗМА ФАРАДЕЯ – МАКСВЕЛЛА И ЭЛЕКТРОМАГНИТНЫЕ ВОЛНЫ МАКСВЕЛЛА – ГЕРЦА

Ранее вы познакомились с великими уравнениями Максвелла. Они достаточно сложны и в них не просто разобраться. Современникам они были непонятны и лишь немногие их приняли всерьез. Их освоение, даже в современной более ясной и простой векторной форме написания и продвинутом общем уровне мышления, требует больших усилий мысли, освоения новых понятий и представлений. В первоначальном же виде у Максвелла они были записаны в кватернионной форме и содержали затемняющие суть дела лишние уравнения. Электрические и магнитные поля существовали до Максвелла отдельно. Они мгновенно порождались соответственно зарядами и токами, которые считались первопричинами электромагнитных явлений. Уравнения Максвелла описывали нечто совершенно другое и новое. Они описывали электромагнитные поля, которые, меняясь и взаимодействуя друг с другом, могут существовать и распространяются сами по себе в пространстве с конечной скоростью. Из уравнений Максвелла можно было извлечь волновое уравнение, говорящее о существовании гармонической бегущей электромагнитной волны. Ее скорость оказалась близкой к скорости света, что позволяло Максвеллу высказать гипотезу об электромагнитной природе световых волн.

Теория Максвелла, английского физика, противоречила существующей европейской, с нашей точки зрения, квазистационарной электродинамике, хорошо математически разработанной и, казалось, полностью согласующейся с экспериментом. Приверженец европейской теории Герц решил опровергнуть заумную теорию Максвелла, поставив основательные и хорошо продуманные эксперименты. По существу, речь шла об отстаивании европейской теории электромагнетизма, успешно развиваемой такими гигантами науки, как Ампер, Лаплас, Гаусс, Нейман, Вебер, Гельмгольц, против «вздорной» теории Фарадея–Максвелла. Первые шли от зарядов и токов, от описания электромагнитных явлений на их основе. И, как мы сейчас знаем, их ждал тупик. Заряды и токи не позволяют во всех случаях построить состояние электродинамической системы и поэтому понять, изучить и предсказать ее поведение. Для конструирования состояния нужны принципиально новые полевые представления, нужно электромагнитное поле, для которого заряды и токи – лишь его особенности. В теории Фарадея–Максвелла электромагнитные явления могут протекать без зарядов и токов. Состоянием динамической системы электродинамики Фарадея–Максвелла являются электромагнитные поля. И если в механике сплошных сред поля смещений и скоростей возникали естественно и не были сопряжены с новым взлетом мысли, то в электродинамике это не так. В ней конструируются поля совершенно непонятной природы, чуждой человеческому восприятию.

Этот скачок человеческой мысли осуществил М. Фарадей. Именно он придумал электрические и магнитные поля и с их помощью описывал непонятные электромагнитные явления. Максвелл уловил в гений самоучки Фарадея и, как он

сам пишет, лишь придал его представлениям математическое описание. И именно этот безумный скачок мысли был необходим для создания адекватной математической модели.

Результат экспериментов Герца был неожиданным, он противоречил теории, разделяемой Герцем и любезной ему, и полностью соответствовал теории Максвелла. Несмотря на это, Герц опубликовал результаты своих экспериментов и в связи с ними развил теорию Максвелла, открыв излучатель электромагнитных волн, получивший название вибратора Герца. Тем самым вместе с Максвеллом он стал у истоков возникновения радио, величайшего технического достижения начала XX века. Эксперимент Герца подтвердил теоретические открытия Максвелла, и уже ничто не препятствовало триумфальному шествию уравнений Максвелла.

Сейчас очень непросто воспроизвести поиски Максвелла. Даже если это сделать, они будут едва ли понятны. Поэтому в плане общего рассказа о распределенных динамических системах и бегущих гармонических волнах покажем, как из уравнений Максвелла следует их наличие, следует возможность электромагнитных волн. Итак, наша задача выяснить, допускают ли уравнения Максвелла решение в виде бегущих волн, когда вектора E и H являются только функциями двух переменных, времени t и, допустим, переменной x , электромагнитной волны, существующей самостоятельно, при отсутствии зарядов и токов, в пустом пространстве, т.е. при $\epsilon = \mu = 1$ и $\sigma = 0$. В этих условиях уравнения Максвелла (28.19), (28.20) и (28.21) упрощаются и принимают вид:

$$\begin{aligned} \frac{\partial E}{\partial t} = c \operatorname{rot} H, \quad \frac{\partial H}{\partial t} = -c \operatorname{rot} E, \\ \operatorname{div} E = 0, \quad \operatorname{div} H = 0, \end{aligned} \quad (31.1)$$

где $E = E(x, t)$ и $H = H(x, t)$. Заметим, что, не упрощая задачу поиском сразу плоской волны, можно обнаружить, что из уравнений (31.1) следуют волновые уравнения для полей E и H . Действительно, продифференцируем первое уравнение (31.1) по t , а к второму применим операцию rot :

$$\frac{\partial^2 E}{\partial t^2} = c \operatorname{rot} \frac{\partial H}{\partial t}, \quad \operatorname{rot} \frac{\partial H}{\partial t} = -c \operatorname{rot} \operatorname{rot} E,$$

откуда, пользуясь известной формулой векторного анализа $\operatorname{rot} \operatorname{rot} = \nabla \operatorname{div} - \Delta$, запишем

$$\frac{\partial^2 E}{\partial t^2} = -c \operatorname{rot} \operatorname{rot} E = -c^2 (\nabla \operatorname{div} E - \Delta E),$$

и в силу третьего уравнения (31.1) придем к волновому уравнению для векторного поля

$$\frac{\partial^2 E}{\partial t^2} = c^2 \Delta E.$$

Аналогично находится волновое уравнение для вектора магнитного поля. Далее уже можно находить согласованные уравнениями (31.1) решения этих волновых уравнений. Не ставя себе такой общей задачи, забудем об этом эффектном

использовании векторного анализа и скромно продолжим поиск решения уравнений (31.1) в виде плоской волны, распространяющейся вдоль оси x .

Итак, в уравнениях (31.1) векторы E и H – функции только x и t . Из последних двух уравнений (31.1) непосредственно следует, что

$$\frac{\partial E_x}{\partial x} = 0, \quad \frac{\partial H_x}{\partial x} = 0$$

и, следовательно, компоненты E_x и H_x векторов E и H не зависят от x . Поэтому они нас не интересуют, и мы примем $E_x = H_x = 0$. После этого распишем первые два уравнения (31.1) в компонентах:

$$\begin{aligned} \frac{\partial E_y}{\partial t} &= -c \frac{\partial H_z}{\partial x}, & \frac{\partial E_z}{\partial t} &= c \frac{\partial H_y}{\partial x}, \\ \frac{\partial H_y}{\partial t} &= c \frac{\partial E_z}{\partial x}, & \frac{\partial H_z}{\partial t} &= -c \frac{\partial E_y}{\partial x}. \end{aligned}$$

Эти уравнения разделяются на две независимые пары. Выпишем отдельно одну из них:

$$\frac{\partial E_y}{\partial t} = -c \frac{\partial H_z}{\partial x}, \quad \frac{\partial H_z}{\partial t} = -c \frac{\partial E_y}{\partial x}. \quad (31.2)$$

Дифференцируя первое уравнение по t , а второе по x и сравнивая их, приходим к волновому уравнению

$$\frac{\partial^2 E_y}{\partial t^2} = c^2 \frac{\partial^2 E_y}{\partial x^2}, \quad (31.3)$$

которое, как нам известно, имеет решение в виде бегущей гармонической волны

$$E_y = A \sin\left(\omega t - \frac{\omega x}{c}\right). \quad (31.4)$$

Далее из любого из уравнений (31.2) находим, что

$$H_z = A \sin\left(\omega t - \frac{\omega x}{c}\right). \quad (31.5)$$

Полученное частное решение (31.3), (31.4) удовлетворяет исходным уравнениям (31.1) и представляет собою электромагнитную волну, состоящую из двух перпендикулярных друг другу гармонических волн электрического и магнитного полей, распространяющихся вдоль оси x со скоростью c (рис. 31.1).

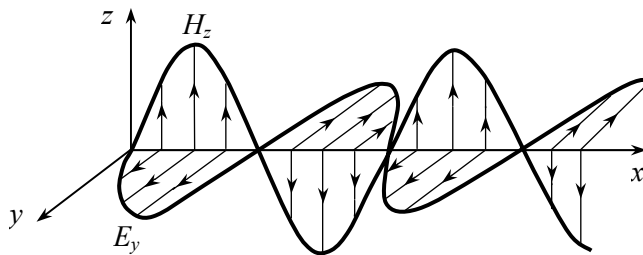


Рис. 31.1. Гармоническая электромагнитная волна

Таким образом, уравнения Максвелла при $\varepsilon = \mu = 1$ и $\sigma = 0$ имеют решение (31.4), (31.5). Остальные компоненты E_x , E_z и H_x , H_z равны нулю. Это решение – плоская электромагнитная гармоническая волна с произвольной частотой ω , волновым числом $2\pi c/\omega$ и скоростью распространения c . Так что таких разных волн много. Среди них – световые волны, радиоволны, ультрафиолетовые и инфракрасные волны, рентгеновские волны. Все эти разные волны имеют одну и ту же электромагнитную природу и различаются только частотой и длиной волны при одинаковой скорости распространения.

32. ОТРАЖЕНИЕ И ПРЕЛОМЛЕНИЕ ВОЛН

Мы познакомились с распределенными динамическими системами; с основными уравнениями классической математической физики; уравнением теплопроводности и одновременно диффузии, и случайных блужданий; волновым уравнением, описывающим колебания натянутой струны или мембраны, продольные и поперечные колебания упругой среды, изменения давления в газе, звуковые колебания, электромагнитные волны.

Новым по сравнению с дискретными системами являются волны: тепловые затухающие волны, скорость распространения которых зависит от частоты и длины волны; незатухающие волны упругой среды и электромагнитные волны, распространяющиеся с неизменной скоростью. С волнами связаны такие впечатляющие явления, как интерференция, дифракция, причудливая эволюция волнового пакета, с которой, в частности, связаны искажения речи и музыки при передаче их на большие расстояния, отражение и преломление волн и другие. Эти явления вам хорошо знакомы со школьной скамьи и курса общей физики. Но там их изложение носило описательный характер и не опиралось на математические модели. Сейчас мы их можем не только описывать, но и изучать, извлекая новые неизвестные нам количественные и качественные сведения.

В качестве еще одного простого примера рассмотрим явления отражения и преломления волн от раздела упругих сред с разными характеристиками. Пусть конкретно речь идет о бегущей вдоль натянутой струны волне вида

$$Ae^{i(\omega t - kx)}, \quad (32.1)$$

наталкивающейся на скачкообразное изменение плотности струны. Направим ось Ox вдоль струны и пусть разные плотности струна имеет при $x < 0$ и $x > 0$, т.е. точка $x = 0$ – точка раздела сред. Нам нужно найти переменное поле $u(x, t)$ смещений струны, вызываемое бегущей слева на границу раздела волной (32.1). Математическая модель должна включать волновые уравнения для частей $x < 0$ и $x > 0$ струны и условия согласования их решений при $x = 0$ и то, что слева направо по части струны $x < 0$ бежит волна вида (32.1).

В соответствии с этим пусть $u(x, t) = u_1(x, t)$ при $x < 0$ и $u(x, t) = u_2(x, t)$ при $x > 0$. Для $u_1(x, t)$ и $u_2(x, t)$ имеют место уравнения:

$$\frac{\partial^2 u_1(x, t)}{\partial t^2} = a_1^2 \frac{\partial^2 u_1}{\partial x^2} \quad (x < 0) \quad (32.2)$$

и

$$\frac{\partial^2 u_2(x, t)}{\partial t^2} = a_2^2 \frac{\partial^2 u_2}{\partial x^2} \quad (x > 0), \quad (32.3)$$

где между $u_1(x, t)$ и $u_2(x, t)$ при $x = 0$ имеют место сопряжения

$$\begin{aligned} u_1(x, t) \Big|_{x=0} &= u_2(x, t) \Big|_{x=0}, \\ \frac{\partial u_1(x, t)}{\partial x} \Big|_{x=0} &= \frac{\partial u_2(x, t)}{\partial x} \Big|_{x=0}. \end{aligned} \quad (32.4)$$

Первое соотношение (32.4) очевидно, так как в точке $x = 0$ струна не разрывается. Второе условие не столь очевидно. Оно следует из одинаковости натяжения N струны слева и справа от точки $x = 0$. Слева перпендикулярно оси Ox на точку стыковки струн действует сила $-N \sin \alpha_1$, где α_1 – угол струны слева от точки $x = 0$ с осью Ox . Аналогично справа действует сила $+N \sin \alpha_2$ (рис. 32.1). Точка струны, отвечающая $x = 0$, имеет нулевую массу, поэтому силы $-N \sin \alpha_1$ и $+N \sin \alpha_2$ уравниваются друг друга, т.е. $\alpha_1 = \alpha_2$ и, следовательно, касательные к струнам слева и справа в точке $x = 0$ совпадают. А это и означает выполнение второго из соотношений (32.4).

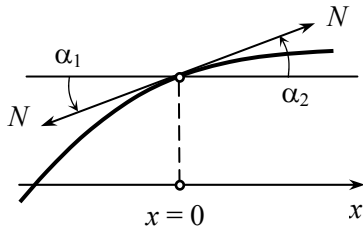


Рис. 32.1. Иллюстрация к преломлению и отражению бегущей гармонической волны на границе раздела $x = 0$ разных однородных сред

Еще нужно отразить тот факт, что слева на границу раздела приходит бегущая волна вида (32.1), в которой волновое число $k = \omega / a_1$.

Несмотря на простоту задачи, без физической подсказки поиск ее решения не представляется простым. Физическая подсказка состоит в том, что нам известно, что при этом возникают отраженная и преломленная волны. Какие они – вопрос, но естественно предположить, что такие же, как и падающая волна, бегущие гармонические волны. В соответствии с этим примем, что

$$\begin{aligned} u_1(x, t) &= A e^{i(\omega t - \omega x / a_1)} + A_1 e^{i(\omega_1 t + \omega_1 x / a_1)}, \\ u_2(x, t) &= A_2 e^{i(\omega_2 t - \omega_2 x / a_2)}, \end{aligned} \quad (32.5)$$

где A , A_1 и A_2 – амплитуды падающей, отраженной и преломленной волн, a_1 и a_2 определены уравнениями (32.2) и (32.3); A_1 , A_2 и ω_1 , ω_2 – неизвестные, подлежащие определению. Из условий сопряжения имеем:

$$\begin{aligned} A e^{i\omega t} + A_1 e^{i\omega_1 t} &= A_2 e^{i\omega_2 t}, \\ -i \frac{\omega}{a_1} A e^{i\omega t} + i \frac{\omega_1}{a_1} A_1 e^{i\omega_1 t} &= -i \frac{\omega_2}{a_2} A_2 e^{i\omega_2 t}, \end{aligned}$$

где оба соотношения выполняются при любом t .

Из этих соотношений следует, что $\omega_1 = \omega_2 = \omega$. Это можно было сообразить и сразу, ведь условия (32.4) должны соблюдаться при всех t . Принимая $\omega_1 = \omega_2 = \omega$, приходим к условиям

$$\begin{aligned} A + A_1 &= A_2, \\ -\frac{\omega}{a_1} A + \frac{\omega}{a_1} A_1 &= -\frac{\omega}{a_2} A_2, \end{aligned}$$

откуда

$$A_1 = \frac{1 - a_1 / a_2}{1 + a_1 / a_2} A, \quad A_2 = \frac{2}{1 + a_1 / a_2} A. \quad (32.6)$$

Требуемые решения $u_1(x, t)$ и $u_2(x, t)$ найдены. Рассмотрим зависимости амплитуд A_1 и A_2 отраженной и преломленной волн от a_1 и a_2 . Для этого на рис. 32.2 представлен график зависимости A_1/A и A_2/A от a_1/a_2 . Напомним, что a_1^2 и a_2^2 в волновых уравнениях (32.2) и (32.3) – это отношения общей для них силы натяжения к плотностям струны при $x < 0$ и $x > 0$. Из графика рис. 32.2 следует, что для однородной струны, когда $a_1/a_2 = 1$, как и должно быть, отраженной волны нет, а преломленная – продолжение падающей. При исчезающе малой или очень большой плотности струны в части $x > 0$ отношение a_1/a_2 равно 0 и, соответственно, ∞ . В первом случае имеет место отражение волны с той же амплитудой и фазой, во втором – с той же амплитудой, но противоположной фазой. Амплитуда преломленной волны, соответственно, удваивается, либо равна нулю. Эти крайние случаи можно рассматривать как отражение от свободного и закрепленного конца $x = 0$ струны с $x < 0$. В первом случае падающая волна отражается, изменив только направление движения, а во втором еще и перевертывается (рис. 32.3).

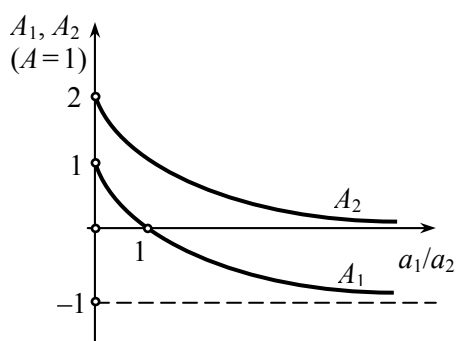


Рис. 32.2. График амплитуд A_1 и A_2 соответственно отраженной и преломленной гармонических волн

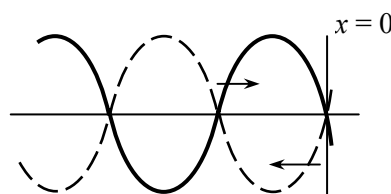


Рис. 32.3. Явление перевертывания при отражении бегущей гармонической волны от «закрепления». Преломленной волны нет

33. СТОЯЧИЕ ВОЛНЫ И КОЛЕБАНИЯ ОГРАНИЧЕННОЙ СТРУНЫ

В этой лекции вы познакомитесь с новым типом колебаний распределенных динамических систем – сплошной среды или электромагнитного поля – стоячими волнами. Они возникают в ограниченных упругих телах или объемах. Стоячая волна может возникнуть и как суперпозиция (интерференция) гармонических волн $Ae^{i(\omega t - kx)}$ и $Ae^{i(\omega t + kx)}$, бегущих навстречу друг другу, так как

$$Ae^{i(\omega t - kx)} + Ae^{i(\omega t + kx)} = 2Ae^{i\omega t} \cos kx. \quad (33.1)$$

У волны (33.1) колебания в каждой точке x совершаются с одной и той же частотой ω и фазой, но с разной амплитудой, зависящей от x и изменяющейся по закону $2A \cos kx$. В соответствии с (33.1) общий вид стоячей одномерной волны можно записать в виде

$$u(x, t) = X(x)T(t), \quad (33.2)$$

где $X(x)$ – функция только x , а $T(t)$ – только времени t .

Во многих случаях колебания распределенной системы представимы в виде суперпозиции стоячих волн. В качестве примера мы рассмотрим произвольные колебания ограниченной струны с закрепленными концами и их частный случай, когда эти колебания вызваны сосредоточенным ударом по ней в каком-то месте.

Общая математическая модель следующая. Задано волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} = a^{-2} \frac{\partial^2 u}{\partial x^2} \quad (0 \leq x \leq l) \quad (33.3)$$

и начальные и граничные условия на функцию $u(x, t)$. Начальные условия

$$u(x, t) \Big|_{t=0} = \varphi(x), \quad \frac{\partial u}{\partial t} \Big|_{t=0} = \psi(x), \quad (33.4)$$

граничные –

$$u(x, t) \Big|_{x=0} = 0, \quad u(x, t) \Big|_{x=l} = 0. \quad (33.5)$$

Метод получения требуемого решения был впервые предложен Фурье и назван его именем. Он состоит в том, что сначала ищутся решения вида стоячих волн (33.2) уравнения (33.3) при граничном условии (33.5). Таких решений оказывается счетное число. После этого решение задачи (33.3)–(33.5) ищется в виде суперпозиции полученных решений вида (33.2). При этом оказывается, что суперпозиция всегда может быть подобрана требуемым образом, т.е. так, чтобы удовлетворялись начальные условия (33.4); краевые условия, как и само уравнение (33.2), удовлетворяются автоматически в силу того, что каждое слагаемое суперпозиции им удовлетворяет.

Подставляем (33.2) в (33.3) и находим, что

$$\frac{T''}{T} = a^2 \frac{X''}{X}.$$

Так как левая часть этого равенства не зависит от x , а правая от t , то они равны постоянной, пусть $-\lambda$. Это приводит к уравнениям

$$T'' + \lambda T = 0, \quad X'' + a^2 \lambda X = 0. \quad (33.6)$$

Нас интересуют только те решения уравнений (33.6), для которых функция $u(x, t)$, определяемая согласно (33.2), удовлетворяет краевым условиям (33.5). Это возможно только при положительном λ . При $\lambda > 0$ общие решения (33.6) имеют вид:

$$T = \sin \sqrt{\lambda} t + B \cos \sqrt{\lambda} t, \quad X = C \sin a \sqrt{\lambda} x + D \cos a \sqrt{\lambda} x,$$

и, следовательно,

$$u(x, t) = TX = (A \sin \sqrt{\lambda} t + B \cos \sqrt{\lambda} t)(C \sin a \sqrt{\lambda} x + D \cos a \sqrt{\lambda} x). \quad (33.7)$$

Это решение должно удовлетворять краевым условиям (33.6). Из них следует, что

$$D = 0, \quad C \sin a \sqrt{\lambda} l = 0. \quad (33.8)$$

Принять $C = 0$ нельзя, так как в результате приходим к тривиальному нулевому решению, поэтому из (33.8) следует, что

$$\sqrt{\lambda} = \frac{\pi}{al} n \quad (n = 1, 2, 3, \dots).$$

Итак, приходим к решениям вида

$$u(x, t) = \left(A \sin \frac{\pi n}{al} t + B \cos \frac{\pi n}{al} t \right) \sin \frac{\pi n}{l} x. \quad (33.9)$$

Это стоячие волны с дискретными длинами волн $2l/n$ и соответствующими частотами $\omega_n = \pi n/al$. У этой стоячей волны все точки колеблются синхронно с временной частотой $\omega_n = \pi n/al$ и амплитудой, гармонически меняющейся с изменением x пропорционально $\sin(\pi n x/l)$. Вид нескольких стоячих волн для $n = 1, 2, 3$ показан на рис. 33.1.

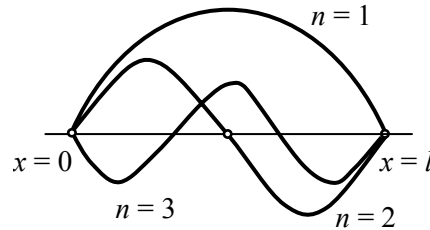


Рис. 33.1. Гармонические стоячие волны: $n=1$ – основной тон, $n=2$ – первая гармоника, $n=3$ – вторая гармоника

Далее будем искать решение исходной задачи (33.3)–(33.5) в виде бесконечной суммы всех стоячих волн (33.7):

$$u(x, t) = \sum \left(A_n \sin \frac{\pi n}{al} t + B_n \cos \frac{\pi n}{al} t \right) \sin \frac{\pi n}{l} x, \quad (33.10)$$

где A_n и B_n пока произвольны. Начальные условия (33.4) приводят к условиям

$$\begin{aligned} \sum B_n \sin \frac{\pi n}{l} x &= \varphi(x), \\ \sum \frac{\pi n}{al} A_n \sin \frac{\pi n}{l} x &= \psi(x), \end{aligned} \quad (33.11)$$

которые можно рассматривать как разложения в ряды Фурье функций $\varphi(x)$ и $\psi(x)$ соответственно. Так как при $n \neq m$

$$\int_0^l \sin \frac{\pi n}{l} x \sin \frac{\pi m}{l} x dx = 0,$$

а при $n = m$ тот же интеграл равен $l/(2\pi)$, то, умножая каждое из равенств (33.11) на $\sin \pi m/l x$ и интегрируя по x от 0 до l , находим, что

$$\begin{aligned} B_m &= \frac{2\pi}{l} \int_0^l \varphi(x) \sin \frac{\pi m}{l} x dx, \\ A_m &= \frac{2a}{m} \int_0^l \psi(x) \sin \frac{\pi m}{l} x dx. \end{aligned} \tag{33.12}$$

Так красиво и просто решается общая задача (33.3)–(33.5). Решение получено в виде суперпозиции дискретной последовательности стоячих волн. Каждая стоячая волна характеризуется частотой $\omega_n = n\pi/al$ и длиной волны $2l/n$. Ее вид определяется формулой (33.10). Частоты ω_n образуют спектр частот возможных колебаний ограниченной струны. Наименьшая частота ω_1 отвечает основному тону, частота ω_2 – первой гармонике, ω_3 – второй и т.д. (см. рис. 33.1). Спектр колебаний струны определяется его спектром частот и соответствующими им амплитудами.

Если речь идет о рояльной струне, то состав гармоник ее колебаний определяет основной тон звука и его окраску, зависящую от гармоник. При одинаковом основном тоне окраска звука может быть совершенно разной. Так, один и тот же нотный звук у рояля, скрипки и гитары звучит совершенно по-разному. Это при одинаковых основных тонах. Так что окраска и красота звука определяется всем его спектром. Основной тон определяется длиной, натяжением и плотностью струны. Спектр же зависит от того, как возбудили ее колебания: где ее ударил молоточек рояля, как и где проводится смычок скрипки и какой он сам, наканифолен ли и как натянут и из чего сделан? Конечно, дело не только в том, каков спектр колебаний самой струны, но и в том, как эти колебания превращаются в звук. Но этот сложный вопрос мы оставим в стороне и рассмотрим только, как зависит спектр колебаний струны от места ее удара $x = \xi$. Разница звучаний рояля Бекштейна и Стенвея проистекает, прежде всего, от различия мест удара молоточка по струне.

При составлении математической модели колебаний первоначально неподвижной струны, получившей сосредоточенный на очень малой длине удар, передавший ей импульс p , будем исходить из общей модели (33.3)–(33.5). В этой общей модели надлежит только конкретизировать вид функций $\varphi(x)$ и $\psi(x)$.

Удар будем считать мгновенным и точечным. В соответствии с этим за время удара струна не успеет переместиться и поэтому $\varphi(x) = 0$. Так как струне при ударе передается импульс p , то

$$\int_0^l \rho \psi(x) dx = p, \tag{33.13}$$

где ρ – постоянная плотность струны, а $\psi(x)$ – скорость струны после удара в точке x . Удар мы предполагаем сосредоточенным в исчезающе малой окрестности точки $x = \xi$. Вне этой исчезающе малой окрестности скорость $\psi(x)$ непосредственно после удара равна нулю и поэтому условие (33.13) можно записать в виде

$$\int_{\xi-\varepsilon}^{\xi+\varepsilon} \psi(x) dx = \frac{p}{\rho} \quad (\varepsilon > 0) \tag{33.14}$$

и $\psi(x) = 0$ при $x \notin [x - \varepsilon, x + \varepsilon]$, где ε исчезающе мало. При неограниченном уменьшении длины отрезка, по которому ударяет молоточек, на этом маленьком отрезке функция $\psi(x)$ неограниченно растет, но все время соблюдается условие (33.13). Таким образом, функция $\varphi(x) = 0$, а функция $\psi(x)$ всюду равна нулю, кроме исчезающе маленькой окрестности точки $x = \xi$, и для нее имеет место условие (33.13). Используем эти сведения для вычисления постоянных A_m и B_m по формулам (33.12). Согласно первой из них, $B_m = 0$, а согласно второй,

$$A_m = \frac{2a}{m} \int_{\xi-\varepsilon}^{\xi+\varepsilon} \psi(x) \sin \frac{\pi m}{l} x dx \approx \frac{2ap}{m\rho} \sin \frac{\pi m}{l} \xi, \tag{33.15}$$

где формулы тем точнее, чем меньше $\varepsilon > 0$.

Это рассуждение, позволяющее у общего решения (33.10) найти конкретные значения коэффициентов A_n и B_n ($n = 1, 2, \dots$), можно формализовать и упростить с помощью уже знакомой вам обобщенной δ -функции. Мгновенный удар молоточка сообщает струне, точнее ее точке $x = \xi$, импульс p . Математическая модель сосредоточенного в точке импульса p – это $p\delta(x - \xi)$, т.е.

$$\psi(x) = \frac{p}{\rho} \delta(x - \xi). \tag{33.16}$$

Подставляя (33.16) в формулы (33.12) для A_m , формально приходим к (33.15), где знак приближения заменяется равенством.

В обоих случаях приходим к тому, что колебания струны, получившей мгновенный сосредоточенный в точке $x = \xi$ удар импульса p , имеют вид:

$$u(x, t) = \sum \frac{2ap}{n\rho} \sin \frac{\pi n}{l} \xi \sin \frac{\pi n}{l} x \sin \frac{\pi n}{al} t. \tag{33.17}$$

Согласно (33.17), максимальные амплитуды основной частоты $\pi/(al)$ и гармоник $2\pi/(al)$, $3\pi/(al)$, ... соответственно равны

$$\frac{2ap}{\rho} \sin \frac{\pi}{l} \xi, \quad \frac{ap}{\rho} \sin \frac{2\pi}{l} \xi, \quad \frac{2ap}{3\rho} \sin \frac{3\pi}{l} \xi,$$

что наглядно представлено на рис. 33.2 для основного тона $n = 1$ и первых трех гармоник $n = 2$, $n = 3$ и $n = 4$.

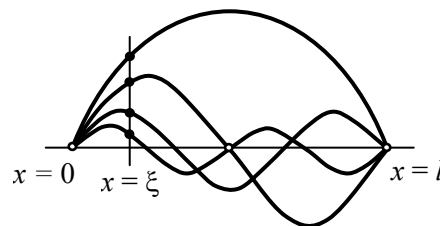


Рис. 33.2. Графическая иллюстрация к зависимости амплитуды от порядка гармоники при импульсном ударе струны в точке $x = \xi$

Из рис. 33.2, в частности, видно, что основной тон наилучше представлен при ударе по центру струны, но при этом отсутствует вторая и четвертая гармоники. Для того чтобы хорошо представить несколько первых гармоник, можно ударить недалеко от концов струны. Так и делается в рояле, причем чем это расстояние меньше, тем больше гармоник представлено. У рояля Стенвея это расстояние меньше, чем у Бекштейна, и поэтому его звук «звонче», но зато звук у рояля Бекштейна «бархатнее» и «мягче».

Звук скрипичной струны и фортепианной одного и того же основного тона сильно различаются, и это отличие объясняется также различностью представительности гармоник.

Наше ухо устроено так, что звук воспринимается фактически через его спектр. Музыкальные звуки до, ре, ми и т.д. имеют спектр, состоящий из основного тона и гармоник. Музыкальные аккорды имеют спектр, составленный из спектров, входящих в него звуков. Шум имеет очень много спектральных частот, такой спектр называют сплошным в противоположность дискретному спектру, когда частот не очень много.

Почему определенные последовательности звуков вызывают у нас такие сильные эмоции – пожалуй, одна из величайших загадок природы. Хотя некоторые необходимые для этого требования известны и составляют предмет теории музыки.

34. МИКРОЧАСТИЦА

До сих пор мы рассматривали различные математические модели окружающего нас мира, подчас непонятного и загадочного, но все же того, в котором мы живем. Но помимо этого мира существуют миры, скрытые от нас, которые недоступны нам, и в этом смысле существуют лишь в нашем воображении. Таков микромир – мир молекул, атомов и элементарных частиц. Мир, о котором человек догадался давно, но который стал доступен изучению только в последнее столетие, вместе с созданием новой неклассической физики – квантовой механики.

Математические модели объектов микромира – квантовой механики – еще в большей мере, чем модели классической физики, носят математический характер, не только оторванный от нашего непосредственного опыта, повседневных представлений и интуиции, но и непринимаемый ими. Несогласования имели место и раньше при рождении законов механики и электродинамики. В механике повседневные наблюдения, казалось бы, бесспорно свидетельствуют о том, что силы вызывают движение, в то время как они только его изменяют. Более сложные электрические явления казались порождаемыми зарядами и токами и совершенно неприемлемой была идея о существовании электромагнитных волн, ибо при этом нет ни зарядов, ни токов. Для квантовой механики такие неприятия проявлялись значительно сильнее. Микромир никогда и ни в чем непосредственно не воспринимался нашими органами чувств и оказался невообразимо отличным от привычного макромира. Вместе с тем, в конечном счете, именно микромир лежит в основе всех проявлений макромира. Поэтому микромир все же можно обнаружить с помощью создаваемых человеком приборов и специальных условий опыта: наблюдая спектры излучения и поглощения атомов, радиоактивность, дифракцию электронов, фотоэффект и другое.

Приобретаемые таким образом новые данные и их теоретическое осмысление противоречили не только многовековому опыту, но и сложившимся к тому времени теоретическим представлениям об окружающем физическом мире, получившим впоследствии наименование классической физики. Конечно, это относится не только к микромиру, но и другим разделам новой физики нашего века – физике больших скоростей, близких к скорости света, огромных расстояний в миллионы световых лет, гигантских масс и огромных плотностей, сверхвысоких температур, невиданных магнитных полей и др.

Укажем на некоторые явные отличия проявлений микромира, говорящих о необходимости кардинального пересмотра представлений классической физики и создания кардинально новых математических моделей.

Человечество давно, со времен древних греков, догадывалось, что материальный мир состоит из мельчайших частиц – атомов – и все происходящее с ним порождается их движениями и взаимодействиями. При этом само собой разумеется, что хотя они и чрезвычайно маленькие, но наделены такими же свойствами, как и обычные тела: они могут перемещаться, ударяться друг о друга, зацепляться и т.д. Создаваемая поначалу кинетическая теория газов, представления о жидких и твердых телах и многообразные химические превращения не противоречили этим

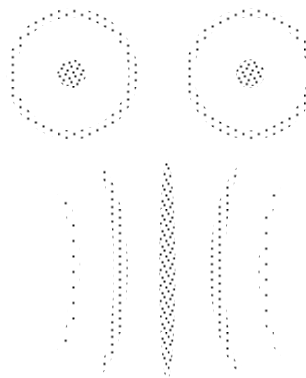
наивным представлениям. Но на пороге XX века начали обнаруживаться несоответствия. Постепенно эти несоответствия проявили свою полную несовместимость с классической физикой.

Например, выяснилось, что в некоторых случаях бессмысленно говорить о траекториях движения микрочастиц, что представление о наличии таких траекторий входит в явное противоречие с опытом, что в микромире может в классическом понимании нарушаться закон сохранения энергии, а колеблющийся в атоме электрический заряд – электрон – почему-то не излучает, как это положено согласно классической электродинамике, а если и излучает, то только волновые пакеты определенных частот и энергий. Вы, конечно, слышали об этих и других особенностях микромира: о дифракции электронов, о радиоактивном распаде, спектрах излучения и поглощения атомов, квантовых генераторах – лазерах, фотоэффекте, ядерных реакциях и реакторах, атомной бомбе, непомерно большой энергии излучения Солнца, необъяснимых закономерностях таблицы Менделеева и, возможно, многом другом.

Для доказательства несостоятельности научных представлений и теорий не требуется много противоречий, нужно много соответствий, но противоречия достаточно одного. На одном мы и остановимся, на дифракции электронов.

Электрон по первоначальным представлениям – это мельчайшая заряженная частица с очень маленьким зарядом и массой. Его полет можно отклонить магнитным полем в полном соответствии с силой Лоренца, действующей на движущийся в магнитном поле электрический заряд. Его можно обнаружить после удара о фотопластинку, на ней он оставит после проявления черную точку. В электрическом поле он движется в полном соответствии с законами Ньютона. Но вот поставим на пути потока электронов постоянной интенсивности преграду с двумя отверстиями, а после преграды экран, который фиксирует попадание в него электронов. Естественно считать, что на экране оставят следы только те электроны, которые пролетают через отверстия. Остальные будут задержаны препятствием и на него не попадут. При этом одни электроны летят через одну дырочку, а другие – через другую. Пролетев через дырочки в преграде, они долетят до экрана и оставят на нем свои следы. Из сказанного с очевидностью следует, что если мы будем в течение некоторого времени T пропускать электроны через одну дырочку, закрыв другую, а затем опять в течение того же времени так же пропускать через другую, то это должно дать на экране ту же картину попаданий в него электронов, как если их пропускать через обе дырочки в течение того же времени T . Различие можно видеть только в том, что пролетевшие через две дырочки электроны взаимодействуют друг с другом. Но это кулоновское взаимодействие очень мало. Кроме того, оно, как и всякое другое, исключается, если время между полетами электронов более времени их движения от препятствия до экрана, т.е. если поток электронов достаточно малой интенсивности. При описанном раздельном пропускании электронов то через одну дырочку, то через другую или сразу через две, но в течение в два раза меньшего времени, на экране должны быть одинаковые картины попаданий в него электронов. Если электроны летят по определенным траекториям, как обычные маленькие частицы, то это, безусловно, так. Что же дает эксперимент? Он дает нечто совершенно неожиданное, изображенное на рис. 34.1.

Рис. 34.1. Дифракция электронов при последовательном прохождении через две дырочки и одновременном: черные точки соответствуют «ударам» о фотоэмульсионный слой отдельных электронов



Эти картинки разные, совершенно не похожие друг на друга. Вторая картинка похожа на то, что через два отверстия пролетали не частицы – электроны, а проходил свет – электромагнитные волны – с характерной для него картинкой дифракции. Так что же такое летящий электрон – это летящая частица или волна, для которой имеет место явление дифракции? Скорее всего, ни то, ни другое, они – микрочастицы, объекты новой незнакомой нам природы. Верно лишь то, что в одних случаях они проявляют себя подобно частицам, в других ведут себя как волны, а в третьих – ни как волны и ни как частицы, а как микрочастицы.

История синтеза математических моделей микромира – сложный, противоречивый и запутанный процесс, очень далекий от логического вывода. В нем много фантастических гипотез, непостижимых догадок и озарений с последующей фильтрацией, уточнением и взаимными увязками. Изложение квантовой механики даже для физиков обходит сколько-нибудь полное описание этого процесса. Тем более, естественно это сделать для математиков, начав сразу с описания итоговой математической модели, ее формализма, который затем иллюстрируется на примерах.

Освоение необычных математических моделей микромира требует выработки новых представлений, понятий и новой интуиции на основе постижения математической модели путем проникновения в ее логику и гармонию, исходя из чисто математических соображений. Такой путь постижения истины тоже доступен человеческому интеллекту и именно в этой возможности его мощь и сила, вызвавшая гигантский скачок естественных наук в последние столетия.

Основанием необходимости этих необычных математических моделей служит невозможность трактовки наблюдаемых явлений микромира с позиций классической физики, а основанием их справедливости являются качественные и количественные совпадения получаемых с их помощью выводов и результатов с данными разнообразных наблюдений и экспериментов (дифракция электронов на кристаллических решетках, спектры атомов и молекул, фотоэффект, спектры равновесного излучения, химические свойства веществ и таблица Менделеева, явления физики твердого тела и полупроводников, квантовые генераторы и многое, многое другое). К этому следует добавить, что все же необычные квантовые модели переходят в классические при изменении параметров от их значений в микромире к значениям нашего макромира. И в этом можно видеть стыковку законов нашего макромира и микромира.

Лучше всего воспринять модель микромира, о которой будет речь ниже, как игру, так же, как в свое время вы слушали игру Коена «жизнь». Только теперь меняются не фишки на шахматной доске, а некая функция ψ и находимые с ее помощью физические величины. Такой подход позволит вам легко принять ее правила и, потрудившись, научиться играть. Когда вы научитесь «играть», математическая модель микромира вам станет привычной, а со временем и понятной.

Если основной математический аппарат классической физики – это дифференциальные уравнения, то для квантовой механики – это еще и совсем другой, существенно отличный от дифференциальных уравнений, математический аппарат теории линейных операторов. Он был создан математиками до того, как был востребован физиками, исходя из внутренних потребностей и мотивов развития самой математики. Теория линейных операторов вам знакома из курса функционального анализа, но, несомненно, действия с линейными операторами вам менее привычны, чем с дифференциальными уравнениями.

Математические модели классической физики и квантовой механики существенно разные, но принцип детерминизма в его достаточно широком понимании уцелел. Как и в классической физике, в основе описания объектов микромира лежит состояние, но теперь в виде некоторой комплексной функции ψ от координат микрочастиц \vec{x} и, возможно, их спинов. При этом состояние микросистемы – это сама комплексная функция ψ перечисленных переменных, а не значения входящих в нее переменных.

Изменение во времени функции ψ подчиняется дифференциальному уравнению Шредингера

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi, \quad (34.1)$$

где H – оператор рассматриваемой системы, \hbar – так называемая постоянная Планка, равная очень маленькой величине $1,05 \cdot 10^{-27}$ эрг·с.

Функция ψ , как уже отмечалось – это функция вектора \vec{x} – координат микрочастиц, спинов и времени t . Оператор H – это так называемый оператор Гамильтона. Он действует по переменным вектора \vec{x} . Назван он так потому, что получается по определенным правилам из функции Гамильтона микросистемы, рассматриваемой как классическая механическая система.

То, что именно функция является описанием состояния, имело место и в классической механике при описании сплошных сред, и в электродинамике при описании электрического и магнитного полей. Не ново и то, что эта функция удовлетворяет некоторому уравнению в частных производных. Не так уж ново и то, что функция ψ комплексная: такое имеет место в приложении теории функций комплексного переменного к задачам гидродинамики и электродинамики. Существенное отличие состоит не в этом. Оно состоит в том, как по функции состояния ψ находятся возможные значения наблюдений физических величин, характеризующих рассматриваемую микросистему; как функция ψ связана с наблюдаемыми физическими величинами: координатами, импульсами, моментами, энергиями, спинами, частотами, длинами волн и др.

Математический формализм квантовой механики. В квантовой механике постулируется, что каждой физической величине отвечает эрмитов – комплекс-

ный самосопряженный – линейный оператор, собственные значения которого – единственные возможные значения этой физической величины.

Напомним, что эрмитовость оператора L означает, что для любых функций ψ_1 и ψ_2 имеет место (звездочка означает комплексное сопряжение)

$$\int \psi_1 * L\psi_2 d\bar{x} = \int \psi_2 (L\psi_1) * d\bar{x}. \quad (34.2)$$

Собственные значения λ любого эрмитова оператора L действительные. Числа λ и соответствующие собственные функции φ находятся из условия

$$L\varphi = \lambda \varphi \quad (\varphi \neq 0). \quad (34.3)$$

Множество всех собственных функций эрмитова оператора полное, а собственные функции φ_1 и φ_2 , отвечающие разным собственным значениям $\lambda_1 \neq \lambda_2$, ортогональны

$$\int \varphi_1 * \varphi_2 d\bar{x} = 0. \quad (34.4)$$

Итак, как уже указывалось, каждой физической величине l сопоставляется по правилам, которые будут указаны ниже, эрмитов оператор L . Возможные значения физической величины l – это только его собственные значения. В результате измерений эти значения могут наблюдаться. При этом в идентичных условиях результаты измерений могут быть разными, и фактический физический смысл имеет лишь их среднее значение (математическое ожидание). Это среднее значение $\langle l \rangle$, согласно квантово-механическому формализму, находится по состоянию ψ квантовой системы по формуле

$$\langle l \rangle = \int \psi * L\psi d\bar{x}, \quad (34.5)$$

где функция состояния ψ предполагается нормированной, т.е.

$$\int \psi * \psi d\bar{x} = 1.$$

Сказанным исчерпывается формальное описание математической модели квантовой системы. Эта математическая модель включает понятие состояния – функцию ψ , уравнение Шредингера изменения состояния во времени и формализм отыскания возможных значений любой физической величины и ее среднестатистического значения по функции состояния и операторам этих физических величин, согласно (34.5).

Фактическое построение математической модели квантовой системы требует написания функций ψ и H и уравнения Шредингера, а также знания операторов, отвечающих интересующим нас физическим величинам.

Ее исследование в рамках изложенного состоит в решении уравнения Шредингера, возможно, при некоторых дополнительных сведениях о функции ψ ; отыскании собственных значений и собственных функций операторов интересующих нас физических величин и вычислении с помощью найденной функции ψ их средних значений по формуле (34.5).

Укажем операторы, отвечающие координатам x, y и z , импульсам p_x, p_y и p_z , кинетической T и потенциальной V энергиям, а также оператору Гамильтона H , для частицы массы m в потенциальном поле $V(x, y, z)$ без учета ее спина.

Операторы, отвечающие координатам x, y и z , – это умножение на x, y и, соответственно, z . Так что оператор, отвечающий координате x , от функции ψ – это $x\psi$; для y и z – это $y\psi$ и $z\psi$. Оператор от ψ , отвечающий функции $V(x, y, z)$, – это $V(x, y, z)\psi$.

Импульсам p_x, p_y и p_z отвечают дифференциальные операторы

$$-i\hbar \frac{\partial}{\partial x}, -i\hbar \frac{\partial}{\partial y}, -i\hbar \frac{\partial}{\partial z}. \quad (34.6)$$

Вектору импульса $p(p_x, p_y, p_z)$ отвечает вектор – оператор

$$-i\hbar \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) = -i\hbar \nabla, \quad (34.7)$$

где ∇ – значок оператора градиента.

Кинетической энергии T частицы массы m соответствует оператор

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) = -\frac{\hbar^2}{2m} \nabla^2 = -\frac{\hbar^2}{2m} \Delta, \quad (34.8)$$

где Δ – значок оператора Лапласа.

Потенциальная энергия $V(x, y, z)$ является функцией координат, и поэтому соответствующий ей оператор – это умножение на $V(x, y, z)$.

Осталось указать оператор Гамильтона частицы массы m . Согласно так называемому принципу соответствия классической функции Гамильтона

$$\frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) + V(x, y, z)$$

и квантового оператора Гамильтона, он записывается в виде

$$H = -\frac{\hbar^2}{2m} \Delta + V(x, y, z). \quad (34.9)$$

Теперь мы можем записать уравнение Шредингера для микрочастицы массы m (без спина). Функция состояния будет

$$\psi = \psi(\bar{x}, t) = \psi(x, y, z, t), \quad (34.10)$$

а уравнение Шредингера, согласно (34.1) и (34.9), имеет вид

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) + V(x, y, z)\psi. \quad (34.11)$$

Прежде чем перейти к конкретным примерам, укажем на два важных следствия, вытекающих из описанной математической модели квантовой системы. Одно следствие касается того, что можно указать, не только какие значения может принимать та или иная физическая величина и не только ее среднее значение, но и вероятности, с которыми эти значения могут приниматься.

Для того чтобы это обнаружить, разложим в формуле среднего значения (34.5) функцию ψ по собственным нормированным функциям оператора L . Ради простоты пусть спектр собственных значений оператора L дискретен, и тогда разложение нормированной функции ψ по собственным функциям $\varphi_1, \varphi_2, \dots$ с собственными значениями l_1, l_2, \dots , которые есть возможные значения величины l , имеет вид:

$$\psi = c_1\varphi_1 + c_2\varphi_2 + \dots, \quad (34.12)$$

где

$$c_s = \int \psi(\bar{x}) \varphi_s^*(\bar{x}) d\bar{x}.$$

Подставляя (34.12) в (34.5), непосредственно находим, что

$$\langle l \rangle = \sum |c_s|^2 l_s, \quad (34.13)$$

где в силу нормированности функции ψ

$$\sum |c_s|^2 = 1.$$

Как известно, математическое ожидание величины l , которая может принимать значения l_1, l_2, \dots с вероятностями p_1, p_2, \dots , равно

$$Ml = \sum l_s p_s. \quad (34.14)$$

Сравнивая (34.14) с (34.13), находим, что вероятность p_s принятия значения l_s равна

$$p_s = |c_s|^2. \quad (34.15)$$

В частности, отсюда следует, что если функция ψ совпадает с одной из собственных функций оператора L , пусть с φ_s , то $p_s = 1$ и поэтому в этом случае физическая величина может принимать только одно значение l_s , а вероятности других значений l_j ($j \neq s$) равны нулю.

Таким образом, в общем случае функции ψ величина l может принимать любое из значений l_1, l_2, \dots с вероятностями p_1, p_2, \dots соответственно. Это не исключает возможности в частных случаях (частных видах функции ψ) принятия величиной l лишь некоторых значений или даже только одного. Только в последнем случае результат измерения физической величины l предскажем.

Таков первый важный вывод из сформулированной математической модели квантовой системы.

Проиллюстрируем его простым примером, когда речь идет о значениях координат x, y и z микрочастицы. Этим величинам отвечают операторы умножения на x, y и z . Поэтому, согласно формуле (34.5),

$$\langle x \rangle = \int \psi^* x \psi dx dy dz = \int x |\psi|^2 dx dy dz = \int x \left(\int |\psi|^2 dy dz \right) dx$$

и аналогично для $\langle y \rangle$ и $\langle z \rangle$. Отсюда непосредственно следует, что

$$\iint |\psi|^2 dy dz$$

является плотностью вероятностей значений координаты x , а $|\psi|^2$ — плотность вероятностей значений x, y и z .

Тем самым найден статистический смысл функции квантового состояния ψ : квадрат ее модуля $|\psi|^2$ — это плотность вероятностей принятия координатами значений x, y и z .

Однако, согласно предыдущему, функция ψ дает статистическое описание не только значений координат x, y и z , но и любых других физических величин, операторы которых нам известны.

Второй вывод, который можно сделать из определения математической модели, относится к так называемому принципу неопределенности Гейзенберга о возможных результатах одновременного измерения физических величин (l и k с операторами L и K) и состоит в утверждении, что однозначный результат измерения двух физических величин возможен только при коммутативности операторов L и K , т.е. когда

$$LK - KL = 0.$$

В противном случае либо обе величины l и k не имеют однозначных значений, либо если одна однозначна, то другая неопределенна. Полнее, для дисперсий величин l и k имеет место неравенство

$$Dl Dk \geq \frac{1}{4} \int \psi^* \left(\frac{LK - KL}{i} \right)^2 \psi d\bar{x}. \quad (34.16)$$

Поясним это важное неравенство. В нем l и k – произвольные физические измеряемые величины, а L и K – соответствующие им операторы. Оно является следствием очевидного неравенства

$$\int \left| L\psi + \frac{\lambda}{i} K\psi \right|^2 d\bar{x} \geq 0,$$

справедливого при любом действительном λ .

Выше формулой (34.5) были определены средние значения физических величин l и k . Аналогично можно определить их дисперсии, как средние значения величин $(l - \langle l \rangle)^2$ и $(k - \langle k \rangle)^2$. Так,

$$Dl = \int \psi^* (L - \langle l \rangle)^2 \psi d\bar{x},$$

где функция $\psi(\bar{x}, t)$, являющаяся состоянием микрочастицы, как и в формуле (34.5), предполагается нормированной на единицу.

Например, операторы координаты x и импульса p_x не коммутируют, так как

$$\left(-i h \frac{\partial}{\partial x} \right) x + x i h \frac{\partial}{\partial x} = -i h - i h x \frac{\partial}{\partial x} + x i h \frac{\partial}{\partial x} = -i h,$$

и поэтому, согласно формуле (34.16),

$$DxDp_x \geq \frac{h^2}{4}, \quad (34.17)$$

т.е. произведение дисперсий результатов измерений координаты x и соответствующего импульса p_x не может быть меньше $h^2/4$.

Это существенно отлично от того, что имеет место в классической механике, где, наоборот, для определения состояния требуется задание вместе с координатой x и импульса p_x . В квантовой же механике они одновременно не могут быть заданы принципиально. Это возможно лишь в случае коммутативности соответствующих операторов. Так, например, коммутируют между собой операторы координат x, y и z , а также импульсов p_x, p_y и p_z , и поэтому одновременное однозначное измерение только всех координат или только всех импульсов возможно, но если одновременно определены три координаты или три импульса, то не определены все импульсы и соответственно все координаты.

Таким образом, одновременная определенность всех физических величин, к которой мы привыкли в классической механике, в квантовой механике не имеет места. Напомним, что кроме этого кардинального различия с классикой, имеет место и еще другое существенное различие, о котором уже говорилось. Это дискретность физических величин, их квантованность, поскольку возможные значения физической величины определяются спектром собственных значений ее оператора, а этот спектр может быть дискретным.

Перейдем к рассмотрению движений свободной квантовой частицы в некоторых простых одномерных полях. Рассмотрим мы и гармонический осциллятор, и квантовый вариант задачи Ньютона – атом водорода и водородоподобный ионизированный атом гелия или лития.

