Как математики изучают хаос

Академик РАН Я. Г. Синай

В августе 2000 года я получил приглашение выступить с циклом Хедриковских лекций на заседании Математической ассоциации Америки. Аудитория состояла из математиков, преподающих в различных университетах США, профессиональные интересы которых разнородны и, как правило, далеки от тематики данных лекций. Этот текст представляет собой введение в математическую теорию хаоса и более или менее соответствует фактическому содержанию прочитанных лекций.

Лекция 1. Хаос и беспорядок. Гиперболичность

В течение последних тридцати – сорока лет теория хаоса стала весьма популярной. В любом книжном магазине, где есть научный отдел, сегодня можно найти множество книг, в которых объясняется, что такое хаос. Это свидетельствует о том, что хаос — интересное и содержательное явление, причем не только для математика, но и для физика, инженера, экономиста, специалиста, работающего в разных областях приложений науки.

На самом деле то, что ныне зовётся хаосом, существовало в математике задолго до наших дней под именем эргодической теории. Однако физики не могли использовать эти слова, ибо термин «эргодичность» напоминал им эргодическую гипотезу и основания статистической механики. Согласно широко распространённой в теоретической физике точке зрения, жизнь человеческая слишком коротка, чтобы тратить ее на обоснования. Насколько я знаю, слова «детерминированный хаос» были впервые произнесены более тридцати лет назад российским физиком Чириковым и покойным профессором Фордом — физиком из Технологического института штата Джорджия. Этот термин понадобился потому, что в те годы появилось много результатов, относящихся к эргодической теории и ее приложениям.

Теория вероятностей также изучает хаос. Казино, азартные игры, карты — классические примеры хаоса, но это — недетерминированный хаос. Теория хаоса использует методы и основные результаты теории вероятностей, не являясь, однако, ее частью. В теории вероятностей всегда предполагают, что имеется случайный механизм, делающий исходы статистических экспериментов непредсказуемыми, и цель теории — изучать статистику будущего развития и делать разумные прогнозы.

В теории хаоса нет случайного механизма. Наоборот, предполагается наличие детерминированной динамики, детерминированных законов движения и т. п. В этом смысле теория хаоса может рассматриваться как часть теории динамических систем. Малые шумы, которые неизбежно присутствуют в конкретных

Математическое просвещение, сер. 3, вып. 5, 2001(32–46)

ситуациях, приводят лишь к эффектам более высокого порядка малости. Законы динамики определяют индивидуальные траектории движения; теория хаоса изучает статистические свойства этих траекторий.

Если задана траектория, можно поставить вопрос, имеется ли у нее некоторое усредненное по времени поведение. Вот типичные вопросы теории хаоса:

a) существуют ли средняя частота затмений в нашей Солнечной системе?

b) является ли Солнечная система устойчивой, иначе говоря, останутся ли расстояния между планетами ограниченными в будущем?

c) можно ли предсказывать погоду?

После того, как у данной динамической системы установлено существование временны́х средних, для типичных в определенном смысле траекторий можно поставить вопрос об уклонениях траекторий от их среднего поведения. С точки зрения теории вероятностей это вопрос о справедливости центральной предельной теоремы в ситуациях, когда случайность (в обычном смысле слова) отсутствует.

Уже это простое обсуждение показывает, что в теории хаоса имеется большой параметр t — время, и нас интересуют различные вопросы, формулируемые в виде предельных теорем при $t \to \infty$. Удивительным открытием пятидесятых – шестидесятых годов было осознание того, что многие явления и идеи, связанные с хаосом, могут относиться к случаю, когда фазовое пространство маломерно и законы динамики выглядят весьма просто. Соответствующую теорию иногда называют малоразмерным хаосом. С другой стороны, имеется множество важных проблем, касающихся систем статистической механики с большим числом степеней свободы. В этих системах два параметра: t — время и N — число степеней свободы, и оба стремятся к бесконечности. Один из важнейших вопросов здесь — о стреле времени и необратимости. Если встать на точку зрения Демокрита, согласно которой окружающая среда состоит из слабо связанных между собой одинаковых частиц, то можно задаться вопросом, в какой мере основанные на этой картине модели отражают направленность времени. Конечно, это очень трудный вопрос, до сих пор остающийся открытым. Приведем пример. Если в некотором объеме имеется газ, содержащий 10 частиц, то в процессе эволюции можно обнаружить моменты времени, когда газ занимает только половину объема. Для 100 частиц вероятность этого события гораздо меньше, хотя и положительна. Необходимо наблюдать динамику существенно дольше, прежде чем подобное событие произойдёт. Мы можем выбрать некоторый порог и считать, что событие, вероятность которого меньше этого порога, никогда не случается. На самом деле это — очень серьезный шаг; но теория описывает результаты эксперимента только после того, как этот порог выбран. Здесь следует вспомнить знаменитый численный эксперимент Ферми – Паста – Улама. Они обнаружили модель, в которой здравый смысл предсказывал потерю памяти о предыдущих состояниях и сходимость к равновесию, а реальное поведение оказалось совсем иным.

Если система со многими степенями свободы находится в термодинамическом равновесии, то это — состояние пространственного беспорядка, поскольку локальные конфигурации частиц не подчиняются никаким детерминистским законам, а имеют устойчивую предельную статистику при $N \to \infty$. Одним из главных достижений теории хаоса стало открытие, что большой параметр t в динамике играет ту же роль, что и большой параметр N в статистической механике. В этом смысле детерминированный хаос есть динамический беспорядок. Соответствующая теория называется в динамике термодинамическим формализмом. Ниже мы приводим пример теоремы из этой теории, относящийся к предмету нашего обсуждения.

