В. А. Зорич

Математический анализ задач естествознания

Электронное издание

Москва
Издательство МЦНМО
2018

УДК 517
ББК 22.16
386

Зорич В. А.
Математический анализ задач естествознания.
Электронное издание.
М.: МЦНМО, 2018.
155 с.
ISBN 978-5-4439-3225-5

Эта книга содержит записи годового экспериментального спецкурса естественнонаучного содержания. В нём представлены три темы:

— анализ размерностей физических величин с примерами приложений, включая модель турбулентности по Колмогорову;

— функции очень многих переменных и явление концентрации: нелинейный закон больших чисел, геометрический смысл распределений Гаусса и Максвелла, теорема Котельникова—Шеннона;

— классическая термодинамика и контактная геометрия: два начала термодинамики на языке форм, распределения и теорема Фробениуса, метрика Карно—Каратеодори.

Спецкурс предназначен в первую очередь математикам, но может быть также полезен студентам и специалистам иных специальностей.

В приложении помещена общедоступная статья автора «Математика как язык и метод».

Подготовлено на основе книги:
Зорич В. А. Математический анализ задач естествознания. — Новое изд., доп. — М.: МЦНМО, 2017. — 160 с. — ISBN 978-5-4439-1225-7

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11,
тел. (499)241-08-04.
http://www.mccme.ru

ISBN 978-5-4439-3225-5

Оглавление

Тема II
Многомерная геометрия и функции очень многих переменных

Предисловие к первому изданию

Это небольшой курс естественнонаучного содержания, предназначенный математикам. Он допускает развития разной направленности.

Здесь представлено кое-что из наследия Галилея, Ньютона, Эйлера, Бернулли, Карно, Клаузиуса, Максвелла, Больцмана, Гиббса, Пуанкаре, Эйнштейна, Планка, Шрёдингера, Каратеодори, Колмогорова, Котельникова, Шеннона...

Название «Математический анализ задач естествознания», разумеется, отражает лишь тенденцию, а не обещание какой-то всеобщности типа «всё, сразу и даром». Произведён очень условный отбор трёх тем.

Заметим, что те, кто особенно чтим нами, математиками, Архимед, Ньютон, Лейбниц, Эйлер, Гаусс, Пуанкаре, были не только математиками, но учёными, натурфилософами.

В математике решение важных конкретных научных вопросов и построение абстрактной общей теории — процессы неразрывные, как вдох и выдох. Длительное нарушение этого баланса опасно.

Не стоит попадать в положение рыбака, продолжающего с увлечением дёргать леску и ловить рыбку на уже оторвавшейся от берега льдине.

Герман Вейль отмечал, что «подлинно реалистическая математика должна мыслиться в одном строю с физикой как часть теоретической конструкции единого реального мира...». Между прочим, это пока ещё указано даже в наших дипломах кандидатов и докторов физико-математических наук.

В заключение я хотел бы поблагодарить всех, кто способствовал правке первоначального текста, особенно В. И. Арнольда, не пропустившего ни единого абзаца стостраничной распечатки[1], сделавшего массу острых наблюдений, сопровождая их комментариями[2]. Ес-

[1]Это не преувеличение: файл, предназначенный для В. И., случайно был послан на принтер, который некоторые буквы кириллицы систематически заменял другими (кстати, иногда при этом получались довольно забавные вещи). Так вот, В. И. заодно повсюду исправил и дефекты печати.

[2]Например, что рассуждения темы I «мракобесны». Об этом я должен предупредить читателя. См. также сноску на с. 29.

ли я не все замечания или пожелания коллег здесь учёл, то это вовсе не значит, что я ими пренебрёг, скорее, кое-что счёл предметом размышления и обсуждения.

В. Зорич
Москва, 2007

Предисловие к новому изданию

В этом новом издании книги, помимо исправления замеченных погрешностей предыдущего издания, сделано несколько изменений и добавлений. В основном это касается темы II, точнее её второго параграфа, посвящённого многомерной геометрии и феномену концентрации меры, имеющему разнообразные проявления (помимо самой геометрии, например, в теории вероятностей, статистической физике, теории передачи информации по каналу связи при наличии помех, теории больших информационных систем...). Эта тематика стала предметом многочисленных публикаций. Для сведения заинтересованного читателя мы пополнили соответствующую библиографию совсем свежей литературой.

Тема III, посвящённая термодинамике, здесь почти не изменена. На самом же деле она расширилась фактически до размера небольшой независимой книжки. Книга, возможно, будет опубликована под названием «Математические аспекты классической термодинамики».

Тема I осталась практически неизменной, если не считать нескольких новых комментариев.

В. Зорич
Москва, 2017

Тема I

Анализ размерностей физических величин

Два вводных слова

Абстрактное (отвлечённое) число, например 1 или $2\frac{2}{3}$, и арифметика абстрактных чисел, например, то, что $2+3=5$, вне зависимости от того, складываются ли яблоки или слоны, — великое достижение цивилизации, сопоставимое с приобретением письменности. К этому мы так привыкли, что уже и не замечаем чуда, лежащего где-то в самой основе удивительной эффективности математики.

Если же вы знаете, к какому объекту относится число, то, как правило, сразу одновременно появляются и дополнительные возможности, и ограничения. Вспомним детские стишки: «И вышло у меня в ответе: два землекопа и две трети». Да, число $2\frac{2}{3}$ допустимо в арифметике, но недопустимо в этой конкретной ситуации.

Как пользоваться тем, что в конкретных задачах мы имеем дело не с отвлечёнными числами, а с размерными величинами?

Есть ли какая-то наука на этот счёт? — Небольшая, но есть. Её (а также опасности неумелого её использования) знает каждый квалифицированный естествоиспытатель. О ней и пойдёт речь.

К сведению читателя: нумерация формул в каждой главе сквозная и независимая.

Глава I

Элементы теории

§1. Размерность физической величины (начальные представления)

1.1. Измерение, единица измерения, процесс измерения

Все перечисленные понятия фундаментальны и подвергались анализу лучшими представителями науки, физики и математики в первую очередь. Такой анализ приводит к анализу понятий пространства, твёрдого тела, движения, времени, причинности и т. п.

Мы здесь не собираемся погружаться так глубоко. Отметим, однако, что каждая теория хорошо моделирует некоторую сферу явлений только в каких-то масштабах. В каких именно, мы, к сожалению, порой узнаём только тогда, когда теория перестаёт согласовываться с действительностью. Именно в эти моменты мы обычно вновь возвращаемся к фундаменту теории, подвергая его тщательному анализу и должной реконструкции.

А сейчас давайте начнём с накопления доступного нам полезного конкретного материала.

1.2. Основные и производные единицы

В быту мы постоянно пользуемся какими-то единицами измерения длины, массы, времени, силы, скорости, энергии, мощности... Какие-то из них мы выделяем как основные, а какие-то при этом уже оказываются производными.

Примеры основных единиц:

L — единица длины *метр* (м или m);
M — единица массы *килограмм* (кг или kg);
T — единица времени *секунда* (с или s).

Примеры производных единиц:

v — скорость (м/с или m/s) $[v] = LT^{-1}$;
V — объём (м^3 или m^3) $[V] = L^3$;

a — ускорение (м/с^2 или m/s^2) $[a]=LT^{-2}$;
l — световой год $[l]=[cT]=L$;
F — сила, $F=ma$, $[F]=[ma]=MLT^{-2}$.

Намечается общее наблюдение о формуле $L^{d_1}M^{d_2}T^{d_3}$ размерности механической физической величины; $\{d_1, d_2, d_3\}$ — вектор размерности в базисе $\{L, M, T\}$.

Разовьём эту векторную алгебро-геометрическую аналогию.

1.3. Зависимые и независимые единицы

Пример. Единицы величин v, a, F независимы и тоже могут быть приняты за основные, так как

$$[L] = v^2a^{-1}, \quad [M] = Fa^{-1}, \quad [T] = va^{-1}.$$

Намечается аналогия с векторным пространством, его базисом и системами независимых векторов. (Более глубокий смысл этой аналогии откроется ниже.)

§ 2. Формула размерности физической величины

2.1. Изменение числовых значений физической величины при изменении размеров основных единиц

Пример. Если расстояния измерять в километрах (т. е. вместо одного метра за единицу измерения длин принять в 1000 раз большую единицу, 1 км), то одна и та же физическая длина будет иметь разные численные значения по отношению к этим двум единицам длины, а именно: 1 км$=10^3$ м, L км$=10^3L$ м, 1 м$=10^{-3}$ км, L м$=10^{-3}L$ км. Таким образом, изменение единицы длины в α раз приводит к изменению числового значения L м всех измеряемых длин в α^{-1} раз, т. е. значение L заменится на $\alpha^{-1}L$.

Это же относится и к возможному изменению единиц массы и времени (тонна, грамм, миллиграмм; час, сутки, год, миллисекунда...).

Значит, если физическая величина в базисе $\{L, M, T\}$ имеет размерность $L^{d_1}M^{d_2}T^{d_3}$, т. е. $\{d_1, d_2, d_3\}$ — её вектор размерности, то изменение единиц измерения длины, массы и времени в α_1, α_2, α_3 раз соответственно, по-видимому, должно приводить к изменению числового значения этой величины в $\alpha_1^{-d_1}\alpha_2^{-d_2}\alpha_3^{-d_3}$ раз.

2.2. Постулат инвариантности отношения значений одноимённых физических величин

Пример. Площадь треугольника есть функция $y = f(x_1, x_2, x_3)$ длин трёх его сторон. Возьмём другой треугольник и подсчитаем его площадь $\widetilde{y} = f(\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3)$.

При изменении размера единицы длины числовые значения y и $\widetilde{y}$ изменятся. Однако их отношение $\widetilde{y}/y$ при этом останется неизменным.

Пусть теперь имеется две однотипные физические величины, $y = f(x_1, x_2, \ldots, x_m)$, $\widetilde{y} = f(\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_m)$, зависящие не от трёх, а от любого конечного набора только длин, или масс, или времён...

В теории размерностей принимается следующий основной постулат.

Постулат абсолютности отношений:

$$\frac{\widetilde{y}}{y} = \frac{f(\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_m)}{f(x_1, x_2, \ldots, x_m)} = \frac{f(\alpha\widetilde{x}_1, \alpha\widetilde{x}_2, \ldots, \alpha\widetilde{x}_m)}{f(\alpha x_1, \alpha x_2, \ldots, \alpha x_m)}. \tag{1}$$

Иными словами, постулируется, что при изменении масштаба основной единицы (длины, массы, времени...) однотипные физические величины $y, \widetilde{y}$ (все площади, все объёмы, все скорости, все силы...) меняют своё числовое значение в одно и то же число раз (своё для каждого типа величин).

2.3. Функция размерности и формула размерности физической величины в данном базисе

Равенство (1) показывает, что отношение

$$\frac{f(\alpha x_1, \alpha x_2, \ldots, \alpha x_m)}{f(x_1, x_2, \ldots, x_m)} =: \varphi(\alpha)$$

зависит только от α. Это позволяет указать вид функции $\varphi(\alpha)$.

Заметим сначала, что

$$\frac{\varphi(\alpha_1)}{\varphi(\alpha_2)} = \varphi\Big(\frac{\alpha_1}{\alpha_2}\Big). \tag{2}$$

Действительно,

$$\frac{\varphi(\alpha_1)}{\varphi(\alpha_2)} = \frac{f(\alpha_1 x_1, \alpha_1 x_2, \ldots, \alpha_1 x_m)}{f(\alpha_2 x_1, \alpha_2 x_2, \ldots, \alpha_2 x_m)} = \frac{f\Big(\frac{\alpha_1}{\alpha_2}x_1, \frac{\alpha_1}{\alpha_2}x_2, \ldots, \frac{\alpha_1}{\alpha_2}x_m\Big)}{f(x_1, x_2, \ldots, x_m)} = \varphi\Big(\frac{\alpha_1}{\alpha_2}\Big).$$

Здесь крайние равенства непосредственно следуют из определения функции φ; а учитывая независимость φ от переменных

$(x_1, x_2, \ldots, x_m)$, можно перейти к переменным $\alpha_2{}^{-1}(x_1, x_2, \ldots, x_m)$ и получить среднее равенство.

Считая, что φ — регулярная функция, продифференцируем соотношение (2) по α_1, и, положив затем $\alpha_1 = \alpha_2$, придём к уравнению

$$\frac{1}{\varphi(\alpha)} \frac{d\varphi}{d\alpha} = \frac{1}{\alpha} \varphi'(1).$$

Поскольку $\varphi(1) = 1$, решение этого уравнения при таком условии имеет вид

$$\varphi(\alpha) = \alpha^d.$$

Итак, найденный вид функции φ является следствием принятого постулата теории размерностей физических величин. В формулировке постулата мы для простоты начальных рассмотрений считали, что все переменные $(x_1, x_2, \ldots, x_m)$ имеют одинаковый характер (длины, или массы, или времени, или скорости...). Но ясно, что любую функциональную зависимость физической величины от набора физических переменных можно представить в виде

$$y = f(x_1, x_2, \ldots, x_{m_1}, y_1, y_2, \ldots, y_{m_2}, \ldots, z_1, z_2, \ldots, z_{m_k}),$$

где однотипные физические переменные собраны в группы и обозначены общим символом (мы не хотели вводить нумерацию групп).

Это означает, что

$$\frac{f(\alpha_1 x_1, \alpha_1 x_2, \ldots, \alpha_1 x_{m_1}, \ldots, \alpha_k z_1, \alpha_k z_2, \ldots, \alpha_k z_{m_k})}{f(x_1, x_2, \ldots, x_{m_1}, \ldots, z_1, z_2, \ldots, z_{m_k})} = \alpha_1{}^{d_1} \cdot \ldots \cdot \alpha_k{}^{d_k},$$

если переменные разных групп допускают независимое изменение масштабов. В таком случае в рамках действия постулата мы нашли закон

$$f(\alpha_1 x_1, \alpha_1 x_2, \ldots, \alpha_1 x_{m_1}, \ldots, \alpha_k z_1, \alpha_k z_2, \ldots, \alpha_k z_{m_k}) = \\ = \alpha_1{}^{d_1} \cdot \ldots \cdot \alpha_k{}^{d_k} f(x_1, x_2, \ldots, x_{m_1}, \ldots, z_1, z_2, \ldots, z_{m_k}) \quad (3)$$

изменения численного значения физической величины

$$y = f(x_1, x_2, \ldots, x_{m_1}, \ldots, z_1, z_2, \ldots, z_{m_k})$$

при изменении масштабов единиц физически независимых переменных.

Набор $(d_1, \ldots, d_k)$ называют *вектором размерности* или просто *размерностью* величины y по отношению к выделенным независимым физическим единицам измерения.

Функцию $\varphi(\alpha_1, \ldots, \alpha_k) = {\alpha_1}^{d_1} \cdot \ldots \cdot {\alpha_k}^{d_k}$ называют *функцией размерности*.

Символом $[y]$ в зависимости от контекста обозначают либо вектор размерности, либо функцию размерности.

Физическую величину называют *безразмерной*, если её вектор размерности нулевой.

Например, если $\varphi(\alpha_1, \ldots, \alpha_k) = {\alpha_1}^{d_1} \cdot \ldots \cdot {\alpha_k}^{d_k}$ — функция размерности физической величины y по отношению к независимым физическим величинам $\{x_1, \ldots, x_k\}$, то отношение

$$\Pi := \frac{y}{{x_1}^{d_1} \cdot \ldots \cdot {x_k}^{d_k}}$$

является безразмерной величиной относительно $\{x_1, \ldots, x_k\}$.

§3. Основная теорема теории размерностей

3.1. П-теорема[1]

Теперь рассмотрим общий случай зависимости

$$y = f(x_1, x_2, \ldots, x_k, \ldots, x_n) \tag{4}$$

физической величины y от переменных величин $x_1, x_2, \ldots, x_n$, среди которых только первые k являются физически (размерно) независимыми в смысле обсуждаемой нами теории размерностей физических величин.

Приняв $x_1, x_2, \ldots, x_k$ за единицы измерения соответствующих величин, т. е. меняя масштаб, полагая $\alpha_1 = {x_1}^{-1}, \ldots, \alpha_k = {x_k}^{-1}$, мы из равенства (3) получим соотношение

$$\Pi = f(1, \ldots, 1, \Pi_1, \Pi_2, \ldots, \Pi_{n-k}) \tag{5}$$

между безразмерными величинами

$$\Pi = \frac{y}{{x_1}^{d_1} \cdot \ldots \cdot {x_k}^{d_k}}, \quad \Pi_i = \frac{x_{k+i}}{{x_1}^{d_{i_1}} \cdot \ldots \cdot {x_k}^{d_{i_k}}}, \quad \text{где } i = 1, \ldots, n-k.$$

[1]Её также называют теоремой Бакингема, связывая её появление с работами: *Buckingham E.* On physically similar systems; illustrations of the use of dimensional equations // Phys. Rev. 1914. Vol. 4. P. 345—376; и *Buckingham E.* The principle of similitude // Nature 1915. Vol. 96. P. 396—397.

В неявной форме П-теорема и принцип подобия содержатся также в работе: *Jeans J. H.* Proc. Roy. Soc. 1905. Vol. 76. P. 545, не говоря о том, что законами подобия по существу владели уже Ньютон и Галилей. Историю вопроса можно найти в работе [5] вместе со ссылкой на работу [6], где уже есть всё, кроме названия П-*теорема*. В этой связи укажем ещё на любопытные публикации [7] и [8].

Равенство (5) можно переписать в виде

$$y = {x_1}^{d_1} \cdot \ldots \cdot {x_k}^{d_k} \cdot f(1, \ldots, 1, \Pi_1, \Pi_2, \ldots, \Pi_{n-k}). \tag{6}$$

Итак, пользуясь масштабной однородностью зависимостей между физическими величинами, выраженной в сформулированном выше постулате, мы можем перейти от соотношения (4) к соотношению (5) между безразмерными величинами, уменьшив при этом число переменных; или мы можем перейти к равносильному (5) соотношению (6), выделив явно всю размерную составляющую величины y по отношению к максимальной системе $x_1, x_2, \ldots, x_k$ размерно независимых переменных.

Возможность такого перехода от общего соотношения (4) к более простым соотношениям (5) или (6) и составляет содержание так называемой П-*теоремы* — основной теоремы теории размерностей физических величин, которую мы сейчас доказали.

3.2. Принцип подобия

Содержание, смысл, возможности П-теоремы и подводные камни, связанные с ней, будут раскрываться ниже по мере рассмотрения конкретных примеров её использования.

Но одна идея весьма эффективного (и эффектного) приложения доказанной теоремы лежит прямо на поверхности и очевидна: не ломая самолёты, корабли и прочие объекты, многие эксперименты можно проводить в лаборатории над моделями, пересчитывая затем с помощью П-теоремы результаты (например, экспериментально найденную безразмерную зависимость (5)) к конкретным натуральным размерам объектов (по формуле (6)).

Глава II

Примеры приложений

Рассмотрим теперь некоторые примеры, на которых прояснятся различные стороны доказанной П-теоремы.

§1. Период обращения тела на круговой орбите (законы подобия)

Тело массы m удерживается на круговой орбите радиуса r центральной силой F. Надо найти период обращения

$$P = f(r, m, F).$$

Здесь и далее фиксируем стандартный для механики базис основных физических единиц (длина, масса, время), которые, следуя Максвеллу, мы будем обозначать $\{L, M, T\}$. (В термодинамике символ T используется для обозначения абсолютной температуры, но, если не оговорено иное, мы пока используем это обозначение для единиц времени.)

Найдём векторы размерности величин P, r, m, F в базисе $\{L, M, T\}$. Запишем их столбцами следующей таблицы:

	P	r	m	F
L	0	1	0	1
M	0	0	1	1
T	1	0	0	−2

Поскольку функция размерности, как мы показали, всегда имеет степенной вид, перемножению таких функций отвечает сложение соответствующих показателей степени, т. е. в итоге отвечают линейные операции над векторами размерности соответствующих физических величин.

Значит, используя стандартную линейную алгебру, можно находить системы независимых величин по матрице, образованной их векторами размерности, а также, раскладывая вектор размерности какой-то величины по векторам размерности выбранных независимых величин, находить формулу размерности этой величины в системе независимых величин конкретной задачи.

Так, в нашем случае величины r, m, F независимы, ибо матрица, образованная их векторами размерности $[r]$, $[m]$, $[F]$, невырожденна. Найдя разложение $[P]=\frac{1}{2}[r]+\frac{1}{2}[m]-\frac{1}{2}[F]$, на основании формулы (6) главы I можно сразу написать, что

$$P=\left(\frac{mr}{F}\right)^{1/2}\cdot f(1,1,1).$$

Таким образом, с точностью до постоянного множителя $c=f(1,1,1)$ (который можно найти в единичном лабораторном эксперименте), мы нашли зависимость P от r, m, F.

Конечно, зная закон Ньютона $F=m\cdot a$, в данном случае легко можно было бы найти и окончательную формулу, где $c=2\pi$.

Однако всё, что мы использовали, — это лишь общее указание о существовании зависимости $P=f(r,m,F)$.

§ 2. Гравитационная постоянная. Третий закон Кеплера и показатель степени в законе всемирного тяготения Ньютона

Вслед за Ньютоном найдём показатель степени α в законе всемирного тяготения

$$F=G\frac{m_1m_2}{r^\alpha}.$$

Используем предыдущую задачу и известный Ньютону третий закон Кеплера, который применительно к круговым орбитам означает, что квадраты периодов обращения планет (относительно центрального тела массы M) относятся как кубы радиусов их орбит. В силу результата предыдущей задачи и закона всемирного тяготения с пока ещё не найденным показателем α имеем

$$\left(\frac{P_1}{P_2}\right)^2=\left(\frac{m_1r_1}{F_1}\right)\Big/\left(\frac{m_2r_2}{F_2}\right)=\left(\frac{m_1r_1}{m_1\frac{M}{r_1^\alpha}}\right)\Bigg/\left(\frac{m_2r_2}{m_2\frac{M}{r_2^\alpha}}\right)=\left(\frac{r_1}{r_2}\right)^{\alpha+1}.$$

Но по закону Кеплера $\left(\frac{P_1}{P_2}\right)^2=\left(\frac{r_1}{r_2}\right)^3$.

Значит, $\alpha=2$.

§3. Период колебаний тяжёлого маятника (включение g)

После сделанных при решении первой задачи подробных разъяснений мы можем теперь позволить себе более компактное изложение, с остановками только на каких-то новых обстоятельствах.

Найдём период колебаний маятника. Точнее: груз массы m, закреплённый на конце невесомого подвеса длины l, отклоняется от положения равновесия на некоторый начальный угол φ_0, освобождается и под действием силы тяжести начинает совершать периодические колебания. Мы ищем период P этих колебаний.

Написать, что $P = f(l, m, \varphi_0)$, было бы ошибочно, ибо на Земле и на Луне маятник имеет разные периоды колебаний ввиду разницы в силе тяжести на поверхности этих тел. Сила тяжести на поверхности тела, например Земли, характеризуется величиной g ускорения свободного падения у поверхности этого тела.

Поэтому вместо невозможного соотношения $P = f(l, m, \varphi_0)$ следует считать, что $P = f(l, m, g, \varphi_0)$.

Напишем векторы размерности всех этих величин в базисе $\{L, M, T\}$:

	P	l	m	g	φ_0
L	0	1	0	1	0
M	0	0	1	0	0
T	1	0	0	−2	0

Видно, что векторы $[l]$, $[m]$, $[g]$ независимы и $[P] = \frac{1}{2}[l] - \frac{1}{2}[g]$.

В силу П-теоремы в форме соотношения (6) главы I отсюда следует, что

$$P = \left(\frac{l}{g}\right)^{1/2} \cdot f(1, 1, 1, \varphi_0).$$

Мы нашли, что $P = c(\varphi_0) \cdot \sqrt{\frac{l}{g}}$, где безразмерный множитель $c(\varphi_0)$ зависит только от безразмерного угла φ_0 начального отклонения (измеряемого в радианах).

Точное значение $c(\varphi_0)$ тоже можно найти, хотя на сей раз это уже не так просто. Это можно сделать, решая уравнение колебаний тяжёлого маятника и используя эллиптический интеграл

$$F(k, \varphi) := \int_0^{\varphi} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}.$$

А именно, $c(\varphi_0) = 4K\left(\sin\left(\frac{1}{2}\varphi_0\right)\right)$, где $K(k) := F\left(k, \frac{1}{2}\pi\right)$.

§ 4. Расход объёма и массы на водосливе

На широком уступе, имеющем форму ступеньки, находящаяся на верхней платформе вода, падая под действием силы тяжести, образует водопад. Глубина воды на верхней платформе известна и равна h. Надо найти удельный (за единицу времени на единице ширины уступа) объёмный расход воды V.

Если мы правильно себе представляем механизм явления, то $V = f(g, h)$.

Поскольку это явление определяется гравитацией, помимо размерной постоянной g — ускорения свободного падения — можно бы для осторожности ввести и плотность жидкости ϱ, т. е. считать, что

$$V = f(\varrho, g, h).$$

Проделаем стандартную процедуру нахождения векторов размерностей:

	V	ϱ	g	h
L	2	−3	1	1
M	0	1	0	0
T	−1	0	−2	0

Видно, что векторы $[\varrho]$, $[g]$, $[h]$ независимы и $[V] = \frac{1}{2}[g] + \frac{3}{2}[h]$. В силу П-теоремы теперь получаем, что

$$V = g^{1/2}h^{3/2} \cdot f(1, 1, 1).$$

Таким образом, $V = c \cdot g^{1/2}h^{3/2}$, где c — постоянный множитель, подлежащий определению, например в лабораторном эксперименте.

Удельный расход массы Q при этом, очевидно, равен ϱV. К этой же формуле можно было бы прийти, применив процедуру метода размерностей к соотношению $Q = f(\varrho, g, h)$.

§ 5. Сила сопротивления при движении шара в невязкой среде

Шар радиуса r перемещается со скоростью v в невязкой среде, имеющей плотность ϱ. Надо найти силу сопротивления, действующую при этом на шар. (Конечно, можно было бы считать, что на покоящийся шар со скоростью v набегает поток, — типичная ситуация испытаний в аэродинамической трубе.)

Напишем общее соотношение $F=f(\varrho, v, r)$ и проанализируем его по размерностям:

	F	ϱ	v	r
L	1	−3	1	1
M	1	1	0	0
T	−2	0	−1	0

Видно, что векторы $[\varrho]$, $[v]$, $[r]$ независимы и $[F]=[\varrho]+2[v]+$ $+2[r]$.

В силу П-теоремы теперь получаем, что

$$F=\varrho v^2 r^2\cdot f(1,1,1). \tag{1}$$

Таким образом, $F=c\cdot\varrho v^2 r^2$, где c — безразмерный постоянный коэффициент.

§6. Сила сопротивления при движении шара в вязкой среде

Прежде чем переходить к формулировке этой задачи, напомним понятие вязкости среды и найдём размерность вязкости.

Если на поверхность густого мёда положить лист бумаги, то, чтобы перемещать лист вдоль поверхности, придётся прикладывать определённые усилия. В первом приближении сила F, прикладываемая к прилипшему к поверхности мёда листу, будет пропорциональна площади S листа, скорости v его перемещения и обратно пропорциональна расстоянию h от поверхности до дна, где мёд тоже прилип и остаётся неподвижным, несмотря на перемещение, происходящее наверху (как река).

Итак, $F=\eta\cdot Sv/h$. Коэффициент η в этом соотношении зависит от среды (мёд, вода, воздух...) и называется *коэффициентом вязкости среды* или просто *вязкостью*.

Отношение $\nu=\eta/\varrho$, где, как всегда, ϱ — плотность среды, часто встречается в задачах гидродинамики и называется *кинематической вязкостью среды*.

Найдём размерности этих величин в стандартном базисе $\{L, M, T\}$. Поскольку $[\eta]=[FhS^{-1}v^{-1}]$, функция размерности, отвечающая вязкости, в указанном базисе имеет вид $\varphi_\eta=L^{-1}M^1T^{-1}$, а вектор размерности — $[\eta]=(-1,1,-1)$.

Для кинематической вязкости $[\nu]=[\eta/\varrho]$, поэтому $\varphi_\nu=L^2M^0T^{-1}$ и $[\nu]=(2,0,-1)$.

Теперь попробуем решить предыдущую задачу о силе сопротивления, возникающей при движении того же шарика, но уже в вязкой среде.

Исходная зависимость теперь будет выглядеть так: $F=f(\eta,\varrho,v,r)$.

Проанализируем её по размерностям:

	F	η	ϱ	v	r
L	1	−1	−3	1	1
M	1	1	1	0	0
T	−2	−1	0	−1	0

Видно, что векторы $[\varrho],[v],[r]$ независимы; $[F]=[\varrho]+2[v]+2[r]$ и $[\eta]=[\varrho]+[v]+[r]$.

В силу П-теоремы теперь получаем, что

$$F=\varrho v^2r^2\cdot f(\mathrm{Re}^{-1},1,1,1), \tag{2}$$

где осталась неизвестная функция $f(\mathrm{Re}^{-1},1,1,1)$, зависящая от одного безразмерного параметра

$$\mathrm{Re}=\frac{\varrho vr}{\eta}=\frac{vr}{\nu}, \tag{3}$$

играющего ключевую роль в вопросах гидродинамики.

Эта безразмерная величина Re (показатель соотношения сил инерции и вязкости) называется *числом Рейнольдса* по имени английского физика и инженера Осборна Рейнольдса (Osborn Reynolds), впервые обратившего на неё внимание в своих работах 1883 г. по турбулентности.

Оказывается, при увеличении числа Рейнольдса, например при росте скорости течения или при уменьшении вязкости среды, характер течения претерпевает структурные превращения (называемые бифуркациями), эволюционируя от спокойного устойчивого ламинарного течения к турбулентности и хаосу.

Весьма поучительно приостановиться на этом месте и удивиться кажущемуся принципиальному совпадению результатов (1) и (2) последних двух задач.

Удивление исчезает, если повнимательнее рассмотреть переменную величину $f(\mathrm{Re}^{-1},1,1,1)$.

В предположениях, которые на современном языке равносильны условию относительной малости числа Рейнольдса, Стокс (Stokes) ещё в 1851 г. нашёл, что $F=6\pi\eta vr$. Это не противоречит формуле (2), а лишь говорит о том, что при малых числах Рейнольдса функция $f(\mathrm{Re}^{-1}, 1, 1, 1)$ асимптотически ведёт себя как $6\pi\mathrm{Re}^{-1}$. Действительно, подставляя это значение в формулу (2) и вспоминая определение (3) числа Рейнольдса, получаем указанную формулу Стокса.

§7. Упражнения

1. Если существуют оркестры, то, наверное, скорость звука слабо зависит (или не зависит) от длины волны?

(Вспомните природу звуковой волны, введите модуль упругости среды E и исходя из зависимости $v = f(\varrho, E, \lambda)$ покажите, что $v = c \cdot (E/\varrho)^{1/2}$.)

2. Каков закон изменения скорости распространения ударной волны, возникающей при очень сильном взрыве в атмосфере?

(Введите энергию взрыва E_0. Давлением перед ударной волной можно пренебречь; упругость воздуха уже роли не играет. Найдите сначала закон $r = f(\varrho, E_0, t)$ распространения ударной волны.)

3. Получите формулу $v = c \cdot (\lambda g)^{1/2}$ скорости распространения волны по глубокому водоёму под действием силы тяжести (здесь c — числовой коэффициент, g — ускорение свободного падения, λ — длина волны).

4. Скорость распространения волн в мелкой воде не зависит от длины волны. Приняв это наблюдение как факт, покажите, что она пропорциональна квадратному корню из глубины водоёма.

5. Формула, используемая для определения количества жидкости, протекающей по цилиндрической трубке (например по артерии), имеет вид

$$v = \frac{\pi \varrho P r^4}{8\eta l},$$

где v — скорость потока, ϱ — плотность жидкости, P — разность давлений на концах трубки, r — радиус сечения трубки, η — вязкость жидкости, l — длина трубки.

Выведите эту формулу (с точностью до числового множителя), проверив предварительно согласованность размерностей обеих её частей.

6. а) Животным пустыни приходится преодолевать большие расстояния между источниками воды. Как зависит максимальное время, которое может бежать животное, от размера животного L?

(Считайте, что испарение происходит только с поверхности, величина которой пропорциональна L^2.)

б) Как зависит скорость бега животного по ровному месту и в гору от размера животного?

(Считайте, что развиваемая мощность и соответствующая интенсивность тепловыделения (пусть испарения) пропорциональны, а сила сопротивления горизонтальному движению (воздух и пр.) пропорциональна квадрату скорости и площади фронтальной поверхности.)

в) Как зависит от размера животного расстояние, которое оно может пробежать?

(Сопоставьте ответы на два предыдущих вопроса.)

г) Как зависит от размера животного высота прыжка?

(Критическая нагрузка, которую выдерживает не слишком высокая колонна, пропорциональна площади сечения колонны. Допустите, что ответ на вопрос упирается только в прочность костей, а возможности мышц как раз прочности костей и соответствуют.)

Здесь всюду речь идёт о животных масштаба человека: верблюды, лошади, собаки, зайцы, кенгуру, тушканчики в их привычной среде обитания. См. в этой связи цитируемые ниже книги Арнольда и Смита.

7. Вслед за лордом Рэлеем найдите период малых колебаний капелек жидкости под действием их поверхностного натяжения, считая, что всё происходит вне гравитационного поля (в космосе).

(Ответ: $c\cdot(\varrho r^3/s)^{1/2}$, где ϱ — плотность жидкости; r — радиус капли; s — поверхностное натяжение, $[s]=(0,1,-2)$.)

8. Найдите период обращения двойной звезды.

Имеется в виду, что два тела с массами m_1 и m_2 вращаются по круговым орбитам вокруг их общего центра масс. Система находится в пустом пространстве и удерживается силами взаимного тяготения этих тел.

(Если возникнут недоумения, вспомните о гравитационной постоянной и её размерности.)

9. «Откройте» закон смещения Вина $\varepsilon(\nu,T)=\nu^3F(\nu/T)$ и закон Рэлея—Джинса $\varepsilon(\nu,T)=\nu^2TG(\nu/T)$ распределения интенсивности излучения абсолютно чёрного тела как функции частоты и абсолютной температуры.

[Основной закон Вина (не закон смещения, приведённый выше) имеет вид $\varepsilon(\nu,T)=\nu^3\exp(-a\nu/T)$ и справедлив при $\nu/T\gg 1$, а закон Рэлея—Джинса $\varepsilon(\nu, T) = 8\pi\nu^2 kT/c^3$ справедлив при относительно малых значениях ν/T.

Оба эти закона (удельной интенсивности излучения в интервале частот от ν до $\nu + d\nu$) объединяет знаменитая, открывающая эпоху квантовой физики, формула Планка (1900 г.)

$$\varepsilon(\nu, T) = \frac{8\pi}{c^3}\nu^2 \frac{h\nu}{e^{h\nu/(kT)} - 1},$$

где c — скорость света, h — постоянная Планка, k — постоянная Больцмана ($k = R/N$, R — газовая постоянная, N — число Авогадро). Законы Вина и Рэлея—Джинса получаются из формулы Планка соответственно при $h\nu \gg kT$ и $kT \gg h\nu$.]

Пусть ν_T — та частота, при которой функция $\varepsilon(\nu, T) = \nu^3 F(\nu/T)$ достигает максимума при фиксированном значении температуры T. Вслед за Вином проверьте, что имеет место замечательный закон смещения $\nu_T/T = \mathrm{const}$. Найдите эту константу, используя формулу Планка.

10. Приняв гравитационную постоянную G, скорость света c и постоянную Планка h за основные единицы (их значения см., например, в книге [12]), найдите универсальные планковские единицы длины $L^* = (hG/c^3)^{1/2}$, времени $T^* = (hG/c^5)^{1/2}$ и массы $M^* = (hc/G)^{1/2}$.

(Значения $G = 6{,}67 \cdot 10^{-11}$ Н $\cdot$ м^2/кг^2, $c = 2{,}997925 \cdot 10^8$ м/с, $h = 6{,}625 \cdot 10^{-34}$ Дж $\cdot$ с, другие физические константы, а также сведения о единицах измерения можно найти, например, в книгах [4а], [4б], [4в].)

Много задач, разобранных примеров, поучительных рассуждений и предостережений, связанных с анализом размерностей и принципами подобия, можно найти в книгах [1—3].

§8. Заключительные замечания

Уже то немногое, что было сказано об анализе размерностей и его применении, позволяет сделать следующие наблюдения.

Эффективность использования метода в основном зависит от правильного понимания самой природы явления, к которому он применяется. (Кстати, вначале только люди уровня Ньютона, братьев Бернулли и Эйлера умели применять анализ бесконечно малых, обходясь без парадоксов, что требовало дополнительной прозорливости[1].)