Во всех этих случаях уравнение Шредингера, согласно (34.10), имеет вид

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) + V(x, y, z) \psi. \quad (34.18)$$

Будем искать его решения вида

$$\psi(x, y, z, t) = A(t) \bar{\psi}(x, y, z). \quad (34.19)$$

После подстановки (34.19) в (34.18) и разделения переменных t и x, y, z получаем

$$\frac{i\hbar}{A} \frac{\partial A}{\partial t} = \frac{1}{\bar{\psi}} \left\{ -\frac{\hbar^2}{2m} \Delta \bar{\psi} + V \bar{\psi} \right\} = E, \quad (34.20)$$

где E – постоянная. Перед нами два уравнения: одно относительно $A(t)$, второе относительно $\bar{\psi}(x, y, z)$. Решение первого уравнения легко находится и имеет вид

$$A(t) = e^{-(iE/\hbar)t}, \quad (34.21)$$

решение второго уравнения

$$-\frac{\hbar^2}{2m} \Delta \bar{\psi} + V \bar{\psi} = E \bar{\psi} \quad (34.22)$$

требует конкретизации функции V . Если нам удалось найти решение уравнения (34.22), удовлетворяющее естественным условиям, налагаемым на функцию $\bar{\psi}$ (непрерывность, ограниченность и еще какие-нибудь условия, следующие из постановки задачи), то решением (точнее одним из решений) исходного уравнения (34.20) будет

$$\psi(x, y, z, t) = e^{-(iE/\hbar)t} \bar{\psi}(x, y, z). \quad (34.23)$$

Это решение ψ при любом t удовлетворяет уравнению (34.22) (если в него вместо $\bar{\psi}$ подставить ψ) и поэтому является собственной функцией оператора энергии

$$-\frac{\hbar^2}{2m} \Delta + V. \quad (34.24)$$

Следовательно, в найденном состоянии ψ энергия системы имеет определенное значение E , так как E , согласно (34.20), является собственным значением оператора энергии, отвечающим собственной функции ψ .

Состояния квантовой системы, отвечающие определенным значениям энергии E , называются стационарными. Так что ψ – стационарное состояние с энергией E . Заметим, что

$$|\psi|^2 = |\bar{\psi}|^2, \quad \left| e^{-(iE/\hbar)t} \right| = 1, \quad (34.25)$$

поэтому плотности вероятности значений координат в стационарном состоянии не зависят от времени (потому оно и названо стационарным).

Для квантовой системы имеет место принцип суперпозиции. Это следует из линейности относительно функции состояния уравнения Шредингера. Поэтому суперпозиция стационарных состояний является тоже некоторым состоянием, удовлетворяющим уравнению Шредингера. Верно и обратное: всякое допустимое решение уравнения Шредингера может быть представлено суперпозицией стационарных состояний. Тем самым изучение возможных состояний квантовой системы сводится в этом смысле к отысканию всех ее стационарных состояний.

Спектр собственных значений оператора энергии может быть дискретным, непрерывным или смешанным. Каждому невырожденному значению спектра отвечает единственная нормированная собственная функция. При вырожденности собственного значения их может быть несколько и даже бесконечно много. Однако в любом случае все собственные функции можно ортогонализировать и нормировать.

Условие ортогональности двух различных собственных функций ψ_1 и ψ_2 , отвечающих дискретному или непрерывному спектру, состоит в том, что

$$\int \psi_1 \psi_2^* d\bar{x} = 0. \quad (34.26)$$

Условие нормировки по форме записывается по-разному для собственного значения, отвечающего дискретному или непрерывному спектру.

Для случая дискретного спектра условие нормированности функции ψ_s , отвечающей собственному значению E_s , уже было написано, оно имеет вид

$$\int \psi_s \psi_s^* d\bar{x} = 1. \quad (34.27)$$

Для непрерывного спектра, собственные функции которого зависят от параметра E , т.е. $\psi = \psi_E$, условие нормировки можно записать в виде

$$\int \psi_E \psi_{E'}^* d\bar{x} = \delta(E - E'). \quad (34.28)$$

Эта запись означает, что при $E \neq E'$

$$\int \psi_E \psi_{E'}^* d\bar{x} = 0, \quad (34.29)$$

что соответствует ортогональности. При $E = E'$ интеграл (34.29) обращается в бесконечность, но так, что все же существует интеграл от него по E , равный единице, т.е.

$$\int \psi_E \psi_E^* d\bar{x} = 1, \quad (34.30)$$

где интегрирование по E происходит по непрерывной части спектра.

Для того чтобы сблизить записи (34.26)–(34.31), запишем их в виде

$$\int \psi_i \psi_j^* d\bar{x} = \delta_{ij}, \quad \int \psi_E \psi_{E'}^* d\bar{x} = \delta(E - E'), \quad (34.31)$$

где символы δ_{ij} и $\delta(E - E')$ обращаются в нуль соответственно при $i \neq j$ и $E \neq E'$, а при $i = j$ и $E = E'$ первый равен 1, а для второго равен единице интеграл

$$\int \delta(E - E') dE = 1.$$

Напомним, что символ δ_{ij} (функция от целочисленных значений i и j) был введен Кронекером и называется символом Кронекера, а символ $\delta(E - E')$ был введен в теоретическую физику Дираком и называется функцией Дирака. Впервые эта функция использовалась Хевисайдом в созданном им новом, непонятом современниками, операционном исчислении.

После этого затянувшегося, но необходимого предисловия перейдем, наконец, к рассмотрению указанных выше квантовых систем.

Микрочастица (например, электрон) описывается операторами координат, импульса, моментов импульса, энергии и спина, кроме того, она имеет массу и заряд. Если микрочастица имеет заряд, то на нее действует электрическое, а при движении – и магнитное поля. Такое богатство описания и возможных влияний приводит к гигантскому разнообразию явлений в поведении микрочастиц. Последующее имеет целью познакомиться лишь с некоторыми из них, наиболее простыми, не приводящими к математическим трудностям и все же достаточно представительными для того, чтобы почувствовать отличия микромира от привычного макромира. В последующем изложении примеры относятся только к микрочастице, для которой отсутствует или несущественен спин и отсутствует магнитное поле.

Свободная микрочастица. Для свободной микрочастицы $V = 0$, и поэтому уравнение (34.22) принимает вид

$$-\frac{\hbar^2}{2m} \Delta \bar{\psi} = E \bar{\psi}. \quad (34.32)$$

Непосредственно проверяя, можно убедиться, что оно имеет решение вида

$$\bar{\psi} = \exp \frac{i}{\hbar} (p_x x + p_y y + p_z z), \quad (34.33)$$

где p_x, p_y, p_z – произвольные постоянные, удовлетворяющие условию:

$$\frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) = E. \quad (34.34)$$

Поэтому, согласно (34.23), соответствующее решение уравнения Шредингера для свободной частицы запишется в виде:

$$\psi = \exp \frac{i}{\hbar} (p_x x + p_y y + p_z z - Et). \quad (34.35)$$

Это знаменитая функция. Она получила наименование волны де Бройля и с нее началась волновая трактовка микрочастицы. Чтобы это понять, заметим, что, с одной стороны, функция ψ описывает комплексную плоскую волну с частотой $\omega = E/\hbar$ и волновым числом $k = |\bar{p}|/\hbar = \sqrt{2mE}/\hbar$, распространяющуюся вдоль вектора \bar{p} , а с другой, – она представляет собой микрочастицу с энергией E и импульсом \bar{p} . Последнее следует из того, что функция ψ является собственной функцией операторов импульсов

$$-ih \frac{\partial}{\partial x}, \quad -ih \frac{\partial}{\partial y}, \quad -ih \frac{\partial}{\partial z}$$

и одновременно оператора кинетической энергии $-\hbar^2 \Delta / 2m$ с собственными числами p_x, p_y, p_z и E соответственно. Согласно смыслу функции состояния ψ , плотность вероятности ее местонахождения пропорциональна $|\psi|^2$, который, как видно из (34.35), при всех x, y, z и t равен единице. Таким образом, волна де Бройля как функция состояния описывает частицу с определенным импульсом и энергией, но располагающуюся равномерно во всем пространстве, в полном соответствии с принципом неопределенности.

Микрочастица в потенциальной яме. Пусть в прямоугольной потенциальной яме с неограниченными стенками находится микрочастица. В этом случае в уравнении Шредингера

$$ih \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V \psi \quad (34.36)$$

функция V зависит только от одной координаты x :

$$V = \begin{cases} 0 & \text{при } x \in (0, a), \\ \infty & \text{при } x \notin (0, a), \end{cases} \quad (34.37)$$

и ее график имеет вид, изображенный на рис. 34.2.

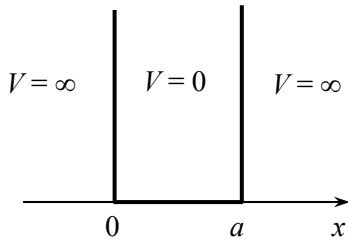


Рис. 34.2. Иллюстрация к поведению электрона в бесконечно глубокой потенциальной яме

Решение уравнения Шредингера (34.36) можно искать в виде

$$\psi = \psi_1(x)\psi_2(y, z). \quad (34.38)$$

После подстановки (34.38) в (34.36) получаемое уравнение можно записать в виде:

$$\frac{1}{\psi_1} \left(ih \frac{\partial \psi_1}{\partial t} + \frac{\hbar^2}{2m} \frac{\partial^2 \psi_1}{\partial x^2} - V \psi_1 \right) = \frac{-1}{\psi_2} \left(ih \frac{\partial \psi_2}{\partial t} + \frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi_2}{\partial y^2} + \frac{\partial^2 \psi_2}{\partial z^2} \right) \right) \quad (34.39)$$

с разделяющимися переменными, что позволяет последующее рассмотрение свести к поиску решения одномерного уравнения Шредингера:

$$ih \frac{\partial \psi_1}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi_1}{\partial x^2} + V \psi_1. \quad (34.40)$$

Этот прием будет использоваться и в последующих задачах.

При $x \notin (0, a)$ из уравнения (34.40) следует, что $\psi_1 = 0$, поскольку при этом $V(x) = \infty$. При $x \in (0, a)$ оно принимает вид одномерного уравнения свободной частицы (ψ_1 заменено на ψ)

$$ih \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2},$$

решение которого легко находится и записывается в виде суперпозиции функций вида

$$\exp i \left(\frac{p_x}{\hbar} x - \frac{E}{\hbar} t \right), \quad E = \frac{p_x^2}{2m},$$

которая должна удовлетворять крайевым условиям равенства нулю при $x = 0$ и $x = a$. Этим крайевым условиям удовлетворяют только функции вида:

$$c \exp i \left(\frac{p_x}{\hbar} x - \frac{E}{\hbar} t \right) - c \exp i \left(-\frac{p_x}{\hbar} x - \frac{E}{\hbar} t \right) = 2ic \exp \left(-\frac{iEt}{\hbar} \right) \sin \frac{p_x}{\hbar} x, \quad (34.41)$$

для которых

$$\frac{p_x a}{\hbar} = \pi k \quad (k = \pm 1, \pm 2, \dots),$$

то есть

$$p_x = \frac{\pi \hbar}{a} k, \quad E_k = \frac{\pi^2 \hbar^2}{2ma^2} k^2. \quad (34.42)$$

Мы уже знаем, что функции (34.41) являются собственными функциями оператора кинетической энергии $-\hbar^2/(2m) \cdot \partial^2/\partial x^2$ с собственными значениями E_k , определяемыми формулой (34.42). Функция (34.41) состоит из суперпозиции двух функций, которые являются собственными для оператора импульса $-\hbar \partial/\partial x$ с собственными значениями, равными $p_k = \pm \pi \hbar k/a$.

Таким образом, функции состояния ψ формулы (34.41) отвечает квантовая частица с энергией E_k и импульсами p_k и $-p_k$, принимаемыми равновероятно.

Выясним еще, какова плотность вероятности координаты x квантовой частицы, находящейся в потенциальной яме с неограниченными краями для функций ψ вида (34.41).

Для этого необходимо функцию формулы (34.41) нормировать и найти квадрат ее модуля. Обозначим эту функцию ψ_k . Непосредственные вычисления дают, что

$$\int \psi_k \psi_k^* dx = \frac{1}{\sqrt{2a}} \left(\psi_k = e^{-i(E/\hbar)t} \sin \frac{p_k}{\hbar} x \right),$$

и поэтому для нормированной функции ψ_k

$$|\psi_k|^2 = \frac{1}{2a} \sin^2 \frac{k\pi}{a} x. \quad (34.43)$$

Соответствующие графики плотности вероятности изображены на рис. 34.3. (для $k = 1, 2$ и 3).

Возможные функции состояния ψ являются всевозможными суперпозициями с комплексными коэффициентами найденных функций ψ_k , т.е.

$$\psi(x, t) = \frac{1}{2a} \sum c_k \exp \left(-i \frac{\pi^2 \hbar k^2}{2ma^2} t \right) \sin \frac{k\pi}{a} x,$$

для которой условие нормировки состоит в равенстве единице суммы квадратов модулей коэффициентов c_k . При этом квантовая частица имеет энергию E_k и импульс p_k или $-p_k$ с вероятностями $1/2|c_k|^2$.

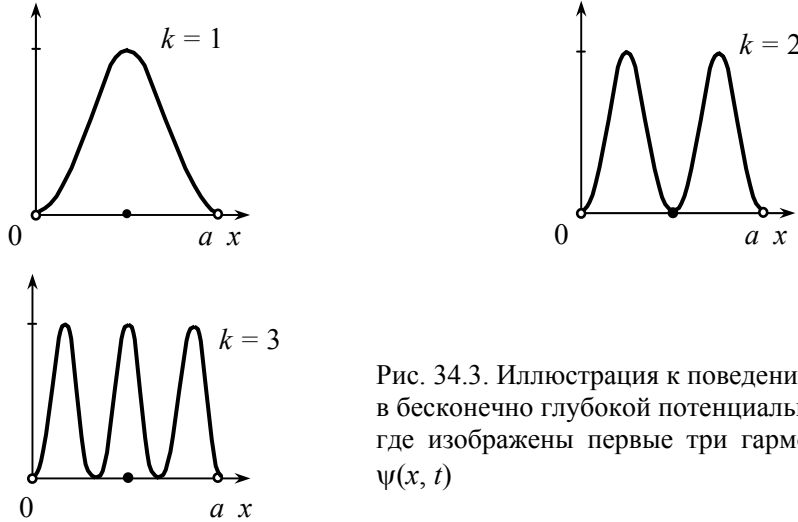


Рис. 34.3. Иллюстрация к поведению электрона в бесконечно глубокой потенциальной яме, где изображены первые три гармоники функции $\psi(x, t)$

Что же отличает движение квантовой частицы в потенциальной яме от движения классической? Классическая частица будет двигаться в ямке от стенки к стенке, идеально отскакивая от них. Может иметь любую энергию E . Половину времени попеременно и периодически она будет иметь импульс $\sqrt{2mE}$, и другую половину $-\sqrt{2mE}$. Никакая суперпозиция движений иметь места, конечно, не может. По существу, мало общего, только соотношение $p_x^2 = 2mE$ между энергией и импульсом.

Мы решили задачу об одномерном движении квантовой частицы в бесконечной прямоугольной потенциальной яме, определяемой функцией (34.37), которую обозначим $V_a(x)$. Располагая этим решением

$$\psi_k(x) = \frac{1}{\sqrt{2a}} \exp\left(-i \frac{E_k}{\hbar} t\right) \sin \frac{\pi k}{a} x,$$

можно указать решение и для частицы, находящейся в трехмерной прямоугольной яме (прямоугольном ящике), для которого потенциальная функция имеет вид:

$$V = V_a(x) + V_b(y) + V_c(z).$$

А именно, искомые функции $\psi(x, y, z)$ будут

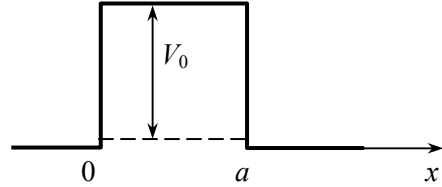
$$\psi_{k_1 k_2 k_3}(x, y, z) = \frac{e^{-iEt/\hbar}}{\sqrt{8abc}} \sin \frac{\pi k_1}{a} x \sin \frac{\pi k_2}{b} y \sin \frac{\pi k_3}{c} z,$$

где k_1, k_2, k_3 – целые числа, большие нуля, и

$$E_{k_1 k_2 k_3} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{k_1^2}{a^2} + \frac{k_2^2}{b^2} + \frac{k_3^2}{c^2} \right).$$

Диффузия микрочастицы через потенциальный барьер. Допустим, что микрочастицы движутся в направлении оси Ox . Пусть энергия каждой из них E и, следовательно, они описываются найденной выше волной де Бройля. Поставим на пути этих микрочастиц потенциальный барьер (рис. 34.4) с уровнем энергии V_0 .

Рис. 34.4. Иллюстрация к явлению прохождения микрочастицы сквозь потенциальный барьер ширины a , в частности, и тогда, когда энергия микрочастицы меньше энергетической высоты барьера V_0



В классическом случае, если $E > V_0$, частицы пройдут через потенциальный барьер и продолжат свое движение с прежней скоростью вдоль оси Ox . Если же $E < V_0$, то частицы полностью отразятся и ни одна из них через барьер не пройдет.

Оказывается, что для макрочастиц все будет не так. Только часть из них при $E > V_0$ «просочится» через барьер, а часть отразится обратно. Частичное отражение будет иметь место и при $E < V_0$. Остальные преодолели барьер, хотя их энергия E меньше энергетической высоты барьера V_0 . Если рассуждать в терминах волнового дуализма, то такое поведение довольно естественно: падающая на потенциальный барьер волна де Бройля частично отразится, а частично, преломившись, пройдет через него.

Посмотрим, что дает в этом случае математический формализм модели микрочастицы. Будем искать стационарное решение уравнения Шредингера с энергией E . Учитывая, что потенциальная энергия $V(x)$ на рис. 34.4 описывается формулой

$$V(x) = \begin{cases} 0 & \text{при } x < 0, \\ V_0 & \text{при } 0 \leq x \leq a, \\ 0 & \text{при } x > a, \end{cases}$$

найдем, что уравнение (34.22), вытекающее из уравнения Шредингера, для разных x имеет различный вид, так что

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{\partial^2 \bar{\psi}}{\partial x^2} &= E \bar{\psi} \quad \text{при } x < 0, \\ -\frac{\hbar^2}{2m} \frac{\partial^2 \bar{\psi}}{\partial x^2} &= (E - V_0) \bar{\psi} \quad \text{при } 0 \leq x \leq a, \\ -\frac{\hbar^2}{2m} \frac{\partial^2 \bar{\psi}}{\partial x^2} &= E \bar{\psi} \quad \text{при } x > a. \end{aligned} \quad (34.44)$$

Решение $\bar{\psi}$ этих уравнений должно быть ограничено и непрерывно вместе с первой производной.

Вводя для краткости обозначения

$$\frac{2mE}{\hbar^2} = k^2, \quad \frac{2m(E - V_0)}{\hbar^2} = k_1^2,$$

из уравнений (34.44) непосредственно находим, что $\bar{\psi}$ будет таково:

$$\begin{aligned}\bar{\psi} &= Ae^{ikx} + Be^{-ikx} \quad \text{при } x < 0, \\ \bar{\psi} &= Ce^{ik_1x} + De^{-ik_1x} \quad \text{при } 0 \leq x \leq a, \\ \bar{\psi} &= Ee^{ikx} + Fe^{-ikx} \quad \text{при } x > a.\end{aligned}\tag{34.45}$$

Условия гладкости функции $\bar{\psi}(x)$ при $x = 0$ и $x = a$ приводят к четырем условиям:

$$\begin{aligned}A + B &= C + D, \\ ikA - ikB &= ik_1C - ik_1D, \\ Ce^{ik_1a} + De^{-k_1a} &= Ee^{ika} + Fe^{-ika}, \\ ik_1Ce^{ik_1a} - ik_1De^{-ik_1a} &= ikEe^{ika} - ikFe^{-ika}\end{aligned}\tag{34.46}$$

для шести неизвестных комплексных постоянных A, B, C, D, E и F . Так как уравнений только четыре, то их решение неоднозначно. Это и понятно, так как функция $\bar{\psi}$ может быть определена с точностью до постоянного множителя и, кроме того, по физическому смыслу задачи из бесконечности справа налево не могут двигаться микрочастицы, что еще не было учтено. В силу этого можно принять, что $A = 1$ и $F = 0$, после чего уравнения (34.46) решаются однозначно и

$$\begin{aligned}B &= \frac{(k^2 - k_1^2)[\exp(ik_1a) - \exp(-ik_1a)]}{(k - k_1)^2 \exp(ik_1a) - (k + k_1)^2 \exp(-ik_1a)}, \\ E &= \frac{-4kk_1}{(k - k_1)^2 \exp(ik_1a) - (k + k_1)^2 \exp(-ik_1a)}, \\ C &= \frac{k + k_1 + (k_1 - k)B}{2k_1}, \quad D = \frac{k_1 - k + (k + k_1)B}{2k_1}.\end{aligned}$$

Из этих формул следует, что для отраженной и преломленной волн де Бройля

$$\begin{aligned}|B|^2 &= \frac{4(k^2 - k_1^2) \sin^2 k_1a}{4(k^2 + k_1^2)^2 \sin^2 k_1a + 16k^2k_1^2 \cos^2 k_1a}, \\ |E|^2 &= \frac{16k^2k_1^2}{4(k^2 + k_1^2)^2 \sin^2 k_1a + 16k^2k_1^2 \cos^2 k_1a}\end{aligned}$$

и, следовательно, наряду с отраженными от барьера частицами существуют проходящие через барьер частицы вне зависимости от того, больше или меньше энергия частицы энергетической высоты V_0 барьера. Это совершенно новое явление «просачивания» или диффузии через энергетический барьер микрочастицы, не имеющее классического аналога и противоречащее классике.

Атом водорода. Атом водорода – это электрон в центральном кулоновском поле электрически положительного заряда ядра. Классическим аналогом описания атома водорода является известная нам проблема двух тел, решенная Ньютоном. В классическом случае движение тела характеризуется тремя законами Кеплера об эллиптичности орбиты, постоянстве секторной скорости и соотношении периодов обращения и больших полуосей орбит. Для микрочастицы все не так:

нет орбиты, энергия и импульс квантованы и одновременно заведомо не имеют определенных значений. Как оказывается, возможные значения энергии

$$E_n = \frac{me^4}{2h^2} \frac{1}{n^2} \quad (n = 1, 2, 3, \dots) \quad (34.47)$$

зависят от целого числа n .

С математической точки зрения отыскание квантовых значений энергии E_n – необычайно эффективное решение сложной и трудной задачи о собственных значениях и функциях линейного дифференциального оператора.

Столь же эффективно подтверждение получаемых результатов наблюдениями частот спектральных линий атома водорода. Напомним, что частота ω испускаемого атомом фотона (волны де Бройля) при переходе электрона с энергетического уровня E_n на E_m определяется соотношением

$$h\omega = E_n - E_m \quad (n < m),$$

что приводит к наблюдаемым сериям спектральных линий Лаймана

$$\omega_{1m} = E_1 - E_m/h \quad (m = 2, 3, 4, \dots),$$

Бальмера

$$\omega_{2m} = (E_2 - E_m)/h \quad (m = 3, 4, 5, \dots),$$

Фитца–Пашена

$$\omega_{3m} = (E_3 - E_m)/h \quad (m = 4, 5, 6, \dots)$$

и другими.

Дальнейшей целью является изложение постановки задачи и описание результатов ее решения для атомов водорода или водородоподобного иона с зарядом ядра z и одним электроном.

В рассматриваемом случае потенциальная энергия равна

$$E_n = \frac{me^4}{2h^2} \frac{1}{n^2} \quad (n = 1, 2, 3, \dots),$$

а уравнение Шредингера целесообразно записать в сферических координатах $0 \leq r < \infty$, $0 \leq \vartheta \leq \pi$, $0 \leq \varphi < 2\pi$:

$$\begin{aligned} -i\hbar \frac{\partial \Psi}{\partial t} &= -\frac{\hbar^2}{2m} \Delta \Psi + V(r)\Psi = \\ &= -\frac{\hbar^2}{2m} \left\{ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right] \right\} \Psi + V(r)\Psi. \end{aligned}$$

Поиск стационарных решений этого уравнения Шредингера приводит к задаче о собственных функциях и значениях E .

Отыскание непрерывных конечных при всех r , ϑ и φ (включая $r = \infty$) решений этого уравнения довольно сложная, красиво решаемая задача. Ее решение – результат труда многих выдающихся ученых – сейчас излагается в учебниках по квантовой механике и математической физике. Основные результаты этого исследования состоят в следующем. При $E > 0$ спектр непрерывен и соответствующие собственные функции не обращаются в ноль на бесконечности. Это соответствует

возможности неограниченного ухода микрочастицы от ядра атома. Напротив, при $E < 0$ спектр дискретен и соответствующие собственные функции в бесконечности зануляются, что означает, что электрон остается на конечном расстоянии от ядра. Дискретные собственные значения равны

$$E_n = -\frac{z^2 m e^4}{2h^2} n^{-2} \quad (n = 1, 2, 3, \dots).$$

Каждому из этих собственных значений отвечает конечное число собственных функций вида

$$\bar{\Psi}_{nlm} = R_{nl}(r) Y_{lm}(\vartheta, \psi) \quad (l = 0, 1, 2, \dots, n-1; m = 0, \pm 1, \dots, \pm l).$$

Как уже отмечалось, это полностью соответствует наблюдаемым спектрам атома водорода.

Квантовый линейный осциллятор. В предыдущих лекциях мы уделили значительное внимание классическому линейному гармоническому осциллятору, описываемому уравнением

$$m\ddot{x} + kx = 0$$

и функцией Гамильтона

$$H = \frac{p_x^2}{2m} + \frac{m\omega^2 x^2}{2} \quad \left(p_x = m\dot{x}, \quad \omega^2 = \frac{k}{m} \right).$$

Соответствующий квантовый осциллятор описывается уравнением Шредингера:

$$ih \frac{\partial \Psi}{\partial t} = -\frac{h^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + \frac{m\omega^2}{2} x^2 \Psi \quad (\Psi = \Psi(x, t)),$$

отыскание стационарных решений которого с энергией E приводит к уравнению

$$-\frac{h^2}{2m} \frac{d^2 \bar{\Psi}}{dx^2} + \frac{m\omega^2}{2} x^2 \bar{\Psi} = E \bar{\Psi}.$$

Найти его собственные значения и функции непросто. Это непростое решение дает, что

$$E_n = h\omega \left(n + \frac{1}{2} \right) \quad (n = 0, 1, 2, \dots)$$

и соответствующие нормированные собственные функции равны

$$\bar{\Psi}_n(x) = \frac{1}{\sqrt{x_0}} H_n \left(\frac{x}{x_0} \right) \exp \left(-\frac{x^2}{2x_0^2} \right),$$

где

$$x_0 = \sqrt{\frac{h}{m\omega}}, \quad H_n(y) = \frac{(-1)^n}{\sqrt{2^n n! \sqrt{\pi}}} e^{y^2} \frac{d^n e^{-y^2}}{dy^n}.$$

Первые три функции $\bar{\Psi}_0$, $\bar{\Psi}_1$ и $\bar{\Psi}_2$ изображены на рис. 34.5.

Согласно сказанному, значения энергии эквидистантно квантованы, и минимально возможная энергия больше нуля и равна $h\omega/2$. Никак не похожи и графики $|\Psi_n|^2$ (см. рис. 34.5) на график вероятности обнаружения в случайный момент вре-

мени классического осциллятора вместе с координатой x . Этот последний график изображен на рис. 34.6.

Рис. 34.5. Виды первых трех функций $\bar{\psi}$ для стационарных решений уравнения Шредингера для квантового линейного осциллятора

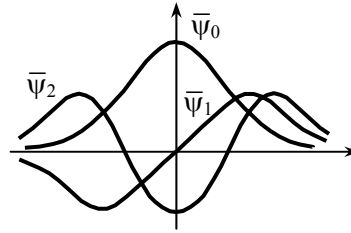
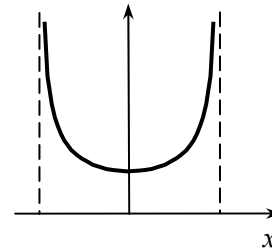


Рис. 34.6. Плотность вероятности величины x классического линейного осциллятора. Вертикальные асимптоты соответствуют экстремальным возможным значениям x ($x = \pm a$, где a – амплитуда колебаний). (Ср. с рис. 34.5)



Квантовое уравнение Ньютона. Движение микрочастицы описывается изменением ее функции состояния и ни в коей мере не напоминает движение по траектории. Ведь если известно место частицы, то неопределенна ее скорость, т.е. нет траектории. Вместе с тем, быстро летящая микрочастица оставляет в пузырьковой камере след от своего движения. Что же представляет собой этот след?

Дело в том, что в этом случае величины дисперсий координаты и скорости (импульса) настолько малы, что след «случайного блуждания» от образования одной капельки к другой все же представляется нам траекторией, поскольку он достаточно тонок. Ширина этой «траектории» гигантских размеров по сравнению с атомом, но в нашем масштабе макромира она тонкая.

Согласимся с тем, что все время неопределенности положения и скорости очень малы, но почему же движение в целом соответствует нашим классическим представлениям? На этот вопрос в свое время ответил П. Эренфест, обнаружив, что для среднего значения $\langle x \rangle$ координаты x имеет место аналог уравнения Ньютона, который можно назвать квантовым уравнением Ньютона:

$$m \frac{d^2}{dx^2} \langle x \rangle = - \left\langle \frac{\partial V}{\partial x} \right\rangle \approx - \frac{\partial V(\langle x \rangle)}{\partial x}.$$

Вывод его не очень сложен, содержится в учебниках по квантовой механике и опирается только на изложенные вам ранее положения модели микрочастицы. Для макротел $\langle x \rangle$ неотличимо от x , поскольку дисперсия x очень мала.

35. ПРОСТРАНСТВО И ВРЕМЯ

В предыдущих лекциях рассматривалось много разных математических моделей. Почти во всех них фигурировали пространственные переменные и время. Что это такое, не объяснялось, их смысл казался самоочевидным. И. Кант считал представления о пространстве и времени врожденными. Но оказалось, что они все же требуют пояснений и заслуживают особого внимания, и пришло время задуматься над тем, что за пространственные переменные и время стоят в великих математических моделях механики и физики, а, следовательно, и во всех математических моделях, которые на них основаны. Обратите внимание: вопрос не о том, что такое пространство и время вообще, вопрос конкретен: что за переменные стоят в математических моделях естествознания и техники? И будет поставлен еще конкретнее: как связаны между собой эти переменные в разных инерциальных системах, движущихся относительно друг друга с постоянными скоростями?

Ньютон считал, что существует выделенная абсолютная система отсчета. В ней координаты измеряются принятым масштабом, а время – абсолютными часами, что существуют абсолютное пространство и абсолютное время.

Окружающее нас пространство еще Евклид подверг математическому изучению, результатом которого была евклидова геометрия и наше евклидово пространство. Однако уже Гаусс усомнился, что наше пространство евклидово и хотел измерить углы большого треугольника, образованного вершинами гор. Если бы сумма углов этого треугольника отличалась от π , то это свидетельствовало бы о неевклидовости нашего пространства. Появившиеся затем геометрии Лобачевского и Римана подвели теоретическую базу под эти сомнения Гаусса, но измерения в масштабах Земли не дали достоверных отклонений от евклидовости.

Сомнения о характере геометрии нашего пространства зародились сравнительно давно, но что казалось бесспорным, так это то, что пространство – это пространство, а время – это время, они как бы существуют отдельно сами по себе и никак не связаны друг с другом. И вот именно это оказалось неверным. Выяснилось это в начале двадцатого века. Об этом и будет речь дальше. Выяснилось, что нужно говорить не о трехмерном пространстве и одномерном времени, а о едином четырехмерном пространственно-временном мире.

Координатные системы, в которых справедливы законы Ньютона, называют инерциальными. Инерциальные системы – это системы, движущиеся относительно друг друга с постоянной скоростью. Если направить оси Ox и $O'x'$ вдоль этой скорости относительного движения систем $Oxyz$ и $O'x'y'z'$, то координаты и время одной и той же точки в этих разных системах отсчета будут связаны соотношениями:

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t, \quad (35.1)$$

получившими название преобразований Галилея. Факт справедливости уравнений Ньютона во всех инерциальных системах и, в частности, в системах $Oxyz$ и $O'x'y'z'$ означает, что уравнения Ньютона инвариантны (сохраняют свой вид) при переходе от переменных x, y, z, t к x', y', z', t' , согласно (35.1). Действительно, уравнения

$$m\ddot{x} = F_x, \quad m\ddot{y} = F_y, \quad m\ddot{z} = F_z \quad (35.2)$$

после замены переменных, согласно (35.1), переходят в такие же уравнения для x' , y' , z' и t' . Обоснованность одинаковости сил F_x , F_y и F_z в обеих координатных системах можно видеть в том, что это так для взаимодействий, зависящих от расстояний или величин деформаций упругих тел. Итак, уравнения Ньютона и все, вытекающее из них, т.е. все уравнения механических систем, инвариантны по отношению к преобразованиям Галилея. Эта инвариантность есть следствие справедливости законов Ньютона во всех инерциальных системах координат, связанных между собою преобразованиями Галилея.

Теперь от механики перейдем к электродинамике. Весь наш опыт и специально поставленные эксперименты свидетельствуют, что уравнения Максвелла имеют место в инерциальных системах, т.е. они тоже должны быть инвариантны относительно преобразований Галилея. Не будем трогать всю систему уравнений Максвелла. Ограничимся одним из ее следствий – волновым уравнением для плоской электромагнитной волны, распространяющейся вдоль оси Ox и электрическим полем, направленным по оси Oy . Такое уравнение имеет вид:

$$\frac{\partial^2 E_y}{\partial t^2} = c^2 \frac{\partial^2 E_y}{\partial x^2}. \quad (35.3)$$

Это волновое уравнение, записанное в системе $Oxyz$, после преобразований Галилея (35.1) должно перейти опять-таки в волновое уравнение. Во что должно перейти E_y , мы не знаем и не будем задумываться. Ясно, что оператор

$$\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} \quad (35.4)$$

должен перейти в оператор

$$\frac{\partial^2}{\partial t'^2} - c^2 \frac{\partial^2}{\partial x'^2}. \quad (35.5)$$

Первый из них зануляется на функции $E_y(x, t)$, второй – на неизвестной нам функции $E'_y(x', t')$.

Непосредственно находим:

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{\partial}{\partial x'} \cdot \frac{\partial x'}{\partial x} + \frac{\partial}{\partial t'} \cdot \frac{\partial t'}{\partial x} = \frac{\partial}{\partial x'}, \quad \frac{\partial^2}{\partial x^2} \rightarrow \frac{\partial^2}{\partial x'^2}, \\ \frac{\partial}{\partial t} &= \frac{\partial}{\partial x'} \frac{\partial x'}{\partial t} + \frac{\partial}{\partial t'} \frac{\partial t'}{\partial t} = -v \frac{\partial}{\partial x'} + \frac{\partial}{\partial t'}, \\ \frac{\partial^2}{\partial t^2} &= -v \frac{\partial^2}{\partial x'^2} \frac{\partial x'}{\partial t} - v \frac{\partial^2}{\partial x' \partial t'} \frac{\partial t'}{\partial t} + \frac{\partial^2}{\partial t \partial x'} \frac{\partial x'}{\partial t} + \frac{\partial^2}{\partial t'^2} \frac{\partial t'}{\partial t} = \\ &= v^2 \frac{\partial^2}{\partial x'^2} - 2v \frac{\partial^2}{\partial x' \partial t'} + \frac{\partial^2}{\partial t'^2}, \end{aligned}$$

и поэтому оператор (35.4) переходит после замены переменных (35.1) – преобразования Галилея – в оператор

$$\frac{\partial^2}{\partial t'^2} - 2v \frac{\partial^2}{\partial x' \partial t'} - (c^2 - v^2) \frac{\partial^2}{\partial x'^2}, \quad (35.6)$$

отличный от волнового оператора (35.5).

С одной стороны, это неожиданно и означает, что либо неверны утверждения о справедливости уравнений Ньютона и Максвелла в инерциальных системах, либо неверно преобразование Галилея, и тогда неверны уравнения Ньютона, либо, наконец, неверны уравнения Максвелла. Какую из этих трех возможностей выбрать?

С другой стороны, то, что оператор (35.4) не перешел при преобразовании Галилея в такой же оператор (35.5), довольно очевидно: ведь система $O'x'y'z'$ движется со скоростью v относительно системы $Oxyz$, и поэтому при таком переходе скорость c распространения электромагнитных волн должна измениться, должна измениться и форма волны. Убедимся в этом, найдя гармонические волны $e^{i(\omega t' - kx')}$, зануляющие оператор (35.6):

$$\begin{aligned} & \left\{ \frac{\partial^2}{\partial t'^2} - 2v \frac{\partial^2}{\partial x' \partial t'} + (v^2 - c^2) \frac{\partial^2}{\partial x'^2} \right\} e^{i(\omega t' - kx')} = \\ & = [-\omega^2 - 2v\omega k - (v^2 - c^2)k^2] e^{i(\omega t' - kx')} = 0, \end{aligned}$$

откуда

$$(\omega + vk)^2 = c^2 k^2$$

или

$$\omega/k = -v \pm c.$$

Это означает, что в системе $O'x'y'z'$ при $v < c$ в одну сторону волна распространяется со скоростью $c - v$, а в другую – со скоростью $-c - v$.

Для выяснения, разная ли скорость распространения света в разных инерциальных системах и зависит ли она от движения источника света, проводились специальные очень точные и основательные опыты. Знаменитый опыт Майкельсона – Марли показал, что скорость света во всех системах одинакова и равна $c \approx 300\,000$ км/с. Это говорит о том, что преобразования Галилея неверные и надлежит найти преобразование, которое сохраняет вид волнового оператора. Такие преобразования впервые были написаны Лоренцем и названы его именем, но он не связывал их с общей проблемой пространства и времени.

Преобразования Лоренца должны заменить преобразования Галилея и могут подсказать, как уточняются уравнения Ньютона. Эти уточнения очень маленькие – и в механических наблюдениях, и экспериментах с $v \ll c$ себя не проявляли.

Столь же незначительны отличия преобразований Лоренца от преобразований Галилея. Но при скоростях, близких к скорости света, достижимых в современных ускорителях и наблюдаемых у космических частиц, не учитывать их уже нельзя.

Напишем знаменитые уравнения Лоренца:

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad y = y', \quad z = z', \quad t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \quad (35.7)$$

и перейдем к их выводу. Он не очень сложен. Сложнее понять и осмыслить, что из них следует. Некоторые простейшие следствия мы рассмотрим, а сейчас только отметим, что одно и то же событие, происходящее в некоторой точке и какой-то момент времени, имеет в разных инерциальных системах не только различные координаты, но и времена. Два события, одновременные в одной инерциальной системе, не одновременны в другой и различие в их временах зависит как от их относительной скорости v , так и удаленности r мест, в которых они происходят.

Действительно, пусть в одно и то же время t произошли события в точках x и $x + r$. Это в одной системе отсчета. Назовем ее K . Тогда в другой K' , движущейся относительно K со скоростью v , они произошли во времена

$$t'_1 = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \quad \text{и} \quad t'_2 = \frac{t - v(x+r)/c^2}{\sqrt{1 - v^2/c^2}},$$

различающиеся на величину

$$t'_1 - t'_2 = \frac{vr}{\sqrt{1 - v^2/c^2}}.$$

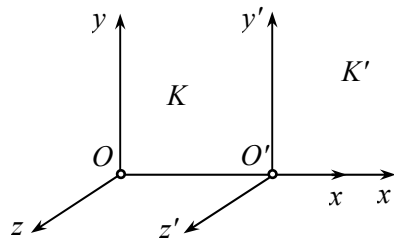
Мы выведем уравнения Лоренца (35.7) из свойств однородности и изотропности пространства, однородности времени и инвариантности волнового оператора – постоянства скорости распространения электромагнитных волн – во всех инерциальных системах. Как показал А. Пуанкаре, по отношению к преобразованиям Лоренца инвариантно не только следующее из уравнений Максвелла волновое уравнение, но и сами уравнения Максвелла.

Будем исходить из общего вида связи координат и времени x, t и x', t' одного и того же события в разных инерциальных системах координат K и K' , движущихся относительно друг друга с постоянной скоростью v : система K' относительно K (ее начало координат) со скоростью v , а система K относительно K' со скоростью $-v$. Движение их относительно друг друга происходит вдоль совпадающих осей Ox и $O'x'$ (рис. 35.1). В момент времени $t = t' = 0$ начала координат O и O' обеих систем совпадают. Будем исходить из общего вида связей координат и времени систем K и K' , удовлетворяющего условиям однородности и изотропности пространства и однородности времени:

$$x' = ax + bt, \quad y' = y, \quad z' = z, \quad t' = \alpha x + \beta t, \quad (35.8)$$

точнее, попробуем удовлетворить всем вышесформулированным условиям, исходя из такого вида преобразований, подбирая параметры a, b, α и β . Преобразование Галилея соответствует $a = 1, b = -v, \alpha = 0, \beta = 1$, и, как мы уже знаем, оно не подходит, поскольку требованиям однородности и изотропности удовлетворяет, а волновой оператор не сохраняет. Итак, нам надлежит найти значения четырех параметров a, b, α и β .

Рис. 35.1. Две инерциальные системы K и K' , движущиеся относительно друг друга с постоянной скоростью вдоль оси Ox , совпадающей с $O'x'$



Начнем с требования сохранения волнового оператора. Повторяя уже знакомые вам вычисления, получим:

$$\begin{aligned}\frac{\partial^2}{\partial x^2} &= a^2 \frac{\partial^2}{\partial x'^2} + 2a\alpha \frac{\partial^2}{\partial x' \partial t'} + \alpha^2 \frac{\partial^2}{\partial t'^2}, \\ \frac{\partial^2}{\partial t^2} &= b^2 \frac{\partial^2}{\partial x'^2} + 2b\beta \frac{\partial^2}{\partial x' \partial t'} + \beta^2 \frac{\partial^2}{\partial t'^2},\end{aligned}$$

и поэтому при замене переменных (35.8) волновой оператор (35.4) переходит в оператор

$$\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} = (b^2 - c^2 a^2) \frac{\partial^2}{\partial x'^2} + (2b\beta - 2a\alpha c^2) \frac{\partial^2}{\partial x' \partial t'} + (\beta^2 - c^2 \alpha^2) \frac{\partial^2}{\partial t'^2}.$$

Преобразованное к новым переменным волновое уравнение, отвечающее оператору (35.4), будет совпадать с волновым уравнением, соответствующим оператору (35.5), если их операторы различаются множителем, т.е. при выполнении условий

$$\frac{b^2 - c^2 a^2}{-c^2} = \frac{2(b\beta - a\alpha c^2)}{0} = \frac{\beta^2 - c^2 \alpha^2}{1},$$

которые влекут два условия:

$$b\beta - a\alpha c^2 = 0, \quad c^2 a^2 - b^2 = c^2(\beta^2 - c^2 \alpha^2). \quad (35.9)$$

Два условия на параметры a , b , α и β следуют из того, что начало координат O' движется относительно O со скоростью v , а O относительно O' – со скоростью $-v$. Точка O' имеет в системе K' координату $x' = 0$. Поэтому $ax + bt = 0$, где x – ее координата в системе K спустя время t , т.е. скорость движения точки O' в системе K равна $v = -b/a$. Аналогично точка O в системе K' имеет координату $x' = bt$ в момент времени $t' = \beta t$. Поэтому ее скорость в системе K' равна b/β , что ведет к еще одному соотношению $b\beta = -v$.

Таким образом, к двум соотношениям (35.9) добавляются еще два:

$$-b/a = v, \quad b/\beta = -v. \quad (35.10)$$

Как оказывается, эти четыре соотношения зависимые и не позволяют найти все четыре параметра. Но их все можно выразить через один, например a :

$$b = -av, \quad \beta = a, \quad \alpha = -\frac{av}{c^2}. \quad (35.11)$$

В соответствии с этими соотношениями (35.11) искомое преобразование (35.8) записывается в виде:

$$x' = ax - avt, \quad t' = -\frac{av}{c^2}x' + at. \quad (35.12)$$

Смотря на (35.12), есть соблазн положить $v = 0$ и найти, что $a = 1$. Это верно, но только в том смысле, что при $v = 0$ $a = 1$. А что будет при других $v \neq 0$?

Используем изотропность пространства и то, что переход от системы K' к K должен иметь такой же вид (35.12), но с заменой v на $-v$. В соответствии с этим разрешим уравнения (35.12) относительно x , t и найдем

$$x = \frac{x' + vt'}{a(1 - v^2/c^2)}, \quad t = \frac{x'v/c^2 + t'}{a(1 - v^2/c^2)}. \quad (35.13)$$

Преобразование (35.12) должно совпадать с (35.13), если в последнем v заменить на $-v$, а x', t' заменить на x, t и обратно $-x, t$ на x', t' . Одинаковость влечет условия:

$$a = \frac{1}{a(1 - v^2/c^2)}, \quad av = \frac{v}{a(1 - v^2/c^2)}, \quad \frac{av}{c^2} = \frac{v/c^2}{a(1 - v^2/c^2)},$$

которые сводятся к одному

$$a^2 \left(1 - \frac{v^2}{c^2} \right) = 1.$$

Отсюда

$$a = \pm \left(1 - \frac{v^2}{c^2} \right)^{-1/2}.$$

Но мы уже знаем, что при $v = 0$ $a = 1$. Поэтому $a = (1 - v^2/c^2)^{-1/2}$ и преобразование (35.8) оказывается имеющим вид уже написанных ранее преобразований Лоренца (35.7).

Перепишем его и обратное ему, используя сокращенное обозначение $\gamma(v^2) = (1 - v^2/c^2)^{-1/2}$, в виде

$$x' = \gamma(v^2)(x - vt), \quad t' = \gamma(v^2) \left(t - \frac{v}{c^2} x \right) \quad (35.14)$$

и

$$x = \gamma(v^2)(x' + vt'), \quad t = \gamma(v^2) \left(t' + \frac{v}{c^2} x' \right). \quad (35.15)$$

Напомним, что v – скорость движения системы K' по отношению к K или, что то же, $-v$ – скорость движения системы K по отношению к системе K' .

Еще раз обратим внимание, что из преобразований Лоренца следует, что одно и то же событие – нечто происходящее в каком-то месте и в какое-то время – имеет в разных инерциальных системах отсчета не только разные координаты, что нам привычно, но и разные времена. Действительно, пусть в системе отсчета K событие произошло в точке x в момент t . Тогда, согласно (35.7), в системе K' оно имело место в точке $x' = \gamma(v^2)(x - vt)$ в момент $t' = \gamma(v^2)(t - vx/c^2)$. Это означает, что абсолютного времени, как и координат, нет. Каждая инерциальная система имеет свой отсчет места и свои часы.

Теперь уже не столь неожиданно и понятно, что события, одновременные в одной системе, не одновременны в другой. Понятие одновременности не абсолютное, а относительное, так как в каждой системе отсчета свои часы. Эти часы соотносятся друг с другом очень странно: часы системы K' по отношению к часам системы K отстают, но и часы системы K , наблюдаемые в системе K' , так же отстают. Этот вывод, естественный с точки зрения логики, учитывающей симметрию систем K и K' , совершенно неприемлем с точки зрения нашей привычной оценки хода разных часов.

Будем наблюдать из системы K часы системы K' , находящиеся в точке x' , в течение времени τ по часам системы K , т.е. от момента t_1 до t_2 , где $t_2 - t_1 = \tau$. Отвечающие этим временам t_1 и t_2 показания часов системы K' , которые находятся в точке x' , пусть t'_1 и t'_2 . Тогда, согласно второму из уравнений (35.15),

$$t_1 = \gamma(v^2) \left(t'_1 + \frac{vx'_1}{c^2} \right), \quad t_2 = \gamma(v^2) \left(t'_2 + \frac{vx'_2}{c^2} \right)$$

и, следовательно,

$$\tau = t_2 - t_1 = \gamma(v^2)(t'_2 - t'_1) = \gamma(v^2)\tau', \quad (35.16)$$

т.е. часы системы K' в системе K идут медленнее в $\gamma^{-1}(v^2)$ раз. Аналогично часы в системе K по часам системы K' также идут настолько же медленнее. Замедление хода движущихся часов по отношению к неподвижным имеет прямые экспериментальные подтверждения. Такое подтверждение дают измерения длительности жизни коротко живущих микрочастиц, летящих с большими скоростями. Быстро движущаяся частица в неподвижных часах живет дольше, и это удлинение времени жизни согласуется с тем, которое следует из преобразований Лоренца, т.е. больше в $\gamma(v^2)$ раза. Время жизни летящей частицы – это время в часах координатной системы, движущейся вместе с частицей. Такое время называют собственным временем частицы. Из равноправия всех инерциальных систем собственное время жизни частицы не зависит от ее скорости движения, оно во всех инерциальных системах одно и то же.

Некоторое подтверждение имеется и для часов, путешествующих на искусственном спутнике, по отношению к земным часам путешествующие часы отстают.

Вместе с тем из этого подтвержденного экспериментом факта замедления движущихся часов вытекает широко обсуждаемый в свое время парадокс близнецов. Он состоит в том, что из двух родившихся близнецов одного оставили на Земле, а другого отправили на космическом корабле со скоростью, близкой к скорости света. Тогда по возвращении на Землю, путешествующий близнец может быть юношей, а оставшийся на Земле – стариком. Физики считают, что так и будет. Но парадокс в том, что можно то же самое рассуждение провести с точки зрения путешествующего близнеца, по отношению к которому путешествовал близнец, оставшийся на Земле, и поэтому он должен быть стариком, а близнец на Земле – юношей. Этот парадокс разрешается тем, что полной симметрии между близнецами нет: путешествующий близнец тормозил и, повернув обратно, разогнался и что происходит при этом с его собственными часами, наша теория ответа не дает. Близнец же на Земле ничего такого не делал.

Рассмотрим теперь другое явление сокращения размеров движущегося отрезка. Пусть в системе K' вдоль оси Ox' расположен отрезок $[x'_1, x'_2]$ длины $l' = x'_2 - x'_1$. Длину l' он имеет в масштабе системы K' . Этот отрезок, назовем его L' , движется со скоростью v относительно системы K . Примем следующую естественную процедуру измерения его длины в системе K .

Засечем в системе K координаты x_1 и x_2 концов отрезка L' в какой-нибудь один и тот же для обоих концов момент времени t в системе K . За его длину в системе K примем величину $l = x_2 - x_1$. Найдем соотношение между l и l' .

Согласно первому соотношению (35.15) и второму (35.14), имеем:

$$\begin{aligned}x_1 &= \gamma(v^2)(x'_1 + v t'_1), & x_2 &= \gamma(v^2)(x'_2 + v t'_2), \\t'_1 &= \gamma(v^2)\left(t - \frac{v}{c^2}x_1\right), & t'_2 &= \gamma(v^2)\left(t - \frac{v}{c^2}x_2\right),\end{aligned}$$

откуда

$$\begin{aligned}l &= x_2 - x_1 = \gamma(v^2)[x'_2 - x'_1 + v(t'_2 - t'_1)] = \\&= \gamma(v^2)\left[l' - \gamma(v^2)\frac{v^2}{c^2}(x_2 - x_1)\right] = \gamma(v^2)\left[l' - \gamma(v^2)\frac{v^2}{c^2}l\right]\end{aligned}$$

и поэтому

$$\left(1 + \gamma^2 \frac{v^2}{c^2}\right)l = \gamma(v^2)l'$$

или

$$l = \gamma^{-1}(v^2)l'. \quad (35.17)$$

Это последнее соотношение означает, что при измерении в системе K отрезка длины l' системы K' происходит его уменьшение в $\gamma^{-1}(v^2)$ раз.

Преыдушие следствия из преобразований Лоренца поясняли нам особенности физического пространства и времени. Конечно, это одновременно и то пространство и время, в котором мы живем, но чтобы их обнаружить, мы должны иметь дело со скоростями порядка скорости света или обладать способностью регистрировать ничтожные по величине эффекты.

Следующим неожиданным, но тоже для скоростей, много меньших скорости света, едва заметным эффектом является неверность правила сложения переносной и относительной скоростей. В свое время мы вывели формулу разгона ракеты и установили, что с неограниченным уменьшением ее массы она неограниченно увеличивает свою скорость. Этот вывод следовал из того, что с каждым выбросом массы dm со скоростью c относительно ракеты она получала приращение в скорости, равное $c dm/m$. Это было приращение скорости ракеты в системе отсчета, связанной с ракетой. Из преобразований Лоренца следует, что в системе отсчета, связанной с Землей, с которой она запущена, ракета такого приращения скорости не получит, оно будет меньше, и за счет этого уменьшения скорость ракеты не может превзойти скорость света, но может приблизиться к ней.