Пусть $z = (z_1, \dots, z_d)$ — точка d-мерной решетки, z_i — целые числа. Для большинства наших задач достаточно рассматривать $d = 1$. Но важно, однако, иметь общую картину для произвольного d. Предположим, что $z \in V$, где V — ограниченная область, и в каждой точке $z \in V$ есть переменная $\sigma(z)$, принимающая конечное число значений. Типичным примером является $\sigma(z) = \pm 1$, когда в каждой точке выбирается случайный знак. Рассмотрим функцию $U(\sigma(z);\, \sigma(z'),$ $|z - z'| \leqslant r)$, которую назовём потенциалом взаимодействия переменной $\sigma(z)$ с её соседями в радиусе r. Зафиксируем конфигурацию $\bar{\sigma}(\partial V) = \{\sigma(z) \mid z \notin V,$ $\mathrm{dist}(z, V) \leqslant r\}$. Сумма

$$\sum_{z \in V} U(\sigma(z);\, \sigma(z'), |z - z'| \leqslant r) = U(\sigma(z), z \in V \mid \bar{\sigma}(\partial V))$$

называется энергией конфигурации $\sigma(z), z \in V$ при граничном условии $\bar{\sigma}(\partial V)$. Потом такие конфигурации будут использоваться для кодирования траекторий динамических систем. Типичный пример — так называемая ферромагнитная модель Изинга, где $\sigma(z) = \pm 1$, $r = 1$ и

$$U(\sigma(z);\, \sigma(z'), |z - z'| \leqslant 1) = \sum_{z' : |z - z'| = 1} (\sigma(z) - \sigma(z'))^2 + h\sigma(z).$$

Последняя сумма при $h = 0$ есть дискретный вариант известного интеграла Дирихле

$$\int_V (\nabla\sigma, \nabla\sigma)\, dz.$$

Однако во многих отношениях модель Изинга далека от её непрерывного аналога.

Выражение

$$\pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)) = e^{-\beta U(\sigma(z), z \in V \mid \bar{\sigma}(\partial V))}$$

называется статистическим весом конфигурации $\sigma(z), z \in V$. Параметр β — обратная температура. Введем обозначение

$$\Xi(V \mid \beta, \bar{\sigma}(\partial V)) = \sum_{\{\sigma(z), z \in V\}} \pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)).$$

Распределение Гиббса в объеме V при граничном условии $\bar{\sigma}(\partial V)$ и обратной температуре β — это вероятностное распределение на конфигурациях $\sigma(z)$, $z \in V$, где

$$p(\sigma(z), z \in V) = \frac{1}{\Xi(V \mid \beta, \bar{\sigma}(\partial V))} \pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)).$$

Можно без труда обобщить все предыдущие понятия на случай, когда множество значений $\sigma(z)$ бесконечно или даже континуально, а потенциал U имеет бесконечный радиус взаимодействия r. Такие выражения часто возникают из динамики.

Первый вопрос — о пределе гиббсовского распределения при $V \to \infty$ и зависимости этого предела от граничных условий. Пределы, получающиеся таким способом, могут быть определены с помощью так называемых уравнений Добрушина – Лэнфорда – Рюэля. О ситуациях, когда эти пределы зависят от граничных условий, говорят как о ситуациях, имеющих дальний порядок; он связан с фазовыми переходами первого рода. Сейчас есть разнообразные методы доказательства существования дальнего порядка. В частности, дальний порядок есть в упомянутой выше модели Изинга при $d > 1$, $h = 0$ и достаточно больших β. Однако задачи о фазовых переходах достаточно далеки от теории хаоса. Причина в том, что дальнего порядка заведомо нет при $d = 1$, потому что в этом случае распределения Гиббса — марковские цепи с конечной памятью. Впрочем, кое-что интересное может происходить, если $r = \infty$. Рассмотрим пример теоремы, имеющей приложения в динамике и теории хаоса.

Предположим, что задана функция $p(\sigma(0) \mid \sigma(-1), \sigma(-2), \ldots, \sigma(-n), \ldots)$, удовлетворяющая условиям:

a) $p(\sigma(0) \mid \sigma(-1), \sigma(-2), \sigma(-3), \ldots, \sigma(-n), \ldots) > \text{const} > 0;$

b) $\displaystyle\sum_{\sigma(0)} p(\sigma(0) \mid \sigma(-1), \sigma(-2), \sigma(-3), \ldots, \sigma(-n), \ldots) = 1.$

Тогда $p(\sigma(0) \mid \sigma(-1), \sigma(-2), \ldots, \sigma(-n), \ldots)$ можно рассматривать как условную вероятность $\sigma(0)$ при заданной предыстории $\sigma(-1), \sigma(-2), \ldots, \sigma(-n), \ldots$ Зададимся вопросом: как много вероятностных распределений P совместимы с этой условной вероятностью? Последнее означает, что

$$\lim_{n\to\infty} \frac{P(\sigma(0), \sigma(-1), \ldots, \sigma(-n))}{P(\sigma(-1), \ldots, \sigma(-n))} = p(\sigma(0) \mid \sigma(-1), \ldots, \sigma(-n), \ldots).$$

Если p зависит только от конечного числа переменных, т. е. $p = p(\sigma(0) \mid \sigma(-1), \sigma(-2), \sigma(-3), \ldots, \sigma(-n))$, то обычная теория цепей Маркова говорит, что P единственно. Для того, чтобы единственность имела место и при $n = \infty$, функция p должна достаточно хорошо приближаться функциями конечного числа переменных. Положим

$$\alpha_n = \sup_{\sigma,\sigma',\sigma''} \left| \frac{p(\sigma(0) \mid \sigma(-1), , \ldots, \sigma(-n), \sigma'(-n-1), \ldots)}{p(\sigma(0) \mid \sigma(-1), , \ldots, \sigma(-n), \sigma''(-n-1), \ldots)} - 1 \right| .$$

Теорема. *Если $\sum_n \alpha_n < \infty$, то P единственно.*

Если $\sum_n \alpha_n = \infty$, то, вероятно, может проявиться эффект дальнего порядка и P может стать неединственным.