Анализ размерностей особенно полезен, когда законы явления ещё не описаны. Именно в этой ситуации он даёт, хотя и очень общие, полученные полуэвристическим путём, связи, полезные, однако, для понимания механизма явления и выбора направления дальнейших исследований и уточнений. (Мы потом продемонстрируем это на примере подхода Колмогорова к описанию всё ещё таинственного фундаментального явления турбулентности.)

Основной постулат теории размерностей связан с линейной теорией преобразований подобия, теорией измерений, понятием твёрдого тела, однородностью пространства и т. п. В гиперболической геометрии Лобачевского подобных фигур, как известно, вообще нет, хотя локально эта геометрия вполне допускает евклидово приближение. Значит, как и все законы, постулат теории размерности и она сама применимы в каких-то масштабах, зависящих от задачи. Заранее эти масштабы редко бывают известны и чаще всего обнаруживаются тогда, когда возникают несоответствия.

Метод показывает, что чем больше размерно независимых величин мы имеем, тем больше упрощается и конкретизируется исследуемая функциональная зависимость. Но чем больше открывается физических связей, тем меньше остаётся размерно независимых величин. (Например, расстояние можно теперь измерять световыми годами.) Получается, что чем больше мы знаем, тем меньше нам даёт общий анализ размерностей. В противовес этому проникновение в принципиально новые сферы обычно сопровождается появлением

---

[1]Процитирую оправданные опасения В. И. Арнольда по поводу возможной переоценки П-теоремы: «Такой подход крайне опасен потому, что открывает возможность безответственных спекуляций (под названием „теория размерностей“) там, где нужно было бы экспериментально проверять соответствующие законы подобия, ведь они вовсе не следуют из размерностей описывающих изучаемое явление величин, а являются глубокими и не очевидными естественнонаучными фактами». Впрочем, это относится и к неумелому использованию таблицы умножения, статистики или теории катастроф.

В своей книжке «Математическое понимание природы» (М.: МЦНМО, 2009), обсуждая подобную физическую теорию адиабатических инвариантов, В. И. Арнольд на странице 117 подмечает: «Теория адиабатических инвариантов — это странный пример физической теории, противоречащей на вид легко проверяемым математическим фактам. Несмотря на такое неприятное свойство этой „теории“, она доставила замечательные физические открытия тем, кто не побоялся использовать её выводы (хотя они и были математически необоснованными)».

Словом: «Думайте сами, решайте сами, иметь или не иметь».

новых размерно независимых величин (об этом, об алгебраическом аспекте размерностей и многом другом сказано, например, в книге [15]).

Если не считать задач 9 и 10, мы здесь ограничились рассмотрением явлений, описываемых в рамках величин классической механики. Для начала этого, возможно, и хватит. Но истинное наслаждение можно получить, читая в оригинале рассуждения учёных, мыслителей и вообще профессионалов, способных к масштабному, многоплановому и одновременно единому взгляду на мир или предмет. Это относится к разным сферам. Это сродни симфонии. Это захватывает!

[Если вы смирились с «мракобесностью» анализа размерностей, но изложенное ещё не кажется вам достаточно безумным, то, улыбнувшись, приведу следующую выдержку из книжки известного физика (которого не называю, чтобы ненароком не подвергнуть его доброе имя нападкам людей менее свободомыслящих).

«Физики начинают изучение явления с введения подходящих единиц измерения. Неразумно измерять радиус атома в метрах или скорость электрона в километрах в час, необходимо найти подходящие единицы. Уже из одного такого введения единиц сразу же возникают важные следствия. Так, из заряда электрона e и его массы m нельзя составить величину, имеющую размерность длины. Это значит, что в классической механике атом невозможен — электрон не может двигаться стационарно. Положение изменилось с появлением постоянной Планка $\hbar$ ($\hbar = h/(2\pi)$). Как видно из определения, $\hbar = 1{,}054 \cdot 10^{-34}$ Дж · с имеет размерность энергии, умноженной на время[1]. Теперь можно составить величину размерности длины $a_0 = \hbar/(me^2)$.

Если в это соотношение подставить значения входящих сюда постоянных, то мы должны получить величину порядка размеров атома; получается $0{,}5 \cdot 10^{-10}$ м. Так из простой размерной оценки нашёлся размер атома.

Легко увидеть, что $e^2/\hbar$ имеет размерность скорости, она приблизительно в сто раз меньше скорости света. Если поделить эту величину на скорость света c, получится безразмерная величина $\alpha = e^2/(\hbar c) = 1/137$, характеризующая взаимодействие электрона

[1]Размерность h можно усмотреть из формулы Планка, приведённой в задаче 9.

с электромагнитным полем. Эту величину называют *постоянной тонкой структуры*.

Мы сделали оценки для атома водорода. Нетрудно получить их для атома с зарядом Ze. Движение электрона в атоме определяется его взаимодействием с ядром, которое пропорционально произведению заряда ядра на заряд электрона. Поэтому для ядра с зарядом Ze в формулах для α и a_0 нужно заменить e^2 на Ze^2. В тяжёлых элементах при $Z \sim 100$ скорость электронов приближается к скорости света».]

Сделаем наконец несколько практических замечаний.

Анализ размерностей является хорошим контролёром:

а) если размерности левой и правой частей равенства не совпадают, то надо искать ошибку;

б) если под знаком отличной от степенной функции (например, под знаком логарифма или экспоненты) появилась не безразмерная величина, то надо искать ошибку (или избавляющее от этого преобразование);

в) складывать можно только величины одинаковой размерности.

(Если $v = at$ — скорость, а $s = \frac{1}{2}at^2$ — путь в равноускоренном движении, то формально, конечно, верно, что $v + s = at + \frac{1}{2}at^2$. Однако с физической точки зрения это равенство приводится к двум: $v = at$ и $s = \frac{1}{2}at^2$. Бриджмен, в цитируемой книге которого приведён этот пример, указывает здесь на полную аналогию с равенством векторов, порождающим равенства одноимённых координат.)

Глава III

Дальнейшие приложения: гидродинамика и турбулентность

§1. Уравнения гидродинамики (общие сведения)

Основными классическими характеристиками движущейся сплошной среды (жидкости, газа), как известно, являются скорость $v = v(x, t)$, давление $p = p(x, t)$ и плотность среды $\varrho = \varrho(x, t)$ как функции точки x области течения и времени t.

Аналогом уравнения Ньютона $ma = F$ в случае движения идеальной сплошной среды является уравнение Эйлера

$$\varrho \frac{dv}{dt} = -\nabla p. \tag{1}$$

Если среда вязкая, то в правой части к силе, связанной с перепадом давления, добавится сила внутреннего трения и получится уравнение

$$\varrho \frac{dv}{dt} = -\nabla p + \eta \Delta v, \tag{2}$$

где η — вязкость среды.

Последнее уравнение было выведено в 1827 г. Навье (Navier) в некотором специальном случае, а затем последовательно обобщено Пуассоном (Poisson, 1831), Сен-Венаном (Saint-Venant, 1843) и Стоксом (Stokes, 1845). Поскольку v — векторное поле, это векторное уравнение равносильно системе уравнений на координаты поля v. Эта система называется системой уравнений Навье—Стокса и часто обозначается как NS-система.

При $\eta = 0$ мы возвращаемся к уравнению Эйлера, относящемуся к идеальной (невязкой) жидкости, для которой можно пренебречь потерями энергии при внутреннем трении.

К NS-уравнению добавляется ещё общее уравнение неразрывности

$$\frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho v) = 0, \tag{3}$$

выражающее в дифференциальной форме закон сохранения массы (изменение количества вещества в любой области течения совпадает с его потоком через границу этой области).

Для однородной несжимаемой жидкости $\varrho \equiv \text{const}$, $\operatorname{div}(v) = 0$, уравнение неразрывности выполнено автоматически, а уравнение (2) Навье—Стокса принимает вид

$$\frac{dv}{dt} = -\nabla\left(\frac{p}{\varrho}\right) + \nu\triangle v, \tag{4}$$

где $\nu = \eta/\varrho$ — кинематическая вязкость среды.

Если течение происходит в присутствии массовых сил (например, в гравитационном поле), то в правую часть NS-уравнения (2), естественно, добавится плотность f этих сил.

Далее, если полную производную, стоящую в левой части NS-уравнения, раскрыть, учитывая, что $\dot{x} = v$, то получим иную форму записи NS-уравнения

$$\partial_t v + (v\nabla)v = \nu\triangle v + \frac{1}{\varrho}(f - \nabla p). \tag{5}$$

Если течение стационарно, т. е. поле скорости v не зависит от времени, то последнее уравнение приобретает вид

$$(v\nabla)v = \nu\triangle v + \frac{1}{\varrho}(f - \nabla p). \tag{6}$$

Мы не имеем здесь целью далее погружаться в огромную область работ, связанных с уравнением Навье—Стокса. Напомним лишь, что следующая задача включена в список проблем века (и за её решение даже назначена неплохая премия): если начальные и граничные условия трёхмерного NS-уравнения (2) были гладкими, то будет ли гладкость решения сохраняться вечно или за конечное время могут возникнуть особенности? (В двумерном случае вопроса, как и особенностей, уже нет.)

С физической точки зрения, вероятно, больший интерес представляют иные проблемы, например вопрос о том, как из NS-уравнения получить удовлетворительное описание турбулентных течений и как происходит переход к турбулентности и хаосу.

Итак, уравнения динамики сплошной среды существуют (для реальных сред их дополняют термодинамическими уравнениями состояния). В ряде случаев эти уравнения динамики можно решить явно. В других случаях удаётся провести расчёт конкретных течений на компьютере. Но в целом тут ещё многое требует дальнейших и, возможно, принципиально новых исследований.

Вернёмся теперь к самому началу и представим себе, что у нас ещё нет уравнений Навье—Стокса, а мы всё же интересуемся течением сплошной среды. Пусть, например, поток однородной несжимаемой жидкости, имея на бесконечности скорость u, набегает на предмет, имеющий некий характерный размер l. Нас интересует стационарный режим обтекания, т. е. возникающее при этом векторное поле скорости $v = v(r)$ как функция радиус-вектора r точки x пространства относительно какой-то фиксированной системы декартовых координат. Пусть ϱ — плотность жидкости, η — её вязкость, а $\nu = \eta/\varrho$ — кинематическая вязкость.

Считая, что $v = f(r, \eta, \varrho, l, u)$, попробуем привлечь анализ размерностей:

	v	r	η	ϱ	l	u
L	1	1	−1	−3	1	1
M	0	0	1	1	0	0
T	−1	0	−1	0	0	−1

Отсюда в силу П-теоремы вытекает следующее соотношение между безразмерными величинами:

$$\frac{v}{u} = f\left(\frac{r}{l}, 1, \frac{\varrho u l}{\eta}, 1, 1\right), \tag{7}$$

в котором мы обнаруживаем число Рейнольдса $\mathrm{Re} := \frac{\varrho u l}{\eta} = \frac{ul}{\nu}$.

Мы получили бы то же самое, считая, что $v = f(r, \nu, \varrho, l, u)$, или если бы исходили из соотношения $v = f(r, \nu, l, u)$, в котором плотность спрятана в кинематическую вязкость ν.

Таким образом, можно менять значения ϱ, u, l, η, но если при этом не меняется их комбинация, выраженная числом Рейнольдса Re, то характер течения останется тем же, с точностью до замены масштаба измерения длины и времени (или длины и скорости). Замечательно!

§2. Потеря устойчивости течения и замечания о бифуркациях в динамических системах

Характер течения при различных значениях числа Рейнольдса, вообще говоря, различен. По мере возрастания числа Рейнольдса происходят перестройки (бифуркации) топологии течения. Его характер меняется от устойчивого ламинарного течения к турбулентности и хаосу: при $\mathrm{Re} \asymp 10^0$ течение ламинарное; затем при $\mathrm{Re} \asymp 10^1$ появляются первое критическое значение Re_1 и первая бифуркация (первая перестройка топологии течения) и т. д. Последовательность

$\mathrm{Re}_1 < \mathrm{Re}_2 < \ldots < \mathrm{Re}_n < \ldots$ критических значений быстро сходится, что, впрочем, является довольно универсальным явлением.

(Впервые определённая универсальность была обнаружена в 1978 г. Фейгенбаумом (M. J. Feigenbaum), она носит его имя (*универсальность Фейгенбаума*) и состоит в том, что существует предел

$$\lim_{n\to\infty} \frac{R_{n+1} - R_n}{R_n - R_{n-1}} = \delta^{-1},$$

где $\{R_n, n \in \mathbb{N}\}$ — последовательность критических значений, при которых происходит перестройка динамической системы, называемая бифуркацией удвоения периода, а $\delta = 4{,}6692\ldots$)

Итак, последовательность чисел Re_n имеет предел Re_∞.

При дальнейшем увеличение параметра задачи Re за значение Re_∞ возникает режим, который в гидродинамике называют турбулентностью. При очень больших значениях числа Рейнольдса движение становится совсем хаотическим (как бы не детерминированным случайным процессом).

До сих пор открыт принципиальный вопрос о том, в какой мере вообще можно подойти к описанию турбулентности, исходя из классических гидродинамических уравнений Навье—Стокса.

Здесь вспоминаются слова Ричарда Фейнмана, который по другому поводу, если я не ошибаюсь, говорил, что, может быть, уравнение Шрёдингера уже и содержит в себе формулу жизни, но это не отменяет биологию, которая исследует живую клетку, не дожидаясь, пока существование жизни будет обосновано через уравнение Шрёдингера.

Проявив себя истинным естествоиспытателем, А. Н. Колмогоров в 1941 г. предложил модель развитой турбулентности, которая, хотя и уточнялась впоследствии, осталась базовой и заслуживающей специального обсуждения. Именно модель Колмогорова мы выбрали для демонстрации нетривиального применения анализа размерностей при исследовании явления, никаким фундаментальным описанием которого мы пока не располагаем.

Конечно, серьёзным новым общим достижением теории динамических систем было открытие странных аттракторов (Лоренц, E. N. Lorenz, 1963). Это позволило конкретизировать идеологию возникновения хаотических проявлений детерминированной системы как её чувствительности к малым изменениям начальных условий, а также дать возможное общединамическое объяснение явления турбулентности (Рюэль и Такенс, D. Ruelle & F. Takens, 1971). (См. также комментарий к статье [5]).

§3. Турбулентность (начальные представления)

В сборнике статей о турбулентности [12] академик О. М. Белоцерковский вспоминает, что, когда он учился на физическом факультете Московского университета, лекции по электричеству им читал профессор С. Г. Калашников, который на первой же лекции рассказал следующее. Однажды на экзамене он (Калашников) спросил студента: «Что такое электричество?» Студент заёрзал, засуетился и говорит: «Вот ещё вчера знал, а забыл». На это Калашников заметил: «Был один человек, который знал, и тот забыл!» Ситуация с турбулентностью примерно такая же, хотя об этом размышляют люди разных специальностей, и, конечно, физики, математики, и астрономы в первую очередь.

Быстрая река, обтекая опоры моста, образует вихри, которые рисовал проникавший во всё Леонардо да Винчи. Вихри можно наблюдать и в воздухе по облакам пыли за мчащейся машиной или, что гораздо приятнее, просто по облакам, причудливо деформирующимся на глазах. Космические вихри порождают галактики. Вода из крана перестаёт течь спокойно, когда кран открывают слишком сильно. Маленькому самолёту не разрешат взлёт сразу за крупным лайнером. А через иллюминатор самолёта интересно наблюдать кажущиеся сверху крохотными океанские суда, за которыми тянется отчётливо различимый турбулентный след.

В науке термин *турбулентность* закрепился, по-видимому, после Максвелла, лорда Кельвина и Рейнольдсa к концу XIX века, хотя его адекватно употреблял уже Леонардо да Винчи.

§4. Модель Колмогорова

4.1. Многомасштабность турбулентных движений

Рассмотрим, как и выше, обтекание тела, т. е. поток, набегающий на тело, имеющее характерный размер l. Если скорость потока большая или вязкость среды маленькая, т. е. при очень больших значениях числа Рейнольдса, в некотором объёме за телом возникнет область развитой турбулентности. В ней течение крайне неустойчиво, хаотично, носит характер пульсаций разных масштабов, распределённых по области турбулентности. Изменение поля скоростей $v = v(x, t)$ напоминает здесь стационарный случайный процесс, когда стационарно не само поле $v = v(x, t)$, а некоторые его осреднённые вероятностные характеристики (выражаемые, например, гистограммами распределения вероятностей тех или иных связанных с потоком величин).

А. Н. Колмогоров заметил, что при очень больших значениях числа Рейнольдса в области турбулентности картина течения хоть и сложна, но локально однородна и изотропна.

Турбулентность можно (или нужно) рассматривать как проявление взаимодействия движений разных масштабов. На движения больших масштабов накладываются пульсации меньших масштабов, и они переносятся движениями больших масштабов (пассажиры поезда, перемещаясь по вагону-ресторану, конечно, участвуют в движении масштабов расстояния между городами, но при этом есть на что посмотреть и в масштабах вагона).

Поясним. Клетка живёт своей жизнью. Сообщества клеток, взаимодействуя, образуют определённую ткань. Группа тканей составляет орган. Группа органов — организм. Организм садится в машину и едет на работу. На улицах города возникает транспортный поток. Все эти потоки вместе с городами и странами переносятся в пространстве вращающейся Землёй и т. д. Ещё ближе к теме мог бы быть пример, относящийся к многомасштабной жизни океана или атмосферы.

Когда мы говорим о перемещении пассажира, то, разумеется, имеем в виду характеристики его движений в масштабах вагона. Мы не различаем индивидуального пассажира, когда говорим о скорости в масштабах движения всего поезда.

Турбулентность — это прежде всего относительные движения в участке жидкости, а не абсолютное переносное движение, в котором он участвует как элемент движения более крупного масштаба.

Если так, то в турбулентности перед нами целый спектр движений различных масштабов и мы, например, хотим указать какие-то характерные параметры движений различных масштабов: распределение энергии турбулентного потока по масштабам движений, относительные скорости движений различных масштабов, скорость расхождения частиц в турбулентном потоке и т. п.

4.2. Развитая турбулентность и инерционный интервал

Теперь перейдём к более конкретному разговору.

Рассмотрим поток, например, как и выше, обтекающий тело характерного размера l, при больших значениях числа Рейнольдса ($\mathrm{Re} \gg 1$). Возникающую при $\mathrm{Re} \gg 1$ турбулентность принято называть *развитой турбулентностью*.

Развитую турбулентность в движениях масштабов $\lambda \ll l$ и вдали от твёрдых стенок (т. е. на расстоянии от них, много большем, чем величина λ) будем вслед за Колмогоровым считать изотропной и однородной.

Условие $\mathrm{Re} \gg 1$ можно трактовать как малость вязкости.

Вязкость сказывается только в самых мелкомасштабных движениях некоторого масштаба λ_0, так как внутреннее трение возникает лишь между близкими участками жидкости. При переносе больших масс жидкости вязкость несущественна, так как диссипируемая в трении энергия в этом случае пренебрежимо мала по сравнению с кинетической энергией инерционного движения большой массы.

Промежуток масштабов $\lambda_0 \ll \lambda \ll l$ называют *инерционным интервалом*. В движениях этих масштабах вязкостью можно пренебрегать.

Величина λ_0 называется *внутренним масштабом турбулентного движения*, а l — *внешним масштабом движения*.

4.3. Удельная энергия

Установившийся турбулентный режим поддерживается затратой внешней энергии, которая диссипируется за счёт вязкости.

Пусть ε — удельная мощность диссипации, точнее, это энергия, диссипируемая единицей массы жидкости в единицу времени. В соответствии с этим определением величина ε имеет следующий вектор размерности в стандартном базисе $\{L, M, T\}$: $[\varepsilon] = (2, 0, -3)$.

[Сила $F = ma$ имеет размерность $(1, 1, -2)$. Энергия, работа, потенциальная энергия $F \cdot h$ имеют размерность $(2, 1, -2)$. Значит, $[\varepsilon] = (2, 0, -3)$.]

Кинетическая энергия потока, набегающего со скоростью u, уменьшается вследствие диссипации энергии при внутреннем трении в вязкой жидкости. Изменение Δu средней скорости основного движения происходит на протяжённости порядка l (до встречи с телом и после этого).

Величина ε должна определяться этой потерей кинетической энергии основного движения, т. е. должна быть функцией $\varepsilon = f(\varrho, \Delta u, l)$.

На основании П-теоремы тогда заключаем, что по порядку величины

$$\varepsilon \sim \frac{(\Delta u)^3}{l}. \tag{8}$$

Аналогично для перепада давления можно получить соотношение

$$\Delta p \sim (\Delta u)^2 \varrho. \tag{9}$$

4.4. Число Рейнольдса движений данного масштаба

Поскольку интерес представляют движения разных масштабов, с каждым масштабом λ связывают соответствующее ему число Рейнольдса

$$\mathrm{Re}_\lambda := \frac{v_\lambda \cdot \lambda}{\nu}. \tag{10}$$

В этих терминах внутренний масштаб турбулентности, т. е. величина λ_0, должен определяться условием, что по порядку величины $\mathrm{Re}_{\lambda_0} \sim 1$, поскольку большие значения числа Рейнольдса равносильны малой вязкости.

4.5. Закон Колмогорова—Обухова

Найдём теперь средние скорости v_λ движений масштаба λ (или, что то же самое, изменение средней скорости турбулентного потока на протяжении расстояний порядка λ).

В инерционном интервале, когда $\lambda_0 \ll \lambda \ll l$, можно считать, что $v_\lambda = f(\varrho, \varepsilon, \lambda)$.

На основании П-теоремы тогда заключаем, что по порядку величины

$$v_\lambda \sim (\varepsilon\lambda)^{1/3}. \tag{11}$$

Это соотношение называется *законом Колмогорова—Обухова*.

(А. М. Обухов — ученик А. Н. Колмогорова 1940-х годов, впоследствии академик, директор Института физики атмосферы в Москве. Другой ученик А. Н. Колмогорова той же поры А. С. Монин, тоже стал академиком, но директором Института океанографии. По этому поводу А. Н. Колмогоров шутил, что один его ученик заведует океаном, а другой — атмосферой.)

4.6. Внутренний масштаб турбулентности

Найдём теперь внутренний масштаб λ_0 турбулентного течения. Как мы знаем, его надо искать из условия, что $\mathrm{Re}_{\lambda_0} \sim 1$.

Для числа Рейнольдса Re основного движения масштаба l в соответствии с общим определением числа Рейнольдса и в согласии с формулой (10) имеем $\mathrm{Re} \sim (\Delta u \cdot l)/\nu$. Учитывая соотношения (8) и (11), получаем

$$\mathrm{Re}_\lambda \sim \frac{v_\lambda \cdot \lambda}{\nu} \sim \frac{(\varepsilon\lambda)^{1/3}\lambda}{\nu} \sim \frac{\Delta u \cdot (\lambda)^{4/3}}{\nu l^{1/3}} = \mathrm{Re}\Big(\frac{\lambda}{l}\Big)^{4/3}.$$

Полагая $\mathrm{Re}_\lambda \sim 1$, находим

$$\lambda_0 \sim \frac{l}{\mathrm{Re}^{3/4}}. \tag{12}$$

Для соответствующей скорости имеем

$$v_{\lambda_0} \sim (\varepsilon\lambda_0)^{1/3} \sim \frac{\Delta u}{l^{1/3}} \cdot \frac{l^{1/3}}{\mathrm{Re}^{1/4}} = \frac{\Delta u}{\mathrm{Re}^{1/4}}. \tag{13}$$

4.7. Энергетический спектр турбулентных пульсаций

Масштабу длины λ, как длине волны, сопоставим волновое число $k := 1/\lambda$. Пусть $E(k)dk$ — кинетическая энергия в пульсациях (движениях) с волновым числом k в промежутке dk, отнесённая к единице массы жидкости.

Найдём плотность $E(k)$ этого распределения. Поскольку $E(k)dk$ имеет размерность энергии, отнесённой к единице массы, а $[dk] = (-1, 0, 0)$, находим, что $[E(k)] = (3, 0, -2)$.

Комбинируя ε и k, из соображений размерности вслед за Колмогоровым находим

$$E(k) \sim \varepsilon^{2/3} k^{-5/3}. \tag{14}$$

Считая, что v_λ определяет порядок величины кинетической энергии движений всех масштабов, не превосходящих λ, отсюда можно вновь получить закон Колмогорова—Обухова

$$v_\lambda^2 \sim \int\limits_{k=1/\lambda}^{\infty} E(k)dk \sim \varepsilon^{2/3} k^{-2/3} \sim (\varepsilon\lambda)^{2/3} \quad \text{и} \quad v_\lambda \sim (\varepsilon\lambda)^{1/3}.$$

4.8. Турбулентное перемешивание и расхождение частиц

Две частицы, находящиеся в турбулентном потоке на взаимном расстоянии λ, через промежуток времени t окажутся разведёнными на расстояние $\lambda(t)$. Найдём скорость $\lambda'(t)$ расхождения частиц.

Как и при выводе закона Колмогорова—Обухова, считаем, что $\lambda' = f(\varrho, \varepsilon, \lambda)$, и в полном соответствии с соотношением (11) находим

$$\frac{d\lambda}{dt} \sim (\varepsilon\lambda)^{1/3}. \tag{15}$$

Как видно, скорость расхождения растёт вместе с λ. Это объясняется тем, что в рассматриваемом процессе участвуют лишь движения масштабов, меньших λ. Движения больших масштабов переносят частицы, не приводя к их расхождению.

Тема II

Многомерная геометрия и функции очень многих переменных

Введение

Почти весь объём многомерного тела сосредоточен у его границы. Например, если с 1000-мерного арбуза радиуса один метр снять корку толщиной один сантиметр, то останется меньше тысячной доли всего арбуза.

Это явление локализации или концентрации меры имеет многочисленные порой неожиданные проявления.

Например, любая более или менее регулярная функция на многомерной сфере почти постоянна в том смысле, что если взять случайно и независимо пару точек сферы и подсчитать значения функции в этих точках, то они с большой вероятностью окажутся почти совпадающими.

С точки зрения математика, привыкшего к функциям одного, двух и нескольких (не очень многих) переменных, это может показаться неправдоподобным. Но именно это обеспечивает стабильность основных параметров среды нашего обитания (температуры, давления...), лежит в основе статистической физики, изучается в теории вероятностей под названием законов больших чисел и имеет много применений (например, при передаче информации по каналу связи при наличии помех).

Явление концентрации меры в некотором отношении поясняет как статистическую устойчивость значений термодинамических величин, породившую эргодическую гипотезу Больцмана, так и замечательные эргодические теоремы, возникшие в её обоснование.

Принцип концентрации изложен в главе II, которую можно читать независимо от главы I, где приведены примеры областей знаний, в которых функции очень многих переменных появляются естественным образом.

В главе I (которую, разумеется, тоже можно читать независимо) мы подробнее остановимся на одном менее популярном примере — передаче информации по каналу связи. Выведем и обсудим теорему отсчётов — формулу Котельникова — основу современной цифровой записи сигнала. В главе III мы дополним эти рассмотрения теоремой Шеннона о скорости передачи информации по каналу связи при наличии помех.

Глава I

Некоторые примеры функций очень многих переменных в естествознании и технике

§1. Цифровая запись сигнала (КИМ — кодово-импульсная модуляция)

1.1. Линейный прибор и его математическое описание (свёртка)

Хорошей математической моделью многих приборов и аппаратов является линейный оператор.

Слова «линейный прибор с инвариантными во времени свойствами» на математическом языке означают, что это линейный оператор A, действующий на функции от времени и коммутирующий с оператором сдвига T, т. е. $AT = TA$, где $T = T_\tau$, а $(T_\tau f)(t) = f(t-\tau)$.

Например, если A — проигрыватель, то он завтра должен выдать с того же диска ту же музыку, что и сегодня, лишь с естественным сдвигом во времени.

В соответствии с принятой у радиоинженеров терминологией, функцию f, на которую действует оператор A, будем называть *сигналом*, точнее — *входным сигналом* или *входом*, а результат Af обработки сигнала f прибором A будем называть *сигналом на выходе* или *выходом* и обозначать $\widetilde{f}$.

Поскольку непрерывную функцию f можно хорошо аппроксимировать ступенчатой функцией, легко понять, что если знать отклик такого прибора A на элементарную ступеньку, то можно найти его отклик и на любой входной сигнал f.

В идеале ступенька превращается в единичный импульс — δ-функцию. Если написать тождество $f(t) = \int f(\tau)\delta(t-\tau)\,d\tau$, то сразу найдём, что

$$Af(t) = \int f(\tau)A\delta(t-\tau)\,d\tau = \int f(\tau)\widetilde{\delta}(t-\tau)\,d\tau =: f*\widetilde{\delta},$$

где $*$ — символ операции свёртки функций. Значит, $Af = f*\widetilde{\delta}$.

Функция $A\delta = \widetilde{\delta}$, т. е. отклик прибора на единичный импульс (δ-функцию), называется *аппаратной функцией прибора* и часто обозначается символом E.

Итак, с математической точки зрения прибор A есть не что иное, как просто оператор свёртки $Af = f * \widetilde{\delta} = f * E$.

Решение уравнений в свёртках имеет, таким образом, весьма конкретные прямые приложения (например, восстановление переданного сигнала f по принятому сигналу $Af = \widetilde{f}$).

1.2. Фурье-двойственное (спектральное) описание линейного прибора

Напомним, что частоту ν периодического процесса обычно измеряют количеством полных циклов в единицу времени (один Герц — это одно полное колебание в секунду; обозначается 1 Гц или 1 Hz). Угловая или круговая частота $\omega = 2\pi\nu$ отличается от частоты ν только множителем 2π, отвечающим за переход к измерению в радианах в единицу времени.

Подсчитаем отклик $\widetilde{f}$ прибора на входной сигнал $f = e^{i\omega t}$ (тем самым, в силу формулы Эйлера $e^{i\omega t} = \cos\omega t + i\sin\omega t$ узнаем отклик прибора и на простейшие гармонические колебания $\sin\omega t$ круговой частоты ω; комплексный язык удобен):

$$Af = f * \widetilde{\delta} = f * E = \int f(t-\tau)\widetilde{\delta}(\tau)\,d\tau = \\ = \int e^{i\omega(t-\tau)}E(\tau)\,d\tau = \Big(\int E(\tau)e^{-i\omega\tau}\,d\tau\Big)e^{i\omega t} = P(\omega)e^{i\omega t}.$$

Мы получили колебание той же частоты, что и на входе, но, возможно, изменённое по амплитуде в $|P(\omega)|$ раз и по фазе на аргумент $\arg P(\omega)$ комплексного числа $P(\omega)$.

Величина P как функция от ω называется *спектральной характеристикой прибора*. Спектральная характеристика прибора, как видно, является (с точностью до нормирующего множителя) преобразованием Фурье $\widehat{E}$ аппаратной функции E этого прибора: $P = 2\pi\widehat{E}$. Условимся, что $p(\nu) := P(2\pi\nu) = P(\omega)$.

Напомним, что в терминах частот ω и ν преобразование Фурье $\widehat{f}$ и $\check{f}$ соответственно и интеграл Фурье функции f (формула обращения преобразования Фурье в пространстве L_2) имеют вид

$$f(t) = \int \widehat{f}(\omega)e^{i\omega t}\,d\omega, \quad \text{где } \widehat{f}(\omega) = \frac{1}{2\pi}\int f(t)e^{-i\omega t}\,dt;$$
$$f(t) = \int \check{f}(\nu)e^{i2\pi\nu t}\,d\nu, \quad \text{где } \check{f}(\nu) = \int f(t)e^{-i2\pi\nu t}\,dt.$$

Поскольку преобразование Фурье обратимо, по функции P (или по p) восстанавливается функция E. Значит, спектральная характеристика или спектральная функция P (или p) прибора, так же как и его аппаратная функция E, полностью определяет прибор A.

Вычислим теперь Af, зная P. Представляя f интегралом Фурье, находим представление Af в виде интеграла Фурье:

$$f(t) = \int \widehat{f}(\omega)e^{i\omega t}\,d\omega \quad \text{и} \quad Af(t) = \int \widehat{f}P(\omega)e^{i\omega t}\,d\omega.$$

В частности, если $f=\delta$, то

$$E(t) = \widetilde{\delta}(t) = A\delta(t) = \frac{1}{2\pi}\int P(\omega)e^{i\omega t}\,d\omega \quad \Big(= \int \widehat{E}(\omega)e^{i\omega t}\,d\omega\Big).$$

1.3. Функции и приборы с финитным спектром

Приборы, с которыми реально приходится иметь дело, подобно ушам и глазам, способны воспринимать сигнал только в определённом диапазоне частот. Поэтому представляется естественным обратить особое внимание на функции и приборы с финитным спектром.

Если спектр (преобразование Фурье $\widehat{f}$) функции f финитен (тождественно равен нулю вне некоторого компакта), то в её представлении в виде интеграла Фурье участвует интеграл, взятый только по конечному промежутку:

$$f(x) = \int_{-a}^{a} \widehat{f}(\omega)e^{i\omega x}\,d\omega.$$

Функции с финитным спектром ввиду их важности были предметом самостоятельного математического исследования.

Если функция g принадлежит пространству $L_2(\mathbb{R})$, то, как известно из теории преобразования Фурье, функция

$$f(x) = \int_{-a}^{a} g(\omega)e^{i\omega x}\,d\omega$$

тоже принадлежит пространству $L_2(\mathbb{R})$. Более того, применив неравенство Коши—Буняковского—Шварца, легко увидеть, что она ограничена на вещественной оси и распространяется на всю комплексную плоскость как целая функция и при этом

$$|f(x+iy)| \leqslant ce^{a|y|}.$$

Класс целых функций такого вида называют классом Винера и обозначают W_a (см., например, [4]).

Теорема Пэли—Винера (Paley—Wiener) утверждает, что $f \in W_a$ тогда и только тогда, когда эта функция допускает представление

$$f(z) = \int_{-a}^{a} g(\omega) e^{i\omega z}\, d\omega, \quad \text{где } g \in L_2(\mathbb{R}).$$

1.4. Идеальный фильтр и его аппаратная функция

Теперь обратимся к прибору с финитным спектром. Простейшим и основным примером тут может служить прибор, спектральная функция которого $P(\omega)$ равна единице на промежутке

$$[-a, a] = [-\Omega, \Omega]$$

и равна нулю вне этого промежутка.

Такой прибор, как мы теперь понимаем, будет пропускать без искажений все гармоники частоты, не большей Ω ($|\omega| \leqslant \Omega = a$), и не будет реагировать на более высокие частоты.

Такой прибор называют *фильтром низких частот с верхней граничной частотой* Ω.

Найдём аппаратную функцию такого фильтра низких частот с верхней граничной частотой a (поднормированную осредняющим множителем):

$$E_a(t) = \frac{1}{2a} \int_{-a}^{a} e^{i\omega t}\, d\omega = \frac{\sin at}{at}.$$

Важность этой функции в электротехнике и теории передачи информации по каналу связи привела к тому, что появилось следующее специальное обозначение:

$$\operatorname{sinc} x := \frac{\sin x}{x}.$$

Некоторое объяснение того, почему вдруг на конце тут появилась буква c, будет дано ниже.

1.5. Теорема отсчётов (формула Котельникова—Шеннона)

Проделаем следующие выкладки сначала в наиболее простом варианте, не затемняющем деталями суть дела. Мы возьмём регулярную функцию f с финитным спектром $\widehat{f} = \frac{1}{2\pi}\varphi$, сосредоточенным на отрезке $[-\pi, \pi]$, разложим φ в ряд Фурье и подсчитаем коэффициенты:

$$f(t) = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \varphi(x)e^{itx}\,dx; \quad \varphi(x) = \sum_{-\infty}^{\infty} c_n e^{-inx};$$

$$c_n = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} \varphi(x)e^{inx}\,dx = f(n).$$

Подставляя в первый интеграл вместо функции φ её ряд Фурье и проводя почленное интегрирование (считая функцию φ «хорошей»), находим

$$f(t) = \sum_{-\infty}^{\infty} c_n \left(\frac{1}{2\pi}\int\limits_{-\pi}^{\pi} e^{ix(t-n)}\,dx\right) =$$

$$= \sum_{-\infty}^{\infty} f(n)\frac{\sin\pi(t-n)}{\pi(t-n)} = \frac{\sin\pi t}{\pi}\sum_{-\infty}^{\infty} f(n)\frac{(-1)^n}{t-n}.$$

Теперь вполне аналогично для общей функции с финитным спектром

$$f(t) = \frac{1}{2a}\int\limits_{-a}^{a} \varphi(x)e^{itx}\,dx,$$

раскладывая функцию φ в ряд Фурье на отрезке $[-a, a]$, находим

$$\varphi(x) = \sum_{-\infty}^{\infty} c_n e^{-i\frac{\pi}{a}nx}, \quad \text{где } c_n = \frac{1}{2a}\int\limits_{-a}^{a} \varphi(x)e^{i\frac{\pi}{a}nx}\,dx = f\Big(\frac{\pi}{a}n\Big),$$

и получаем следующее представление функции с финитным спектром, известное под названием *формулы Котельникова*, или *формулы Котельникова—Шеннона*, или *теоремы отсчётов*:

$$f(t) = \sum_{-\infty}^{\infty} f\Big(\frac{\pi}{a}n\Big)\frac{\sin a\big(t-\frac{\pi}{a}n\big)}{a\big(t-\frac{\pi}{a}n\big)}.$$

Эта формула была получена Котельниковым в 1933 г. в работе [1] и переоткрыта Шенноном в 1949 г. в работе [2]. Как интерполяционная формула (специальный случай формулы Лагранжа для целых функций конечной степени) она уже была известна математикам (см., например, [4] и цитируемую там литературу). Заслуга Котельникова и Шеннона состоит в интерпретации этой формулы с точки зрения кодирования сигнала и передачи информации по каналу связи.