Итак, пусть система K' движется, как и ранее, относительно системы K со скоростью v , и в системе K' относительно нее движется точка со скоростью u' . Тогда скорость u этой же точки относительно системы K будет $v + u'$. Это по классическим представлениям о пространстве и времени. В механике скорость v называется переносной, а u' — относительной, движение точки в системе K — абсолютным, а движение этой же точки в системе K' — относительным. В этой терминологии абсолютная скорость равна сумме переносной и относительной. Посмотрим, что будет при наших новых воззрениях на пространство и время. Пусть движущаяся точка M в системах K и K' находится в точках x и x' в моменты времени t и t' . Теперь t не обязано совпадать с t' . Значения x , t и x' , t' связаны между собой преобразованиями Лоренца (35.14) и (35.15). Первое из них дает переход от системы

K к K' , а второе, наоборот, – от K' к K . Система K' движется относительно K со скоростью v , и точка M в системе K' движется со скоростью $u' = dx'/dt'$. Скорость этой же точки M в системе K равна $u = dx/dt$. Нам надлежит найти связь между этими скоростями u и u' , зная, что x, t и x', t' связаны соотношениями (35.14) и (35.15). Непосредственно имеем:

$$u = \frac{dx}{dt} = \frac{d\{\gamma(v^2)(x' + vt')\}}{d\{\gamma(v^2)(t' + vx'/c^2)\}} = \frac{dx' + vdt'}{dt' + (v/c^2)dx'} = \frac{u' + v}{1 + vu'/c^2}.$$

Таким образом, действительно скорости v и u' не складываются. Имеет место более сложная зависимость

$$u = \frac{v + u'}{1 + u'v/c^2}, \tag{35.18}$$

которая при $v \ll c$ и $u' \ll c$ близка к простому сложению, но при $v = c$ и $u' = c$ $u = c$, а не $2c$, и отличие весьма велико.

Можно заметить, что при $v < c$ и $u' < c$ и поэтому любое число прибавок к скорости ракеты скоростей, меньших скорости света, никогда не позволяет ей превзойти скорость света.

Естественным следствием изменений представлений о пространстве и времени является их объединение, необходимость их совместного рассмотрения. Ранее были отдельно трехмерное пространство и одномерное время, теперь четырехмерное пространство пространства и времени. Точками этого четырехмерного пространства являются события. В разных системах отсчета одно и то же событие описывается по-разному в системе Kx, y, z, t , и в системе $K'x', y', z', t'$, причем, могут отличаться друг от друга не только x и x', y и y', z и z' , но и t и t' . Преобразования Лоренца представляют некоторый частный случай, когда оси систем K и K' параллельны и движутся относительно друг друга вдоль оси Ox , совпадающей с Ox' . К этому частному случаю сводятся любые две системы \tilde{K} и \tilde{K}' , движущиеся друг относительно друга с постоянной скоростью, так как переход от \tilde{K} к K , как и от \tilde{K}' к K' – это некоторые переходы от одной пространственной системы к другой, и переход от \tilde{K} к \tilde{K}' можно представить как последовательность переходов от \tilde{K} к K , от K к K' и от K' к \tilde{K}' (рис. 35.2).

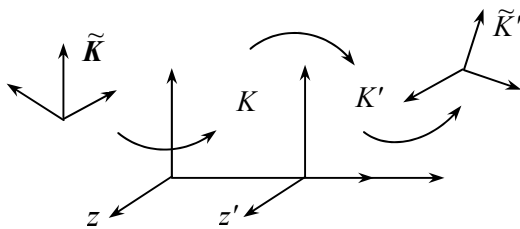


Рис. 35.2. Последовательные переходы от одной инерциальной системы к другой: системы \tilde{K} , K и K' , \tilde{K}' неподвижны относительно друг друга; K и K' движутся относительно друг друга с постоянной скоростью

При преобразованиях координат, связанных с переходом от одной системы к другой, неподвижной относительно нее, сохраняются расстояния

$$dr^2 = dx^2 + dy^2 + dz^2. \tag{35.19}$$

Оказывается, аналогично этому в четырехмерном пространстве-времени сохраняется величина

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2. \tag{35.20}$$

Для того чтобы в этом убедиться, достаточно показать, что это имеет место при преобразованиях Лоренца (35.15) от системы K к K' , так как при преобразованиях от системы \tilde{K} к K и от K' к \tilde{K}' это очевидно имеет место. Непосредственно используя (35.15), находим

$$\begin{aligned} dx^2 + dy^2 + dz^2 - c^2 dt^2 &= \\ &= [d(\gamma(v^2)(x' + vt'))]^2 + dy'^2 + dz'^2 - c^2 \left[d\left(\gamma(v^2) \left(t' + \frac{v}{c^2} x' \right) \right) \right]^2 = \\ &= \gamma^2(v)(dx'^2 + 2v dx' dt' + v^2 dt'^2) + dy'^2 + dz'^2 - \\ &- c^2 \gamma^2(v) \left(dt'^2 + 2 \frac{v}{c^2} dt' dx' + \frac{v^2}{c^4} dx'^2 \right) = dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2. \end{aligned}$$

Итак,

$$ds^2 = ds'^2. \tag{35.21}$$

Величину ds можно назвать расстоянием в четырехмерном пространстве событий. Преобразования, связанные с переходом от одной системы \tilde{K} к другой \tilde{K}' , его сохраняют и являются в этом смысле аналогами движений в трехмерном пространстве. Эту аналогию можно сделать явной, если время t заменить на $i\tau/c$ и, соответственно, v на $-icv$. При этом квадратичная форма ds^2 примет вид:

$$ds^2 = dx^2 + dy^2 + dz^2 + d\tau^2, \tag{35.22}$$

а преобразование Лоренца окажется вида:

$$x' = \frac{x + v\tau}{\sqrt{1 + v^2}}, \quad \tau' = \frac{\tau - vx}{\sqrt{1 + v^2}},$$

или, полагая $(1 + v^2)^{-1/2} = \cos \theta$ и соответственно $v(1 + v^2)^{-1/2} = \sin \theta$ в виде:

$$x' = x \cos \theta + \tau \sin \theta, \quad \tau' = -x \sin \theta + \tau \cos \theta, \tag{35.23}$$

представляющем собою переход от системы $Ox\tau$ к системе $Ox'\tau'$, повернутой относительно нее на угол θ (рис. 35.3).

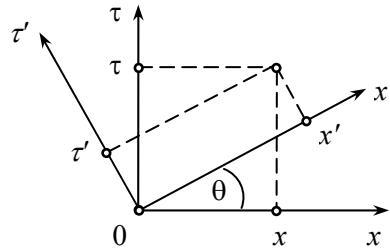


Рис. 35.3. Геометрическая интерпретация преобразования Лоренца

Таким образом, переход от одной инерциальной системы к другой можно представить как обычное преобразование в евклидовом пространстве от одной ортогональной системы координат к другой.

В связи с инвариантностью ds^2 , выражаемой (35.21), мне хотелось бы сказать несколько слов о причинности в новом пространственно-временном мире. Спросим себя, какие события $S_1(x_1, y_1, z_1, t_1)$ и $S_2(x_2, y_2, z_2, t_2)$ в разных точках (x_1, y_1, z_1, t_1)

и (x_2, y_2, z_2, t_2) и в разные времена t_1 и t_2 могут быть причинно связаны, и какие – нет? Согласно самому смыслу принципа причинности, событие S_2 может быть следствием события S_1 только в случае $t_2 \geq t_1$. При обратном неравенстве это невозможно. При переходе от одной инерциальной системы к другой времена t_2 и t_1 могут изменяться и может нарушаться неравенство $t_2 \geq t_1$. Величина

$$ds^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2$$

при переходе от одной инерциальной системы к другой не меняется, поэтому, если $ds^2 < 0$, то $t_2 - t_1$ не может обратиться в нуль, и первоначальное неравенство $t_2 > t_1$ сохраняется. Следовательно, наличие причинной связи между такими событиями может иметь место во всех системах отсчета. Этот случай характеризуется тем, что время прохождения светом расстояния

$$r = [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2}$$

меньше $t_2 - t_1$. Напротив, при $ds^2 > 0$ это время больше, чем $|t_2 - t_1|$. При этом с переходом к другой системе отсчета $t_2 - t_1$ может сменить знак, т.е. такие события не могут быть причинно связанными. Невозможность причинной связи между событиями, для которых $ds^2 > 0$, следует еще и из того, что действие, устанавливающее эту причинную связь, должно было бы распространяться быстрее скорости света. Тем самым мы приходим к выводу, что дальное действие противоречит соблюдению принципа причинности во всех инерциальных системах. Уравнения Максвелла устранили мгновенность кулоновского взаимодействия электрических зарядов, то же сделала современная теория гравитации и гравитационных волн с ньютоновским взаимодействием масс. Ни то, ни другое не может осуществляться быстрее скорости света.

Таким образом, принцип причинности сохраняется и в отсутствии абсолютного времени.

Уравнения Максвелла инвариантны по отношению к преобразованиям Лоренца и не инвариантны по отношению к преобразованиям Галилея. Уравнения Ньютона, наоборот, инвариантны по отношению к преобразованиям Галилея и не инвариантны по отношению к преобразованиям Лоренца. Различия преобразований Галилея и Лоренца порядка v^2/c^2 и очень малы для скоростей, наблюдаемых во времена Ньютона механических систем. Сегодня это уже не так. Наблюдаемые космические частицы имеют скорости, близкие к скорости света c , так же уже сравнимы со скоростью света скорости частиц современных мощных ускорителей. Наблюдения за их движением обнаруживают расхождения с уравнениями Ньютона. Уравнения Ньютона нужно подправить так, чтобы они оставались справедливыми и при больших скоростях v . Выяснить, каковы эти поправки, можно, исходя из требования их инвариантности по отношению к преобразованию Лоренца. Опуская все рассуждения, подведем окончательный итог.

Уравнение

$$\frac{d\bar{p}}{dt} = \bar{F},$$

где \bar{p} – количество движения или импульс, сохраняется, но сам импульс, ранее равный $m\bar{v}$, теперь равен

$$\bar{p} = \frac{m_0 \bar{v}}{\sqrt{1 - v^2/c^2}},$$

где m_0 – так называемая масса покоя. Полная энергия E частицы равна

$$E = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}},$$

а энергия покоя E_0 и кинетическая энергия T соответственно равны

$$E_0 = m_0 c^2, \quad T = E - E_0,$$

причем кинетическая энергия T

$$T = E - E_0 = \frac{m_0 v^2}{2} + \frac{3}{8} m_0 \frac{v^4}{c^2} + \dots,$$

и при $v \ll c$ близка к ее классическому значению $m_0 v^2/2$.

Таким образом, изменения, произошедшие с уравнением Ньютона, сводятся к тому, что масса m заменяется на массу

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} = m_0 \left(1 + \frac{1}{2} \frac{v^2}{c^2} + \dots \right),$$

зависящую от скорости и отличающуюся от прежней массы на величину порядка v^2/c^2 . Механика Ньютона, уточненная применительно к большим скоростям движущихся тел, получила наименование релятивистской механики. Быстро движущаяся частица со скоростью порядка скорости света названа релятивистской частицей.

36. РАЗГОН РЕЛЯТИВИСТСКИХ ЧАСТИЦ В ЦИКЛОТРОНЕ

Циклотрон – устройство для получения пучков заряженных частиц с очень большими скоростями, приближающимися к предельной скорости света. Идея циклотрона очень проста. Заряженная частица в однородном постоянном магнитном поле, имеющая начальную скорость, перпендикулярную к нему, движется по окружности с частотой ω в классическом приближении, не зависящей от ее скорости. Независимость ее частоты вращения от скорости позволяет осуществить резонансный разгон частицы переменным электрическим полем постоянной частоты ω . Реально разгон заряженной частицы производится не на всем ее пути, а только при прохождении ею узкой щели с разностью напряжений $V \sin \omega t$ (рис. 36.1). Пересекая щель в момент t , частица заряда e получает дополнительную энергию $eV \sin \omega t$. Совершив пол-оборота за время π/ω , она снова пересекает щель и снова получает добавку энергии, равный $-eV \sin(\omega t + \pi) = eV \sin \omega t$. Этот разгон при $eV \sin \omega t > 0$ мог бы продолжаться до тех пор, пока излучение частицы и потери энергии при ее движении не сравняются с величиной подкачиваемой энергии и не возникнет вращение с постоянной скоростью. Но этому мешает релятивистский эффект увеличения массы частицы с ростом скорости и следующее из него замедление угловой скорости. При этом условия резонанса нарушаются, резко падает подкачка энергии и прекращается дальнейший разгон частицы.

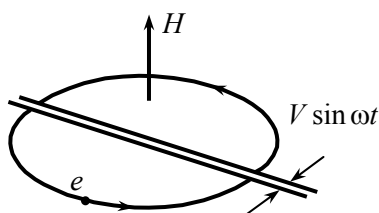


Рис. 36.1. Принципиальная схема циклотрона

Такова вкратце идея разгона частицы в циклотроне и проблема, возникающая при ее реализации, если мы хотим достичь скоростей, приближающихся к скорости света.

Поиск пути преодоления этой трудности естественно сопроводить построением и исследованием математической модели циклотрона, сосредоточив свое внимание только на механизме резонансного разгона заряженной частицы с учетом релятивистских эффектов.

Начнем с выяснения зависимости частоты вращения частицы от ее заряда e , массы m и величины магнитного поля H . На движущуюся со скоростью v перпендикулярно магнитному полю заряженную частицу действует сила Лоренца, равная по величине $e v H/c$. Эта сила, перпендикулярная магнитному полю и скорости движения частицы, не меняя величины ее скорости, создает центростремительное ускорение, соответствующее движению частицы по окружности некоторого радиуса R с частотой ν , так что

$$R\nu^2 = \nu v = \frac{e\nu H}{mc}$$

или

$$v = \frac{eH}{mc}. \quad (36.1)$$

Конечно, к такому же результату можно прийти непосредственно из уравнения Ньютона с учетом зависимости массы от скорости v

$$\frac{d}{dt}(m\bar{v}) = \frac{e}{c}[\bar{v}, H], \quad m = m_0 \left(1 - \frac{v^2}{c^2}\right)^{-1/2}, \quad (36.2)$$

предполагая, что начальная скорость \bar{v} ортогональна постоянному однородному магнитному полю \bar{H} . В формулах (36.1) и (36.2) масса m зависит от скорости. Учитывая, что энергия E частицы равна mc^2 , соотношение (36.1) можно записать в виде

$$v = \frac{eHc}{E}, \quad (36.3)$$

явно обнаруживающем зависимость частоты ν от скорости движения v . В классическом приближении, когда m постоянно, этой зависимости нет. Таким образом, если мы хотим сохранить равенство $\omega = \nu$ при изменении скорости частицы v , необходимо соответствующим образом менять частоту ω или величину магнитного поля H . Искусственное осуществление этих связей затруднительно, но, оказывается, нелинейные эффекты резонансного разгона позволяют реализовать необходимую связь между ω и H автоматически. Эта возможность следует из математической модели, построение которой мы продолжим.

Пусть E и \bar{E} – значения энергии одной и той же частицы после следующих друг за другом проходах ею через щель, тогда

$$\bar{E} = E - f(E) + eV \sin \omega t, \quad (36.4)$$

где t – момент времени прохода частицей через щель, после которого ее энергия равна \bar{E} , $f(E)$ – потери энергии частицы при движении между проходами через щели, а ω (считаемая постоянной) – частота изменения напряжения в щели: пренебрежимо малой ширины. Далее напишем соотношение между временем t и временем \bar{t} следующего прохода через щель

$$\bar{t} = t + \pi \frac{\bar{E}}{eHc}. \quad (36.5)$$

Если бы от момента t до \bar{t} энергия частицы оставалась постоянной, то в формуле (36.5) \bar{E} совпадало бы с \bar{E} . Но при движении от щели к щели она несколько уменьшает свою энергию, как уже было принято в формуле (36.4), на величину $f(\bar{E})$. Поэтому естественно приближенно принять, что

$$\bar{E} = \bar{E} - 1/2 f(\bar{E}).$$

Величины подкачки энергии при проходах щели в моменты времени t и \bar{t} соответственно будут $eV \sin \omega t$ и $-eV \sin \omega \bar{t} = eV \sin(\omega \bar{t} + \pi)$. Для осуществления резонансного разгона надо, чтобы для некоторого целого $n > 0$ имело место

$$\omega \bar{t} = \omega t + (2n - 1)\pi. \quad (36.6)$$

В простейшем случае $n = 1$, мы им ограничимся. Соотношение (36.6) означает, что фазы гармонически колеблющегося напряжения $\varphi = \omega t$ и $\bar{\varphi} = \omega \bar{t} + \pi$ в щелях в моменты прохода через них частицы совпадают – различаются на целое число 2π .

Из этого требования (36.6) при $n = 1$, согласно (36.5) и принятой величине \bar{E} , следует, что

$$\bar{E} - \frac{1}{2} f(\bar{E}) = \frac{1}{\omega} eHc. \quad (36.7)$$

Условие (36.7) определяет энергию частицы E^* , как корень уравнения (36.7). Далее из (36.4) при $\bar{E} = E$ приходим к соотношению

$$f(E^*) = eV \sin \omega t, \quad (36.8)$$

выполнение которого возможно только при

$$f(E^*) < eV. \quad (36.9)$$

При выполнении условия (36.9) уравнение (36.8) имеет два решения: $0 < t_1^* < \pi/2$ и $\pi/2 < t_2^* = \pi - t_1^* < \pi$. Найденным значениям E^* , t_1^* и E^* , t_2^* отвечают стационарные движения заряженной частицы, при которых последовательные значения энергии частицы E и \bar{E} и, соответственно, фаз φ и $\bar{\varphi}$ совпадают и равны E^* и φ_1^* или φ_2^* (фазы совпадают, если различаются на целое число 2π). В этих стационарных состояниях заряженная частица движется периодически с частотой $\nu = \omega$, т.е. имеет место резонанс.

Согласно уравнению (36.7), определяющему энергию E^* частицы в стационарном движении, она растет с увеличением магнитного поля H , но только до тех пор, пока выполняется условие существования стационарного режима (36.9), поскольку потери на движение частицы не могут в стационарном режиме превосходить максимально возможную ее подкачку со стороны электрического поля в щели.

Для того, чтобы стационарный режим и эта зависимость E^* от H могли реализоваться, стационарный режим должен быть устойчивым. Это условие не только необходимое, но и достаточное при достаточно медленном квазистационарном изменении H . Итак, мы пришли к необходимости исследовать устойчивость найденных стационарных движений.

Пусть dE и dt – малые возмущения стационарных значений E^* и φ^* , а $d\bar{E}$ и $d\bar{t}$ – отвечающие им изменения значений \bar{E} и \bar{t} при следующем проходе частицей щели. Согласно (36.4) и (36.5), находим, что

$$\begin{aligned} d\bar{E} &= dE - f'(E^*) dE + eV\omega \cos \omega t^* dt, \\ d\bar{t} &= dt + \frac{\pi}{eHc} \left(d\bar{E} - \frac{1}{2} f'(E^*) d\bar{E} \right). \end{aligned} \quad (36.10)$$

Полагая $d\bar{E} = z dE$ и $d\bar{t} = z dt$, приходим к характеристическому уравнению

$$\xi = \left| \begin{array}{cc} 1 - f'(E^*) - z & eV\omega \cos \omega t^* \\ \left[1 - \frac{1}{2} f'(E^*) \right] \frac{\pi}{Hc} z & 1 - z \end{array} \right| = 0$$

или

$$z^2 + \left[-2 + f' - \left(1 - \frac{1}{2} f' \right) \frac{\pi V \omega}{Hc} \cos \omega t^* \right] z + 1 - f' = 0, \quad (36.11)$$

определяющему возможные значения z .

Условия устойчивости состоят в том, чтобы оба корня уравнения (36.11) были по модулю меньше единицы. Для этого коэффициенты характеристического уравнения (36.11) τ и ν при первой и нулевой степенях z должны удовлетворять неравенствам

$$\nu < 1, \quad 1 - \tau + \nu > 0, \quad 1 + \tau + \nu > 0.$$

Соответствующая подстановка приводит к неравенствам:

$$\begin{aligned} 1 - f'(E^*) &< 1, \\ 1 + 2 - f' + \left(1 - \frac{1}{2} f' \right) \frac{\pi V \omega}{Hc} \cos \omega t^* + 1 - f' &> 0, \\ 1 - 2 + f' - \left(1 - \frac{1}{2} f' \right) \frac{\pi V \omega}{Hc} \cos \omega t^* + 1 - f' &> 0. \end{aligned}$$

Первое и второе неравенства выполняются в силу положительности f' и малости f' и $\pi \omega V / Hc$, последнее неравенство приводит к условию

$$\cos \omega t^* < 0,$$

которое определяет, какой из двух возможных стационарных режимов устойчив. Этот устойчивый стационарный режим группирует частицы и позволяет, медленно увеличивая H , наращивать их энергию. При приближении $E_{кр}$ к критическому предельному значению, определяемому уравнением $f(E_{кр}) = eV$, происходит нарушение устойчивости и срыв квазистационарного разгона.

Идея изложенного выше способа разгона релятивистских заряженных частиц впервые была выдвинута В.И. Векслером в 1944 году, а изложенное теоретическое исследование принадлежит А.А. Андронову и Г.С. Горелику (1945 г.) [1]. Основным результатом исследования они представили в виде резонансной кривой со скосом и явлением гистерезиса, характерными для нелинейного резонанса.

37. МАТЕМАТИКА КАК ЯЗЫК, КАК ОПЕРАЦИОННАЯ СИСТЕМА И МОДЕЛИ

В этой главе речь пойдет о том, что такое математика, как она устроена, и почему так огромна и решающа ее роль в точном естествознании и точной науке. Но все это уже после того, как вы познакомились со многими математическими моделями и увидели, как с их помощью познается окружающий мир, подчас на основе очень простых моделей и простых методов их исследования.

Так что такое математика, точнее, что такое математика функционально, в чем ее роль и значение? Ответ на этот вопрос уже был дан во введении: математика – это язык, по своему назначению и роли, т.е. функционально, во многом такой же, как обычные языки: русский, английский, немецкий, французский и другие. Еще Галилей сказал: «Великая книга природы написана математическими символами». Из этого следует, что только тот, кто знает и владеет этим символическим языком, может надеяться читать книгу природы и понимать окружающий его мир: природу, технику и общество. Позднее эта же мысль весьма лаконично была выражена великим физиком В.Гиббсом в его реплике: «Математика тоже язык», и поэтому ее роль столь же велика в изучении окружающего мира, как и обычных языков, в обыденной жизни. С этим нельзя не согласиться, в общем, это так, но все же язык математики чем-то особенный, и вот об этой его особенности и пойдет речь ниже.

Обычный язык – это словарь и звуковые, и графические способы кодирования слов, грамматика, всевозможные фразы и их наборы в виде описаний, сообщений, приказаний, рассказов, инструкций, повестей, романов, ... Нечто похожее есть и в математическом языке. Аналогии слов и грамматики, определяющей правила их сочетания, можно видеть в числах, векторах, матрицах, функциях и действиях с ними: сложении, вычитании, умножении, делении, дифференцировании, интегрировании, ..., а аналог словесных описаний – в математических моделях. Аналогия словесных описаний с математическими моделями достаточно прямая и очевидная, а вот действий и правил, и сочетания слов – не столь ясная. Здесь проявляется специфика математического языка, поскольку его «грамматика» – это не только правила сочетания элементов – слов, но и правила преобразования одних слов в другие.

Первую часть математики – элементы и действия с ними – естественно называть операционной системой, вторую часть – математическими моделями.

Понять назначение и общие принципы построения операционной системы можно, проследив историю ее становления и развития. Подчеркнем итоговую логику этого становления, а не фактическую конкретную историю, которая подчас весьма путана и непоследовательна.

Началась операционная система с целых положительных чисел. Как они возникли, теряется в далеком прошлом. Но римские и арабские целые числа уже хорошо известны. Практические потребности уже тех давних времен привели к изобретению действий с целыми числами: сложению, вычитанию, умножению и делению. Они возникли как замена длительных и трудновыполнимых пересчетов

количеств материалов, товаров, денег, поступающих из разных мест в разное время и в разные места. Эти операции с числами были не всегда выполнимы, и, чтобы это устранить, придумали отрицательные и дробные числа, которым присвоили определенный смысл. Так, если положительное число – это наличность, то отрицательное – это долг. С иррациональными числами произошла длительная заминка: греки не присваивали числа длине гипотенузы равнобедренного прямоугольного треугольника с единичными катетами, она не измерялась числом, что греками своеобразно обосновывалось отсутствием у катета и гипотенузы общей меры, т.е. отрезка, который бы укладывался в каждом из них целое число раз. Это, непреодолимое для греческой мысли препятствие, в дальнейшем длительно игнорировалось, и лишь в прошлом веке было преодолено Г. Кантором, К. Вейерштрассом и Р. Дедекиндом. По существу, это было пополнение рациональных чисел действительными в результате операции предельного перехода, как требование ее замкнутости.

Возникновение комплексных чисел имело ту же причину: устранение неразрешимости квадратных уравнений, в связи с необходимостью извлекать квадратный корень из отрицательного числа.

Гамильтон нашел обобщение комплексных чисел и изобрел кватернионы. Для кватернионов сохранялись все правила действия с действительными и комплексными числами, кроме свойства коммутативности. Затем было доказано, что это единственное обобщение такого рода, хотя подобных кватернионам гиперкомплексных чисел в дальнейшем было придумано много.

Кватернион – это гиперкомплексное число вида

$$a + \alpha i + \beta j + \gamma k, \quad (37.1)$$

где a, α, β, γ – обычные действительные числа; i, j, k – единичные числа новой природы, некие символы или единичные элементы, для которых имеет место следующая таблица умножения:

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1

Операции сложения, вычитания и умножения на действительное число для кватернионов такие же, как для комплексных чисел, а умножение кватернионов определяется приведенной таблицей. Из нее видно, что кватернионы, для которых $\beta = \gamma = 0$, или $\alpha = \beta = 0$, или $\alpha = \gamma = 0$ – это обычные комплексные числа. Умножение кватернионов не коммутативно. Деление кватернионов возможно в силу того, что произведение кватерниона (37.1) и кватерниона $a - \alpha \cdot i - \beta \cdot j - \gamma \cdot k$ – действительное число, так что операция деления числа (37.1) на число $a_1 + \alpha_1 i + \beta_1 j + \gamma_1 k$ после умножения числителя и знаменателя на сопряженный кватернион $a_1 - \alpha_1 \cdot i - \beta_1 \cdot j - \gamma_1 \cdot k$ сводится к делению его на действительное число, что сводится к делению на него чисел a, α, β и γ .

Кватернионы были использованы Дж.К.Максвеллом для записи придуманных им уравнений электромагнитного поля. Привычную для нас векторную форму написания уравнений Максвелла получили лишь после того, как кватернионное исчисление породило понятие вектора и векторного исчисления. Произошло это примерно так. Подобно тому, как в комплексном числе выделяется действительная и мнимая части, у кватерниона $a + \alpha \cdot i + \beta \cdot j + \gamma \cdot k$ выделили действительную часть a и векторную часть $\alpha \cdot i + \beta \cdot j + \gamma \cdot k$ и, далее, у произведения двух векторов

$$(\alpha i + \beta j + \gamma k) (\alpha_1 i + \beta_1 j + \gamma_1 k) = -\alpha\alpha_1 - \beta\beta_1 - \gamma\gamma_1 + \begin{vmatrix} i & j & k \\ \alpha & \beta & \gamma \\ \alpha_1 & \beta_1 & \gamma_1 \end{vmatrix}$$

действительную часть со знаком минус, назвали их скалярным произведением, а векторную – векторным произведением. Это и привело к возникновению векторного исчисления с действиями сложения, вычитания, скалярного и векторного произведений и, конечно, еще умножения и деления на действительное число. Понятие вектора породило линейную алгебру, линейные преобразования векторов, матрицы и матричное исчисление.

Напомним, что кватернионы – не единственные придуманные гиперкомплексные числа, но они единственные в своем роде, где по сравнению с действительными и комплексными числами нарушена только коммутативность произведения, ассоциативность и дистрибутивность сохранены.

Потребность описания процессов изменения привела к изобретению функций и действий с ними, включая дифференцирование, интегрирование и операторные преобразования, т.е. к возникновению математического и функционального анализа сначала функций только действительного переменного, затем многих переменных и комплексного переменного. Операторы и преобразования стали восприниматься, в свою очередь, как элементы, с которыми также возможны разные действия.

На этом процесс расширения математической операционной системы не остановился, он продолжался и продолжается, но сказанного достаточно, чтобы сформулировать основной стимул ее развития и предъявляемые к ней требования. Стимул развития – это стремление описать все, что требуется и удастся, а требования – это замкнутость действий и операций, их принципиальная выполнимость. Замкнутость – это существенная особенность именно математического языка.

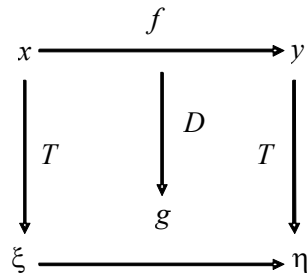
Эти принципы построения операционной системы – возможно более широкое описание на основе натуральных чисел и действий с ними, и требование замкнутости вновь изобретенных операций – в некоторой мере дают ответ на недоуменный вопрос А. Эйнштейна: «Как это может быть, что математика, являясь после всего продуктом мышления людей, не зависимым от опыта, так замечательно приспособлена к объектам действительности?».

Сказанным содержание математической операционной системы не ограничивается: еще есть множества, действия и соотношения между ними. Действия – это объединение, пересечение и дополнение. Соотношения – это принадлежность и непринадлежность, включение и невключение. Отношения больше и меньше

имеются для действительных чисел и модулей чисел, функций и операторов. Конечно, для всех них имеются отношения равенства и тождественности.

Множества и действия с ними можно присоединить к описанной ранее операционной системе, а можно считать еще одной математической операционной системой. Некоторое разделение все же желательно, так как позволяет обнаружить особенности каждой из них. Эти особенности и существенные различия есть даже среди изоморфных операционных систем, не различимых с точки зрения предшествующего их описания.

Две операционные системы A и B изоморфны, если между их элементами и действиями возможно установление взаимно однозначных соответствий T и соответственно D таких, что для любого элемента (элементов) $x \in A$ и допустимого действия $f \in A$ имеет место диаграмма коммутативности, где $y = fx$, $\eta = g\xi$, $Tx = \xi$, $Ty = \eta$, $Df = g$.



Изоморфные системы в силу этой диаграммы коммутативности, казалось бы, отличаются друг от друга только обозначениями. Но на самом деле это не так, потому что в предыдущем изложении не учитывалось, как осуществляются действия. А это, оказывается, очень важно. Действия можно задать неким абстрактным соответствием в виде таблицы умножения, но это реально осуществимо только для конечных чисел элементов, а их бесконечно много. Так что должна быть какая-то, возможно более простая, процедура реализации этого абстрактного соответствия.

Поясним сказанное на примере римской и арабской записей одних и тех же целых чисел. Так, числа 156 и 267 в римской записи – это CLVI и соответственно CCLXVII. Перемножить числа 156 и 267 может школьник младших классов, а перемножить числа в римской записи, CLVI и CCLXVII, очень трудно и практически невозможно, если числа четырехзначные. Греки этому обучались в течение многих лет.

Сказанное относится не только к операции умножения двух чисел, еще более ярко это видно на операциях дифференцирования и интегрирования. Даже для элементарных функций интегрирование не всегда выполнимо. Но эти операции можно сделать очень простыми и всегда легко выполнимыми в операционной системе, придуманной Хевисайдом. Продифференцировать или проинтегрировать функцию $f(t)$ в ней очень просто, и результаты, соответственно, $pF(p) - f(0)$ и $F(p)/p$, где $F(p)$ – запись функции $f(t)$ в системе Хевисайда, p – комплексное число.

Операционная система Хевисайда изоморфна нашей привычной, и этот изоморфизм устанавливается известным преобразованием Лапласа. Правда, в ней

плохо с умножением: оно выполнимо очень сложно, а в привычной для нас системе – просто.

Осталось еще добавить, что операционных систем типа Хевисайда довольно много. Совсем другая изоморфная операционная система, в которой почти все действия, правда, приближенно, легко и быстро выполнимы, – это операционная система современной ЭВМ с хорошим математическим обеспечением, памятью и быстродействием.

Зачем же нам нужны другие изоморфные операционные системы, отличные от привычной общематематической? Ответ можно сформулировать так. В ряде случаев решить некоторую задачу в обычной математической операционной системе не удастся, но если перевести ее в другую, ей изоморфную, то там решение найти можно. Решив задачу в изоморфной системе, можно вернуться в исходную систему и получить требуемое решение. Эта возможность – прямое следствие диаграммы коммутативности. Действительно, допустим, действие f над элементом x невыполнимо в операционной системе A , но оно выполнимо в операционной системе B . Согласно диаграмме коммутативности, действию f в изоморфной системе отвечает действие g , и, как непосредственно видно из этой диаграммы,

$$y = f x = T^{-1} \eta = T^{-1} g \xi = T^{-1} g T x ,$$

где все операции T , g и T^{-1} выполнимы, причем операция T соответствует переходу от исходной системы A к используемой B , а операция T^{-1} – возврату в исходную систему A .

На этом краткое описание математических операционных систем закончим и перейдем к аналогам фраз, описаний, рассказов и т.п. в математическом языке. Этот аналог уже был указан – это математические модели, т.е. изоморфные отображения описаний каких-то реальных или мыслимых объектов с помощью математической операционной системы. Чтобы прояснить сказанное, определим математическую модель абстрактно и формально, как некоторое множество элементов математической операционной системы и связей, устанавливаемых ее действиями и отношениями. В соответствии с этим определением, уравнение $x^2 + px + q = 0$ – модель, соотношение $y = ax + b$ – тоже модель, $x < y$ – тоже модель. Модель и $dx/dt = z + x$, модель и $x < y$. Но понятие модели имеет и содержательный смысл. Чтобы выяснить, в чем он состоит, заметим, что понятие изоморфизма имеет место и для моделей.

Две модели изоморфны, если между их элементами, действиями и отношениями между ними можно установить взаимнооднозначное соответствие, для которого справедлива диаграмма коммутативности. Речь может идти о моделях в одной и той же операционной системе или разных.

Теперь осталось сказать о математической модели и математическом моделировании самое главное, но, к сожалению не формализуемое: природа, как окружающий нас мир, – тоже операционная система, так как ее удается с той или иной полнотой представить в виде разнообразных элементов и связей между ними и текущим временем, называемых законами природы и принципами ее устройства. С этой точки зрения объекты природы – это ее природные модели, и они могут быть изоморфны математическим моделям. Вернее, человечество создало матема-

тический язык так, чтобы это имело место. И как только удастся построить изоморфную объекту природы математическую модель, изучая ее, мы постигаем и природный объект. Неформальным в описанном является выделение в объектах природы элементов, подчас выдумываемых и изобретаемых, и установление связей между ними и текущим временем, т.е. неформально то, как мы постигаем устройство мира и операционную систему природы и как мы постигаем изоморфную ей математическую операционную систему, модели которой могут быть изучены, исследованы и поняты. Поэтому наука в целом и математическое моделирование как часть ее – это не только точная наука, но и неформализуемое искусство. В истории науки и отдельных исследованиях формализуемое и неформализуемое, точная наука и искусство тесно переплетаются между собой и трудно отделимы друг от друга. Но это разделение, например, было очень явным и четким, когда М. Фарадей постиг электромагнитные явления, выдумав магнитные и электрические поля, а Дж.К. Максвелл построил соответствующую математическую модель – знаменитые уравнения Максвелла. В механике были придуманы силы и материальные точки, обладающие массой и находящиеся в каких-то местах, и имеющие скорость движения, Ньютон построил математическую модель их движения – уравнения Ньютона, затем Лагранж – уравнения Лагранжа для систем материальных точек с голономными идеальными связями между ними, но без необходимости знания вызываемых ими сил.

По-видимому, в основе не только нашего осознанного мышления, но и неосознанного постижения мира нашими органами чувств и мозгом лежит построение изоморфных моделей. Как устроены эти модели, как они строятся и познаются – одна из важнейших и не раскрытых тайн природы. С помощью этого интуитивного мышления человек осознанно придумывает еще более мощные, сознательно обоснованные математические модели и необходимую для их построения и исследования операционную систему.

38. ГЕОМЕТРИЧЕСКОЕ, ФИЗИЧЕСКОЕ, АНАЛОГОВОЕ, МАТЕМАТИЧЕСКОЕ И СТАТИСТИЧЕСКОЕ ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Все перечисленные в заголовке виды моделирования основаны на изоморфизмах. Геометрическое и физическое моделирование – на геометрическом и соответственно физическом подобии, вызываемом изменением масштаба длины и соответственно масштабов базовых физических величин, например длины, времени и массы. Аналоговое моделирование возможно в силу одинаковости видов некоторых закономерностей для величин различной физической природы, например, механической и электрической, процессов теплопроводности, диффузии и образования электрического поля. Об изоморфизме реальных объектов и математических моделей уже говорилось. Имитационное моделирование исходит из того, что неизвестная, подлежащая отысканию величина оказывается совпадающей с математическим ожиданием специальным образом организованного случайного процесса.

Геометрическое моделирование используется с глубокой древности: законы геометрического подобия были известны уже грекам. Физическое моделирование возникло более века тому назад из-за крупных катастроф с морскими судами и затем стимулировало его еще и развитие авиации. Аналоговое моделирование, точнее, лежащая в его основе одинаковость законов природы для объектов разной физической природы, были ясны уже Дж. У. Рэлею (J. W. Rayleigh), в частности, в виде аналогии между томсоновским электрическим колебательным контуром и механическим гармоническим осциллятором. Именно эти аналогии и менее полные привели к возникновению теории колебаний как науки об общих закономерностях эволюционных процессов в механике, физике, химии, биологии и науке об обществе. Математическое моделирование в наше время стало ведущим и широко используемым в немалой степени благодаря не только теоретическим успехам, но и появлению универсальных быстродействующих вычислительных машин.

В некоторых типах задач успешно используется имитационное моделирование, называемое еще методом Монте-Карло или статистическим моделированием.

Всем этим видам моделирования посвящено значительное количество научной литературы. Здесь об этих видах идет речь лишь для того, чтобы показать их общую основу – изоморфизм разных реальных объектов и процессов между собой и представленными мысленно или записями на бумаге, или на том или ином носителе информации. После сказанных общих слов далее поговорим несколько подробнее о физическом и имитационном моделировании, а затем, отдельно, о математическом моделировании как науке и искусстве.

Физическое моделирование [45]. Физическое моделирование не предполагает фактического наличия математической модели, достаточно опереться на более скромные сведения и понимание изучаемого объекта: надо знать, какие физические величины полностью определяют поведение рассматриваемого объекта, все и ни одного лишнего. Однако основывается оно на том, что имеется матема-

тическая модель и входящие в нее физические величины имеют размерность, а математическая модель такова, что изменения всех базовых масштабов физических величин ее не меняют – масштабы сокращаются. Это эквивалентно наблюдаемой нами естественной независимости законов природы от принимаемых масштабов измерения физических величин, что в принципе могло бы быть и не так: в неевклидовых геометриях не имеют места привычные законы геометрического подобия.

Пусть определяющие физические величины f_1, \dots, f_m , а физическая размерность величины f_i есть $L^{\alpha_i} S^{\beta_i} M^{\gamma_i}$, где L, S и M – масштабы длины, времени и массы. Эти данные запишем в виде таблицы:

	L	S	M
f_1	α_1	β_1	γ_1
f_2	α_2	β_2	γ_2
...
f_m	α_m	β_m	γ_m

Изменения масштабов L, S и M соответственно в K_L, K_S и K_M раз преобразует каждую определяющую величину f_i в величину

$$\bar{f}_i = K_L^{\alpha_i} K_S^{\beta_i} K_M^{\gamma_i} f_i, \quad i = \overline{1, m}. \tag{38.1}$$

Новые физические определяющие величины $\bar{f}_1, \dots, \bar{f}_m$ фактически отвечают тому же самому объекту и определяются той же самой математической моделью. Отсюда следует, что объект, характеризуемый величинами f_1, \dots, f_m , подобен объекту, характеризуемому величинами $\bar{f}_1, \dots, \bar{f}_m$, тогда и только тогда, когда для некоторых K_L, K_S и K_M выполняется соотношение (38.1). Эти же формулы (38.1) позволяют пересчитывать определяющие величины одного объекта в другой с такими же определяющими величинами и одинаковыми математическими моделями, которые не обязательно знать. Кроме величин f_1, \dots, f_m , объект может описываться и другими величинами, но они должны определяться ими, и поэтому опять-таки пересчитываться в соответствии с (38.1).

Найдем теперь условия физического подобия, т.е. выясним, когда существуют такие K_L, K_S и K_M , при которых имеют место формулы (38.1).

Формулы (38.1), логарифмируя, можно записать в виде:

$$\alpha_i \ln K_L + \beta_i \ln K_S + \gamma_i \ln K_M = \ln \frac{\bar{f}_i}{f_i}, \quad i = \overline{1, m}. \tag{38.2}$$

Система линейных уравнений (38.2) относительно $\ln K_L, \ln K_S$ и $\ln K_M$ имеет решение, если ранг матрицы

$$\begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \cdot & \cdot & \cdot \\ \alpha_m & \beta_m & \gamma_m \end{vmatrix} \tag{38.3}$$

равен рангу расширенной матрицы

$$\begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 & \ln \frac{\bar{f}_1}{f_1} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_m & \beta_m & \gamma_m & \ln \frac{\bar{f}_m}{f_m} \end{vmatrix}. \quad (38.4)$$

Последнее будет иметь место, если из линейной зависимости строк матрицы (38.3)

$$\begin{aligned} x_1 \alpha_1 + \dots + x_m \alpha_m &= 0, \\ x_1 \beta_1 + \dots + x_m \beta_m &= 0, \\ x_1 \gamma_1 + \dots + x_m \gamma_m &= 0, \end{aligned} \quad (38.5)$$

где $x_1^2 + x_2^2 + \dots + x_m^2 \neq 0$, следует выполнение соотношений

$$x_1 \ln \frac{\bar{f}_1}{f_1} + \dots + x_m \ln \frac{\bar{f}_m}{f_m} = 0$$

или

$$f_1^{x_1} \dots f_m^{x_m} = \bar{f}_1^{x_1} \dots \bar{f}_m^{x_m}, \quad (38.6)$$

где величины этого равенства справа и слева в силу (38.5) имеют нулевую размерность. Выполнение этих соотношений для любого решения x_1, \dots, x_m системы (38.5) является, таким образом, необходимым и достаточным критерием физического подобия. Ясно, что достаточно выполнения этих условий для независимых решений x_1, \dots, x_m системы (38.5). Каждому независимому решению x_1, \dots, x_m отвечает свое условие подобия, и их столько, сколько независимых решений у системы (38.5). Таковы условия подобия. При их выполнении формулы пересчета находятся как решения относительно K_L, K_S и K_M линейных уравнений (38.2). Эти формулы пересчета неоднозначны с точностью до условий подобия, одинаковых у объекта и его модели.

Перейдем к примерам.

Пример 1. Пусть предполагаемая к строительству стержневая статическая определяемая конструкция сделана из стержней длин l_1, l_2, \dots, l_n , сочлененных в узловых точках, где она нагружена некоторыми силами F_1, \dots, F_m , и нас интересуют усилия G_1, \dots, G_n во всех стержнях с целью выяснения, достаточно ли они прочны. Решить задачу мы хотим с помощью построения уменьшенной модели и экспериментального определения сил в стержнях. Прежде всего, очевидно, что конструкция, предполагаемая к строительству, и модель должны быть геометрически подобны, так как в противном случае соотношения $l_i/l_j, i \neq j$, для них были бы различными. По той же причине должна соблюдаться пропорциональность для прилагаемых к конструкции усилий. При выполнении этих условий обосновано, что определяющими являются характерный размер и приложенное усилие, например, l_1 и F_1 . Для них таблица имеет вид:

		L	S	M
x_1	l_1	1	0	0
x_2	F_1	1	-2	1

Для указанной модели уравнения (38.5) запишутся в виде:

$$x_1 + x_2 = 0, \quad -2x_2 = 0, \quad x_2 = 0,$$

они не имеют ненулевого решения. Поэтому физическое подобие имеет место между любыми описанными конструкциями с пропорциональными внешними силами, а усилия G_1, \dots, G_n определяются через аналогичные усилия $\overline{G}_1, \dots, \overline{G}_n$ в модели по формуле

$$G_s = \frac{F_1}{\overline{F}_1} \overline{G}_s, \quad s = \overline{1, n}. \tag{38.7}$$

Пример 2. Пусть теперь в такой же конструкции, не обязательно статически определяемой, стержни упругие с коэффициентами упругости k_1, \dots, k_n . Из таких же соображений следует необходимость геометрического подобия с моделью и одинаковые соотношения между внешними усилиями и коэффициентами упругости. Определяющими теперь будут характерные величины: размер l , внешнее воздействие F и коэффициент упругой жесткости k , например, l_1, F_1, k_1 . Как и ранее, составим таблицу

		L	S	M
x_1	l	1	0	0
x_2	F	1	-2	1
x_3	k	0	-2	1

и уравнения для отыскания критериев подобия:

$$x_1 + x_2 = 0, \quad -2x_2 - 2x_3 = 0, \quad x_2 + x_3 = 0.$$

Теперь эти уравнения имеют ненулевое решение $x_1 = l, x_2 = -1, x_3 = l$, и ему соответствует критерий подобия $lF^{-1}k$. Для подобия эти величины у проектируемой конструкции и модели должны быть одинаковыми, а пересчет усилий в стержнях по-прежнему определяется формулой (38.7).

Пример 3. В качестве следующего примера возьмем маятник. Маятник как механическая система определяется длиной l , массой m и ускорением силы тяжести g . Этому соответствуют таблица

		L	S	M
x_1	L	1	0	0
x_2	m	0	0	1
x_3	g	1	-2	0

и уравнения:

$$x_1 + x_3 = 0, \quad -2x_3 = 0, \quad x_2 = 0,$$

не имеющие ненулевого решения, поэтому все маятники между собой подобны. Этот вывод непосредственно следует из математической модели маятника

$$J\ddot{\varphi} + lmg \cdot \sin \varphi = 0,$$

так как путем замены масштаба времени $t \rightarrow \sqrt{lmg/J} \tau$ это уравнение приводится к виду

$$\frac{d^2\varphi}{d\tau^2} + \sin \varphi = 0.$$

Это означает, что все маятники между собой физически подобны.

Допустим, мы хотим, пользуясь этим физическим подобием, найти период колебаний T . Непосредственно это сделать нельзя, так как период T зависит не только от величин l , m и g , но еще от амплитуды a колебания маятника, и в приведенную выше таблицу ее нужно ввести. После этого внесения появятся еще одна переменная, x_4 , и соответствующие уравнения запишутся в виде:

$$x_1 + x_3 + x_4 = 0, \quad -2x_3 = 0, \quad x_2 = 0.$$

Теперь уже ненулевое решение есть: $x_1 = -1, x_2 = 0, x_3 = 0, x_4 = 1$ и ему отвечает критерий физического подобия a/l . При его выполнении соотношение между периодами колебаний T и \bar{T} определяется отношением одномерных с ними величин $\sqrt{lmg/J}$ и $\sqrt{\bar{l}\bar{m}\bar{g}/\bar{J}}$, т.е.

$$T = \bar{T} \cdot (lmg/J)^{1/2} \cdot (\bar{l}\bar{m}\bar{g}/\bar{J})^{-1/2}$$

или

$$T = (lmg/J)^{1/2} \cdot f(a/l),$$

так как $\bar{T} \cdot (\bar{l}\bar{m}\bar{g}/\bar{J})^{-1/2}$ не имеет размерности и зависит только от отношения $\bar{a}/\bar{l} = a/l$. Как известно, при $a/l \ll 1, f(a/l) = 2\pi$.

Пример 4. В качестве последнего примера определим критерий подобия для стационарного течения вязкой жидкости в поле силы тяжести, когда ее течение определяется характерным размером l , скоростью v , плотностью ρ , вязкостью ν и ускорением силы тяжести g . При этих предположениях соответствующие таблица и уравнения, определяющие критерий физического подобия, имеют вид:

		L	S	M
x_1	l	1	0	0
x_2	v	1	-1	0
x_3	ρ	-3	0	1
x_4	ν	2	-1	0
x_5	g	1	-2	0

$$x_1 + x_2 - 3x_3 + 2x_4 + x_5 = 0, \quad -x_2 - x_4 - 2x_5 = 0, \quad x_3 = 0.$$

Уравнения для x_1, \dots, x_5 имеют два независимых решения: $x_1 = 1, x_2 = 1, x_4 = -1$ и $x_1 = -1, x_2 = 2, x_5 = -1$, которым отвечают знаменитые числа Рейнольдса и Фроуда

$$\text{Re} = \frac{lv}{\nu} \quad \text{и} \quad \text{Fr} = \frac{v^2}{lg}.$$

Примером сравнительно недавнего сенсационного применения моделирования на основе размерностей физических величин и общих упрощенных представлений о моделируемом явлении является определение засекреченной энергии первого ядерного взрыва на основе открытой демонстрации его кино съемки.

Взрыв принимался точечным, а возникающая ударная волна – сферической. В этих упрощенных представлениях радиус взрывной волны R является функцией времени, зависящей от энергии взрыва E и начальной плотности окружающего воздуха ρ . Соответствующая таблица размерностей энергии E , плотности ρ , времени t и радиуса R взрывной волны имеет вид:

	E	ρ	t	R
L	2	-3	0	1
M	1	1	0	0
T	-2	0	1	0

Поиск безразмерной комбинации определяющих величин E , ρ и t приводит к системе уравнений

$$2x - 3y = 0, \quad x + y = 0, \quad -2x + z = 0,$$

имеющей только нулевое решение $x = y = z = 0$ и поэтому все точечные взрывы со сферической ударной волной подобны между собой.

В соответствии с этим, согласно приведенной выше таблице, находим, что радиус R выражается через определяющие физические величины E , ρ и t формулой вида

$$R = CE^{1/5} \rho^{-1/5} t^{2/5},$$

где C – некоторая константа, которая может быть найдена по единичному конкретному эксперименту или теоретическому расчету, обнаруживших, что $C = 1$. Теперь найденная формула позволяет найти энергию ядерного взрыва E , если известен радиус ударной волны при каком-нибудь t . Но можно найти E и более надежно, используя серию значений времени t и соответствующих значений R . Логарифмируя найденную формулу зависимости R от E , ρ и t и учитывая, что $C = 1$, находим, что

$$\frac{5}{2} \ln R = \frac{1}{2} \ln \frac{E}{\rho} + \ln t.$$

Согласно этому, график зависимости $5/2 \cdot \ln R$ от $\ln t$ изображается прямой, отсекающей на оси ординат отрезок длины $1/2 \cdot \ln E/\rho$. На эту прямую должны лечь данные кинофильма ядерного взрыва и, исходя из этого, она может быть найдена.

Имитационное моделирование. Краткий рассказ об имитационном моделировании начнем с исторического примера отыскания числа π путем случайных бросаний иголки на разграфленный на одинаковые полосы лист бумаги. Этой забавой в прошлом увлекались и известные ученые, кидая иголку тысячи раз: чем больше раз кинешь, тем точнее найдешь число π .

Под случайным бросанием понималось, что иголка, попадая на лист бумаги, после остановки располагается относительно параллельных линий, нанесенных на бумагу так, что ближайшее расстояние x ее центра и угол φ с ними – независимые случайные величины с постоянными плотностями вероятностей $1/a$, где $2a$ – ширина полос, и, соответственно, $1/\pi$. Иголка длины $2a$ будет пересекаться с одной из линий, если

$$a \cos \varphi \geq x. \quad (38.8)$$

В соответствии с принятыми обозначениями, вероятность пересечения иголки с одной из линий равна (рис. 38.1)

$$\iint_D \frac{1}{a\pi} dx d\varphi, \quad (38.9)$$

где D – область, определяемая неравенствами $0 \leq x \leq a$, $-\pi/2 \leq \varphi \leq \pi/2$ и (38.8). Интеграл (38.9) легко находится и равен $2/\pi$.

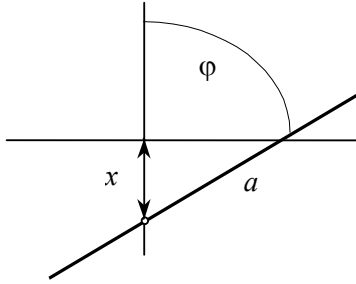


Рис. 38.1. Пересечение иглы с одной из линий

Пусть теперь иголка случайно брошена N раз и пересеклась с одной из линий n раз. Согласно закону больших чисел, частота пересечения n/N стремится к найденной вероятности $2/\pi$, так что приближенно и тем точнее, чем больше число бросаний N ,

$$\pi \approx \frac{2N}{n}.$$

Известное неравенство Чебышева позволяет оценить вероятность того, что ошибка будет не больше некоторого малого ε . Согласно этому неравенству,

$$P\left(\left|\frac{n}{N} - M \frac{n}{M}\right| > \varepsilon\right) < \frac{Dn/N}{\varepsilon^2}, \quad (38.10)$$

где Mn/N и Dn/N – соответственно математическое ожидание и дисперсия частоты пересечений n/N . Частота

$$\frac{n}{N} = \frac{1}{N} (\xi_1 + \dots + \xi_N), \quad (38.11)$$

где $\xi_i = 1$, если в i -м бросании иголка пересеклась с одной из линий, и $\xi_i = 0$ – в противном случае. Случайные величины ξ_i независимы, и для каждой из них вероятность быть единицей равна $p = 2/\pi$. Далее, согласно (38.10) и (38.11), непосредственно находим, что

$$P\left(\left|\frac{n}{N} - p\right| > \varepsilon\right) < \frac{p(1-p)}{N\varepsilon^2}, \quad (38.12)$$

т.е. вероятность отклонения частоты n/N от величины p больше, чем на ε , не более чем $p(1-p)/(N\varepsilon^2)$, и стремится к нулю при $N \rightarrow \infty$.