В задачах теории хаоса P обычно единственно. Следующее, что нам нужно, — это так называемые временны́е корреляционные функции $E_p(\sigma(0)\sigma(z))$ и скорость их сходимости к пределу при $z \to \infty$.

Теория хаоса состоит из двух основных частей. В первой из них главная задача — доказать, что некоторая динамическая система (или класс динамических систем) неустойчива или гиперболична. Вторая — это теория, показывающая, что неустойчивые динамические системы имеют особые свойства, которые можно ассоциировать с хаосом.

Идея о связи неустойчивости и хаоса прослеживается уже в статьях и книгах Пуанкаре (особенно в его учебнике по теории вероятностей, где он обсуждает происхождение случайных событий), в работах Адамара, Биркгофа, Морса.

В более близкие к нам времена (тридцать – сорок лет назад) эта тема была центральной в работах физиков Н. С. Крылова, Б. Чирикова, Г. Заславского, Дж. Форда и др. В математике топологические аспекты этой идеи развивались С. Смейлом, Д. Аносовым, М. Шубом, Дж. Пейлисом и др. Математическая теория хаоса появилась в работах Д. Рюэля, Р. Боуэна, В. Оседедца, Я. Песина, моих и др.

Вывод основных свойств хаоса из неустойчивости является сейчас хорошо разработанной математической теорией. Мы обсудим ее в лекции 3. Во многих случаях основная трудность состоит в том, чтобы для данной динамической системы доказать достаточно сильную неустойчивость. В этой лекции мы обсудим общее понятие динамической неустойчивости.

Пусть мы имеем систему обыкновенных дифференциальных уравнений

$$\frac{dx_i}{dt} = f_i(x_1, \ldots, x_d), \quad 1 \leqslant i \leqslant d, \tag{1}$$

здесь d — размерность фазового пространства. Функции f_i достаточно гладкие и регулярные, чтобы для любого начального условия $x(0) = (x_1(0), \ldots, x_d(0))$ система (1) имела единственное решение $x(t) = (x_1(t), \ldots, x_d(t))$, определенное для всех t, $-\infty < t < \infty$.

Рассмотрим малое возмущение начальных условий $x'(0) = (x'_1(0), \ldots, x'_d(0))$, $x'_i(0) = x_i(0) + \delta x_i(0)$. Вектор $\delta x(0) = (\delta x_1(0), \ldots, \delta x_d(0))$ характеризует размер возмущения. Возмущенное решение $x'(t)$ отклоняется от начального решения $x(t)$ и можно задать вопрос: насколько велико время t, при котором $\|x'(t) - x(t)\| = O(1)$, т. е. $x'(t) - x(t)$ по порядку сравнимо с характерным размером фазового пространства? (В конкретных ситуациях обычно легко определить точное значение этой величины.) Чем меньше время t, тем более неустойчива динамика. Предположим, что $\delta x(0) \to 0$ вдоль некоторого направления и $t \sim \lambda \ln \|\delta x(0)\|$ для некоторого $\lambda > 0$. Конечно, λ зависит от выбранного направления, которое называется в этом случае неустойчивым. Последняя формула означает, что $\|\delta x(t)\| \asymp e^{\lambda t}\|\delta x(0)\|$. Однако правая часть значима только при таких t и $\delta x(0)$, что величина $e^{\lambda t}\|\delta x(0)\|$ сравнима по порядку с характерным размером фазового пространства. Если $t \to -\infty$, то соответствующее направление называется устойчивым (терминология предложена Смейлом). Число λ в этом выражении называется показателем Ляпунова. Множество всех возможных λ называется ляпуновским спектром (данной траектории).

Точка $x(0)$ называется гиперболической, если $\mathbb{R}^d = R^1(x(0)) + E^{(u)}(x(0)) + E^{(s)}(x(0))$, где $R^1(x(0))$ — одномерное подпространство, порожденное вектором $f(x(0)) = (f_1(x(0)), \ldots, f_d(x(0)))$, $E^{(u)}$ и $E^{(s)}$ — подпространства, порождённые неустойчивыми и устойчивыми направлениями соответственно. Понятия гиперболической неподвижной точки и периодической орбиты хорошо известны в анализе, гидродинамике, прикладной математике. Теория хаоса имеет дело с общими гиперболическими траекториями. Система (1) называется гиперболической, если каждая точка $x(0)$ — гиперболическая. Гиперболичность может быть равномерной и неравномерной. Современное понятие гиперболичности введено Аносовым и Смейлом.

Пусть $\dim E^{(u)}(x(0)) = k$, $\dim E^{(s)}(x(0)) = l$ и эти размерности не зависят от $x(0)$. C^1-подмногообразие $\Gamma^{(u)}(x)$ называется неустойчивым многообразием точки $x = x(0)$, если $\dim \Gamma^{(u)}(x) = k$, касательная плоскость к $\Gamma^{(u)}(x)$ в точке x совпадает с $E^{(u)}(x)$ и $\Gamma^{(u)}(x) = \{y : \mathrm{dist}(x(t), y(t)) \to 0$ при $t \to -\infty\}$. Аналогично, l-мерное C^1-подмногообразие $\Gamma^{(s)}(x)$ называется устойчивым многообразием точки $x = x(0)$, если $\dim \Gamma^{(s)}(x) = l$, касательная плоскость к $\Gamma^{(s)}(x)$ в точке x совпадает с $E^{(s)}(x)$ и $\Gamma^{(s)}(x) = \{y : \mathrm{dist}(x(t), y(t)) \to 0$ при $t \to \infty\}$. Аналогичные определения можно дать в более общем случае гладких векторных полей на гладких многообразиях или гладких итераций на гладких многообразиях. Во всех этих случаях имеется однопараметрическая группа преобразований фазового пространства, образованная сдвигами вдоль траекторий. Такие группы и называются динамическими системами.