Разумеется, формулу можно преобразовать и к виду

$$f(t) = \frac{\sin at}{a}\sum_{-\infty}^{\infty} f\Big(\frac{\pi}{a}n\Big)\frac{(-1)^n}{t-\frac{\pi}{a}n},$$

однако нам сейчас важнее исходная формула.

1.6. Кодово-импульсная модуляция сигнала (КИМ)

Формула Котельникова—Шеннона показывает, что регулярный сигнал с финитным спектром частот, лежащих в промежутке $[-a,a]$, может быть полностью восстановлен по набору его дискретных значений, отсчитываемых с интервалом $\Delta = \frac{\pi}{a}$. (Отсюда и название «теорема отсчётов»).

Более того, сигнал восстанавливается посредством комбинации, составленной из сдвигов аппаратной функции простейшего фильтра низких частот с верхней граничной частотой a, т. е. это ещё и относительно просто технически реализуемая формула дискретного кодирования и передачи сигнала. (В принципе, например, периодический сигнал можно задать коэффициентами его ряда Фурье, а аналитическую функцию — коэффициентами её тейлоровского разложения, но это не всегда удобно с точки зрения возможности аппаратной реализации.)

Описанная идея дискретного кодирования сигнала лежит в основе современных цифровых технологий записи, хранения, передачи и воспроизведения информации (музыка, видео, библиотеки, поисковые системы и т. д.).

Заметим теперь, что указанное выше специальное обозначение

$$\operatorname{sinc} x := \frac{\sin x}{x}$$

следует интерпретировать как «синус считающий»: c здесь от слова count.

1.7. Пропускная способность идеального канала связи

Теоретическая разработка понятия информации и её количественного описания была стимулирована появлением в начале XX в. телеграфа и радиосвязи. Ещё до появления формулы Котельникова специалисты нащупывали подходы к решению этих вопросов (см., например, сборник статей [3]). Так, в честь Найквиста (Nyquist) отсчётный интервал $\Delta = \frac{\pi}{a}$ не без оснований часто (по-видимому, вслед за Шенноном) называют *интервалом Найквиста*.

За время Δ передаётся одно отсчётное значение функции f, т. е. Δ^{-1} — это количество отсчётных значений, передаваемых в единицу времени (секунду).

Формула Котельникова—Шеннона, по-видимому, даёт нам, не знающим ещё, что мы называем информацией и как её измерять, явную связь между шириной используемой полосы частот и скоростью передачи информации по каналу связи. Но об этом мы ещё будем говорить ниже.

1.8. Оценка размерности TV-сигнала

Пусть, например, как и Шеннон, мы рассматриваем TV-сигнал, имеющий частоту $W = 5$ MHz (1 мегагерц $= 10^6$ Герц) и продолжительность $T = 1$ час. Подсчитаем, какой длины должен быть вектор отсчётных значений, отвечающий такому сигналу, т. е. найдём количество $N = T/\Delta = 2WT$ отсчётных значений:

$$N = 2 \cdot 5 \text{ МГц} \cdot 1 \text{ час} = 2 \cdot 5 \cdot 10^6 \cdot 60^2 = 3{,}6 \cdot 10^{10}.$$

Это вектор в пространстве $\mathbb{R}^N$ огромной размерности. Геометрия такого пространства имеет свою специфику. О ней в основном и пойдёт речь в главе II.

§2. Некоторые другие области многопараметрических явлений и пространств большой размерности

Для того чтобы понять, что цифровая запись сколь-нибудь сложного сигнала или сообщения требует большого количества знаков, конечно, не обязательно знать всё изложенное выше, включая формулу Котельникова—Шеннона. Мы остановились на этом подробнее потому, что здесь вполне содержательные вещи можно рассказать, не требуя от читателя почти никакой серьёзной предварительной подготовки. С другой стороны, эти сведения далеко не так часто известны широкому кругу математиков, как то, что мы намерены лишь упомянуть ниже в качестве иных и даже более фундаментальных областей, где описываемые величины и функции обычно тоже зависят от очень большого числа параметров, т. е. по существу тоже связаны с пространством большой размерности.

2.1. Молекулярная теория вещества

Термодинамические функции, например давление газа, с точки зрения статистической физики, а не феноменологической термодинамики, зависят от огромного количества переменных. (Вспомните, например, число Авогадро $N_A = 6{,}022 \cdot 10^{23}$ моль$^{-1}$.)

Удивительно (во всяком случае до сделанных ниже разъяснений, связанных с многомерной геометрией и принципом концентрации), что при этом (потом выяснится, что благодаря этому) значения этих функций могут оставаться стабильными.

Ниже мы коснёмся кинетической теории и даже фактически получим знаменитое распределение Максвелла.

2.2. Фазовое пространство в классической гамильтоновой механике

Фазовое пространство сколь-нибудь сложной механической системы, как правило, многомерно.

Многие функции, связанные с системой, оказываются тем самым функциями очень большого числа переменных.

2.3. Термодинамические ансамбли Гиббса

Объединив идеи термодинамики и гамильтоновой механики, Гиббс (Gibbs) внедрил в статистическую физику замечательную математическую структуру — гамильтонову систему, наделённую мерой, эволюционирующей (иногда к равновесию, что интересно) под действием гамильтонова потока в фазовом пространстве системы.

Статистическая физика (работами Больцмана) и задача Ньютона трёх и многих тел (работами Пуанкаре) инициировали появление и бурное развитие математической теории динамических систем с её многочисленными, разнообразными и порой всё ещё открытыми вопросами. (Например, вопросы об описании фазовых переходов, турбулентности, хаоса.)

2.4. Теория вероятностей

И, конечно, большие числа (испытаний), их осредняющее взаимодействие, характеристики редких заметных отклонений и т. п. составляют фундамент философии понятий и ядро целой области математики — теории вероятностей.

Ниже мы на миг коснёмся теории вероятностей, рассматривая в качестве примера применения излагаемых геометрических соображений получение вслед за Гауссом нормального распределения при обработке погрешностей наблюдений.

Глава II

Принцип концентрации и его проявления

§1. Шар и сфера в евклидовом пространстве $\mathbb{R}^n$ при $n \gg 1$

1.1. Концентрация объёма шара при $n \to \infty$

Рассмотрим шар $B^n(r)$ радиуса r в евклидовом пространстве $\mathbb{R}^n$ большой размерности n. Пусть $\operatorname{Vol} B^n(r)$ — его объём. Тогда

$$\frac{\operatorname{Vol} B^n(r+\Delta)}{\operatorname{Vol} B^n(r)} = \frac{(r+\Delta)^n}{r^n} = \left(1+\frac{\Delta}{r}\right)^n.$$

Это значит, что если размерность n велика, то стоит увеличить радиус шара всего на $\Delta = \frac{1}{n} r$, как его объём увеличится больше чем вдвое.

Например, если в пространстве $\mathbb{R}^{1000}$ у арбуза радиуса 1 м толщина корки 1 см, то после удаления корки от арбуза останется меньше тысячной доли его объёма.

Таким образом, подавляющая часть объёма многомерного шара сосредоточена в малой окрестности его границы.

1.2. Термодинамический предел

Как известно, Максвелл, занимаясь статистической физикой, точнее кинетической теорией, первым нашёл закон (*закон Максвелла*) распределения молекул данного объёма газа по скоростям и, как следствие, по кинетической энергии.

Пусть в данном объёме при данной температуре имеется n молекул газа массы m; v_i — скорость i-й молекулы, а E_n — суммарная кинетическая энергия молекул. С ростом n величина E_n растёт и имеет порядок n: $E_n \asymp n$.

В формальной записи сказанное означает, что

$$\frac{1}{2} m v_1^2 + \ldots + \frac{1}{2} m v_n^2 = E_n; \qquad \sum_{i=1}^{n} v_i^2 = \frac{2E_n}{m} \asymp n.$$

Ищутся статистические характеристики такого ансамбля частиц, когда $n \to \infty$ при условии, что $E_n \asymp n$. Предельный переход с таким условием называют *термодинамическим предельным переходом* или коротко, но менее точно, *термодинамическим пределом*. (Точные определения и современную математическую трактовку термодинамических пределов см., например, в книге [20].)

С математической точки зрения мы имеем здесь дело с $(3n-1)$-мерной сферой в $\mathbb{R}^{3n}$, радиус которой растёт как $n^{1/2}$ при $n \to \infty$.

В случае статистического исследования закона распределения независимых ошибок Δ_i измерения при выводе закона Гаусса считаем дисперсию D конечной и имеем абсолютно похожую ситуацию:

$$\frac{1}{n}(\Delta_1^2 + \ldots + \Delta_n^2) \asymp D; \quad \sum_{i=1}^{n} \Delta_i^2 \asymp n.$$

Если молекулы, как и ошибки наблюдения, считать равноправными, а соответствующие им точки сферы равномерно распределёнными по её поверхности, то и статистика молекул газа, и статистика ошибок измерения приводятся к статистике проекции $(n-1)$-мерной сферы на прямую в пространстве $\mathbb{R}^n$ при $n \to \infty$, когда радиус сферы растёт как $n^{1/2}$.

Мы уточним сказанное и проведём соответствующие вычисления ниже.

1.3. Концентрация площади сферы

В евклидовом пространстве $\mathbb{R}^n$ рассмотрим гиперсферу $S^{n-1}(r)$ радиуса r с центром в начале координат. Введём в $\mathbb{R}^n$ сферические координаты, отсчитывая угол ψ от положительного направления координатной оси, которую мы будем обозначать как ось x. Значит, $x = r\cos\psi$, $dx = -r\sin\psi\, d\psi$ и $\sin\psi = \dfrac{(r^2-x^2)^{1/2}}{r} = \left(1-\left(\dfrac{x}{r}\right)^2\right)^{1/2}$.

Площадь элементарного сферического слоя, отвечающего промежутку углов $(\psi, \psi + d\psi)$, даётся формулой

$$\begin{aligned}&\sigma_{n-2}(r\sin\psi) r\, d\psi = c_{n-2}(r\sin\psi)^{n-2} r\, d\psi = \\ &= c_{n-2} r^{n-2} \sin^{n-3}\psi\, r\sin\psi\, d\psi = c_{n-2} r^{n-2}(1-(x/r)^2)^{\frac{n-3}{2}}(-dx). \qquad (1)\end{aligned}$$

Здесь $\sigma_{n-2}(\varrho)$ — площадь $(n-2)$-мерной сферы радиуса ϱ, а $c_{n-2} = \sigma_{n-2}(1)$.

По формуле (1) находим теперь площадь сферического слоя, проектирующегося в отрезок $[a, b] \subset [-r, r]$ оси x:

$$c_{n-2} r^{n-2} \int_a^b (1-(x/r)^2)^{\frac{n-3}{2}}\, dx. \qquad (2)$$

Доля, которую эта площадь составляет от площади всей сферы $S^{n-1}(r)$ радиуса r, равна

$$P_n[a,b] := \frac{\int\limits_a^b (1-(x/r)^2)^{\frac{n-3}{2}}\,dx}{\int\limits_{-r}^{r} (1-(x/r)^2)^{\frac{n-3}{2}}\,dx}. \tag{3}$$

Переходя здесь к термодинамическому пределу, когда $n \to \infty$, а $r = \sigma n^{1/2}$, получим нормальное распределение

$$P_n[a,b] := \frac{\int\limits_a^b e^{-\frac{x^2}{2\sigma^2}}\,dx}{\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,dx}. \tag{4}$$

Как уже было отмечено выше, оно одновременно появляется и как распределение Гаусса в статистической теории обработки результатов наблюдений (теории ошибок), и как распределение Максвелла в статистической физике (в кинетической теории). Здесь же содержится и центральная предельная теорема теории вероятностей.

Если же фиксировать r и устремить n к бесконечности, то величина $P[a,b]$ в формуле (3) будет экспоненциально быстро стремиться к нулю при $0 < a < b \leqslant r$.

Проведём подтверждающие это выкладки.

Напомним сначала классические результаты относительно асимптотики интеграла Лапласа

$$F(\lambda) := \int\limits_{I=[a,b]} f(x)e^{\lambda S(x)}\,dx.$$

Пусть обе функции f и S определены и регулярны на промежутке интегрирования I, причём функция S вещественна и имеет единственный абсолютный максимум на I, который достигается в точке $x_0 \in I$, и $f(x_0) \neq 0$.

Тогда при $\lambda \to +\infty$ асимптотика интеграла Лапласа такая же, как если бы интеграл был распространён только на сколь угодно малую окрестность точки x_0 в I (это так называемый *принцип локализации*).

После такой локализации вопрос сводится к рассмотрению специального случая, когда $I = [x_0, x_0 + \varepsilon]$ или/и $I = [x_0 - \varepsilon, x_0]$, а ε —

сколь угодно малое положительное число. Используя теперь тейлоровские разложения, находим, что при $\lambda \to +\infty$

$$F(\lambda) = \frac{f(x_0)}{-S'(x_0)} e^{\lambda S(x_0)} \lambda^{-1}\bigl(1+O(\lambda^{-1})\bigr), \tag{A}$$

если $x_0 = a$ и $S'(x_0) \neq 0$ (значит, $S'(x_0) < 0$, поскольку $a < b$);

$$F(\lambda) = \sqrt{\frac{\pi}{-2S''(x_0)}}\, f(x_0) e^{\lambda S(x_0)} \lambda^{-1/2}\bigl(1+O(\lambda^{-1/2})\bigr), \tag{B}$$

если $x_0 = a$, $S'(x_0) = 0$, $S''(x_0) \neq 0$ (значит, $S''(x_0) < 0$);

$$F(\lambda) = \sqrt{\frac{2\pi}{-S''(x_0)}}\, f(x_0) e^{\lambda S(x_0)} \lambda^{-1/2}\bigl(1+O(\lambda^{-1/2})\bigr), \tag{C}$$

если $a < x_0 < b$, $S'(x_0) = 0$, $S''(x_0) \neq 0$ (т. е. $S''(x_0) < 0$).

Следовательно, в соответствии с формулой (C) при $n \to \infty$ имеем

$$\int_{-r}^{r} (1-(x/r)^2)^{\frac{n-3}{2}}\, dx = \int_{-r}^{r} e^{\frac{n-3}{2}\log(1-(x/r)^2)}\, dx \sim r\sqrt{\frac{2\pi}{n}}, \tag{5}$$

а в соответствии с формулой (A) при $n \to \infty$ и $\delta > 0$ находим

$$\int_{\delta r}^{r} (1-(x/r)^2)^{\frac{n-3}{2}}\, dx = \int_{\delta r}^{r} e^{\frac{n-3}{2}\log(1-(x/r)^2)}\, dx \sim r\frac{1}{n\delta}(1-\delta^2)^{\frac{n-1}{2}}. \tag{6}$$

Поэтому, каким бы малым ни было $\delta > 0$, при $n \to \infty$ мы имеем

$$P_n[\delta r, r] \sim \frac{1}{\delta\sqrt{2\pi n}}(1-\delta^2)^{\frac{n-1}{2}} \sim \frac{1}{\delta\sqrt{2\pi n}} e^{-\frac{1}{2}\delta^2 n} \to 0. \tag{7}$$

Это означает, что подавляющая часть площади многомерной сферы S^{n-1} сосредоточена в малой окрестности экватора.

Указанное обстоятельство позволяет объяснить тот, казалось бы, парадоксальный факт, что выбранные наугад два единичных вектора в пространстве $\mathbb{R}^n$ большой размерности n с очень большой вероятностью окажутся почти ортогональными (например, вероятность того, что их скалярное произведение сколь-нибудь заметно отклонится от нуля, быстро убывает с ростом величины отклонения). Уточним сказанное. Все направления в пространстве $\mathbb{R}^n$ считаются равноправными. Векторы пары случайны и независимы. Если один из них выбран, то второй с большой вероятностью окажется в окрестности экватора, сопряжённого первому вектору, о чём говорит формула (7). Она же позволяет дать оценку вероятности отклонения от ортогональности.

1.4. Изопериметрическое неравенство и почти постоянство функции на сфере очень большой размерности

Отметим и объясним ещё один не менее парадоксальный и более тонкий факт, связанный с многомерностью.

Пусть S^m — единичная сфера в евклидовом пространстве $\mathbb{R}^{m+1}$ очень большой размерности $m+1$. Пусть на сфере задана достаточно регулярная (например, из некоторого фиксированного класса Липшица) вещественнозначная функция. Берём случайно и независимо друг от друга пару точек на сфере и вычисляем в них значения нашей функции. С большой вероятностью эти значения будут почти одинаковы и близки к некоторому числу M_f.

[Это пока ещё гипотетическое число M_f называют *медианным значением функции* или *медианой функции*. Его также называют *средним значением функции в смысле Леви*. Мотивировка терминов вскоре прояснится вместе с точным определением числа M_f.]

Объясним это явление.

Введём некоторые обозначения и соглашения. Договоримся расстояние между точками сферы S^m понимать в смысле её геодезической метрики ϱ. Через A_δ обозначим δ-окрестность в S^m множества $A \subset S^m$. Нормируем стандартную меру сферы, заменив её равномерно распределённой вероятностной мерой μ, т. е. $\mu(S^m) = 1$.

Справедливы следующие утверждения (доказательства см. в книге [3]).

Для любых $0 < a < 1$ и $\delta > 0$ существует

$$\min\{\mu(A_\delta) : A \subset S^m,\ \mu(A) = a\},$$

и он достигается на сферической шапочке A^o меры a.

Здесь $A^o = B(r)$, где $B(r) = B(x_o, r) = \{x \in S^m : \varrho(x_o, x) < r\}$ и $\mu(B(r)) = a$.

При $a = \dfrac{1}{2}$, т. е. когда A^o — полусфера, получаем такое следствие.

Если подмножество $A \subset S^{n+1}$ таково, что $\mu(A) \geqslant 1/2$, то $\mu(A_\delta) \geqslant 1 - \sqrt{\pi/8}\, e^{-\delta^2 n/2}$. (При $n \to \infty$ здесь $\sqrt{\pi/8}$ можно заменить на $1/2$.)

Обозначим через M_f такое число, для которого

$$\mu\{x \in S^m : f(x) \leqslant M_f\} \geqslant \frac{1}{2} \quad \text{и} \quad \mu\{x \in S^m : f(x) \geqslant M_f\} \geqslant \frac{1}{2}.$$

Его-то и называют *медианой* или *средним в смысле Леви значением функции* $f : S^m \to \mathbb{R}$. (Если M_f-уровень функции f на сфере имеет нулевую меру, то мера каждого из указанных двух множества будет в точности равна половине μ-площади сферы.)

Лемма Леви [2], непосредственно вытекающая из приведённого выше утверждения и его следствия, гласит следующее.

Если $f \in C(S^{n+1})$ и $A = \{x \in S^{n+1} : f(x) = M_f\}$, то

$$\mu(A_\delta) \geqslant 1 - \sqrt{\pi/2}e^{-\delta^2 n/2}.$$

Пусть теперь $\omega_f(\delta) = \sup\{|f(x) - f(y)| : \varrho(x, y) \leqslant \delta\}$ — *модуль непрерывности функции f.*

Значения функции f на множестве A_δ близки к M_f. Точнее, если $\omega_f(\delta) \leqslant \varepsilon$, то $|f(x) - M_f| \leqslant \varepsilon$ на A_δ. Таким образом, лемма Леви показывает, что «хорошие» функции на самом-то деле почти постоянны на почти всей области определения S^m, когда её размерность m очень велика.

Интересно, однако, посмотреть на это явление концентрации, например, с точки зрения эргодических теорем при термодинамическом предельном переходе, а также с точки зрения возможности геометрического толкования закона больших чисел (уже и нелинейного) и предельных распределений теории вероятностей. Весьма красивы и возможные физические приложения (см. например, статью [24]).

1.5. Концентрация меры, проявления и формализация

Феномен *концентрации меры* имеет многочисленные проявления (в многомерной геометрии, теории вероятностей, статистической физике, теории передачи информации по каналу связи при наличии помех, теории больших информационных систем...). Даже в геометрии он проявляется, конечно, не только на рассмотренной выше многомерной сфере (см., например, [3] и [11]). Разумеется, однако, концентрация меры имеет место не для всех пространств, функций и мер.

Явление стабилизации значений функций очень многих переменных, по-видимому, уже было знакомо таким физикам, как Максвелл и Больцман. Совершенно отчётливо его отмечает Лоренц [15]. В теории вероятностей вся идеология законов больших чисел в геометрическом аспекте есть то же проявление феномена концентрации меры. Сам термин *принцип концентрации меры*, если я не ошибаюсь, в математику ввёл Виталий Мильман, который дал основанное на нём новое доказательство известной теоремы Дворецкого о почти евклидовых подпространствах в линейных нормированных пространствах большой размерности. См. в этой связи книгу [3], которая также содержит написанное М. Громовым геометрическое приложение, включающее изопериметрическое неравенство на римановых многообразиях. Там же имеется библиография с указанием первоисточников, например работы П. Леви [2], где в явной форме сформулирован принцип концентрации. В каком-то виде это явление отмечал уже и Пуанкаре в своих лекциях по теории вероятностей [1]. Сейчас тематика, связанная с принципом концентрации меры, стала предметом многочисленных исследований. Источники [4]—[14] — это очень неполная, но относительно свежая библиография, которую мы приводим для сведения заинтересованного читателя.

Современные математические исследования, относящиеся к явлению концентрации меры, обычно начинаются с введения следующих исходных объектов: (X, d, μ) — пространство (множество) X с метрикой d и вероятностной мерой μ. Предмет исследования — взаимодействие метрики и меры. Этому служит следующая *функция концентрации меры*:

$$\alpha_\mu(r) := \alpha_{(X,d,\mu)}(r) := \sup\Big\{1-\mu(A_r)\colon \mu(A) = \frac{1}{2}\Big\},$$

где супремум берётся по измеримым множествам $A \subset X$, а

$$A_r := \{x \in X\colon\ d(x, A) < r\}.$$

Проведённые выше рассмотрения, относившиеся к многомерной сфере, теперь позволяют читателю понять смысл и конкретное содержание этих, казалось бы, абстрактных понятий.

Связь концентрации меры на пространстве (X, d, μ) с концентрацией значений 1-липшицевой функции f на метрическом пространстве (X, d) около её медианного значения M_f совсем прозрачна: если $A := \{x \in X\colon\ f \leqslant M_f\}$, то

$$A_r \subset \{x \in X\colon f \leqslant M_f + r\} \quad \text{и} \quad \mu(A_r) \leqslant \mu\{x \in X\colon f \leqslant M_f + r\}.$$

Повторив это рассуждение с обратными знаками неравенства, приходим к следующей заключительной оценке:

$$\mu(\{x \in X\colon |f - M_f| \geqslant r\}) \leqslant 2\alpha_\mu(r).$$

Можно также показать, что функция концентрации меры позволяет дать следующую оценку возможного отклонения медианного значения M_f 1-липшицевой функции f от значения E_f её математического ожидания: $|M_f - E_f| \leqslant 2 \int\limits_0^\infty \alpha_\mu(t)\, dt$.

§ 2. Некоторые замечания

2.1. Различные средние

Каково взаимоотношение стандартного среднего значения $\bar{f}$ функции и введённого выше её среднего значения M_f в смысле Леви?

Переходя, если надо, к функции $f - M_f$, без ограничения общности можно считать, что для функции f медианное значение M_f равно нулю. Обозначим через $|S^n|$ площадь (стандартную евклидову n-меру) сферы S^n, и пусть T — верхняя грань значений функции $|f|$ на этой сфере (т. е. максимальное значение $|f|$ в случае непрерывной функции).

Пусть ε — малое положительное число. Выкладки

$$|\bar{f}| \leqslant \frac{1}{|S^n|}\left(\int\limits_{|f(x)| \leqslant \varepsilon T} |f|(x)\, dx + \int\limits_{|f(x)| > \varepsilon T} |f|(x)\, dx \right) \leqslant$$
$$\leqslant \varepsilon T + T \frac{1}{|S^n|} \int\limits_{|f(x)| > \varepsilon T} dx = \left(\varepsilon + \frac{1}{|S^n|} \int\limits_{|f(x)| > \varepsilon T} dx \right) T$$

показывают, что значение $|\bar{f}|$ мало в сравнении с T, если последний интеграл в круглых скобках мал в сравнении с $|S^n|$. Иными словами, это так, если площадь области $D_{\varepsilon T} \subset S^n$, где $|f(x)| > \varepsilon T$, мала в сравнении с площадью всей сферы. По условию $M_f = 0$, поэтому область $D_{\varepsilon T}$ лежит вне некоторой δ-окрестности медианного уровня функции, около которого, как было показано, сосредоточена подавляющая доля всей поверхности сферы, если, конечно, сама эта окрестность не слишком маленькая. Величина δ, характеризующая относительную ширину этой окрестности, как раз и зависит от модуля непрерывности функции f, связывающего величины ε и δ.

Если рассматривать функции с фиксированной константой Липшица L на сфере радиуса r, то для них $T \asymp Lr$, $\delta \asymp L^{-1}\varepsilon$ и при $n \gg 1$ с малой относительной погрешностью, конечно, имеет место равенство $\bar{f} = M_f$.

Однако с ростом n следует дать расти и L. Например, для термодинамики типичны так называемые *сумматорные функции* вида $f(x_1)+\ldots+f(x_n)$ (сумма энергий частиц и т. п.). Уже для простейшей линейной функции $x_1+\ldots+x_n$ имеем $L=\sqrt{n}$ (для доказательства достаточно сместиться из начала координат в точку $(1, \ldots, 1)\in\mathbb{R}^n$).

В интересных для термодинамики случаях, как мы уже отмечали, естественно считать, что радиус сферы $S^n(r)$ растёт вместе с ростом n, причём $r\asymp\sqrt{n}$. Естественно поэтому считать также, что размах значений сумматорной функции на такой сфере будет порядка $L\cdot r=\sqrt{n}\cdot\sqrt{n}=n$. В этом случае тоже можно установить справедливость описанного явления концентрации, этой формы закона больших чисел.

Отметим в заключение, что в $\mathbb{R}^n$ стандартная единица объёма, n-мерный «куб», с ростом размерности n становится всё более крупной: диаметр n-мерного «куба» равен $\sqrt{n}$.

Вписанный в такой «куб» шар занимает ничтожную долю объёма куба, если $n\gg 1$.

Полезно также заметить (это используется в теории кодирования, см. главу 3), что если взять два шара одинакового радиуса и расположить их так, что расстояние между центрами равно радиусу, то шары будут пересекаться, однако если $n\gg 1$, то область их пересечения будет иметь объём, ничтожно малый по сравнению с объёмом каждого из шаров.

2.2. Многомерный куб и принцип концентрации

Рассмотрим стандартный единичный n-мерный замкнутый промежуток $I^n\subset\mathbb{R}^n$, который для краткости позволим себе называть n-мерным кубом. Почистим куб I^n, как апельсин, удалив из него некоторую $\frac{\delta}{2}$-окрестность границы куба. Останется снова куб, но уже с ребром длины $r=1-\delta$. Его объём r^n мал, если $n\gg 1$. Мы удалили объём $1-r^n=1-(1-\delta)^n$, который составляет основную часть всего объёма 1 единичного куба.

Если бы мы имели n независимых случайных величин x_i, принимающих значения на единичном отрезке $[0, 1]$ и имеющих распределения вероятностей $p_i(x)$, которые равномерно по i отделены от нуля, например если все $p_i(x)$ одинаковы, то с ростом n подавляющая часть случайных точек $(x_1, \ldots, x_n)\in I^n$ окажется в непосредственной близости границы куба.

Такой принцип концентрации в соответствующей формулировке, конечно, справедлив и для областей общего типа в пространстве $\mathbb{R}^n$ при $n \gg 1$.

Добавим ещё несколько слов о многомерном кубе.

В евклидовом пространстве $\mathbb{R}^n$ рассмотрим n-мерный единичный куб

$$I^n := \left\{x = (x^1, \dots, x^n) \in \mathbb{R}^n : |x^i| \leqslant \frac{1}{2},\ i = 1, 2, \dots, n\right\}.$$

Его объём равен 1, хотя диаметр равен $\sqrt{n}$. (Напомним, кстати, что объём n-мерного шара радиуса r в евклидовом пространстве $\mathbb{R}^n$ выражается формулой

$$V_n(r) = \frac{\pi^{\frac{n}{2}}}{\frac{n}{2}\Gamma\left(\frac{n}{2}\right)} r^n,$$

из которой следует, что главный член асимптотики величины r_n радиуса n-мерного шара единичного объёма при $n \to +\infty$ имеет вид $\sqrt{\frac{n}{2\pi e}}$, т. е. это величина порядка $\sqrt{n}$.) Будем считать стандартную меру вероятностной равномерно распределённой мерой на кубе I^n.

Пусть $a = (a^1, \dots, a^n)$ — единичный вектор, а $x = (x^1, \dots, x^n)$ — произвольная точка куба I^n.

Справедливо следующее неравенство (оценка вероятности типа неравенства Бернштейна):

$$P_n\left\{\left|\sum_{i=1}^{n} a^i x^i\right| \geqslant t\right\} \leqslant 2\exp(-6t^2).$$

Интерпретируя сумму $\sum\limits_{i=1}^{n} a^i x^i$ как скалярное произведение $\langle a, x\rangle$, понимаем, что оно может быть большим (порядка $\sqrt{n}$), если вектор a направить не вдоль какого-то одного ребра куба, а вдоль главной диагонали, смешивая все координатные направления в равной степени. Взяв $a = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$, на основании указанной оценки заключаем, что объём n-мерного куба I^n с ростом n концентрируется в относительно малой окрестности гиперплоскости, проходящей через начало координат ортогонально вектору $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$.

В отличие от случая шара или сферы, на кубе регулярная функция может терять свойство концентрации значений, понимаемое в прежнем смысле, когда вероятность отклонения значения функции от медианного значения убывала экспоненциально: взять хотя бы функцию, значение которой в точке совпадает со значением первой координаты этой точки. Проекция шара на любую прямую устроена одинаково, а в случае куба это не так. Равноправность вхождения координат точки куба в функцию, определённую на кубе, достигается как раз тогда, когда функция фактически зависит только от проекции точки на главную диагональ (одну из равноправных главных диагоналей) куба. Именно такие функции обычно интересны при рассмотрении больших систем равноправных переменных (частиц в термодинамике или случайных величин в теории вероятностей). Для таких функций на кубе

$$I^n := \Big\{x = (x^1, \ldots, x^n) \in \mathbb{R}^n \colon |x^i| \leqslant \frac{1}{2},\ i = 1, 2, \ldots, n\Big\},$$

как мы видим, тоже будет действовать принцип концентрации значений. Типичным окажется значение в центре куба.

При исследовании концентрации значений липшицевых функций на кубе полезно иметь в виду следующее обстоятельство: какие бы два множества в кубе I^n ни взять, если их объёмы больше некоторой величины ε, то расстояние между ними не превосходит величины $d(\varepsilon)$, зависящей только от ε, но не от размерности куба. Нетривиальность утверждения в том, что при этом диаметр куба $\sqrt{n}$ может быть сколь угодно большим. То же, конечно, относится и к шару единичного объёма.

Мы видели на примере сферы, что типичные значения функции лежат вблизи значения, которое было названо медианным. Большую роль тут играла геометрия сферы и равномерность распределения меры. Для областей, полученных малым гладким возмущением сферы, принцип остаётся, но для больших деформаций сферы или меры, конечно, в прежней форме он не имеет места.

Есть ли его разумное обобщение, в котором рост размерности в конце концов всё же оказывается определяющим фактором?

(Например, любое достаточно многомерное выпуклое тело допускает почти сферические сечения растущей размерности, и это «почти» тем лучше, чем больше исходная размерность.)

2.3. Принцип концентрации, термодинамика, эргодичность

Все основные термодинамические функции (например, давление), которые с точки зрения статистической физики являются типичными значениями функций, зависящих от огромного числа переменных (фазовых координат отдельных молекул), повседневно демонстрируют нам законы больших чисел и принципы концентрации.

Более того, если эволюцию состояния термодинамической системы мы рассматриваем как движение точки по поверхности уровня энергии или в очень многомерной области, ограниченной такой поверхностью, то в силу принципа концентрации большую часть времени точка будет находиться в области медианных значений любой достаточно регулярной функции, заданной на этой поверхности или в ограниченной ею области пространства. При этом практически всё равно, считать ли значение средним по поверхности или по ограниченной ею области пространства.

Пусть движение происходит в области многомерного фазового пространства $\mathbb{R}^n$, определяемой условием $H \leqslant E$ (где H — гамильтониан системы). Область уже не шар. Слой между уровнями H и $H + \Delta$ уже имеет неравномерную толщину: $\Delta/|\nabla H|$. При сопоставлении средних по области и средних по поверхности это следует учесть. Если при интегрировании по области в качестве одной из переменных интегрирования фигурирует криволинейная координата H, то на поверхности равномерная стандартная мера $d\sigma$ заменится на $d\sigma/|\nabla H|$.

В силу теоремы Лиувилля это инвариантная мера соответствующей системы Гамильтона (микроканоническая по Гиббсу).

Однако функцию $|\nabla H|$ тоже можно бы считать почти постоянной, если уровень $H = E$ похож на сферу.

Эргодическая гипотеза статистической механики, как известно, идёт от Больцмана и связана с двумя способами наблюдения (измерения) величин: брать среднее значение величины по всей совокупности однотипных объектов или следить за случайно выбранным типичным объектом и брать среднее по времени от значения этой величины вдоль траектории этого объекта.

Довольно скоро было понято, что для выводов теории столь сильное предположение, как эргодичность динамической системы, совсем не обязательно. Важна только концентрация меры, связанная с тем, что типичные системы статистической термодинамики содержат очень много частиц (очень многомерны). Лоренц это отмечает совершенно явно: «Максимумы же вероятности, входящие в теории, которыми мы занимаемся, чрезвычайно остры». (См. лекции Лоренца «Статистические теории в термодинамике», указанные в списке литературы.) Это и значит, что плотности распределения вероятности состояний системы и значений наблюдаемых величин очень концентрированы, поэтому мы получаем стабильные значения измеряемых функций состояния системы, хотя в отличие от феноменологической термодинамики, статистическая механика не исключает больших уклонений, указывая одновременно малость их вероятности.

2.4. Принцип концентрации и предельные распределения

Принцип концентрации — это геометрический аналог, возможно и нелинейный, закона больших чисел. Последний в классической теории вероятностей обычно относится к линейной комбинации случайных величин, например к их сумме.

Предельные теоремы теории вероятностей устанавливают закон больших чисел, а также предельные законы распределения вероятностей.

Выше, рассматривая принцип концентрации на примере площади многомерной сферы $S^n(r) \subset \mathbb{R}^{n+1}$, мы заодно получили и предельный закон распределения этой площади, когда размерность n сферы и её радиус $r = \sqrt{n}$ неограниченно растут. Получаемый здесь результат соответствует центральной предельной теореме теории вероятностей.

Для возмущённой сферы, как мы только что отметили, по-видимому, тоже действует принцип концентрации, поэтому должны быть и предельные распределения, которые, в частности, должны соответствовать различным известным вариантам центральной предельной теоремы.

Это заодно может дать некоторую адекватную формулировку и подтверждение общего суждения о том, что если на большую систему накладывается какое-то глобальное ограничение (например, ограничение полной энергии), то это в определённом смысле предопределит её микроскопическую вероятностную структуру.

Прежде чем перейти к следующей главе, отметим, что различные аспекты обсуждавшихся здесь вопросов явно или неявно присутствуют во многих исследованиях, быть может, формально относящихся к разным областям математики или её приложениям. См. в этой связи, например, книги и статьи [1—3, 15—24].