Таким образом, π приближенно и со сколь угодно большой точностью можно найти путем случайных бросаний иголки. Но этих бросаний нужно довольно много. Так, чтобы найти три точных знака π , т.е. при $\varepsilon < 10^{-3}$, с вероятностью 0,99

нужно бросить иголку примерно 10^8 раз. Конечно, сегодня таким способом вычислять число π по меньшей мере неразумно. Но есть примеры, где это не только возможно, но, пожалуй, и единственно приемлемо. Этот пример – вычисление многомерных определенных интегралов от функций многих переменных, например десяти, вида

$$\int_D f(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (38.13)$$

по ограниченной области D , определяемой некоторыми неравенствами, позволяющими выяснить, лежит ли точка (x_1, \dots, x_n) в этой области или нет.

Пусть $a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n$ и $A \leq f(x_1, \dots, x_n) \leq B$. Первые n неравенств и неравенство $A \leq z \leq B$ определяют $(n + 1)$ -мерный параллелепипед Π в $(n + 1)$ -мерном пространстве R_{n+1} переменных x_1, \dots, x_n и z . Величина интеграла есть не что иное, как $(n + 1)$ -мерный объем области

$$0 \leq z \leq f(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in D \quad (38.14)$$

в пространстве R_{n+1} . Назовем ее \bar{D} . Область \bar{D} лежит внутри параллелепипеда Π . Обозначим объемы областей \bar{D} и Π через \bar{V} и V . \bar{V} – это интересующее нас значение интеграла (38.13), а V равно $(B - A)\Pi(b_i - a_i)$. Отношение \bar{V}/V не что иное, как вероятность попадания случайной точки $M(x_1, \dots, x_n, z)$ в область \bar{D} , если она имеет постоянную плотность вероятности в параллелепипеде Π , равную $\{(B - A)\Pi(b_i - a_i)\}^{-1}$. Это означает, что величину \bar{V}/V можно найти путем независимых случайных «бросаний» точки M точно так же, как это имеет место с бросанием иголки для определения числа π . Конечно, «бросаний» нужно достаточно много. Но что замечательно: необходимое число «бросаний» не зависит от размерности n интеграла. Осуществить же случайное бросание можно с помощью датчиков случайных чисел, и при каждом «бросании» непосредственно определяется, в соответствии с заданием области D неравенствами (38.14), когда «брошенная» точка попала в область \bar{D} .

39. ОБЩАЯ СХЕМА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Математическое моделирование включает решение двух трудоемких задач: составление математической модели и ее исследование. Эти задачи очень разные, требующие различных знаний, навыков и интуиции, но, несмотря на это, они тесно взаимосвязаны, поскольку при составлении модели следует учитывать трудности и возможности ее исследования, а при исследовании может обнаружиться необходимость корректировки построенной модели. Кроме того, при исследовании необходимо понимать, что учтено и что не учтено в модели, т.е. на какие вопросы она может дать ответы, а на какие фактически не может.

Математические модели, которые ранее рассматривались в этой книге, можно назвать простыми. Простыми по написанию, простыми для исследования, хотя и не всегда простыми с точки зрения получаемых результатов. Результаты исследования моделируемого объекта могли быть неожиданными и требующими объяснения. Модели просты за счет того, что они учитывают только некоторые стороны объекта, не стремясь к всесторонней и полной адекватности с реальным объектом, и поэтому, естественно, могут объяснить только то, что в ней учтено и заложено, и зачастую только качественно, а количественно подчас весьма приближенно. Несмотря на это, а может быть, благодаря этому, роль простых моделей в науке очень велика, поскольку они ведут к пониманию существа происходящих процессов – пониманию, очень важному для ученого, инженера, конструктора и изобретателя. Но это необходимое понимание не всегда достаточно, и во многих случаях нужны более точные количественные данные. Это требует дополнительного уточнения и усложнения модели, что, как правило, влечет значительно более сложные численные исследования. Эта вычислительная сторона математического моделирования, несмотря на ее важность, в этой книге фактически не затронута, в ней идет речь о том, как этих вычислений избежать, если это возможно и допустимо.

Построенные модели и их исследование могут требовать не только достаточных знаний, но и определенного искусства. Это искусство в некотором смысле лежит вне науки, поскольку является результатом индивидуального опыта и обучения показами на примерах. Искусству тоже можно учиться. Ведь учат сочинять стихи, музыку, рисовать, играть на скрипке, танцевать. Следует признать, что современное традиционное изложение математики, теоретической и отчасти прикладной, не способствует выработке этого искусства, и в результате лишь отдельные, одаренные от природы люди, могут вовремя преодолеть барьер математического формализма. Более успешно он преодолевается в научных школах, благодаря примерам, общению, беседам, обсуждениям и показам. Как, возможно, вы уже догадались и вспомнили, что именно эту цель облегчения взятия барьера, разделяющего теоретическую математику и прикладное математическое моделирование, ставит перед собой эта книга.

Итак, пусть перед вами стоит задача построения и исследования модели, т.е. задача математического моделирования. С чего следует начать? Прежде всего не

жалейте времени на то, чтобы осмыслить, уточнить и конкретизировать поставленную перед вами задачу, посмотреть на нее с разных точек зрения. Нужно возможно точнее и полнее ответить на вопрос, что вы хотите узнать, на какие вопросы ответить, и какой объект следует для этого рассматривать, что в нем обязательно надо учесть, а чем можно или, может быть, нужно пренебречь? Этот важнейший этап математического моделирования можно назвать *исходной постановкой задачи и определением объекта исследования*. Не думайте, что в дальнейшем вам не придется к нему вернуться. Очень может быть, что придется, и чуть ли ни тогда, когда вы будете считать задачу решенной.

Затем вы можете приступить ко второму этапу, состоящему в словесном или мысленном описании объекта, повелевающих и управляющих им законов природы. После этого снова подумайте, нельзя ли чем-то пренебречь или, наоборот, что-то еще нужно все-таки учесть. В итоге вы *завершите мысленное или словесное описание объекта исследования и выяснение законов природы, которые определяют его поведение и свойства*. Далее такое описание следует изоморфно отобразить в математическую модель, стараясь это сделать возможно проще и удобнее для последующего исследования. При этом не следует упускать возможностей упрощений или декомпозиций, ведущих к расчленению задачи исследования на последовательные этапы. В результате вы приходите *к первоначальной математической модели*.

Теперь следует подумать о том, как ее изучить и исследовать. Возможно, что это обдумывание приведет вас к некоторым изменениям модели, способствующим простоте, удобству и экономичности ее исследования. Далее нужно провести исследование и сделать выводы о свойствах объекта, его поведении и возможных в нем явлениях и сопоставить с тем, что вы знаете об объекте. Если что-то не совпадает, вызывает сомнение или не вполне устраивает, – вернитесь к одному из предшествующих этапов или, может быть, даже к исходной постановке задачи. Если вас все устраивает, можно переходить к формулировке выводов и рекомендаций. Конечно, то, что было сказано, – это общая схема, и вовсе не обязательно буквально ей следовать.

В заключение этого краткого и общего описания сложного и многообразного математического моделирования, которое иллюстрировалось много раз, перечислим основные его этапы, а затем приведем некоторые, также общие, полезные при моделировании соображения:

1. Постановка задачи и определение объекта исследования.
2. Идеализация и упрощение объекта.
3. Мысленно-словесное описание объекта и определяющих его поведение и свойства законов природы.
4. Дальнейшие упрощения и, наоборот, учет чего-то нового.
5. Составление первоначальной математической модели.
6. Внесение корректив в математическую модель.
7. Выбор и реализация методов исследования математической модели.
8. Сопоставление результатов исследования модели с тем, что известно об объекте.
9. Выводы и рекомендации.

Отметим, что от каждого из этапов возможны возвраты к любому предыдущему.

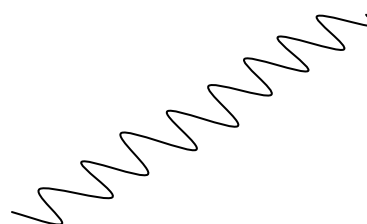
А теперь некоторые полезные общие соображения:

1. Чем проще модель, тем меньше возможность ошибочных выводов.
2. Модель должна быть простой, но не проще, чем это возможно.
3. Пренебрегать можно чем угодно, нужно только знать, как это повлияет на решение.
4. Модель должна быть грубой, малые возможные и допустимые поправки не должны кардинально и существенно менять ее поведение.
5. Модель и расчет не следует делать точнее исходных данных. Выводы не должны существенно меняться в пределах ошибок исходных данных и ошибок счета. Если этого не удастся достичь за счет модификаций модели, необходимы более точные исходные данные.
6. При анализе результатов исследования модели важны не только конкретные численные результаты, но и понимание, почему и как все происходит, как все это зависит от параметров. Во всяком случае, к этому следует стремиться, хотя в некоторых случаях это удастся далеко не сразу.

40. МОДЕЛИ ВИБРОПОГРУЖЕНИЯ

Как происходит вибропогружение 18-метрового тонкого металлического шпунта в мерзлый грунт, я видел сам, а создатель этого чуда, инженер Д.Д. Баркан, чиркнув карандашом поперек шпунта, показал мне, как это происходит: шпунт вибрировал и погружался (рис. 40.1). Причину погружения шпунта Баркан видел в наблюдаемом им экспериментально явлении разжижения грунта под влиянием вибраций. Проблема, которая его «заела», состояла в том, что вибраторы очень быстро выходили из строя, и их, поломанных и негодных, лежала невдалеке громадная куча. Это было на строительстве плотины Горьковской ГЭС зимой 1961 года. Прислали меня от ГИФТИ (Горьковский исследовательский физико-технический институт), чтобы помочь в беде, что напрямую я сделать, естественно, не мог, но по возвращении домой вскоре сообразил, в чем причина чудесного погружения шпунта в мерзлый грунт, как в воду. Она была в усреднении сил сопротивления погружению шпунта под влиянием его высокочастотных – 50 герц – вибраций. На этой основе были построены модели, о которых рассказывается ниже, и найдены зависимости скорости погружения от ее параметров. Эти результаты были использованы в создании последующих, еще более эффективных конструкций систем вибропогружения, где, в частности, вибраторы уже так быстро не ломались.

Рис. 40.1. Осциллограмма вибропогружения шпунта, полученная движением карандаша слева направо поперек него



Я позволю себе кратко воспроизвести, как осмысливалось это необычное явление и как сформировалась приемлемая математическая модель, результаты исследования которой можно было положить в основу инженерного конструирования.

Сначала я прочитал работы Д.Д. Баркана о разжижении грунта под влиянием вибраций и использовании этого явления для вибропогружения шпунта, свай и другого. Изложенное подкреплялось описанием экспериментов по погружению тяжелого шарика в вибрирующий грунт, но они не очень согласовывались с теоретическими соображениями и данными экспериментов по реальным вибропогружениям шпунтов и свай. Кроме того, мне казалось очевидным, что это разжижение не может происходить в промерзшем грунте, а шпунт его преодолевал весьма успешно.

Первая идея была в том, что шпунт погружается в грунт, как змея, в силу бегущих вдоль него продольных волн. Но эту идею пришлось отвергнуть, так как вследствие очень большой скорости распространения бегущих волн вдоль шпунта и небольшой его длины он практически двигался, как целое. Тогда я принял са-

мую естественную модель, что с боков шпунт испытывает сухое кулоновское трение, а своим торцом трамбуется и раздвигает грунт. Эта модель представлена схематически на рис. 40.2, а соответствующая математическая модель имеет вид

$$m \ddot{x} = P + F \cos \omega t - F_{\text{бок}} - F_{\text{лоб}}, \quad (40.1)$$

где $F \cos \omega t$ – гармоническая сила вибратора, действующая на шпунт; P – вес шпунта вместе с вибратором; $F_{\text{бок}}$ и $F_{\text{лоб}}$ – силы бокового и лобового сопротивлений; x – перемещение шпунта, отсчитываемое вниз.

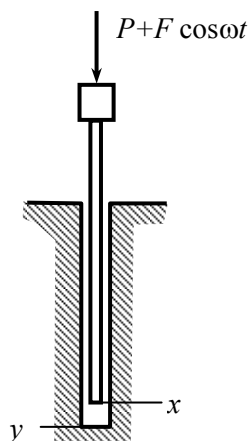


Рис. 40.2. Упрощенная модель шпунта с вибратором, погружаемого в грунт, с учетом лишь бокового и лобового сопротивлений, описанных формулами (40.2) и (40.3)

Грунт с боков от шпунта принимался неподвижным, и тогда

$$F_{\text{бок}} = Q \text{sign}(\dot{x}). \quad (40.2)$$

Уровень грунта под торцом шпунта имеет координату y , и ее увеличение требует некоторого усилия R . Если x – координата нижнего торца шпунта, y – уровень продавливаемого им грунта, то

$$F_{\text{лоб}} = \begin{cases} R & \text{при } x = y \text{ и } \dot{x} > 0; \\ 0 & \text{при } x < y \text{ либо } x = y \text{ и } \dot{x} \leq 0, \end{cases} \quad (40.3)$$

причем

$$\dot{y} = \begin{cases} \dot{x} & \text{при } x = y \text{ и } \dot{x} > 0; \\ 0 & \text{при } x < y \text{ либо } x = y \text{ и } \dot{x} \leq 0. \end{cases} \quad (40.4)$$

Сегодня решение уравнений (40.1)–(40.4) на ЭВМ не представляло бы затруднений, но в то время это была сложная, трудоемкая вычислительная задача. Но оказалось возможным приближенно аналитически найти ее решение, предполагая, согласно запомнившейся мне осциллограмме движения шпунта, показанной Барканом, что

$$x = a \cos \omega t + vt, \quad (40.5)$$

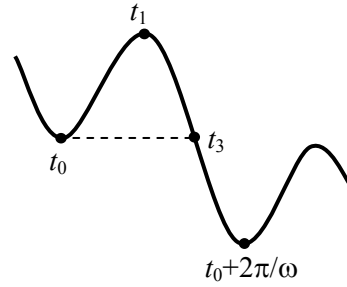
где a – амплитуда его колебаний, ω – частота, а v – скорость погружения. Принимая это предположение и интегрируя уравнение (40.1) по периоду (с учетом, что имеет место (40.5)), непосредственно находим, что

$$\int_0^{2\pi/\omega} m\ddot{x} \cdot dt = P \frac{2\pi}{\omega} - \int_0^{2\pi/\omega} F_{бок} dt - \int_0^{2\pi/\omega} F_{лоб} =$$

$$= \frac{2\pi}{\omega} P + \left(t_1 - t_0 + t_1 - t_0 - \frac{2\pi}{\omega} \right) Q + \left(t_0 + \frac{2\pi}{\omega} - t_3 \right) R = 0. \tag{40.6}$$

Смысл моментов времени t_0, t_1 и t_3 указан на осциллограмме движения шпунта (40.5), представленной на рис. 40.3.

Рис. 40.3. Экспериментально полученный график колебаний сваи с указанием определенных моментов времени t_0, t_1, t_3 и $t_0+2\pi/\omega$



Они находятся из очевидных уравнений, и при этом приближенно соотношение (40.6) принимает вид

$$\frac{2\pi}{\omega} P - \frac{2v}{a\omega^2} Q - \frac{2v}{a\omega^2} R = 0,$$

откуда

$$v = \frac{\pi a \omega}{Q + R} P. \tag{40.7}$$

Этот «удивительный» результат можно трактовать как то, что под влиянием вибраций силы $F_{бок}$ и $F_{лоб}$ усредняются и оказываются эквивалентными силе вязкого трения с коэффициентом $(Q + R)/(\pi a \omega)$. Этот коэффициент вязкого трения убывает с ростом амплитуды колебаний a и частоты ω . Очень грубо, принимая, что $F \gg Q + R$, амплитуда $a = F/(M\omega^2)$ и тогда коэффициент вязкого трения

$$h = \frac{Q + R}{\pi a \omega} = \frac{(Q + R) M \omega}{\pi F}. \tag{40.8}$$

Таким образом, действительно, при вибропогружении силы сопротивления грунта подобны вязкому трению, но это не оттого, что грунт якобы разжижается, разжижение не требуется.

Рассматриваемая модель весьма проста, она хорошо объясняет суть явления вибропогружения, но она не дает хороших совпадений с экспериментом. Дело в том, что грунт обладает еще упругостью, как с боков от шпунта, так и в его торце. Эта упругость проявляется в виде некоторых, ярко или неярко выраженных, резонансных явлениях: шпунт в грунте обладает некоторой собственной частотой, он подобен осциллятору. Более полная модель, учитывающая упругость грунта, изображена на рис. 40.4.

Теперь уже модель описывается тремя переменными x, y и z : x – положение шпунта, y – смещение бокового грунта, z – смещение доннышка ямки в грунте. Как и прежде, проскальзывание шпунта и продавливание грунта происходит при си-

лах Q и, соответственно R , но при этом боковинка и доньшко грунта имеют упругость k_6 и k_n соответственно. Математическая модель теперь состоит, как и ранее, из уравнения (40.1) для x , уравнений для y и z и формул, определяющих силы $F_{бок}$ и $F_{лоб}$,

$$\dot{y} = \begin{cases} \dot{x} & \text{при } |k_6 y| < Q, \\ 0 & \text{при } |k_6 y| = Q; \end{cases}$$

$$\dot{z} = \begin{cases} \dot{x} & \text{при } k_n(x-z) = R \text{ и } \dot{x} \geq 0, \\ 0 & \text{при } k_n(x-z) < R; \end{cases} \quad (40.9)$$

$$F_{бок} = k_6 y, \quad F_{лоб} = \begin{cases} k_n(x-z) & \text{при } x > z, \\ 0 & \text{при } x \leq z. \end{cases}$$

Едва ли сегодня имеет смысл исследовать эту модель (40.9) приближенно тем же, что и ранее, очень простым методом. Формулы получаются громоздкие и не всегда приемлемые. Просчет же на ЭВМ легко и быстро осуществим.

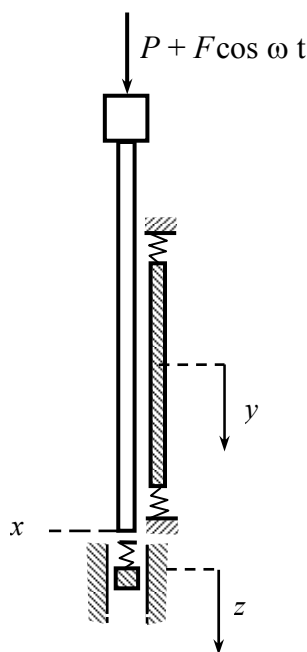


Рис. 40.4. Улучшенная модель, описанная уравнениями (40.9), вибропогружаемого шпунта с упрощенным учетом упругих свойств грунта под действием боковых и фронтальных воздействий; масса грунта не учитывается

Существенно новым является другая зависимость амплитуды колебаний шпунта от частоты и возможность ее существенного возрастания при приближении частоты к резонансной частоте колебаний шпунта. В этой уточненной модели при вибрациях сил сопротивления погружению в вязкое трение эффект перехода остается тем же самым, а с ростом амплитуды «вязкое трение», как и ранее, уменьшается, так что выгодно работать в режиме резонанса. Вот, пожалуй, и все. Может быть, следует отметить, что дальнейшее совершенствование этого метода вибропогружения привело к переходу от центробежных вибраторов – генераторов силы $F \cos \omega t$ – к виброударным механизмам, которые в некотором отношении оказались предпочтительнее. Простейшая модель виброударника была рассмотрена методом точечных отображений [12, 25], а изложенное опубликовано в [33].

Более полное математическое исследование описанных моделей вибропогружения также методом точечных отображений содержится в [6]. Позднее близкие задачи рассматривались И.И. Блехманом [13].

Можно еще отметить, что рассматривались и более полные модели с учетом вязкости грунта и его массы. Но чего-то существенно нового, кроме очевидного, это для приложений и понимания не давало.

41. ОСНОВНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СОВРЕМЕННОЙ НАУКИ И ТЕОРИЯ КОЛЕБАНИЙ

Настоящая книга – это лекции автора, читаемые им студентам младших (2–3) курсов, магистрантам, аспирантам и по другим поводам, обработанные, как единое целое, о математических моделях и моделировании. Но по мере приближения к переднему краю науки, к тому, что сделано в XX веке, последовательность и цельность изложения уступает место фрагментарности и, компенсируя ее, впервые делаются ссылки на литературу. Эта глава не лишена этого недостатка: она, по существу, – переработанное изложение докладов автора 1998 и 2001 годов на международных конференциях, посвященных памяти Я.З. Цыпкина и его и моего учителя А.А. Андропова, как ученых, имеющих мировую известность. Оба доклада тематически связаны между собой и настоящей книгой: в основе теории колебаний лежит математическая модель «динамическая система», которая и есть основная модель современной науки и основная модель настоящей книги. Александр Александрович Андронов, ученик Л.И. Мандельштама, признанный в мире создатель новой науки – теории нелинейных колебаний, и поэтому естественно связать изложение содержания, роли и места теории колебаний в науке именно с ним.

Динамическая система как основная математическая модель современной науки. Научное понимание мира человечеству пришло с небес. Невозможно представить себе другие объекты, нежели Солнце, Луна и планеты на фоне величественного звездного ночного или дневного неба, одновременно столь привлекающие внимание, значительные, загадочные и, вместе с тем, обладающие яркими простыми закономерностями. Не погода, приходящая тоже с небес, а именно Солнце, Луна, планеты и звезды, потому что капризы погоды неуловимы, а то, что предсказуемо, тесно связано с Солнцем и звездами, т.к. обнаруживается неизменная причинная связь. Птолемей, Коперник, Галилей, Ньютон – вот привычные вехи в истории познания неба и, наконец, Лаплас, перенесший причинную обусловленность небесных явлений на явления земные, провозгласивший причинную обусловленность и предсказуемость всех явлений мира и его изменений, всего, что в нем происходит. И это сделал тот самый Лаплас, который разрабатывал теорию вероятностей и ввел в эту науку самую известную случайную величину с распределением, носящим его имя.

Детерминизм Лапласа основывался на потрясающих воображение успехах астрономии и механистической трактовке мира, и повседневно наблюдаемая случайность его не поколебала. Это потом все явления мира разделили на предсказуемые и непредсказуемые, случайные, резко противопоставив их друг другу. А сегодня принцип детерминизма имеет всеобщий смысл, в расширенном понимании он справедлив и для марковских случайных процессов, статистической физики и квантовой механики. Он лежит в основе нашего понимания мира и воплощен в модели динамической системы, включающей описание, называемое состоянием, и однозначный оператор его прогнозирования. Состояние может не быть непо-

средственно наблюдаемым и измеряемым, но оно как-то связано с наблюдаемыми и измеряемыми величинами.

Следующая историческая веха – Анри Пуанкаре. Будущее следует за настоящим и им определяется. Этот принцип детерминизма реализуется в описаниях природных явлений дифференциальными уравнениями. До Пуанкаре усилия исследователей были направлены на получение их решений: аналитических, приближенных и численных. Пуанкаре стал изучать всю совокупность решений в их геометрической интерпретации в фазовом пространстве. Это повлекло новую топологическую точку зрения на совокупность всех решений, как на разбиение фазового пространства на фазовые траектории, и привело к качественной теории дифференциальных уравнений и теории динамических систем. Фазовый и бифуркационный портреты динамической системы позволили составить полное качественное представление о динамике, свойствах и явлениях в изучаемом объекте, которые могут в нем происходить с изменением времени и параметров.

Построение математических моделей эволюционных процессов в виде динамических систем – это проблема отыскания, изобретения самодостаточного описания – состояния – и нахождения оператора его изменения со временем. В отыскании этого самодостаточного описания человечество проявило громадную изобретательность. В механике состояние – это положения и скорости всех материальных точек, в электродинамике – придуманные М. Фарадеем фантастические электрические и магнитные поля, в квантовой физике – знаменитая функция ψ , для случайных марковских процессов и описываемых ими физических явлений – плотность вероятностей. Состояние конструируется человеком как самодостаточное описание, по которому возможен прогноз такого же описания. Изучения биологических и общественных явлений, возможно, приведут к совершенно неожиданным конструкциям состояния.

Для систем управления состояние – не только самодостаточное описание для прогноза, но и полная информация для организации управления. Понятия управляемости и наблюдаемости исходят из состояния, его управляемости и его наблюдаемости.

Изучению динамической системы как математической модели уже более ста лет. С одной стороны, результаты этого исследования огромны, и мы очень много узнали, а с другой, – и сегодня достаточно полное изучение конкретной многомерной системы – очень трудная и часто неразрешимая задача. Регулярный успех достигнут только для размерности системы $n \leq 2$; для $n = 3$ успех возможен лишь благодаря новой вычислительной технике, а при $n \geq 4$ – это сделать трудно и часто невозможно.

До последнего времени существовала пропасть между известными, конкретными системами, в подавляющем большинстве с простой структурой фазового портрета и простыми движениями, и теоретически открытыми в работах А. Пуанкаре и Д. Биркгофа сложными структурами и движениями, рекуррентными и устойчивыми по Пуассону, непериодическими и нерекуррентными. Устойчивые по Пуассону – это те, которые мы сегодня называем хаотическими и стохастическими, и странными аттракторами. Новое, что принесли 60–80 годы XX века, состоит не в открытии новых движений и странных аттракторов, а в осмыслении и наглядном показе, как эти сложные движения возникают в динамических систе-

мах, описываемых простыми по виду дифференциальными уравнениями, и как и почему эти движения, несмотря на соблюдение теоремы о единственности решения дифференциальных уравнений, оказываются непредсказуемыми и случайными.

Мне представляется, что первым ударом по устоявшимся ограниченным представлениям о поведении динамических систем, ударом, послужившим спусковым крючком, были работы начала 60-х годов XX века математика С. Смейла. До него представления о диссипативных динамических системах в значительной мере ограничивались тем, что наблюдалось у грубых двумерных систем [2–4]. Смейл показал, что в многомерном случае, $n > 2$, существуют грубые динамические системы с очень сложными движениями; с бесконечным числом периодических движений; в пространстве динамических систем могут быть области негрубых систем и, следовательно, негрубость не есть необщность [46].

Эти неожиданные факты в короткое время вызвали бум, который в 70–80 годах вовлек в свой вихрь новых людей, в основном физиков и механиков, не знающих ни классической теории динамических систем Пуанкаре–Биркгофа, ни исследований геодезических линий (траекторий свободного движения материальной точки) на поверхности отрицательной кривизны, начатых еще Ж. Адамаром, ни символической динамики, ни более поздних работ по теории нелинейных колебаний и исследованию конкретных систем. Им казалось, что все происходит впервые, для них это так и было.

Если говорить о произвольных динамических системах, то несомненно, что среди них очень много и, возможно, подавляющее большинство таких, которые мы никогда не сможем изучить. Но скорей всего, что это и не очень нужно. То, что является общим и частным с формально математической точки зрения пространства динамических систем, не всегда совпадает с тем, что интересно и встречается в реальных системах. Мне кажется, что хорошим примером этого различия может служить вопрос об общности и частоте встречаемости устойчивых равновесий. Однажды А.А. Андронов сказал мне, увлекающемуся пространством коэффициентов полиномов, что с точки зрения этого пространства при больших размерностях устойчивость – очень редкая вещь. Я с этим согласился, но на предложение этим заняться никак не откликнулся. Позднее этот вопрос увлек других исследователей, показавших, что это действительно так. Вместе с тем мне представляется, что это не имеет никакого отношения к частоте встречаемости устойчивых равновесий в природе, и особенно технике. Все дело в том, как выбрано пространство параметров, в котором выделяется область устойчивости, какие параметры выбирает природа, и какие конструирует человек. Вхождение варьируемых параметров определяется физической структурой системы. В свое время были интересные и отчасти забытые работы И.И. Гальперина по выделению структурных схем линейных систем управления, устойчивых при любых значениях параметров, устойчивых в силу своей структуры. Сегодня упорно во всем мире ищутся системы управления, малочувствительные к изменениям своих параметров, грубые – по терминологии Андронова–Понтрягина, робастные – по-английски. Мне кажется, что в этом вопросе представляет интерес отыскание систем, робастных по своей структуре.

Поведение нашего мира, по-видимому, существенно отличается от изучаемых нами динамических систем с компактным фазовым пространством. Свойство повторяемости, имеющее место у установившихся движений динамических систем, не присуще нашему миру. В нем, скорее всего, все происходит наново. Эпиклезиаст был не прав, говоря, что все, что было – и будет. Прав был Демокрит. Но на ограниченных промежутках времени в ограниченных масштабах упрощающая идеализация, принимаемая теорией динамических систем, вполне приемлема.

По-видимому, самая простая, эффективная и практически значимая классификация движений динамических систем – это разделение их на устойчивые и неустойчивые, на устойчивые и неустойчивые по Ляпунову и орбитно устойчивые и неустойчивые. Различие между устойчивостью по Ляпунову и орбитной устойчивостью весьма значительно. Устойчивость по Ляпунову – это близость законов изменения состояния во времени для невозмущенного и возмущенного движений. Орбитная устойчивость – это неограниченное пребывание возмущенного движения в малой окрестности невозмущенного. Законы изменения состояний возмущенного и невозмущенного движений при этом могут значительно различаться. Из сказанного следует, что из устойчивости по Ляпунову всегда следует орбитная устойчивость, а обратное не имеет места.

Простые установившиеся движения динамических систем – это устойчивые состояния равновесия и периодические движения, сложные – устойчивые по Пуассону, отличные от простых. Они обладают ослабленным свойством периодичности, состоящим в повторяемости со временем с любой степенью точности любого из пройденных ранее состояний, они орбитно устойчивы, но вообще неустойчивы по Ляпунову.

Устойчивость и неустойчивость по Ляпунову в общем случае с точки зрения значений показателей Ляпунова – это экспоненциальная устойчивость или неустойчивость. Предполагая эту общность, сложные движения – это орбитно устойчивые движения, экспоненциально неустойчивые по Ляпунову, т.е. такие, от которых почти всякое возмущенное движение локально экспоненциально удаляется, а в целом все время блуждает по некоторой области – аттрактору. Однако устойчивое по Пуассону и экспоненциально устойчивое или асимптотически устойчивое движение обязательно просто периодическое.

Почти все остальные движения асимптотически приближаются к простым или сложным установившимся движениям, состояниям равновесия, периодическим движениям и движениям, устойчивым по Пуассону.

В дополнение к сказанному я хотел бы кратко ответить на вопросы предсказуемости и непредсказуемости движений динамических систем и механизма возникновения непривычных, сложных, непредсказуемых и случайных движений. Они уже затрагивались ранее при рассмотрении стохастического осциллятора – «часов наоборот». Ниже о том же, но в более общем плане.

В отсутствие неизвестных возмущений и погрешностей счета все движения предсказуемы. Фактическая непредсказуемость возникает, с одной стороны, от наличия неустраняемых случайных или неизвестных возмущений, а с другой – столь же неизбежных погрешностей вычислений. Различие между этими двумя факторами состоит в неповторяемости первого и повторяемости второго при использовании идентичных цифровых вычислительных средств. Однако результи-

рующее действие этих неустранимых факторов погрешностей может быть существенно разным.

Возможно, что с уменьшением неизбежных погрешностей неограниченно растет и достижимая точность прогноза для всех $t > 0$. В этом случае мы склонны считать, что имеет место предсказуемость. Но возможен и совершенно другой случай, когда, сколь бы ни были малы неизбежные погрешности, ошибка прогнозирования может быть не меньше некоторой конечной величины, которая может существенно превосходить допустимую. В этом случае мы вынуждены признать, что имеет место непредсказуемость. Именно это имеет место для сложных движений в силу их экспоненциальной неустойчивости по Ляпунову.

Динамическую систему можно рассматривать и с привычной для специалистов по теории управления точки зрения, как звено с входом и выходом, однако при этом следует считать, что в звене могут автономно происходить динамические процессы, которые возмущаются входным случайным воздействием. При таком подходе в зависимости от этих автономных движений звено может быть преобразователем, усилителем или генератором стохастичности [28]. Преобразователь стохастичности – это то, что хорошо известно и настойчиво изучалось. Усилитель и генератор стохастичности – нечто совершенно неожиданное, ранее никак не учитываемое.

Дисперсия выхода усилителя стохастичности может быть намного большей дисперсии входного воздействия, и даже с локальным коэффициентом усиления, стремящимся к бесконечности при уменьшении случайного входа. При этом стохастические свойства выхода, по-прежнему, как и в случае преобразователя стохастичности, определяются исчезающе малым случайным входом, но выход неограниченно уменьшается вместе с входом. Для генератора стохастичности при сколь угодно малости случайного входа дисперсия выхода остается больше некоторой конечной величины, а его статистическое описание не зависит от входа. Оно полностью определяется самой невозмущенной системой, ее детерминированным описанием.

Наконец, о механизме возникновения сложных, устойчивых по Пуассону, движений усилителей и генераторов стохастичности. В свое время рождение теории автоколебаний было неразрывно связано с открытием Андроном связи между ними и предельными циклами Пуанкаре. В какой-то мере это повторилось и со сложными движениями, но теперь благодаря не введенным и изучаемым Пуанкаре предельным циклам, а открытым им гомоклиническим и гетероклиническим кривым [41], открытым в связи с безнадежно и длительно изучаемой им ограниченной проблемой трех тел. Гомоклинические и гетероклинические фазовые кривые или движения динамической системы – это двоякоасимптотические движения при $t \rightarrow \infty$ и $t \rightarrow -\infty$, стремящиеся к одному и тому же или, соответственно, разным неустойчивым седловым периодическим движениям. Вот эти, естественно и общим образом возникающие гомоклинические и гетероклинические кривые (движения), асимптотические, казалось бы, к несущественным неустойчивым седловым периодическим движениям, как выяснилось, могут порождать сложные, орбитно устойчивые и неустойчивые по Ляпунову движения [22–25, 28].

О сложности этих движений писал уже А. Пуанкаре, но он, по-видимому, не подозревал, что они могут порождать сложные установившиеся движения. Слож-

ные движения, как оказалось, могут порождать двоякоасимптотические движения не только к периодическим движениям, но и состояниям равновесия, или к тем и другим (будем называть их также гомоклиническими и гетероклиническими).

Широко известно, что устойчивое периодическое движение может родиться от теряющего устойчивость состояния равновесия (бифуркация Андронова). Аналогично, сложное движение, орбитно устойчивое и неустойчивое по Ляпунову, может родиться от замкнутых контуров, составленных из гомоклинических и гетероклинических кривых.

Выше речь шла об описании и поведении изолированной системы, не подверженной влиянию внешних воздействий, ее еще называют автономной. Только в этом случае состояние и его изменение определяются самой системой. В неизолированной системе, подверженной внешним по отношению к системе воздействиям, это уже не так: изменение состояния зависит не только от самого состояния, но и внешних воздействий. Динамическую систему с внешним воздействием называют еще автоматом, с которым мы встречаемся при описании игр, целесообразного поведения и распознавания (перцептрон). Автомат предполагает наличие выхода, определяемого его состоянием, либо состоянием и внешним воздействием. Это сближает понятие автомата со звеном с входом и выходом, а в случае его линейности – с коэффициентом передачи и амплитудно-фазовой частотной характеристикой. Если внутреннее устройство звена нам неизвестно, а можно регистрировать только его вход и выход, то такое звено называют еще черным ящиком. Неизолированная, неавтономная система, безусловно, более общая, чем изолированная и автономная, но несмотря на это, именно изолированная, автономная система остается основной моделью, поскольку она описывает базовые основополагающие законы природы, и поскольку путем расширения рассматриваемого объекта всегда, как мы думаем, можно достичь его изолированности и автономности. С этой точки зрения неавтономность есть результат выделения некоторой части в изолированной автономной системе.

А.А. Андронов и теория колебаний как наука об эволюционных процессах и явлениях. Александр Александрович Андронов создал новую науку – теорию нелинейных колебаний, всемирно известную научную школу и оставил учеников, продолживших его дело [37]. Теория колебаний сегодня – это широкая всеобъемлющая наука об эволюционных процессах в природе, технике и обществе, в механике, физике, астрономии, химии, биологии, и во всем, что нас окружает, и в нас самих. Буквально и в восприятии многих людей название «Теория колебаний» не адекватно тому, что я сказал. Но так распорядилась история: название пришло от колебаний маятника, механического и электрического осцилляторов, и история же определила мощь и значимость теории колебаний как науки всеобъемлющей, воплощающей в себе великую идею изоморфизма и единства природы и мира.

Механика – теоретическая и прикладная – включает в себя и теорию колебаний, и теорию устойчивости, и отчасти даже теорию управления. Во всяком случае, эти разделы имеются в четырехтомнике «Механика в СССР за 50 лет» и в программах традиционных съездов по механике. Более того, основоположник современной теории устойчивости А.М. Ляпунов был академиком по отделению механики. Лагранж, создавший теорию малых колебаний, основывался на механике.

Знаменитый трактат Рэлея, от которого, по нашим представлениям, ведет свое начало наука о колебаниях, назывался «Теория звука», т.е. относился к механике. Известная книга Уиткера «Аналитическая механика» включает не только теорию колебаний и теорию устойчивости, но и теорию динамических систем. Знаменитый труд Пуанкаре, основополагающий в теории динамических систем, называется «Новые методы небесной механики». Широко известные инженерные книги по теории колебаний Ден Гартога и С.П. Тимошенко рассказывают о колебаниях механических систем. Если этим ограничиться, то создается впечатление, что теория колебаний целиком относится к механике. Однако сказанное говорит лишь об огромной роли механики в возникновении и становлении теории колебаний. Уже в лекциях Л.И. Мандельштама 1930–31 и 1931–32 учебных годов, прочитанных на физфаке МГУ и изданных в 1955 году, теория колебаний предстает как самостоятельная наука. В них идет речь о колебаниях не только в механике, но и во многих разделах физики. Благодаря успешным исследованиям школы Мандельштама, в 1931 году состоялась первая Всероссийская конференция по теории колебаний. Именно на этой конференции А.А. Андронов выступает с широким и обстоятельным докладом о математических проблемах теории колебаний. Следующей решающей вехой в становлении теории колебаний как самостоятельной науки является книга А.А. Андронova, А.А. Витта и С.Э. Хайкина «Теория колебаний», вышедшая в 1937 году. Следует отметить и замечательную книгу Г.С. Горелика «Колебания и волны», написанную в 1950 году. Первая книга заложила основы теории нелинейных колебаний, вторая еще более расширила охват теорией колебаний разнообразных физических явлений.

Значительный вклад в становление теории колебаний как самостоятельной науки внесла и школа Крылова–Боголюбова–Митропольского. Именно она инициировала и организовала длительную традиционную серию международных конференций и симпозиумов по теории нелинейных колебаний с разделами «Аналитические и качественные методы исследования и их приложения» и проводила весьма длительную серию научных школ по этой тематике. В более поздний период с 1972 года, когда уже самостоятельность теории колебаний как науки не вызывала сомнений, следует отметить научные школы, организуемые в г. Горьком (Нижем Новгороде): Научно-исследовательский радиофизический институт (НИРФИ), затем Институт прикладной физики (ИПФ РАН), а также конференции, проводимые Научно-исследовательским институтом прикладной математики и кибернетики (НИИ ПМК). Имевший место на них бум, вызванный «открытием» хаотических и стохастических колебаний, охватил всю страну. Теорию колебаний стали ассоциировать с нелинейной физикой, нелинейной динамикой и синергетикой. Теория колебаний стала наукой об общих закономерностях эволюционных процессов в механике, физике, радиотехнике, радиофизике, гидродинамике, теории управления, химии, биологии, науке об обществе и т.д. Одновременно теория колебаний обрела прочную математическую опору в качественной теории дифференциальных уравнений и теории динамических систем, не только используя их, но и существенно определяя их проблематику и направления исследований.

Из сказанного видно, сколь многообразны связи теории колебаний с другими науками, и что другие науки – не только объект приложений ее методов, но и неиссякаемый плодотворный источник новых вопросов, задач и проблем.

Ядром теории колебаний являются математические модели законов природы и происходящих в ней процессов, включая технику и общество. В основе этих моделей лежит понятие состояния и оператора его изменения. Состояние изображается, оператор открывается. Фарадей изобрел электрические и магнитные поля – состояние, Максвелл написал уравнения, определяющие оператор.

В зарождении и развитии каждой науки основную роль играют так или иначе установленные факты, методы изучения и общее понимание. В теории колебаний факты – это накопленные знания математических моделей и характера протекания описываемых ими процессов и сопровождающих их явлений; методы – способы изучения этих математических моделей в виде обычных дифференциальных уравнений или уравнений в частных производных, или в виде динамических систем того или иного вида; общее понимание – это прежде всего наглядные представления о возможных эволюционных процессах, в частности представления о фазовом пространстве, состояниях равновесия и периодических движениях, автоколебаниях, хаосе и странном аттракторе, волнах, волновых и модовых взаимодействиях, дисперсии, резонансе, синхронизации, бифуркациях и бифуркационном портрете, локальной и глобальной устойчивости, области притяжения, мягком и жестком возбуждениях, гистерезисе, принципе суперпозиции, скользящих движениях и разрывах и т.д.

Основные типы эволюционных процессов можно разделить на:

- 1) порядок, синхронизацию и предсказуемость на основе асимптотической устойчивости по Ляпунову;
- 2) хаос, непредсказуемость и стохастичность на основе экспоненциальной неустойчивости при орбитной и глобальной устойчивости;
- 3) гомеостаз, реализацию цели или программы и самоорганизацию на основе направленных обратных связей и управления.

В этом кратком перечислении преодоленные в XX веке многовековые заблуждения, огромные новые знания, научное мировоззрение и понимание эволюционных процессов в природе, технике и обществе. В этом не только вчера и сегодня теории колебаний, но и предвидимое завтра. Одно из важнейших различий вчера и сегодня – это то, что вчера теория колебаний в основном изучила и освоила простейшие установившиеся движения – состояния равновесия и периодические движения, хотя неявно сложные движения проявляли себя в статистической термодинамике и физике, турбулентности жидкостей, газов и плазмы, а сегодня стало возможным все это рассматривать с единых позиций, как проявление все той же динамической системы. Кроме того, это еще и изучение процессов, основанных на направленных связях и управлении, изучение выходящего за рамки бурно развивающейся теории автоматического регулирования. Важно это не только для понимания природы, человека, общества, но и их эволюции, создания интеллектуальных машин.

Многие очень крупные открытия в науке совершаются дважды: первый раз абстрактно-теоретически, отдельными людьми, и это открытие никого не волнует и не трогает, его не замечают, и второй раз практически жизненно, когда оно овладевает многими умами и потрясает людей, открывая им что-то ранее неведомое, неожиданное и очень важное. В теории колебаний так было с открытием автоколебаний и с открытием хаотических и стохастических колебаний.

У крупных открытий есть и третья фаза, когда они становятся общеизвестными и как бы «исчезают». Так произошло с великим открытием, что Земля – шар и ничем не поддерживается, что Земля вращается вокруг Солнца, что две массы или два заряда взаимодействуют друг с другом, хотя между ними пустота. Так произошло и с автоколебаниями, происходит сейчас с хаотическим и стохастическим движениями, так произошло даже со специальной теорией относительности: она уже никого не удивляет, и студенты воспринимают ее как должное и даже обычное.

Первый раз устойчивые предельные циклы открыл Анри Пуанкаре. Пуанкаре – по типу мышления великий геометр – изобразил решения дифференциальных уравнений в виде кривых в пространстве их переменных, и именно поэтому естественно и легко обнаружил предельные циклы. Они были естественны потому, что в некоторых общих ситуациях замыкание кривой было неизбежным, и эту неизбежность, во всяком случае на плоскости, Пуанкаре сделал очевидной, благодаря введению секущего отрезка и функции последования на нем (рис. 41.1). Была очевидна не только неизбежность замыкания кривой, т.е. периодичность решения дифференциального уравнения, но и возможность его устойчивости. И именно эта естественность и простота до примитивности способствовали тому, что никто на это не обратил особого внимания, и никому не пришло в голову, что это интересно, очень важно и несет открытие человечеству.

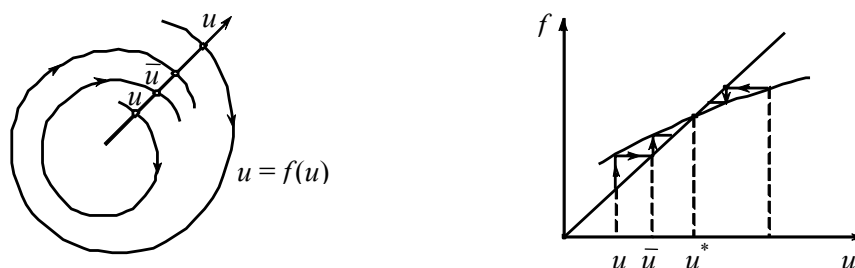


Рис. 41.1. Секущий отрезок, функция последования и диаграмма Ламерея; u^* – неподвижная точка, отвечающая устойчивому предельному циклу

Второй раз, спустя значительное время, устойчивые предельные циклы в виде автоколебаний реальных систем открыл А.А. Андронов [5]. Открыл не как абстрактные устойчивые замкнутые кривые, а как практически жизненно важные явления в механике, электричестве, астрономии, химии и биологии; как явления в природе и технике: в далеких звездах цефеидах, охватывающей весь мир радиопередаче, звучании музыкальных инструментов, шимми автомобиля и флаттере крыла самолета, могущих их разрушить, в вибрации инструмента при механической обработке, препятствующей ее точности, и многом, многом другом.

Это второе открытие вызвало революцию во взглядах инженеров и ученых. Подчас оно воспринималось с недоверием, с большим трудом, и нередко активно отвергалось, как это, например, имело место в химии с известной реакцией Белоусова, хотя эта реакция демонстрировалась в прозрачной колбе с химическим раствором, периодически меняющим свой цвет. Даже в механике, где, казалось бы, все достаточно просто, наблюдалось длительное неприятие и, например, шимми автомобиля трактовалось как резонанс с воздействием неровностей дороги, а виб-

рации при точении трактовались как тот же резонанс, вызванный отрывами стружки, и т.п. Признание автоколебаний было без помех, пожалуй, только в зарождающейся радиотехнике, но там податься было некуда, на них все основано.

А.А. Андронов не просто указал на идентичность предельных циклов А. Пуанкаре и автоколебаний в реальных объектах. Он нашел их, изучил во многих реальных конкретных системах, обнаружил и исследовал явления захвата и затягивания – сегодня синхронизации – при периодическом воздействии на автоколебательную систему и при взаимодействии ее мод, опираясь сначала на метод малого параметра А. Пуанкаре, а затем в сильно нелинейных системах – на метод точечных отображений. Он изучил, как возникают и исчезают автоколебания при изменении параметров (рис. 41.2–41.5) – это широко сейчас известная бифуркация рождения автоколебания – предельного цикла – из теряющего устойчивость равновесия и другие (от замкнутой сепаратрисы седла, при слиянии и исчезновении устойчивого и неустойчивого равновесий) [2–4].

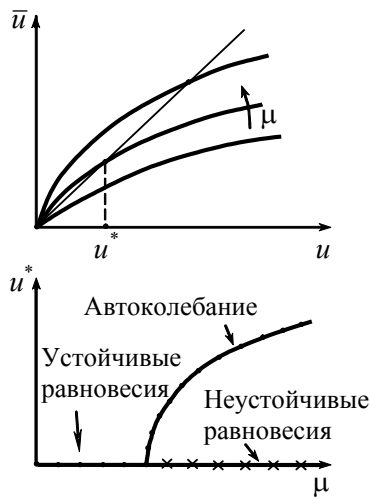


Рис. 41.2. Бифуркация рождения автоколебания (мягкое возбуждение) при изменении параметра μ . Диаграмма Ламерея и бифуркационная диаграмма

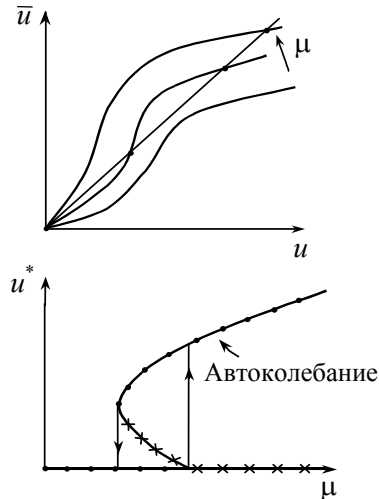


Рис. 41.3. Жесткое возбуждение автоколебаний

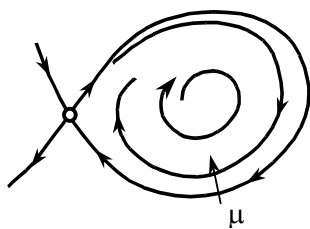


Рис. 41.4. Рождение автоколебания от сепаратрисы седла

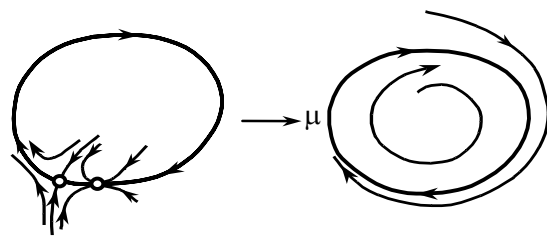


Рис. 41.5. Рождение автоколебания при слиянии седла и узла

Эти открытия были сделаны применительно к гладким системам второго порядка. Обобщение на многомерный случай и более общие системы (кусочно-гладкие и сингулярно возмущенные) было делом последующих исследований, как и использование новых возможностей вычислительной техники (рис. 41.6–41.8).

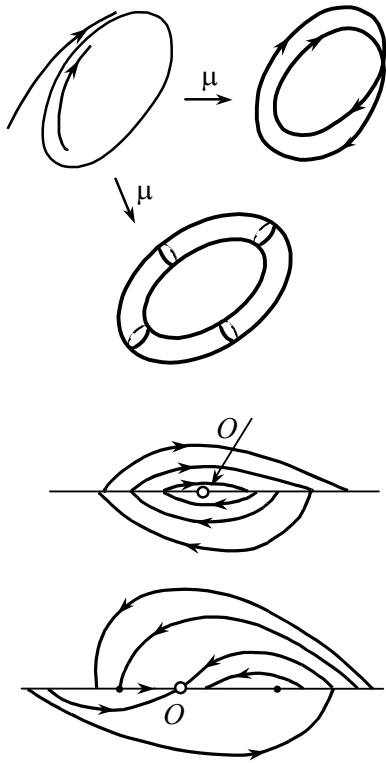


Рис. 41.6. Новые бифуркации автоколебания в многомерном случае (удвоение периода, рождение тора)

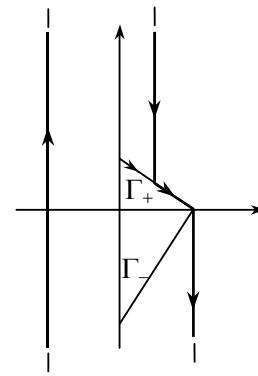


Рис. 41.8. Разрывные автоколебания в предельной сингулярно возмущенной системе

Рис. 41.7. Новые возможные равновесия O в разрывных кусочно-гладких системах (в точке равновесия правые части дифференциальных уравнений в нуль не обращаются)

Все это привело к тому, что вместо мифических механизмов возникновения колебаний человечество научилось открывать истинные причины колебаний и устранять их там, где они вредны и опасны, и вызывать, где нужны и полезны. В современной технике обоих случаев очень много, и барьер неприятия автоколебаний, возникающих как бы из ничего, был преодолен. До этого были понятны затухающие и в идеализации незатухающие колебания маятника, грузика на пружинке, упругой среды и т.д., понятны вынужденные колебания, вызываемые действием переменной силы или напряжения, было понятно и удивительное явление резонанса, но как могут возникать колебания «из ничего», без вызывающего их воздействия, было непонятно и неприятно, и этот психологический барьер был преодолен, и сегодня автоколебания всем привычны и понятны.