Утверждение, что при некоторых дополнительных условиях гиперболические точки имеют устойчивые и неустойчивые многообразия, называется теоремой Адамара – Перрона или теоремой об инвариантных многообразиях.

Ниже нам потребуется также понятие инвариантной меры. Нормированная мера μ называется инвариантной для данной динамической системы, если для любой ограниченной измеримой функции φ

$$\int \varphi(x(0))\, d\mu(x(0)) = \int \varphi(x(t))\, d\mu(x(0)).$$

Мера μ абсолютно непрерывна (по отношению к мере Лебега), если $d\mu(x) = p(x) dx_1 \cdot \ldots \cdot dx_d$. В этом случае функция p называется плотностью меры. Мы обсудим роль инвариантных мер в лекции 3.

Лекция 2. Основные примеры гиперболических динамических систем

1. Одномерные отображения. Пусть f — определенная на $[0,1]$ кусочно C^2-гладкая функция, для которой $f(x) \in [0,1]$ при $0 \leqslant x \leqslant 1$. Такая функция порождает итерации T_f^n: $T_f(x) = f(x)$, $T_f^n(x) = f(T_f^{n-1}(x))$. Отображение T_f — гиперболическое, если $f'(x) > 1$ при всех x. Это неравенство может выполняться только для разрывных f, причем в точках разрыва имеются в виду левая и правая производные. Интересно отметить, что в этом и других случаях разрывность динамики не представляет серьезной трудности. Простейшая ситуация возникает, когда f обладает марковским свойством: отрезок $[0,1]$ можно разбить на конечное число подинтервалов $[x_i, x_{i+1}]$ так, чтобы на каждом интервале $f(x_i) = 0$, $f(x_{i+1}) = 1$, т.е. $T_f[x_i, x_{i+1}] = [0,1]$. У гиперболического T_f всегда есть абсолютно непрерывная инвариантная мера. В марковском случае это — теорема Реньи, она может быть доказана при помощи теоремы из первой лекции. Общее утверждение доказано Ледрапье, Келлером, Хофбауэром и другими.

Самый известный случай в теории одномерных отображений — квадратичные отображения $x \mapsto rx(1-x)$, $0 \leqslant x \leqslant 1$. Здесь r — параметр, $0 \leqslant r \leqslant 4$. Эти отображения изучались в голоморфной динамике, имеющей дело с итерациями голоморфных функций (Фату, Жюлиа, Карлесон, Милнор, Любич, Макмюллен, Дуади, Хаббард и многие другие). Они оказались важны для так называемой универсальности Фейгенбаума, описывающей возникновение хаоса.

Для хаоса важно, чтобы у отображения была абсолютно непрерывная инвариантная мера. Центральным результатом здесь является следующая теорема Якобсона.

Теорема. *Множество значений r, при которых у T_f есть абсолютно непрерывная инвариантная мера, имеет положительную меру Лебега.*

Недавно Любич доказал, что $[0,4] = C_1 \cup C_2 \cup C_3$, где $l(C_1) = 0$, $l(C_2) > 0$ и для всех $r \in C_2$ преобразование T_f имеет абсолютно непрерывную инвариантную меру, а для всех $r \in C_3$ у T_f есть периодическая траектория, притягивающая почти все точки отрезка $[0,1]$. Множество C_3 открыто и всюду плотно. Множество значений r, при которых T_f гиперболично, имеет полную меру.

2. Двумерные отображения. Простейший пример двумерного гиперболического отображения — групповой автоморфизм двумерного тора, $T(x_1, x_2) = (ax_1 + bx_2, cx_1 + dx_2)$, где $x = (x_1, x_2)$ — точка $\mathbb{T}^2$, координаты берутся по модулю 1, коэффициенты a, b, c, d — целые и $ad - bc = 1$. Отображение T гиперболично, если собственные числа λ_1, λ_2 матрицы $\left\| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right\|$ таковы, что $|\lambda_1| > 1 > |\lambda_2|$. Неустойчивые и устойчивые многообразия $\Gamma^{(u)}(x)$, $\Gamma^{(s)}(x)$ — прямые, идущие вдоль направлений собственных векторов, соответствующих λ_1 и λ_2.

3. Стандартные отображения. Рассмотрим периодическую гладкую функцию $V(x)$ периода 1, имеющую две невырожденные критические точки. Отображения двумерного тора $T(x_1, x_2) = (x_1', x_2')$, где $x_1' = x_1 + kV(x) \pmod 1$, $x_2' = x_2 + x_1' \pmod 1$, k — параметр, называются стандартными отображениями. Они были введены Б. В. Чириковым и британским физиком Тейлором и широко используются в физике плазмы, нелинейных колебаниях и т. п. Стандартные отображения при малых значениях k дают простейшие примеры применения теории КАМ (Колмогоров – Арнольд – Мозер). В некотором смысле стандартные отображения являются двумерными аналогами квадратичных отображений. Как только стандартные отображения стали популярными, появилась гипотеза, что множество гиперболических точек имеет положительную меру Лебега для всех $k > 0$. Эта очень трудная проблема все еще остается открытой.