Глава III

Связь при наличии помех

§1. Дискретная запись непрерывного сигнала — конкретизация

1.1. Энергия и средняя мощность сигнала

Напомним, что теорема отсчётов (формула Котельникова)

$$f(t)=\sum_{-\infty}^{\infty} f\Big(\frac{k}{2W}\Big)\frac{\sin 2\pi W\Big(t-\dfrac{k}{2W}\Big)}{2\pi W\Big(t-\dfrac{k}{2W}\Big)} \tag{1}$$

восстанавливает сигнал — функцию $f\in L_2(\mathbb{R})$ с финитным спектром частот ν, не превышающих W Герц, — по совокупности отсчётных значений $f(t_k)$ в точках $t_k=k\Delta$, где $\Delta=\dfrac{1}{2W}$ — отсчётный интервал времени (интервал Найквиста), зависящий от W.

Чем шире полоса частот, тем сложнее может быть функция f, тем чаще надо брать отсчёты для её адекватного дискретного кодирования и восстановления, но зато тем больше информации она (т. е. такой сигнал) может нести.

Базовой в разложении (1) является функция $\operatorname{sinc} t=\dfrac{\sin t}{t}$, которая, как мы уже знаем, имеет постоянный спектр, равный 1, на единичном промежутке частот и является аппаратной функцией идеального фильтра низких частот с единичной полосой пропускания. Таким образом, функция отсчётов sinc реализуется как ответ такого фильтра на единичный импульс, осуществлённый в момент $t=0$.

Соответственно функция

$$e_k(t)=\operatorname{sinc} 2\pi W\Big(t-\frac{k}{2W}\Big)=\frac{\sin 2\pi W\Big(t-\dfrac{k}{2W}\Big)}{2\pi W\Big(t-\dfrac{k}{2W}\Big)}$$

имеет спектр $\check{e}_k(\nu)=\dfrac{1}{2\pi W}\exp\Big(-i\dfrac{\pi}{W}k\nu\Big)$ в полосе частот $0\leqslant\nu\leqslant W$ ($|\nu|\leqslant W$).

Из ортогональности функций $\check{e}_k$ на промежутке $[-W,W]$ (или любом промежутке длины $2W$) и равенства Парсеваля для преобразования Фурье можно заключить, что сами функции e_k ортогональны в пространстве $L_2(\mathbb{R})$, причём $\|e_k\|^2 = \frac{1}{2W}$.

Значит из равенства $f = \sum\limits_{-\infty}^{\infty} x_k e_k$ можно заключить, что

$$\|f\|^2 = \frac{1}{2W} \sum_{-\infty}^{\infty} x_k^2.$$

На практике сигнал f имеет некоторую конечную продолжительность T, т. е. $f(t) \equiv 0$ вне отрезка $0 \leqslant t \leqslant T$. Это условие несовместимо с условием финитности спектра функции f. Однако можно считать, что значения $f(t)$ функции малы вне отрезка $[0, T]$ и отсчётные значения функции вне этого отрезка полагаются равными нулю.

Тогда равенство $f = \sum\limits_{-\infty}^{\infty} x_k e_k$ заменится на $f(t) = \sum\limits_{k=1}^{2WT} x_k e_k(t)$, где $t \in [0, T]$, $x_k = f(k\Delta)$, $\Delta = \frac{1}{2W}$.

Такой сигнал f записывается вектором $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ его отсчётных значений, и при этом $n = 2WT$.

При тех же условиях равенство Парсеваля

$$\int_{-\infty}^{\infty} f^2(t)dt = \sum_{-\infty}^{\infty} x_k^2 \|e_k\|^2 = \frac{1}{2W} \sum_{-\infty}^{\infty} x_k^2$$

заменится на равенство

$$\int_0^T f^2(t)dt = \sum_1^n x_k^2 \|e_k\|^2 = \frac{1}{2W} \sum_1^n x_k^2 = \frac{1}{2W} \|x\|^2.$$

Интеграл здесь, с точностью до конкретного размерного множителя, даёт энергию (работу) сигнала f (например, когда f реализуется перепадом напряжения на единичном сопротивлении). Значит, средняя мощность P сигнала f на временно́м промежутке $[0, T]$ есть

$$P = \frac{1}{T} \int_0^T f^2(t)dt = \frac{1}{2WT} \|x\|^2 = \frac{1}{n} \|x\|^2.$$

Таким образом, $\|x\|^2 = nP = 2WTP$, и можно интерпретировать P как среднюю мощность, которая приходится на одну координату вектора x, т. е. на одно отсчётное значение сигнала f.

Сигналы f продолжительности T с финитным спектром в полосе частот W, средняя мощность которых не больше P, в векторном представлении $x=(x_1, \ldots, x_n)$ оказываются, таким образом, расположенными внутри шара $B(o, r)=B(r)\subset\mathbb{R}^n$ радиуса $r=\sqrt{2WTP}=$ $=\sqrt{nP}$ с центром в начале координат евклидова пространства $\mathbb{R}^n$ размерности $n=2WT$.

1.2. Квантование по уровням

Измерение отсчётного значения сигнала f производится с некоторой пороговой (предельной) точностью ε. Если амплитуда любого передаваемого сигнала не больше A (т. е. $|f|(t)\leqslant A$ при $t\in[0, T]$), то, разделив отрезок $[-A, A]$ равномерной сеткой точек (уровней) с шагом ε, мы в качестве $f(t)$ будем брать ближайшую к значению $f(t)$ точку этой сетки. Значения $f(t)$ оказываются проквантованными по уровням, количество которых $\alpha=\dfrac{2A}{\varepsilon}$ (считаем α целым и бóльшим 1). Отвечающее сигналу f слово $x=(x_1, \ldots, x_n)$, состоящее из n букв x_k, будет записано в алфавите, имеющем α различных букв. Всего имеется α^n таких различных слов x. Если $n=2WT$, а W и T — большие числа, то число α^n огромно.

1.3. Идеальный многоуровневый канал связи

При этих условиях за время T можно различить $M=\alpha^{2WT}$ (и не более) различных сигналов $f\sim x=(x_1, \ldots, x_n)$, т. е. указать один определённый сигнал/слово/сообщение из M возможных.

Двоичная запись $x^0=(x_1^0, \ldots, x_m^0)$, различающая M объектов, требует $m=\log_2 M$ символов $0, 1$ (считаем m целым). Информацию о значении очередной координаты двоичного вектора (если координаты равноправны, а их возможные значения $0, 1$ равновероятны) принимают за элементарную единицу информации и называют *бит*. Если мы могли бы без каких-либо ошибок получать и передавать векторы (слова), кодирующие наши M сообщений, то за время T мы могли бы различить M объектов (сигналов, сообщений). Скорость передачи информации (о выборе одного из M возможных объектов) по такому идеальному каналу связи (и при такой кодировке), измеренная в битах в секунду, была бы равна $\dfrac{1}{T}\log_2 M=2W\log_2\alpha$.

1.4. Помехи (белый шум)

Будем теперь работать с векторами $x=(x_1, \ldots, x_n)\in\mathbb{R}^n$. Здесь $n=2WT\gg 1$, и мы знаем, что $\|x\|^2=2WTP=nP$, где P — средняя мощность на одну координату вектора x, т. е. средняя мощность соответствующего x сигнала f.

Допустим, как это на самом деле и бывает, что в канале связи имеются помехи. Они породят вектор помех $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$, и на приёмном конце канала связи вместо вектора x будет получен смещённый вектор $x+\xi$. Таким образом, около каждой точки $x \in \mathbb{R}^n$ возникает область неопределённости $U(x)$, в точки которой помехи могут сместить x.

Помехи могут быть разной природы и в соответствии с этим могут иметь разные характеристики. Будем считать, что у нас помеха случайна, не зависит от x и является белым (тепловым) шумом, т. е. вектор $\xi \in \mathbb{R}^n$ случаен, а его координаты — независимые случайные величины, одинаково распределённые по нормальному гауссовскому закону (с нулевым математическим ожиданием и дисперсией σ^2). Пусть N — средняя (на отсчётное значение) мощность помехи. Тогда $\|\xi\|^2 = nN = 2WTN$ (здесь обозначение N от слова noise, в то время как P произошло от power) и $\sqrt{N} = \sigma$ — стандартное отклонение от нуля случайного значения каждой из координат вектора ξ. Как и прежде, мы считаем, что $2WT = n \gg 1$.

§2. Пропускная способность канала связи с шумом

2.1. Грубая оценка пропускной способности канала с шумом

Совместная средняя мощность сигнала и шума не больше $P+N$, поэтому координаты вектора $x+\xi$ по модулю в среднем не должны превосходить $\sqrt{P+N}$ и он должен лежать внутри шара радиуса $\sqrt{n(P+N)}$.

Поскольку ожидаемое смещение каждой из координат вектора (сообщения) x под действием помехи (белого шума) имеет порядок $\sigma = \sqrt{N}$, число хорошо различимых на приёмном конце значений каждой из координат пропорционально $\sqrt{\dfrac{P+N}{N}}$. Коэффициент пропорциональности k зависит от того, как истолковывается термин «хорошо различимый». Если требуется улучшить разрешение, то k надо уменьшить.

За время T имеется $n = 2WT$ независимых значений (отсчётов) координат, поэтому полное число K различимых сигналов будет

$\left(k\sqrt{\frac{P+N}{N}}\right)^{2WT}$. Число $\log_2 K$ двоичных единиц, которые можно передать за время T, будет, таким образом, составлять $WT\log_2 k^2\frac{P+N}{N}$. Значит, скорость передачи будет $W\log_2 k^2\frac{P+N}{N}$ бит в секунду.

2.2. Геометрия сигнала и помехи

Теперь напомним следующее. На вектор $x\in\mathbb{R}^n$ накладывается помеха $\xi\in\mathbb{R}^n$ в виде белого шума. Это означает, что дан вектор $x\in\mathbb{R}^n$ и не зависящий от x случайный, имеющий равномерное распределение по направлениям в пространстве $\mathbb{R}^n$ вектор $\xi\in\mathbb{R}^n$. Размерность n пространства $\mathbb{R}^n$ огромна. Тогда из принципа концентрации, который мы обсуждали в главе II, следует, что с ничтожно малой вероятностью ошибки вектор ξ будет почти ортогонален вектору x (т. е. скалярное произведение и корреляцию векторов x и ξ следует считать равными нулю).

Добавим к этому, что в силу концентрации основной части объёма многомерного шара в малой окрестности его граничной сферы можно считать, что если случайная точка лежит в таком шаре, то, скорее всего, она расположена почти на его границе.

Итак, в нашей ситуации, когда $n=2WT\gg 1$, имеем основание считать, что

$$\|x\|^2=nP,\quad \|\xi\|^2=nN,\quad \|x+\xi\|^2=n(P+N).$$

Области неопределённости $U(x)$, возникающие на приёмном конце около каждой точки $x\in\mathbb{R}^n$ в результате действия помехи, в нашем случае следует считать шарами $B(x,r)$ радиуса $r=\sqrt{n\sigma}=\sqrt{nN}$.

Сколько при таких условиях имеется различимых сигналов x, лежащих в шаре $B(o,\sqrt{nP})$? Понятно, что не более чем отношение объёма шара $B(o,\sqrt{n(P+N)})$ к объёму шара радиуса $\sqrt{nN}$. Таким образом, для числа M различимых сигналов имеем оценку сверху

$$M\leqslant\left(\sqrt{\frac{P+N}{N}}\right)^{2WT}=\left(\frac{P+N}{N}\right)^{WT}, \tag{2}$$

откуда для скорости C передачи информации получаем оценку

$$C=\frac{\log_2 M}{T}\leqslant W\log_2\frac{P+N}{N}=W\log_2\left(1+\frac{P}{N}\right). \tag{3}$$

Тут стоит приостановиться и сделать следующее замечание. Если попытаться в шар радиуса $\sqrt{n(P+N)}$ упаковать как можно больше шаров радиуса $\sqrt{nN}$ с вроде бы предполагаемым условием, что вкладываемые шары являются как бы жёсткими и не пересекаются, а могут лишь примыкать друг к другу, то при $n=2WT\gg 1$ количество таких шаров будет катастрофически мало́ по сравнению с указанным выше отношением объёмов. Теорема Шеннона (см. ниже), к доказательству которой мы теперь всё подготовили, утверждает, что, тем не менее, при достаточно больших временах T можно добиться скорости передачи, сколь угодно близкой к указанной верхней оценке величины C, имея при этом сколь угодно малую вероятность ошибки при передаче сообщений.

Именно возможность делать ошибки, хотя бы сколь угодно редко, позволяет освободиться от условия, что вкладываемые шары не пересекаются. Если размерность пространства велика, то, как мы отмечали в части II, центры шаров можно сближать, шары будут пересекаться, но относительный объём пересечения при этом может быть очень мал, даже когда центры шаров одинакового радиуса сблизились на расстояние радиуса. При сближении центров количество погружаемых шаров растёт, но растёт и вероятность ошибки при декодировании принятого сигнала.

Учёт взаимодействия указанных обстоятельств и составляет геометрическую основу теоремы Шеннона.

2.3. Теорема Шеннона

Теорема. *Пусть P — средняя мощность передатчика, и пусть помеха есть белый шум с мощностью N в полосе W. Применяя достаточно сложную систему кодирования, можно передавать двоичные цифры со скоростью*

$$C = W\log_2\frac{P+N}{N}$$

со сколь угодно малой частотой ошибок. Никакой метод кодирования не допускает передачи с большей средней скоростью при произвольно малой частоте ошибок.

Возьмём число M, равное числу, стоящему в правой части оценки (2). Это большое число, и, пренебрегая его дробной частью, будем считать, что оно целое. В шаре $B(o,\sqrt{nP})\subset\mathbb{R}^n$, где должны находиться передаваемые векторы (слова, сигналы) x, выберем наугад M точек. «Наугад» значит, что точки случайны, независимы и вероятность попадания точки в какую-то область пропорциональна объёму области, т. е. равна отношению этого объёма к объёму всего шара $B(o,\sqrt{nP})$. (Если случайный выбор M точек шара повторить много раз, то, как правило, точки будут распределяться так.) У нас $n=2WT$, а $M=\left(\frac{P+N}{N}\right)^{WT}$, поэтому одна точка будет приходиться на объём величины $\frac{1}{M}|B|$, где $|B|$ — объём всего шара $B(o, \sqrt{nP})$, а значит, вероятность попадания в эту же область ещё одной точки из наших M точек будет равна $\frac{1}{M}=\left(\frac{N}{P+N}\right)^{WT}$. При $T\to+\infty$ эта вероятность, конечно, стремится к нулю даже независимо от соотношения положительных величин P и N.

Если наши M точек выбраны случайно, то, учитывая, что $n= =2WT\gg 1$ и объём шара $B(o, \sqrt{nP})$ сконцентрирован у его граничной сферы, можно с ничтожной относительной погрешностью (которая тем меньше, чем больше T) считать, что все выбранные точки окажутся в сколь угодно малой окрестности этой граничной сферы.

Напомним ещё, что вектор ξ помехи, как мы показали выше, при $n\gg 1$ со сколь угодно близкой к единице вероятностью ортогонален вектору x сигнала. Итак, при $n=2WT\gg 1$ (т. е. при $T\to+\infty$) мы имеем[1] $\|x\|^2=nP$, $\|\xi\|^2=nN$, $\|x+\xi\|^2=n(P+N)$.

Теперь проведём заключительное рассуждение и выполним конкретную оценку. Предположим, что мы сделали типичный случайный выбор M точек в шаре $B(o, \sqrt{nP})$. Будем считать их отвечающими M различным сообщениям, которые мы собираемся посылать через канал связи. Такой выбор точек в шаре плюс такое их соответствие сообщениям означает определённое кодирование предполагаемых к передаче сообщений. Мы заранее согласуем с принимающим устройством выбранный код. Если бы не было помех, то приёмник, приняв не искажённый сигнал x, однозначно бы его дешифровал в отвечающее ему в силу согласованного кода сообщение.

При наличии в канале шума на принимающее устройство вместо x поступает сигнал $x+\xi$. Приёмник ищет ближайшую к $x+\xi$ точку шара $B(o, \sqrt{nP})$ среди M точек фиксированного кода и считает её переданным сигналом. Ошибка при этом возможна, т. е. можно прочитать не то сообщение, которое было послано. Но это возможно лишь тогда, когда в $(\|\xi\|=\sqrt{nN})$-окрестность точки $x+\xi$ помимо x попадёт ещё какая-то из M точек кода.

Оценим сверху вероятность такого события. Для этого оценим объём пересечения $(\sqrt{nN})$-окрестности точки $x+\xi$ с шаром

[1] Если P и N интерпретировать как дисперсии D_x, D_ξ сигнала и помехи, то мы здесь имеем классическое теоретико-вероятностное соотношение для дисперсии суммы независимых случайных величин: $D^2_{x+\xi}=D_x{}^2+D_\xi{}^2$.

$B(o, \sqrt{nP})$. Поскольку $\|x+\xi\|^2 = n(P+N)$, это простая геометрическая задача. Рассмотрим двумерную плоскость, проходящую через начало координат o и точки $a = x$, $b = x + \xi$. Треугольник oab — прямоугольный с прямым углом при вершине a, известными длинами катетов $|oa| = \sqrt{nP}$, $|ab| = \sqrt{nN}$, и гипотенузы $|ob| = \sqrt{n(P+N)}$. Вычисляя его площадь двумя способами, легко находим длину h высоты, опущенной из вершины a на гипотенузу: $h = \sqrt{n\frac{PN}{P+N}}$. Если теперь взять шар радиуса h с центром в основании этой высоты, то он, очевидно, покроет всю интересующую нас область пересечения шара $B(o, \sqrt{nP})$ и $(\sqrt{nN})$-окрестности точки $b = x + \xi$. Значит, вероятность того, что помимо точки x, лежащей на границе этой области, в неё попадёт ещё какая-нибудь из M кодовых точек, меньше, чем отношение объёма шара радиуса h к объёму шара радиуса $\sqrt{nP}$. Тем самым эта вероятность меньше чем $\left(\frac{N}{P+N}\right)^n = \left(\frac{N}{P+N}\right)^{WT}$, и она стремится к нулю при $T \to +\infty$.

Таким образом, при достаточно больших значениях T со сколь угодно малой вероятностью ошибки в таком канале связи за время T можно различить один из $M = \left(\frac{P+N}{N}\right)^{WT}$ различных объектов, точнее, идентифицировать одно из M возможных различных сообщений. В переводе на двоичные единицы это $\log_2 M$ бит информации за время T. Значит, мы действительно можем достичь скорости передачи, сколь угодно близкой к общей верхней оценке скорости, указанной в неравенстве (3).

Теорема доказана.

§3. Обсуждение теоремы Шеннона, примеры и дополнения

3.1. Комментарий Шеннона

Лучшим коротким комментарием к этой теореме, проливающим свет на некоторые её стороны, которые при первом чтении редко бывают замеченными, являются следующие слова самого Шеннона [2, с. 103].

«Мы будем называть систему, передающую без ошибок со скоростью C, идеальной системой. Такая система не может быть осуществлена ни при каком конечном процессе кодирования, но к ней можно насколько угодно приблизиться. По мере приближения к идеалу происходит следующее: 1. Скорость[1] передачи двоичных чисел приближается к $C = W\log_2\left(1+\frac{P}{N}\right)$. 2. Частота ошибок приближается к нулю. 3. Передаваемый сигнал по своим статистическим свойствам приближается к белому шуму. Это справедливо, грубо говоря, потому, что применяемые функции сигнала должны быть распределены случайно внутри шара радиуса $\sqrt{2WTP}$. 4. Пороговый эффект становится острым. Если помеха превзойдёт значение, для которого построена система, частота ошибок возрастёт очень быстро. 5. Требуемые задержки в передатчике и приёмнике неограниченно возрастают. Конечно, в широкополосной системе задержка в одну миллисекунду может уже рассматриваться как бесконечная».

Здесь, возможно, требует пояснения лишь первая фраза п. 5, что заодно пояснит и реальный смысл величины C, фигурирующей в теореме как скорость передачи. Чтобы записать двоичными символами каждый из M различных объектов, требуется $\log_2 M$ двоичных знаков. Индивидуальное сообщение из M возможных посылается или поступает только после того, как передан или, соответственно, принят весь его двоичный код длины $\log_2 M$. На это вместе требуется время $2T$, которое и подразумевает задержку в передаче самого сообщения, в то время как средняя скорость передачи двоичных знаков, т. е. бит в секунду, при $T \to +\infty$ действительно приближается к указанному в теореме верхнему пределу.

Дальше мы приведём несколько примеров, которые, возможно, прояснят ещё некоторые аспекты затронутого круга вопросов.

3.2. Слабый сигнал в большом шуме

Из построения оптимального кода видно (и это явно отмечено Шенноном в п. 3 приведённой выше цитаты), что по статистическим свойствам такой код похож на белый шум. Это значит, что установить (обнаружить) контакт с очень умной внеземной цивилизацией, посылающей нам сигналы, неотличимые от шума, довольно трудно.

Но давайте рассмотрим следующую ситуацию. К нам поступает очень слабый периодический сигнал на фоне больших помех, которые, однако, случайны. Например, в канале большой белый шум. Можно ли выделить полезный сигнал f из шума? Допустим, что мы знаем или смогли узнать период T сигнала f. Давайте прослушаем

[1] Полное совпадение с формулой Мещерского—Циолковского!

и запишем сигнал вместе с шумом $n \gg 1$ раз. Потом воспроизведём все эти записи синхронно, т. е. сложим. Тогда случайный шум будет сам себя подавлять, а сигнал будет усиливаться. Так что со случайным шумом иногда можно пытаться бороться, используя его же.

3.3. Избыточность языка

С шумом в канале связи мы очень часто справляемся подобным путём, который наука называет усложнением кода или его избыточностью.

В п. 4 Шеннон отмечает острый пороговый эффект оптимального кода. Мы к этому ещё вернёмся чуть позже, но кое-что на примере поясним уже здесь.

Если в телефонном разговоре вы что-то диктуете собеседнику, а он какое-то слово не расслышал или не знает, вы начинаете повторять или сообщать слово по буквам, а буквы сообщать, произнося целые слова типа Анна, Мария, Балбес, Аристотель...

Вы боретесь с помехами, кодируя А, М, Б... заведомо избыточным кодом. Оптимально экономный код, конечно, замечателен, но так же опасен, как всякий максимум потенциальных возможностей, — он неустойчив.

Любой разговорный язык, как мы легко замечаем, избыточен (примерно на 50 %), но зато пригоден для повседневного общения.

3.4. Тонкие измерения на грубом приборе

Как измерить толщину листа бумаги на приборе, где только что с точностью до 0,5 см измерили ваш рост? Вспомним приведённый выше пример слабого периодического сигнала в большом шуме канала связи. Если есть возможность взять несколько пачек этой бумаги и, сложив их, найти, например, что тысяча листов имеет толщину 20 см с абсолютной погрешностью в пределах 0,5 см, то, считая листы примерно одинаковыми, найдём, что толщина одного листа 0,2 мм с возможной ошибкой в пределах 0,005 мм.

Идеи рассмотренных примеров распространяются и на создание надёжных конструкций (приборов, аппаратов) из ненадёжных элементов.

3.5. Код Шеннона—Фано

Спрятанная в многочисленных деталях доказательства вероятностная структура оптимального кода теоремы Шеннона отчётливо различима во всех подробностях в следующей обнажённой идее оптимального кода, называемого кодом Шеннона—Фано.

Для экономии места и слов мы рассмотрим простой показательный пример, возможность обобщения которого будет очевидна.

У нас есть алфавит из четырёх букв, из которых составляются слова. По каналу связи с одинаковой скоростью и надёжностью можно передавать двоичные символы 0 и 1. Ими можно закодировать буквы алфавита, например так: $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$. После этого можно передавать текст по буквам. Забудем пока про помехи и позаботимся об экономности кода, что влияет на скорость передачи информации.

Допустим, что статистический анализ языка установил, что четыре буквы алфавита встречаются в текстах с разной частотой. Пусть, например, их вероятности соответственно $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$.

Тогда разумнее поступить так. Сначала разделим буквы на две равновероятные группы (здесь это первая буква и все остальные), которые различим символами 0 и 1 соответственно. Потом повторим такую же процедуру в каждой из групп и её подгрупп, пока подгруппы не сведутся к единственному элементу. Это и есть идея кода Шеннона—Фано. В нашем случае этот код выглядит так: (0), $(1, 0)$, $(1, 1, 0)$, $(1, 1, 1)$.

Сравним два приведённых кода на достаточно длинном тексте из T букв. В первом случае нам потребуется передать $2T$ двоичных символов. Во втором случае

$$\left(\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3\right)T=\frac{7}{4}T.$$

Более того, даже без знаков препинания в оптимальном коде можно восстановить последовательность букв (расшифруйте, например, 10011000111110). Но стоит сделать ошибку в передаче или приёме хотя бы одного двоичного символа, как текст становится нечитаемым.

3.6. Статистические характеристики оптимального кода

Рассмотренный пример, демонстрирующий идею кода Шеннона—Фано, показывает, что оптимальный код стремится равномерно распределить информацию по передаваемым символам. Этого можно достичь при передаче длинных сообщений, если предварительно осуществлять их статистическую обработку.

Опасности оптимально экономного кода теперь тоже ясны.

Отметим ещё, что теорема Шеннона относится к помехе, предполагавшейся белым шумом. Помехи могут быть разной природы, как случайными, так и детерминированными. При этом даже случайные помехи могут иметь разные статистические характеристики. Разумеется, в конкретной ситуации надо действовать конкретно. Общая теория даёт указания о разумном порядке таких действий, но не решает все проблемы разом.

3.7. Кодирование и декодирование — ε-энтропия и δ-ёмкость

Ранее мы уже упомянули о квантовании непрерывного сигнала по дискретному набору уровней. Стандартная процедура, позволяющая перейти к дискретному конечному описанию любого компактного подмножества метрического пространства, состоит в построении его конечной ε-сети, т. е. такой конечной совокупности точек, что любая точка компакта может быть с точностью до ε-смещения аппроксимирована (заменена) одной из точек сети. Величина ε характеризует точность приближения, допускаемую или допустимую ошибку. Если используемые приборы не способны различать объекты в масштабах, меньших ε, то без особой нужды нет смысла возиться со всеми точками компакта, достаточно заменить компакт какой-то его ε-сетью. Саму ε-сеть можно считать дискретным кодом, описывающим компакт с точностью до ε.

Разумеется, желательно иметь по возможности наиболее экономную ε-сеть, т. е. содержащую возможно меньше точек. Когда ε стремится к нулю, количество N_ε точек в такой наиболее экономной ε-сети, вообще говоря, неограниченно растёт. Скорость этого роста связана со спецификой компакта и метрики пространства.

Величину $\log_2 N_\varepsilon$ Колмогоров назвал *ε-энтропией* компакта.

Если, например, взять единичный куб I^n или любую ограниченную область в евклидовом пространстве $\mathbb{R}^n$, то, как нетрудно проверить, предел отношения $\log N_\varepsilon$ и $\log\frac{1}{\varepsilon}$ при $\varepsilon \to 0$ даст размерность n пространства.

Кстати, это обстоятельство можно превратить в определение размерности и иметь возможность говорить не только о целых размерностях.

Другой, более интересный пример использования понятия ε-энтропии, о котором здесь стоит упомянуть, связан с тринадцатой проблемой Гильберта [3]. Огрубляя ситуацию, вопрос можно сформулировать в следующем рекламном виде: существуют ли функции многих переменных? Поточнее: всякую ли функцию нескольких переменных можно собрать из функций меньшего числа переменных, т. е. представить в виде суперпозиции (композиции) конечного числа таких функций?

А. Н. Колмогоров и В. И. Арнольд доказали, что всякая непрерывная функция многих переменных может быть представлена в виде суперпозиции непрерывных функций одной и двух переменных, причём, как заметил Колмогоров, в качестве функции двух переменных достаточно иметь лишь функцию $x+y$ (см. [6, 7]).

Но ещё до этого А. Г. Витушкин показал, что не каждая гладкая функция многих переменных является суперпозицией столь же гладких функций меньшего числа переменных, см. [4]. Чтобы сформулировать это точно, рассмотрим вслед за Витушкиным число $v = n/p$ — отношение количества переменных функции к порядку её непрерывных производных. Оно служит показателем сложности функции в смысле Витушкина. Обозначим, как всегда, через $C_n^{(p)}$ класс p-гладких функции от n переменных, определённых на единичном n-мерном кубе $I^n \subset \mathbb{R}^n$. Пусть $k < n$. Спрашивается, когда любую функцию класса $C_n^{(p)}$ можно представить суперпозицией функций класса $C_k^{(q)}$? Витушкин показал, что для этого необходимо, чтобы было выполнено соотношение $\left(\frac{n}{p} = v\right) \leqslant \left(\widetilde{v} = \frac{k}{q}\right)$.

Доказательство Витушкина использовало, в частности, оценки Олейник чисел Бетти алгебраических многообразий, полученные в связи с её и Петровского исследованиями шестнадцатой проблемы Гильберта о числе и расположении овалов вещественной алгебраической кривой.

Колмогоров дал иное, прямое и, по-видимому, самое естественное объяснение (доказательство) результата Витушкина, связанное как раз с информацией и энтропией [9].

Пространства $C_n^{(p)}$, $C_k^{(q)}$ бесконечномерны, но, как показал Колмогоров, если в этих пространствах брать компакты, состоящие из всех функций, производные которых ограничены фиксированной константой, то их ε-энтропия при $\varepsilon \to 0$ будет расти как $\left(\frac{1}{\varepsilon}\right)^{\frac{n}{p}}$ и $\left(\frac{1}{\varepsilon}\right)^{\frac{k}{q}}$ соответственно. Если все функции из $C_n^{(p)}$ допускают представление в виде суперпозиции конечного числа функций класса $C_k^{(q)}$, то

$$\left(\frac{1}{\varepsilon}\right)^{\frac{n}{p}} = O\left(\left(\frac{1}{\varepsilon}\right)^{\frac{k}{q}}\right).$$

Таким образом, должно выполняться неравенство

$$\left(\frac{n}{p} = v\right) \leqslant \left(\widetilde{v} = \frac{k}{q}\right).$$

По поводу самой проблемы Гильберта стоит всё же заметить, что в рамках алгебраических функций (которые, возможно, тоже имел в виду Гильберт, говоря в своей тринадцатой проблеме о представлении решений алгебраического уравнения седьмой степени) проблема ещё остаётся открытой. См. в этой связи источники [3, 9].

Мы не собираемся углубляться здесь в эти вопросы и всего лишь упомянули пример иного нетривиального использования понятия дискретного кода и ε-энтропии.

Вернёмся ещё ненадолго к дискретному кодированию и добавим несколько слов о декодировании. Экономная ε-сеть может служить экономным дискретным кодом объекта (метрического компакта), описывающим его с точностью до ε. Предположим, что такой экземпляр кода есть на обоих концах канала связи. Если при передаче сообщений нет ошибок, то на приёмном конце получают сигнал о той точке ε-сети, которую выбрали на передающем конце. Но если по тем или иным причинам в канале связи возможны ошибки и переданная точка на приёмном конце может быть воспринята со смещением в пределах её δ-окрестности, то, очевидно, возможны ошибки при декодировании, о чём мы уже и выше говорили. Если мы намерены полностью исключить возможность ошибки, то от экономного кода в виде ε-сети придётся отказаться. Теперь уже наоборот, надо искать максимальное количество n_δ точек компакта, удалённых друг от друга не менее чем на 2δ. Только такой код (он, очевидно, будет 2δ-сетью) при указанных условиях может обеспечить безошибочность передачи.

Если величина $\log_2 N_\varepsilon$, как мы уже знаем, называется *ε-энтропией* компакта, то величина $\log_2 n_\delta$ называется его *δ-ёмкостью* или *δ-вместимостью*.

Подсчёт ε-энтропии и δ-ёмкости различных классов функций и дальнейшие ссылки можно найти в работах [10, 11].

Некоторые дальнейшие сведения, относящиеся к обработке сигналов, можно найти в работах [11—16].

§ 4. Математическая модель канала связи с помехами

4.1. Простейшая модель и постановка вопроса

Как обычно, для экономии всего на свете рассмотрим сначала простейшую модель, которая, однако, уже содержит почти всё самое ценное из нужного нам сейчас и при желании легко может быть обобщена.

В канале связи передатчик посылает на приёмник символы 0 и 1. Помехи приводят к тому, что приёмник порой может расшифровать посланный символ 0 как 1, а 1 как 0. Пусть p — вероятность правильного прохождения переданного символа.

По каналу посылаются сообщения (текст, слова), состоящие из последовательности букв — символов 0, 1 нашего простейшего двухбуквенного алфавита. Считаем, что канал действует на каждую букву слова независимо, т. е. это *канал без памяти*.

Какова пропускная способность такого канала связи?

Вопрос поставлен. Интуитивно понятно, что он разумен. Одновременно понятно, что для ответа на него надо себе уяснить, что именно имеется в виду[1].

Выше, прежде чем доказывать теорему Шеннона, мы (вслед за Шенноном) должны были сначала разобраться в смысле и точном содержании некоторых допущенных нашей интуицией к употреблению терминов и понятий.

Выполним и теперь (опять вслед за Шенноном) эту работу. Конечно, предыдущий опыт нам её сильно облегчит. Собственно, во многом она и будет его абстрактным оформлением.

Что же касается более общей абстрактной модели канала связи с шумом, то, очевидно, вместо алфавита из двух букв можно иметь любой конечный (не однобуквенный) алфавит и иметь вероятности p_{ij} перехода i-й посланной буквы в j-ю на приёмном конце. Матрица (p_{ij}) вероятностей переходов моделирует некоторый канал связи с помехами.

[1]Рассказывали, что кого-то из визитёров известного Принстонского института перспективных исследований (IAS) посадили в пустовавший в тот момент кабинет Гёделя. Уезжая, визитёр оставил на столе благодарственную записку хозяину с выражением сожаления, что ему с Гёделем не удалось познакомиться поближе. Через какое-то время он получает вежливое письмо от прочитавшего записку Гёделя, который просит уточнить, что именно тот имел в виду.

Если буквы искажаются с разной вероятностью, то скорость передачи информации может зависеть (и зависит) от разумности кода, используемого для записи передаваемых сообщений. Ясно, что лучше почаще использовать те символы, которые меньше подвергаются искажениям. Кроме того, как мы уже знаем по опыту кода Шеннона—Фано, полезно учесть статистические особенности текста самого сообщения, подлежащего передаче.

Пропускная способность канала, заданного матрицей (p_{ij}), по-видимому, должна быть просто верхней гранью возможной скорости передачи относительно всего, что от самого канала (прибора) уже не зависит, например, верхней гранью по всем возможным кодировкам посылаемых текстов.

Ясно, что один и тот же прибор разные пользователи могут использовать с разной степенью эффективности. Возможности самого прибора должны оцениваться в предположении, что он используется максимально эффективно.

После того как максимальная скорость при оптимальном коде будет достигнута, конечно, могут появиться новые вопросы. Мы видели, например, какие опасности скрываются за предельно экономными кодами. Но отложим это всё в сторону. Сейчас надо постепенно разобраться со скоростью передачи информации и с тем, что же мы вообще понимаем под терминами *информация* и *количество информации*.

4.2. Информация и энтропия (предварительные рассмотрения)

Как мы уже отмечали, появление телеграфа и радиосвязи стимулировало разработку понятия информации и её количественного описания.

Мерой информации, по-видимому, разумно считать меру изменения неопределённости, связанную с полученной информацией.

В простейшей ситуации, когда имеются две равноправные возможности, например, когда случайная величина может иметь ровно два равновероятных значения 0 и 1 (выключено, включено), сообщение о её конкретном значении (состоянии) ликвидирует неопределённость. Напомним, что мера такой информации принимается за единицу и называется *битом* (bit — сокращение от binary digit — двоичная цифра).

Чтобы идентифицировать один из M объектов, задавая вопросы, на которые отвечают только «да» (1) или «нет» (0), как известно, требуется $\log_2 M$ двоичных символов (алгоритм деления пополам). Такая система (случайная величина) способна запасти $m = \log_2 M$ бит информации (в соответствии с мерой её неопределённости). Точнее, если все M возможных значений (состояний) рассматриваемой случайной величины равновероятны, то идентификация (выделение, сообщение о реализации) одного из них при указанном соответствии равносильна заданию $\log_2 M$ двоичных цифр, т. е. сообщению $\log_2 M$ бит информации.

Теперь формальнее: пусть X — произвольная дискретная случайная величина, которая может принимать M различных значений x_i с вероятностями p_i соответственно.

Как учесть вероятности? Какую меру неопределённости (и информации) разумно связывать с такой случайной величиной?

Запишем только что полученный нами результат в следующем виде:

$$m = \log M = M \cdot \frac{1}{M} \log M = \sum \frac{1}{M} \log M = -\sum \frac{1}{M} \log \frac{1}{M},$$

где $\frac{1}{M}$ будем трактовать как вероятность появления (реализации, выбора) конкретного из этих M объектов. (Здесь и далее $\log = \log_2$.)