Такова сравнительно простая и ясная история открытия автоколебаний и связанных с ними понятий устойчивости, фазового портрета, бифуркаций, бифуркационного портрета, синхронизации, грубости и других. Нечто похожее, но в большем масштабе, в соответствии с увеличением числа вовлеченных в науку людей и трудностью проблемы, произошло и с открытием сложных установившихся движений: хаотических и стохастических. Но теперь число непосредственных участников огромно, все переплетено и запутано, а причина неприятия неоднозначна. Описать все это сколько-нибудь полно невозможно. История эта проти-

воречива: с одной стороны, движения как бы не принимались, а с другой, – на них базируются статистическая термодинамика и физика, не отрицается и турбулентность, так что дело не в непризнании, а непонимании, как детерминизм может порождать непредсказуемость и случайность и связанную с ними сложность. Но и это не совсем так: общее объяснение этому давали и Пуанкаре, и Крылов, и другие. Не было ясного приемлемого механизма, как это происходит – наверное, это главное. С этой точки зрения, опять первый Пуанкаре.

Открытие Пуанкаре – это гомоклинические и гетероклинические кривые и связанное с ними необычайно сложное и запутанное поведение соседних кривых, которые Пуанкаре даже не пытался изобразить. Именно в этой непостижимой сложности увидел Пуанкаре камень преткновения в ее решении, но он не придал этому большого значения, точнее, не продолжил исследование. Почему так случилось, трудно сказать, возможно, не пришло время, отвлекли другие проблемы.

Гомоклинические и гетероклинические кривые уже не так просты, как предельные циклы, секущий отрезок и функция последования. Вместо секущего отрезка Пуанкаре прибег к секущей плоскости и порождаемому на ней точечному отображению T .

На рис. 41.9 седловое периодическое движение Γ и гомоклиническая к нему кривая γ . На рис. 41.10 их сечение двумерной плоскостью, на которой O – седловая неподвижная точка пересечения Γ с секущей, а точки $\dots, M_{-1}, M_0, M_1, \dots$ – последовательные точки пересечения секущей с гомоклинической кривой γ и одновременно часть точек пересечения инвариантных кривых S^- и S^+ седловой точки O . Сложность, о которой говорил Пуанкаре, – это поведение кривых S^- и S^+ : ни одна из них не пересекается сама с собой, но они пересекаются друг с другом в бесконечном числе точек и неограниченно продолжаются. Как они это делают, Пуанкаре описывать не стал.

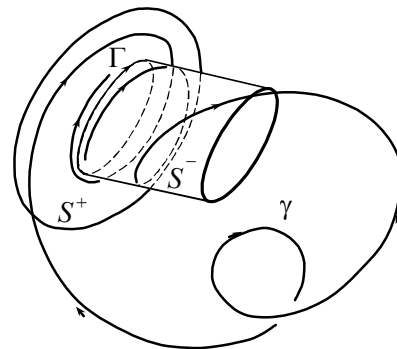


Рис. 41.9. Седловое движение Γ и гомоклиническая кривая Пуанкаре γ

Можно заметить, что если выделить показанную на рис. 41.10 область D , то области $\dots, \sigma_{-1}, \sigma_0, \sigma_1, \dots$ и области $\dots, \nu_{-1}, \nu_0, \nu_1, \dots$, при точечном отображении T последовательно переходят друг в друга, образуя один поток входящих в область D точек, и другой – выходящих из нее. Чтобы понять всю сложность происходящего с ними внутри области D , укажем, что в области G (рис. 41.11), составленной из окрестности $\varepsilon(\Gamma)$ седловой замкнутой кривой Γ и окрестности $\varepsilon(\gamma)$ гомоклинической кривой γ имеется бесчисленное множество разных седловых кривых. Если окрестности $\varepsilon(\Gamma)$ и $\varepsilon(\gamma)$ достаточно малы, то всякой бесконечной в обе стороны последовательности целых чисел вида

$$\dots, j-2, j-1, j_0, j_1, j-2, \dots \quad (j-2 > N)$$

отвечает единственная седловая кривая, которая последовательно оборачивается $\dots, j-2, j-1, j_0, j_1, j-2, \dots$ раз возле кривой Γ в $\varepsilon(\Gamma)$, а между ними возвращается в $\varepsilon(\Gamma)$, проходя вдоль гомоклинической кривой γ в ее окрестности $\varepsilon(\gamma)$ [22–24]. Из этого следует, что, кроме гомоклинической кривой γ в G существует бесконечное множество других гомоклинических кривых, возле каждой из которых есть, в свою очередь, такие же кривые, и так до бесконечности.

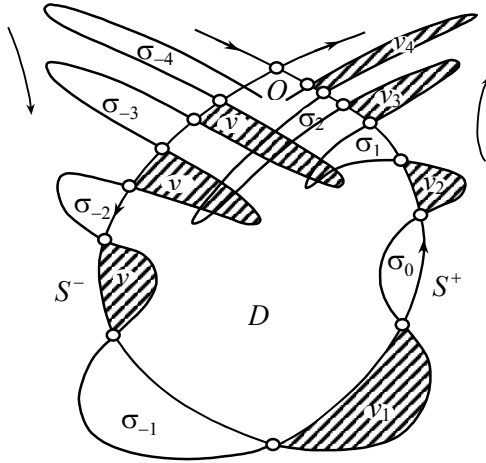


Рис. 41.10. Гомоклиническая структура на секущей поверхности

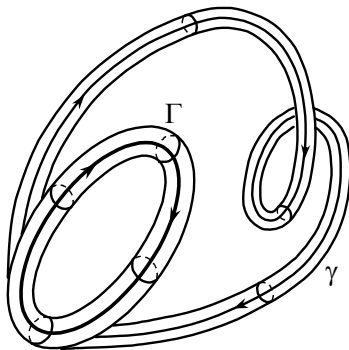


Рис. 41.11. Окрестности гомоклинической структуры

Теперь, глядя «сквозь магический кристалл», общий механизм образования сложных движений можно представить в виде последовательно-параллельного прохождения фазовых траекторий вблизи гомоклинических и гетероклинических кривых [24]. При этом происходит сложное и тонкое их разделение, подобно тому, как это происходит с входным потоком областей $\dots, \sigma_{-2}, \sigma_{-1}, \sigma_0, \sigma_1, \dots$. Если область σ_i пересекается с v_{-j} , то она выбрасывается наружу (вовне области D), как точка областей $v_{-j+1}, \dots, v_{-1}, v_0, v_1, \dots$. Вариантов тут очень много, но механизм один. Наглядно и схематически этот механизм похож на то, что происходит с горошинами на доске Гальтона, если бы их снизу собирать и сыпать в верхнюю воронку снова и снова.

Со сложными движениями столкнулся и Ж. Адамар, изучая геодезические линии на поверхности отрицательной кривизны в динамической интерпретации

свободного движения по ней материальной точки. Это открытие было воспринято математиками и привело к символической динамике и ее статистическому описанию. Здесь можно обратить внимание на не выявленную в то время связь с движениями в гомоклинической структуре.

Сложные движения присутствуют и в теории центральных движений Д. Биркгофа [23], но скорее это то, что может быть, а не как это может быть.

Все сказанное выше можно отнести к первому этапу открытия сложных движений. Второй этап массового повального увлечения сложными хаотическими и стохастическими движениями детерминированных динамических систем вызван, пожалуй, не столько новыми теоретическими открытиями этого первого этапа, сколько реальной возможностью их получения и наблюдения во многих конкретных системах с помощью компьютера. Компьютер стал решающим доводом их существования и реальности, это породило своеобразное компьютерное понимание и толкование их происхождения и классификации. После сказанного хочу изложить взгляды А.А. Андропова на сложные движения, относящиеся к первому этапу.

А.А. Андронов знал о сложных центральных движениях Биркгофа и говорил о них в опубликованном докладе первой Всероссийской конференции по колебаниям в 1931 году. Он говорил об общей классификации движений динамических систем, центральных движениях Биркгофа, включая рекуррентные и устойчивые по Пуассону – сегодня хаотические и стохастические – и о том скромном месте, которое занимают среди них автоколебания. Поставив ряд насущных проблем теории колебаний, он не ставил задачу существенного расширения круга исследуемых движений. История показала, что это было бы несвоевременным, нужны были другие, более мощные подходы и средства, чем те, которыми располагала математика и наука в то время. Расширение этих исследований произошло много позднее, в 60–80 годах XX века. Выступление А.А. Андропова относилось к тому времени, когда основным освоенным инструментом исследования было сочетание метода малого параметра Пуанкаре и теории устойчивости Ляпунова. В их рамках сложные движения не обнаруживаются.

А.А. Андропова привлекала новая общая дисциплина – динамика машин, он считал, что теория колебаний должна привести к ней и позволит ее создать. Он оставил наброски содержания этой предполагаемой книги, которую ему не было суждено написать. Машина, как он ее понимал, для того чтобы функционировать, должна обладать свойством грубости, а ее рабочее движение должно быть асимптотически устойчиво по Ляпунову. С этих позиций единственные подлежащие изучению установившиеся движения – это асимптотически устойчивые равновесия и периодические движения. Возможно, именно поэтому в широко известной книге «Теория колебаний» [2] в примечании к введению он написал: «Однако, по-видимому, требование грубости системы запрещает все остальные рекуррентные движения, кроме состояний равновесия и периодических». Заметьте – речь идет не об устойчивых по Пуассону непериодических движениях, а только о рекуррентных. Рекуррентных – потому, что устойчивые по Пуассону он, по-видимому, считал вообще не подходящими и не нужными для машин. Понятие же грубости – еще до работы с Понтрягиным – понималось очень широко и имело несколько ас-

пектов: физическая реализуемость, математическая корректность модели, приближенная вычислимость и даже общность.

Сказанное относится к 1930-м годам. Позднее, во второй половине 40-х годов, А.А. Андронов проявил интерес к сложным движениям в конкретных системах. Его аспиранты Н.А. Фуфаев и А.С. Алексеев обнаружили и исследовали их в электромагнитном прерывателе и регуляторе температуры с опережением. Дальше эти исследования были продолжены его учениками и учениками учеников [27]. Они внесли значительный и существенный вклад как в первый этап, так и во второй – этап повального увлечения сложными хаотическими и стохастическими движениями, которые сегодня обнаруживают всюду и о которых сейчас знают все.

Следует отметить, что наиболее значительные результаты в общей теории динамических систем были получены в 60-е годы XX века С. Смейлом, как он сам пишет, под влиянием работ А.А. Андропова.

В заключение приведу шуточный пример двух крупнейших открытий XX века в познании динамики эволюционных процессов. Представьте нашу жизнь в виде скачущего по лестнице мячика сначала вверх, потом вниз. Вверх – потому, что он молод и коэффициент k отскока больше единицы, вниз – потому, что повзрослел и коэффициент отскока уже меньше единицы. На рис. 41.12 приведены диаграммы Ламерея для последовательных значений скоростей отскока от ступенек лестницы. Подъем стохастичен, спуск – регулярен.

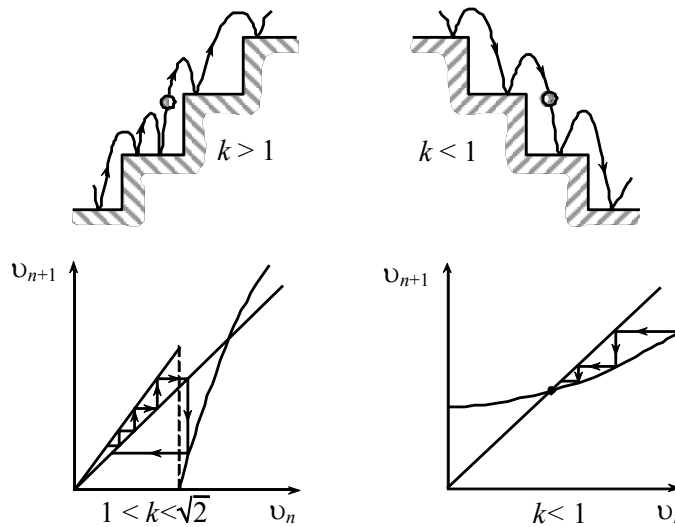


Рис. 41.12. Прыгающий по лестнице мячик с $k > 1$ и $k < 1$

42. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КАК ПЛОДОТВОРНАЯ ИДЕЯ НАУЧНОГО ИССЛЕДОВАНИЯ. D-РАЗБИЕНИЕ

Поскольку эта книга о математических моделях, то естественно отметить и обратить внимание на все аспекты и методы их использования. А они действительно широки и незаменимы, поскольку только с помощью модели можно найти ответы на многие вопросы, если их удастся сформулировать на математическом языке. Этот ответ может следовать только из каких-то известных нам сведений, а эти сведения, сформулированные на математическом языке, и есть математическая модель. Поэтому все внутриматематические исследования, как и исследования всех математически поставленных задач естествознания, техники и прочего – это тоже математическое моделирование. Так, теория целых положительных чисел – это изучение математической модели, состоящей из элементов a_0, a_1, a_2, \dots , связанных соотношениями

$$a_n + a_0 = a_n, \quad a_{n+1} = a_n + a_1, \quad n = 0, 1, 2, \dots$$

Среди давно сложившихся и широко используемых весьма эффективных внутриматематических моделей можно указать на модели геометрий, топологий, всевозможных групп, линейных и нелинейных уравнений, дифференциальных уравнений. К ним относится, конечно, и основная математическая модель в настоящей книге – динамическая система, которую можно трактовать и как полугруппу, а при взаимной однозначности ее оператора, как одномерную группу Ли. Пример того, как простые математические модели указывают путь решения очень сложной и, казалось бы, неприступной проблемы, уже был – это распознавание образов и перцептрон. Там такими моделями были модели объекта, образа и распознавания образов соответственно в виде точки многомерного пространства, множества точек в нем и узнавания принадлежности нового объекта этому множеству. Теперь будет описан другой пример, в котором речь идет о проблеме устойчивости движения. Рассказ состоит из двух частей. В первой части речь идет о том, как под влиянием настоятельных потребностей приложений сформировалась эта простая модель, во второй – к каким новым научным результатам она привела.

Теория управления как наука – детище XX века, но зародилась она еще во второй половине XIX века в работах гениального физика Д.К. Максвелла (1868) и выдающегося инженера И.А. Вышнеградского (1877). Появились они под давлением острой практической потребности обеспечения устойчивости и стабилизации скорости вращения паровых машин – основного источника механической энергии начавшейся промышленной революции. Создал паровую машину Д. Уатт вместе с центробежным регулятором скорости вращения, и вот эти регуляторы начали обнаруживать непокорный нрав: чем, казалось бы, лучше их делали и большего хотели от них получить, тем хуже они работали, подчас приводя к авариям. Возникшая инженерная проблема была достаточно масштабной: только в Англии было порядка ста тысяч паровых машин, а затем к ним присоединились и паровые турбины. Непокорный нрав регуляторов – неустойчивость – породил математическую теорию устойчивости, начавшуюся с упомянутых работ Максвелла

и Вышнеградского, впитавшую затем идеи и методы великих А. Пуанкаре и А.М. Ляпунова и разросшуюся вширь и вглубь в результате интенсивной деятельности многих выдающихся исследователей разных стран.

Историю теории автоматического управления можно рассматривать, как историю разрешения конфликта между устойчивостью и качеством управления. Именно этот конфликт проявился в центробежных регуляторах Уатта, именно он привел к изодромному регулятору, релейным системам, адаптивному и робастному управлению. Выбор стратегии управления ограничивается требованием к динамике реализующей ее системы, что влечет во многих случаях необходимость предварительного теоретического исследования, важнейшим элементом которого является определение условий устойчивости. Для линейных систем, во всяком случае с формальной точки зрения, проблема устойчивости полностью решается известными критериями Рауса и Гурвица, возникшими вскоре под влиянием работ Максвелла и Вышнеградского. Поначалу эти алгебраические критерии устойчивости Рауса и Гурвица казались исчерпывающими. Потом выяснилось, что они не поддаются обобщению на не алгебраические уравнения и не вполне удовлетворяют практическим потребностям, особенно в вопросах конструирования новых систем управления, машин и устройств. Инженерам нужен был метод исследования устойчивости, позволяющий понять, в каких пределах они могут выбирать и менять параметры конструируемой системы, не нарушая ее устойчивости, какой свободой они обладают для выполнения остальных требований. Именно поэтому уже в первой половине XX века возникли новые критерии устойчивости, среди которых широкую известность получили графические критерии Михайлова и Найквиста. Отметим, что уже Вышнеградский получил критерий устойчивости в виде алгебраического неравенства, представив его в виде области устойчивости на плоскости двух параметров – неравномерности и специально введенного вязкого трения. На этой диаграмме конфликт между устойчивостью и качеством стабилизации скорости вращения паровой машины получил наглядное геометрическое воплощение: определяющие их величины должны находиться в некоторой области D , вид которой таков, что уменьшение одного параметра требует увеличения другого. Из сказанного следует, что требовалась новая постановка математической задачи, более общая и адекватная потребностям приложений, чем воспринятая математиками проблема Рауса – Гурвица, состоящая в отыскании необходимых и достаточных условий, что все корни полинома лежат в левой полуплоскости (условий $\operatorname{Re} z_\nu < 0$ при всех ν). Центральным моментом появления нового подхода к описанному выше кругу задач, помимо ошеломляющей загадочности не вполне обоснованного, но верного критерия Найквиста, является недоумение по поводу того, как это может быть, что корни полинома найти нельзя, а условия того, что они лежат на левой полуплоскости или внутри единичного круга – можно. Выводы критериев Рауса, Гурвица, Найквиста, Михайлова и других, известные в середине 1940-х годов, ответа на этот естественный вопрос не давали. Вместе с тем ответ на него прост и сразу вознаграждается [20, 27]. Дело в том, что хотя корни действительно найти нельзя, легко установить, когда они выходят за пределы левой полуплоскости или любой заданной области G комплексной плоскости (заметим, что сами по себе они исчезнуть или возникнуть при изменении коэффициентов не могут, а могут только непрерывно перемещаться по плоскости

комплексного переменного z). Для того чтобы установить, когда это имеет место, достаточно располагать параметрическим уравнением границы области G в виде $z = f(\omega)$, $\omega \in \Omega$. Имеется в виду, что z пробегает границу области G , когда действительное переменное ω пробегает множество Ω . При этом корень характеристического уравнения $\chi(z) = 0$ при его изменении пересекает границу области G тогда и только тогда, когда при некотором $\omega \in \Omega$

$$\chi(f(\omega)) = 0. \quad (42.1)$$

Пусть теперь система, у которой надлежит исследовать устойчивость равновесия по параметрам u , имеет характеристическое уравнение

$$\chi(z, u) = 0, \quad (42.2)$$

а устойчивость означает, что все корни этого уравнения лежат в области G комплексной плоскости Z . Далее, пусть точка z принадлежит границе области G тогда и только тогда, когда при некотором $\omega \in \Omega$

$$z = f(\omega). \quad (42.3)$$

Из этого следует, что граница области устойчивости состоит из точек, для которых

$$\chi(f(\omega), u) = 0, \quad \omega \in \Omega. \quad (42.4)$$

Множество точек N_ω , удовлетворяющих условию (42.4), содержит все граничные точки области устойчивости, но может содержать и другие точки. Это множество N_ω в целом определяет разбиение пространства параметров u на области $D(j)$ с разными числами j корней уравнения (42.2) в области G . Назовем это *D*-разбиением. В нем нас интересует только область устойчивости $D(0)$. Оказывается, что для отыскания области $D(0)$ полезно рассматривать все *D*-разбиение в целом. В этом и состоит новая идея исследования устойчивости. В ее основе лежит модель, определяемая условием (42.4), а моделью устойчивости является принадлежность точки области $D(0)$. Описанная модель не предполагает характеристическое уравнение (42.2) полиномом по z . Это может быть любая функция, и область G может быть не только левой полуплоскостью, а рассматриваемая система может описываться обыкновенными дифференциальными уравнениями. Ее возможности значительно шире.

Теперь последовательно используем эту математическую модель устойчивости для:

- 1) построения области устойчивости по комплексному параметру w ;
- 2) обнаружения, что область устойчивости для полинома $z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n$ по параметрам a_1, a_2, \dots, a_n односвязна и выпукла по каждому из них;
- 3) обоснования и обобщения известного критерия Найквиста;
- 4) построения области устойчивости по двум действительным параметрам для ультрацентрифуги;
- 5) решения проблемы о мере робастной устойчивости.

Этим перечислением не ограничиваются полезные использования описанной модели устойчивости [31]. Более того, понятие устойчивости может быть заменено другим с подобными общими свойствами.

1. Пусть ради простоты комплексный параметр w входит в характеристическое уравнение линейно, так, что оно может быть записано в виде

$$f(z) + wg(z) = 0. \quad (42.5)$$

Граница области устойчивости $D(0)$ входит в границу D -разбиения, которую обозначим N_ω . Она является отображением границы области $\operatorname{Re} z \leq 0$, т.е. мнимой оси плоскости Z , для точек которой $z = i\omega$, $-\infty < \omega < \infty$, и, следовательно, на плоскости W имеет место уравнение

$$w = -\frac{f(i\omega)}{g(i\omega)}, \quad -\infty < \omega < \infty, \quad (42.6)$$

при этом левая сторона мнимой оси отображается в левую же сторону кривой N_ω , определяемой уравнением (42.6). Заштрихуем эту левую сторону, считая, что кривая N_ω пробегается в направлении возрастания параметра ω , тогда при переходе на комплексной плоскости W со штрихованной стороны кривой N_ω на нештрихованную один корень уравнения (42.5) становится чисто мнимым и переходит с левой полуплоскости $\operatorname{Re} z_v < 0$ на правую. Кривая N_ω разбивает плоскость W на какие-то области, среди которых могут быть и области устойчивости. Если известно число s корней слева от мнимой оси в какой-нибудь точке w^* , то во всех точках области, к которой она принадлежит, число корней s одно и то же. Более того, если из точки w^* мы перемещаемся по кривой γ в любую другую точку \bar{w}^* , то в ней число s уменьшается на число пересечений кривой N_ω со стороны штриховки и увеличивается на число пересечений ее против штриховки. Области устойчивости $D(0)$ отвечает $s = 0$. Остальные области с $s > 0$ обозначим через $D(s)$. С помощью этих обозначений сказанное иллюстрируется на рис. 42.1, где стрелка на кривой N_ω указывает направление ее пробегания, согласно (42.4), при росте ω , а пересекающая ее стрелка – переход из одной области $D(s)$ в другую, $D(s + 1)$. При этом s возрастает на единицу, при обратном переходе s убывает на единицу. Вывод из сказанного состоит в том, что для отыскания области устойчивости $D(0)$ по комплексному параметру w достаточно построить кривую N_ω и заштриховать ее левую сторону и еще знать число корней уравнения (42.5) справа от мнимой оси для какого-нибудь значения $w = w^*$.

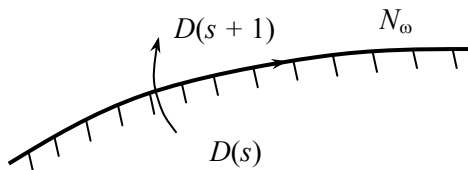


Рис. 42.1. Иллюстрация к правилу штриховки кривой N_ω D -разбиения

2. В качестве примера использования сказанного в п. 1 построим область устойчивости по параметру w для характеристического уравнения вида:

$$z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n - wz^m = 0, \quad 0 \leq m \leq n, \quad (42.7)$$

где a_1, a_2, \dots, a_n – заданные действительные числа, для которых уравнение (42.7) при $w = 0$ имеет все корни слева от мнимой оси, т.е. точка $w = 0$ принадлежит области устойчивости $D(0)$. Кривая N_ω описывается уравнением

$$w = \frac{1}{(i\omega)^m} (z^n + a_1 z^{n-1} + \dots + a_n)_{z=i\omega} = \frac{1}{(i\omega)^m} (i\omega - z_1) \cdots (i\omega - z_n), \quad (42.8)$$

где z_1, z_2, \dots, z_n – корни уравнения (42.7) при $w = 0$. Все они лежат слева от мнимой оси. Из (42.8) непосредственно видно, что при изменении ω от $-\infty$ до ∞ $\arg w$ возрастает на величину πn , а $|w|$ все время отличен от нуля и при $\omega = 0$ обращается в бесконечность. Левая сторона и штриховка кривой N_ω при этом все время обращена в сторону точки $w = 0$, и всякий луч, выходящий из точки $w = 0$, пересекает кривую N_ω только со стороны штриховки. Поэтому областью устойчивости по действительному параметру $\operatorname{Re} w$ или параметру $a_m + \operatorname{Re} w$ является единственный отрезок, содержащий точку $w = 0$. Из этого следует, что область устойчивости $D(0)$ по параметрам a_1, a_2, \dots, a_n в пространстве этих параметров выпукла по каждому из этих параметров, т.е. пересекается с любой прямой, параллельной любой из осей Oa_s этого пространства по единственному отрезку, либо вообще с ней не пересекается.

3. Знаменитый критерий Найквиста утверждает, что если разомкнутое линейное звено с входом x и выходом y устойчиво, т.е. если с некоторого момента t_0 $x = 0$, то выход $y \rightarrow 0$ при $t \rightarrow \infty$, то замкнутая система, получаемая из разомкнутой, тоже устойчива, если точка $w = 1$ комплексной плоскости W лежит вне кривой

$$w = K(i\omega), \quad (42.9)$$

амплитудно-фазовой частотной характеристики разомкнутого звена. Этот критерий оказался весьма удобным и применимым во многих случаях, но и несколько загадочным, поскольку не имел убедительного и полного обоснования. С позиций, изложенных выше, критерий Найквиста и его обобщения непосредственно следуют из рассмотрения области устойчивости по комплексному параметру w для характеристического уравнения замкнутой системы, которое может быть записано в виде

$$P(z) - wQ(z) = 0, \quad (42.10)$$

где $P(z)$ и $Q(z)$ – некоторые, не имеющие общих нулей, целые функции, определяемые тем, что

$$K(z) = \frac{P(z)}{Q(z)}.$$

Уравнение (42.10) при $w = 1$ является характеристическим для замкнутой системы, а при $w = \infty$ – для разомкнутой. Граница области устойчивости N_ω для (42.10) определяется уравнением (42.9), а точка $w = \infty$ принадлежит области устойчивости, и поэтому, если точка $w = 1$ не отделена от точки $w = \infty$ кривой N_ω , то замкнутая система устойчива. Это и есть критерий Найквиста. Из приведенного краткого рассуждения непосредственно следуют обобщения критерия Найквиста на случай, когда точка $w = 1$ не охватывается или охватывается кривой N_ω (42.9), но так, что, выходя из нее и двигаясь к точке $w = \infty$, числа пересечений этой кривой с разных сторон штриховки и нештриховки равны. Очевидно и обобщение на случай, когда разомкнутое звено имеет известное число степеней неустойчивости, т.е. известно число s корней справа от мнимой оси у уравнения (42.10) при $w = \infty$.

4. Ультрацентрифуга – это очень быстро вращающееся тело, которое не может быть жестко закреплено, потому что малейший дисбаланс приводит к громадным силам инерции. Закрепление должно быть мягким, самоцентрирующимся, но тогда оно не обеспечивает устойчивого вращения центрифуги. Эту устойчивость можно обеспечить введением специальных демпфирующих устройств. Как твердое тело, центрифуга имеет 4 степени свободы, две поступательных и две вращательных и еще две степени свободы демпфера, поэтому дифференциальные уравнения ее движения 12 порядка. Такова же степень характеристического уравнения, от корней которого зависит устойчивость. Устранение неустойчивости надлежит реализовать за счет подбора двух параметров демпфера – k и h . Двенадцатый порядок действительного характеристического уравнения ввиду наличия круговой симметрии может быть уменьшен до шести, но при этом характеристический полином будет уже комплексным. Это комплексное характеристическое уравнение записывается в виде:

$$\chi(z) = (k + \nu z^2 + \varepsilon + hz)(P_4(z) - iP_3(z)) + P_2(z) = 0,$$

где P_3 и P_4 – действительные полиномы степеней 3 и 4 соответственно.

Граница области устойчивости $D(0)$ на плоскости параметров k , h определяется, как и ранее, параметрическими уравнениями

$$(k - \omega^2 + \varepsilon + i h \omega)(P_4(i\omega) - iP_3(i\omega)) + P_2(i\omega) = 0,$$

которые могут быть разрешены относительно k и h и записаны в виде:

$$k = f(\omega), \quad h = g(\omega), \quad -\infty < \omega < \infty, \quad (42.11)$$

позволяющем быстро с помощью ЭВМ построить кривую N_ω , нанести на ней штриховку и выделить область устойчивости $D(0)$. Одна из возможных картинок граничной кривой N_ω , определяемой уравнениями (42.11), представлена на рис. 42.2. Из нее следует, что область устойчивости $D(0)$ ограничена двумя участками этой кривой N_ω : верхний отвечает высокочастотному нарушению устойчивости, а нижний – низкочастотному, соответственно большим и малым значениям частоты ω . Из этого же рис. 42.2 видно, что выбор и обеспечение нужных значений параметров довольно деликатен и без этой картинки едва ли возможен. Ведь каждый неудачный выбор – это авария и разрушение центрифуги.

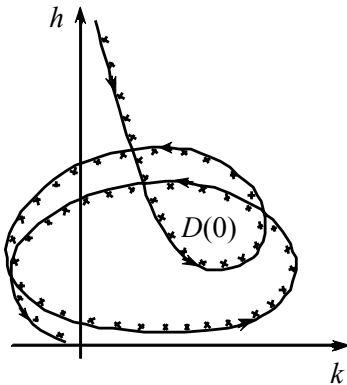


Рис. 42.2. D -разбиение и область устойчивости ультрацентрифуги по параметрам k и h

5. Переходим к последнему пункту: робастной устойчивости и мере робастной устойчивости. Проблемы робастной устойчивости некоторое время бурно об-

суждались и изучались. Робастная устойчивость – это когда устойчивость сохраняется при достаточно больших изменениях параметров. Как следует из предыдущего, эта проблема полностью решается, если речь идет об одном комплексном параметре или двух действительных, когда построение области устойчивости $D(0)$ не вызывает больших затруднений. Но совсем иначе обстоит дело, когда параметров много, и наглядно представить область устойчивости очень трудно, если вообще возможно. В этой ситуации, естественно, возникает вопрос, когда область простого вида лежит в области устойчивости (простые области – это шар, эллипсоид, параллелепипед и др.), и более того, как сделать эту область возможно большей. Эти вопросы привели к следующей общей математической задаче: найти максимальное $\rho = \rho^*$, при котором все функции комплексного переменного z вида

$$\chi(z) = a_1\varphi_1(z) + \dots + a_n\varphi_n(z), \tag{42.12}$$

где

$$\rho = (k_1^m |a_1 - a_1^*|^m + \dots + k_n^m |a_n - a_n^*|^m)^{1/m} \leq \rho^*, \tag{42.13}$$

имеют корни только в заданной области G комплексной плоскости Z . Здесь $\varphi_1(z), \dots, \varphi_n(z)$ – заданные функции z , a_1^*, \dots, a_n^* – некоторые номинальные значения действительных параметров a_1, \dots, a_n , при которых все корни функции (42.12) лежат в области G , k_1, \dots, k_n – произвольные положительные числа и $m \geq 1$. Заметим, что при $m = 2$ и $m = \infty$ неравенство (42.13) определяет простые области – эллипсоид и параллелепипед.

Согласно принятой нами модели устойчивости, найти меру робастной устойчивости ρ^* сравнительно просто: она равна минимуму величины ρ , определяемой (42.13), по всем параметрам a_1, \dots, a_n , лежащим на границе области устойчивости, т.е. удовлетворяющим при каком-нибудь $\omega \in \Omega$ условию

$$\chi(z(\omega), a_1, \dots, a_n) = 0,$$

где $z(\omega)$ – любая точка границы области G при подходящем $\omega \in \Omega$. Не будем решать эту задачу условной оптимизации, это не имеет отношения к основной теме книги. Отметим, что ее решение возможно и не очень сложно [29]. Так, например, для полинома или квазиполинома, записанного в виде

$$\chi(z) = \sum (a_{j2s} z^s \operatorname{sh}(\tau_j z) + a_{j2s+1} z^s \operatorname{ch}(\tau_j z)),$$

когда область G – левая полуплоскость, и полинома вида

$$\chi(z) = \sum a_{2s} (z^{n-s} + z^{-n+s}) + a_{2s+1} (z^{n-s} - z^{-n+s}),$$

когда область G – единичный круг,

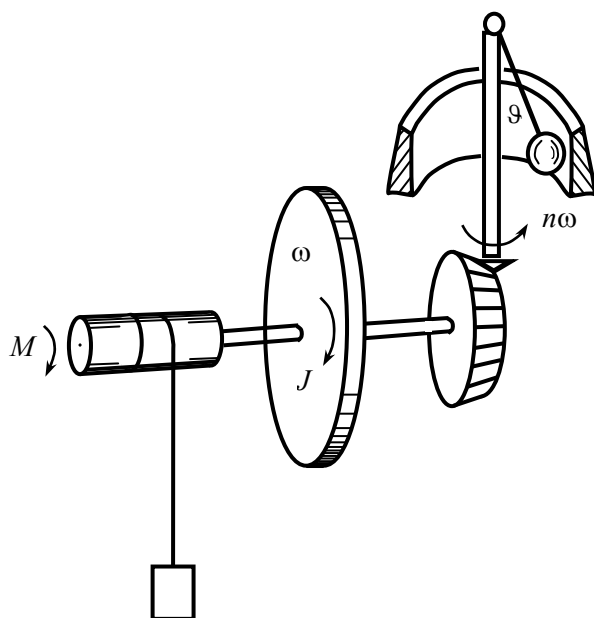
$$\rho^* = \min_{\omega \in \Omega} \left\{ \frac{|\chi_1|^m}{(\Phi_1^2(\gamma))^{m+1}} + \frac{|\chi_2|^m}{(\Phi_2^2(\gamma))^{m+1}} \right\},$$

где Φ_k – векторы с компонентами $(k_j \varphi_{kj})^{r/2}$, $\gamma = m/(m-1)$, $\varphi_{kj} = \partial \chi_k / \partial \alpha_j$.

43. ИДЕАЛИЗАЦИЯ, МАТЕМАТИЧЕСКАЯ КОРРЕКТНОСТЬ И РЕАЛЬНОСТЬ

Эта тема или проблема необычайно широка и многообразна. С ней сталкивается всякий, строящий и исследующий новую или частично новую модель, не имеющую испытанных аналогов или прототипов. Последующее ограничивается двумя конкретными примерами: центробежный фрикционный регулятор скорости вращения и известная проблема парадокса Пенлеве. Их объединяет то, что в обоих случаях речь идет об идеализации реальных упругих тел и податливых связей идеальными абсолютно твердыми и жесткими.

Фрикционный регулятор скорости вращения. На рис. 43.1 схематически изображен очень простой фрикционный регулятор скорости вращения. По идее



своего устройства он должен стабилизировать скорость вращения благодаря тому, что с ее ростом возрастает сила трения об ограничительное кольцо, а при ее убывании сила трения уменьшается.

Центробежная сила, прижимающая шарик к кольцу, равна $(n\omega)^2 r$, и поэтому в весьма упрощенных предположениях

$$J\dot{\omega} = M - k(n\omega)^2 r \cdot rn, \quad (43.1)$$

где введенные обозначения ясны из рисунка.

Из (43.1) непосредственно находим, что скорость вращения имеет устойчивое равновесное значение

$$\omega^* = \left(\frac{M}{k r^2 n^3} \right)^{1/2}. \quad (43.2)$$

Рис. 43.1. Схема фрикционного регулятора скорости вращения

Из (43.2) следует, что дифференциальная неравномерность равновесной скорости вращения ω^* по моменту M равна

$$\frac{d\omega^*}{dM} = \frac{1}{2\omega^* k r^2 n^3} \quad (43.3)$$

и уменьшается с ростом коэффициента передачи шестеренок n . Описанный регулятор, в частности, использовался при астрономических наблюдениях для компенсации вращения Земли, и при этом величина момента M зависела еще и от момента, необходимого для вращения астрономических приборов: рефракторов, рефлекторов и др. Как видно из (43.3), требуемая точность вращения могла быть

обеспечена за счет выбора достаточно большого n . Но при этом обнаружилось, что центробежный регулятор может быть неустойчив и не выполняет своего назначения. Сегодня причину неустойчивости регулятора было бы естественно искать в падающей характеристике трения, т.е. в том, что коэффициент трения зависит от скорости вращения, и что $dk/d\omega < 0$ при скорости $\omega = \omega^*$. Этот механизм неустойчивости рассматривался ранее. Конечно, так может быть, но неустойчивость, как оказалось, возможна и при $dk/d\omega \geq 0$. Для того, чтобы понять ее происхождение и устранить ее, необходимо учесть реальную податливость кольца, изменение r с изменением давления шарика регулятора на ограничительное кольцо. Это удивительно и интуитивно не чувствуется: причем здесь податливость кольца; кажется, что можно считать его абсолютно жестким. Но это не так, и выясняется только после учета податливости.

Учтем неабсолютную жесткость ограничения, налагаемого кольцом на движение шарика центробежного регулятора, зависимость момента сухого трения от скорости и силу тяжести. В качестве обобщенных координат примем угол поворота $n\phi$ вертикальной штанги центробежного регулятора и угол ϑ отклонения шарика от нее. Функция Лагранжа без учета сил трения, момента $M(\phi)$ и давления кольца на шарик, но с учетом силы тяжести, записывается в виде:

$$L = \frac{1}{2} \{ J_0 \dot{\phi}^2 + J_1(\vartheta) n^2 \dot{\phi}^2 \} + \frac{1}{2} B \dot{\vartheta}^2 - V(\vartheta), \tag{43.4}$$

дополнительные обобщенные силы равны

$$Q_\phi = M(\phi) - k(n\phi) N(\vartheta) l \sin \vartheta, \\ Q_\vartheta = -l N(\vartheta),$$

где $N(\vartheta)$ – силы нормального давления кольца на шарик, остальные обозначения очевидны.

Теперь непосредственно записываем нелинейные уравнения движения Лагранжа в виде:

$$\{ J_0 + n^2 J_1(\vartheta) \} \ddot{\phi} + n^2 J_1'(\vartheta) \dot{\vartheta} \dot{\phi} = M(\phi) - k(n\phi) n l \sin \vartheta N(\vartheta), \\ B \ddot{\vartheta} - \frac{1}{2} n^2 J_1'(\vartheta) \dot{\phi}^2 + V'(\vartheta) = -l N(\vartheta). \tag{43.5}$$

Из них интересующее нас равномерное вращение $\dot{\phi} = \omega^*$, $\vartheta = \vartheta^*$ находится из уравнений:

$$M(\omega^*) - k(n\omega^*) n l \sin \vartheta^* N(\vartheta^*) = 0, \\ -\frac{1}{2} n^2 J_1'(\vartheta^*) \omega^{*2} + V'(\vartheta^*) = -l N(\vartheta^*). \tag{43.6}$$

Пусть его решение ω^* , ϑ^* нам известно. Заметим, что из (43.6) можно найти дифференциальную неравномерность регулирования

$$\frac{d\omega^*}{dm} = \frac{l N' - 1/2 \cdot (n^2 J'' \omega^{*2}) + V''}{n^3 J_1' \omega^* k N' + (l N' - 1/2 \cdot (n^2 J'' \omega^{*2}) + V'') (k' N n^2 l \sin \vartheta^* - M)}.$$

При $N' \rightarrow \infty$ она неограниченно убывает с ростом n . В этом нет ничего нового. Исследование устойчивости равномерного вращения приводит к линеаризованным уравнениям ($\omega = \omega^* + \xi$, $\vartheta = \vartheta^* + \eta$):

$$\begin{aligned} (I_0 + n^2 J_1) \dot{\xi} - M' \xi + J_1' n^2 \omega^* \dot{\eta} + k'(n \omega^*) n^2 N l \sin \vartheta^* \xi &= 0, \\ B \ddot{\eta} - \frac{1}{2} J_1'' \omega^{*2} \eta - J_1 n^2 \omega^* \xi + l N \eta + V'' \eta &= 0 \end{aligned}$$

и совершенно новым и неожиданным условиям устойчивости

$$\begin{aligned} n^2 k' N l \sin \vartheta^* - M' &> 0, \\ (-M' + n^2 k' N \sin \vartheta^*) \left(-\frac{1}{2} n^2 \omega^{*2} J_1'' + l N' + V'' \right) + n^2 \omega^* J_1' n N l \sin \vartheta^* &> 0, \quad (43.7) \\ (-M' + n^2 k' N l \sin \vartheta^*) n J_1' - l N' \sin \vartheta (I_0 + n^2 J) k &> 0, \end{aligned}$$

из которых видно (последнее неравенство), что при $N' \rightarrow \infty$ неизбежно наступает неустойчивость. Таким образом, с одной стороны, учет даже ничтожно малой упругости по углу ϑ в ограничении кольцом приводит к неустойчивости, а с другой стороны, неравенства (43.7) показывают, что если искусственно ввести не очень большую жесткость в контакте шарика с кольцом, то можно добиться устойчивости центробежного регулятора.

Математическая некорректность модели (43.1) состоит в том, что сколь угодно малая податливость кольца ведет к резкому качественному изменению поведения модели. Эта некорректность модели обнаруживается не столь непосредственно, как это имело место ранее, например, для модели Вольтерра «хищник–жертва». Теперь некорректность связана с неочевидным увеличением размерности модели, переходом от дифференциальных уравнений первого порядка к третьему. В случае модели «хищник–жертва» эта некорректность обнаруживалась в рамках исходной модели при малых изменениях правых частей дифференциальных уравнений второго порядка. Вместе с тем, то, что фактически происходит с фрикционным регулятором при его неустойчивости из-за слишком большой жесткости контакта с ограничивающим кольцом, дает некоторую подсказку, что же еще следует учесть. Нарушение устойчивости ведет к прыганию шарика вдоль кольца, прыганию, при котором шарик от него отделяется, и это подсказывает, что нужно учесть еще одну степень его свободы по углу ϑ , что и было сделано.

Парадокс Пенлеве и автоколебания при кулоновском трении. В самом конце XIX века французский механик П. Пенлеве, читая парижским студентам учебный курс о трении и стремясь к общей теории, обнаружил, что для некоторых простых механических систем с трением уравнения движения, составляемые по общим правилам механики, неразрешимы. Он опубликовал книгу о трении [42], где привел и эти удивительные примеры, что через десять лет вызвало бурную дискуссию с участием выдающихся ученых того времени: Р. Мизеса, Л. Прандтля, Ф. Клейна, Г. Гамеля, Л. Лекорню и др. и, конечно, самого П. Пенлеве. Обнаруженный П. Пенлеве казус получил название парадокса Пенлеве. Он и сегодня, спустя сто лет, не имеет решения, позволяющего с уверенностью сказать, как поведут себя очень простые механические системы с кулоновским трением. Невозможность точного предсказания вызвана тем, что в системах с парадоксом Пенле-

ве обнаруживается существенная зависимость их поведения от весьма непривычных в таких задачах малых и трудно определяемых параметров и закономерностей [30, 35].

Последующее имеет целью ознакомление с существом парадокса Пенлеве и обнаружение того, что он влечет неожиданную и неизвестную ранее возможность возникновения неустойчивости и автоколебаний, несмотря на то, что характеристика трения не имеет падающего участка. Возникающие колебания в естественном предельном описании являются разрывными или, по существующей математической терминологии, контрастными структурами со скачками и всплесками, которые с физической точки зрения можно назвать «внутренними» ударами. Разрывные колебания в виде периодических контрастных структур возникают потому, что математическая модель систем с парадоксом Пенлеве является сингулярно возмущенной системой дифференциальных уравнений.

На рис. 43.2 приведены некоторые известные примеры механических систем с парадоксом Пенлеве: система Пенлеве–Клейна о движении стержня между параллельными направляющими L_1 и L_2 , из которых одна (L_1) гладкая, а другая (L_2) шероховатая, под действием постоянной силы F_0 – пример *a*); тормозная колодка, прижимаемая с силой F_0 к колесу, вращаемому моментом M – пример *б*); круг (цилиндр), вращаемый моментом M и вжимаемый в угол силой F_0 , одна из сторон которого шероховатая, а другая гладкая – пример *в*).

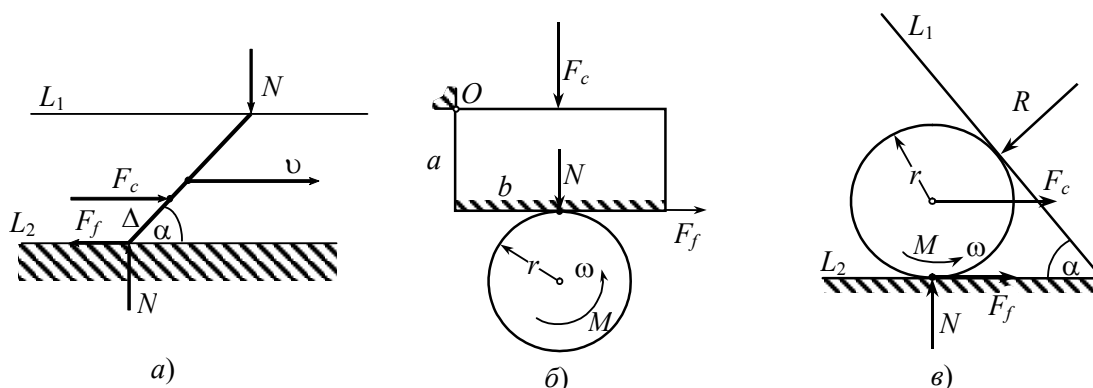


Рис. 43.2. Некоторые известные примеры механических систем с парадоксом Пенлеве

Уравнения движения во всех трех случаях приводятся к одному и тому же виду:

$$\dot{V} = 1 - F, \quad F = |N| \operatorname{sign} v, \quad \chi N - \lambda + F = 0, \quad (43.1)$$

где v – линейная либо угловая скорость, F и N – силы трения и нормального давления, χ и λ – параметры ($\chi > 0, -1 \leq \lambda \leq 1$) [30].

Как динамическая система, уравнения (43.1) содержат одну фазовую переменную v . При заданном начальном условии $v = v_0$ для получения их решения $v(t), F(t), N(t)$ необходимо из двух последних уравнений (43.1) найти F . Но именно это при $\chi < 1$ невозможно: решения может не быть или их оказывается два. В случае $\chi < 1$ уравнения

$$F = |N| \operatorname{sign} v, \quad \chi N - \lambda + F = 0 \quad (43.2)$$

при $\text{sign } v = \text{sign } \lambda$ имеют два решения, а в противном случае его нет. Это непосредственно видно, если на плоскости переменных N и F , рис.43.3, нанести графики функций, отвечающие уравнениям (43.2).

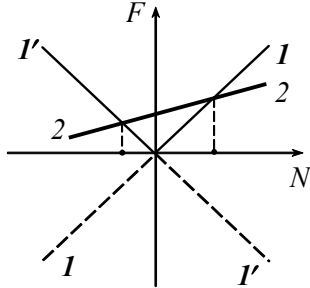


Рис. 43.3. Иллюстрация к парадоксу Пенлеве

Здесь ломаные линии I и I' отвечают первому из уравнений (43.2) соответственно при $v > 0$ и $v < 0$, а прямая 2 соответствует второму уравнению в (43.2) и предполагает $0 < \chi < 1, \lambda > 0$. Согласно рис. 43.3, в этом случае при $v > 0$ решений два, а при $v < 0$ их нет. При $0 < \chi < 1$ и $\lambda < 0$ ситуация аналогичная. В этой невозможности найти нужное единственное решение системы (43.2) и состоит парадокс Пенлеве. Он указывает на то, что классическое описание (43.1) при $0 < \chi < 1$ дефектно.

Уже на упомянутой дискуссии было высказано мнение, что необходимо учесть реальную упругость твердых тел. Различные способы такого учета применительно к задачам Пенлеве–Клейна (см. рис. 43.2,а) делались во многих работах, они действительно избавляли от парадокса, но приводили к разным результатам. В каком-то смысле они делались вслепую, без понимания истинной причины парадокса Пенлеве, хотя тот или иной способ учета упругости приводил к устранению парадокса Пенлеве. А причина была, и ее выяснение требовало нового понимания силового взаимодействия в случае сухого трения, более полной трактовки закона Ньютона о действии и противодействии, требовало обнаружения направленности действия, понимания, что в законе Кулона нормальная сила N , вызывающая силу трения F , имеет направленный характер от N к F и реализуется не за счет силы N , а некоторого стороннего источника энергии. Понимание этого давало новую трактовку соотношениям (43.2) как замкнутого цикла направленных воздействий от N к F (первое соотношение в (43.2)) и от F к N (второе соотношение в (43.2)). Этот направленный цикл взаимодействий изображен на рис. 43.4.

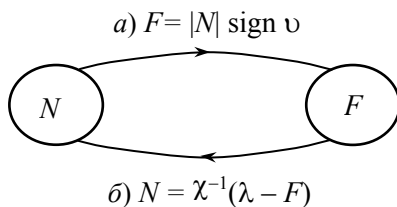


Рис. 43.4. Цикл мгновенных направленных связей в системе с парадоксом Пенлеве

При $\chi < 1$ он влечет парадокс Пенлеве, потому что в этом цикле взаимодействия мгновенные и его коэффициент усиления, равный $1/\chi$, при $\chi < 1$ больше единицы. В реальной же системе парадоксальности нет при всех $\chi > 0$, потому что связи от N к F и от F к N не мгновенные в силу фактической вязкоупругой податливости твердых тел и связей между ними. Отметим, что этот же замкнутый цикл,

вызывающий парадокс Пенлеве, используется в технике в так называемых системах самоторможения, когда тормозят не дополнительными силовыми воздействиями, а за счет направленных связей этого цикла: сила трения вызывает нормальное давление, оно увеличивает силу трения, в результате – быстрое само по себе возникающее торможение и остановка. Торможение мгновенное, подобно удару, если тормозится конструкция абсолютно жесткая.

Такое понимание парадокса Пенлеве позволяет предложить простейший способ учета вязкоупругих инерционных податливостей реальных твердых тел и связей между ними в виде временных релаксационных задержек, записав уравнения (43.1) в виде:

$$\dot{v} = 1 - F, \quad \tau_1 \dot{F} + F = |N| \operatorname{sign} v, \quad \tau_2 \dot{N} + N = \chi^{-1}(\lambda - F), \quad (43.3)$$

где $\tau_1 > 0$ и $\tau_2 > 0$ – малые временные задержки.

Введение этих задержек и запись сингулярных уравнений (43.1) в виде сингулярно возмущенных дифференциальных уравнений (43.3) устраняет парадокс Пенлеве, но теперь вместо системы (43.1) первого порядка появляется система третьего порядка, и ее изучение далеко не просто. Можно исследование упростить, предполагая τ_1 и τ_2 стремящимися к нулю, поскольку они на самом деле малы, но при этом, как оказывается, результат зависит от соотношения τ_1/τ_2 , и, если его не фиксировать, то вообще никакого предельного поведения у системы (43.3) нет.

Конечно, модель (43.3) не точна, она феноменологическая, и реальный характер задержек в некоторых случаях может быть сложнее, но все же это, пусть и грубое, но уточнение парадоксальных уравнений (43.1), и оно позволяет понять, что же может происходить в системе с парадоксом Пенлеве. Отметим, что некоторые модельные способы учета вязкоупругих свойств твердых тел приводят к похожим уравнениям и результатам. Но перед этими конкретными способами уточнения уравнений (43.1) у уравнений (43.3) есть определенное преимущество: уравнения (43.3) более общие и содержат два существенных малых параметра τ_1 и τ_2 , а модельные конструкции учитывают только один из них. Это, в частности, объясняет, почему конкретные уточнения приводят к разным результатам.

Перейдем к исследованию сингулярно возмущенных уравнений (43.3), точнее, к некоторым фрагментам этого исследования, позволяющим составить представление о возможных, порождаемых ими, движениях и явлениях. Начнем с рассмотрения крайних частных случаев уравнений (43.3): $\tau_1 = 0$ и $\tau_2 = 0$. При этом порядок уравнений (43.3) уменьшится. Исключая переменную F , запишем их для возможности сравнения в одних и тех же фазовых переменных V и N :

$$\dot{v} = 1 - \lambda + \chi N, \quad \tau \dot{N} = -N [1 + \chi^{-1} \operatorname{sign}(v N)] + \chi^{-1} \lambda \quad (43.4)$$

и соответственно

$$\dot{v} = 1 - |N| \operatorname{sign} v, \quad \tau \dot{N} = -N [1 + \chi^{-1} \operatorname{sign}(v N)] + \chi^{-1} \lambda. \quad (43.5)$$

При $\tau \rightarrow 0$ обе модели (43.4) и (43.5) переходят в парадоксальную модель (43.1), отвечающую абсолютной жесткости твердых тел. Они не только содержат одни и те же фазовые переменные v и N , но и зависят от одних и тех же параметров χ и λ и испытывают бифуркации на одних и тех же линиях, и в силу этого

имеют одинаковые разбиения плоскости параметров на области с различными фазовыми портретами, которые не всегда одинаковые. Плоскость параметров разбивается на шесть областей: A, B, C, D, E и G (рис. 43.5).