4. Геодезические потоки на поверхностях постоянной отрицательной кривизны и других многообразиях. Эти динамические системы являются естественными аналогами групповых автоморфизмов Tor^2. Зададим на верхней полуплоскости $H = \{z : \operatorname{Im} z > 0\}$ комплексной плоскости риманову метрику $ds^2 = (dx^2 + dy^2)/y^2$.

Хорошо известно, что H с такой метрикой имеет гауссову кривизну -1. Чтобы построить пример геодезического потока, возьмем группу $\Gamma = \left\{ \left\| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right\| \right\}$, a, b, c, d — целые и $ad - bc = 1$. Можно выбрать фундаментальную область

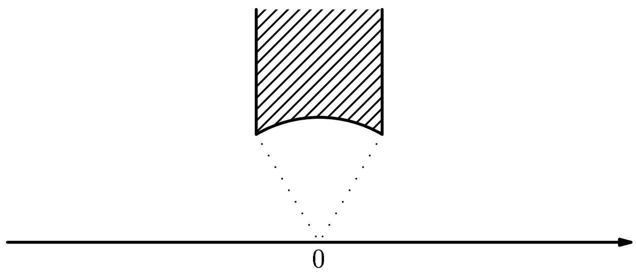

Рис. 1.

$Q = H/\Gamma$, как изображено на рисунке 1. Легко подсчитать, что площадь Q конечна.

Для построения других примеров нужно брать различные дискретные подгруппы группы $SL(2,\mathbb{R})$ действительных матриц $\|\begin{smallmatrix} x & y \\ z & w \end{smallmatrix}\|$, $xw - zy = 1$.

Область Q играет важную роль в теории чисел, теории представлений, квантовом хаосе. Фазовое пространство M геодезического потока — это трехмерное многообразие единичных касательных векторов к Q. Геодезический поток — это динамическая система в M, перемещающая каждый вектор с единичной скоростью вдоль определяемой этим вектором геодезической.

Геодезические потоки можно ввести на любом римановом многообразии. В случае многообразий отрицательной кривизны они имеют особые свойства равномерно сильной гиперболичности. Устойчивые и неустойчивые многообразия — хорошо известные в гиперболической геометрии объекты, называемые орисферами. В двумерном случае устойчивые и неустойчивые многообразия — это одномерные семейства векторов, ортогональных к окружностям, касающимся прямой $\operatorname{Im} z = 0$, или единичные векторы, ортогональные к горизонтальным прямым. Орисферы (в двумерном случае — орициклы) можно получать как пределы окружностей (см. рис. 3).

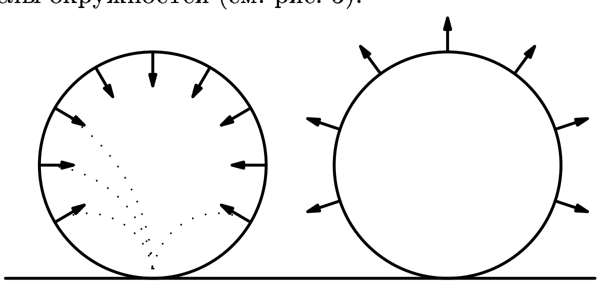

Рис. 2.

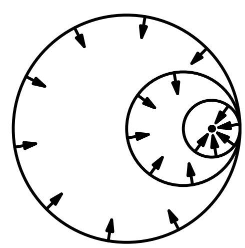

Рис. 3.

5. Системы Аносова. Динамические системы с однородной гиперболичностью всех траекторий называются системами Аносова (в честь российского математика Аносова, который их ввел). Они включают в себя групповые автоморфизмы торов и их нелинейные возмущения, геодезические потоки на компактных многообразиях отрицательной кривизны и некоторые другие системы.

6. Рассеивающие, или гиперболические биллиарды. Биллиарды — это динамические системы, соответствующие равномерному движению точки положительной массы внутри некоторой области Q. Точка отражается от границы так, что угол падения равен углу отражения. Динамика такого движения разрывна. Свойства биллиарда зависят от свойств его границы ∂Q. Биллиард называется гиперболическим или рассеивающим, если его граница ∂Q — вогнута (см. рис. 4).

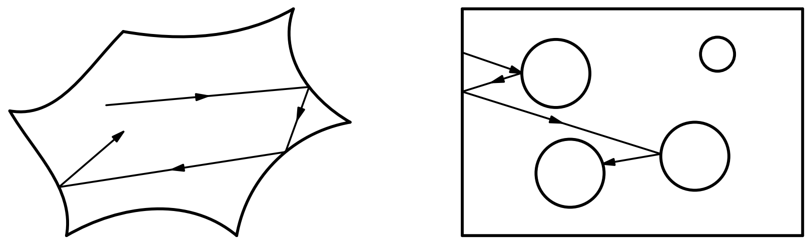

Рис. 4.

Как и в случае геодезических потоков на многообразиях отрицательной кривизны, типичная точка имеет устойчивое и неустойчивое многообразия. Примеры для двумерного случая приведены на рис. 5. Особые точки на этих многообразиях, имеющие вид клюва, соответствуют траекториям, которые в некоторый момент касаются границы. В этих точках динамика разрывна. Знаменитая система упруго сталкивающихся твердых шаров может также быть представлена как биллиардная система. Бунимович доказал гиперболичность биллиарда внутри стадиона (стадион Бунимовича). Газу из твердых сфер и другим биллиардам посвящена большая книга под редакцией Д. Саса, которая недавно вышла в издательстве Шпрингер.

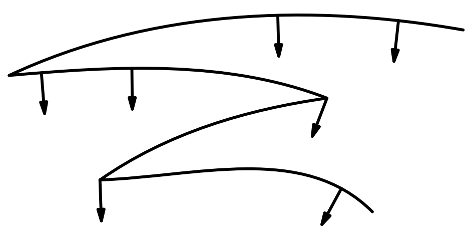

Рис. 5.