Тогда, наверное, в общем случае мы должны прийти к величине $-\sum\limits_{i=1}^{M} p_i \log p_i$. Мы сейчас подкрепим это предположение.

Величина $H(X) = -\sum p_i \log p_i$ называется *энтропией* дискретной случайной величины X. (По непрерывности считаем, что $0 \log 0 = 0$.)

Поэкспериментируем с ней. Если вероятность p_i события x_i мала, то, надо полагать, информация о том, что произошло очень редкое событие, очень большая $(-\log p_i)$. С другой стороны, если событие редкое, то на протяжении длительного периода наблюдений оно появляется со своей информацией в зубах крайне редко (в доле p_i от всего времени наблюдений). Поэтому осреднённая по большому промежутку наблюдений информация, которую приносит такое событие (такое значение x_i случайной величины X), будет равна $-p_i \log p_i$.

Так что если $-\log p_i$ есть мера неопределённости и информации, связываемая с событием x_i, вероятность которого p_i, то $-p_i \log p_i$ есть среднестатистическое количество информации, которое приносит появление такого события, а тогда

$$H(X) = -\sum_{i=1}^{M} p_i \log p_i$$

(математическое ожидание $-\log p_i$) есть среднее количество информации, которое несёт единичное событие (значение) случайной величины X.

Обратите внимание на то, что мы не интересуемся здесь тем, в чём именно состоит реальное событие x_i, хотя для других целей это, может быть, и есть самое главное.

Фиксируем теперь в точной формальной записи статистический характер энтропии: для любых положительных чисел ε, δ найдётся такое число $n_{\varepsilon\delta}$, что при $n \geqslant n_{\varepsilon\delta}$ выполняется неравенство

$$P\left\{\left|-\frac{1}{n}\sum_{i=1}^{n} \log p_{x_i} - H(X)\right| < \delta\right\} > 1-\varepsilon, \tag{4}$$

где, как обычно, P — вероятность указанного в скобках события, но теперь x_i, $i=1, \ldots, n$, — n независимых значений случайной величины X, а p_{x_i} — вероятности этих значений.

Как связана энтропия с кодированием?

Рассмотрим сообщения/слова/векторы $\bar{x} = (x_1, \ldots, x_n)$, образованные n последовательными независимыми значениями случайной величины X. Вероятность $p_{\bar{x}}$ появления слова $\bar{x}$ равна

$$p_{\bar{x}} = p_{x_1} \cdot \ldots \cdot p_{x_n}.$$

В силу соотношения (4) при $n \geqslant n_{\varepsilon\delta}$ с вероятностью, большей чем $1-\varepsilon$, будем иметь

$$2^{-n(H(X)+\delta)} \leqslant p_{\bar{x}} \leqslant 2^{-n(H(X)-\delta)}. \tag{5}$$

Слово $\bar{x}$ называют *δ-типичным*, если для него выполнены эти оценки. Ясно, что существует не более $2^{n(H(X)+\delta)}$ таких δ-типичных слов, а если $n \geqslant n_{\varepsilon\delta}$, то их ещё и не меньше чем $(1-\varepsilon)2^{n(H(X)-\delta)}$, и при этом всё множество не δ-типичных слов имеет вероятность, не большую чем ε.

В принципе теперь уже можно использовать двоичные последовательности длины $n(H(X)+\delta)$, чтобы закодировать все δ-типичные слова. Даже если все остальные слова закодировать одним символом, вероятность ошибки при передаче слов $\bar{x}$ длины n, вызванная таким кодом, будет меньше ε.

С другой стороны (и этот эффект неустойчивости экономных кодов мы уже отмечали), любой код, использующий в той же ситуации двоичные последовательности относительно чуть меньшей длины $n(H(X)-\delta)$ (например, $2\delta n$ из $n(H(X)+\delta)$ посланных символов потерялось в шуме), будет иметь асимптотически неисчезающую вероятность ошибки, стремящуюся к единице при $n \to +\infty$.

Итак, связь энтропии и кодирования информации состоит, например, в том, что эффективное кодирование асимптотически при $n \to +\infty$ требует $N \sim 2^{nH(X)}$ слов и энтропия $H(X)$ может интерпретироваться как мера количества информации в битах на передаваемый символ, т. е. на одно значение случайной величины X.

Отсюда, в частности, следует, что энтропия источника информации не должна превышать пропускную способность канала связи, если мы хотим адекватно и без задержек передавать поступающую информацию по этому каналу связи.

4.3. Условная энтропия и информация

Вернёмся постепенно к передаче информации по каналу связи. Передатчик посылает сообщение $\bar{x} = (x_1, \ldots, x_n)$, приёмник получает $\bar{y} = (y_1, \ldots, y_n)$. Как по полученному восстановить посланное? Если никаких искажений нет, т. е. всегда $y_i = x_i$, то и вопроса нет. Поэтому будем считать, что канал характеризуется некоторой матрицей (p_{ij}) вероятностей перехода посланного сигнала x_i в принятый сигнал y_j.

Поставим вопрос иначе. Какую информацию о сообщении $\bar{x}$ содержит сообщение $\bar{y}$? Или по-другому: как меняется (уменьшается) неопределённость $\bar{x}$, когда мы узнаём $\bar{y}$?

Обратимся к условным вероятностям и введём понятие условной энтропии $H(X \mid Y)$ случайной величины X на входе канала связи относительно случайной величины Y на выходе.

Тогда вслед за Шенноном можно будет рассмотреть величину

$$I(X; Y) = H(X) - H(X \mid Y), \tag{6}$$

и считать её эффективным *количеством информации*, которое в среднем переносится одним посылаемым сигналом (значением случайной величины X) в таком канале связи.

Значит, *пропускная способность канала связи* определится как

$$C = \sup_{\{p_x\}} I(X; Y), \tag{7}$$

где супремум берётся по всевозможным кодам, т. е. по всевозможным распределениям вероятностей $\{p_x\}$ входной случайной величины X, имеющей фиксированный конечный набор значений (алфавит).

Итак, определим условную энтропию $H(X\,|\,Y)$ одной случайной величины X относительно другой случайной величины Y.

Пусть $\{p_x\}$, $\{p_y\}$ и $\{p_{x,y}\}$ — распределения вероятностей случайных величин X, Y и совместной случайной величины $Z=(X,Y)$ соответственно.

Если вероятность появления на входе значения x_i случайной величины X равна p_{x_i}, а вероятность $p(y_j\,|\,x_i)$ перехода x_i в y_j дана и равна p_{ij}, то вероятность p_{x_i,y_j} совместного события $z_{ij}=(x_i,y_j)$ равна $p(y_j\,|\,x_i)p_{x_i}$, а общая вероятность p_{y_j} появления на выходе значения y_j случайной величины Y равна $\sum\limits_i p(y_j\,|\,x_i)p_{x_i}$.

Для облегчения записей формул и без потери их ясности мы не будем больше писать дополнительные нижние индексы. Например, стандартные формулы условной вероятности будем писать так: $p_{x,y}=p(y\,|\,x)p_x$ или $p_{x,y}=p(x\,|\,y)p_y$, поскольку $p_{x,y}=p_{y,x}$.

Найдём сначала условную энтропию $H(X\,|\,Y=y)$ случайной величины X при условии, что случайная величина Y имеет значение y. Иными словами, мы сейчас найдём, какой становится энтропия (неопределённость) X при условии, что случайная величина Y приняла значение y.

$$H(X \mid Y=y) = -\sum_x p(x \mid y)\log p(x \mid y) = -\sum_x \frac{p_{x,y}}{p_y}\log\frac{p_{x,y}}{p_y}.$$

Теперь можно найти и интересующую нас условную энтропию $H(X|Y)$ случайной величины X относительно случайной величины Y:

$$\begin{aligned}H(X \mid Y) &= -\sum_y p_y H(X \mid Y=y) = -\sum_y p_y \sum_x \frac{p_{x,y}}{p_y}\log\frac{p_{x,y}}{p_y} = \\ &= -\sum_{x,y} p_{x,y}\log p_{x,y} + \sum_y p_y \log p_y = H(X,Y) - H(Y).\end{aligned}$$

Здесь $H(X,Y)$ — совместная энтропия пары $Z=(X,Y)$ случайных величин X и Y; распределение вероятностей пары есть $\{p_{x,y}\}$.

Мы нашли, что $H(X|Y)=H(X,Y)-H(Y)$. Но поскольку $p_{x,y}=p_{y,x}$ и $p_{x,y}=p(y\,|\,x)p_x=p(x\,|\,y)p_y=p_{y,x}$, справедливы также соотношения $H(X,Y)=H(Y,X)$ и $H(X,Y)=H(X|Y)+H(Y)=H(Y|X)+H(X)$.

Итак,

$$H(X, Y) = H(X \mid Y) + H(Y) = H(Y \mid X) + H(X) = H(Y, X). \qquad (8)$$

Учитывая формулу (6) (определение Шеннона) количества информации, находим, что

$$I(X; Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X, Y). \qquad (9)$$

Поскольку $H(X, Y) = H(Y, X)$, отсюда следует, что

$$I(X; Y) = I(X; Y). \qquad (10)$$

4.4. Интерпретация потери информации в канале с шумом

Остановимся коротко на итоговом неформальном осмыслении введённых понятий и обнаруженных взаимоотношений.

Энтропия $H(X) = -\sum\limits_x p_x \log p_x$ дискретной случайной величины X является некоторой статистической, осреднённой её характеристикой. Если $-\log p_x$ трактовать как выраженную в двоичных единицах (битах) меру неопределённости, редкости события (значения x случайной величины X) и пропорционально этому измерять количество информации, содержащейся в сообщении о наступлении события x, то $H(X)$ будет математическим ожиданием этой величины $-\log p_x$.

Энтропия есть некая средняя мера неопределённости случайной величины X, приходящаяся на одно её значение. В другой трактовке это средняя мера информации, которая приходится на одно значение случайной величины. При этом предполагается, что мы получаем длинную серию независимых значений случайной величины X и осредняем количество полученной информации по количеству принятых значений случайной величины. По умолчанию предполагается также, что значения передаются и поступают равномерно — одно значение в какую-то одну и ту же единицу времени. Именно поэтому в случае, когда речь идёт о передаче информации по каналу связи, энтропию источника предпочитают трактовать как среднее количество информации, которое создаёт источник за единицу времени.

Если канал способен пропустить без искажений такой поток информации, то всё в порядке. Если же могут возникать ошибки, то возникают новые проблемы. На примере теоремы Шеннона видно, что в конкретной ситуации приходится иметь дело с конкретными физическими параметрами (полоса частот, уровень сигнала, уровень шума, статистические характеристики помехи и т. п.). Правильно учесть и задействовать эти параметры — самостоятельная важная задача.

Мы рассмотрели некоторую абстрактную модель канала связи с помехами и пришли к полезному понятию условной энтропии $H(X\,|\,Y)$. Её смысл в том, чтобы оценить средний уровень неопределённости случайной величины X, который остаётся, если у вас есть возможность наблюдать за состоянием случайной величины Y. Если X и Y независимы, то ясно, что наблюдение за Y ничего не говорит о X и $H(X\,|\,Y)=H(X)$. С другой стороны, если $X=Y$ (например, когда идёт безошибочная передача по каналу связи), то $H(X\,|\,Y)=0$.

Таким образом, величину $H(X\,|\,Y)$ в задаче передачи информации по каналу связи можно трактовать как среднюю на передаваемое значение (на символ или на единицу времени) потерю информации в этом канале связи.

Тогда именно

$$I(X;Y)=H(X)-H(X\mid Y)$$

естественно принять за среднюю меру информации, которая проходит по каналу связи при передаче по нему значений случайной величины X, кодирующих исходные сообщения. Содержательная часть сообщений нас не интересует. Информацию мы измеряем в битах, а скорость её создания, передачи или приёма измеряем в битах на символ или в битах на единицу времени.

Значения, которые может принимать случайная величина X, можно считать алфавитом, в котором записываются (кодируются) подлежащие передаче сообщения. Сообщения считаются достаточно длинными, чтобы в задаче вообще можно было использовать статистические характеристики. Распорядиться алфавитом для кодирования сообщений, как мы видели на примере кода Шеннона—Фано, можно по-разному. Оптимальный код для передачи сообщений выбирается с учётом характеристик используемого канала связи.

Пропускная способность (7) канала связи есть максимальная средняя скорость передачи информации по этому каналу связи, которой можно достичь или к которой можно сколь угодно приблизиться при передаче длинных текстов сообщений, предварительно разумно закодированных используемым в канале алфавитом.

4.5. Расчёт пропускной способности абстрактного канала связи

Формулой (7) мы определили пропускную способность абстрактного канала связи. Только что мы в общих чертах обсудили содержание введённых понятий. Теперь, в заключение, проведём всё же один конкретный расчёт. Найдём пропускную способность нашего абстрактного канала связи в простейшем примере, с которого мы вообще начали весь этот абстрактный разговор. Напомним условия.

В канале связи передатчик посылает на приёмник символы 0 и 1. Помехи приводят к тому, что приёмник порой может расшифровать посланный символ 0 как 1, а 1 как 0. Пусть p — вероятность правильного прохождения переданного символа.

По каналу посылаются сообщения (текст, слова), состоящие из последовательности букв — символов 0, 1 нашего простейшего двухбуквенного алфавита. Считаем, что канал действует на каждую букву слова независимо, т. е. это канал без памяти.

Какова пропускная способность такого канала связи?

В данном случае матрица вероятностей перехода упрощена до предела не только двухбуквенностью алфавита, но и тем, что оба передаваемых символа 0 и 1 имеют одинаковую вероятность прохождения без искажения. Итак, случайная величина X на входе канала может принимать два значения. Пусть кодировка посылаемого сообщения такова, что вероятности появления значений 0, 1 соответственно равны p_0, p_1.

На выходе канала случайная величина Y тоже может принимать только эти два значения 0 или 1, но, возможно, с другими вероятностями q_0, q_1. Найдём их.

Значение 0 на выходе получится с вероятностью p, если на входе 0, и с вероятностью $1-p$, если на входе 1. В свою очередь, на входе 0 с вероятностью p_0 и 1 с вероятностью p_1. Поэтому вероятность появления на выходе значения 0 равна $pp_0+(1-p)p_1$. Соответственно, 1 на выходе появляется с вероятностью $pp_1+(1-p)p_0$.

Распределение вероятностей совместной случайной величины $Z=(X,Y)$ тоже легко выписать:

$$(0,0)\sim pp_0, \quad (0,1)\sim(1-p)p_1, \quad (1,0)\sim(1-p)p_0, \quad (1,1)\sim pp_1.$$

Теперь можно подсчитать энтропии $H(X)$, $H(Y)$, $H(X,Y)$ и по второй из формул (9) найти скорость передачи информации. В нашем случае получается, что

$$I(X;Y) = H(Y) - h(p),$$

где $h(p) = -p\log p - (1-p)\log(1-p) = H(X,Y) - H(X) = H(Y\,|\,X)$.

Максимальное значение величина $I(X;Y)$ достигает при $H(Y)=1$, т. е. когда распределение на выходе равномерно: $q_0 = q_1 = \frac{1}{2}$. Но $q_0 = pp_0 + (1-p)p_1$, а $q_1 = pp_1 + (1-p)p_0$, поэтому условие $q_0 = q_1 = \frac{1}{2}$ выполняется в точности тогда, когда входное распределение равномерно: $p_0 = p_1 = \frac{1}{2}$.

(Мы здесь воспользовались тем, что, как нетрудно проверить, опираясь на выпуклость логарифма, если дискретная случайная величина X имеет M различных значений, то $0 \leqslant H(X) \leqslant \log M$, причём равенство слева реализуется на вырожденных распределениях, когда какое-то одно значение принимается с вероятностью 1, а остальные с вероятностью 0, а равенство справа достигается на равномерном распределении.)

Итак, мы нашли, что пропускная способность рассматриваемого простейшего модельного канала связи с помехами равна $C = 1 - h(p)$. Здесь $h(p) = H(Y\,|\,X)$ характеризует потерю информации на передаваемый символ. (См. также [18].)

Вычисление скорости, разумеется, всегда ведут с точностью до постоянного множителя, отвечающего выбранной единице времени.

Например (см. [17]), пусть канал физически способен пропускать 100 двоичных символов 0, 1 в единицу времени, причём каждый передаваемый символ может замениться противоположным с вероятностью 0,01. В этом случае $h(p) = h(1-p) = h(0{,}01) \approx 0{,}0808$, и $C = 100(1 - 0{,}0808) = 91{,}92 \approx 92$ двоичных знаков — бит в единицу времени.

Обратите внимание на то, что результат не равен 99.

Имея за плечами накопленный опыт, можно теперь попытаться самостоятельно доказать следующую интуитивно понятную теорему Шеннона.

Теорема. *Пусть имеется источник информации X, энтропия которого в единицу времени равна $H(X)$, и канал связи с пропускной способностью C. Тогда если $H(X) > C$, то ни при каком кодировании передача сообщений без задержек и искажений невозможна. Если же $H(X) < C$, то всегда можно достаточно длинные сообщения закодировать так, чтобы они передавались без задержек и при этом вероятность ошибок была бы сколь угодно близка к нулю.*

Тема III

Классическая термодинамика и контактная геометрия

Введение

«Эта термодинамика образует ныне замечательную научную систему, детали которой ни по красоте, ни по блестящей законченности не уступают всей системе в целом; она заслуживает имя термодинамики классической». Так великий Лоренц писал о классической термодинамике, открывая свои «Статистические теории в термодинамике» [2].

Мы здесь остановимся на некоторых математических аспектах термодинамики. Опишем два начала термодинамики на языке дифференциальных форм (глава I). Дадим представление о связи классической термодинамики с контактной геометрией (глава II). Наконец, добавим кое-что из статистической физики и скажем несколько слов о квантово-механической стороне термодинамики (глава III).

Глава I

Классическая термодинамика (начальные представления)

§1. Два начала термодинамики

1.1. Энергия и вечный двигатель

Кто не знает слов и словосочетаний «энергия», «закон сохранения энергии», «вечный двигатель»? Многие даже слышали, что такой двигатель вроде бы невозможен.

А ведь это (после некоторого уточнения) и есть первый из двух фундаментальных законов термодинамики, называемых обычно двумя началами термодинамики (и научного мировоззрения).

Закон сохранения энергии неотрывен от нашего нынешнего сознания. Даже трудно поверить, что в такой форме он появился лишь во второй половине XIX столетия (Майер, Джоуль, Клаузиус, Гельмгольц). Подробности см., например, в [4].

Наиболее популярная формулировка первого начала термодинамики состоит как раз в том, что в природе не существует и неосуществим механизм (источник механической энергии, вечный двигатель), единственным продуктом циклической деятельности которого было бы неограниченно повторяющееся совершение механической работы (например, поднятие груза в 1 миллиграмм на высоту 1 миллиметр).

1.2. Вечный двигатель второго рода и энтропия

«Известно основное значение, которым обладает не только в этом отделе науки, но и среди наших общих знаний о вселенной, второе начало термодинамики, или начало Карно—Клаузиуса. Во всяком случае можно сказать, что оно царствует более чем над половиной физики». Это тоже слова Лоренца из уже упомянутой во введении книги [2].

Второе начало термодинамики, хотя и не засекречено, известно не такому широкому кругу обывателей, как первое, хотя оно имеет ещё более многочисленные и любопытные по непохожести, но эквивалентные по содержанию формулировки. Самая простая и, казалось бы, банальная (формулировка Клаузиуса) утверждает, что при контакте тел тепло (энергия) переходит от более горячего тела к более холодному (первое остывает, а второе нагревается) и само по себе не может перейти в обратном направлении.

Другая форма второго начала, созвучная приведённой выше формулировке первого начала, утверждает, что не существует вечного двигателя второго рода, т. е. циклически действующей машины, единственным результатом рабочего цикла которой было бы превращение взятой из теплового резервуара (океана тепла) тепловой энергии в механическую энергию (Уильям Томсон; с 1892 г. за научные заслуги лорд Кельвин).

Не вникая здесь глубже в известные и представленные в любом солидном учебнике термодинамики детали, добавим лишь, что исторически второе начало, как ни странно, формально было открыто раньше первого родоначальником термодинамики Сади Карно (1824 г.). Он отвечал на вопрос Джеймса Уатта (1765 г.) о возможном коэффициенте полезного действия тепловой (паровой) машины. (Был период появления и распространения паровых машин и двигателей. Уатт спрашивал вполне конкретно: сколько нужно угля, чтобы такая машина выполнила заданную механическую работу.)

Работа Карно (забытая и обнаруженная в 1834 г. Клапейроном) была развита в 1850—60-х годах Клаузиусом и привела к появлению введённого им в 1865 г. второго после энергии фундаментального понятия термодинамики — *энтропии* термодинамической системы.

В некотором смысле два начала термодинамики в результате порождают две фундаментальные характеристики состояния термодинамической системы: его *энергию* и *энтропию*.

Второе начало можно привести к утверждению о том, что любая замкнутая (изолированная от внешних воздействий) термодинамическая система эволюционирует в сторону роста энтропии её состояния.

Например, газ, освобождённый из содержавшего его сосуда, распространится по всей комнате, т. е. перейдёт из более организованного (собранного в сосуд) состояния в менее организованное (и более вероятное, чем если бы он вдруг сам вновь собрался в сосуде)[1].

[1]Зоммерфельд, подчёркивая роль второго начала, цитирует [6] Роберта Эмдена: «В гигантской фабрике естественных процессов принцип энтропии занимает место директора, который предписывает вид и течение всех сделок. Закон сохранения энергии играет лишь роль бухгалтера, который приводит в равновесие дебет и кредит».

К концу XIX — началу XX столетия, с одной стороны, родилась статистическая термодинамика (Максвелл, Больцман, Гиббс, Пуанкаре, Эйнштейн), с другой стороны — началось математическое оформление, формализация классической термодинамики (Максвелл, Гиббс, Планк, Каратеодори).

§ 2. Математическое оформление двух начал термодинамики

В каждой науке есть свои любимые игрушки. В термодинамике это газ в цилиндре под поршнем. Поршень можно двигать, меняя объём газа. Стенки цилиндра могут проводить тепло, а могут и наоборот — изолировать газ от теплообмена с внешней средой.

Играя на бумаге в этот прибор (ключевой элемент как паровой машины, так и современного двигателя внутреннего сгорания), С. Карно провёл один из первых в физике гениальных (и недорогих) мысленных экспериментов. Двигая поршень, подогревая, а когда надо — охлаждая или теплоизолируя цилиндр, он придумал циклический процесс, называемый теперь (после некоторых сделанных Клапейроном вариаций) циклом Карно. Карно нашёл ответ на вопрос Уатта о возможном коэффициенте полезного действия любой тепловой машины, а заодно, как это случается, сделал великое открытие, которое (после разработки Клаузиусом) стало вторым началом термодинамики. (Идея первого начала по существу тоже присутствовала в рассуждениях Карно[1].)

Поиграем и мы чуть-чуть. Это поможет нам как принять полезные математические схемы, так и не потерять потом физическое содержание в возникающих абстракциях.

Классический закон Клапейрона утверждает, что между объёмом V газа, его давлением P и температурой T, характеризующими равновесное термодинамическое состояние газа, имеется связь

$$\frac{PV}{T} = C$$

(называемая уравнением состояния), где постоянная C зависит только от количества газа, см. [7, 8a].

[1]Профессиональные историки науки могут сказать, и не без основания, что мы тут несколько упрощаем дело и что в работах Карно не всё было так уж гладко. Объективности ради отмечаем это, отсылая заинтересованного читателя к оригинальным работам Карно и к работам их комментаторов.

Если надавить на поршень, понемногу (чтобы не выходить за равновесные состояния) поменяв объём газа на некоторую величину dV (отрицательную, если мы сжимаем газ), то мы совершим работу $-P\,dV$. Часть её пойдёт на увеличение внутренней энергии E газа (который мы как пружину сжали), а часть в виде тепла δQ может уйти во внешнюю среду, если наш цилиндр проводит тепло. Если же он адиабатичен, т. е. не проводит тепло, то теплообмена с внешней средой нет, тепловых потерь не будет ($\delta Q = 0$) и просто внутренняя энергия газа E увеличится в точности на величину $dE = -P\,dV$ проделанной нами работы. Естественно считать, что сам газ при этом совершает работу $\delta W = P\,dV$. (Например, если бы он выдавил поршень, увеличив объём на $dV > 0$, то, очевидно, он совершил бы работу $\delta W = P\,dV$.)

В общем случае энергетический баланс состоит в том, что

$$\delta Q = dE + \delta W. \tag{1}$$

Заметим, и это весьма важно, что в отличие от dE дифференциальные формы δQ и δW не являются точными. Это не дифференциалы функций. Работа, которую надо совершить, например, при изменении объёма газа вдвое, зависит не только от начального и конечного значений объёма, так же как и возникающий при этом интегральный теплообмен с внешней средой. Обе эти величины существенно зависят от условий, в которых совершается переход из одного термодинамического состояния в другое. Например, если процесс адиабатический, то теплообмена вообще нет. В таком процессе (на таком пути перехода) интеграл от формы δQ равен нулю. На другом пути, соединяющем те же термодинамические состояния, интеграл от формы δQ, как правило, отличен от нуля, если стенки цилиндра проводят тепло. Ясно, что это автоматически относится и к дифференциальной форме работы δW. Именно поэтому мы употребили различные символы дифференциала в фундаментальном равенстве (1).

Всё, что пока было сказано (без обсуждений и точных формулировок), имело целью воскресить в памяти читателя самые общие сведения из термодинамики. Они служат мотивировкой того формального определения термодинамической системы, которое мы даём ниже.

2.1. Дифференциальная форма теплообмена

В конце XIX — начале XX в. трудами Максвелла, Гиббса, Планка, Каратеодори и других учёных, физиков и математиков, богатый материал термодинамики был подвергнут математической систематизации. Возникли, как это бывает в достаточно продвинутой формализованной теории, основные положения, понятия, принципы, аксиомы, теоремы, компактно включившие в себя многочисленные конкретные достижения, полученные трудом многих исследователей.

Вот к такому математически более систематизированному изложению мы теперь и переходим.

Будем считать, что равновесное состояние нашей термодинамической системы определяется набором параметров

$$(\tau, a_1, \ldots, a_n) =: (\tau, a),$$

где $\tau > 0$ будет играть роль *абсолютной температуры* T, а $a = (a_1, \ldots$ $\ldots, a_n)$ — набор *внешних параметров*, которые можно менять, как объём газа под поршнем в рассмотренном выше «игрушечном» примере.

Саму термодинамическую систему в рассматриваемой математической модели отождествим с фундаментальной дифференциальной формой

$$\omega := dE + \sum_{i=1}^{n} A_i\, da_i, \tag{2}$$

называемой *формой теплообмена* или *формой притока тепла*. Здесь E — *внутренняя энергия системы* (по определению), а A_i — *обобщённая сила*, отвечающая вариации координаты a_i (т. е. $\sum\limits_{i=1}^{n} A_i\, da_i$ отвечает работе δW системы, связанной с изменением внешних параметров, а сама форма ω соответствует дифференциальной форме теплообмена δQ равенства (1)). Величины E и A_i, естественно, зависят от $(\tau, a_1, \ldots, a_n)$. Эти зависимости входят в определение термодинамической системы. Формально они и составляют это определение. Соотношения $A_i = f_i(\tau, a_1, \ldots, a_n)$ называют *уравнениями состояния*.

Форма ω, определяющая термодинамическую систему, должна удовлетворять некоторым требованиям, которые мы сейчас укажем.

2.2. Два начала термодинамики на языке дифференциальных форм

Рассмотрим ориентированный путь γ в пространстве равновесных состояний (τ, a), отвечающий переходу системы из одного равновесного состояния (τ_0, a_0) в другое (τ_1, a_1). Тогда интегралы

$$\int\limits_\gamma \omega = \int\limits_\gamma \delta Q, \quad \int\limits_\gamma \sum_{i=1}^{n} A_i\, da_i = \int\limits_\gamma \delta W$$

дадут соответственно количество тепла, полученное системой, и работу системы, связанную с внешними силами, при таком переходе, а интеграл от dE (см. (1), (2)) даст изменение внутренней энергии системы.

Второе начало термодинамики, открытое Карно, трудами Клаузиуса свелось к тому, что для любого осуществимого термодинамического цикла γ имеет место замечательное неравенство

$$\int\limits_\gamma \frac{\delta Q}{T} \leqslant 0,$$

где T — реальная физическая абсолютная температура, причём равенство $\int\limits_\gamma \frac{\delta Q}{T} = 0$ реализуется в точности тогда, когда замкнутый путь γ идёт в пространстве равновесных состояний термодинамической системы.

Это означает, что дифференциальная форма $\frac{\delta Q}{T}$ является дифференциалом некоторой функции S равновесного состояния системы. Именно её Клаузиус и назвал *энтропией* термодинамического состояния системы. Пока что функция S определена с точностью до аддитивной постоянной. Но чаще всего бывает нужна лишь разность значений этой функции при переходе из одного состояния в другое. В этом случае аддитивная постоянная никакой роли не играет. (Есть, однако, причины, по которым следует считать, что при $T = 0$ энтропия системы во всех состояниях одна и та же. Это теорема Нернста, или, как говорят, третье начало термодинамики. Поэтому при $T = 0$ полагают $S = 0$. На этом и многом другом мы здесь не останавливаемся. См. с. 123, 127.) Итак абсолютная температура T как функция состояния термодинамической системы замечательна тем, что T^{-1} является интегрирующим множителем для дифференциальной формы δQ теплообмена, превращающим её в точную форму — в дифференциал функции S — энтропии системы.

Формализуя сказанное, будем считать, что в математической модели определяющая термодинамическую систему форма ω такова, что *форма $\tau^{-1}\omega$ является точной*. Функцию S, для которой $dS = \tau^{-1}\omega$, естественно, назовём *энтропией* системы.

Рассмотрим теперь адиабатический процесс. Сейчас это путь γ в пространстве равновесных состояний (τ, a), идущий вдоль нулевых пространств $(\ker\omega)$ дифференциальной формы ω. (В старых обозначениях это сводится к тому, что $\delta Q = 0$ вдоль γ, т. е. нет теплообмена с внешней средой.)

Тогда

$$\int_\gamma \omega = \int_\gamma dE + \int_\gamma \sum_{i=1}^n A_i\, da_i = E_1 - E_0 + \int_\gamma \sum_{i=1}^n A_i\, da_i = 0, \tag{3}$$

что является законом сохранения механической энергии, который подразумевает первое начало термодинамики.

Завершая определение математической модели термодинамической системы как дифференциальной формы (2), мы теперь трансформируем первое начало термодинамики в требование, чтобы *форма работы* $\sum\limits_{i=1}^n A_i\, da_i$ *при постоянном значении температуры* τ *была замкнута*. (В силу леммы Пуанкаре она тогда и точна в любой односвязной области параметров a. Значит, замкнутый рабочий цикл любой механической машины, работающей в такой области параметров при одной и той же температуре, не может стать источником механической энергии — вечным двигателем.) Итак, указана простейшая математическая модель термодинамической системы.

Дифференцируя форму (2), с учётом сказанного находим

$$d\omega = d\sum_{i=1}^n A_i\, da_i = \sum_{i=1}^n \frac{\partial A_i}{\partial \tau}\, d\tau \wedge da_i. \tag{4}$$

Если вспомнить, что $\omega = \tau\, dS$, то из формулы (4) при $1 \leqslant i \leqslant n$ получаем

$$\frac{\partial A_i}{\partial \tau} = \frac{\partial S}{\partial a_i}. \tag{5}$$

2.3. Термодинамика без тепла

С формально-логической точки зрения после того, как мы в школе узнали, что никакого особого теплорода или переливающейся тепловой жидкости не существует, что тепло — это энергия, что имеется механический эквивалент теплоты, позволяющий забыть о калориях, и т. д., после всего этого естественно строить формальную теорию без избыточных, зависимых понятий, каким в фундаментальном равенстве (1) оказалось тепло, фигурирующее в его левой части.

Поэтому, развивая последовательный формализм классической термодинамики, как это на первый взгляд ни парадоксально, отказываются от самого понятия теплоты (см., например, [11a]).

Именно правая часть соотношения (1) и только она определила исходную в математических построениях дифференциальную форму ω в формуле (2). В формуле (2), в отличие от соотношения (1), стоит не равенство, а определение формы ω. Эта форма выражена в терминах энергии системы и её работы, т. е. только в терминах энергии.

Интеграл этой формы вдоль пути теперь уже по определению можно назвать *потоком тепла* или *теплом*, полученным системой из внешней среды при указанном изменении состояния термодинамической системы.

2.4. Адиабатические переходы и аксиома Каратеодори

Особую роль в этих построениях имеют *адиабатические процессы*, происходящие без теплообмена с внешней средой. Это, как мы уже отмечали при выводе равенств (3), равносильно тому, что путь γ в пространстве состояний идёт, касаясь ядра $\ker\omega$ дифференциальной формы ω в текущей точке. Иными словами, вектор скорости движения всегда находится в соответствующем пространстве $\ker\omega$.

Вспоминая, что $\omega = \tau\, dS$, можно заключить, что равновесные адиабатические процессы происходят при неизменном значении энтропии. Путь γ лежит на поверхности уровня функции S. Более того, само распределение ядер $\ker\omega$ касательно к поверхностям уровня энтропии.

Итак, через равновесные состояния невозможно адиабатически перейти из одного термодинамического состояния системы в другое, если эти состояния имеют разные значения энтропии.

В своей аксиоматике классической термодинамики Каратеодори [11a] принял это обстоятельство в качестве исходного эквивалента второго начала термодинамики. А именно, введя форму (3), вместо приведённых выше математических требований он принимает более знакомое физикам положение: *в любой окрестности равновес-*

ного термодинамического состояния системы имеются равновесные состояния, в которые нельзя перейти адиабатическим процессом[1].

После этого начинается математика, которая приводит к тому, что было изложено выше. Эта математика имеет геометрический аромат и интересна сама по себе теми вопросами, которые она ставит и решает. Об этом будет сказано в следующей главе.

А здесь мы добавим только следующее, что известно физикам, но может быть неизвестно кому-то из математиков (не Каратеодори). Медленные равновесные адиабатические переходы между термодинамическими состояниями действительно сохраняют энтропию и обратимы. Но без теплообмена могут совершаться и неравновесные переходы между равновесными состояниями. Например, пусть газ находится в половине теплоизолированного сосуда (термоса), разгороженного перегородкой. Вы удаляете перегородку. Газ заполняет весь сосуд. Через какое-то время он, успокаиваясь, приходит в своё новое равновесное состояние. Можно показать, что энтропия этого состояния больше энтропии исходного равновесного состояния газа, когда он занимал только половину объёма сосуда.

Подсчитаем энтропию идеального одноатомного газа. Напомним известный из школы закон Клапейрона (Клапейрона—Менделеева)

$$P = \frac{n}{N}\frac{RT}{V} = nk\frac{T}{V},$$

где P — давление, V — объём, T — абсолютная температура газа; R — газовая постоянная, N — число Авогадро, n — количество молекул газа, $k = \frac{R}{N}$ — постоянная Больцмана.

Напомним также, что средняя кинетическая энергия молекулы газа равна $\frac{3}{2}kT$, поэтому внутренняя энергия E всего газа равна $\frac{3}{2}nkT$ и $T = \left(\frac{3}{2}nk\right)^{-1}E$.

[1]Это положение почти в той же форме и с анализом некоторых его следствий присутствует уже в лекциях Пуанкаре [4] (раздел 193). Таким образом, это положение, выделенное в качестве исходного, можно было бы называть также аксиомой Каратеодори—Пуанкаре или, в хронологическом порядке, аксиомой Пуанкаре—Каратеодори.

Нарушение этого положения непосредственно ведёт к построению вечного двигателя второго рода. Например, если взять два состояния на одной изотерме, то адиабатический переход из состояния с большей энтропией в состояние с меньшей энтропией невозможен.

Теперь, учитывая, что $\delta Q = dE + P\,dV$, а $dS = \frac{\delta Q}{T}$, находим

$$dS = \left(\frac{3}{2}nk\right)\frac{dE}{E} + nk\frac{dV}{V},$$

откуда с точностью до аддитивной постоянной получаем

$$S = nk\ln\left(VE^{\frac{3}{2}}\right) = k\ln\left(V^n E^{\frac{3}{2}n}\right).$$

В частности, если при сохранении внутренней энергии газа ему предоставить возможность занять вдвое больший объём, то энтропия нового состояния газа увеличится на $nk\ln 2$.

Интересно, что без теплообмена вернуться обратно уже не получится даже с нарушением равновесности промежуточных состояний.