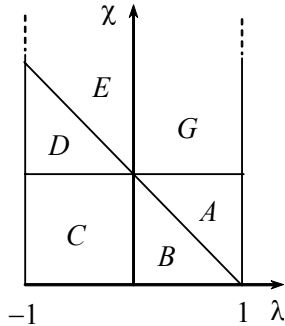


Рис. 43.5. Шесть областей в плоскости параметров

Первые три соответствуют парадоксальности ($\chi < 1$), остальные не парадоксальны. Так, в области A фазовые портреты для моделей (43.4) и (43.5) показаны на рис. 43.6 и 43.7.

В обоих случаях финальным является равноускоренный разгон Γ_+ . В первом случае при $\tau \rightarrow 0$ финальному движению могут предшествовать один или два удара, сопровождающиеся неограниченным значением силы N и конечным скачком скорости v . При этом коэффициент скачка скорости v равен $\mu = (\chi - 1)(\chi + 1)^{-1}$. Во втором случае возможны мгновенные остановки, т.е. удары с коэффициентом $\mu = 0$. Наглядные представления об ударах можно получить, изобразив соответствующие рис. 43.6,а и 43.7,а предельные фазовые портреты при $\tau \rightarrow 0$ (рис. 43.6,б и 43.7,б).

На рис. 43.6,б точка M мгновенно переходит в точку M' , M' в M'' , ... , и, наконец, в \bar{M} , после чего наступает равноускоренный разгон. На рис. 43.7,б аналогично точка M переходит в M' , M' в \bar{M} , а точка K переходит в K' , K' в $\bar{K} = \bar{M}$. Из послеударной точки $\bar{M} = \bar{K}$ начинается равноускоренный разгон с нулевой начальной скорости.

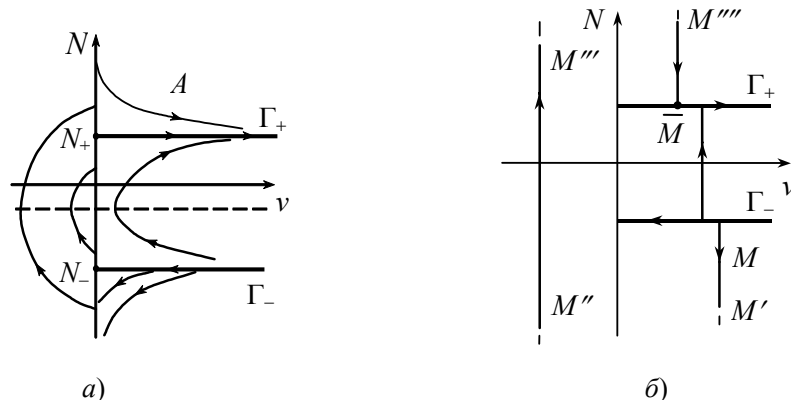


Рис. 43.6. Фазовые портреты для модели (43.3) в области A

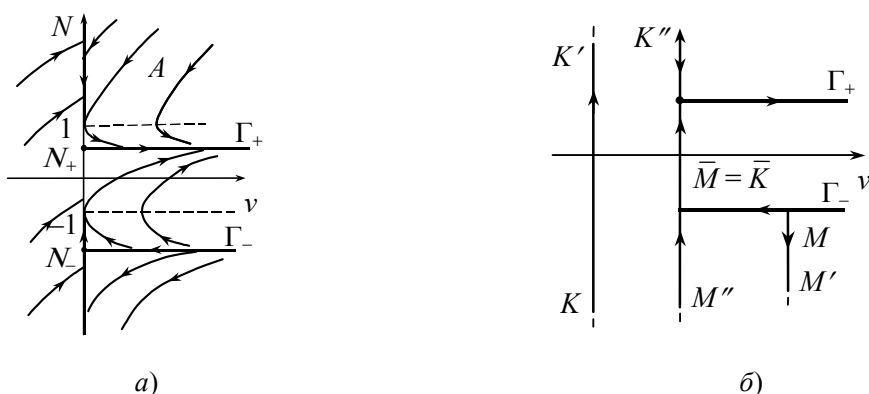


Рис. 43.7. Фазовые портреты для модели (43.5) в области A

В области B фазовые портреты для моделей (43.4) и (43.5) изображены соответственно на рис. 43.8 и 43.9.

На каждом из них по два установившихся предельных движения: устойчивое состояние равновесия O и равноускоренный разгон Γ_+ . Между устойчивыми состояниями равновесия O есть существенное различие: в первом случае область притяжения (см. рис. 43.8) при $\tau \rightarrow 0$ неограниченно уменьшается и исчезает, во втором (см. рис. 43.9.) – сохраняется (она заштрихована). При $\tau \rightarrow 0$ возможны удары с указанными ранее величинами коэффициента скачка скорости v и бесконечным значением N . Удары и предельные движения в виде устойчивого равновесия возможны и в парадоксальной области C . В непарадоксальных областях D, E и G ударов при $\tau \rightarrow 0$ не возникает, и предельные установившиеся движения одинаковые и единственные. Подчеркнем, что в парадоксальном случае B имела место неединственность предельного движения, и оно могло быть разным при $\tau \rightarrow 0$ в разных моделях (43.4) и (43.5).

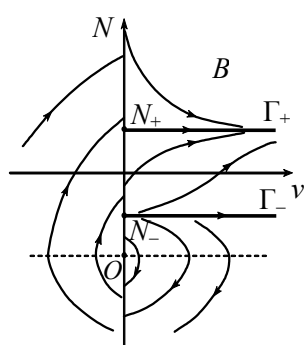


Рис. 43.8. Фазовые портреты для модели (43.4) в области B

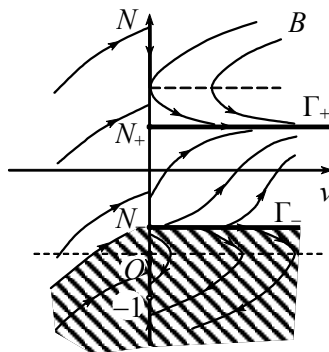


Рис. 43.9. Фазовые портреты для модели (43.5) в области B

Таковы особенности поведения упрощенных моделей систем с парадоксом Пенлеве. Уточнение модели с учетом обоих временных задержек позволяет обнаружить большее разнообразие возможных движений и неожиданную возможность автоколебаний, которые при $\tau \rightarrow 0$ могут принимать вид так называемых контрастных структур или неограниченно увеличивать свою частоту при сохранении конечности амплитуд колебаний сил N и F . Не вдаваясь в детали довольно сложного

и в значительной степени численного компьютерного исследования, кратко опишем основные факты, причем сделаем это для модели (43.3) с учетом еще и вязкого трения, т.е. для модели вида:

$$\dot{V} = 1 - F, \quad \tau_1 \dot{F} + F = |N| \operatorname{sign} V + hV, \quad \tau_2 \dot{N} + N = \chi^{-1}(\lambda - F). \quad (43.6)$$

Хорошо известно, что сухое трение в механических системах может приводить к неустойчивости и автоколебаниям, но для этого нужно, чтобы характеристика трения имела падающий участок. Как оказывается, последнее не обязательно, когда механическая система с сухим трением парадоксальна по Пенлеве. В парадоксальной системе неустойчивость и автоколебания возможны из-за наличия в ней цикла направленных воздействий с подкачкой энергии, и эту его роль не только не может нарушить добавок вязкого трения, но в некотором смысле ее усиливает.

В изложении нового факта состоит цель дальнейшего. При малых, но конечных задержках τ_1 и τ_2 обнаруживается неустойчивость и рождение автоколебаний в соответствии с бифуркацией Андронова. При τ_1 и τ_2 , стремящихся к нулю, автоколебания принимают вид контрастных структур, т.е. сопровождаются разрывными скачками скорости V («внутренние» удары), при которых силы F и N обращаются в бесконечность.

Равновесия системы (43.6) определяются уравнениями:

$$\lambda - F = 0, \quad \chi N - \lambda + F = 0, \quad F = |N| \operatorname{sign} V + hV. \quad (43.7)$$

Из них следует, учитывая, что $\lambda \leq 1$, что равновесие имеется только при $h > 0$, $\chi > 1 - \lambda$, $V > 0$ и для него

$$F^* = 1, \quad N^* = \chi^{-1}(\lambda - 1), \quad V^* = h^{-1}[1 - \chi^{-1}(1 - \lambda)]. \quad (43.8)$$

Согласно (43.6), характеристическое уравнение имеет вид:

$$\tau_1 \tau_2 z^3 + (\tau_1 + \tau_2) z^2 + (1 + \sigma \chi^{-1} + h \tau_2) z + h = 0,$$

где $\sigma = \operatorname{sign} N^* < 0$, и единственное условие устойчивости состоит в выполнении неравенства

$$(\tau_1 + \tau_2)(1 + \sigma \chi^{-1} + h \tau_2) - \tau_1 \tau_2 h > 0,$$

которое при исчезающих малых положительных τ_1 и τ_2 сводится к единственному требованию: $\chi > 1$. Таким образом, единственное равновесие (43.8) существует при $\chi > 1 - \lambda$, оно устойчиво при $\chi > 1$, неустойчиво при $\chi < 1$ и теряет устойчивость через фокус.

Компьютерное исследование показывает, что при потере устойчивости рождается устойчивое периодическое движение, возникают автоколебания. При уменьшении τ_1 и τ_2 этот предельный цикл вытягивается, частично удаляясь в бесконечность, и приобретает вид контрастной структуры со скачками или всплесками (рис. 43.10, а, б). Описанное происходит в области параметров $1 > \chi > 1 - \lambda$, $\lambda > 0$, $h > 0$.

В другой области параметров, $\chi < 1 - \lambda$, $\lambda > 0$, $h > 0$, при τ_1 и τ_2 , стремящихся к нулю, также возникают автоколебания в виде периодических контрастных структур (рис. 43.10, в). Механизм их возникновения отличен от приведенного

выше и состоит в наличии в рассматриваемой сингулярно возмущенной системе (43.6) такой структуры фазового пространства, при которой

- 1) существуют, по крайней мере, два многообразия медленных движений, из которых только одно устойчиво;
- 2) существует область начальных значений, при которых возможен уход фазовой траектории в бесконечность при $\tau_1 \rightarrow 0$, $\tau_2 \rightarrow 0$;
- 3) существует особенность структуры, которая обеспечивает возврат фазовых траекторий из бесконечности;
- 4) существует механизм срыва фазовой точки с устойчивого многообразия.

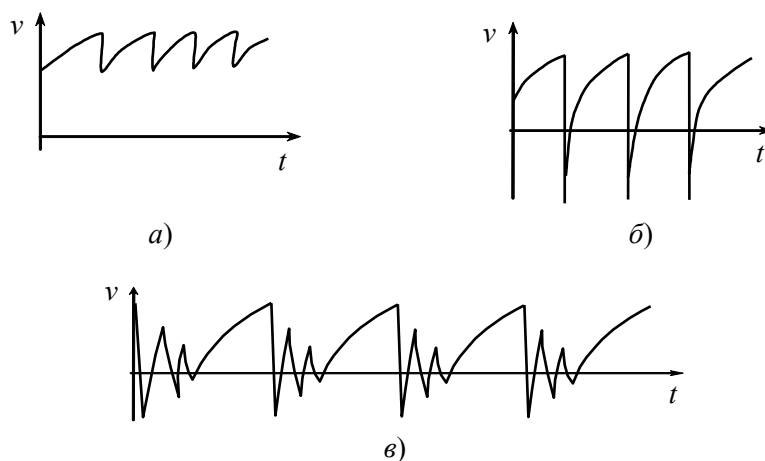


Рис. 43.10. Автоколебания в системе (43.6) при $\lambda > 0$, $h > 0$: а), б) $1 > \chi > 1 - \lambda$, в) $\chi < 1 - \lambda$

В более простом случае, $\tau_2 = 0$, когда система (43.6) становится двумерной, типичный пример возникновения автоколебаний типа контрастных структур в результате выполнения условий 1)–4) приведен на рис. 43.11. Здесь Γ_+ и Γ_- – устойчивое и неустойчивое многообразия медленных движений, определяемые вырожденными, $\tau_1 = \tau_2 = 0$, уравнениями (43.6) при $\chi < 1 - \lambda$, $\lambda > 0$, $h > 0$, КС – периодическое движение типа контрастной структуры со всплесками при $\tau_1 \rightarrow 0$.

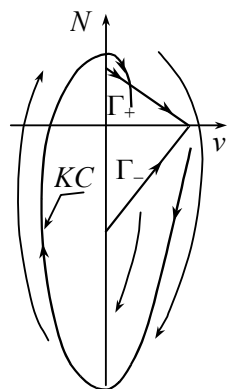


Рис. 43.11. Пример возникновения автоколебаний типа контрастных структур в случае $\tau_2 = 0$

При $h = 0$ периодических контрастных структур при стремлении τ_1 и τ_2 к нулю не возникает, но при конечных τ_1 и τ_2 имеются автоколебания, которые при τ_1 и τ_2 , стремящихся к нулю, неограниченно увеличивают свою частоту. При этом

амплитуда колебаний скорости v стремится к нулю, а амплитуды колебаний сил F и N остаются конечными.

Итак, сухое кулоновское трение само по себе или в сочетании с вязким трением ($h > 0$) при наличии парадокса Пенлеве может вызвать автоколебания, которые в случае $h > 0$ при увеличении жесткости входящих в систему твердых тел и связей между ними переходят в разнообразные периодические контрастные структуры – разрывные периодические автоколебания. Их наличие обусловлено замкнутым циклом направленных связей со сторонним источником энергии и подобно тому, что имеет место в релейных системах автоматического регулирования с реле, имеющим такую же характеристику, как кулоновское трение. Подчеркнем, что средняя скорость автоколебаний всегда положительна, и внешняя сила F_c и момент M совершают на каждом периоде определенную работу. Специфика этих автоколебаний в том, что они могут быть близки к разрывным колебаниям и иметь вид разнообразных контрастных структур со скачками и всплесками. Особенность систем с парадоксом Пенлеве еще и в том, что их поведение существенно зависит от малых параметров и соотношений между ними. В рассматриваемых моделях (43.3) и (43.6) малые параметры – это τ_1 и τ_2 , а соотношение между ними – это отношение τ_1/τ_2 [46].

44. ДИНАМИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ И ГЛОБАЛЬНОЙ ПОИСКОВОЙ ОПТИМИЗАЦИИ С АДАПТИВНОЙ МОДЕЛЮ

Основой для описания эволюционных процессов и явлений в природе и технике, построения их математических моделей и исследования является динамическая система. Все модели, о которых рассказывается в этой книге, – динамические системы.

Последующее повествование имеет целью обратить внимание на возможность, естественность и целесообразность динамической трактовки некоторых вычислительных, поисковых и алгоритмических процедур на примерах известного и широко используемого метода наименьших квадратов (МНК) и глобальной поисковой оптимизации с адаптивной стохастической моделью оптимизируемой функции. Первый пример демонстрирует возрастание возможностей и эффективности приложений МНК в результате его динамического представления, второй – широту понятия и возможностей описания в виде динамической системы.

Универсальная рекуррентная форма метода наименьших квадратов. Метод наименьших квадратов – это метод оценивания неизвестных параметров теоретических моделей по косвенным измерениям. Авторами его считаются К. Гаусс и А. Лежандр. Классический «метод уравнивания Гаусса» был установлен в 1795 году и успешно использован Гауссом в 1801 году для определения траектории астероида Церера. Первое изложение элементов метода и само название «метод наименьших квадратов» даны в 1806 году Лежандром в работе, посвященной вычислению орбит комет. Однако вероятностное обоснование метода в 1809–1810 годах и глубокая разработка вычислительных проблем принадлежат Гауссу. Дальнейшее развитие этот метод получил в работах многих известных математиков: Лапласа, Чебышева, Неймана, Рао и др. Особенно большой вклад внесли А.А. Марков и А.Н. Колмогоров, которым принадлежат строгое обоснование и установление границ содержательной применимости МНК.

Главные достоинства решения задачи оценивания с использованием МНК связаны с ее априорной разрешимостью и такими замечательными свойствами получаемых оценок, как несмещенность, эффективность и состоятельность. Благодаря этому МНК является одним из наиболее известных и многообразно используемых математических методов обработки наблюдений и экспериментальных данных. Он позволяет на основе приемлемых гипотез и математических моделей определять неизвестные параметры и закономерности не только в задачах прямой обработки данных, но и в задачах фильтрации, идентификации, распознавания образов, сжатия описания (кодирования), автокорреляционного анализа и др. Но существуют две трудности применения классического МНК. Первая связана с обращением матрицы, подчас большой размерности, без уверенности в ее корректности. Вторая трудность вызвана тем, что задача анализа данных, осуществляющая преобразование от «исходных данных» к «результату», обычно требу-

ет многочисленных и многократных изменений в выборке данных и в принимаемых гипотезах в отношении изучаемых процессов и явлений. МНК в своей классической форме при любых изменениях используемых данных, принимаемых моделей и гипотез требует полного повторения всех вычислений. Именно поэтому уже в 1821 году Гаусс предложил рекуррентный вариант процедуры, позволяющий корректировать ранее вычисленную оценку с учетом вновь поступивших дополнительных измерений без необходимости повторять все предшествующие вычисления, для случая, когда наблюдение представляется скаляром. В 1950 году Плакетт обобщил эту идею на случай векторной величины. Рекуррентная форма МНК получила очень широкое распространение, особенно в теории идентификации, адаптивного управления и современной теории фильтрации. Дальнейшее расширение рекуррентной формы МНК для случая поиска единственного решения с минимальной нормой путем псевдообращения матрицы данных осуществлено в работах А. Алберта. Универсальная рекуррентная форма МНК предложена в работе [32]. В ней МНК рекуррентен не только в отношении добавляемых и удаляемых данных, но и используемых моделей: базисных функций, отыскиваемых параметров и линейных ограничений на них.

В основе МНК лежит минимизация квадратичного функционала

$$J(a_1, a_2, \dots, a_m) = \sum_{j=1}^N \left(\sum_{i=1}^m a_i \varphi_i(x^j) - b_j \right)^2 \quad (44.1)$$

при линейных ограничениях

$$L_k = \sum_{i=1}^m c_{ki} a_i - d_k = 0 \quad (k = 1, 2, \dots, l) \quad (44.2)$$

по искомым значениям параметров a_1, a_2, \dots, a_m . Здесь $x^1, b_1; x^2, b_2; \dots; x^N, b_N$ – используемые данные. Гипотеза, положенная в основу применения МНК, состоит в том, что при искомым значениях параметров a_1, a_2, \dots, a_m и лишенных погрешностей используемых данных имеют место зависимости, зануляющие функционал J и обращающие все L_k в нуль. Эти предположения, которые могут быть точными или принимаемыми приближениями, определяются выбранными базисными функциями $\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x)$ и линейными ограничениями (44.2) и вместе составляют модель, положенную в основу конкретного применения МНК.

Методами линейной алгебры и оптимизации (метод множителей Лагранжа для условной оптимизации) можно найти в виде явных формул значения параметров, минимизирующие функционал (44.1) при условиях (44.2). Это МНК в своей исходной классической нереккуррентной форме. О трудностях и неудобствах такого прямого его использования сказано выше.

Преобразование, определяемое МНК, от входных данных и управляющих команд исследователя к выходным результатам можно представить схематически, как показано на рис. 44.1. Прямоугольник символизирует вычислительную процедуру МНК; A, B и C – входы для данных, необходимых для вычислений (A – исходные данные, B – базисные функции, C – линейные ограничения), D – выход, дающий значения искомым параметрам и достигнутое значение минимизируемого функционала J . Классическая форма МНК, как она описана выше, предполагает,

что сначала поступают все входные данные, а затем выполняются вычисления и выдаются требуемые выходные данные.

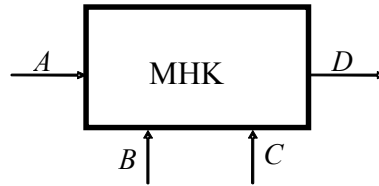


Рис. 44.1. Представление универсального МНК в виде преобразователя с входом и выходом

Универсальная рекуррентная форма МНК состоит в том, что все время на входы A , B и C могут подаваться извне новые данные или исключаться старые, выход D выдает соответствующие им значения искомым параметров и происходит это не путем многократного повторения вычислительной процедуры МНК, а рекуррентно на основе вводимого состояния θ в виде совокупности пяти матриц, в частности, содержащих значения параметров a_s и функционала J . При этом

$$\begin{aligned}\theta^{k+1} &= T(u^{k+1})\theta^k, \\ a^{k+1} &= f\theta^{k+1}.\end{aligned}\quad (44.3)$$

Здесь θ^k и θ^{k+1} – состояния на k -м и следующем $k+1$ -м тактах вычислений, a^{k+1} – вектор искомым значений параметров на $k+1$ шаге, u^{k+1} – вектор входа используемых данных и управляющих воздействий исследователя, поступающих после k -го шага. Вектор u^{k+1} представляет собой одно из трех (шести, объединенных парно), перечисляемых ниже возможных входных воздействий:

- 1) введение новых входных данных x^j , b_j или исключение старых;
- 2) введение новой функции $\varphi_l(x)$ или исключение какой-то прежней с соответствующим введением нового параметра a_i или исключением прежнего;
- 3) добавление нового линейного ограничения или исключение старого.

Реализация оператора управляемой динамической системы (44.3) не требует обращения матриц, а только их сложения и умножения. Вместе с тем описание этого оператора $T(u)$ достаточно громоздко и поэтому не приводится.

Описанная универсальная рекуррентная форма и ее представление в виде управляемой динамической системы придает МНК новые широкие возможности практического использования и необходимую адаптивность к решаемой задаче. В частности, не требуется обращения матриц, обходится проблема плохой обусловленности информационной матрицы, легко исключается ошибка начального приближения, неизбежная при прежней частичной рекуррентности, значительно эффективнее поиск подходящей модели, устранение ошибочных или случайно сильно уклоняющихся данных. Появляется возможность построения приспособленного базиса и сжатия описания и кодирования, текущей обработки потоковых данных и др.

Поисковая глобальная оптимизация с адаптивной моделью [27]. Алгоритм поисковой глобальной оптимизации, минимизации функции $Q(x)$ на множестве X определяется правилами назначения следующей точки:

$$x^n = f(x^1, Q^1, \dots, x^{n-1}, Q^{n-1}), \quad Q^i = Q(x^i), \quad (44.4)$$

и остановки поиска

$$g(x^1, Q^1, \dots, x^n, Q^n) > 0. \quad (44.5)$$

После остановки в качестве «минимума» функции $Q(x)$ на X берется значение

$$Q_{\min} = \min [Q^1, \dots, Q^n] = Q^s, \quad (44.6)$$

в точке $x_{\min} = x^s$. При этом функция $Q(x)$ предполагается заданной в том смысле, что для любого $x \in X$ может быть найдено значение $Q(x)$. Выбор следующей после x^1, x^2, \dots, x^n точки x^{n+1} может опираться на разные соображения. Ниже будем исходить из наибольшей вероятности того, чтобы в новой выбранной точке x^{n+1} вероятность неравенства

$$Q_{\min} - Q(x) > \eta > 0 \quad (44.7)$$

была максимальной. Для возможности такого выбора принимается стохастическая модель функции $Q(x)$, которая строится по уже известным значениям функции $Q(x)$ в точках x^1, x^2, \dots, x^n . Согласно этой модели, вероятность того, что $Q(x) < \bar{Q}$ равна

$$P(Q(x) < \bar{Q} / x^1, Q^1; \dots; x^n, Q^n) = \sum C_s(x) \varphi \left(\frac{\bar{Q} - Q^s}{\gamma_s(x)} \right). \quad (44.8)$$

Функции $\varphi(\xi)$, $C_s(x)$ и $\gamma_s(x)$ выбираются из доступных общих представлений о функции $Q(x)$. В соответствии с ними $\varphi(\xi)$ – неубывающая функция ξ , меняющаяся от 0 до 1 при изменении ξ от $-\infty$ до $+\infty$, $C_s(x)$ – функции x такие, что $C_s(x) \rightarrow 0$ при возрастании расстояния ρ_s точки x от точки x^s и $C_s(x^k) = \delta_{sk}$ (символ Кронекера), $\gamma_s(x)$ – возрастающие функции расстояния ρ_s , например $D_s \rho_s$, где D_s – коэффициент, оценивающий скорость изменения функции от точки x к точке x^s . В качестве конкретных видов функций $\varphi(\xi)$ и $C_s(x)$ можно, например, принять

$$C_s(x) = \rho_s^{-1} \left(\sum_{i=1}^n \rho_i^{-1} \right)^{-1}, \quad \varphi(\xi) = \begin{cases} \frac{1}{2} e^{\xi} & \text{при } \xi \leq 0, \\ 1 - \frac{1}{2} e^{-\xi} & \text{при } \xi \geq 0. \end{cases} \quad (44.9)$$

Подчеркнем, что стохастическая модель (44.8) вобрала в себя не только априорные представления о функции $Q(x)$, но и сведения, полученные о ней после n шагов поиска. С каждым следующим шагом поиска эта модель уточняется. Именно поэтому она названа адаптивной стохастической моделью оптимизируемой функции. Следующая точка x^{n+1} находится как точка максимума вероятности неравенства (44.7), определяемой как максимум (44.8) при $\bar{Q} = Q_{\min} - \eta$. Поэтому описанный метод оптимизации имеет практический смысл лишь, если задача максимизации функции (44.8) много проще такой же задачи максимизации функции $Q(x)$, что может иметь место при значительных трудностях или больших необходимых затратах для вычисления функции $Q(x)$, когда эти вычисления требуется по возможности сократить.

Определение следующей точки поиска на каждом шаге требует задания величины $\eta > 0$ в неравенстве (44.7). Эту величину η можно рассматривать как управляющий параметр поиска. Интуитивно ясно, что в начале поиска надлежит стремиться к возможно более быстрому уточнению модели и лишь после этого отыскивать минимум, что в процессе поиска их следует целесообразно сочетать. В начале поиска можно надеяться на значительное уменьшение значений Q , в дальнейшем они менее вероятны. Из этого следует, что вначале нужно выбирать большие значения η , а затем их уменьшать. Как можно предположить, при больших η происходит уточнение модели, а при малых η небольшие уменьшения минимума. Но все это не более, чем общие эвристические соображения, подкрепляемые экспериментами. Более обоснованный выбор η на каждом шаге можно сделать, если располагать примерным графиком зависимости

$$P(Q(x) < Q_{\min} - \eta)$$

от числа η . Эта вероятность убывает с η , и это убывание может быть медленным или быстрым. При медленном убывании естественно увеличивать η до разумного предела, при очень быстром убывании его следует выбирать меньшим, предпочитая малый, но достижимый результат, несколько большему, но весьма мало вероятному. Мне представляется, что здесь мы сталкиваемся именно с тем случаем, когда природная интуиция человека и его опыт играют решающую роль, и едва ли попытки уточнения теории, во всяком случае в рамках принятой модели и метода поиска, могут быть реально полезны. Но мы уклонились от основной темы, и нам надлежит описанный процесс поиска минимума привести к управляемой динамической системе, т.е. надлежит определить понятие состояния, пространства состояний и указать оператор изменения состояния, который будет зависеть от управляющей переменной η . Определение следующей точки поиска x^{n+1} требует знания всей предшествующей последовательности значений $x^1, Q^1; x^2, Q^2; \dots; x^n, Q^n$ и отыскания максимума по x функции (44.8) при заданном $\bar{Q} = Q_{\min} - \eta$. Напомним, что x_{\min} – это значение x^s , при котором Q^s – максимум из всех известных значений Q^1, Q^2, \dots, Q^n . Тогда в качестве состояния следует принять всю последовательность значений $x^1, Q^1; \dots; x^n, Q^n$. Пространство состояний – это множество всевозможных таких последовательностей для всех $n \geq 1$ с $x^l \in X$ и $Q^l = Q(x^l)$, а оператор определяется тем, что x^{n+1} есть максимум функции (44.8) при $\bar{Q} = Q_{\min} - \eta$, где $\eta > 0$ – задаваемая на каждом шаге управляющая переменная [32].

45. ИГРОВАЯ МОДЕЛЬ ЧЕЛОВЕЧЕСКОГО ОБЩЕСТВА

Модель организации и функционирования современного общества [38, 39]. Математические модели, которые рассматривались до сих пор, – это пока весьма упрощенные, но модели существующих или заведомо осуществимых объектов и процессов. Когда речь шла о моделировании, предполагалось, что это моделирование реальных объектов, но математические модели могут относиться и к несуществующим объектам, воображаемым, гипотетическим и даже фантастическим. В одних случаях осуществимость их не вызывает сомнения, в других – это не так. Моделируя полет на Луну, можно понять, реализуем ли он и что для этого требуется. Но могут быть модели, для которых реализация весьма проблематична, даже если в принципе и возможна. Такая ситуация имеет место при моделировании будущего устройства общества людей или возможных в будущем попыток моделирования новых генетических конструкций живых организмов и человека. Такое, назовем его фантастическим, моделирование тоже имеет право на жизнь и, более того, интересно и может быть полезным.

Игровая модель человеческого общества, пожалуй, не только фантастическая модель, но и непривычная, поскольку по необходимости ее построение опирается не на точные знания и точные науки, а на общие гуманитарные и естественнонаучные соображения и доводы, но следующие из них выводы неожиданны и интересны и согласуются с некоторыми определяющими тенденциями, наблюдаемыми в современном обществе. Даже при математическом изучении общества гуманитарные соображения неизбежны, и роль математики состоит в том, чтобы помочь сделать их более четкими и отделить то, что принимается, от того, что из этого следует, и вывести то, что нужно и интересно.

Дальнейшее представляет собой попытку постижения организации и функционирования современного общества, тенденций и перспектив его эволюции и развития с позиций точной науки: математического моделирования, теории управления и теории колебаний.

Математического моделирования, потому что мы хотим построить математическую модель и на ней основывать дальнейшие выводы. Общество и его деятельность являются очень сложными, и в понимании их мы довольствуемся различными соображениями, и кажется, что точная наука здесь ничем помочь не может, а уж математика и подавно. Но это не так в силу феномена простых математических моделей, которые описывают изучаемый объект в целом, отражая лишь интересующие нас важнейшие его особенности и стороны.

Теории управления потому, что функционирование общества основано на целесообразной синхронизации действий отдельных людей, создаваемой и поддерживаемой не без использования управления.

Теории колебаний потому, что она – всеобъемлющая наука об эволюционных процессах, а функционирование общества – это эволюционный процесс.

Общество и составляющие его люди в своей организации и своем функционировании похожи на живой организм, состоящий из клеток. Сообщество клеток обеспечивает жизнь всего организма, а жизнедеятельность клеток обеспечивается

гомеостазом среды обитания клеток в организме. Нечто аналогичное имеет место и в человеческом обществе. Для его жизни необходимо обеспечить соответствующую среду, назовем ее базовым уровнем общества. По мере совершенствования многоклеточного живого организма, как необходимые условия жизни, помимо гомеостаза среды, возникли еще гуморальные и нервные управления. В современном обществе они имеют аналог в организации жизни и действий людей, осуществляемые государством и общественными организациями. В живом организме все происходит как бы само собой, в силу законов природы, ее тенденций к самоорганизации, еще недостаточно понятых нами. В человеческом обществе, в отличие от дикой природы, многое организуется на сознательном уровне при целенаправленном вмешательстве людей в свою жизнь. Отмеченная аналогия между живым организмом и человеческим обществом очевидна, но столь же, если не более, разительны отличия. Общество по отношению к живому организму находится на следующем уровне эволюции жизни, когда «клетка» становится мыслящей и приобретает возможность прямого воздействия на его организацию и деятельность. Такого результативного, прямого и быстрого воздействия у клеток на организм нет. Это существенное различие влечет гигантские ускорения эволюции жизни, совершенствования и развития общества людей по сравнению с живым организмом клеток.

Над тем, как функционирует человеческое общество, начали думать давно, но только сейчас, как мне представляется, появилась необходимая научная база, и проблема назрела. До каких-то пор смена и формирование общественных, социальных и экономических структур происходили в основном подобно развитию всего живого на Земле – эволюционно биологически, не без проб, ошибок, тупиков и отбора. Вместе с тем, на этот процесс оказывали влияние традиции, верования, общие представления и естественно-научные взгляды. В изучении общества со времен древней Греции наметились два подхода. Их можно обозначить «от человека» и «от государства». Восторжествовал государственный подход как в описании исторических событий, так и текущей жизни общества. В основе всего происходящего и судеб общества ставились государство и власть, их интересы и цели. Лишь в последнее время наметился перелом, и отчасти учитываются интересы всех людей в целом. Все это довольно естественно, поскольку государство и власть выполняли необходимую всему обществу жизнеобеспечивающую функцию защиты от внешней агрессии, порабощения и истребления. Развитие материальной производственной основы перешагнуло эти рамки и приняло характер всеобщей научно-технической и информационной революции, и они в значительной мере определяют структуру современного общества. Представляется, что сегодня нечто подобное возможно и в социально-экономической, и общественной сферах. Если угодно, пришло время признать, что сегодня нет более важного и интересного объекта научного изучения, чем человеческое общество. Изучения не локальных вопросов экономичности, рентабельности, доходности и эффективности предприятий, производства, сельского хозяйства, оборота товаров и денег, что само по себе тоже очень важно, а механизмов функционирования общества и общих принципов его организации и управления.

Глобальная математическая модель и основанные на ней теории функционирования, организации и управления обществом очень нужны и могли бы помочь

человечеству в преодолении угрожающих ему социальных, экологических, экономических кризисов и катастроф, способствовать социальной революции, аналогичной промышленной и информационной. Но, как представляется, изменения не произойдут сами собой, как это имело и имеет место для технической и информационной революций, и есть, по крайней мере, две причины этого. Одна причина – отсутствие непосредственной заинтересованности тех, от кого это зависит, а вторая – недостаточная компетентность тех, от кого это зависит. В сфере материального производства, чтобы преуспеть, необходимо и то, и другое: и заинтересованность, и умение, что совершенно не обязательно для власти.

Заинтересованность, если и может возникнуть, имеет весьма призрачные основания, не следующие из коренных интересов власти, компетентность и вовсе не нужна, точнее, не нужна компетентность, необходимая для социальной революции. Чтобы стать сапожником, надо учиться, для обувной промышленности готовятся кадры в специальных техникумах и институтах. Чтобы стать строителем, тоже надо учиться. Чтобы стать врачом, нужно пройти пятилетний срок специального обучения, потом еще клиническую практику, и затем еще все время учиться и учиться; чтобы стать старшим научным сотрудником, надо не только окончить институт или университет, но и защитить кандидатскую, а то и докторскую диссертации. После всего этого выдержать конкуренцию, требующую личной заинтересованности, позволяющей проявить необходимую активность.

А что нужно изучить, уметь и обязательно знать, чтобы стать членом думы, правительства, областной или городской администрации? Таких жестких требований нет. Их нет потому, что власть сама себе хозяйка, и никто ей не указ. И еще потому, что для проникновения во власть нужны в первую очередь совсем другие качества, чем те, благодаря которым люди совершили научно-техническую революцию, совершают информационную и могли бы совершить научную социально-экономическую. Для того чтобы могла произойти основанная на знаниях и разуме научная социально-экономическая революция, совершенно необходимы открытость деятельности власти, наличие жесточайших критериев оценки делаемого властью и наличие механизма отсека некомпетентных действий, обмана и туфты, постоянного притока во власть деятельных и талантливых людей.

Что же может привести к осуществлению этих требований? Мне представляется, что только осознание неизбежности глубочайшего и жесткого кризиса, осознание наступления катастрофы, угрожающей существованию всех, в том числе и власти, в сочетании с пониманием своих интересов и выгод не только в кратковременном, сиюминутном, но и долгосрочном плане, понимание единства всех людей как вида и их судьбы и, наконец, понимание того, что нужно делать и к чему следует стремиться, чтобы избежать катастрофы.

Величайшим искушением всякой власти является силовое управление обществом и миром. Однако управлять сверхсложным и, самое главное, активным объектом, каким является человеческое общество, силовыми и даже всепокрушающими методами невозможно, не разрушая его, не лишив его совершенно необходимых качеств, способности к самоорганизации, ведущей к его эффективности и процветанию. Силовое управление, в конечном счете, ведет к фактическому уничтожению человеческого общества. Властное управление должно быть не силовым, а способствующим активности всех людей, наилучшему их использованию.

Управление человеческим обществом не может быть полностью централизовано, его основой должна быть его способность к самоорганизации, которой властное управление должно способствовать и наличие чего должно обеспечить. Коллективный разум и деятельность людей не может заменить никакая централизованная власть: он – основа жизни и эволюции общества. Многое из того, что сказано выше в связи с желанием построить и обсудить глобальную математическую модель базового уровня человеческого общества, позволяющую понять принципы его функционирования, организации и управления, бездоказательно и выражает точку зрения автора. Она высказывается не для того, чтобы с ней обязательно соглашались, и далеко не все, что высказано, будет использовано. Это, если угодно, приглашение к обсуждению, к тому, чтобы задуматься.

Вместе с тем математическое рассмотрение предполагает наличие исходных утверждений, которые должны быть приняты. Обычно это общепринятые представления точной науки, и они как бы за кадром: они не обсуждаются, из них исходят, строя дальнейшую теорию. Теперь поступить так нельзя: формулировка исходных положений для общества требует обсуждения и аргументации, они не просты и не очевидны. В данном случае исходные положения – это *игровое восприятие жизни, стремления людей, составляющих общество, и то, что мы хотим от него*. Только после того, как все это будет обсуждено и принято, можно перейти к выводам, которые можно назвать принципами организации и управления обществом.

Исходным положением будет то, что общество создается и существует для человека, для всех людей, и поэтому должно соответствовать природе человека и его стремлениям и желаниям, совместимым между собой и приемлемым для всех людей. Это положение, естественно, нельзя доказать, его можно только аргументировать и не более, всегда может найтись человек, который скажет: «Нет, общество создано только для меня или нас, а государство – это я» или «мир и все люди существуют только для моего народа, и все должны подчиняться и служить ему». Такие точки зрения тоже имеют свои веские аргументы как в природе человека, так и в том, что совершенствование жизни до сих пор осуществлялось через жесткую борьбу за существование. Это нельзя оспаривать, это так было, но на современном этапе развития жизни это уже не так. Этот механизм эволюции изжил себя, его необходимо заменить, потому что он перестает служить в должной мере цели прогресса жизни на Земле. Борьба за существование должна принять другие формы. Так думают сейчас многие, но не все. В этих основах все же легче разобраться, чем в человеческом обществе. В этом и состоит возможная помощь точной науки: сложные и очень сложные вопросы она сводит к менее сложным и даже простым.

Для нас то, что общество в равной мере для всех людей и ради всех людей – аксиома. Борьба за существование как механизм эволюции и совершенствования жизни должна быть смягчена и заменена чем-то другим. Необходимость этой замены вызвана тем, что период чисто физиологического и физического совершенствования живого дополнился, если не сменился, периодом его интеллектуального совершенствования. Интеллектуальное совершенствование открывает новые возможности и пути эволюции и прогресса жизни. А борьба за существование в условиях наличия высокого интеллекта ведет совсем не к тем последствиям, ко-

торые были бы желательны и способствовали бы совершенствованию жизни и успешной деятельности людей.

В формулировании исходных постулатов нет логики, но есть доводы и соображения. От нежелательных последствий борьбы за существование живая природа создавала защитные инстинкты. Человек это может и должен сделать сознательно. Вначале он это делал на основе религиозных представлений и верований, которые как бы спускались с небес, почитаемых и устрашающих. Сегодня это необходимо сделать, опираясь на интеллект и естественные и точные науки.

Необходимая коррекция борьбы за существование – это превращение ее в конкуренцию на основе стремления к достижению поставленной цели и в рамках жестко соблюдаемых правил. Эту трансформацию борьбы за существование можно назвать игрой. Игрой, где все имеют равные возможности, никто не нарушает правил, и победители признательны побежденным, потому что без них не было бы ни победы, ни ее радости, ни самой игры. Без побежденных игровой механизм состязательного совершенствования человечества не работает.

Итак, игра вместо борьбы за существование, равенство прав и возможностей всех людей и, наконец, все это ради людей. Таков итог предварительных соображений о постулатах, которые будут положены в основу дальнейших, уже логических выводов.

Можно предвидеть и слышать веские возражения: «Все это так, но люди в силу своей природы, инстинктов и эгоизма никогда это не будут соблюдать, даже если и примут». К сожалению, это правда. Почти правда, поскольку все же возможны два возражения: одно – в том, что природа наделила человека моралью и этическими представлениями и нормами, душой, и второе – люди все же способны осознать, что прямолинейный эгоизм и стремление к сиюминутной выгоде неизбежно приведут к кризису и катастрофе, что изменения и принятие ограничений в поведении, в конечном счете, просто выгоднее.

Изменению поведения людей может способствовать объяснение реально сложившейся грозной ситуации и возможных путей ее преодоления. Вот именно здесь точная наука, убедительность ее доводов и логики могут помочь. Конечно, народы разных стран и разные люди находятся на разных стадиях способности это понять, осмыслить и изменить свое поведение, и это не произойдет мгновенно, но это должно произойти; трудно поверить, что гигантские возможности самоорганизации природы дадут сбой. Более того, соответствующие тенденции, несомненно, имеются в цивилизованных странах.

Кризис эволюции жизни [38]. Прежде всего, хочется пояснить, почему возникла столь необычная тема и как следует понимать проблему человеческого общества, социальных отношений и их истории.

Историю человеческого общества можно воспринимать как продолжение эволюции жизни на Земле. Проблему эволюции общества людей – как задачу составления математических моделей, способных объяснить наблюдаемую реальность и подсказать возможные пути дальнейшего совершенствования.

Представляется, что человеческое общество настолько сложно и многообразно, а адекватная математическая модель настолько сложна и необозрима, что такой подход безнадежен. В какой-то мере это подтверждается, поскольку пока никаких результатов на этом пути получить не удавалось: все, что написано об об-

щество людей и истории его развития, носит описательно-объяснительный гуманитарный характер.

Рассмотрение простых математических моделей достаточно сложных объектов показало, что иногда с их помощью удается просто ответить на очень сложные вопросы. Таковы модели кризисов сердца, двуногой ходьбы, инфекционных заболеваний, игры в отгадывание, загадок Каспийского моря и др.

Грубое описание человеческого общества простой моделью «производители–продукт–управленцы» было представлено в гл. 12. Эта модель отражает факт объединения конкурирующих между собой людей в общество для более эффективного производства и получения необходимого для жизни продукта. Изложению результатов математического исследования этой модели и вытекающих из него выводов предпослел краткое общее обсуждение проблемы эволюции жизни.

Несомненно, что многое в жизни современного общества – отдельных людей, групп, сообществ, стран и народов, а также всего человечества в целом, – унаследовано от предков. Поэтому, чтобы понять его сегодняшнюю жизнь и пытаться прогнозировать и повлиять на его будущее, нужно осмыслить его прошлое, прослеживая прошлую эволюцию жизни на Земле от достаточно отдаленных времен. Безусловно, зарождение и эволюция жизни таит в себе много загадочного и непознанного. Но вместе с тем можно думать, что основные факторы этой эволюции известны. Это наследственность, изменчивость и отбор – то, на что в итоге указал Дарвин и развито в дарвинизме, и то, что удалось узнать о хромосомах и генах. Как сегодня принято многими исследователями, биологическая наследственность полностью определяется генами. Наследование поведения также носит генетический характер, но лишь отчасти: громадную роль играют традиции, верование, прижизненное обучение и индивидуальный опыт. Изменчивость, как представляется, таинственна и далека от понимания, поскольку поверить в то, что она определяется только случайными мутациями генов, трудно. Трудно поверить, что только случайность и отбор могут привести к наблюдаемому величайшему совершенству и всеобъемлющему использованию возможностей природы. Конечно, время эволюции непостижимо огромно – миллиарды лет. Но, по-видимому, еще что-то нами не познано. Может быть, в основе лежит случайность, как некий генератор, но есть еще фильтр и какая-то направленность. Более того, если иметь в виду не только биологическую эволюцию, но и социальную, то это так, и механизмы фильтрации и направленности, хотя они не выделены и не изучены нами в должной мере, безусловно, имеют место.

Отбор, как таковой, неизменен в том, что отбираются те, кто продолжил жизнь. И до человека именно выжившие в решающей мере определяли эволюцию. Но для человечества это уже не так. Многое определяется накопленными и освоенными человечеством знаниями, традициями, верованиями, моралью и т.д. У животных значительную роль в ускорении эволюции играет половой отбор, у человека – традиции, верования, знания, умения и т.п. Эволюция человечества как бы раздвоилась: на биологическую и социальную, и они друг с другом не согласованы. Они даже подчас антагонистичны. В какой-то мере можно говорить о кризисе эволюционного совершенствования. По-видимому, достаточно обстоятельно впервые на него обратил внимание К. Лоренц в своих «Восьми смертных грехах человечества». Среди них можно выделить генетическое биологическое

вырождение и интеллектуальное оскудение, а также истощение природных ресурсов и непрекращающиеся вражду и войны.

До сих пор природа сама определяла ход эволюции и успешно решила проблему создания *homo sapiens*. Но в эволюции человечества она дает сбой: полезные ранее инстинкты, и прежде всего агрессия, проявляются в войнах, нескончаемой вражде, которые едва ли способствуют разумной эволюции. Они заведомо не помогают реализации наших чаяний о полноценной жизни, основанной не на прямом истреблении, подавлении, неприязни и недоброжелательстве, а на доброте и сочувствии, интересной работе. Можно думать, что такая жизнь лишь плод фантазии, и история коммунизма как бы это подтвердила. Но был не только коммунизм, был еще один эксперимент, давший не менее ужасный результат в виде фашизма. Эти эксперименты, по существу, – попытки реализации двух сторон эволюции человека: биологической и интеллектуальной (ранее названной социальной). Эти крайние эксперименты можно трактовать, как попытки природы уйти от раздвоения, как некое указание, что человечество должно искать выход в разумном согласовании. Нужно трансформировать инстинкт агрессии, нужно умерить требования всеобщего благоденствия, потому что первое ведет к нечеловеческой жизни, а второе – к всеобщей апатии и застою. Едва ли сама природа может выпутаться из этого тупика. Точнее, она может из него выпутаться, используя новые возможности созданного ею человека, опираясь на его интеллект. Успешность такого подхода уже отчасти апробирована: более 200 лет тому назад интеллект небольшой группы людей создал конституцию и определил правила жизни, которые привели к расцвету и ведущей роли современной Северной Америки. Для того чтобы это произошло, потребовалось уникальное сочетание благоприятных обстоятельств. Эти уникальные обстоятельства вряд ли вновь возникнут сами по себе. Более того, действия современной Америки едва ли способствуют разрешению проблемы раздвоения эволюции, едва ли предотвращают угрозы, выявленные К. Лоренцем.

Что же нужно для решения проблемы согласования, для продолжения приемлемой дальнейшей эволюции жизни человечества? Скорее всего, нужен согласованный синтез, и ростки этого синтеза наблюдаются в современном мире – в подмене борьбы за существование игровым восприятием жизни, в замене смертельной схватки игрой, которая дает удовлетворение и радость человеку, не истребляет и не порабощает его, которая сохраняет и стимулирует активность человека без того, чтобы истреблять врага или унижать, лишая интереса к жизни; игрой, которая своей увлекательностью обеспечит активность и эффективность деятельности людей. Замена смертельной схватки игрой широко представлена в животном мире, и ростки ее наблюдаются в современном обществе.

После сказанного вернемся к обсуждению математической модели «производители–продукт–управленцы», даваемого ею объяснения кризиса эволюции жизни с появлением человека и естественно возникающей идеи игрового общества. Ниже мы будем анализировать модель, которую будем называть расширенной, потому что в ней будет учтен параметр m , и параметр μ будет считаться отличным от единицы. Выпишем здесь уравнения рассматриваемой расширенной модели:

$$\begin{aligned} \dot{x} &= (a - bx - ly + cz)x, \\ \dot{y} &= (-dy - mx - ey + fz)y, \\ \dot{z} &= \begin{cases} F = g \frac{1 + \varepsilon_1 y}{1 + \varepsilon_2 y} \frac{\mu x}{1 + \delta z} - hx - ky & \text{при } z > 0, \text{ или } z = 0 \text{ и } F > 0, \\ 0 & \text{при } z = 0 \text{ и } F \leq 0. \end{cases} \end{aligned} \quad (45.1)$$

В этой модели, в отличие от рассмотренной в гл. 12, имеется дополнительный параметр m , учитывающий давление производителей на управленцев и возможность отличия от единицы параметра μ , характеризующего эффективность производителей. Рассмотрение этой модели в сопоставлении с реальной историей содержится в [39].

Как следует из результатов гл. 12 и уравнений (45.1), удалось обнаружить, что с ростом уровня технологии g , помимо производителей x , возникает сначала продукт z и затем управленцы y , и что возникновение управленцев происходит независимо от параметра $\varepsilon_1/\varepsilon_2$, характеризующего их полезность.

При низком уровне технологии, когда $g < h$, имеется единственное устойчивое состояние равновесия $x = x^*$, $y = z = 0$, т.е. в обществе существуют одни производители.

При среднем уровне технологии, когда $h < g < h(1 + \delta d/f)$ появляется накопленный продукт $z = z^*$.

При высоком уровне технологии, когда $g > h(1 + \delta d/f)$, появляются еще и управленцы. При этом устойчивое равновесие $x = x^*$, $y = y^*$, $z = z^*$ может стать неустойчивым и возникнут автоколебания. Это соответствует периодическому изменению всех переменных x , y и z , причем на некоторых участках z может обращаться в нуль, что соответствует глубокому кризису. Из сказанного следует, что возникновение управленцев не зависит от ε_1 и ε_2 , т.е. они возникают просто потому, что общество может их прокормить.

Следует отметить, что значимость второго вывода не легко сразу оценить, а с ним, по-видимому, связана такая многовековая трагичность истории человечества, как проявление кризиса эволюции жизни с приходом человека.

До появления человека вожаком стаи – это самый сильный, умный и полезный, и как только появляется более достойный, он сменяет прежнего вожака. В человеческом обществе с некоторых пор это не так. Согласно модели (12.1) это тоже не так: управленцы – аналог вожака – могут быть не только полезными, но и вредными, и это не мешает их появлению и не ведет к исчезновению, если, конечно, уровень технологии g высок ($g > h(1 + \delta d/f)$).

Это, естественно, ведет к конфликту между управленцами и производителями и, возможно, внутри самих управленцев. Конфликт не нашел отражения в модели, но это наблюдается в течение всей истории человечества и ведет, как правило, к действиям управленцев, способствующих их укреплению, а не отстранению.

Добиться этого удастся одурачиванием общества и подавлением силой, имитацией своей необходимости, присвоением властных полномочий и их злоупотреблением. При этом используются любые средства, вплоть до уничтожения конкурентов, активизации агрессии, раздоров, ненависти, войн, террора на национальной, религиозной, классовой и других основах, придумыванием внешних и внутренних врагов, территориальных и других претензий.

Остановить этот процесс очень трудно, поскольку он охватывает весь мир и сам себя поддерживает. В прошлом веке он привел к двум мировым сокрушительным ужасным войнам, сегодня – к расцвету терроризма и гигантскому отставанию социального совершенствования от фантастических успехов промышленной и информационной революций. Это, по существу, кризис эволюции жизни с появлением человека. Многие видят выход, причем окончательный, в либерально-демократическом обществе. К. Лоренц, напротив, увидел в этом обществе устранимые восемь смертных грехов, могущих привести человечество к катастрофе.

Прежде чем продолжить обсуждение современных последствий дефекта организации общества, обратим внимание на решающую его роль в истории человечества. Дело не только в конфликте и его локальных последствиях, а в извращенном формировании общества и взглядов людей на него, в трансформации необходимого управления в неограниченную власть, в выделении в обществе особой главенствующей части, живущей своей привилегированной жизнью, использующей народ для ее осуществления и лишь отчасти заботящейся о нем в плане своих интересов и целей.

Причина кризиса обнаруживается моделью (45.1), но может ли эта модель объяснить то, что предпринимает человечество для его преодоления. В целом нет, отчасти да. Точнее, в той ее части, которая вызвана стремлением к материальному благополучию и достатку. При этом деятельность людей должна быть направлена на управление параметрами модели.

Управление, в частности, можно осуществить через давление производителей на управленцев (член $mх$ во втором уравнении модели (45.1)). Увеличивая m , можно устранить неприемлемых управленцев или принудить их повысить свою эффективность. Этого же можно добиться, увеличивая h . Желание увеличить потребление производителями требует увеличения эффективности как производителей (увеличение μ), так и управленцев (увеличение $\varepsilon_1/\varepsilon_2$). Сказанное следует из условия

$$\mu g > h \left(1 + \frac{\delta(db + ma)}{fb - mc} \right),$$

отражающего наличие глобально устойчивого равновесия и надлежащей зависимости m от $\varepsilon_1/\varepsilon_2$.

Таким образом, кризис – дефект объекта, описываемого моделью (45.1), – может быть преодолен с появлением у общества возможности управления параметрами модели и непосредственным воздействием, а также способностью этим воспользоваться.

Эти возможности могут реализоваться в двух планах: описанным давлением на управленцев, что требует соответствующих прав, и выборностью управленцев при условиях достоверных сведений о них и зрелости общества.