7. Системы Смейла с аксиомой А. Аттрактор Лоренца. С. Смейл предложил рассматривать динамические системы, у которых сильная однородная гиперболичность достигается на предельном множестве, притягивающем все траектории. Смейл назвал такие системы системами, удовлетворяющими аксиоме А. Они играют большую роль в топологической теории динамических систем.

В 1962 году Е. Лоренц привел пример системы из трех обыкновенных дифференциальных уравнений, которая обладает многими свойствами систем с аксиомой А. Она имеет очень простой вид

$$\begin{aligned} \frac{dx_1}{dt} &= -\sigma x_1 + \sigma x_2, \\ \frac{dx_2}{dt} &= r x_1 - x_2 - x_1 x_3, \\ \frac{dx_3}{dt} &= -b x_3 + x_1 x_2. \end{aligned} \tag{2}$$

Здесь σ, r, b — числовые параметры. Е. Лоренц открыл эту систему при исследовании знаменитой проблемы Бенара в гидродинамике. Он обнаружил при численных расчетах, что система (2) при некоторых значениях σ, r, b проявляет сильные свойства гиперболичности. Десятью годами позже эта система стала весьма популярной, и идеи хаоса начали проникать в физику. Примерно в это же время Рюэль и Такенс (РТ) предложили понятие странного аттрактора применительно к проблеме турбулентности. Строго говоря, система (2) не удовлетворяет определению РТ, однако очень близка к нему. Объяснение гиперболичности (2) было дано Р. Вильямсом и (несколько иным способом) Афрамовичем, Быковым, Шильниковым. Недавно Такер при помощи компьютерных вычислений нашел доказательство гиперболичности системы Лоренца (2).

8. Отображение Эно и аттрактор Эно. Вдохновленный работой Е. Лоренца, французский астроном М. Эно предложил дискретную версию системы Лоренца (2), в численных экспериментах демонстрирующую аналогичные свойства. Она имеет вид $(x_1, x_2) \to (1 + x_2 - a x_1^2, b x_1)$, где a, b — параметры. М. Бенедик и Л. Карлесон доказали замечательные результаты о гиперболичности отображений Эно.

9. Подкова Смейла, гомоклинические и гетероклинические траектории. Всё это — механизмы, порождающие инвариантные множества, состоящие из гиперболических траекторий. Обычно такие множества имеют меру нуль и в некотором смысле исключительны. Тем не менее, траектории на этих множествах проявляют хаотическое поведение. Преимущество подковы Смейла, гомоклинических и гетероклинических траекторий в том, что во многих случаях их гораздо легче обнаружить аналитическими или численными методами.

Лекция 3. Математика устоявшегося хаоса

Из гиперболичности и неустойчивости траекторий легко следует, что в таких системах число возможных типов траекторий растёт экспоненциально. Многие их них — хаотические, потому что общее число нехаотических траекторий растет медленнее, чем экспоненциально. Точная формулировка терминов и понятий, относящихся к этому вопросу, связана с теорией сложности динамических систем. Мы не будем здесь её обсуждать.

Детерминированная динамика хаотична, если она обладает свойствами случайных событий или случайных экспериментов из теории вероятностей. Первой основной теоремой теории вероятностей является закон больших чисел о существовании и устойчивости временны́х средних. Он имеет естественный аналог в динамике. А именно, определим для любой ограниченной функции φ её временнóе среднее

$$\lim_{t\to\infty}\frac{1}{t}\int_u^{u+t}\varphi(x(s))\,ds=\bar\varphi(x(0))$$

при условии, что этот предел существует. Он не зависит от u, но может зависеть от $x(0)$. Эргодическая теорема Биркгофа утверждает, что если μ — инвариантная мера, то временнóе среднее существует для μ-почти всех точек $x(0)$. Таким образом, первый шаг в анализе хаоса состоит в выборе инвариантной меры μ. Обычно динамические системы имеют много типов траекторий: периодические траектории, асимптотические траектории, транзиентные траектории и т. п. Если мы фиксируем μ, то мы фиксируем некоторое множество траекторий с однородным режимом.

Есть общие теоремы о существовании по крайней мере одной инвариантной меры (один из примеров — теорема Боголюбова – Крылова для динамических систем с компактным фазовым пространством). Во многих случаях существование естественных инвариантных мер следует из законов динамики. Например, так называемые гамильтоновы системы, групповые автоморфизмы, геодезические потоки на компактных римановых многообразиях, биллиарды имеют абсолютно непрерывные инвариантные меры.

Известен общий способ выбора естественной инвариантной меры, называемый иногда принципом Рюэля. Возьмём любую нормированную плотность $\rho(x;0)$, $\int\rho(x;0)\,dx_1\,dx_2\cdot\ldots\cdot dx_d=1$. Она порождает меру μ_0, для которой $\dfrac{d\mu_0}{dx}=\rho(x;0)$. Для любого $t>0$ и ограниченной φ

$$\int\varphi(x(t))\,d\mu_0(x(0))=\int\varphi(x(t))\rho(x;0)\,dx=\int\varphi(y)\rho(y;t)\,dy.$$

Плотность $\rho(y;t)$ удовлетворяет уравнению Лиувилля

$$\frac{\partial\rho}{\partial t}+\operatorname{div}(\rho,f)=0,\quad f(y)=(f_1(y),\ldots,f_d(y))$$

или его аналогу для случая итераций. Если существует предел

$$\lim_{t\to\infty}\int\varphi(x(t))\,d\mu_0(x(0))=\int\varphi(y)\,d\mu_\infty(y),$$

где μ_∞ — предельная мера и μ_∞ не зависит от μ_0, то μ_∞ — естественная мера. Теорема Якобсона и результат Бенедикса – Карлесона, упомянутые в предыдущей лекции, гарантируют существование μ_∞ для соответствующих систем. Для гамильтоновых систем и других систем, названных выше, трудной задачей является доказательство того, что μ_∞ совпадает с естественной абсолютно непрерывной мерой. Эта задача сводится к доказательству перемешивающего свойства динамики и будет обсуждаться ниже.