Таким образом, аксиому Каратеодори можно было бы отнести и к неравновесным переходам, совершаемым без обмена теплом. Кстати, без них, по-видимому, даже невозможно достаточно полное адекватное описание равновесной термодинамики (см. например, [11a]).

Глава II

Термодинамика и контактная геометрия

§1. Контактные распределения

1.1. Адиабатический процесс и контактное распределение

Приведённая выше аксиома Каратеодори констатирует физическую закономерность, касающуюся адиабатических переходов между термодинамическими состояниями.

Дадим её полную математическую формулировку и проанализируем, к каким математическим вопросам и последствиям она приводит.

У нас имеется дифференциальная 1-форма ω, определённая формулой (1) главы I.

В каждом касательном пространстве к пространству состояний термодинамической системы имеется гиперплоскость $\ker \omega$, на векторах которой форма ω равна нулю. Эти касательные гиперплоскости $\ker \omega$, ядра формы ω, распределены по всему пространству M состояний. Обозначим это *распределение* символом H. Возникает пара (M, H) — пространство с распределением касательных гиперплоскостей. Путь γ в M будем считать *допустимым*, если он в любой своей точке касается соответствующей плоскости распределения H.

Вопрос: всегда ли можно соединить две точки пространства M допустимым путём? (Считаем, что M — гладкое, связное многообразие.)

Аксиома Каратеодори утверждает, что отвечающая термодинамике форма ω такова, что в окрестности любой точки пространства M есть точки, в которые из нашей точки нельзя попасть допустимым путём.

1.2. Формализация

Забывая теперь о термодинамике, обратимся к возникшим чисто математическим понятиям, структурам, объектам и вопросам.

Пусть M — гладкое, связное n-мерное многообразие и H — распределение касательных гиперплоскостей (размерности $n-1$) на нём.

Плоскости распределения H часто называют *горизонтальными*, а идущие вдоль них *допустимые* пути часто также называют *горизонтальными* или *интегральными*.

Например, если на M транзитивно и свободно (т. е. без неподвижных точек) действует некоторая группа диффеоморфизмов, то касательная гиперплоскость в одной точке, разнесённая по M преобразованиями группы, породит распределение H на M.

Нигде не вырождающаяся в нулевую дифференциальная 1-форма ω на M, как мы видели, порождает распределение $H = \{\ker \omega\}$ на M.

§ 2. Интегрируемость распределений

2.1. Теорема Фробениуса

Распределение H на M называется *интегрируемым*, если у него есть интегральная поверхность, т. е. такое подмногообразие, что в каждой его точке соответствующая плоскость распределения H является касательной плоскостью подмногообразия.

Например, если бы распределение H было одномерным, точнее задавалось векторным полем, то по основной теореме существования решений дифференциальных уравнений оно всегда было бы интегрируемым (через каждую точку проходит интегральная кривая).

В общем случае это, как известно, не так. Пример: $H = \{\ker \omega\}$, где $\omega = y\,dx + dz$ в $\mathbb{R}^3$ (обходу замкнутого контура в плоскости координат x, y отвечает незамкнутая интегральная кривая в $\mathbb{R}^3$ — «виток спирали» с шагом, равным площади, охваченной контуром.) Порождаемое невырожденной гладкой дифференциальной 1-формой ω распределение касательных гиперплоскостей $H = \{\ker \omega\}$ интегрируемо, лишь когда $d\omega = 0$ на соответствующих плоскостях $\ker \omega$.

Это условие равносильно следующему условию Фробениуса интегрируемости распределения $H = \{\ker \omega\}$ в $\mathbb{R}^3$:

$$\omega \wedge d\omega = 0.$$

Для распределения k-мерных плоскостей, заданного на n-мерном многообразии в виде общих нулей $n - k$ всюду независимых гладких дифференциальных 1-форм $\omega_1, \ldots, \omega_{n-k}$, т. е. когда плоскости этого распределения H получены пересечением гиперплоскостей $\ker \omega_1, \ldots, \ker \omega_{n-k}$, критерий интегрируемости распределения H состоит в том, что $d\omega_i = 0$ на плоскостях распределения H для каждой из форм ω_i. Это равносильно выполнению следующих условий (см. например, [4]):

$$\omega_1 \wedge \ldots \wedge \omega_{n-k} \wedge d\omega_i = 0, \quad 1 \leqslant i \leqslant n-k.$$

2.2. Интегрируемость, связность, управляемость

Теперь вернёмся к вопросу соединимости точек пространства допустимыми путями. Такой вопрос возникает и в теории управления, где слово «соединимость» заменяется термином «управляемость».

У нас этот вопрос возник в связи с термодинамикой, адиабатическими процессами и аксиомой Каратеодори.

Этот вопрос непосредственно связан с вопросом интегрируемости соответствующего распределения H гиперплоскостей. А именно, интегрируемость и соединимость взаимно дополнительны: интегрируемость — несоединимость, соединимость — неинтегрируемость.

В самом деле, если распределение H гиперплоскостей интегрируемо, то любой контактный к нему допустимый путь не может покинуть интегральную поверхность распределения. Значит, в окрестности любой точки есть точки многообразия, недостижимые из неё допустимым путём.

Каратеодори доказал, что для распределения гиперплоскостей локально справедливо и обратное утверждение: несоединимость — интегрируемость [11].

(После этого доказывается, что форма ω, задающая распределение $\ker \omega$, локально допускает интегрирующий множитель, что $\tau^{-1}\omega = dS$, что S — энтропия и выполнено второе начало термодинамики.)

Геометров, интересовавшихся физикой, таких как П. К. Рашевский, математическое содержание работы Каратеодори, по-видимому, стимулировало к общему исследованию этого полезного геометрического вопроса о соединимости точек пространства путями, допустимыми для заданного распределения.

Первый общий ответ в терминах скобок Ли был получен П. К. Рашевским в работе [5], но называется теоремой Чоу, см. [6, 7]. Люди знающие говорят «теорема Рашевского—Чоу». Исследование вопроса, его расширение, подключение сюда аспектов теории управления продолжается по сей день, см. например, [7, 8].

В терминах скобок Ли условия интегрируемости и соединимости формулируются следующим образом (см. например, [9—11]).

Пусть $e_1, \ldots, e_k$ — гладкие линейно независимые в каждой точке векторные поля на многообразии M, порождающие распределение H семейства k-мерных касательных плоскостей.

Для интегрируемости распределения H необходимо и достаточно, чтобы скобки Ли $[e_i, e_j]$ порождающих распределение полей не выходили за плоскости распределения.

Для того чтобы локально (а на связном многообразии и глобально) любые точки многообразия были соединимы допустимым для распределения H путём, достаточно, чтобы итерации $[[e_i, e_j]\ldots]$ скобок исходных полей $e_1, \ldots, e_k$ порождали базис всего касательного пространства к многообразию в каждой точке многообразия.

Эти формулировки и условия Фробениуса связывают известные в исчислении дифференциальных форм соотношения. (Вообще, язык дифференциальных форм дуален языку векторных полей.) Если X и Y — гладкие векторные поля, а ω — 1-форма на многообразии, то

$$d\omega(X, Y) = X\omega(Y) - Y\omega(X) - \omega([X, Y]),$$

где $X\omega(Y)$ и $Y\omega(X)$ — производные Ли функций $\omega(Y)$ и $\omega(X)$ вдоль полей X и Y соответственно, а $[X, Y]$ — скобка (коммутатор) этих векторных полей.

В общем случае формы ω порядка m мы имеем

$$d\omega(\xi_1, \ldots, \xi_{m+1}) = \sum_{i=1}^{m+1} (-1)^{i+1} \xi_i \omega(\xi_1, \ldots, \widehat{\xi_i}, \ldots, \xi_{m+1}) + \\ + \sum_{1 \leqslant i < j \leqslant m+1} (-1)^{i+j} \omega([\xi_i, \xi_j], \xi_1, \ldots, \widehat{\xi_i}, \ldots, \widehat{\xi_j}, \ldots, \xi_{m+1}),$$

где символ $\frown$ отмечает выпускаемый член, $[\xi_i, \xi_j]$ — скобка Ли полей ξ_i, ξ_j, а $\xi_i\omega$ — производная функции $\omega(\xi_1, \ldots, \widehat{\xi_j}, \ldots, \xi_{m+1})$ по полю ξ_i.

2.3. Метрика Карно—Каратеодори

Пусть M — риманово многообразие с метрикой g. Пусть распределение H в M таково, что любую пару точек многообразия M можно соединить допустимой для H кривой.

Тогда в M (а точнее, в (M, g, H)) можно ввести новую метрику, в которой расстояние между точками измеряется нижней гранью длин допустимых кривых, соединяющих эти точки.

Эта метрика (по-видимому, с лёгкой руки М. Громова) называется *метрикой Карно—Каратеодори* или CC-*метрикой*.

Она возникает и полезна в различных ситуациях, где появляются тройки вида (M, g, H). Например, в комплексном пространстве комплексные касательные к нетривиальным гиперповерхностям (например, к гиперсфере) образуют неинтегрируемые контактные распределения.

Полезно напомнить, что в силу теоремы Дарбу любая *контактная форма* (т. е. такая гладкая форма ω, что форма $d\omega$ невырожденна на $\ker\omega$) гладкой заменой координат локально приводится к виду $x_1\,dy_1 + \ldots + x_n\,dy_n + dz$. Например, в случае $\mathbb{R}^3$ получаем форму $x\,dy + dz$.

Заметим, что здесь движения по прямым, параллельным оси x, допустимы. В плоскостях $x = c$ допустимы движения вдоль наклонных прямых. Наклон линейно зависит от c. Этого достаточно, чтобы стала очевидной возможность соединить любые две точки пространства $\mathbb{R}^3$ путём, допустимым для распределения

$$H = \{\ker(x\,dy + dz)\}.$$

Конечно, неинтегрируемость распределения, порождённого контактной формой $x\,dy + dz$, можно проверить, сославшись на условие Фробениуса.

О контактных структурах см., например, [10—13].

2.4. Контактная форма Гиббса

Мы уже знаем фундаментальное, отвечающее второму началу термодинамики, соотношение $\omega = \tau\,dS$ для формы (2) главы I. Здесь ω имеет смысл формы теплообмена, τ — абсолютной температуры, а S — энтропии.

Рассмотрим вслед за Гиббсом форму

$$\Omega = -\tau\,dS + \sum_{i=1}^{n} A_i\,da_i + dE,$$

или, конкретнее, как и Гиббс, контактную форму

$$\Omega = -T\,dS + P\,dV + dE,$$

с привычными для классической термодинамики параметрами: T — абсолютная температура, S — энтропия, P — давление, V — объём и E — внутренняя энергия газа, жидкости или иной термодинамической системы.

Основное, выражающее энергетический баланс, соотношение (1) главы I, где теперь $\delta Q = T\,dS$, говорит, что равновесные термодинамические процессы (уже не обязательно адиабатические) происходят так, что на них $\Omega = 0$, т. е. они идут вдоль распределения $\{\ker \Omega\}$ в пространстве $\mathbb{R}^5$ параметров (T, S, P, V, E). Но это не пространство M состояний (τ, a), или (T, V), или, например, (S, V).

Свободными на самом деле здесь остаются только два параметра, и реально равновесные процессы протекают на *лежандровой поверхности* (интегральном подмногообразии M максимальной размерности) *контактной структуры*, определяемой формой Ω в пространстве $\mathbb{R}^5$. Это один из основных геометрических тезисов Гиббса.

В случае общей формы

$$\Omega = -\tau\,dS + \sum_{i=1}^{n} A_i\,da_i + dE$$

мы имеем $2n+3$ переменные величины $(E, S, A_1, \ldots, A_n, \tau, a_1, \ldots, a_n)$, но состояние системы определяют только параметры $(\tau, a_1, \ldots, a_n)$ (или эквивалентный им набор из $n+1$ независимых величин). Так что в общем случае равновесные термодинамические процессы идут вдоль лежандрова $(n+1)$-мерного многообразия (интегрального многообразия максимальной размерности) распределения гиперповерхностей $H = \{\ker \Omega\}$, которое в пространстве $\mathbb{R}^{2n+3}$ порождает (контактная) форма Ω.

2.5. Заключительные замечания

Завершая этот параграф, отметим, что мы коснулись здесь только одного конкретного геометрического вопроса, связанного с классической термодинамикой, правда, поговорили и о том, как на современном математическом языке формулируются фундаментальные начала термодинамики.

Существуют и другие направления современных математических исследований, порождённые классической термодинамикой, например попытка формализовать (аксиоматизировать) понятие энтропии [14] или систематизация и обобщение процедур, связывающих основные термодинамические функции (потенциалы) [15].

Всё же основной, физически содержательный математический пласт классической термодинамики, поднятый Гиббсом [1а], [1б], [1в], связан с контактной геометрией. Представление об этом можно получить, просмотрев том трудов конференции, посвящённой 150-летию со дня рождения Гиббса [2]. Она проходила в том самом Йельском университете, где Гиббс и работал.

Процитирую (в моем русском переводе) только следующую фразу из статьи В. И. Арнольда [3], помещённой в этом сборнике: «Я думаю, Гиббс был первым, кто понял значение контактной структуры для физики и термодинамики».

Фундаментальное термодинамическое равенство

$$dE = T\,dS - P\,dV$$

к 1873 г., когда появился труд Гиббса [1а], без сомнения, уже было стандартной основой при построении термодинамики. Но даже Клаузиус, сформулировавший в 1850 г. второе начало термодинамики, концептуально определивший энтропию в 1854 г. и давший ей имя в 1865 г., никогда не отводил ей центральной роли в изложении термодинамики. Для Клаузиуса и его современников термодинамика сводилась к изучению взаимосвязи теплоты и работы.

Гиббс первым исключил тепло и работу из оснований термодинамики, отдав предпочтение функциям термодинамического состояния: энергии и энтропии.

Термодинамика стала теорией свойств вещества, находящегося в состоянии термодинамического равновесия. Гиббс, в частности, показал, что для равновесия двух фаз 1 и 2 одной и той же субстанции мало равенства температур и давлений. Ещё нужно, чтобы энергии, энтропии и объёмы фаз (отнесённые к единице массы) удовлетворяли соотношению: $E_2 - E_1 = T(S_2 - S_1) - P(V_2 - V_1)$, причём равновесное состояние изолированной термодинамической системы устойчиво, если и только если оно соответствует локальному максимуму энтропии.

Гиббс стремился найти общий графический (геометрический) метод, который бы мог в целом описывать термодинамические свойства среды (во всех её фазах) при обратимых процессах.

В 1873 г. Гиббс ввёл свою знаменитую фундаментальную поверхность, связывающую энергию E термодинамической системы с её энтропией S и объёмом V: $E = E(S, V)$. Он показал, как природа её касательных плоскостей связана с сосуществованием фаз и критических точек.

Эта работа восхитила Максвелла и стимулировала его к построению известной модели Максвелла этой поверхности.

Гиббс показал, что $E(S, V)$ является порождающей (характеристической) функцией, знание которой определяет все термодинамические свойства системы.

Отмечали, что, подобно тому как «Аналитическая механика» Лагранжа была венцом науки восемнадцатого столетия, мемуар Гиббса 1876, 1878 годов «О равновесии гетерогенных сред» [1в] в значительной степени был венцом классического естествознания девятнадцатого столетия.

Глава III

Термодинамика классическая и статистическая

§1. Кинетические теории

Исходная идея статистической физики — получить термодинамику из механики.

«Австрийскому физику Больцману принадлежит честь первого успешного подхода к этой задаче и установления связи между понятием вероятности, определённым образом понимаемой, и термодинамическими функциями, в частности энтропией. Рядом с ним нужно считать одним из основателей этой новой ветви теоретической физики — статистической термодинамики — Уилларда Гиббса. Далее следует упомянуть работы Пуанкаре, Планка и Эйнштейна» (Лоренц, 1915 г., [5]).

1.1. Молекулы и давление

Зарождение кинетической теории газов восходит к Даниилу Бернулли. В его книге «Гидродинамика» (Страсбург, 1738 г.) уже выведено давление газа из изменения импульса молекул газа, сталкивающихся со стенкой сосуда. В 1746 г. Даниил и Иоганн Бернулли даже были удостоены премии Парижской академии наук за работу на эту тему.

В 1856 г. Август Крёниг связал температуру идеального газа со средней кинетической энергией его молекул.

Дальнейшие продвижения в кинетической теории и статистической термодинамике относятся уже ко второй половине XIX — началу XX века и связаны с именами Клаузиуса, Максвелла, Больцмана, Гиббса, Пуанкаре, Эйнштейна.

1.2. Распределение Максвелла

Наивно можно было бы предполагать, что молекулы газа, находящегося в состоянии термодинамического равновесия, все имеют приблизительно одну и ту же скорость (кинетическую энергию) в результате многочисленных соударений и перераспределения импульсов. Тем более впечатляющим было открытие Максвелла (1866 г.), простыми соображениями показавшего, что проекции скоростей молекул идеального газа на любое направление распределены нормально. Соответственно возникает классическое распределение Максвелла молекул по их кинетической энергии.

(Геометрический смысл распределения Максвелла как нормального распределения в отношении скоростей молекул идеального газа кратко отмечен на с. 54.)

1.3. Энтропия по Больцману

Больцман (1868—1871 гг.) развил результат Максвелла и показал, что равновесное распределение частиц идеального газа по энергиям во внешнем силовом поле (например, в поле тяготения) определяется функцией распределения $\exp(-E/kT)$, где E — сумма кинетической и потенциальной энергий частицы, T — абсолютная температура, k — постоянная Больцмана.

Открытое Больцманом в 1872 г. соотношение $S = k \log \Pi$ между энтропией S и в определённом смысле понимаемой вероятностью термодинамического состояния[1] называют принципом Больцмана или формулой Больцмана. Оно даёт статистическую интерпретацию второго начала термодинамики, сводящуюся в конце концов к тому, что термодинамические процессы имеют тенденцию переводить систему из менее вероятного состояния в более вероятное, т. е. в направлении возрастания энтропии. Равновесному состоянию отвечает условный максимум функции S.

Основная заслуга Больцмана состоит в последовательной реализации идеи интерпретировать классическую термодинамику как статистическую механику, см. [1].

Он вывел основное кинетическое уравнение для газа, определяющее при некоторых предположениях его эволюцию как механической системы, и получил свою знаменитую H-теорему о том, что эволюция происходит в сторону возрастания энтропии.

[1] Здесь Π — число микросостояний системы, например состояний молекул газа в данной ёмкости, соответствующих макросостоянию, энтропия которого равна S. Современное значение постоянной Больцмана $k = (1{,}38044 \pm 0{,}00007) \cdot 10^{-23}$ Дж/К.

Больцман указал и широко известное ныне выражение $H = -\int \varrho \log \varrho$ энтропии через плотность распределения, которое вместе с обозначением H перекочевало и в другие сферы, став, например, фундаментальным понятием современной теории передачи информации. (Больцман, в отличие от Клаузиуса, кажется, сначала обозначал энтропию буквой E, отданной ныне энергии. Мы, как теперь принято, обозначаем энтропию через S, оставляя символ H за гамильтонианом гамильтоновой механической системы.)

1.4. Ансамбли Гиббса и термодинамизация механики

Гиббс предложил наиболее общую и стройную математическую основу статистической механики и термодинамики.

Не склонный к восторгам Зоммерфельд называет Гиббса «великим термодинамистом и статистиком». Его первые работы, затерявшиеся в Трудах Академии штата Коннектикут от 1876 и 1878 гг., стали общеизвестны после того, как в 1902 г. Оствальд издал их на немецком языке под названием «Термодинамические этюды» [17, с. 67].

За год до смерти в 1902 г. Гиббс опубликовал свою «Статистическую механику», послужившую математическим фундаментом последующим поколениям, см. [2].

Как мы уже отмечали, Гиббс объединил идеи термодинамики и гамильтоновой механики. Он ввёл замечательную математическую структуру статистической физики — гамильтонову систему, наделённую вероятностной мерой, эволюционирующей под действием гамильтонова потока в фазовом пространстве системы.

Эта модель стала источником многочисленных задач и исследований теории динамических систем, активно ведущихся и сейчас.

Напомним вкратце, что в фазовом пространстве Γ гамильтоновой системы Гиббс вводит называемое теперь *каноническим* распределение вероятностей состояний. Плотность ϱ этого распределения определяется гамильтонианом (энергией) $H = H(q, p)$ системы

$$\varrho := c \exp(-\beta H),$$

где $c = \left(\int_\Gamma \exp(-\beta H)\, dq\, dp \right)^{-1}$ — нормирующий множитель, а в физических системах $\beta = 1/(k\tau)$, k — постоянная Больцмана, τ — абсолютная температура.

(Подробному обсуждению и обоснованию канонического распределения посвящена значительная часть книги [12].)

Каноническое распределение инвариантно относительно действия гамильтонова потока, поскольку инвариантны мера и гамильтониан.

Гамильтониан может зависеть от параметров $a := (a_1, \ldots, a_n)$, т. е. $H = H(q, p, a)$. Это могут быть параметры внешних воздействий типа перемещений поршня и изменения объёма газа.

Гиббс, в частности, указал следующий красивый естественный процесс термодинамизации гамильтоновой механической системы.

Рассмотрим среднюю энергию

$$E(\beta, a) = \int_\Gamma H \varrho \, dq \, dp$$

и осреднённые силы реакции связей, отвечающие параметрам a_i,

$$A_i = -\int_\Gamma \frac{\partial H}{\partial a_i} \varrho \, dq \, dp.$$

Последние соотношения будем интерпретировать как уравнения состояния $A_i = f_i(\beta, a)$, $i = 1, \ldots, n$.

Нетрудно проверить, что 1-форма

$$\omega = dE + \sum_{i=1}^{n} A_i \, da_i$$

удовлетворяет аксиомам двух начал термодинамики (см. с. 124).

1.5. Эргодичность

Исходная идея статистической физики, как уже упоминалось выше, — получить термодинамику из механики. Например, рассматривая газ в сосуде как механическую систему, состоящую из огромного числа слабо взаимодействующих хаотически движущихся частиц, можно найти его внутреннюю энергию, связать среднюю кинетическую энергию частиц с температурой газа, найти давление и энтропию.

Функции классической термодинамики при этом получаются как статистически средние величины, возникающие в результате суммирования огромного числа мелких вкладов отдельных частиц этой многопараметрической механической системы. Тут начинают работать вероятностные идеи, законы больших чисел, принципы концентрации. Это новая идеология. В ней второе начало термодинамики (и многое другое) уже не рассматривается как абсолютная истина, в силу которой сама по себе система никогда не переходит в состояние с меньшей энтропией или тепло никогда не может само по себе перейти от холодного тела к горячему. Теперь все молекулы в принципе могут собраться в углу комнаты, но вероятность этого события оказывается ничтожной, как и время его возможного существования. Мы практически не в состоянии наблюдать это событие.

Возникает много фундаментальных вопросов. Например, что, собственно, и как мы наблюдаем и измеряем?

Когда мы хотим представить себе все возможные состояния данной системы, можно либо следить за её эволюцией во времени, либо представить себе целый *ансамбль* систем, которые суть копии нашей системы, взятые в один и тот же момент времени и представляющие все возможные состояния нашей системы.

С позиций механики такой ансамбль просто есть фазовое пространство системы, в котором каждая индивидуальная точка представляет систему в конкретном состоянии. Эволюция системы изобразится траекторией точки в фазовом пространстве. Движение точек порождает движение всего фазового пространства, удачно названное *фазовым потоком*.

Классическая теорема Лиувилля говорит, что *фазовый поток гамильтоновой системы сохраняет фазовый объём*, т. е. течёт как несжимаемая жидкость.

Естественно поэтому считать, что вероятность найти систему в одном из состояний, отнесённых к какой-то области фазового пространства, доступной движению, пропорциональна объёму этой области.

С другой стороны, можно проследить всю эволюцию индивидуальной системы и вероятность состояния считать пропорциональной времени, в течение которого оно существует.

Больцман в этой связи сначала даже считал, что система в своей эволюции пробегает все возможные состояния на поверхности уровня данной энергии и что поэтому оба подхода к измерениям средних и статистическому описанию состояний системы эквивалентны. Но это бы означало, что фазовая траектория проходит через все точки указанной поверхности, что неверно. Тем не менее сам принципиальный вопрос об измерении средних остаётся. Он приобрёл несколько формулировок, объединённых термином *эргодическая гипотеза Больцмана*, и привёл к ряду замечательных математических теорем, по сей день украшающих теорию динамических систем.

Напомним, например, эргодическую теорему Биркгофа (1931).

Пусть V — инвариантная часть фазового пространства относительно фазового потока, сохраняющего меру μ, и f — μ-измеримая функция в V. Если мера части V конечна, то для μ-почти всех точек $p \in V$ существует временнóе среднее вдоль траекторий; оно является μ-измеримой функцией $\widetilde{f}$, среднее значение $\widetilde{A}$ которой в V совпадает со средним значением A самой функции f в V.

Если, кроме того, V метрически неразложимо относительно этого фазового потока, то функция $\widetilde{f}$ постоянна μ-почти всюду

(т. е. тогда средние по траекториям для μ-почти всех точек $p \in V$ одинаковы и совпадают с пространственным средним A функции f в V).

(Говорят, что пространство X с мерой μ, инвариантной относительно действующего на X потока, *метрически неразложимо* относительно потока, а поток *эргодичен* в X, если X не представляется как объединение непересекающихся множеств X_1, X_2, инвариантных относительно потока и имеющих положительные меры.)

Следует заметить, что для целей термодинамики важны не столько индивидуальные эргодические теоремы о динамических системах на фиксированных пространствах конечной размерности, сколько, быть может, даже более грубые асимптотические факты о средних, связанные с предельным переходом (термодинамическим пределом), когда количество частиц и соответственно размерность фазового пространства неограниченно возрастают. В этой связи имеет смысл внимательнее проанализировать возможные интерпретации, проявления, применения и развитие рассмотренного в теме II принципа концентрации.

1.6. Парадоксы, проблемы, трудности

Обычно даже очень красивая теория — это упрощённая модель явления, хорошо работающая только в каких-то пределах, которые, как правило, мы узнаём, выйдя на границы дозволенного и столкнувшись с новыми проблемами.

Для статистической теории, сводящей термодинамику к механике и предрекающей вслед за вторым началом термодинамики эволюцию замкнутых систем в сторону роста энтропии, первым серьёзным пробным камнем был вопрос, поставленный учителем Больцмана и называемый теперь *парадоксом Лошмидта* (1876 г.). Он формулируется предельно просто и состоит в следующем.

Любая гамильтонова механическая система допускает обращение времени. Как же такая система может приводить к избранному направлению эволюции? Ведь если поменять направление времени, то изменится и направление эволюции?!

Вопрос этот обсуждался лучшими умами своего времени (см. например, книги [1, 4] и классический обзор [6]).

Масло в огонь добавил Цермело. (*Парадокс Цермело* 1896 г. Того самого Эрнеста Цермело, который математикам больше известен по наделавшей шуму аксиоме Цермело.) Он, апеллируя к теореме Пуанкаре о возвращении, сказал, что молекулы не только могут, но и непременно должны будут в какой-то момент все собраться в углу комнаты.

И вообще, любая окрестность любой точки фазового пространства гамильтоновой механической системы, блуждающая под действием фазового потока в ограниченной области пространства, бесконечное число раз будет возвращаться, пересекаясь со своим исходным положением. (Это столь же простое, сколь и важное наблюдение Пуанкаре 1883 г. почти очевидно, если вспомнить теорему Лиувилля о том, что фазовый поток течёт как несжимаемая жидкость, сохраняя фазовый объём. Если бы все образы исходной окрестности, рассматриваемые через равные интервалы времени, попарно не пересекались, то их общий объём был бы бесконечным. Если же есть пересечение m-го и n-го образов и $m < n$, то есть пересечение $(n - m)$-го образа окрестности с ней самой.)

Надо сказать, что «механическая теория теплоты прожила трудную жизнь и окончательно приобрела права гражданства в науке лишь к концу XIX столетия, прежде всего благодаря работам Макса Планка 1887—1892 гг.» [20, с. 236].

Отметим ещё, что классики, как правило, лучше других понимали и достоинства, и недостатки своих теорий, на что сами же чаще всего и указывали. Приведём, например, слова Каратеодори о классической термодинамике. (Слова Каратеодори, написавшего работу [18, с. 236], которую, кстати, высоко оценил Планк, стимулировавший последующую публикацию работы [19].)

«Можно поставить вопрос: каким образом должна быть построена феноменологическая термодинамика, чтобы при расчётах использовать лишь непосредственно измеряемые величины, т. е. объём, давление и химический состав тела? Теория, которая при этом возникает, логически неоспорима и математически совершенна, так как она, исходя из реально наблюдаемых фактов, обходится минимальным количеством гипотез. Однако именно это преимущество делает её мало пригодной с точки зрения исследователя, не только потому, что в ней температура появляется в качестве производной величины, но прежде всего потому, что гладкие стены искусственно воздвигнутых в этой теории зданий не позволяют установить какой-либо связи между миром видимой и осязаемой материи и миром атома».

Вот понимание необходимости и важности статистической термодинамики.

Удивительно, что первым реальным подтверждением атомизма и молекулярного строения вещества, как считается, оказалось броуновское движение, открытое английским ботаником Робертом Броуном (Броун, Браун, Brown) всего-то в 1828 г. Математические основы теории броуновского движения с важными для физики выводами были заложены Эйнштейном в знаменитом 1905 году его интеллектуального взрыва работой [3а] и последовавшей за ней работой [3б], которая уже так и называлась «К теории броуновского движения».

Во многих отношениях интересны и показательны следующие вводные строки работы [3а].

«В этой работе будет показано, что согласно молекулярно-кинетической теории теплоты взвешенные в жидкости тела микроскопических размеров вследствие молекулярного теплового движения должны совершать движения такой величины, что легко могут быть обнаружены под микроскопом. Возможно, что рассматриваемые движения тождественны с так называемым броуновским движением; однако доступные мне данные относительно последнего настолько неточны, что я не мог составить об этом определённого мнения.

Если рассматриваемое движение вместе с ожидаемыми закономерностями действительно будет наблюдаться, то классическая термодинамика не может считаться вполне справедливой уже для микроскопически различимых областей, и тогда возможно точное определение истинных атомных размеров. Если же, наоборот, предсказание этого движения не оправдается, то это будет веским аргументом против молекулярно-кинетического представления о теплоте».

([Статистическое понимание второго начала термодинамики чувствовали и отстаивали Клаузиус, Максвелл, Больцман, Гиббс. Но их объяснения основывались на мысленных экспериментах, исходивших из постулата существования молекул. Лишь после открытия броуновского движения интерпретация второго начала термодинамики как абсолютного закона становится невозможной. Броуновские частицы, поднимаясь и опускаясь за счёт теплового движения молекул, воочию демонстрируют нам вечный двигатель второго рода. Поэтому в конце XIX в. исследование броуновского движения приобрело огромное теоретическое значение и привлекло внимание многих физиков-теоретиков, включая Эйнштейна.] Так сказано в книге [20][1].)

Работы Эйнштейна по броуновскому движению потом стали почвой для обширной математической теории.

Гиббс тоже не оставил следующие поколения без работы. Не говоря о математических проблемах, связанных с современным пониманием равновесных состояний (как инвариантных мер) и эволюции к ним, напомним лишь *парадокс Гиббса* из классической термодинамики, приводящий к неизбежности даже квантовой трактовки состояний.

В середине горизонтального цилиндра стоят две полупроницаемые подвижные перегородки A и B, в таком порядке. Левая половина цилиндра заполнена газом a, проникающим через мембрану A и не проникающим через B. Правая часть цилиндра заполнена газом b, проникающим через мембрану B и не проникающим через A. Если постепенно сдвигать перегородку A до упора налево, то газ a останется там, где был, а газ b распространится по всему цилиндру. Энтропия его состояния увеличится, поэтому увеличится и общая энтропия системы двух этих газов, равная сумме энтропий их состояний. Теперь проделаем то же самое с перегородкой B, медленно отодвинув её до упора направо. Газ a распространится на весь цилиндр. Энтропия системы ещё вырастет на ту же фиксированную величину ($k\ln 2$).

[1] Как говорил Бор, «не для спора, но для выяснения истины»: приведённая выше цитата из работы Эйнштейна показывает, что даже тут не всё было так просто.

Будем теперь повторять эксперимент, непрерывно сближая характеристики газов. В пределе получим один газ. Тогда начальное состояние и конечное состояние системы неразличимы. А их энтропии оказались разными?!

По-видимому, физические характеристики не непрерывны?! Например, это может относиться к энергии. Попробуем это учесть.

§ 2. Квантовая статистистическая термодинамика (два слова)

2.1. Счёт состояний и условный экстремум

Квантовые расчёты тут в некотором отношении даже прозрачнее классических. Проведём их, следуя Шрёдингеру [16].

Рассмотрим ансамбль из N одинаковых, но перенумерованных систем, каждая из которых может находиться в одном из перенумерованных состояний $1, 2, \ldots, l, \ldots$ Пусть $\varepsilon_1 \leqslant \varepsilon_2 \leqslant \ldots \leqslant \varepsilon_l \leqslant \ldots$ — значения энергии индивидуальной системы в этих состояниях и $a_1, a_2, \ldots$ $\ldots, a_l, \ldots$ — количество систем ансамбля, находящихся в состояниях $1, 2, \ldots, l, \ldots$ соответственно.

Такой набор $a_1, a_2, \ldots, a_l, \ldots$ может реализоваться многими способами. А точнее, число способов равно

$$P = \frac{N!}{a_1!\,a_2!\ldots a_l!\ldots}.$$

Совокупность чисел $a_1, a_2, \ldots, a_l, \ldots$ должна удовлетворять условиям

$$\sum_l a_l = N, \quad \sum_l \varepsilon_l a_l = E,$$

где E — полная совокупная энергия систем ансамбля.

Будем искать максимальное значение P при указанных ограничениях, что даст нам наиболее вероятный набор $a_1, a_2, \ldots, a_l, \ldots$ чисел заполнения.

(Поясним, что в интересных для термодинамики случаях, когда N очень велико, имеет место явление концентрации. Можно показать, что общее число всех возможных при наших условиях состояний ансамбля почти совпадает с максимальным значением P, которое мы намерены искать. Значит, мы действительно найдём и наиболее вероятный набор $a_1, a_2, \ldots, a_l, \ldots$ чисел заполнения.)

При больших значениях n по формуле Стирлинга $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$. Поэтому можно считать что $\log(n!) \approx n(\log n - 1)$. (Здесь $\log = \ln$.)

Тогда, действуя методом Лагранжа отыскания условного экстремума, напишем, что

$$\sum_l \log a_l\, da_l + \lambda \sum_l da_l + \mu \sum_l \varepsilon_l\, da_l = 0,$$

откуда следует, что при любом l выполняется равенство

$$\log a_l + \lambda + \mu\varepsilon_l = 0 \quad \text{и} \quad a_l = e^{-\lambda-\mu\varepsilon_l},$$

причём λ и μ подчинены условиям

$$\sum_l e^{-\lambda-\mu\varepsilon_l} = N, \quad \sum_l \varepsilon_l e^{-\lambda-\mu\varepsilon_l} = E.$$

Обозначая через $E/N = U$ среднюю энергию, приходящуюся на одну систему ансамбля, можем записать полученный результат в следующем виде:

$$\frac{E}{N} = U = \frac{\sum_l \varepsilon_l e^{-\mu\varepsilon_l}}{\sum_l e^{-\mu\varepsilon_l}} = -\frac{\partial}{\partial\mu}\log\sum_l e^{-\mu\varepsilon_l},$$

$$a_l = N\frac{e^{-\mu\varepsilon_l}}{\sum_l e^{-\mu\varepsilon_l}} = -\frac{N}{\mu}\frac{\partial}{\partial\varepsilon_l}\log\sum_l e^{-\mu\varepsilon_l}.$$

Дополнительные рассмотрения, выясняющие физический смысл величины μ в термодинамической ситуации, приводят к тому, что

$$\mu = \frac{1}{kT},$$

где k — постоянная Больцмана, а T — абсолютная температура. Возникает важная величина, называемая *статистической суммой*:

$$Z = \sum_l e^{-\frac{\varepsilon_l}{kT}}.$$

Теперь можно написать, как именно числа заполнения распределены по энергетическим уровням при данной абсолютной температуре:

$$a_l = N\frac{e^{-\frac{\varepsilon_l}{kT}}}{Z}.$$

Энтропия системы с точностью до аддитивной постоянной имеет вид

$$S = k\log Z + \frac{U}{T} = k\log\sum_l e^{-\frac{\varepsilon_l}{kT}} + \frac{1}{T}\frac{\sum_l \varepsilon_l e^{-\frac{\varepsilon_l}{kT}}}{\sum_l e^{-\frac{\varepsilon_l}{kT}}}.$$

Остальные термодинамические функции (потенциалы) тоже допускают аналогичные выражения.