Либерально-демократические общества по своей идее такие возможности предоставляют. Отсутствие возможности эффективного управления приводит к революционному пути его реализации.

Расширенная модель (45.1) описывает динамический процесс под многообразным воздействием природы и людей. Это хаотическое случайное воздействие

предсказуемо только на небольших участках времени и то далеко не всегда. Но, наверное, можно говорить о его общих тенденциях.

Что включает в себя модель (12.1) и что из нее непосредственно следует, было сказано. То же самое можно сказать и о расширенной модели (45.1), включающей элементы управления, охарактеризовать мотивы и цели управления и известные пути его реализации. Управленцы заинтересованы в эффективности производителей μ , совершенствовании технологии g , уменьшении доли производителей h в производимом продукте и увеличении своей доли k . Производители, напротив, заинтересованы в увеличении своей доли h и эффективности управления $\varepsilon_1/\varepsilon_2$.

Реализация желаний управленцев ими исторически узурпирована: они власть. Производители должны добиваться реализации своих желаний и ограничения желаний управленцев. Это сделать можно через профсоюзы и другие общественные организации, через оппозицию, через контроль за деятельностью управленцев, опирающийся на независимые средства массовой информации и правоохранительные органы, выборность власти и досрочное ее отстранение. Ослабление узурпации власти, коррупции и сговора возможно путем разумного разделения ее полномочий.

Как уже отмечалось, либерально-демократическое общество позволяет реализовать более или менее эффективное обоюдостороннее управление объектом – моделью (45.1), но оно в недостаточной мере удовлетворяет многогранным потребностям человека. Более того, оно ведет к восьми смертным грехам человечества, указанным К. Лоренцем. Кроме того, равновесное противостояние в либерально-демократическом обществе склонно к потере устойчивости.

Что же можно предложить человечеству, удовлетворяющее его страстям, его врожденным наклонностям и устремлениям? Что больше всего любит и ценит человек, что может удовлетворять его постоянно, что сделает возможным преодолеть кризис и создать устойчивую и одновременно морально и этически приемлемую организацию социальной жизни, обеспечить организацию общества, для которого либерально-демократическое общество – исходный плацдарм, общества, которое сможет обуздать неумеренное потребление и расточительство, создать жизненно необходимые соблюдения требований экологии и уважения ко всем людям, победить необузданную тягу к власти и деньгам? Подсказка пришла не от математики, а от великих провидцев человеческого духа и жизни.

Игровое восприятие жизни и игровая модель общества. Несомненно, все любят играть, кто как и с кем, или с чем. Жизнь человека можно воспринимать и описывать как игру, подчас комичную, трагичную, интересную, скучную, увлекательную, постылую, или какую-нибудь другую, но почти всегда желанную. Наверное, все помнят слова из арии Елецкого (опера «Пиковая дама»): «Что наша жизнь – игра...». Этой фразой открывается по телевидению показ игры «Что, где, когда?» в элитарном клубе, где каждый может зарабатывать большие деньги своим интеллектом.

Известный писатель М. Пришвин написал: «Жизнь желанная – это игра, все, кто может, играют, а кто не может – трудится в надежде когда-нибудь поиграть».

Природа наделила человека многими качествами, которые способствуют его активному, подчас самозабвенному участию в игре: самолюбием, честолюбием, желанием первенствовать и добиваться своего, и т.п. Победа в игре может вы-

звать незабываемые по приятности и привлекательности эмоции. Люди способны многими годами истязать себя тренировками ради единого мига ощущения восторга и торжества победы на Олимпийских играх или других состязаниях.

Еще А. Эйнштейн сказал: «Успех – это неустанный труд плюс умение относиться к жизни как к игре».

Так что игра не только желанна и любима, но она эффективна и способствует успеху в жизни. Игра – приемлемая форма и образ жизни людей.

Все привыкли, что в игре кто-то выигрывает, а кто-то проигрывает. Все выиграть не могут. Это кажется бесспорным, но есть много игр, для которых это не так. Прежде всего, не так для игр человека с природой, это не так и для социальных игр, т.е. игр человека в обществе.

Кроме того, все привыкли, что играя надо свои «карты» прятать, скрывать свои намерения. Но оказывается, что и это не всегда так. Может быть, совсем наоборот, и выгодно показать свои «карты» и даже договориться с партнерами о совместных действиях.

Дело в том, что игры понимаются гораздо шире, чем те, в которые мы ради удовольствия играем между собой. Математика описывает игру, как простейшую одноходовую, когда каждый игрок делает свой ход и каждый получает выигрыш, зависящий от его хода и ходов остальных игроков. Сказанное можно наглядно представить в виде таблицы, где в первой строке представлены все игроки I_1, I_2, \dots, I_n ; во второй – ходы, которые они могут сделать, и при этом каждый игрок I_s выбирает любой ход x_s из своего множества Ω_s разрешенных ему ходов; в последней строке стоят получаемые в результате сделанных ходов выигрыши каждого из игроков: игрок I_s , сделав ход $x_s \in \Omega_s$, получает выигрыш $f_s(x_1, x_2, \dots, x_n)$. Это простейшая одноходовая игра n лиц, но уже ее достаточно, чтобы продемонстрировать сказанное выше об играх. Для этого воспользуемся обсужденным ранее примером Ю.Б. Гермейера. В придуманной им игре каждый игрок I_s выбирает любое число x_s ($s = 1, 2, \dots, n$), $n \geq 3$, заключенное между 0 и 1, и получает выигрыш

$$f_s(x_1, \dots, x_n) = x_s + \sum_{j \neq s} (1 - x_j). \quad (45.2)$$

Пусть каждый игрок знает только свою функцию выигрыша (45.2). Он видит, что может максимально способствовать своему выигрышу, выбрав $x_s = 1$. И так поступают все, т.е. все выбирают $x_1 = x_2 = \dots = x_n = 1$, и в результате этого выбора, согласно (45.2), каждый получает выигрыш в одну единицу. Если же все игроки раскроют свои функции выигрыша и каждый о них узнает, то, немного подумав, они могут сообразить, что выиграют значительно больше, если договорятся, что все будут в «ущерб себе» выбирать $x_s = 0$ ($s = 1, 2, \dots, n$). После этого выигрыш каждого окажется равным $n - 1$, т.е. больше прежнего, равного 1, поскольку $n \geq 3$. Вместе с тем каждый игрок имеет искушение не выполнить договоренность, выбрав вместо нуля единицу, и выиграть n вместо $n - 1$. При этом все остальные игроки получают на единицу меньше, и, скорее всего, это приведет к развалу договоренности, и все окажутся в убытке, получив 1 вместо n .

Нетрудно догадаться, почему в этой игре все не так, как в привычных для нас играх. В этой игре сумма всех выигрышей положительна, в то время как в привычных играх она равна нулю, и выигрыш одних получается только за счет проигрыша других. При игре двух игроков с нулевой суммой выигрышей игра носит антагонистический характер со всеми вытекающими отсюда следствиями. Но в игре с природой, за счет природы, ее ресурсов и даров, сумма выигрышей всех с ней играющих может быть положительной.

Теперь можно сформулировать игру, в которую мы играем, живя в обществе и участвуя во всеобщей трудовой или иной деятельности.

Каждый человек I_s по желанию или в силу тех или иных обстоятельств и соображений выбирает из множества возможных игр O игру, в которую он намеревается играть. Выбрав игру O_s , он одновременно выбирает множество своих возможных действий Ω_s и функцию выигрыша $f_s(x_1, \dots, x_n)$. Игра состоит в последовательном выборе каждым из игроков множества Ω_s , действия x_s и получении соответствующего выигрыша $f_s(x_1, \dots, x_n)$, зависящего как от его действия x_s , так и действий остальных игроков. Каждый игрок в течение своей жизни делает много ходов и может менять игры, но последнее обычно происходит редко.

Обратим внимание, что действия, которые выбирает игрок I_s из множества Ω_s , определяются не только множеством Ω_s , но и реальными возможностями игрока. Иными словами, выбираемое действие должно принадлежать некоторому присущему данному игроку множеству ω_s , определяемому всей его предысторией и, в частности, предшествующей игрой и ее результатами. В принципе может меняться со временем и функция выигрыша.

Принципы организации и управления обществом. Постараемся сформулировать, чего каждый человек, живя в обществе, добивается и хочет, и чего хотят люди от общества. Кажется, что сделать это очень трудно и даже невозможно: людей очень много, и они все разные. Но не нужно входить в детали, ограничимся тем, что является общим для всех или почти всех. Ведь мы хотим взять в основу и обсудить простую модель, а не модель, достаточно адекватную обществу во всех его проявлениях. Общим для всех людей является:

- 1) стремление к возможно большему удовлетворению своих потребностей;
- 2) желание объединиться в общество для более эффективного производства продукта и большего удовлетворения своих потребностей, несмотря на естественную конкуренцию;
- 3) стремление отдельных людей к максимальному удовлетворению своих потребностей, что, в конечном счете, должно привести к возможно большему удовлетворению потребностей всех членов общества, так, чтобы успех каждого одновременно был успехом всех.

Обсудим условия осуществимости принципов 1)–3) (стремлений людей и их желаний) в рамках игровой модели общества, точнее, условия, необходимые для их выполнения, и назовем их принципами организации и управления обществом.

Прежде всего, отметим, что равенство прав и возможностей всех людей общества заложено в самой игровой модели. Некоторое неравенство все же имеется в силу неодинаковости множеств ω_s , но это неравенство столь же неизбежно, как и различие данных, которые получает человек от природы, от того, каким он ро-

дился и как проявлял себя. Возможно, что общество должно как-то компенсировать это естественное неравенство, что и наблюдается.

Далее, согласно 1)–3), совершенно необходимо соблюдение правил игры, в частности, только при их соблюдении имеет место равенство прав и возможностей. Только при соблюдении этих правил игра привлекательна, желанна и может быть эффективной. С современной точки зрения, соблюдение правил игры можно трактовать как соблюдение прав и обязанностей каждого человека. Для соблюдения этих прав и успешности игры, которая требует возможно большей информации о ней, необходима открытость функционирования общества и происходящих в нем процессов принятия решений, от которых зависят функции выигрыша каждого из игроков. Не будем детализировать, что и в какой мере нужно знать или уметь прогнозировать. Для нас сейчас важно лишь то, что для этого необходима открытость общества. Итак, открытость общества – это одна из необходимых гарантий соблюдения прав человека и возможности успешной игры всех людей. Возможно, что эта многогранная и существенная роль открытости общества послужила тому, что Дж. Сорос выдвинул ее как основную характеристику желаемого общества и стремление к его созданию. И еще одно необходимое условие вытекает из того, что чем шире множество допустимых ходов, тем шире возможности оптимизации выигрышей. Это требование максимально допустимой широты множеств Ω_s допустимых действий можно трактовать как принцип минимальности принуждения, минимальности ограничений действий людей. В результате мы пришли к трем принципам организации и управления обществом: соблюдения прав человека, открытости общества и минимальность принуждения.

Обсудим, как обеспечить выполнение привнесенной и сформулированной в п. 3 цели общества. Ранее потребление и продукт могли трактоваться весьма широко. Теперь это только деньги, сами по себе или их эквивалент. Денежную часть выигрыша f_s обозначим \bar{f}_s . Введем еще фактический денежный выигрыш φ_s каждого игрока. Ясно, что общий денежный выигрыш F всего общества равен $F = \sum \bar{f}_s = \sum \varphi_s$. Выигрыш каждого будет способствовать выигрышу всех остальных игроков и общему выигрышу, если принять, что $\bar{f}_s = \alpha_s F$, где $\alpha_s > 0$, $\sum \alpha_s = 1$, но при этом все игроки лишены возможности оценивать реальный эффект своих действий. Кроме того, игроки теряют интерес к игре, так как выигрыш игрока фактически не зависит от его действий. Этого можно было бы избежать, приняв $\bar{f}_s = \beta_s \varphi_s$, где $\beta_s > 0$, $\sum \beta_s \varphi_s = \sum \varphi_s$, но при этом не выполняется способствование выигрыша каждого выигрышу всех.

Оба требования могут быть соблюдены, если принять, что

$$\bar{f}_s = \alpha_s F + \beta_s \varphi_s, \quad (45.3)$$

где $\alpha_s > 0$, $\beta_s > 0$. При этом коэффициенты β_s не должны быть слишком малыми, а коэффициенты $\alpha_s > 0$ – слишком большими. Это необходимо для того, чтобы изменения функции выигрыша, по которым игрок может судить об эффективности своих действий, в основном определялись функцией φ_s , точнее, позволяли понять, как меняется φ_s .

Цель общества (п. 3) предполагает еще максимизацию общего выигрыша F через максимизацию всех индивидуальных выигрышей φ_s , т.е. максимум функции F или близкая к нему величина F должна достигаться при оптимизации каждой из функций φ_s по переменной x_s . Можно заметить, что если каждая из функций φ_s зависит только от x_s , то из оптимизации всех φ_s по x_s следует оптимизация общего выигрыша F , как функции всех переменных x_1, x_2, \dots, x_n . Можно надеяться, что нечто близкое будет и в том случае, когда в функции φ_s ведущей, самой существенной переменной является x_s . Интересно заметить, что игры с выполнением такого условия наиболее привлекательны.

Подведем итог тому, что мы узнали о принципах функционирования, организации и управления современным обществом. Принцип функционирования общества – это игра, принципы организации – это:

1.1) открытость, свободный доступ к информации о функционировании, процессах и событиях, происходящих в обществе;

1.2) принцип наименьшего принуждения, минимальность ограничений в действиях членов общества;

1.3) принцип прав человека, т.е. правил жизненных игр, обязательных и одинаковых для всех.

Управление обществом должно устанавливать правила игры, т.е. величины выигрышей \bar{f}_s и множества допустимых действий Ω_s , причем делать это так, чтобы:

2.1) каждый из игроков был лично заинтересован в игре и выигрыше, располагал достаточной информацией и возможностью ее оптимизации;

2.2) общий выигрыш F способствовал всем индивидуальным выигрышам \bar{f}_s ;

2.3) оптимизация индивидуальных выигрышей была бы одновременно в достаточной мере и оптимизацией общего выигрыша F .

Кроме того, управление обществом должно обеспечить соблюдение правил установленной игры и способствовать их совершенствованию. Стремление людей к совершенствованию потребления может осуществляться не только за счет более целесообразных действий в играх, но и за счет создания новых игр и новых действий. Новые игры и допустимые действия – это новые орудия труда, типы и способы производства, добычи и обработки, новые технологии, производственные, сельскохозяйственные и другие процессы, новые условия жизни и труда, способствующие более полному удовлетворению потребностей человека. Совершенствование игр происходит на основе познавательной деятельности людей, благодаря их любознательности и интеллекту, желанию понять, исследовать и улучшить. Игра должна охватывать в должной мере и эти виды деятельности человека, что ведет к большей привлекательности совместной игровой жизни и способствует общему прогрессу.

Базой этих нововведений является познавательная, исследовательская и поисковая деятельность людей на основе преемственности знаний и умений, обеспечиваемых воспитанием и образованием. Инстинкты и природа человека способствуют этим видам его деятельности, но не во все времена они пользовались поддержкой общества и включались в денежный выигрыш. Общество в этом вопросе

проявляло явно недостаточное внимание к своей выгоде, что лишь отчасти компенсировалось царской милостью, меценатской и благотворительной деятельностью. У этого недопонимания есть веское основание: отсутствие денежной оценки таких действий, отдаленность их эффекта и заинтересованность в них общества в целом, а не конкретных людей. Это привело к тому, что одним из важнейших факторов стимуляции научных и технических исследований стали вооруженное противостояние и конфликты. Неприемлемость этой ситуации очевидна, но исправить ее очень непросто. Агрессивность и стремление к главенствованию тоже в природе человека, и еще недавно являлись, как представляется, основными механизмами эволюции жизни на Земле. Вооруженные конфликты тоже игры, в историческом плане великие, но едва ли приемлемые в настоящем и будущем. С точки зрения рассматриваемой игровой модели общества они противоречат его цели, а включение игр с познавательными действиями (наука, техника, культура, обучение и др.) для нее естественно, поскольку расширяет привлекательность игр, соответствует врожденным устремлениям и склонностям многих людей и существенно, если не решающе, расширяет эффективность игр всех людей и общества в целом.

Идеальная игра людей в обществе. Исходя из игровой модели общественной жизни людей, их стремлений, сформулированных в п. 1) и 2), и привнесенной цели общества 3), мы пришли к необходимости соблюдения ряда общих принципов. Нельзя сказать, что они неожиданны или новы. В той или иной мере они высказывались ранее и даже отчасти осуществляются в цивилизованных странах. Но наряду с этими высказывались и реализовывались совершенно другие принципы. В отношении выигрыша (45.2), (45.3) можно сказать, что он имеет место в любом обществе. Общество с $\alpha_s = 1/n$ и $\beta_s = 0$ можно назвать идеальным коммунистическим, а общество с $\alpha_s = 0$ и $\beta_s = 1$ – диким капиталистическим.

В цивилизованном обществе прямой заработок – это слагаемое $\beta_s \varphi_s$ в формуле (45.3), а слагаемое $\alpha_s F$ – это то, что член общества получает отчасти или полностью в виде образования, медицинского, культурного, социального, правоохранительного обеспечений и поддержек и других бесплатных услуг.

Новое, что мы узнали: каких принципов следует придерживаться, если принять исходные постулаты, и из каких соображений определяются коэффициенты α_s и β_s ; выполнение каких функций обеспечивается каждым из этих принципов и в чем содержание каждого из них. В целом они призваны способствовать привлекательности и эффективности игр и соблюдению их правил.

В частности, были сформулированы требования к всеобщей игре (в том числе и экономической) людей, составляющих общество. В своей полноте они едва ли выполнимы, но это и не требуется, достаточно лишь некоторое приближенное их соблюдение. Вместе с тем изучение игр, удовлетворяющих этим идеальным требованиям, представляет определенный интерес и позволяет сформулировать принципы, на основе которых в обществе становятся возможными приближенная организация и самоорганизация.

Напомним, что игра состоит в том, что каждый из игроков I_s , $s = \overline{1, n}$, выбирает значение своей переменной x_s и получает при этом выигрыш $f_s(x_1, \dots, x_n)$.

Естественной стратегией каждого из игроков является максимизация своего выигрыша по доступной ему переменной x_s . Назовем игру идеальной, если эта стратегия максимизации личного выигрыша ведет к максимизации общего выигрыша всех игроков. Это требование можно сформулировать таким образом, чтобы максимум функции $f_s(x_1, \dots, x_n)$ по всем x_s достигался в той же точке $O(x_1^*, \dots, x_n^*)$, что и максимум функции $F(x_1, \dots, x_n) = \sum_s f_s(x_1, \dots, x_n)$.

Будем считать, что каждый из игроков умеет увеличивать свою функцию выигрыша по своей переменной x_s , меняя ее в соответствии с уравнениями

$$\dot{x}_s = \varepsilon_s(t) \frac{\partial f_s}{\partial x_s}, \quad s = \overline{1, n}, \quad (\varepsilon_s(t) > 0) \tag{45.4}$$

или аналогичными с малыми задержками.

Условия успешности индивидуальных поисков наибольших выигрышей. Вопрос, который решается ниже: когда стратегия (45.4) приводит к достижению идеальной игры и когда точка O единственная?

Возьмем функцию V вида

$$V = \frac{1}{2} \sum_s \varepsilon_s(t) \left(\frac{\partial f_s}{\partial x_s} \right)^2 \tag{45.5}$$

и найдем ее производную по времени при изменении переменных x_s согласно (45.4):

$$\dot{V} = \frac{1}{2} \sum_s \dot{\varepsilon}_s(t) \left(\frac{\partial f_s}{\partial x_s} \right)^2 + \sum_{s,k} \xi_s \frac{\partial^2 f_s}{\partial x_s \partial x_k} \xi_k, \tag{45.6}$$

где $\xi_i = \varepsilon_i(t) \partial f_i / \partial x_i$, $i = \overline{1, n}$.

Сходимость поисковой стратегии (45.4) к выполнению условия идеальной игры может иметь место, если функция V стремится к нулю. Последнее будет выполнено, если

$$\lim_{t \rightarrow \infty} \int_0^t \delta(t) dt = -\infty, \tag{45.7}$$

где $\delta = \dot{V} / V$. Это условие является необходимым и достаточным.

Из (45.5) и (45.6) видно, что для выполнения (45.7) естественно принять, что квадратичная форма

$$\Phi = \sum_{s,k} \xi_s \frac{\partial^2 f_s}{\partial x_s \partial x_k} \xi_k \tag{45.8}$$

определенно отрицательная, так что

$$\Phi \leq -\chi (\xi_1^2 + \dots + \xi_n^2), \quad \chi > 0. \tag{45.9}$$

Далее из (45.5), (45.6) и (45.9) непосредственно находим, что

$$\delta(t) \leq A - B,$$

где $A < \sum |\dot{\varepsilon}_s(t) / \varepsilon_s(t)|$, $B \leq -2\chi \min \varepsilon_s(t)$, при $0 < \underline{\varepsilon} \leq \varepsilon_s(t) \leq \bar{\varepsilon} \leq \infty$ условие (45.7) выполняется, поскольку

$$\int_0^t \delta(\tau) d\tau < \sum \left| \ln \frac{\varepsilon_s(t)}{\varepsilon_j(0)} \right| - 2\chi \int_0^t \min \varepsilon_s(\tau) d\tau \leq n \max \left\{ \left| \ln \frac{\underline{\varepsilon}}{\bar{\varepsilon}} \right|, \left| \ln \frac{\bar{\varepsilon}}{\underline{\varepsilon}} \right| \right\} - 2\chi \underline{\varepsilon} t .$$

Из последней оценки следует, что

$$V(t) < V(0)e^{a-bt}, \quad a, b > 0. \tag{45.10}$$

Согласно (45.10), приход из любой точки области $V(x_1, \dots, x_n) < c < \infty$ в область $V(x_1, \dots, x_n) \leq d$, где $d > 0$, совершается за время, не большее некоторого $T < \infty$. Отсюда следует, что все точки области $V(x_1, \dots, x_n) < c < \infty$ за конечное время переходят в точки одной и той же области $V(x_1, \dots, x_n) \leq d$, они, в свою очередь, при достаточно малом $d > 0$ переходят в единственную точку $O(x_1^*, \dots, x_n^*)$, для которой

$$\frac{\partial f_1}{\partial x_1} = \frac{\partial f_2}{\partial x_2} = \dots = \frac{\partial f_n}{\partial x_n} = 0, \tag{45.11}$$

и других таких точек в области $V(x_1, \dots, x_n) < c$ нет. Поскольку c – любое конечное положительное число, то других таких точек нет вообще. Точка $O(x_1^*, \dots, x_n^*)$ единственная, и для нее, согласно (45.8),

$$\frac{\partial^2 f_s}{\partial x_s^2} < 0, \quad s = \overline{1, n}, \tag{45.12}$$

и поэтому в ней имеет место максимум каждой из функций f_s по переменной x_s .

Итак, уравнения (45.11) имеют единственное решение – точку $O(x_1^*, \dots, x_n^*)$, и она же есть глобально устойчивое состояние равновесия поисковой стратегии, описываемой системой уравнений (45.4). Кроме того, эта же точка является точкой максимума всех выигрышей $f_s(x_1^*, \dots, x_n^*)$ по соответствующей переменной x_s .

Допустимость небольших временных задержек. Пусть стратегия (45.4) реализуется с некоторыми временными задержками, так что

$$\dot{x}_s = u_s, \quad T_s \dot{u}_s + u_s = \varepsilon_s \frac{\partial f_s}{\partial x_s}, \tag{45.13}$$

где ради простоты примем ε_s постоянными, большими $\underline{\varepsilon} > 0$. Для доказательства сформулированного в заглавии утверждения достаточно взять в качестве используемой ранее функции V другую функцию вида

$$V = \frac{1}{2} \left\{ \sum_s \varepsilon_s \left(\frac{\partial f_s}{\partial x_s} \right)^2 + A \left(x_s - \varepsilon_s \frac{\partial f_s}{\partial x_s} \right)^2 \right\}, \tag{45.14}$$

где $A > 0$ затем подбирается нужным образом.

Когда стратегия максимизации индивидуальных выигрышей приводит к глобальному максимуму общего выигрыша? Ответ на этот вопрос сводится к тому, чтобы установить, когда точка $O(x_1^*, \dots, x_n^*)$, являющаяся единственным глобально устойчивым равновесием системы (45.4) и одновременно единственной точкой, в которой все выигрыши f_s максимальны по переменной x_s , будет еще и

единственным глобальным максимумом функции общего выигрыша $F = f_1 + \dots + f_n$.

Единственность максимума, и, следовательно, его глобальность, функции F будет иметь место, если квадратичная форма

$$\Phi_1 = \sum \xi_s \frac{\partial^2 F}{\partial x_s \partial x_k} \xi_k$$

является отрицательно определенной и справедлива оценка вида (45.9). Можно заметить, что

$$\Phi_1 = \sum_{s,k} \xi_s \frac{\partial^2 f_s}{\partial x_s \partial x_k} \xi_k + \sum_{s,k,p \neq s} \xi_s \frac{\partial^2 f_p}{\partial x_s \partial x_k} \xi_k. \quad (45.15)$$

Первая квадратичная форма этой суммы двух слагаемых определенно отрицательна в силу (45.9), ее главные квадратичные члены – это $(\partial^2 f_s / \partial x_s^2) \xi_s^2$ и естественно предположить, что этот член и самый существенный – больший по величине, поскольку переменная ξ_s , скорее всего, самая существенная у выигрыша f_s . Аналогичный член второй квадратичной формы (45.15), $(\partial^2 f_s / \partial x_s^2) \xi_s^2$, выглядит значительно внушительней, но он не более, чем суммарный результат жесткостей влияния хода игрока I_s на выигрыши всех остальных игроков, которых не так уж много и которые могут усредняться. Сказанное – не более, чем довод для принятия условия отрицательной определенности квадратичной формы Φ_1 и оценки $\Phi_1 \leq -\chi_1 (\xi_1^2 + \dots + \xi_n^2)$, которая обеспечивает предполагаемую единственность и глобальность максимума общего выигрыша F , а его совпадение с точкой $O(x_1^*, \dots, x_n^*)$ будет при

$$\frac{\partial F}{\partial x_s} = \sum_{p \neq s} \frac{\partial f_p}{\partial x_s} = 0, \quad s = \overline{1, n}. \quad (45.16)$$

Если трактовать $\partial f_p / \partial x_s$ как жесткость влияния игрока I_s на игрока I_p , то сумма (45.16) – сумма всех жесткостей давления игрока I_s на остальных игроков I_p ($p \neq s$). По сравнению с чем эта сумма жесткостей должна быть мала, чтобы можно было ожидать близости максимума общего выигрыша к общему выигрышу в точке $O(x_1^*, \dots, x_n^*)$? Этот вопрос рассматривается в следующем разделе.

Основной принцип совместной идеальной игры в обществе. Оптимальный общий выигрыш находится из уравнений

$$\frac{\partial F}{\partial x_s} = \frac{\partial f_s}{\partial x_s} + \sum_{p \neq s} \frac{\partial f_p}{\partial x_s} = 0, \quad s = \overline{1, n}, \quad (45.17)$$

т.е. добавочный член в уравнении (45.17) тем менее существенен, чем больше $\partial^2 f_s / \partial x_s^2$ по сравнению с $\sum_{p \neq s} \partial f_p / \partial x_s$. Это следует из первого приближения для изменения ds по отношению к x_s^* :

$$\frac{\partial^2 f_s}{\partial x_s^2} dx_s + \sum_{p \neq s} \frac{\partial f_p}{\partial x_s} = 0, \quad s = \overline{1, n}. \quad (45.18)$$

Итоговый вывод о близости точки глобального максимума общего выигрыша F к точке $O(x_1^*, \dots, x_n^*)$ максимума личных выигрышей можно сформулировать как требование малости сумм жесткостей влияния хода каждого игрока на общий выигрыш всех остальных игроков. Этого можно достичь за счет малости каждого из влияний отдельно, а можно и за счет малости суммарного влияния, учитывая возможность их компенсации. Сказанное можно сформулировать как некий принцип поведения в общей идеальной экономической игре людей в обществе: с одной стороны, каждый игрок должен стремиться к своему личному выигрышу, неуклонно его оптимизируя, а с другой – по возможности в целом не вредить и не способствовать выигрышам других. Как способствование, так и препятствование выигрышам других, оказывается, уменьшает возможный общий выигрыш. Действительно, первое приближение общего выигрыша над суммой индивидуальных максимизированных выигрышей, согласно (45.16), равно

$$dF = \sum_s \frac{\partial F}{\partial x_s} dx_s = \sum_s \frac{\partial F}{\partial x_s} \left[- \sum_{p \neq s} \frac{\partial f_p}{\partial x_s} \left(\frac{\partial^2 f_s}{\partial x_s^2} \right)^{-1} \right] = - \sum_s \left(\sum_{p \neq s} \frac{\partial f_p}{\partial x_s} \right)^2 \left(\frac{\partial^2 f_s}{\partial x_s^2} \right)^{-1} > 0$$

и всегда положительно, и тем больше, чем больше квадраты сумм жесткостей влияния. Можно заметить, что это превышение, во всяком случае, в первом приближении, тем меньше, чем «острее» максимумы личных выигрышей (больше по величине вторая производная от f_s по x_s).

О целесообразной организации и самоорганизации экономических игр в обществе. Под целесообразной организацией понимается близость к идеальной игре. Как можно судить из предыдущего, оптимальное экономическое функционирование при идеальной игре требует от ее участников только умения и успешности в оптимизации своих личных выигрышей. Но оптимальная игра едва ли реальна, да и кто должен ее организовывать и определять? Значительно реальнее некоторое к ней приближение, достигаемое целесообразной стратегией самих экономических игр и экономической политики. Основной принцип этой самоорганизации, ведущей к близости экономической игры к оптимальной, состоит в том, что каждый игрок должен возможно больше стремиться к своему личному выигрышу и одновременно организовывать ее так, чтобы уменьшить суммарное влияние на других игроков, достигаемое тем, что он сам, выбирая игру, стремится максимально возможно избавиться от влияния остальных игроков. Такая самоорганизация естественно предполагает весьма широкие возможности выбора игры, удовлетворяющей этим требованиям, что, в частности, требует достаточно хорошей информированности и понимания последствий своих действий.

Тривиальный случай идеальной игры – это когда каждая из функций выигрыша f_s зависит только от x_s . Этого положения можно достичь путем объединения игроков с сильными зависимостями в «одного», а можно путем организации взаимоотношений, которые не допускают больших влияний друг на друга. Следует отметить и еще одну особенность идеальной игры: успешность оптимизации выигрышей целиком зависит от личных действий каждого игрока, когда его прогнозы проще и надежней, и он может сосредоточиться только на эффективности своих действий, используемых средств и технологий. Может показаться, что в своей полной осуществимости идеальная игра как бы исключает конкуренцию, но

это не так, потому что она присутствует, но не на уровне базовых игровых действий, а в сопоставлении их результатов. Это отдельная важная тема.

Проблема включения управленцев и власти в общее игровое взаимодействие. Ранее определена функциональная структура низшего базового уровня общества, как общества игрового, и основные принципы его организации и управления, следующие из требований эффективности игры и ее общей приемлемости. Прежде чем перейти к теме этого раздела, позволим себе небольшое повторение. Как уже отмечалось, общество и составляющие его люди в своей организации и функционировании похожи на живой организм, состоящий из клеток. Сообщество клеток обеспечивает жизнь всего организма, а жизнедеятельность клеток обеспечивается гомеостазом среды их обитания. Нечто аналогичное имеет место и в человеческом обществе: для его жизни необходимо обеспечение соответствующей среды, называемой базовым уровнем общества. Однако по мере совершенствования многоклеточного живого организма, как необходимое условие жизни, помимо гомеостаза среды, возникли еще гуморальное и нервные управления. В обществе их аналогом можно считать управление жизнью людей, осуществляемое государством и его властными и полномочными структурами. В живом организме все происходит как бы само собой, в силу законов природы и тенденций ее самоорганизации. В человеческом обществе многое организуется при целенаправленном вмешательстве людей. Теперь естественно возникает очень важный вопрос: где и как пролегает граница в современном обществе между централизованным управлением и тем, что можно назвать самоорганизацией, возникающей на основе базового уровня общества, между сознательным централизованным управлением и тем, что можно назвать самоорганизацией, требующей лишь обеспечения базового уровня? Кроме того, сама организация базового уровня требует наличия механизмов формирования игр и соблюдения их правил. Как это делается, выходит далеко за рамки рассматриваемой игровой модели, но сама проблема власти возникает уже внутри нее, и с этих позиций можно сформулировать сущность проблемы и пути ее разрешения.

На наличие этой проблемы указывала и ранее рассмотренная математическая модель «производители – продукт – управленцы», поскольку в ней управленцы возникали не потому, что они нужны и полезны, а потому, что общество может их содержать. В этом их трагическое сходство с паразитическими элементами общества. С позиций игровой модели проблема состоит в том, чтобы включить управленцев и оба вида власти в равноправное всеобщее игровое взаимодействие, в общую для всех игру. Сам статус в обществе управленцев и власти как бы препятствует равноправию, и вместе с тем его необходимо соблюсти. В этом и состоит, на мой взгляд, центральная проблема современного общества – проблема власти. Как это делается, математическая модель ответить не может, но она позволяет ответить на вопрос, что для этого нужно сделать. А сделать нужно так, чтобы ни один из управленцев или власти, выполняя свои функции, т.е. играя в одну из игр, не имел возможностей за счет своих властных полномочий изменять правила своей игры или позволять себе их не исполнять. Для этого нужно целенаправленное и полное разделение *полномочий* между отдельными исполнителями и структурами управленческих и властных полномочий. Разделения, исключающего не только непосредственную возможность использования власти в своих целях, но и

сговора. Организация власти, удовлетворяющая этим требованиям, автоматически приведет к высокому профессиональному уровню и компетентности управленцев и власти, так как только при их наличии они смогут успешно играть и выдержать естественную общую конкуренцию. В какой-то мере именно в этом направлении происходят изменения в цивилизованных обществах: разделение властей, открытость намерений и действий, выборность и сменяемость.

Безусловно, рассмотренная выше игровая модель до примитивности проста по сравнению с реальной жизнью людей на Земле с ее многообразными непостижимыми сложными взаимоотношениями как отдельных людей, так и их коллективов, стран и народов. Все это объять и адекватно отобразить в математической модели невозможно. Да и не о такой модели идет речь. Речь идет только о модели базового уровня общества, о фундаменте, на котором оно стоит, и среде, в которой функционирует. О среде, которая позволит и будет способствовать активной и эффективной деятельности людей, деятельности желанной всеми, в целом почти всеми приемлемой и почти всех устраивающей. Все остальное, что сверх базового уровня и его обеспечения, создается как бы само собой активной деятельностью людей и их способностью к самоорганизации. Все остальное люди создадут в процессе участия во всеобщей игре, включая и помогающее им математическое моделирование возможных сторон их деятельности и жизни.

К модели базового уровня не нужно предъявлять требования полноты и адекватности реальному обществу людей. Нельзя, потому что сколько-нибудь полная модель настолько сложна, что едва ли позволит сделать какие-то общие выводы. От модели базового уровня нужно и можно требовать, чтобы конструируемый уровень обеспечивал и порождал достойную жизнь людей на Земле с ее условиями, окружающей природой и свойствами людей, был основой, обеспечивающей эффективное, удовлетворяющее людей функционирование общества.

Принципы организации, функционирования и управления, сформулированные выше, – это лишь необходимые условия. Они будут достаточными, если эта основа – базовый уровень – устойчив и самодостаточен.

Устойчивость требует приемлемости и желанности такого базового уровня для подавляющего числа людей; возможности его реализовать и поддерживать, фактической невозможности его дезорганизации и разрушения.

Самодостаточность можно видеть в том, что наличие базового уровня достаточно, для того чтобы в его рамках его поддерживать и соблюдать, т.е. когда факторы устойчивости базового уровня в нем самом.

Я не думаю, что базовый уровень нужно и можно насильственно внедрить и организовать, нет, напротив, я думаю, что он – тот, о котором шла речь выше – будет организовываться сам по себе благодаря естественной деятельности людей, и начало этой организации явно видно. Для ускорения и облегчения этого эволюционного процесса, устранения ложных действий и тупиков совсем не лишним является понимание существа его целей и перспектив. С точки зрения высказанного основного постулата желанности и наибольшей эффективности игрового восприятия жизни, устойчивость и самодостаточность должны иметь место. При этом, по-видимому, самая трудная проблема человечества в осуществлении включения всех видов власти в общую игру, а также наличие стран и народов с большим отставанием в развитии.

Конечно, может быть, что все будет совсем не так, и то, что высказано – лишь гипотеза, которой не суждено осуществиться. Именно поэтому в самом начале рассматриваемая модель названа фантастической. Все же, пожалуй, еще нужно несколько слов сказать об одной особенности процесса самоорганизации общества.

В человеческом обществе аналогом гуморальных и нервных управлений являются осознанные властные управления, более или менее централизованные в виде государства и его структур и институтов. Это управление, как в организме, так и в обществе возникает в процессе их самоорганизации. Властные структуры также необходимы не только для поддержания базового уровня, их полномочия значительно шире и это обстоятельство приводит к проблеме организации дополнительных властных управленческих структур, проблеме включения их в общее игровое взаимодействие. Это ставит вопрос о том, где и как пролегает граница между властным управлением и тем, что можно назвать стихийной и естественной самоорганизацией, для которой требуется только обеспечение базового уровня. Не отвечая на очень непростой вопрос о границе, отметим, что решение проблемы этой дополнительной власти также нужно видеть в надлежащем разделении властей, требовании открытости ее (власти) намерений и действий, ее предсказуемости в пределах возможного, эффективной контролируемости, которые в целом обеспечивают ее включение в общую игру.

46. СИНТЕЗ И ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ КВАЗИИНВАРИАНТНОГО ЛИНЕЙНОГО И НЕЛИНЕЙНОГО УПРАВЛЕНИЯ

Научные и технические открытия лишь иногда воспринимаются современниками с пониманием и восторгом или одобрением. В современном политизированном и конкурирующем мире чаще всего они игнорируются и остаются без внимания, с трудом пробивая себе дорогу. Ярким примером такого открытия может служить идея инвариантного регулятора, высказанная Г.В. Щипановым в публикации 1939 года [49], вызвавшая последующую бурную дискуссию. Идея состояла в том, что регулятор должен не устранять возникшую ошибку, а предотвращать ее, сделать объект управления невосприимчивым к внешним воздействиям, инвариантным по отношению к ним. С тех пор прошло более полувека, а дискуссия и обсуждаемая проблема не получили завершения, хотя идея инвариантного управления время от времени привлекала исследователей, что нашло отражение в справочнике по теории автоматического регулирования, в обстоятельной книге о проблеме Г.В. Щипанова [50], в недавнем цикле работ В.А. Якубовича [51–53] и др. Причину этого можно видеть в резолюции специально созданной академической комиссии и последующих административных мерах. Работа Г.В. Щипанова была оценена как целиком неверная и лженаучная. Г.В. Щипанов был уволен, а ИАТ (Институт автоматики и телемеханики), издавший его работу, подвергся критике [50]. Это была трагедия талантливого человека, специалиста по авиаприборам и гироскопам, по книгам которого учились современники. Это была упущенная возможность создания нового типа регуляторов и систем управления. Косвенно в дальнейшем этому способствовала успешно развивающаяся теория оптимального управления, которая, казалось бы, должна была указать лучшие системы, но инвариантных управлений она обнаружить не могла, поскольку таких, о которых думал Г.В. Щипанов, нет, а есть квазиинвариантные. Это практически одно и то же, но среди квазиинвариантных управлений нет оптимального, и поэтому теорией оптимального управления они не могли быть обнаружены. Это не значит, что теория оптимизации не может быть полезна, а означает только то, что постановка задач оптимизации должна быть существенно изменена. Решающими все же были огульный характер критики без указания конкретной ошибки, непонимание и восприятие математической реализации идеи Г.В. Щипанова как некоего парадокса и, конечно, устрашающие административные меры.

Познакомился с обсуждаемой проблемой я благодаря М.А. Айзерману, который в конце 1940-х годов привез в г. Горький тяжелую кипу статей и документов по этой проблеме и дискуссии, и с согласия, а может быть и по просьбе А.А. Андропова, вручил ее мне. Я заинтересовался проблемой, но не понял, а вскоре тяжело заболел, забыл о ней и вспомнил лишь недавно, после того как совместно со В.Н. Смирновой одолели широко известный парадокс Пенлеве в механике. Несмотря на различие задач, в них много общего. В обоих требуется настороженное отношение к корректности идеализаций, математических выкладок и выводов; в обоих проявляется неожиданная роль малых величин и параметров. В итоге ока-

залось, что обоснование Щипановым возможности идеального, инвариантного регулятора не корректно и он не реализуем, но возможен квазиинвариантный регулятор. Практически это различие едва ли существенно и сегодня не так ошеломляюще ново. Известна нечувствительность скользящих движений релейных систем автоматического регулирования [34], известны высокоточные регуляторы с большим коэффициентом усиления обратной связи [21].

Содержание настоящей главы относится к изучению и синтезу квазиинвариантного управления (не только стабилизации, но и слежения) для произвольного минимального фазового объекта, описываемого математической моделью вида

$$\begin{aligned} A_n(p)x &= -B_m(p)(u + \xi(t)), \\ C_r(p)u &= D_q(p)(x - f(t) + \eta(t)), \end{aligned} \quad (46.1)$$

где $A_n(p)$, $B_m(p)$, $C_r(p)$ и $D_q(p)$ – действительные полиномы от $p = d/dt$ степеней n , m , r , и q соответственно; $\xi(t)$, $\eta(t)$ – ограниченные неизвестные возмущения в объекте и управлении; $f(t)$ – заданная функция отслеживания ($f(t) = 0$ – стабилизация); x и u – одномерные переменные объекта и управления. Зависимости (46.1) имеют направленный характер: x определяется переменными u и $\xi(t)$ первого уравнения; x , $f(t)$ и $\eta(t)$ определяют u согласно второму уравнению. Исследование модели (46.1) включает формулировку условий ее физической реализуемости, установление условий устойчивости, оценки ошибок управления и слежения, выявления роли внутренних возмущений $\eta(t)$ и частичной нейтрализации внешнего возмущения $\xi(t)$, что позволило сформулировать условия квазиинвариантности и возможности сколь угодно точного управления. Вместе с тем высокая точность управления, требующая малости оператора $C_r(p) = \mu$, может привести к недопустимо большим значениям управления при тех или иных непредвиденных, возможно и незначительных, нарушениях. Уйти от этого дефекта квазиинвариантного точного управления можно за счет синтеза его с обычным управлением в соответствующей нелинейной системе, используя возможности обоих подходов к управлению: подходу, основанному на устранении возникшей ошибки управления, и подходу, предотвращающему возникновение этой ошибки.

О физической реализуемости математической модели системы управления. Систему управления (46.1) назовем физически реализуемой, если она может быть адекватно отображена реальной физической системой или в сочетании с вычислениями компьютера. Какими же при этом должны быть входящие в уравнение (46.1) функции и операторы? В дальнейшем под физически реализуемыми понимаются гладкие и кусочно-гладкие ограниченные функции с конечным числом разрывов. При этом разрывы предполагаются как идеализации быстрых и очень быстрых гладких изменений. Относится это прежде всего к функциям $\xi(t)$, $\eta(t)$, затем к $f(t)$ и управлению u , к функции $x(t)$ и, возможно, к некоторым ее производным. Требование физической реализуемости приводит к тому, что корректная математическая модель (46.1) должна удовлетворять некоторым дополнительным условиям. О них и пойдет речь ниже.

Начнем с реализуемости направленного от ξ к x оператора

$$A_n(p)x = -B_m(p)\xi(t). \quad (46.2)$$

Прямые вычисления $B_m(p)\xi(t)$ требуют наличия m производных от ξ , а ξ может не иметь ни одной. Как же вычислить x ? Оказывается, что вычисление и реализация x возможна при $m < n$ и невозможна при $m > n$. Общее решение уравнения (46.2) складывается из общего решения однородного уравнения ($\xi = 0$) и частного решения неоднородного уравнения (46.2) при нулевых начальных условиях. Ради простоты примем корни $(\lambda_1, \lambda_2, \dots, \lambda_n)$ полинома $A_n(p)$ не кратными, запишем решение неоднородного уравнения (46.2) в изображениях $\tilde{x}(p)$ и $\tilde{\xi}(p)$ и соответствующий оригинал

$$\begin{aligned}\tilde{x}(p) &= \frac{B_m(p)}{A_n(p)} \tilde{\xi}(p) = \sum_{s=1}^n C_s \frac{1}{p - \lambda_s} \tilde{\xi}(p), \\ x(t) &= \int_0^t \sum_{s=1}^n C_s e^{\lambda_s(t-\tau)} \xi(\tau) d\tau, \quad C_s = \frac{B_m(\lambda_s)}{A_n'(\lambda_s)}.\end{aligned}\tag{46.3}$$

Напомним, что для этого требуется, чтобы $m < n$. Из соотношения

$$\begin{aligned}\frac{B_m(p)}{A_n(p)} &= \sum_{s=1}^n \frac{C_s}{p - \lambda_s} = \left(\prod_{s=1}^n (p - \lambda_s) \right)^{-1} \left(\left(\sum_{s=1}^n C_s \right) p^{n-1} - \right. \\ &\quad \left. - \left(\sum_{s=1}^n C_s \sum_{i \neq s} \lambda_i \right) p^{n-2} + \left(\sum_{s=1}^n C_s \sum_{i, j \neq s} \lambda_i \lambda_j \right) p^{n-3} + \dots \right)\end{aligned}$$

следует, что при $m < n - 1$ $\sum_{s=1}^n C_s = 0$, при $m < n - 2$ $\sum_{s=1}^n C_s \lambda_s = 0$, при $m < n - 3$ $\sum_{s=1}^n C_s \lambda_s^2 = 0$ и т.д.

Теперь согласно (46.3) вычислим производные $x(t)$:

$$\dot{x}(t) = \sum_{s=1}^n \left(C_s \xi(t) + \int_0^t C_s \lambda_s e^{\lambda_s(t-\tau)} \xi(\tau) d\tau \right),$$

причем при $m = n - 1$ $\sum_{s=1}^n C_s \neq 0$, и второй производной нет, т.к. $\xi(t)$ не дифференцируема. При $m \leq n - 2$ $\sum_{s=1}^n C_s = 0$, и имеется вторая производная

$$\ddot{x}(t) = \sum_{s=1}^n \left(C_s \lambda_s \xi(t) + \int_0^t C_s \lambda_s^2 e^{\lambda_s(t-\tau)} \xi(\tau) d\tau \right),$$

у которой нет производной при $m = n - 2$, поскольку $\sum_{s=1}^n C_s \lambda_s \neq 0$, но она есть при $m \leq n - 2$, т.к. при этом $\sum_{s=1}^n C_s \lambda_s = 0$. Продолжая эти вычисления с уменьшением m , находим, что $x(t)$ имеет $n - m$ производных и не более, если $\xi(t)$ не дифференцируема. Если же $\xi(t)$ имеет $k > 0$ производных, то $x(t)$ при этом имеет $n - m + k$ производных. Отсюда следует, что при $k > 0$ можно находить $x(t)$ и при $m \geq n$.

Применим полученные сведения к математической модели (46.1). Будем исходить из недифференцируемости $\xi(t)$, $\eta(t)$ и управления u , но удовлетворяющим принятым условиям физической реализуемости. Это приводит к некоторым соотношениям между величинами m , n , r и q , а именно, $m \leq n - 1$ для физической реализуемости x и его производной. При требовании физической реализуемости l производных от x $m \leq n - l$. Далее, для регулятора при требовании физической реализуемости управления u $q \leq l + r$, где $l = n - m$ – число производных, требуемых от x или $x - f$. Подчеркнем, что это только необходимые требования к системе управления. Впереди еще требования устойчивости и малости ошибок стабилизации и слежения.

Условия устойчивости. Требование устойчивости состоит в том, что при $\xi = \eta = f = 0$ любые решения x уравнений (46.1) стремятся к нулю с ростом времени t . Исключая из уравнений (46.1) управление, при $\xi = \eta = f = 0$ приходим к уравнению для x вида

$$[C_r(p)A_n(p) + B_m(p)D_q(p)]x = 0, \quad (46.4)$$

и, следовательно, для устойчивости требуется, чтобы все корни полинома от p , стоящего в (46.4) перед x , имели отрицательную действительную часть. Для дальнейшего требуется, чтобы это условие выполнялось при малости полинома $C_r(p)$. В простейшем случае $r = 0$, $C_r(p) = \mu$, где μ – малое число, требование устойчивости состоит в надлежащем выборе знака числа μ и отрицательности действительных частей всех корней полиномов $B_m(p)$ и $D_q(p)$. Для полинома $B_m(p)$ это имеет место в силу минимальной фазовости объекта, а для $D_q(p)$ – это требование к его выбору. Степень произведения полиномов $B_m(p)$ и $D_q(p)$ должна быть не меньше $n - 1$. Весьма привлекателен вариант минимальной степени q , когда $q = n - 1 - m$. При этом требуемое число q производных от x не превосходит их числа, равного $n - m$. Желание ослабить требование на функцию f возможно за счет увеличения r .

Вернемся к определению знака μ и определению корней полинома

$$\mu A_n(p) + B_m(p)D_q(p), \quad (46.5)$$

когда $m + q = n - 1$ и $\mu \rightarrow 0$. При $\mu \rightarrow 0$ ($n - 1$) корней полинома (46.5) степени n стремятся к корням полиномов $B_m(p)$ и $D_q(p)$, а еще один корень λ_n должен стремиться к $-\infty$. Последнее достигается надлежащим выбором знака μ , $\text{sgn } \mu = \text{sgn } \{a_n b_m d_q\}$, где a_n, b_m, d_q – старшие коэффициенты полиномов $A_n(p)$, $B_m(p)$ и $D_q(p)$. Значение действительного корня λ_n , который должен при $\mu \rightarrow 0$ стремиться к $-\infty$, можно задать, выбирая μ из уравнения $\mu A_n(\lambda) + B_m(\lambda)D_q(\lambda) = 0$, причем при $\lambda \rightarrow -\infty$ $\mu \rightarrow 0$, а знак μ совпадает с указанным выше.

Рассмотренный случай $r = 0$ и $q = n - 1 - m$ можно назвать простейшим и привлекательным, но он требует при слежении наличия q производных у функции $x - f(t)$. Как уже указывалось, преодолеть это ограничение можно, повышая степень r полинома $C_r(p)$. При этом вместо q производных требуется $q - r$.

Ошибка стабилизации. В этом и следующем параграфах рассматривается ошибка управления – стабилизации и слежения – в простейшем случае $r = 0$, со-

ответственно $C_r(p) = \mu$, при нулевых начальных условиях или спустя достаточное время функционирования устойчивой системы управления. Из (46.1) при $f(t) = 0$ и $\eta(t) = 0$ в изображениях непосредственно находим, что

$$\tilde{x}(p) = -\frac{\mu B_m(p) \tilde{\xi}(p)}{\mu A_n(p) + B_m(p) D_q(p)}. \quad (46.6)$$

Пусть $\lambda_1, \lambda_2, \dots, \lambda_n$ – корни знаменателя (46.6) и имеет место устойчивость, тогда $(n - 1)$ корней, $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$, имеют отрицательную действительную часть, а λ_n – большое отрицательное число, причем $\lambda_n \rightarrow -\infty$ при $\mu \rightarrow 0$. Ради простоты ограничимся общим случаем не кратных корней. Тогда оригинал $\tilde{x}(p)$ представляется в виде

$$x(t) = \mu \sum_{s=1}^n K_s \int_0^t e^{\lambda_s(t-\tau)} \xi(\tau) d\tau,$$

где

$$K_s = \frac{B_m(\lambda_s)}{\mu A_n'(\lambda_s) (B_m(\lambda_s) D_q(\lambda_s))'}.$$

Предполагая $|\xi(t)| \leq M$ и учитывая ограниченность всех коэффициентов $K_s \lambda_s^{-1}$ некоторой величиной $C < \infty$, непосредственно находим, что

$$|x(t)| \leq \mu K \cdot M, \quad (K < \infty). \quad (46.7)$$

Таким образом, ошибка стабилизации при исходных нулевых начальных условиях и ограниченном возмущении $\xi(t)$ неограниченно уменьшается с уменьшением μ . Можно поинтересоваться, как ведет себя управление u и за счет чего малая ошибка стабилизации.