Теперь предположим, что для данной динамической системы выбрана инвариантная мера μ. Теорема Биркгофа позволяет временны́м средним $\bar\varphi(x(0))$ зависеть от $x(0)$. Мера μ называется эргодической, если $\bar\varphi(x(0))=\int\varphi(y)\,d\mu(y)$. В этом случае эргодическая теорема принимает вид обычного закона больших чисел. Терминология происходит из статистической механики, а понятие эргодичности в немного ином виде было предложено Л. Больцманом.

Обычно очень трудно доказать, что данная инвариантная мера эргодична. Например, мы сейчас знаем благодаря КАМ теории, что многие гамильтоновы системы, описывающие динамику взаимодействующих частиц, неэргодичны. Однако для гиперболических динамических систем имеется общее рассуждение Э. Хопфа, которое устанавливает эргодичность. Мы опишем это рассуждение, но, конечно, это не будет полным строгим доказательством. Для простоты ограничимся случаем компактного фазового пространства. Достаточно рассмотреть непрерывные функции φ. Рассмотрим для точки x её локальное устойчивое многообразие $\gamma^{(s)}(x)$, т. е. малую окрестность x, принадлежащую устойчивому многообразию $\Gamma^{(s)}(x)$, $\gamma^{(s)}(x) \subset \Gamma^{(s)}(x)$. Если существует $\bar{\varphi}(x)$, то и $\bar{\varphi}(y)$, $y \in \gamma^{(s)}$ также существует, причём $\bar{\varphi}(y) = \bar{\varphi}(x)$. Таким образом, пока x движется вдоль локальных устойчивых многообразий, временны́е средние в положительном направлении времени остаются теми же самыми. При движении x вдоль траектории временны́е средние сохраняются по очевидным причинам. Одно из следствий эргодической теоремы Биркгофа утверждает, что для почти всех y

$$\lim_{t\to\infty} \frac{1}{t} \int\limits_{-t}^{0} \varphi(x(s))\, ds = \lim_{t\to\infty} \frac{1}{t} \int\limits_{0}^{t} \varphi(x(s))\, ds$$

и оба предела существуют. Это показывает, что временны́е средние одни и те же для почти всех точек локальных неустойчивых многообразий. В гиперболических системах можно перейти от начальной точки к любой точке в малой окрестности начальной, двигаясь вдоль траекторий (1) и локальных устойчивых и неустойчивых многообразий. Поэтому временно́е среднее $\bar{\varphi}(y)$ постоянно в малой окрестности x. Как правило, не слишком трудно перейти от локальной эргодичности к глобальной.

Э. Хопф использовал это рассуждение, чтобы доказать эргодичность геодезических потоков на компактных многообразиях отрицательной кривизны. Позднее оно применялось к системам Аносова, системам с аксиомой А и т. д. Однако оказалось, что применение рассуждения Хопфа к рассеивающим гиперболическим биллиардам — трудная задача (Бунимович, Чернов, Синай, Шимани, Сас и др.).

Эргодическая теорема для марковских цепей утверждает, что при естественных условиях любое начальное распределение сходится к единственному стационарному распределению. Аналогом в динамике является упомянутый выше принцип Рюэля. В случае абсолютно непрерывной инвариантной меры с плотностью ρ это означает, что

$$\int \varphi(x)\rho(x,t)\, dx \to \int \varphi(x)\rho(x)\, dx, \quad t\to\infty. \tag{3}$$

В общей эргодической теории последнее свойство называется перемешиванием. Конечно, отсюда не следует поточечная сходимость $\rho(x,t)$ к $\rho(x)$. Простые примеры показывают, что функция $\rho(\cdot,t)$ при больших t становится очень нерегулярной и (3) выполняется вследствие того, что содержит некоторое усреднение. Во многих важных случаях предельная мера μ_∞ сингулярна по отношению к мере Лебега.

Для гиперболических систем доказательство (3) и сходимости к пределу состоят из двух шагов, и оба шага можно естественно интерпретировать на языке термодинамического формализма, упомянутого в лекции 1. Первый шаг — доказательство того, что динамика порождает на каждом $\Gamma^{(u)}(x(0))$ некоторую инвариантную, но не нормированную меру. Механизм опять-таки прост. Берём меру Лебега на $\Gamma^{(u)}(x(-t))$ и её образ на $\Gamma^{(u)}(x(0))$, задаваемый динамикой. Не очень трудно доказать, что возникающие плотности сходятся. В гладких системах с сильной гиперболичностью сходимость экспоненциальна.

Второй шаг состоит в том, чтобы показать, что среднее на любом $\Gamma^{(u)}(x)$ по построенной мере имеет предел. Этот шаг более трудный. Обычно для него нужны специальные разбиения фазового пространства, так называемые марковские разбиения. Мы обсудим их позднее. Возникающие таким путём меры иногда называются SRB-мерами (в честь Синая, Рюэля, Боуэна).