Подсчитаем, ради интереса, как ведёт себя указанная энтропия при $T \to 0$. Пусть для общности нижнему уровню энергии отвечает n возможных состояний системы, т. е. $\varepsilon_1 = \varepsilon_2 = \ldots = \varepsilon_n$, и пусть следующие m уровней тоже одинаковы, т. е. $\varepsilon_{n+1} = \varepsilon_{n+2} = \ldots = \varepsilon_{n+m}$.

Тогда при $T \to +0$ мы имеем

$$S \sim k \log\left(ne^{-\frac{\varepsilon_1}{kT}} + me^{-\frac{\varepsilon_{n+1}}{kT}}\right) + \frac{1}{T}\frac{n\varepsilon_1 e^{-\frac{\varepsilon_1}{kT}} + m\varepsilon_{n+1}e^{-\frac{\varepsilon_{n+1}}{kT}}}{ne^{-\frac{\varepsilon_1}{kT}} + me^{-\frac{\varepsilon_{n+1}}{kT}}},$$

откуда следует, что $S \sim k \log n$ при $T \to +0$. В частности, в физически значимом случае, когда $n = 1$, предел точно равен нулю. Но даже если $n \neq 1$ и n не грандиозно, величина $k \log n$ почти равна нулю, ввиду малости постоянной Больцмана ($k \approx 1{,}38 \cdot 10^{-23}$ Дж/К).

2.2. Уточняющие замечания и дополнения

1. Сначала скажем два слова о только что проделанных вычислениях.

Результаты, которые мы сравнительно элементарно получили выше, основывались на утверждении, что достаточно найти условия максимума функции P, поскольку в малой окрестности максимума лежит подавляющая часть возможных состояний всего ансамбля наших систем, т. е. это и есть наиболее вероятные состояния ансамбля.

Значительно более тщательное, детальное исследование вопроса (но с тем же принципиальным результатом) было проделано в ставшей классической работе Дарвина и Фаулера. Они использовали комплексный анализ, производящие функции, вычеты и асимптотики. Это можно прочитать, например, в шестой главе книги Шрёдингера [16].

2. Добавим, что, следуя Гиббсу, вводят *статистический интеграл*

$$Z(\beta, a) := \int_\Gamma e^{-\beta H}\, dq\, dp,$$

аналогичный рассмотренной выше статистической сумме, где, как и в распределении Гиббса, $H(q, p, a)$ — гамильтониан, а $\beta = 1/(k\tau)$ с прежним смыслом всех переменных.

Тогда внутренняя энергия системы и уравнения состояния системы (см. с. 98, 115) выразятся в виде

$$E=-\frac{\partial\log Z}{\partial\beta},\quad A_i=\frac{1}{\beta}\frac{\partial\log Z}{\partial a_i},\quad 1\leqslant i\leqslant n,$$

а её энтропия — в форме преобразования Лежандра от $\log Z$

$$S=k(\log Z+\beta E)=k\Big(\log Z-\beta\frac{\partial\log Z}{\partial\beta}\Big).$$

Энтропия допускает и другое, известное уже Больцману, представление:

$$S=-k\int\limits_{\Gamma}\varrho\log\varrho$$

через плотность канонического распределения Гиббса. Именно в этом облике она появляется в теории информации уже для произвольной плотности распределения.

Проверим эту формулу. Поскольку $\varrho=e^{-\beta H}/Z$, имеем

$$-\int\limits_{\Gamma}\varrho\log\varrho=\log Z+\beta\int\limits_{\Gamma}He^{-\beta H}/Z.$$

Последний интеграл есть $-\dfrac{\partial\log Z}{\partial\beta}$. Учитывая это и вспоминая указанное выше выражение энтропии как преобразования Лежандра от $\log Z$, убеждаемся в справедливости проверяемой формулы.

3. Продемонстрируем на примере замечательную компактность и эффективность техники статистических интегралов.

Пусть в нашем обычном пространстве $\mathbb{R}^3$ имеется сосуд D объёма V, заполненный идеальным газом, представляющим из себя n слабо взаимодействующих подвижных частиц одинаковой массы m. Состояние каждой частицы определяется её положением (три координаты) и импульсом (ещё три числа). Гамильтониан системы (кинетическая энергия) имеет вид

$$H=\sum_{i=1}^{n}\frac{|p_i|^2}{2m},$$

и статистический интеграл

$$Z=\int\limits_{\Gamma}e^{-\beta H}=\int\limits_{D^n}dq\int\limits_{\mathbb{R}^{3n}}e^{-\beta H}dp=V^n\int\limits_{\mathbb{R}^{3n}}e^{-\beta H}\,dp$$

в сферических координатах легко до конца считается:

$$Z=c\frac{V^nm^{3n/2}}{\beta^{3n/2}},$$

где положительный множитель c зависит только от n.

Таким образом,

$$\log Z = n\log V - \frac{3n}{2}\log\beta + \frac{3n}{2}\log m + \log c.$$

Отсюда сразу находим связь энергии и абсолютной температуры системы

$$E = -\frac{\partial \log Z}{\partial \beta} = \frac{3n}{2\beta} = \frac{3}{2}nk\tau,$$

а также давление

$$P = \frac{1}{\beta}\frac{\partial \log Z}{\partial V} = \frac{nk\tau}{V}$$

как обобщённую силу, отвечающую параметру V.

Если перейти к обозначениям классической термодинамики, заменив τ на T, k на R/N, nk на νR, где R — универсальная газовая постоянная, N — число Авогадро, ν — количество газа в молях, то полученные соотношения превратятся соответственно в закон Джоуля $E = \frac{3}{2}\nu RT$ и закон Клапейрона $PV = \nu RT$.

Обратите внимание на то, что исходно гамильтониан системы нам здесь даже не был дан как функция температуры и внешнего параметра.

Более того, выкладки не зависят от количества n молекул. Последнее может озадачить физика, понимающего, что о давлении и температуре газа, как и о газе вообще, имеет смысл говорить, только когда n достаточно велико.

Такое же совмещение достоинства и странности содержится в аксиоматической модели Каратеодори классической феноменологической термодинамики, о которой мы говорили ранее. На это обратил внимание Планк, высоко ценивший работы Каратеодори.

Не подобное ли обстоятельство, помимо многого другого, смущало и Максвелла, который дал несколько доказательств своего закона распределения молекул газа по скоростям (кинетическим энергиям)? Одно из доказательств, совсем элементарное, проходит при любом количестве n молекул. Другое приводит к результату при $n \to \infty$.

4. Теперь скажем несколько слов о самом предмете и этих записях.

Почти за каждым, даже очень мелким, разделом и подзаголовком этих записей стоит обычно не только сюжет, а целая область исследований, подробное изложение которой является предметом самостоятельного полновесного курса. Мы посмотрели на многое, но с высоты и в темпе птичьего полёта. В этом надо себе отдавать отчёт, пополняя и уточняя сведения по заинтересовавшему вас вопросу в специальной литературе. Определённый выход на неё даст помещённая ниже библиография.

Философия статистической термодинамики, глубинные связи и содержание её понятий и принципов было предметом по сей день цитируемой классической работы-обзора П. и Т. Эренфестов [6].

Современным математическим языком мы обрисовали лишь скелет двух фундаментальных начал термодинамики, не касаясь многообразия их конкретных проявлений и самой физической термодинамики (см. например, [21—24]).

5. Наконец, добавим два слова типа сноски, не помещённой в нужном месте. Мы начали с двух фундаментальных понятий термодинамики, энергии и энтропии, ими же и закончим.

Забавно, что, когда в 1841 г. Роберт Майер пытался опубликовать свою первую работу о законе сохранения энергии, Поггендорф — редактор журнала «Annalen der Physik» отказался её публиковать. Между тем этот отказ по существу послужил на благо, потому что в первой редакции статья содержала столько ошибок, что могла серьёзно скомпрометировать саму идею, лежащую в её основе, см. [20].

Примерно то же, но по другой причине, случилось и с Клаузиусом с его вторым началом термодинамики. Пришлось отстаивать и величину, названную им потом *энтропией*.

(Выбор Клаузиусом термина энтропия объясняют [20] так.

[Эта величина математически строго определена, но физически мало наглядна. Более того, её абсолютное значение остаётся неопределённым, определены лишь её изменения в термодинамически изолированных необратимых системах; в идеальном случае обратимых процессов энтропия остаётся постоянной.

Поскольку её изменение равно нулю для идеальных обратимых процессов и положительно для реальных, энтропия есть мера отклонения реального процесса от идеального. Этим объясняется данное Клаузиусом название этой величины, которое этимологически означает «изменение».]

По правде говоря, мне это не кажется очень убедительным. Для обозначения отличия существуют куда более точные термины, характеризующие отклонение, а не изменение, перемену и превращение. «Эн»эргия — это ведь содержащаяся «В» объекте способность совершить работу, а «эн»тропия — преобразование, превращение, перераспределение, перемешивание «в» чем-то. Недаром ту часть атмосферы, где происходят самые активные повседневно нами наблюдаемые явления, изменения, превращения, называют *тропо*сферой.)

6. Ещё два слова о двух началах термодинамики.

Первое начало термодинамики — это и общий закон сохранения энергии, и, конечно, исходное открытие эквивалентности теплоты и работы, включающее механический эквивалент теплоты.

Второе начало классической термодинамики — это утверждение о направленности эволюции замкнутой системы в сторону роста её энтропии при переходе из одного равновесного состояния в другое. (Газ, содержавшийся в половине сосуда, после снятия перегородки заполнит весь сосуд. Сам по себе он обратно в половину сосуда не соберётся.)

Под влиянием внешних воздействий (например, когда, двигая поршень, мы меняем объём газа в цилиндре) система (газ) уже может менять своё состояние и в ином направлении. Если всё происходит бесконечно медленно, то можно считать, что мы всё время проходим только через равновесные состояния системы и весь процесс обратим.

Математически это выражается в том, что для обратимых процессов интеграл от формы $T^{-1}\delta Q$ или $\tau^{-1}\omega$ приведённого тепла вдоль пути перехода равен приращению энтропии системы.

Если же условие равновесия по ходу процесса нарушалось, то приращение энтропии системы окажется большим, чем этот интеграл. Это неравенство (Клаузиуса) — характеристическое свойство необратимых процессов (например, контактного теплообмена тел с разной температурой).

Для любого осуществимого термодинамического цикла (замкнутого пути) γ неравенство Клаузиуса имеет вид

$$\int_{\gamma} \frac{\delta Q}{T} \leqslant 0,$$

причём равенство здесь реализуется лишь на обратимых переходах γ. Они идут через равновесные состояния термодинамической системы, которые мы и рассматривали выше (см. с. 99).

Отметим, что формально энтропия определена только для равновесных состояний термодинамической системы. Распространение этого понятия на неравновесные состояния обычно сопровождается разбиением системы на микросистемы, находящиеся по предположению в равновесном состоянии, и суммированием энтропий этих систем.

Подчеркнём также, что в формулировке второго начала термодинамики (в форме Клаузиуса о контактном переходе тепла от горячего тела к холодному или в форме Каратеодори об адиабатической недостижимости некоторых состояний) негласно подразумевалось, что речь идёт о термически однородных системах, т. е. таких, термическое равновесие в которых возможно только при одинаковой температуре всех частей системы.

Нарушение этого условия может приводить к кажущимся противоречиям и парадоксам. Возьмите, например, пару одинаковых брусков. Один при температуре 0 градусов, а другой при температуре 100 градусов. Если привести их в тепловой контакт, то равновесие наступит тогда, когда оба будут иметь температуру 50 градусов. Больше тепла передать от второго тела первому не получится. Но если разделить каждый из брусков пополам теплонепроницаемой перегородкой, то небольшими манипуляциями, связанными с последовательными сдвигами и контактами брусков, можно привести дело к состоянию, когда половинки первого бруска будут иметь температуры 75 и 50 градусов соответственно, в то время как половинки исходно горячего бруска будут иметь температуры 50 и 25 градусов. После этого можно перегородки удалить. Температуры каждого из брусков станут постоянными, причём первый брусок будет горячее второго, который был источником тепла. Кажущееся нарушение принципа Клаузиуса!

Приведём другой пример, относящийся уже к принципу Каратеодори. Два идеальных газа с различными теплоёмкостями, взятые в количестве одного моля каждый, разделены адиабатическим скользящим поршнем. Можете проверить (вслед за Т. А. Афанасьевой-Эренфест), что для этой термически неоднородной системы (части которой в равновесном состоянии системы могут иметь разную температуру) форма δQ не имеет интегрирующего множителя и, значит, нет запрета на адиабатические переходы между любыми состояниями системы [22, с. 59].

7. Некоторые даты (минуя Большой Взрыв, рождение Солнца и подвиг Прометея).

Феноменологическая термодинамика

1680—1705	Изобретение паровой машины. Первый патент на паровой двигатель выдан в 1705 г. английскому слесарю Томасу Ньюкомену [T. Newcomen].
1765	Вопрос Джеймса Уатта [J. Watt (1736—1819)] о коэффициенте полезного действия тепловой машины (впоследствии механический эквивалент теплоты).
1824	Работа Сади Карно [S. Carnot (1796—1832)] «О движущей силе огня...» (впоследствии второе начало термодинамики).
1834	Обнаружение Бенуа Клапейроном [B. Clapeyron (1799—1864)] работы С. Карно и закона Клапейрона (объединившего законы Гей-Люссака и Бойля—Мариотта).
1840—1850	Формулировка и уточнение закона сохранения энергии: Юлиус Роберт Майер [J. R. Mayer (1814—1878)], Джеймс Прескотт Джоуль [J. P. Joule (1818—1889)], Герман Гельмгольц [H. Helmholtz (1821—1894)], Рудольф Клаузиус [R. Klausius (1832—1888)].
1850—1925	Оформление классической феноменологической термодинамики. Равновесная термодинамика как наука (Клаузиус [Klausius]). Контактная геометрия термодинамики (Гиббс [Gibbs]). Аксиоматизация термодинамики (Каратеодори [C. Carathéodory]).

Молекулярная теория теплоты

Идея о молекулярном строении вещества древняя. Идея объяснить тепло молекулярным движением помоложе: ..., Фрэнсис Бэ́кон [F. Bacon (1561—1626)], Иоганн Кеплер [Kepler (1572—1630)], Леонард Эйлер [L. Euler (1707—1783)], ...

1738	Даниил Бернулли [D. Bernulli (1700—1782)]. Молекулярное объяснение давления (в труде «Гидродинамика»).
1856	Август Крёниг [A. Krönig (1822—1879)]. Связаны температура и кинетическая энергия.
1857—1865	Рудольф Клаузиус [R. Klausius(1832—1888)]. Развёрнутая механическая теория теплоты. Исследование принципа эквивалентности теплоты и работы (1850).

Понятие внутренней энергии;
формулировка закона сохранения в форме $\delta Q = dE + P\,dV$;
понятие энтропии и второе начало термодинамики (1865).

Статистическая механика, термодинамика, физика

1860—1866 Джеймс Клерк Максвелл [J. C. Maxwell (1831—1879)].
Закон Максвелла распределения молекул по скоростям и кинетической энергии; длина свободного пробега молекул.
Демон Максвелла, вопросы обратимости.

1868—1906 Людвиг Больцман [L. Boltzmann (1844—1906)].
Распределение Больцмана молекул по энергиям в потенциальном поле.
Последовательный статистический подход к термодинамике.
Уравнение Больцмана и эволюция термодинамических систем.
Энтропия и вероятность состояния. H-теорема о росте энтропии.
Проблемы второго начала термодинамики.
Эргодическая гипотеза.

1883—1892 Анри Пуанкаре [H.Poincaré (1854—1912)].
Динамические системы.
Теорема о возвращении.
Газ как бесстолкновительная сплошная среда и её эволюция.

1896 Эрнест Цермело [E. Zermelo (1871—1953)].
Парадоксы термодинамики.

1902 Джозайя Уиллард Гиббс [J. W. Gibbs (1839—1903)].
Математическая теория статистической механики.
Меры, распределения и их эволюция в гамильтоновых системах.
Равновесные состояния как инвариантные меры.

1905 Альберт Эйнштейн [A. Einstein (1879—1955)].
Теория броуновского движения и второе начало термодинамики; размеры атомов, число Авогадро.

Квантовая статистическая механика, термодинамика, физика

1887—1892 Макс Планк [M. Planck (1858—1947)].
Рождение квантовой теории.

1902 Джозайя Уиллард Гиббс [J. W. Gibbs(1839—1903)].
Термодинамический парадокс.

1926 Эрвин Шрёдингер [E. Schrödinger (1887—1961)].
Квантовая статистическая механика...

Литература

Тема I

Главы I, II

Доступные руководства

1. *Bridgman P. W.* Dimensional analysis. New Haven: Yale University Press, 1932. [Рус. перев.: *Бриджмен П.* Анализ размерностей. М.; Ижевск: РХД, 2001. Это переиздание перевода второго английского издания под редакцией акад. С. И. Вавилова. Оно включает также нобелевскую лекцию П. Бриджмена, посвящённую физике высоких энергий.]
2. *Birkhoff G.* Hydrodynamics. A study in logic, fact and similitude. (Revised edition.) Princeton University Press, 1960. [Рус. перев.: *Биркгоф Г.* Гидродинамика. Методы. Факты. Подобие. М.: ИЛ, 1963.]
3. Седов Л. И. Методы подобия и размерности в механике. М.: Наука, 1967.
4. а) Международная система единиц (СИ). М.: Высшая школа, 1964.

   б) Единицы величин: Словарь-справочник. М.: Издательство стандартов, 1990.

   в) Le Système International d'unités (SI) / Edition du Bureau International des Poids et Mesures. Paris, 1970.

К истории вопроса

5. *Görtler H.* Zur Geschichte des Π-Theorems // Z. Angew. Math. Mech. 1975. Bd. 55, № 1. S. 3—8.
6. *Федерман А.* О некоторых общих методах интегрирования уравнений с частными производными первого порядка // Изв. Санкт-Петербург. Политехн. ин-та. Отд. техн., естествозн. и математ. 1911. Т. 16, № 1. С. 97—155.
7. Из писем в редакцию: Н. А. Морозов — основоположник анализа размерности // УФН. 1953. Т. XLIX, № 1. С. 180—181.

Библиография и нумерация в каждой главе свои (исключение — вспомогательная глава II темы I). Список и его разметка условны, далеки от полноты и совершенства, если не сказать произвольны, учитывая море литературы по затрагиваемым здесь вопросам. (Говорят, что на входе в один из баров Техаса висело объявление: «Просим не стрелять в наших музыкантов. Они играют как могут».)

8. *Ди Бартини Р. О.* Некоторые соотношения между физическими константами // ДАН СССР. 1965. Т. 163, № 4. С. 861—864. (Знаменитая пародия на математическую статью, представленная в «Доклады Академии наук» Бруно Понтекорво, кажется, к первоапрельскому номеру.)

Дополнительная литература

9. *Ландау Л. Д., Лифшиц И. М.* Гидродинамика. М.: Наука, 1988.
10. *Кочин Н. Е., Кибель И. А., Розе Н. В.* Теоретическая гидродинамика. Ч. II. М.: Физматлит, 1963.
11. *Whitney H.* Collected papers / Ed. J. Eells, D. Toledo. Vol. II. Boston: Birkhäuser, 1992. P. 530—584. См. также: The mathematics of physical quantities. P. I. Mathematical models for measurement // Am. Math. Monthly. 1968. Vol. 75. P. 115—138; P. II. Quantity structures and dimension analysis // Ibid. P. 237—256.
12. *Арнольд В. И.* Математические методы классической механики. М.: Наука, 1989. (С. 50. Соображения подобия.)
13. *Jarman M.* Examples in quantitative zoology. London: Edward Arnold, 1970. [Рус. перев.: *Джермен М.* Количественная биология в задачах и примерах. М.: Мир, 1972.]
14. *Смит Дж.* Математические идеи в биологии. М.: Мир, 1970.
15. *Манин Ю. И.* Математика и физика. М.: Знание, 1979. (Новое в науке и технике. Серия «Математика, кибернетика»; Вып. 12). [См. также в кн.: *Манин Ю. И.* Математика как метафора. М.: МЦНМО, 2010. С. 149—230.]
16. *Овсянников Л. В.* Групповой анализ дифференциальных уравнений. М.: Наука, 1978.

Глава III

1. *Колмогоров А. Н.* Локальная структура турбулентности в несжимаемой вязкой жидкости при очень больших числах Рейнольдса // ДАН СССР. 1941. Т. 30, № 4. С. 299—303. [Опубл. также в: УФН. 1967. Т. 93. С. 476—481; и в кн.: *Колмогоров А. Н.* Избранные труды. Т. 1: Математика и механика. М.: Наука, 1985. С. 281—287. (См.: Колмогоров. Юбилейное издание: В 3-х кн. Кн. 1. Биобиблиография. М.: Физматлит, 2003. С. 256.) Есть перевод на нем. яз. в сб.: Sammelband zur statistischen Theorie der Turbulenz. Berlin: Akademie-Verlag, 1958. S. 71—76.]
2. *Колмогоров А. Н.* К вырождению изотропной турбулентности в несжимаемой вязкой жидкости // ДАН СССР. 1941. Т. 31, № 6. С. 538—541. [Опубл. также в кн.: *Колмогоров А. Н.* Избранные труды. Т. 1: Математика и механика. М.: Наука, 1985. С. 287—290. Есть перевод на нем. яз. в сб.: Sammelband zur statistischen Theorie der Turbulenz. Berlin: Akademie-Verlag, 1958. S. 147—150.]
3. *Колмогоров А. Н.* Рассеяние энергии при локально изотропной турбулентности // ДАН СССР. 1941. Т. 32. С. 19—21. [Опубл. также в кн.: *Кол-*

могоров А. Н. Избранные труды. Т. 1: Математика и механика. М.: Наука, 1985. С. 290—293. Есть перевод на нем. яз. в сб.: Sammelband zur statistischen Theorie der Turbulenz. Berlin: Akademie-Verlag, 1958. S. 77—81.]

4. *Колмогоров А. Н.* Уточнение представления о локальной структуре турбулентности в несжимаемой вязкой жидкости при больших числах Рейнольдса // Mécanique de la turbulence: Actes du Colloque International du CNRS (Marseille, août-sept. 1961). Paris: Éditions du CNRS, 1962. P. 447—458 (на рус. и франц. яз.). [Опубл. также в кн.: *Колмогоров А. Н.* Избранные труды. Т. 1: Математика и механика. М.: Наука, 1985. С. 348—352. Опубл. также на англ. яз.: A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number // J. Fluid Mech. 1962. Vol. 13, № 1. P. 82—85.]
5. *Колмогоров А. Н.* Математические модели турбулентного движения вязкой жидкости // УМН. 2004. Т. 59, № 1. С. 5—10. [Посмертная публикация. Весь этот выпуск УМН посвящён А. Н. Колмогорову. В нём опубликованы тексты ряда докладов, сделанных на грандиозной международной конференции «Колмогоров и современная математика», проведённой в 2003 г. в Москве и приуроченной к 100-летию со дня рождения А. Н. Колмогорова. Там же на с. 25—44 есть текст доклада В. И. Арнольда «Колмогоров и естествознание», где, в частности, явно сформулированы общие гидродинамические принципы А. Н. Колмогорова, инициированные исследованием турбулентности.]
6. *Обухов А. М.* О распределении энергии в спектре турбулентного потока // Изв. АН СССР. Сер. Геогр. и геофиз. 1941. С. 4—5.
7. *Obukhov A. M.* Some specific features of atmospheric turbulence // J. Fluid Mech. 1962. Vol. 13, № 1. P. 77—81; J. Geophys. Res. 1962. Vol. 67. P. 3011—3014.
8. *Обухов А. М.* Течение Колмогорова и его лабораторное моделирование // УМН. 1983. Т. 38, № 4. С. 101—111.
9. Этюды о турбулентности. М.: Наука, 1994. [Сборник статей авторов, участвовавших в работе семинара А. М. Обухова и А. С. Монина. Посвящён памяти А. М. Обухова. Предисловие академика О. М. Белоцерковского с воспоминанием о профессоре Сергее Григорьевиче Калашникове, читавшем лекции на физфаке МГУ.]
10. *Ландау Л. Д., Лифшиц И. М.* Гидродинамика. М.: Наука, 1988.
11. *Кочин Н. Е., Кибель И. А., Розе Н. В.* Теоретическая гидродинамика. Ч. II. М.: Физматлит, 1963.
12. Гидродинамическая неустойчивость. Сборник статей. М.: Мир, 1964. [Статьи Эбергарда Хопфа (Hopf E.), Гаррета Биркгофа (Биркхофа) (Birkhoff G.), Ж. Кампе-де-Ферье (Kampé de Fériet J.), Роберта Крейчнана (Kraichnan R.) и др. См. библиографию в статьях.]

13. *Лаврентьев М. А., Шабат Б. В.* Проблемы гидродинамики и их математические модели. М.: Наука, 1977.
14. Visions in mathematics — Towards 2000. GAFA 2000 (Tel Aviv, 1999). Geom. Funct. Anal. Special Volume. Basel: Birkhäuser, 2000.
15. *Kupiainen A.* Lessons for turbulence // GAFA 2000 (Tel Aviv, 1999). Geom. funct. anal. Special Volume, Part I. Basel: Birkhäuser, 2000. P. 316—333.
16. *Frish U.* Turbulence. The legacy of A. N. Kolmogorov. Cambridge: Cambridge University Press, 1995. XIV+299 p. [Рус. перев.: *Фриш У.* Турбулентность. Наследие А. Н. Колмогорова / Пер. с англ. А. Н. Соболевского под ред. М. Л. Бланка. М.: Фазис, 1998. XIV+346 с.]
17. Колмогоров. Юбилейное издание: В 3 кн. Кн. 1. Биобиблиография. М.: Физматлит, 2003. С. 225, 78—80, 342.
18. *Ruelle D.* Hasard et chaos. Paris: Editions Odile Jacob, 1991. [Рус. перев.: *Рюэль Д.* Случайность и хаос. М.; Ижевск: РХД, 2001.]
19. *Mallat S.* A wavelet tour of signal processing. Academic Press, 1999. [Рус. перев.: *Малла С.* Вейвлеты в обработке сигналов. М.: Мир, 2005.]
    (См. с. 236 о работах Колмогорова 1941 и 1962 годов по турбулентности.

    Отмечается, что модель пока не учитывает вихревой процесс и наблюдаемую перемежаемость вихревых и относительно спокойных областей течений. Нет объяснения механизма образования вихрей и обмена энергией между мелкомасштабными и крупномасштабными структурами.

    «Понимание свойств гидродинамической турбулентности — важнейшая задача современной физики.

    Ни один формализм не в состоянии построить физико-статистическое описание, основанное на уравнениях Навье—Стокса, которое бы давало нам возможность понять глобальное поведение турбулентных течений, как это сделано в термодинамике».)
20. Цитата из газеты «Известия» от 15 апреля 2005 г., с. 17.
    [Академик Алексей Липанов, директор Института прикладной математики, в интервью корреспонденту «Известий» говорит:

    «Нам удалось создать математическую модель турбулентности — решить проблему, считавшуюся на протяжении столетия неразрешимой, так как процесс турбулентности считался случайным. Нам удалось доказать, показать, что это заблуждение и что данный процесс поддаётся чёткому моделированию. Практическое применение этого открытия очевидно, конструкторы самолётов и кораблей будут теперь пользоваться этой методикой, что позволит предотвратить катастрофы...» Из Калининграда с Международной конференции по избранным вопросам современной математики, приуроченной к 200-летию со дня рождения Карла Якоби.]

    Будем надеяться, что контакт авторов источников 19, 20 состоится и не приведёт к турбулентности.

Кстати, возможно, заодно уж решена и интересовавшая многих гидромехаников, физиков и инженеров практически важная задача расчёта перепада давления на концах трубы с учётом возможного турбулентного сопротивления прогоняемого через трубу потока жидкости или газа?

(В этой связи см., например, статью Баренблатта в упомянутом выше выпуске УМН (см. [5]). В связи с турбулентностью вообще интересующиеся могут посмотреть темы докладов на международном симпозиуме «Hamiltonian Dynamics, Vortex Structures, Turbulence», проведённом Математическим институтом им. В. А. Стеклова в августе 2006 г. в Москве, а также темы докладов Международной конференции «Математическая гидродинамика», проходившей в Математическом институте им. В. А. Стеклова в июне 2006 г.)

Математикам будет полезно посмотреть также следующие источники.

21. *Shiryaev A.* On the classical, statistical, and stochastic approaches to the hydrodynamic turbulence. Thiele centre for applied mathematics in natural science. Research report № 02, January 2007.
22. *Юдович В. И.* Глобальная разрешимость против коллапса в динамике несжимаемой жидкости // Математические события XX века. М.: Фазис, 2003. С. 519—548.
23. *Friedlander S., Yudovich V.* Instabilities in fluid motion // Notices of the AMS. 1999. Vol. 46, № 11. P. 1358—1367.
24. *Arnold V. I., Khesin B. A.* Topological methods in hydrodynamics. New York: Springer-Verlag, 1998. (Applied Mathematical Sciences; Vol. 125). [Дополненный рус. перев.: *Арнольд В. И., Хесин Б. А.* Топологические методы в гидродинамике. М.: МЦНМО, 2007.]

Тема II

Глава I

Первоисточники

1. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи // Всесоюзный энергетический комитет. Материалы к первому Всесоюзному съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. М.: Управление связи РККА, 1933.
2. *Shannon C. E.* A mathematical theory of communication // Bell System Tech. J. 1948. Vol. 27. P. 379—423, 623—656. [Рус. перев.: *Шеннон К.* Работы по теории информации и кибернетике. М.: ИЛ, 1963.]
3. Теория информации и её приложения. М.: Физматлит, 1959.

Сборник переводов включает статьи:

Хартли Р. В. Л. Передача информации. (*Hartley R. V. L.* Transmission of information // BSTJ. 1928. Vol. 7, № 3. P. 535—563.)

Оливер Б. М., Пирс Дж. Р., Шеннон К. Е. Принципы кодово-импульсной модуляции. (*Oliver B. M., Pierce J. R., Shannon C. E.* The philosophy of PCM // Proc. I. R. E. 1948. Vol. 36, № 11. P. 1324—1331.)
Таллер В. Дж. Теоретические ограничения скорости передачи информации. (*Tuller W. G.* Theoretical limitations on the rate of transmission of information // Proc. I. R. E. 1949. Vol. 37, № 5. P. 468—478.)
Шеннон К. Связь при наличии шума. (*Shannon C. E.* Communication in the presence of noise // Proc. I. R. E. 1949. Vol. 37, № 1. P. 10—21.)
Ли И. В., Читем Т. П., Виснер Дж. Б. Применение корреляционного анализа для обнаружения периодических сигналов в шуме. (*Lee Y. W., Cheatham T. P., Wiesner J. B.* Application of correlation analysis to the detection of periodic signals in noise // Proc. I. R. E. 1950. Vol. 38. P. 1165—1171.)

Обзор с развитием и библиографией

4. *Хургин Я. И., Яковлев В. П.* Финитные функции в физике и технике. М.: Наука, 1971.

Глава II

Математические первоисточники

1. *Poincaré H.* Calcul des probabilités. Paris: Gauthier-Villars, 1912. [Рус. перев.: *Пуанкаре А.* Лекции по теории вероятностей. М.; Ижевск: РХД, 1999.]
2. *Levy P.* Problèmes concrets d'analyse fonctionnelle. Paris: Gauthier-Villars, 1951.

Обновлённое изложение и математические применения

3. *Milman V., Schechtman G.* Asymptotic theory of finite dimensional normed spaces / With an appendix by M. Gromov. Springer-Verlag, 1986. (Lecture Notes in Mathematics; Vol. 1200). (Appendix I. *Gromov M.* Isoperimetric inequality in Riemannian manifolds. P. 114—129.)
4. *Talagrand M.* Concentration of measure and isoperimetric inequalities in product Spaces // Inst. Hautes Études Sci. Publ. Math. 1995. Vol. 81. P. 73—205.
5. *Ball K.* An elementary introduction to modern convex geometry // Flavors of geometry. Cambridge: Cambridge Univ. Press, 1997. (Math. Sci. Res. Inst. Publ.; Vol. 31.) P. 1—58.
6. *Мильман В. Д.* Явления, возникающие в высоких размерностях // УМН. 2004. Т. 59, № 1. С. 157—168.
7. *Milman V. D.* Geometrization of probability // Geometry and dynamics of groups and spaces. Basel: Birkhäuser, 2008. (Progress in Mathematics; Vol. 265). P. 647—667.
8. *Milman E.* On the role of convexity in isoperimetry, spectral gap and concentration // Invent. Math. 2009. Vol. 177, № 1. P. 1—43.

9. *Boucheron S., Lugosi G., Massart P.* Concentration inequalities. A nonasymptotic theory of independence / With a foreword by M. Ledoux. Oxford: Oxford University Press, 2013.
10. *Van Handel R.* Probability in high dimension. Lecture notes. Princeton University. 2014.
11. *Громов М. Л.* Красочные категории // УМН. 2015. Т. 70, № 4. С. 3—76.
12. Стохастический анализ в задачах. Ч. I / Под ред. А. В. Гасникова. М.: Изд-во МФТИ, 2016.
13. *Богачёв В. И.* Слабая сходимость мер. М.; Ижевск: ИКИ, 2016.
14. *Зорич В. А.* Геометрия и вероятность // ТВП. 2017. Т. 62, № 2. С. 292—310.

Физические источники и их математическое развитие

15. *Лоренц Г. А.* Статистические теории в термодинамике. М.; Ижевск: РХД, 2001.
16. *Шрёдингер Э.* Лекции по физике. М.; Ижевск: РХД, 2001.
17. *Хинчин А. Я.* Симметрические функции на многомерных поверхностях // Памяти А. А. Андронова: Сборник статей. М.: Изд-во АН СССР, 1955. С. 541—576.
18. *Опойцев В. И.* Нелинейный закон больших чисел // Автомат. и телемех. 1994. № 4. С. 65–75.
19. *Манин Ю. И.* Математика и физика. М.: Знание, 1979. (Новое в науке и технике. Серия «Математика, кибернетика»; Вып. 12). [См. также в кн.: *Манин Ю. И.* Математика как метафора. М.: МЦНМО, 2010. С. 149—230.]
20. *Минлос Р. А.* Введение в математическую статистическую физику. М.: МЦНМО, 2002.
21. *Кац М.* Вероятность и смежные вопросы в физике. М.: УРСС, 2003.
22. *Козлов В. В.* Тепловое равновесие по Гиббсу и Пуанкаре. М.; Ижевск: РХД, 2002.
23. *Рюэль Д.* Случайность и хаос. М.; Ижевск: РХД, 2001.
24. *Kurchan J., Laloux L.* Phase space geometry and slow dynamics // J. Phys. A: Math. Gen. 1996. Vol. 29. P. 1929—1948.

Глава III

Первоисточники

1. *Shannon C. E.* A mathematical theory of communication // Bell System Tech. J. 1948. Vol. 27. P. 379—423, 623—656. [Рус. перев.: *Шеннон К.* Работы по теории информации и кибернетике. М.: ИЛ, 1963.]
2. Теория информации и её приложения. М.: Физматлит, 1959. Сборник переводов включает статью: *Шеннон К.* Связь при наличии шума. (*Shannon C. E.* Communication in the presence of noise // Proc. I. R. E. 1949. Vol. 37, № 1. P. 10—21.)

Отступление об ε-энтропии и проблеме Гильберта

3. *Hilbert D.* Mathematische Probleme // Gesammelte Abhandlungen. 1935. Vol. III. P. 290—329.

Проблемы Гильберта // Сборник статей под общ. ред. П. С. Александрова. М.: Наука, 1969.

См. также: *Гильберт Д.* Избранные труды. Т. 2. М.: Факториал, 1998.

4. *Витушкин А. Г.* К тринадцатой проблеме Гильберта // ДАН СССР. 1954. Т. 95, № 4. С. 701—704.
5. *Витушкин А. Г.* 13-я проблема Гильберта и смежные вопросы // УМН. 2004. Т. 59, № 1. С. 11—24.
6. *Колмогоров А. Н.* Оценки минимального числа точек ε-сетей в различных функциональных классах и их применение к к вопросу о представимости функций нескольких переменных суперпозициями функций меньшего числа переменных // ДАН СССР. 1955. Т. 101, № 2. С. 192—194.
7. *Колмогоров А. Н.* О представлении непрерывных функций нескольких переменных суперпозициями функций меньшего числа переменных // ДАН СССР. 1956. Т. 108, № 2. С. 179—182.
8. *Колмогоров А. Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного и сложения // ДАН СССР. 1957. Т. 114, № 5. С. 953—956.
9. *Arnold V. I.* From Hilbert's superposition problem to dynamical systems // The Arnoldfest (Toronto, ON, 1997). Providence, RI: AMS, 1999. (Fields Inst. Commun.; Vol. 24). P. 1—18.