При нулевых начальных условиях x определяется возмущением ξ . Действительно, из (46.1) при $f = \eta = 0$ находим для изображений x и u , что

$$\tilde{x}(p) = -\frac{B_m(p)}{A_n(p)} (\tilde{u} + \tilde{\xi}), \quad \tilde{u}(p) = \frac{1}{\mu} D_q(p) \tilde{x}, \quad (46.8)$$

откуда

$$\tilde{x}(p) = -\frac{\mu B_m(p) \tilde{\xi}(p)}{\mu A_n(p) + B_m(p) D_q(p)}.$$

Управление u также определяется возмущением ξ :

$$\tilde{u}(p) = -\frac{B_m(p) D_q(p) \tilde{\xi}(p)}{\mu A_n(p) + D_q(p) B_m(p)} = -\tilde{\xi}(p) + \frac{\mu A_n(p) \tilde{\xi}(p)}{\mu A_n(p) + D_q(p) B_m(p)}, \quad (46.9)$$

причем так, что при $\mu \rightarrow 0$ $u \rightarrow -\xi$, нейтрализуя, согласно (46.8), воздействие ξ на x . Это справедливо в предположении, что последнее слагаемое в формуле (46.9) мало, что может не соблюдаться в точках разрыва или очень больших производных от ξ . Нейтрализация возмущения ξ управлением u может нарушаться при больших, не подавляемых малым μ , величинах производной от ξ . Обнаруживается это, исходя из записи последнего слагаемого в (46.9) в виде:

$$\frac{\mu A_n(p)}{p(\mu A_n(p) + D_q(p)B_m(p))} p\tilde{\xi}(p),$$

где первый множитель неограниченно уменьшается с $\mu \rightarrow 0$, а второй, $p\tilde{\xi}(p)$, может при некоторых t иметь очень большие значения. Вместе с тем это не мешает стремлению ошибки стабилизации x неограниченно уменьшаться вместе с уменьшением μ при ограниченности ξ , поскольку весьма кратковременные скачки производной от $\xi(t)$ существенного влияния не оказывают, а вне скачков производная от $\xi(t)$ ограничена.

Ошибка слежения. Рассмотрим теперь возможности отслеживания переменной $x(t)$ квазиинвариантной системы (46.1) заданной функции $f(t)$. В силу линейности модели (46.1) $x(t)$ представляется суперпозицией воздействий $\xi \neq 0, f = \eta = 0, f \neq 0, \xi = \eta = 0$ и $\eta \neq 0, \xi = f = 0$. Первый случай был рассмотрен, и при $\mu \rightarrow 0 x \rightarrow 0$. Нам надлежит рассмотреть второй. При нулевых начальных условиях в изображениях из (46.1) после исключения \tilde{u} находим, что

$$\mu A_n(p)\tilde{x} + B_m(p)D_q(p)(\tilde{x} - \tilde{f}) = 0$$

или

$$[\mu A_n(p) + B_m(p)D_q(p)](\tilde{x} - \tilde{f}) = -\mu A_n(p)\tilde{f}. \quad (46.10)$$

Нас интересует, как ведет себя ошибка слежения $x - f$, изображение которой, согласно (46.10), имеет вид

$$\tilde{x} - \tilde{f} = -\frac{\mu A_n(p)}{\mu A_n(p) + B_m(p)D_q(p)} \tilde{f}. \quad (46.11)$$

Возникшая ситуация аналогична той, которая была при нарушении стабилизации x возмущением ξ с управлением u в отношении компенсации управлением u возмущения ξ . Достижение малости ошибки слежения $x - f$ за счет малости μ требует ограниченности производной от f . Соответствующую оценку можно получить, записав (46.11) в виде

$$\tilde{x} - \tilde{f} = -\frac{\mu A_n(p)}{p(\mu A_n(p) + B_m(p)D_q(p))} p\tilde{f},$$

из чего следует, что

$$|x - f| < \mu K \max |f'(t)|, \quad K < \infty. \quad (46.12)$$

Из полученной оценки (46.12) и ранее полученной (46.8) следует, что система управления, описываемая (46.1), при ограниченности $f'(t)$ и $\xi(t)$ может сколь угодно точно отслеживать переменную x заданную функцию $f(t)$. Вместе с тем при $r = 0$ физическая реализуемость управления u и устойчивость требуют, чтобы $f(t)$ имела $q = n - 1 - m$ производных.

Влияние внутренних помех. Речь идет об учете помехи $\eta(t)$ при $\xi = f = 0$, т.е. при $u = \mu^{-1}D_q(p)(x + \eta(t))$. В этом случае согласно (46.1) x определяется уравнением

$$[\mu A_n(p) + B_m(p)D_q(p)]x = -B_m(p)D_q(p)\eta.$$

Непосредственно видно, что малость μ не влечет уменьшения ошибки стабилизации. Требуется малость η . Величина η является реальным ограничением точности

стабилизации и слежения. Для реализации второго уравнения (46.1), определяющего управление u , требуется измерение x . Погрешность этого измерения является одной из возможных составляющих помехи η . Очевидно, что измерение x не может быть менее точным, чем требуемая точность управления.

Компьютерный эксперимент и переменная (нелинейная) стратегия управления. Ниже приводятся краткие описания компьютерных экспериментов, проведенных И.С. Гельфер для двух объектов

$$\begin{aligned}\ddot{x} \pm 1,5x &= -u - \xi(t), \\ \ddot{\ddot{x}} - \ddot{x} + 2\ddot{x} + 3\dot{x} - 2x &= -u - \xi(t).\end{aligned}$$

В качестве конкретной их реализации можно указать маятник и двойной маятник, управляемые перемещением точки опоры. У первого объекта $n = 2$, $m = 0$; у второго $n = 4$, $m = 0$. Для регулятора в обоих случаях $r = 0$ и, в соответствии с изложенным ранее, $q = 1$ и $q = 3$. Полиномы $D_q(p)$ выбираем устойчивыми. Для первого объекта

$$u = \mu^{-1}(\dot{x} + 2x),$$

для второго

$$u = \mu^{-1}(\ddot{\ddot{x}} + 6\ddot{x} + 11\dot{x} + 6x).$$

Параметр μ и функции $\xi(t)$, $f(t)$, $\eta(t)$ варьировались. В случаях стабилизации ($f(t) = 0$) и слежения ($f(t) \neq 0$) без учета внутренних помех $\eta(t) = 0$, как при гладких, так и кусочно-гладких $\xi(t)$ и $f(t)$ и нулевых начальных условиях, имело место уменьшение ошибки управления x_{\max} вместе с уменьшением μ в пределах от 0,1 до 0,0001 весьма близко к линейной зависимости от μ . Такие же эксперименты, но при учете внутренних помех, $\eta(t) \neq 0$, обнаружили, что ошибка управления уменьшается с убыванием μ лишь до некоторого предела, естественно, тем меньшего, чем меньше $\eta(t)$. Сказанное относится к управлению при нулевых начальных условиях. При ненулевых начальных условиях величина управления может оказаться недопустимо большой. Обойти эту неприятность можно, меняя стратегию управления u в зависимости от величины x . Простейший, но не наилучший вариант состоит в том, чтобы менять μ в зависимости от $|x|$,

$$\mu = \begin{cases} 0,001 & \text{при } |x| \leq 0,1, \\ 0,1 & \text{при } |x| > 0,1 \end{cases}$$

или

$$\mu = \frac{1 + 10|x|}{1 + 1000|x|}.$$

В первом случае после переходного процесса, приводящего к $|x| \leq 0,1$, все происходило так же, как и в описанном выше эксперименте. Второй вариант переменной стратегии управления также не допускал больших значений и спустя некоторое время приводил к малым значениям $|x|$. Опираясь на эти данные, можно указать достаточно хороший вариант меняющегося управления, в котором для уменьшения состояния (некоторого его функционала) используется оптимальное по этому функционалу управление, а при допустимых $|u|$, требующих малости x , точнее, малости $D_q(p)(x - f)$, используется квазиинвариантное управление.

Ошибался ли Щипанов? Почему идеальный регулятор Г.В. Щипанова воспринимается как парадокс? Как уже отмечалось, идеальный регулятор был признан лженаучным, но вместе с тем конкретная ошибка не была указана. Рассуждения Щипанова были очень просты. В простейшем случае, исключая из уравнений (46.1) управление u , приходим к одному уравнению вида

$$C_r(p)A_n(p) + B_m(p)D_q(p)]x = -C_r(p)B_m(p)\xi, \quad (46.13)$$

из которого при $C_r(p) \equiv 0$ следует, что при устойчивости полинома $B_m(p)D_q(p)$ $x \rightarrow 0$ и влияние возмущения ξ устранено. Парадоксальность вывода в том, что при $x = 0$ управление отсутствует, хотя на объект действует возмущение ξ . Сегодня было бы обращено внимание еще на то, что уравнение (46.13) при $C_r(p) \rightarrow 0$ сингулярно возмущенное и устойчивости $B_m(p)D_q(p)$ недостаточно. Нужно, чтобы корни, стремящиеся к бесконечности при $C_r(p) \rightarrow 0$, имели отрицательные действительные части, еще лучше, чтобы действительные части их стремились к $-\infty$. Но сингулярно возмущенные дифференциальные уравнения были изучены значительно позже, а в то время о них ничего не знали. При $C_r(p) = 0$ управление исчезает, а при $x = 0$ и $C_r(p) \equiv 0$ u есть неопределенность вида $0/0$, и математическая модель некорректна. В рамках формальной математики эта неопределенность может быть устранена, если принять, что

$$\frac{0}{0} = \lim_{C_r(p) \rightarrow 0} \frac{D_q(p)x}{C_r(p)}.$$

Но это возможно только для математической модели, а не для реальной системы управления. При $C_r(p) = 0$ подправленная математическая модель неадекватна реальной физической, и это приводит к тому, что инвариантный регулятор нереализуем. Вместе с тем в рамках модели (46.1) реализуемы при $\eta \equiv 0$ квазиинвариантные сколь угодно точные регуляторы. Значительно позднее было обнаружено, что при некоей дополнительной информации о помехе ξ , уже не в рамках модели (46.1), можно реализовать инвариантное управление [51–53]. Таким образом, формально математически, а именно этого придерживался Г.В. Щипанов, в синтезе «идеального регулятора», ошибки нет. Ошибка в неадекватности в одной единственной точке математической модели и ее непосредственной физической реализации. Неадекватность вызвана тем, что допускается умножение на нуль правой и левой частей уравнения (46.13).

Конечно, Г.В. Щипанов опередил свое время, а дискуссия задержала разработку и исследование его идеи, но все же, уже из других соображений, кое-что было реализовано в релейных системах и системах управления с большим коэффициентом обратной связи, обеспечивающей высокую статическую точность.

Особенностью квазиинвариантной системы управления является также наличие большого коэффициента усиления, поскольку малые значения ошибки управления должны привести к компенсации сравнительно большого внешнего возмущения ξ . В скользящем режиме релейной системы с бесконечным усилением она нечувствительна к достаточно малым внешним возмущениям, а большой коэффициент обратной связи обеспечивает высокую точность стабилизации.

Два подхода к проблеме управления и возможность их синтеза в нелинейной системе. Под управлением понимается стабилизация и слежение. Первым исторически сложился подход к реализации управления на основе уменьшения его ошибки. Для регуляторов прямого действия уменьшение ошибки управления и обеспечение устойчивости были антагонистичны. В дальнейшем эта несовместимость уменьшалась введением непрямого регулирования, астатического, издромного и, наконец, на основе теории оптимального управления. Уже отмечалось, что непосредственно оптимальное управление не могло обнаружить квазиинвариантные управления, но это возможно при существенном изменении постановки задачи, при учете новых факторов помех управления. Второй подход, предложенный Г.В. Щипановым, основан на компенсации внешних возмущений. Его реализация требует больших коэффициентов усиления, что может приводить к недопустимо большим значениям управляющих воздействий. Кроме того, большой коэффициент усиления требует точности измерения x , превосходящей точность управления, малости других помех, агрегированных в возмущении $\eta(t)$, которые должны быть значительно меньше требуемой точности управления (малости x или $|x - f|$). В рамках линейных систем управления преодолеть эти неприятности невозможно, но это возможно при нелинейной стратегии управления. Нелинейная стратегия управления позволяет объединить и синтезировать возможности первого подхода с возможностями второго, когда, например, u заменяют на $a \cdot \text{th}(bu)$. В некоторой мере в простейшем варианте это объединение уже имеет разработанную математическую базу в теории абсолютной устойчивости.

Возможен и более изощренный путь устранения «аварийной ситуации» с $|u| > u_{\max}$, убыстряющий ее ликвидацию и возврат к режиму точного управления с $|x| < \varepsilon$, но при $\eta = 0$ или достаточно малом η . Он состоит в переходе к другому управлению, осуществляемому как бы отдельно для решения однородного уравнения с $\xi = 0$ и решения с нулевыми начальными условиями при реальном ξ . Первое осуществляется классическими методами теории управления, второе – исходное квазиинвариантное управление.

Следует отметить, что сказанное о модели (46.1) допускается для более общего вида модели (46.1) и на модели с несколькими управляемыми переменными и несколькими управлениями. Речь идет об обобщении условий физической реализуемости, требований устойчивости и оценки ошибок управления. Возможно и более широкое рассмотрение в плане объединения методов оптимизации по состоянию с квазиинвариантным управлением, в синтезе подлежащей нелинейной системы, что требует расширения и изменения постановки задачи управления [17].

В заключение приведем примеры квазиинвариантного нелинейного управления в случае, когда $B_m(p)$ и $D_q(p)$ – гурвицевы полиномы, $C_r(p) = \mu$. Допустимые значения μ , при которых решение системы устойчиво, могут быть найдены из D -разбиения по параметру μ характеристического полинома $\mu A_n(p) + B_m(p)D_q(p)$. Эти требования после переходного процесса обеспечивают в дальнейшем малость ошибки управления $|x|$. Однако переходный процесс при запуске или сбое не всегда реализуем, так как может потребовать слишком больших

значений управления u , в то время как для поддержания необходимой малости $|x|$ в установившемся процессе нужно, чтобы $|u|$ был лишь не меньше $\max |\xi(t)|$. Разрешить этот конфликт можно изменением коэффициента усиления $|1/\mu|$, меняя μ в зависимости от $|D_q(p)x|$, как это показано на рис. 46.1, где изменение μ происходит при достижении величиной $|\mu^{-1}D_q(p)x|$ значения u_{\max} – максимально допустимого значения $|u|$. При этом управление будет определяться из уравнения $u = \mu_i^{-1}D_q(p)x$, $i = 1, 2, 3$.

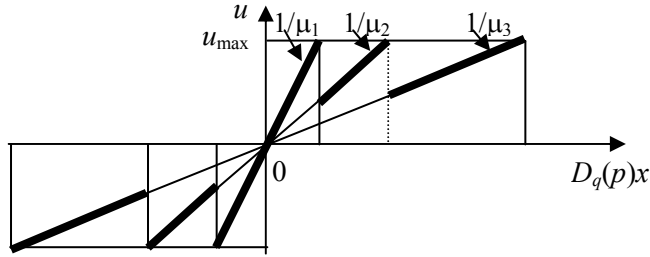


Рис. 46.1. Изменение коэффициента усиления $|1/\mu|$ в зависимости от $|D_q(p)x|$

Ниже описанная возможность уменьшения максимального значения $|u|$ иллюстрируется результатами компьютерных экспериментов для устойчивых и неустойчивых объектов. При этом обнаруживаются не только уменьшение допустимой величины u_{\max} , но и ускорение переходного процесса возврата к требуемой точности управления.

Пример 1. В системе управления объектом вида (46.1)

$$A_4(p) = p^4 - p^3 + 2p^2 + 3p - 2, \quad B_0(p) = 1,$$

$$D_3(p) = p^3 + 6p^2 + 11p + 6, \quad f(t) \equiv 0, \quad \eta(t) \equiv 0,$$

$\xi(t)$ – произвольное ограниченное внешнее воздействие, $|\xi(t)| < \xi_{\max} = 30$. Путем численного построения кривой $N_\omega = -D_3(i\omega)/A_4(i\omega)$ определяем, что допустимые значения μ , обеспечивающие устойчивость работы системы, лежат в интервале $0 < \mu < \mu^* \approx 0,345$.

На рис. 46.2 приведены графики поведения $x(t)$ при воздействии $\xi(t)$ при нулевых начальных условиях и линейном управлении $u = \mu^{-1}D_3(p)x$. График $x(t)$ при $\mu = 0,001$ изображен сплошной кривой, а при $\mu = 0,01$ – пунктирной. Как видно, $|x(t)|$ тем меньше, чем меньше μ и чем быстрее меняется $\xi(t)$. При этом управление $|u|$ в момент перехода $\xi(t)$ на средний, быстро меняющийся участок достигает значений 10^2 при $\mu = 0,01$ и 10^3 при $\mu = 0,001$. При описанном выше кусочно-линейном управлении (см. рис. 46.1) при $u_{\max} = 40$, $\mu_1 = 0,001$ и $\mu_2 = 0,1$ система работает, $|u|$ нигде не превосходит 40 и достигнута точность управления порядка 10^{-3} .

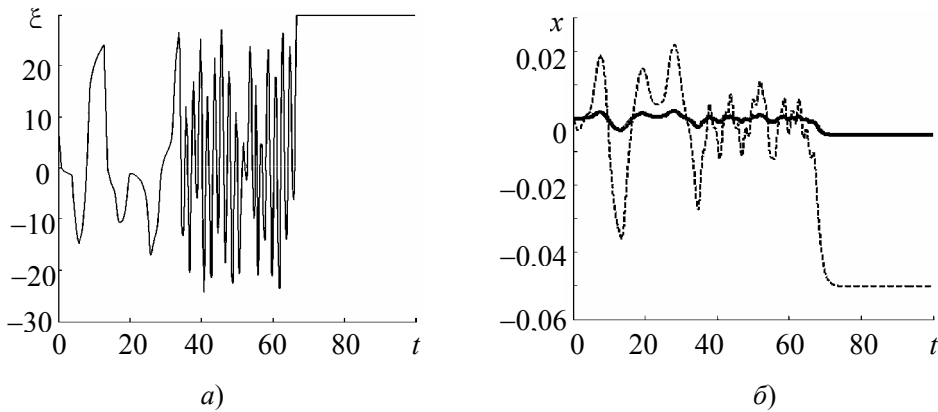


Рис. 46.2. Воздействие $\xi(t)$ (а) и поведение $x(t)$ (б) для этого воздействия при нулевых начальных условиях и линейном управлении

Зададим начальные условия $x_0 = 5$, $\dot{x}_0 = \ddot{x}_0 = 2$, $\ddot{x}_0 = 1$. При использовании линейного управления со значением $\mu = 0,33$, близким к μ^* , оказалось, что нельзя добиться малости $|x|$. Точность управления порядка 10^{-3} достигается при $\mu = 0,001$, но при этом $|u|$ достигает значений 10^5 . Проблема разрешается введением кусочно-линейного управления с $u_{\max} = 200$ и двумя участками с $\mu_1 = 0,001$, $\mu_2 = 0,33$. График $x(t)$ представлен на рис. 46.3,а пунктирной кривой. Здесь и на некоторых других рисунках сплошными линиями изображен график увеличенного в 10 раз значения параметра μ . Сначала выбирается $\mu = \mu_2$, что определяется начальными условиями, и остается ему равным в течение некоторого времени за исключением конечного числа мгновенных переключений на $\mu = \mu_1$ (на графике моменты переключений изображены точками). После переходного процесса устанавливается точность управления порядка 10^{-3} при $\mu = \mu_1$.

На рис. 46.3,б приведены графики $x(t)$ и μ при кусочно-линейном управлении с тремя участками, $\mu_1 = 0,001$, $\mu_2 = 0,1$, $\mu_3 = 0,33$. Видно, что добавление промежуточного участка с $\mu_2 = 0,1$ делает переходный процесс более быстрым и плавным.

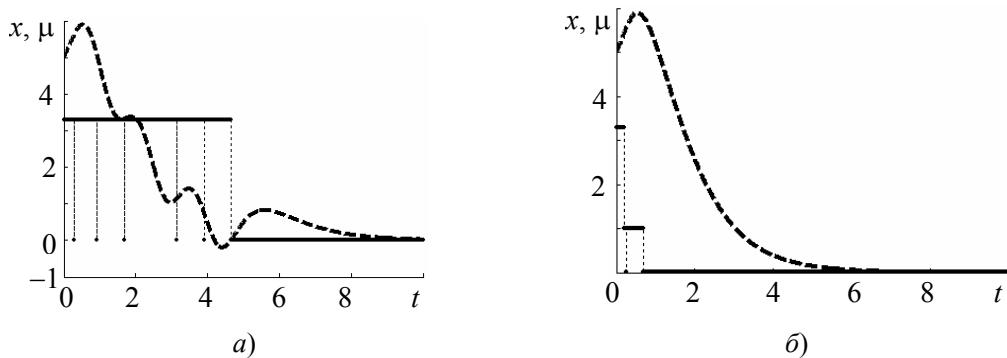


Рис. 46.3. Зависимости $x(t)$ и $\mu(t)$ при кусочно-линейном управлении: а) с двумя участками; б) с тремя участками

На рис. 46.4 представлено управление той же системой при нулевых начальных условиях с $u_{\max} = 80$, двух участках управления с $\mu_1 = 0,001$ и $\mu_2 = 0,05$ и $\xi(t)$, испытывающем в некоторый момент времени скачок величины $3\xi_{\max}$ (график $\xi(t)$ изображен сплошной тонкой кривой). Графики $x(t)$ (пунктирная кривая) и μ (горизонтальные участки) приведены увеличенными в 100, а график $\xi(t)$ уменьшенным в 10 раз. В момент скачка μ переключается с μ_1 на μ_2 и остается равным μ_2 вплоть до окончания действия скачка, затем μ вновь переключается на μ_1 . Значение $|x|$ при скачке возрастает до 0,1, а потом постепенно убывает. Точность управления после скачка вновь устанавливается порядка 10^{-3} , как это было до скачка. Отметим, что управление с использованием лишь одного участка с $\mu_1 = 0,001$ осуществить не удастся. При управлении с тремя участками ($\mu_1 = 0,001$, $\mu_2 = 0,01$, $\mu_3 = 0,05$) несколько улучшается переходный процесс.

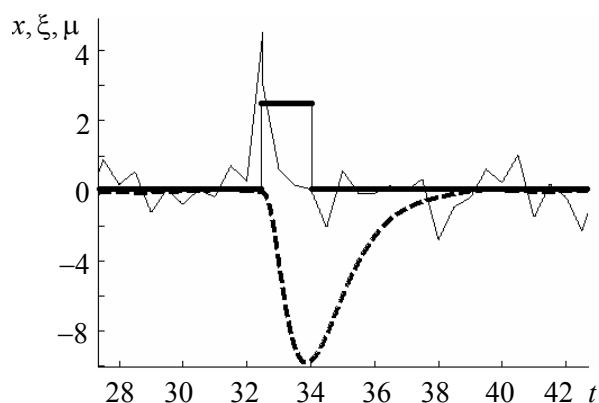


Рис. 46.4. Зависимости $x(t)$, $\xi(t)$ и $\mu(t)$ при кусочно-линейном управлении с двумя участками

Пример 2. Рассмотрим систему управления колебаниями маятника (обычного и перевернутого) с перемещающейся точкой опоры, описываемую дифференциальными уравнениями $(p^2 \pm \omega^2)x = -u - \xi$ (знак «+» соответствует обычному маятнику, «-» – перевернутому), где x – угол отклонения оси маятника от вертикали, $\omega^2 = g/l$, $\xi(t) = -(f(t) + ml\ddot{S}(t))/ml^2$, m – масса маятника, l – расстояние от точки опоры, $S(t)$ – произвольное смещение точки опоры, $f(t)$ – произвольная внешняя сила, $u = M/ml^2$ – управление с помощью момента M . Управление возьмем в виде $u = (ap + b)x/\mu$ так, чтобы полином $D_1(p) = ap + b$ был гурвицев. При $a = 1$, $b = 2$, $\omega^2 = 1,5$ модель примет вид

$$(p^2 \pm 1,5)x = -u - \xi, \quad \mu u = (p + 2)x.$$

Устойчивость полинома $\mu A_2(p) + D_1(p)$ для обычного маятника обеспечивается при $\mu > 0$, а для перевернутого при $0 < \mu < 1,3333\dots$

При ограниченном внешнем воздействии $|\xi(t)| < \xi_{\max} = 10$, с $u_{\max} = 30$ и начальными условиями $x_0 = 5$, $\dot{x}_0 = 2$ осуществить управление системой с точно-

стью порядка 10^{-3} с помощью только одного участка не удастся. Используем кусочно-линейное управление вида, изображенного на рис. 46.1, с двумя участками $\mu_1 = 0,001$, $\mu_2 = 0,5$. Для перевернутого маятника графики $x(t)$ (пунктирная линия) и μ (горизонтальные участки) представлены на рис. 46.5,а. Сначала выбирается участок $\mu = \mu_2$, чтобы преодолеть начальные условия, а потом участок μ_1 . После переходного процесса $|x|$ имеет малость порядка 10^{-3} . При управлении с тремя участками с $\mu_1 = 0,001$, $\mu_2 = 0,1$ и $\mu_3 = 0,5$ (рис. 46.5,б) ускоряется переходный процесс.

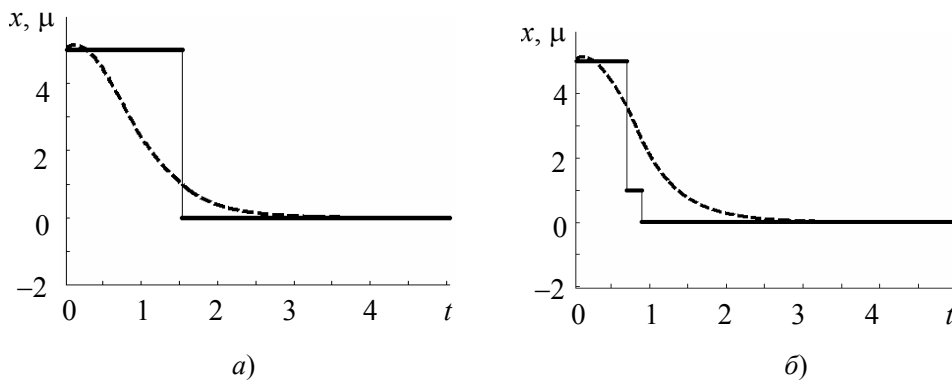


Рис. 46.5. Зависимости $x(t)$ для перевернутого маятника при кусочно-линейном управлении: а) с двумя участками; б) с тремя участками

При нулевых начальных условиях и $\xi(t)$, испытывающем скачок величины $3\xi_{\max}$ (вид $\xi(t)$ изображен тонкой линией на рис. 46.4) управление перевернутым маятником с точностью порядка 10^{-3} при $u_{\max} = 30$ осуществляется с помощью двух участков с $\mu_1 = 0,001$ и $\mu_2 = 0,1$. При увеличении же скачка $\xi(t)$ до $4\xi_{\max}$ и до $5\xi_{\max}$, чтобы осуществить требуемое управление, надо увеличить μ_2 соответственно до 0,5 и 0,9.

На рис. 46.6 представлено кусочно-линейное управление обычным маятником с точностью порядка 10^{-3} при $\mu_1 = 0,001$, $\mu_2 = 1$ и более длительном выбросе по времени $\xi(t)$ до $3\xi_{\max}$.

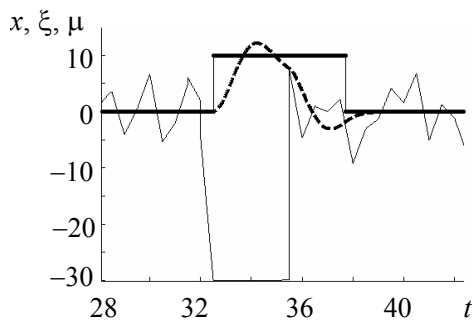


Рис. 46.6. Кусочно-линейное управление обычным маятником; обозначения кривых те же, что и на рис. 46.4

Увеличение $|x|$ до 12 происходит во время выброса $\xi(t)$, по окончании действия которого вновь устанавливается точность управления порядка 10^{-3} . Значение μ при этом переключается с μ_1 на μ_2 и вновь становится равным μ_1 после преодоления выброса $\xi(t)$.

47. АБСТРАКТНАЯ И РЕАЛЬНАЯ МАТЕМАТИЧЕСКИЕ ТЕОРИИ ДИНАМИЧЕСКИХ СИСТЕМ

В настоящее время динамическая система – одна из основных математических моделей. Математическая теория динамических систем, которую разрабатывали величайшие математики XIX и XX веков, очень сложна и обширна. Полное изучение этой теории едва ли необходимо для прикладного пользователя, но ознакомиться с ее основами, безусловно, целесообразно. Краткое изложение современного состояния теории динамических систем является целью этой главы.

Ко времени рождения теории динамических систем господствующее положение в науке занимала идея детерминизма Лапласа: то, что происходит сегодня, определяется тем, что было вчера, а то, что будет завтра, определяется тем, что есть сегодня. Универсальной математической моделью, вытекающей из идеи детерминизма, стала динамическая система.

Теория динамических систем ведет свое начало от А. Пуанкаре, от его труда «Новые методы небесной механики», появившегося в конце XIX века. Пуанкаре создал классическую теорию динамических систем. Большой вклад в развитие теории внес Д. Биркгоф, с появлением его книги «Динамические системы» эта теория стала достоянием мировой науки. Биркгоф, в частности, ввел классификацию всех возможных движений в такой системе. Свой след в развитии теории динамических систем оставили такие величайшие математики, как Пуассон, А.М. Ляпунов, С. Смейл и др. Исчерпывающим источником знаний о динамических системах в России долгое время была книга В.В. Немыцкого и В.В. Степанова «Качественная теория дифференциальных уравнений». Далее изучение динамических систем широко распространилось как в теоретических, так и в прикладных исследованиях. Однако в исследовании конкретных динамических систем успех был достигнут только для размерности, не превышающей второй порядок, который и был подытожен в работах А.А. Андронова с соавторами. Некоторое продвижение в изучении динамических систем больших размерностей все же можно отметить для систем третьего порядка благодаря методу точечных отображений и существенным теоретическим исследованиям в трудах С. Смейла.

Цель исследования динамических систем состоит в определении вида фазовых траекторий и их взаиморасположения, т.е. в построении фазового портрета системы, а также в выяснении зависимости фазового портрета от параметров и в изучении бифуркаций при их изменении. По-видимому, детальное описание фазовых портретов столь разнообразно и огромно, что едва ли оно может быть полным, но можно выделить основные, самые важные характеристики фазового портрета, к которым прежде всего относятся установившиеся движения. В классической теории это устойчивые состояния равновесия, периодические движения, тороидальные образования, многообразия из устойчивых состояний равновесия.

Каковы особенности классической теории динамических систем? Классическая теория базируется на допущении возможности *пренебречь очень малыми воздействиями*. Она основана на двух основных фактах: факте существования решения и его единственности при заданных начальных условиях. Модель, отве-

чающую перечисленным выше условиям, естественно назвать абстрактной или идеальной. Изучению таких абстрактных динамических систем посвящена классическая теория Пуанкаре – Биркгофа. Согласно этой теории существование решения и его единственность представляются достаточными для установления взаимно-однозначного соответствия между фазовыми траекториями и реальными движениями соответствующего объекта, т.е. *в классической теории вычисленные фазовые траектории представляют само движение системы*. То, что фазовые траектории вычислимы и физически реализуемы не вызывало сомнений. При этом само собой подразумевалось, что ошибки счета, как и физические помехи, несущественно малы. Такую трактовку и обоснование можно назвать абстрактными, а классическую теорию – абстрактной теорией динамических систем.

Изучение конкретных динамических систем до появления компьютеров велось в основном аналитическими методами и ограничивалось системами небольшой размерности. При этом получаемые результаты укладывались в рамки классической теории. Ситуация резко изменилась с появлением быстродействующих ЭВМ и возможности наблюдать за поведением фазовых траекторий на дисплее компьютера. Перед исследователями открылся сложный мир фазовых портретов и весьма сложных движений. Появились отдельные примеры, теоретическое исследование которых было затруднено, некоторые конкретные системы с весьма сложными фазовыми портретами не поддавались изучению. Требовались новые подходы. Одновременно обнаруживались явления, которые не находили объяснений. Самым удивительным среди них было обнаружение случайности в поведении фазовых траекторий, когда при одних и тех же начальных условиях могли быть получены совершенно разные решения. Эта случайность обнаружилась и в физическом эксперименте, и при счете на компьютере. Лежащее в основе абстрактной теории утверждение о единственности фазовой траектории с заданными начальными условиями нарушалось, и это не находило объяснения. Возможность сколь угодно точных вычислений всей полутраектории от заданного начального условия и такой же ее физической реализуемости не вызывала сомнений, однако признание абстрактной теорией однозначности, вычислимости и физической реализуемости траекторий приводило к противоречию реального эксперимента с теорией. Остановимся чуть подробнее на результатах исследования особенностей поведения некоторых фазовых траекторий при заданных начальных условиях: с теоретической точки зрения, в плане реализуемости физического процесса и в плане компьютерных вычислений.

1. Абстрактная теория динамических систем предполагает единственность решения для заданных начальных условий. Эта единственность имеет место как для динамических систем, определяемых обыкновенными дифференциальными уравнениями, так и для систем в частных производных, а также для систем, заданных любыми другими способами.

2. Физический эксперимент показывает, что в некоторых случаях реализуемые фазовые траектории с заданными начальными условиями отличаются друг от друга случайным образом, причем различие траекторий может быть весьма значительным. Возможны случаи, когда малейшее изменение начальных условий приводит к существенным изменениям соответствующих фазовых траекторий.

3. Компьютерный эксперимент по вычислению фазовых траекторий при заданных начальных условиях с использованием компьютеров или различных методов счета может приводить к весьма различным траекториям, с некоторого времени ничего общего между собой не имеющим. В таких случаях малейшие изменения начальных условий в пределах, учитываемых компьютером, приводят к аналогичным результатам. Повторный счет на компьютере при одинаковых начальных условиях и одним методом счета приводит к одной и той же траектории.

Эти особенности некоторых фазовых траекторий были восприняты многими исследователями как возможность существования случайных, хаотических или стохастических движений, что находится в явном противоречии с общеизвестной теоремой о единственности решений дифференциальных уравнений и с самим определением динамической системы. Более того, это мнение о наличии случайности в детерминированных системах стало восприниматься как фантастическое открытие, и этому открытию придавалось очень большое значение, а противоречие с безусловно верной теорией игнорировалось. Возникшая вера в наличие случайных движений приняла догматический характер, и этому не помешали не только теорема о единственности решения, но и реальный факт однозначности решения при одинаковом способе счета на компьютере. Вместе с тем, безусловно, весьма значительным был и факт наличия стохастичности при физической реализации динамической системы, и в каком-то смысле именно он одержал победу над теоретическими рассуждениями. После длительного отрицания возможности построения различных траекторий при одних и тех же начальных условиях этот факт был признан как некоторое новое понимание движения динамических систем и был введен новый тип движения, названный хаотическим. В системах второго порядка их не было обнаружено, но в системах третьего порядка и выше они были. Необычайно сложные движения вызвали бум, их открывали всюду. Хаотические движения и хаотические аттракторы получили признание, но их наличие и природа не поддавались объяснению. Определения, что такое хаотический аттрактор, не существовало, и это не удивительно. Ясно, что такое положение требует выяснения истинных причин наблюдаемых явлений.

Истинная причина скрыта именно в базовом для классической теории допущении возможности *пренебрегать очень малыми воздействиями*. Это допущение верно для грубых систем, оно было естественным для систем первого и второго порядка, которые на первых порах были основными объектами, поддающимися изучению. Впрочем, требование вычислительной реализуемости может не выполняться и для двумерных систем, о чем свидетельствует рис. 47.1,а, когда при выборе точки M сколь угодно точные, но разные вычисления могут привести к разным решениям. Правда, этот простой пример реализуется на одномерном отрезке фазовой траектории, и вероятность случайного выбора точки M равна нулю. Аналогичная картина в пространстве представлена на рис. 47.1,б.

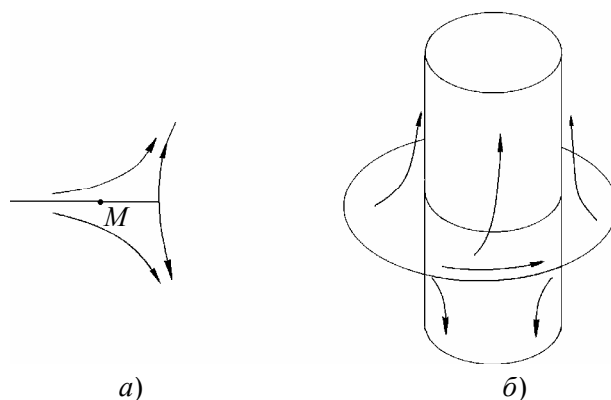


Рис. 47.1. Пример невыполнения требования вычислительной реализуемости: а) для двумерных систем; б) для пространственных систем

В случае многомерной динамической системы такое поведение траекторий может иметь место уже для некоторой области фазового пространства, когда сколь угодно малые отклонения от теоретически принятой идеальной модели динамической системы приводят к значительным изменениям решения за конечное время. Это явление можно назвать *сверхчувствительностью*, и оно наблюдается при экспоненциальной неустойчивости бесконечной фазовой траектории, отвечающей заданному начальному условию. В этом случае имеет место отклонение порядка $\varepsilon e^{\lambda t}$ ($\varepsilon > 0$), и с ростом t это отклонение значительно при сколь угодно малом ε . Этим свойством обладает любая экспоненциально неустойчивая фазовая траектория, в том числе и такие траектории, которые не выходят за пределы некоторой ограниченной области. Подчеркнем, что сверхчувствительность относится как к очень малым неучитываемым возмущениям, так и к погрешностям вычислений. Из сказанного следует, что экспоненциально неустойчивая фазовая траектория не может быть в целом найдена, т.е. *в реальности следует признать факт существования невычислимых и физически нереализуемых фазовых траекторий*. Означает ли это, что фазовая траектория случайна? Компьютер при неизменности метода расчета все время выдает один и тот же результат, не демонстрируя никакой случайности. Физический эксперимент, напротив, при достаточно длительном эксперименте приводит к различным (случайным) результатам. Различия вызваны флуктуациями параметров физической системы и случайными отклонениями, происходящими в реальной системе. При этом при малых возмущениях отклонения могут быть значительными в силу неустойчивости фазовой траектории. Таким образом, для асимптотически неустойчивых движений можно говорить о случайности вычисляемых и физически реализуемых траекторий, но *случайностей, порождаемых динамической системой, нет*. Наблюдаемые случайности – следствие неучета малых, но существенных воздействий.

Почему же так разнятся реализации устойчивых и неустойчивых движений? Прежде всего, заметим, что этих различий нет, если на реализуемую фазовую траекторию нет даже очень малых неизвестных воздействий. При выполнении этого условия состояние системы в точности соответствует фазовой траектории математической модели. Различие проявляется с появлением неизвестных воздействий, когда устойчивость уменьшает их влияние, а неустойчивость увеличивает. При

этом уход от первоначальной фазовой траектории экспоненциально нарастает при неустойчивости, а в случае устойчивости, напротив, экспоненциально убывает. Экспоненциальный рост приводит к тому, что в значительной мере изменение состояния определяется очень малыми неучитываемыми воздействиями. Таким образом, при неустойчивости случайные возмущения могут вызвать изменение состояния и отвечающую ему эволюцию фазовой траектории.

Неустойчивые фазовые траектории могут образовывать притягивающие множества, и тогда можно наблюдать *блуждание* фазовой траектории в некоторой ограниченной области. Притягивающее множество блужданий получило название «странного аттрактора» (странного потому, что устойчивое множество состоит из неустойчивых фазовых траекторий, которые могут пересекаться вследствие сколь угодно малых неучитываемых случайных воздействий). В странном аттракторе фазовая точка перемещается по области притяжения под управлением неучитываемых ничтожно малых случайных неуправляемых воздействий. При этом вычисляемая или физически реализуемая фазовая траектория не имеет определенного вида, задаваемого начальными условиями. И хотя блуждающие движения не вычислимы и физически не реализуемы по начальным условиям, в целом они образуют устойчивые притягивающие множества и в этом смысле похожи на установившиеся устойчивые движения (состояния равновесия, периодические движения и др.). Подчеркну, что блуждание прекращается, как только фазовая точка попадает в область притяжения устойчивой фазовой траектории. Если устойчивость асимптотическая, то дальнейшее движение – это движение по асимптотической траектории. Такое возможно, конечно, при достаточно малых возмущениях. В противном случае, когда приход в область притяжения устойчивого движения невозможен, блуждание не прекращается.

Возникает вопрос о вероятностном описании неустойчивых притягивающих движений. Оказывается, что такое описание иногда возможно и в некоторых случаях легко реализуемо. При этом, естественно, предполагается отсутствие внешних возмущений. Основываясь на изучении точечного отображения аттрактора самого в себя, т.е. исходя из задания динамической системы в виде точечного отображения, можно найти плотность вероятности для этих странных аттракторов. В некоторых простейших случаях вычисление плотностей вероятностей достаточно просто (см. гл. 18). Можно предположить, что блуждающие движения в некоторых случаях подчиняются определенным статистическим законам, но полное изучение всех вариантов мне неизвестно.

Итак, с точки зрения реальной теории, не вдаваясь в детали, все движения в аналитической системе можно разделить на переходные и установившиеся, устойчивые и неустойчивые и, как частный случай неустойчивого движения, движения, блуждающие при наличии притягивающего множества. В свою очередь установившиеся движения делятся на регулярные и блуждающие: асимптотически устойчивые фазовые траектории (регулярные движения) не выходят из конечной области устойчивого состояния равновесия или периодического движения, а асимптотически неустойчивая траектория блуждает в некоторой ограниченной области притягивающего множества (хаотического аттрактора).

После сказанного естественно возникает вопрос, каким образом Пуанкаре, Биркгоф и их последователи проигнорировали возможность различных движений

из одного и того же начального состояния при неучете малых воздействий, не заметили существования невычислимых фазовых траекторий. Возможно, это связано с господствовавшим в то время механистическим взглядом на природу, с тем, что эти «странные» явления в изучаемых механических системах не проявили себя. Известно высказывание английского физика лорда Кельвина, который сто лет назад сказал, что «он не может понять ни одного явления до тех пор, пока не представит себе его механическую модель» (Вольтер Б.В. В царстве неустойчивости // Знание – сила. 1993. №4. С. 113.). Хотя это не значит, что таких движений в механических системах нет. Таких примеров не было, но сейчас они известны, и один из них приведен в этой книге: осциллятор с отрицательным трением и ограничительным торможением. С проявлением неустойчивых движений на практике столкнулись химики, когда начали взрываться химические реакторы. «Предсказать, разглядеть сокрытую опасность неустойчивости далеко не всегда можно «методом здравого смысла». В устойчивости есть некая соразмерность между внешними воздействиями и реакцией системы. Сильное воздействие – активнее реакция. Неустойчивая система не подчиняется такой соразмерности, она может возбудиться от малейшей едва заметной причины» (там же, с. 115). Но не химики создали и развивали теорию динамических систем, а математики в то время были далеки от химии и химических реакторов.

Хочу подчеркнуть, что мир состоит как бы из двух частей: устойчивой и неустойчивой. Изменения, происходящие в устойчивом мире, соответствуют изменяющим его воздействиям. Эволюция устойчивого мира прогнозируема. Изучая процессы, происходящие в том или ином объекте, мы пытаемся найти закономерности его изменения, не сомневаясь, что такие закономерности существуют. Но это не так, точнее, это не всегда так. Окружающий нас мир гораздо сложнее, чем мы думаем. В некоторых случаях изменения происходят из-за очень малых случайных воздействий. В неустойчивом мире нет соответствия: очень малые воздействия могут приводить к громадным изменениям, эволюция неустойчивого мира не прогнозируема. Существующая теория динамических систем относится в основном к устойчивому миру. Неустойчивый мир ждет своего исследования, которое, безусловно, расширит наше понимание окружающего мира.

И в заключение подведем итог имеющимся представлениям о хаотических движениях динамических систем. Так называемая стохастичность движений динамической системы не соответствует ее природе, она есть результат неполноты ее описания. Эта неполнота считалась несущественной в силу ее малости, возможность сверхчувствительности не замечалась и не учитывалась. Но вместе с тем блуждания, вызванные не учитываемыми воздействиями, действительно могут обладать статистическими закономерностями.

СПИСОК ЛИТЕРАТУРЫ

1. Андронов А.А., Горелик Г.С. О резонансных явлениях при движении релятивистской частицы в циклотроне // ДАН СССР. 1945. Т. 11. С. 664.
2. Андронов А.А., Витт А.А., Хайкин С.Э. Теория колебаний. М.: Наука, 1981.
3. Андронов А.А., Леонтович Е.А., Гордон И.И., Майер А.Г. Качественная теория динамических систем второго порядка. М.: Наука, 1966.
4. Андронов А.А., Леонтович Е.А., Гордон И.И., Майер А.Г. Теория бифуркаций динамических систем на плоскости. М.: Наука, 1967.
5. Базыкин А.Д. Система Вольтерра и уравнение Михаэлиса–Менте // В кн.: Вопросы математической генетики. Новосибирск: Изд-во СО АН СССР, 1974. С. 165–173.
6. Баталова З.С. Изучение вибропогружения шпунта при лобовом сопротивлении грунта // Инж. журнал. Механика твердого тела. 1966. Сер. 2. С. 165–173.
7. Баталова З.С. О движениях ротора под влиянием внешней гармонической силы // МТТ. 1967. №1.
8. Баталова З.С., Белякова Г.В., Бухалова Н.В. Периодические движения маятника с колеблющейся осью // МТТ. 1987. №6. С. 18–25.
9. Баталова З.С., Ежевская Н.А. О структуре фазового пространства уравнения Дюффинга // Дифференциальные уравнения. 1989. Т. 25, №7. С. 1270–1271.
10. Баталова З.С., Белякова Г.В. Диаграммы устойчивости периодических движений маятника с колеблющейся осью // ПММ. 1988. Т. 52. Вып. 1. С. 55–63.
11. Баутин Н.Н. Динамическая теория часов. М.: Наука, 1990. 190 с.
12. Беспалова Л.В. К теории виброударного механизма // Инж. журнал. ОТН. 1957. №5.
13. Блехман И.И. Исследование процесса вибрационной забивки свай и шпунтов // Инженерный сборник. 1954. Т. 19. С. 55–64.
14. Вольтера В. Математическая теория борьбы за существование. М.: Наука, 1976.
15. Гаузе Г.Ф., Витт А.А. О периодических колебаниях численности популяций; математическая теория релаксационного взаимодействия между хищниками и жертвами и ее применение к популяции двух простейших // Изв. АН СССР. Отд. матем. и естеств. наук. Сер. VII. 1934. Т. 10. С. 1551–1559.
16. Гаузе Г.Ф. Исследования над борьбой за существование в смешанных популяциях // Зоологический журнал. 1935. Т. 14, № 2. С. 243–270.
17. Гельфер И.С. Нелинейность как расширение возможностей квазиинвариантного управления // Нелинейные колебания механических систем: Труды 8-й Всероссийской науч. конф., Н. Новгород, 22–26 сент. 2008. Н. Новгород, 2008. Т. 1. С. 132–136.
18. Колосов Г.Е., Нежметдинова Д.В. Исследование установившихся колебательных процессов в хемостате // Автоматика и телемеханика. 2000. № 1.
19. Ланда П.С. Нелинейные колебания и волны. М.: Наука, Физматлит, 1997.
20. Неймарк Ю.И. Устойчивость линеаризованных систем (дискретных и непрерывных). Л.: ЛКВВИА, 1949. 140 с.
21. Мееров М.В. Синтез структур автоматического регулирования высокой точности. М.: Физматгиз, 1959.
22. Неймарк Ю.И. О движениях, близких к двоякоасимптотическому движению // ДАН СССР. 1967. Т. 172, № 5. С. 1021–1024.

23. Неймарк Ю.И. Структура движений динамической системы в окрестности гомоклинической кривой // Пятая летняя математическая школа (Ужгород, июнь, 1967) / АН СССР. Киев, 1968. С. 400–435.
24. Неймарк Ю.И. Об одном классе динамических систем // Тр. Международ. симпоз. по нелинейным колебаниям / АН СССР. Киев, 1970. Т. 2. С. 363–376.
25. Неймарк Ю.И. Метод точечных отображений в теории нелинейных колебаний. М.: Наука, 1972. 471 с.
26. Неймарк Ю.И., Баталова З.С., Васин Ю.Г., Брейдо М.Д. Распознавание образов и медицинская диагностика. М.: Наука, 1972.
27. Неймарк Ю.И. Динамические системы и управляемые процессы. М.: Наука, 1978. 336 с.
28. Неймарк Ю.И., Ланда П.С. Стохастические и хаотические колебания. М.: Наука, 1987.
29. Неймарк Ю.И. Робастная устойчивость линейных систем // ДАН СССР. 1991. Т. 319, №3. С. 578–580.
30. Неймарк Ю.И., Смирнова В.Н. К столетию проблемы парадокса Пенлеве // Вестник Нижегородского гос. ун-та. Сер. Математическое моделирование и оптимальное управление. 2001. Вып. 2. С. 7–33.
31. Неймарк Ю.И. Вопросы устойчивости в теории управления в работах научной школы А.А. Андропова // Вестник Нижегородского гос. ун-та им. Н.И. Лобачевского. 2001. Вып. 1(23). С. 168–197.
32. Неймарк Ю.И., Теклина Л.Г. Новые технологии применения метода наименьших квадратов. Н. Новгород: Изд-во Нижегородского университета, 2003.
33. Неймарк Ю.И. Теория вибрационного погружения и вибровыдергивания // Инженерный сборник. 1953. Т. 4. С. 13–48.
34. Неймарк Ю.И. О скользящем режиме и периодических движениях релейных систем // Тр. ГИФТИ и радиофизического факультета ГГУ, сер. физ. Уч. зап. 1956. Т. 30. С. 159–192.
35. Неймарк Ю.И., Смирнова В.Н. Идеализация, сингулярность и парадокс Пенлеве. // Доклады РАН. 2002. Т. 386, № 6. С. 742–745.
36. Неймарк Ю.И., Коган Н.Я., Савельев В.П. Динамические модели теории управления. М.: Наука, 1985.
37. Неймарк Ю.И. Создатель нелинейной динамики и физики. К 100-летию со дня рождения академика А.А. Андропова // Вестник РАН. 2001. Т. 71, № 4. С. 231–236.
38. Неймарк Ю.И. Кризис эволюции жизни и коллективная игра людей // Нелинейный мир, 2007. Т. 5, №1–2. С. 65–70.
39. Неймарк Ю.И., Левин А.Я. Играет ли бог в кости? // Прикладная нелинейная динамика. 2009. №3. С. 98–136.
40. Пуанкаре А. Новые методы небесной механики. Избр. труды. Т. 1. М.: Наука, 1971.
41. Пуанкаре А. Избр. труды. Т. II. М.: Наука, 1972.
42. Пенлеве П. Лекции о механике. М.: Гостехиздат, 1954.
43. Рапопорт А.Н. Автоматные модели поисковой оптимизации и управления. Киров: ВятГТУ, 1999. 132 с.
44. Розенблат Ф. Принципы нейродинамики: перцептрон и теория механизмов мозга. М.: Мир, 1965.
45. Седов Л.И. Методы подобия и размерности в механике. М.: Наука, 1972.
46. Смейл С. Структурно устойчивые гомеоморфизмы с бесконечным числом периодических точек // Тр. Международ. симпоз. по нелинейным колебаниям. Киев: АН УССР, 1963. С. 365–366.

47. Цетлин М.Л. Исследование по теории автоматов и моделированию биологических систем. М.: Наука, 1969.
48. Цетлин М.Л. Конечные автоматы и моделирование простейших форм поведения // УМН. 1963. Т. 18, № 4. С. 5–33.
49. Щипанов Г.В. Теория и методы проектирования автоматических регуляторов // Автоматика и телемеханика. 1939. № 1. С. 49–67.
50. Щипанов Г.В. и теория инвариантности / Сост. З.М. Лезина, В.И. Лезин. М.: Физматлит, 2004. 428 с.
51. Якубович В.А. Управляемые регуляторы в задачах инвариантности и отслеживания // ДАН. 1955. Т. 343, № 2. С. 172–175.
52. Якубович В.А. Синтез стабилизирующих регуляторов, обеспечивающих независимость выходной переменной системы управления от внешнего воздействия // ДАН. 2001. Т. 380. № 1. С. 77–80.
53. Якубович В.А., Проскуряков А.В. Задача об инвариантности систем управления. // ДАН. 2003. Т. 389, № 6. С. 742–746.

Неймарк Юрий Исаакович

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
КАК НАУКА И ИСКУССТВО**

Учебник

Издание второе, исправленное и дополненное

Формат 60x84/8. Печать офсетная. Бумага офсетная.
Гарнитура «Таймс». Уч.-изд. л. 39,5. Усл. печ. л. 48,8.
Тираж 400 экз. Заказ 500

Издательство Нижегородского госуниверситета им. Н.И.Лобачевского
603950, Н. Новгород, пр. Гагарина, 23

Типография Нижегородского госуниверситета им. Н.И. Лобачевского
Лиц. № 18-0099 от 04.05.01 г.
603000, Н. Новгород, ул. Б. Покровская, 37