Следующим центральным результатом теории вероятностей является предельная теорема Гаусса, описывающая поведение флуктуаций

$$\frac{1}{t}\int\limits_0^t \varphi(x(s))\,ds - \bar{\varphi},$$

$\bar{\varphi}(y) = \int \varphi(x)\,d\mu(x)$. Если эта предельная теорема выполняется, то указанная разность убывает как $O(\frac{1}{\sqrt{t}})$. У неё нет никакого асимптотического разложения в обычном смысле, но нормированная разность

$$\sqrt{t}\left(\frac{1}{t}\int\limits_0^t \varphi(x(s))\,ds - \bar{\varphi}\right)$$

имеет универсальное гауссово распределение при $t \to \infty$. Универсальность в данном контексте означает, что это распределение почти не зависит от φ и от меры μ. Более точно, меру μ можно заменить на произвольную неинвариантную меру, абсолютно непрерывную по отношению к μ.

В теории вероятностей предельная теорема Гаусса доказывается при различных предположениях о независимости или слабой зависимости случайных величин.

Оказывается, это свойство также выполняется для гиперболических динамических систем. Дадим некоторые пояснения в случае дискретного времени.

Этот подход основывается на чисто геометрическом или, более точно, чисто динамическом понятии марковского разбиения. Замкнутое подмножество C фазового пространства называется параллелограммом, если оно состоит из локальных устойчивых и локальных неустойчивых многообразий и каждое локальное устойчивое многообразие пересекает каждое локальное неустойчивое многообразие в одной точке. Набор параллелограммов $\{C_1, C_2, \ldots, C_r\}$ называется разбиением, если $\bigcup\limits_{i=1}^{r} C_i$ — всё фазовое пространство и $C_i \cap C_j = \partial C_i \cap \partial C_j$. Разбиение $\{C_1, C_2, \ldots, C_r\}$ называется марковским, если для каждого локального неустойчивого $\gamma^{(u)} \subset C_i$ любое пересечение $T\gamma^{(u)} \cap C_j$ есть локальное неустойчивое многообразие в C_j и, аналогично, для каждого локального устойчивого

$\gamma^{(s)} \in C_i$ любое пересечение $T^{-1}\gamma^{(s)} \cap C_j$ — локальное устойчивое многообразие в C_j.

Обычно нетрудно доказать, что гиперболическая система имеет марковское разбиение. Одно из главных достоинств марковских разбиений заключается в том, что порождаемые ими так называемые символические представления динамики оказываются довольно простыми. Для марковских разбиений бесконечное пересечение $\bigcap\limits_{-\infty}^{+\infty} T^k C_{i_k}$ непусто, если

$$C_{i_k} \cap TC_{i_{k+1}} \neq \varnothing \quad \text{при } -\infty < k < +\infty$$

Марковские разбиения позволяют представить фазовое пространство как пространство реализаций стационарного случайного процесса из теории вероятностей и применить вероятностные методы. Во многих случаях SRB-меры можно представить как гиббсовские распределения в смысле лекции 1 .

Вот пример результата, который можно доказать таким методом. Предположим, мы имеем периодическую конфигурацию окружностей на плоскости. Выберем случайное начальное условие (q, v), где $q = (q_1, q_2)$, $v = (v_1, v_2)$, $v_1^2 + v_2^2 = 1$, и рассмотрим динамику биллиардной частицы с начальным условием (q, v). Оказывается (Бунимович, Синай), что распределение нормированного смещения

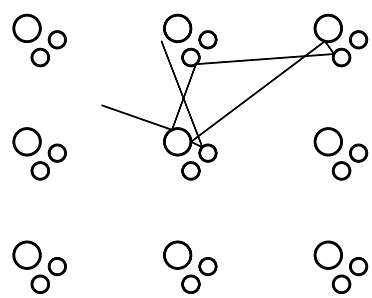

Рис. 6.

$q(t)/\sqrt{t}$ сходится к мере Винера, т. е. детерминированная динамика порождает диффузионный процесс. Обычная картина броуновского движения — это перемещение массивной частицы под воздействием многочисленных независимых столкновений с лёгкими частицами окружающего газа. В нашем случае лёгкая частица движется под воздействием столкновений с массивными неподвижными частицами и диффузия возникает из-за неустойчивости динамики. Это, возможно, перестаёт быть верным при замене окружностей на квадраты.

Многих интересует, в чём состоят основные достижения теории хаоса. Вот несколько ответов.

1. Результаты, показывающие, каким образом обратимая детерминированная динамика может порождать необратимость. Пример был приведен выше.

2. Анализ простых систем со странными аттракторами, наподобие системы Лоренца. Оказывается, сложная динамика может возникать при простых уравнениях движения.

3. Универсальность Фейгенбаума при возникновении хаоса и последовательности бифуркаций удвоения периода. Это замечательное открытие теории хаоса, которое наблюдалось во многих экспериментах. В математике благодаря ему возник так называемый метод ренорм-группы в теории динамических систем.

4. В приложениях теории динамических систем используются такие понятия теории хаоса, как топологическая и метрическая энтропия, ляпуновский спектр и т. п. Эти параметры можно оценить численно с довольно большой точностью, из чего часто удается заключить, что на протяжении больших, но конечных интервалов времени система ведёт себя как гиперболическая система с хаотическими свойствами.

5. Я не думаю, что теория хаоса дала что-либо интересное для гидродинамики, и, в частности, для турбулентности. Вероятно, мы до сих пор не имеем достаточно хорошего понимания динамики в этих случаях.

Этот текст следует более или менее близко Хедриковским лекциям, прочитанным в Лос-Анжелесе в августе 2000 года. Многие важные темы теории хаоса в нём пропущены. Прежде всего я имею в виду замечательные работы Д. Орнстейна и его коллег по проблеме изоморфизма динамических систем и сложности динамики, работы Дж. Мезера, Г. Ксиа о неустойчивости и диффузии Арнольда в гамильтоновых системах и т. д. Все эти темы требуют отдельного изложения.