Русский текст этого доклада есть в книге: Математические события XX века. М.: Фазис, 2003. С. 19—48.

Арнольд В. И. О функциях трёх переменных // ДАН СССР. 1957. Т. 114, № 4. С. 679—681.

10. *Колмогоров А. Н., Тихомиров В. М.* ε-энтропия и ε-ёмкость множеств в функциональных пространствах // УМН. 1959. Т. 14, № 2. С. 3—86.

Дополнительная литература

11. *Витушкин А. Г.* Оценка сложности задачи табулирования. М.: Физматлит, 1959.
12. *Буслаев В. И., Витушкин А. Г.* Оценка длины кода сигналов с конечным спектром в связи с задачами звукозаписи // Изв. АН СССР. Сер. Матем. 1974. Т. 38. С. 867—895.
13. *Блаттер К.* Вейвлет-анализ. Основы теории. М.: Техносфера, 2004. (Серия «Мир математики»).
14. *Малла С.* Вейвлеты в обработке сигналов. М.: Мир, 2005.
15. *Higgins J. R.* Five short stories about the cardinal series // Bull. Amer. Math. Soc. (N. S.). 1985. Vol. 12, № 1. P. 45—89.

16. *Хургин Я. И., Яковлев В. П.* Финитные функции в физике и технике. М.: Наука, 1971.
17. *Вентцель Е. С.* Теория вероятностей. М.: Наука, 1964.
18. *Холево А. С.* Введение в квантовую теорию информации. М.: МЦНМО, 2002.

Тема III

Глава I

Первоисточники, переводы и описания

1. Второе начало термодинамики. М.; Л.: Гостехиздат, 1934. Сб. работ: С. Карно, В. Томсон-Кельвин, Р. Клаузиус, Л. Больцман, М. Смолуховский.
2. *Лоренц Г. А.* Статистические теории в термодинамике. М.; Ижевск: РХД, 2001.
3. *Ehrenfest P., Ehrenfest T.* Begriffliche Grundlagen der statistischen Auffassung in der Mechanik. Enzyclopädie der Math. Wiss. Bd. IV. 1911.
4. *Льоцци М.* История физики. М.: Мир, 1970.

Некоторые руководства и лекции по термодинамике

5. *Planck M.* Vorlesungen über Thermodynamik. Leipzig: Veit, 1917.
6. *Зоммерфельд А.* Термодинамика и статистическая физика. М.: ИЛ, 1955; М.; Ижевск: РХД, 2002.
7. *Леонтович М. Л.* Введение в термодинамику. Статистическая физика. М.: Наука, 1983.
8. а) *Фейнман Р.* Фейнмановские лекции по физике. Кн. 4. Кинетика, теплота, звук. М.: Мир, 1967.
   б) *Фейнман Р.* Статистическая механика. Волгоград: Платон, 2000.
9. *Шрёдингер Э.* Лекции по физике. М.; Ижевск: РХД, 2001.
10. *Ферми Э.* Термодинамика. М.; Ижевск: РХД, 1998.

К формализации классической термодинамики

11. а) *Carathéodory C.* Untersuchungen über die Grundlagen der Thermodynamik // Math. Ann. 1909. Vol. 67. S. 355—386. Включено в: *Carathéodory C.* Gesammelte Mathematische Schriften. Bd. 2. München, 1955. S. 131—166. [Рус. перев.: *Каратеодори К.* Об основах термодинамики // Развитие современной физики. М.: Наука, 1964. С. 188—222. Там же: *Борн М.* Критические замечания по поводу традиционного изложения термодинамики. С. 223—257.]
    б) *Carathéodory C.* Über die Bestimmung der Energie und der absoluten Temperatur mit Hilfe von reversiblen Prozessen // Sitzungsberichte der Preussischen Akademie der Wissenschaften. Physikalisch-mathematische

Klasse. Berlin, 1925. S. 39—47. См. также в собрании сочинений: *Carathéodory C.* Gesammelte Mathematische Schriften. Bd. 2. München, 1955. S. 167—177.

Глава II

Термодинамика и контактная геометрия

1. а) *Gibbs J. W.* Graphical methods in the thermodynamics of fluids // Transactions of Connecticut Academy. 1873. Vol. 2. P. 309—342.

   б) *Gibbs J. W.* A method of geometrical representation of the thermodynamic properties of substances by means of surfaces // Transactions of Connecticut Academy. 1873. Vol. 2. P. 382—404.

   в) *Gibbs J. W.* On the equilibrium of heterogeneous substances // Transactions of Connecticut Academy. 1875—1878. P. 108—248, 344—524.

   г) *Gibbs J. W.* Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics. New Haven, 1902.

   Всё имеется на рус. яз. в книге: *Гиббс Дж. В.* Термодинамика. Статистическая механика. М.: Наука, 1982. В частности, книга содержит статьи «О равновесии гетерогенных веществ» (с. 61—349) и «Основные принципы статистической механики со специальным применением к рациональному обоснованию термодинамики» (с. 350—508).
2. Proceedings of the Gibbs Simposium: Yale University, May 15—17, 1989. Providence, RI: AMS, 1990.
3. *Arnold V. I.* Contact geometry: the geometrical method of Gibbs's thermodynamics // Proceedings of the Gibbs Simposium: Yale University, May 15—17, 1989. Providence, RI: AMS, 1990. P. 163—179.

Математические аспекты

4. *Картан А.* Дифференциальное исчисление. Дифференциальные формы. М.: Мир, 1971.
5. *Рашевский П. К.* О соединимости любых двух точек вполне неголономного пространства допустимой линией // Уч. зап. Моск. пед. ин-та им. К. Либкнехта. Сер. физ. мат. наук. 1938. Т. 3, вып. 2. С. 83—94.
6. *Chow W. L.* Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung // Math. Ann. 1939. Vol. 117. S. 98—105.
7. Sub-Riemannian geometry. Basel: Birkhäuser, 1996. (Progr. Math.; Vol. 144).
8. *Sussmann H. J.* Orbits of families of vector fields and integrabilities of distributions // Trans. Amer. Math. Soc. 1973. Vol. 180. P. 171—188.
9. *Gromov M.* Carnot—Carathéodory spaces seen from within // Sub-Riemannian geometry. Basel: Birkhäuser, 1996. (Progr. Math.; Vol. 144). P. 79—323.

10. *Montgomery R.* A tour of subriemannian geometries, their geodesics and applications. Providence, RI: AMS, 2002. (Mathematical Surveys and Monographs; Vol. 91).
11. *Уорнер Ф.* Основы теории гладких многообразий и групп Ли. М.: Мир, 1987.
12. *Арнольд В. И.* Математические методы классической механики. М.: Наука, 1989.
13. *Арнольд В. И.*, *Гивенталь А. Б.* Симплектическая геометрия. Ижевск: РХД, 2000.
14. *Lieb H.*, *Yngvason J.* The mathematics of the second law of thermodynamics // GAFA 2000 (Tel Aviv, 1999). Geom. funct. anal. Special Volume, Part I. Basel: Birkhäuser, 2000. P. 334—358.
15. *Антоневич А. Б.*, *Бахтин В. И.*, *Лебедев А. В.*, *Саражинский В. Д.* Лежандров анализ, термодинамический формализм и спектры операторов Перрона—Фробениуса // Докл. АН (РАН). 2003. Т. 390, № 3. С. 295—297.

Глава III

Некоторые работы классиков

1. *Boltzmann L.* Wissenschaftliche Abhandlungen. Bd. I—III. New York: Chelsea Publishing Company, 1968. (Reprint of a work first published 1909 in Leipzig). [Рус. перев.: *Больцман Л.* Лекции по теории газов. М.: ОНТИ, 1936.]

   Больцман Л. Избранные труды. (Молекулярно-кинетическая теория газов. Термодинамика. Статистическая механика. Теория излучения. Общие вопросы физики.) М.: Наука, 1984.

   Второе начало термодинамики // Сб. работ: С. Карно, В. Томсон-Кельвин, Р. Клаузиус, Л. Больцман, М. Смолуховский. М.; Л.: Гостехиздат, 1934.
2. *Gibbs J. W.* Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics. New Haven, 1902. [Рус. перев.: *Гиббс Дж. В.* Основные принципы статистической механики со специальным применением к рациональному обоснованию термодинамики // Термодинамика. Статистическая механика. М.: Наука, 1982. С. 350—508.]
3. а) *Einstein A.* Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen // Ann. Phys. 1905. Vol. 17. P. 549—560. [Рус. перев.: *Эйнштейн А.* О движении взвешенных в покоящейся жидкости частиц, требуемом молекулярно-кинетической теорией теплоты // Собрание научных трудов. Т. III. М.: Наука, 1966. С. 108—117.]

   б) *Einstein A.* Zur Theorie der Brownschen Bewegung // Ann. Phys. 1906. Vol. 19. S. 371—381. [Рус. перев.: *Эйнштейн А.* К теории броуновского движения // Собрание научных трудов. Т. III. М.: Наука, 1966. С. 118—127.]

4. *Poincaré H.* Thermodynamique. Deuxiéme édition, revue et corrigée. Paris: Gauthier-Villars, 1908. [Рус. перев.: *Пуанкаре А.* Термодинамика. М.; Ижевск: РХД, 2005.]
5. *Лоренц Г. А.* Статистические теории в термодинамике. М.; Ижевск: РХД, 2001.
6. *Ehrenfest P., Ehrenfest T.* Begriffliche Grundlagen der statistischen Auffassung in der Mechanik. Enzyclopädie der Math. Wiss. Bd. IV. 1911.

Дальнейшие источники

7. *Боголюбов Н. Н.* Проблемы динамической теории в статистической физике. М.; Л.: Гостехиздат, 1946.
8. *Хинчин А. Я.* Математические основания статистической механики. М.; Ижевск: РХД, 2002.
9. *Березин Ф. А.* Лекции по статистической физике. М.: МЦНМО, 2007.
10. *Маслов В. П.* Ультравторичное квантование и квантовая термодинамика. М.: УРСС, 2000.
11. *Кац М.* Вероятность и смежные вопросы в физике. М.: УРСС, 2003. (Включает приложение I с лекцией Г. Е. Уленбека «Уравнение Больцмана».)
12. *Козлов В. В.* Тепловое равновесие по Гиббсу и Пуанкаре. М.; Ижевск: РХД, 2002.
13. *Минлос Р. А.* Введение в математическую статистическую физику. М.: МЦНМО, 2002.
14. а) *Синай Я. Г.* Современные проблемы эргодической теории. М.: Физматлит, 1995.
    б) *Синай Я. Г.* Теория фазовых переходов. М.; Ижевск: РХД, 2001.
15. а) *Рюэль Д.* Случайность и хаос. М.; Ижевск: РХД, 2001.
    б) *Рюэль Д.* Термодинамический формализм. Математические структуры классической равновесной статистической механики. М.; Ижевск: РХД, 2002.
16. *Шрёдингер Э.* Лекции по физике. М.; Ижевск: РХД, 2001.
17. *Зоммерфельд А.* Термодинамика и статистическая физика. М.: ИЛ, 1955; М.; Ижевск: РХД, 2002.
18. *Carathéodory C.* Untersuhungen über die Grundlagen der Thermodynamik // Math. Ann. 1909. Vol. 67. S. 355—386. Включено в: *Carathéodory C.* Gesammelte Mathematische Schriften. Bd. 2. München, 1955. S. 131—166. [Рус. перев.: *Каратеодори К.* Об основах термодинамики // Развитие современной физики. М.: Наука, 1964. С. 188—222. Там же: *Борн М.* Критические замечания по поводу традиционного изложения термодинамики. С. 223—257.]
19. *Carathéodory C.* Über die Bestimmung der Energie und der absoluten Temperature mit Hilfe von reversiblen Prozessen // Sitzungsberichte der Preussischen Akademie der Wissenschaften Physikalisch-mathematische Klasse.

Berlin, 1925. S. 39—47. [См. также в собр. соч.: *Carathéodory C.* Gesammelte Mathematische Schriften. Bd. 2. München, 1955. S. 167—177.]

20. *Льоцци М.* История физики. М.: Мир, 1970.
21. *Фейнман Р.* Фейнмановские лекции по физике. Кн. 4: Кинетика, теплота, звук. М.: Мир, 1967.
22. *Леонтович М. Л.* Введение в термодинамику. Статистическая физика. М.: Наука, 1983.
23. *Ферми Э.* Термодинамика. М.; Ижевск: РХД, 1998.
24. *Ландау Л. Д., Лифшиц Е. М.* Статистическая физика. М.: Наука, 1964.
25. *Рудой Ю. Г.* Математическая структура равновесной термодинамики и статистической механики. М.; Ижевск: ИКИ, 2013.

Приложение
Математика как язык и метод

Что, почему и зачем? Пояснение для очень не маленьких[1]

1. Согласно легенде (популярностью которой, по некоторым сведениям, мы обязаны знаменитому французу Аруэ, известному больше под псевдонимом Вольтер[2]) на голову Ньютона упало яблоко. Поскольку к тому времени в этой голове уже содержались и законы Кеплера, и очень многое другое, результат оказался отличным от того, что получается, когда яблоко падает на другие головы.

Кое-что из того, что произошло в математике как науке после этого события, за минувшие три столетия стало азбукой естествознания.

Лабиринт, рассматриваемый сверху, всегда представляется простым.

2. Намереваясь сказать здесь кое-что о математике вообще, начну с известного примера.

Старше ньютонова яблока легенда о бегущем голышом по Сиракузам Архимеде, восклицающем «Эврика! Эврика!» (Нашёл! Нашёл!). Она имеет несколько вариантов. Мы приведём два, а потом сделаем некоторые наблюдения и выводы.

Процитируем академика М. Л. Гаспарова[3].

«Дело было вот в чём. Сиракузский тиран Гиерон получил от золотых дел мастера золотой венец и хотел проверить, не подмешал ли мастер в золото серебра. Нужно было сравнить объёмы венца и куска чистого золота с тем же весом. Архимед, опускаясь в залитую до краёв ванну и глядя, как переливается через края вытесняемая его телом вода, вдруг понял, что именно так можно легко измерить объёмы двух тел разной формы».

---

[1]В романе Михаила Афанасьевича Булгакова «Мастер и Маргарита» сеанс магии, как известно, заканчивался разоблачением. Популярная пояснительная статья «Математика как язык и метод» имеет к «Математическому анализу задач естествознания» сходное отношение.

[2]Фамилия Arouet с уточнением L(e) J(eun) в латинском написании выглядит как AROVETLI. Перестановкой букв отсюда получен псевдоним VOLTAIRE.

[3]*Гаспаров М. Л.* Занимательная Греция. М.: Фортуна Лимитед, 2002. С. 362.

Несколько более продвинутый вариант состоит в следующем[1].

Мастеру были даны отдельно золото и серебро, которые в сплаве образуют прочное изделие. Архимеду, не повредив венца, надо было узнать, не заменил ли мастер часть золота серебром.

Пусть x_1, x_2 — соответственно количества граммов золота и серебра в готовом венце. Его вес $x_1+x_2=A$ легко измерить и проверить совпадение с общим весом того, что было дано мастеру. Мастер не дурак, совпадение будет. Подвесим венец на динамометр, погрузим его в полную ванну, соберём вылившуюся воду, измерим её объём V и вес P, а также прочитаем новое показание B весов. Свяжем полученные данные. Кому это затруднительно, может опустить пару абзацев.

Плотности p_1, p_2 (массы единиц объёма в терминах веса на Земле) наших драгоценных металлов давно хорошо известны.

Тогда величины $x_1/p_1=V_1$, $x_2/p_2=V_2$ дают объёмы каждого металла в венце. Значит, $V_1+V_2=V$.

Тем самым мы уже имеем соотношения

$$\begin{cases} x_1 + x_2 = A, \\ \dfrac{x_1}{p_1} + \dfrac{x_2}{p_2} = V. \end{cases}$$

Если математика нас научила не только писать, но и решать системы уравнений вида

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1, \\ a_{21}x_1 + a_{22}x_2 = b_2, \end{cases}$$

то мы найдём неизвестные нам x_1, x_2, выполним государственный заказ, получим некоторое вознаграждение для продолжения существования, сможем отвечать на другие подобные вопросы и, что, наверное, главное, получим такое наслаждение от открытия, что, не чувствуя под собой земли, полетим по Сиракузам, повторяя «Эврика! Эврика!»

(Учёные, как правило, народ свободолюбивый, но готовы жизнь отдать за то, чтобы что-то с чем-то связать.)

---

[1]Возможно, он меньше подходил для доступной, высокопрофессиональной, завораживающе увлекательной книги светлой памяти Михаила Леоновича Гаспарова (1935—2005), но больше подходит сейчас нам.

Не поленившись обработать остальные данные эксперимента, мы найдём, что $A - B = P$, т. е. тело, погружённое в воду, теряет в весе столько, сколько весит вытесненная им вода.

Но ведь это же ЗАКОН АРХИМЕДА!

(Естественно, всё можно повторить и с другой жидкостью или даже с газом.)

Это стоит дороже венца сиракузского тирана Гиерона! Спасибо ему за задачу, побочный продукт которой затмевает вопрос.

Да, плоты, лодки, корабли плавали и до Архимеда. Но теперь мы можем проектировать корабль, можем до спуска его на воду предвидеть его грузоподъёмность. Теперь мы можем спроектировать и дирижабль для переноса по воздуху больших конструкций (при строительстве буровых, обсерваторий) в недоступных для наземного транспорта местах и т. д.

3. Мы привели этот пример, чтобы на его фоне сказать несколько слов о специфике математики как науки. Не будем пытаться здесь дать определение математики. Просто понаблюдаем за примером и констатируем кое-что лежащее уже на поверхности.

Математика позволила нам перевести вопрос на какой-то специальный язык. (Какие-то символы, равенства...) Значит, математика имеет атрибуты *языка*.

Но тут есть заметное отличие от просто перевода исходного вопроса, например, с греческого на русский или китайский. По каждому из таких переводов, с одной стороны, худо-бедно можно восстановить исходный текст и содержание исходного вопроса, а с другой стороны, при таком переводе лишь меняется написание, вопрос остаётся.

Переходя к математическому тексту, т. е. к математической записи, мы абсолютно лишаемся возможности вернуться к конкретному частному вопросу, если мы потеряли его исходный текст. Но зато мы получаем определённый математический вопрос (здесь это решение системы уравнений), который, будучи разрешённым, ответит и на наш исходный частный вопрос, и на все подобные вопросы разом.

Математики находят способы (методы) решать и системы уравнений, и многие другие задачи, которые на первый взгляд не интересны никому, кроме самих математиков. На самом же деле, подобно числу, они обслуживают огромную сферу конкретных объектов и явлений.

Итак, математика чаще всего даёт не только специальный *язык* (на котором стараются записать возникший вопрос, отбрасывая всё второстепенное), но также и *метод* решения возникающей уже чисто математической задачи.

Решив её, мы, в частности, получаем ответ и на интересующий нас специальный вопрос.

Теперь мы уже можем привести и оценить следующие высказывания:

Великая книга природы написана языком математики.

(Галилей)

Если обозначения удобны для открытий, то поразительно сокращается путь к истине.

(Лейбниц)

Математика — это искусство называть разные вещи одинаковыми именами.

(Пуанкаре)

Добавим ещё цитату из уже упоминавшейся выше книги Гаспарова:

«На могиле Архимеда по его завещанию вместо памятника было поставлено изображение цилиндра с вписанным в него шаром и начертано открытое им отношение их объёмов — 3 : 2. Полтораста лет спустя, когда в Сицилии служил знаменитый римский писатель Цицерон, он ещё видел этот памятник, забытый и заросший терновником».

Там, где раньше жил грек Архимед, уже давно не Греция. Исчезают не только могилы великих, но целые страны и цивилизации. А закон Архимеда живёт вместе с природой и мирозданием. В этом единении непреходящая ценность и прелесть истинного знания и науки.

В математике, конечно, можно усмотреть и много других сторон. Например, Ломоносов не без оснований отметил, что «математика ум в порядок приводит». Она же учит слышать аргумент и ценить истину.

4. Математика, при всей своей кажущейся абстрактности, питается задачами естествознания и щедро возвращает выращенные на этой плодородной почве плоды. Это как вдох и выдох. Нарушение этого баланса опасно и в науке, и в её преподавании. Высушенная схоластикой наука погибает.

5. И ещё кое-что в этой связи.

Типичное математическое высказывание выглядит так:

Теорема. *Если то-то, то то-то.*

Это же в другой записи: $A \Rightarrow B$ (из A следует B).

Учебники математики обычно очень тщательно исследуют фрагмент ($\Rightarrow B$), т. е. излагают по возможности подробное, логически безукоризненное доказательство того, что B действительно следует из A.

Только очень наивные и неопытные люди, в том числе и некоторые математики, могут позволить себе радоваться такой теореме, если само исходное предположение A малосодержательно, неинтересно, неестественно. Во всяком случае, надо иметь в виду, что A — абсолютно равноправный и весьма существенный неформальный элемент математической теоремы.

Поясним это, напомнив одну из редких фраз, по свидетельству очевидцев произнесённую молчаливым Гиббсом (создателем математических основ классической термодинамики и статистической механики) во время обсуждения (в Йельском университете, где он работал) роли и места математики, её аксиоматического метода, а также её взаимодействия с физикой и естествознанием.

Гиббс встал. Сказал:

«Математик может позволить себе допустить всё, что ему заблагорассудится, но физика-то здравый смысл покидать не должен».

Сел. И на этом закончил своё участие в дискуссии.

Конечно, Гиббс не без иронии выразил то, о чём через полстолетия писал выдающийся математик и мыслитель Герман Вейль:

«Построения математического ума являются одновременно и свободными, и необходимыми. Отдельный математик свободен определять свои понятия и устанавливать свои аксиомы как ему угодно. Но вопрос — заинтересует ли он своих коллег-математиков продуктами своего воображения. Мы не можем не чувствовать, что некоторые математические структуры, развившиеся благодаря совместным усилиям многих учёных, несут печать необходимости, которая не затрагивается случайностями их исторического появления. Каждый, кто созерцает зрелище современной алгебры, будет поражён этой взаимодополнительностью свободы и необходимости».

Поглощение математикой новых задач (вдох), их последующее осмысление, решение, обобщение и на этой основе построение новой абстрактной математической теории (выдох) — естественный замкнутый цикл работы этой науки. В одни исторические периоды доминирует процесс накопления фактического материала, в другие периоды — подведение итогов и распределение всего по местам[1] или логическое упорядочение и формализация[2]. Более того, порой это можно наблюдать и в отдельных областях, и даже в творчестве одного и того же крупного математика (который в одни периоды мог делать и проповедовать одно, а в другие периоды — другое; в этом даже нет лицемерия, если, конечно, не отрицать сам факт).

6. Этапное продвижение в науке часто осуществляется (или, лучше сказать, оформляется) интересным и характерным способом, особенно ярко проявляющимся в таких её абстрактных областях, как теоретическая физика и математика.

Представьте себе песочные часы. Чтобы они работали, их время от времени переворачивают «с ног на голову».

В математике так же. Сначала получат много новых интересных фактов. Обнаружат среди них что-то в том или ином отношении центральное, узловое, связывающее много прежнего. Принимают это за исходный принцип, переворачивая всё с ног на голову (например, сделав теорему аксиомой), и продолжают развитие, опираясь уже на этот новый принцип, объемлющий бо́льшую сферу фактов математики и мироздания.

Например, законы Ньютона выросли из открытий Галилея и Кеплера. Но, приняв законы Ньютона в качестве основы, мы можем получить из них и законы Кеплера, и многое другое. Последующее развитие физики привело к новым, уже вариационным принципам механики, включающим в себя бо́льший круг явлений и взаимодействий, отличных от взаимодействий, описываемых центральными силами.

В некотором смысле в такие моменты переворота происходит смена масштаба. Здесь она состоит как раз в смене и сокращении количества основных принципов при одновременном расширении поля объектов и явлений, которые они охватывают и объединяют.

(Кстати, о перегрузке учебных программ, как правило, говорят те, кто не уследил за своевременной сменой масштаба.

И ещё, возможно, некстати. Есть два способа богатеть: прибирать добро к рукам — война, разбой; создавать ценности — повседневная добросовестная работа.

[1]Например, на рубеже XIX—XX веков в «Энциклопедии математических наук», изданной в Германии по инициативе Ф. Кляйна.

[2]В ещё свежих многотомных трудах Н. Бурбаки.

Там, где часто говорят о величии, скорее используют первый способ.

Выдающийся голландец, учитель многих физиков, Лоренц по поводу Первой мировой войны, как свидетельствует Эйнштейн[1], скромно заметил: «Я счастлив, что принадлежу к нации, слишком маленькой для того, чтобы совершать большие глупости».

Говорят, что ещё сравнительно недавно (а возможно, и сейчас) в букварях японских детей было написано примерно следующее: «Наша страна маленькая и бедная. Нам надо много и хорошо работать, чтобы она стала богаче».

Некогда отсталая провинция России — Финляндия демонстрирует нам эффективность честного труда и уважения законов.

Индуктивный метод науки как-то солиднее и надёжнее, чем принцип «всё, сразу и даром».)

7. Теперь несколько слов о так называемой высшей математике.

Что же характерно для этой высшей математики, оформившейся, условно говоря, в трудах Ньютона и Лейбница и бурно развившейся за три минувших с тех пор столетия?

Эта математика научилась иметь дело не только с постоянными величинами, но и с *процессами в развитии*.

Возникло и постепенно оформилось фундаментальное для науки вообще понятие функциональной зависимости, или *функции*.

Скорость изменения любой переменной величины и ускорение получили адекватную математическую формулировку в терминах *производных*.

Возник новый язык и новое *исчисление* (*дифференциальное*, когда по заданной зависимости ищется относительная скорость изменения величин, и *интегральное*, когда решается обратная задача: например, по записям скорости или ускорения ищется местонахождение подвижного объекта, пусть хоть подводной лодки).

Основы этого дифференциального и интегрального исчисления, как таблица умножения, являются теперь обязательными компонентами любого естественнонаучного образования, если не образования вообще.

Поясним почему. Если

$$x = x(t)$$

[1] *Эйнштейн А.* Собрание научных трудов. Т. 4. М.: Наука, 1967. Статья «Г. А. Лоренц как творец и человек». С. 334—336.

— закон движения, т. е. зависимость координат объекта от времени, то дифференциальное исчисление позволяет найти его скорость $v = x'(t)$ и ускорение $a = x''(t)$.

Если известна масса m объекта и действующая на него сила F, то по закону Ньютона должно быть выполнено соотношение

$$mx'' = F,$$

которое является равенством, содержащим производные (в данном случае вторую производную $x''(t)$ исходной функции $x(t)$). Если нас интересует, как будет двигаться это тело под воздействием заданной силы $F(t)$, то мы будем искать неизвестную в этом случае зависимость $x = x(t)$, удовлетворяющую написанному уравнению.

Появилась необходимость исследования и решения уравнения совсем нового типа — *дифференциального уравнения*.

Это уже опять чисто математический вопрос (отвлечённый от движений планет, эволюции звёздных систем, работы ядерных реакторов, выпечки хлеба, роста банковских сбережений и микроорганизмов, страховых обязательств, популяций рыб и животных, хищников и жертв и так далее, и тому подобное...). Но вопрос, который имеет ко всему этому прямое отношение.

Итак, когда математика строит теорию, предлагающую методы решения определённого класса математических задач, она предоставляет нам аппарат, обслуживающий разом целую новую сферу конкретных явлений. Явления, возможно, были и раньше, как плоты до закона Архимеда. Но теперь мы понимаем их лучше. Точнее, мы построили математическую модель для них, которую мы понимаем и которой с помощью математического аппарата в какой-то степени умеем распоряжаться. Уже одно это, как правило, имеет такие приложения, которые с лихвой окупают щедрые затраты цивилизованного общества на мел для математиков.

Г. Вейль отмечал: «Во всех естественных науках знание основано на наблюдении. Но наблюдение лишь констатирует положение вещей. Как предвидеть будущее? Для этого наблюдение необходимо сочетать с математикой».

Без математики, разумеется, не было бы ни Ньютона, ни Максвелла, ни Эйнштейна, ни Бора..., какими мы их знаем. Значит, не было бы той цивилизации, плодами которой мы так охотно и легко ежедневно пользуемся. Чтобы это стало совсем ясно, представьте себе на миг, что мы (люди) утратили слово—язык—речь[1]. Я не обсуждаю вопрос хорошо-плохо. Я просто хотел пояснить альтернативу и место математики.

К сказанному о математике добавлю напоследок кое-что очень общее. Процитирую страничку из живой, яркой и содержательной книжки Ричарда Фейнмана «Вы, конечно, шутите, мистер Фейнман!» (М.; Ижевск: РХД, 2001. С. 298). Дело было в Швеции в день вручения Нобелевских премий. Далее идёт цитата.

«После ужина мы перешли в комнату, где завязались различные беседы. За столом сидела датская принцесса Какая-то в окружении нескольких человек. Я увидел, что около их стола есть пустой стул, и тоже присел.

Она повернулась ко мне и сказала: „О! Вы один из лауреатов Нобелевской премии. В какой области вы работаете?“

— В физике, — ответил я.

— О, ну об этом никто ничего не знает, поэтому мы не сможем об этом поговорить.

— Напротив, — ответил я. — Мы не можем говорить о физике, потому что кто-то *что-то* о ней знает. Ибо мы *можем* обсуждать только то, о чём никто ничего не знает. Мы можем говорить о погоде; можем обсуждать социальные проблемы; мы можем беседовать о психологии; можем также обсуждать международные финансовые дела, — золотые переводы мы обсуждать *не можем*, поскольку все их понимают, — таким образом, мы все можем говорить только на тему, о которой никто ничего не знает!

Я не знаю, как они это делают. Существует способ принять *ледяное* выражение лица, и она это *сделала*! Она отвернулась, чтобы побеседовать с кем-то другим.

Через некоторое время я понял, что меня полностью исключили из разговора, поэтому я встал и пошёл прочь. Посол Японии, который сидел за столом, вскочил и последовал за мной. „Профессор Фейнман, — сказал он, — есть кое-что, что мне хотелось бы рассказать вам о дипломатии“.

Он пустился в длинную историю о том, как молодой человек в Японии поступает в университет, изучает международные отно-

[1]Возможно, для кого-то убедительнее звучит потеря мобильного телефона, телевидения и прочего, про что Сократ говорил: «Как много есть вещей, без которых можно обходиться».

шения, поскольку считает, что может сделать свой вклад для блага своей страны. Перейдя на второй курс, он начинает испытывать лёгкие приступы сомнения относительно того, что изучает. По окончании колледжа он занимает свой первый пост в посольстве и испытывает ещё бо́льшие сомнения относительно своего понимания дипломатии, пока наконец не осознает, что *никто* ничего не знает о международных отношениях. И тогда он может стать послом! „Поэтому, профессор Фейнман, — сказал он, — когда в следующий раз вы будете приводить примеры того, о чём все говорят, но никто не знает, пожалуйста, включите и международные отношения!“

Он оказался очень интересным человеком, и мы продолжили разговор. Я всегда поражался тому, как по-разному развиваются разные страны и разные люди. Я сказал послу, что есть феномен, который всегда удивлял меня: то, каким образом Япония сумела так быстро достичь такой высокой степени развития, что стала играть в мире столь важную роль. „Какая черта или особенность японцев обеспечила такую возможность?“— спросил я.

Мне очень понравился ответ посла. Он сказал: „Я не знаю. Я могу только предположить, но не знаю, насколько это соответствует истине. Японцы верили, что единственный способ поднять страну — это дать своим детям лучшее образование, чем имели они сами; самым важным для них было уйти от своего положения крестьян и получить образование. Поэтому в семьях огромные усилия прикладывались к тому, чтобы поощрять детей хорошо учиться в школе, чтобы они могли чего-то достичь. Из-за этого стремления постоянно чему-то учиться, через систему образования очень легко распространялись новые идеи из внешнего мира. Быть может, это и есть одна из причин столь быстрого развития Японии“».

Непредвиденный эпилог

Английский перевод книги был уже почти завершён, когда скоропостижно скончался Владимир Игоревич Арнольд. Не знаю, кто будет читать эту книгу, но одного, причём самого яркого, тщательного и профессионального её читателя, которому с дружеской улыбкой явно или неявно были адресованы отдельные места текста, книга лишилась, как лишились и все мы, математики.

О В. И. Арнольде, конечно, ещё будет сказано и написано много. Его научные заслуги оценит объективное время и следующие поколения. Но следующим поколениям не дано быть современниками. Им не будет дано понять, что потеря Арнольда для математики, во всяком случае в России, особенно при нынешнем положении науки, близка к потере среды обитания.[1]

Добавление 2017 г. В самом конце этого 2017 года состоится большая международная конференция, приуроченная к 80-летию со дня рождения Владимира Игоревича Арнольда, а в самом начале этого года в автомобильной катастрофе погиб Виталий Дмитриевич Арнольд — племянник Владимира Игоревича, неуёмной энергии человек, энтузиаст математического просвещения, один из основных организаторов знаменитой теперь ежегодной летней математической школы «Современная математика», где ведущие профессионалы рассказывают на доступном студентам и старшим школьникам уровне о проблемах, которыми они занимаются. В этом году школа посвящена памяти Виталия Арнольда.

[1]Этот эпилог впервые был напечатан (на английском языке) в английском издании этой книги, вышедшем в 2011 году (*Zorich V.* Mathematical Analysis of Problems in the Natural Sciences).

Магазин «Математическая книга»

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга» в Москве по адресу: Б. Власьевский пер., д. 11; тел. (499) 241-72-85; biblio.mccme.ru
Книга — почтой: http://biblio.mccme.ru/shop/order
Книги в электронном виде: http://www.litres.ru/mcnmo/

Мы сотрудничаем с интернет-магазинами

- Книготорговая компания «Абрис»; тел. (495) 229-67-59, (812) 327-04-50; www.umlit.ru, www.textbook.ru, абрис.рф
- Интернет-магазин «Книга.ру»; тел. (495) 744-09-09; www.kniga.ru

Наши партнеры в Москве и Подмосковье

- Московский Дом Книги и его филиалы (работает интернет-магазин); тел. (495) 789-35-91; www.mdk-arbat.ru
- Магазин «Молодая Гвардия» (работает интернет-магазин): ул. Б. Полянка, д. 28; тел. (499) 238-50-01, (495) 780-33-70; www.bookmg.ru
- Магазин «Библио-Глобус» (работает интернет-магазин): ул. Мясницкая, д. 6/3, стр. 1; тел. (495) 781-19-00; www.biblio-globus.ru
- Спорткомплекс «Олимпийский», 5-й этаж, точка 62; тел. (903) 970-34-46
- Сеть киосков «Аргумент» в МГУ; тел. (495) 939-21-76, (495) 939-22-06; www.arg.ru
- Сеть магазинов «Мир школьника» (работает интернет-магазин); тел. (495) 715-31-36, (495) 715-59-63, (499) 182-67-07, (499) 179-57-17; www.uchebnik.com
- Сеть магазинов «Шаг к пятерке»; тел. (495) 728-33-09, (495) 346-00-10; www.shkolkniga.ru
- Издательская группа URSS, Нахимовский проспект, д. 56, Выставочный зал «Науку — Всем», тел. (499) 724-25-45, www.urss.ru
- Книжный магазин издательского дома «Интеллект» в г. Долгопрудный: МФТИ (новый корпус); тел. (495) 408-73-55

Наши партнеры в Санкт-Петербурге

- Санкт-Петербургский Дом книги: Невский пр-т, д. 62; тел. (812) 314-58-88
- Магазин «Мир науки и медицины»: Литейный пр-т, д. 64; тел. (812) 273-50-12
- Магазин «Новая техническая книга»: Измайловский пр-т, д. 29; тел. (812) 251-41-10
- Информационно-книготорговый центр «Академическая литература»: Васильевский остров, Менделеевская линия, д. 5
- Киоск в здании физического факультета СПбГУ в Петергофе; тел. (812) 328-96-91, (812) 329-24-70, (812) 329-24-71
- Издательство «Петроглиф»: Фарфоровская, 18, к. 1; тел. (812) 560-05-98, (812) 943-80-76; k_i_@bk.ru, k_i_@petroglyph.ru
- Сеть магазинов «Учебная литература»; тел. (812) 746-82-42, тел. (812) 764-94-88, тел. (812) 235-73-88 (доб. 223)

Наши партнеры в Челябинске

- Магазин «Библио-Глобус», ул. Молдавская, д. 16, www.biblio-globus.ru

Наши партнеры в Украине

- Александр Елисаветский. Рассылка книг наложенным платежом по Украине: тел. 067-136-37-35; df-al-el@bk.ru

В. А. Зорич

Математический анализ задач естествознания

ISBN 978-5-4439-1225-7
9 785443 912257 >