А.И.Кострикин, Ю.И.Манин

ЛИНЕЙНАЯ АЛГЕБРА И ГЕОМЕТРИЯ

Книга посвящена изложению фундаментальных понятий и аппарата линейной алгебры и родственных ей разделов геометрии. От имеющихся курсов линейной алгебры книга отличается большим вниманием к приложениям и связям с другими областями математики: включено обсуждеиие основных принципов квантовой механики, описана геометрия пространства Минковского, дано введеиие в линейное программирование. Книга содержит современный математический материал, не излагавшийся в традиционных руководствах: язык категорий и категорные свойства линейных пространств, кэлерова метрика, введеиие в теорию многочленов Гильберта.

Для студентов механико-математических специальностей высших учебных заведений.

ОГЛАВЛЕНИЕ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ПРЕДИСЛОВИЕ

На все можно смотреть с разных точек зрения, предмет этой книги — не исключение. Для студента мехмата МГУ линейная алгебра — это то, что читают первокурсникам во втором семестре. Для профессионала-алгебраиста, воспитанного в духе Бурбаки, линейная алгебра — это теория алгебраических структур частного вида — линейных пространств и линейных отображений, или, в более новомодном стиле, теория линейных категорий.

С более широкой точки зрения, содержание линейной алгебры состоит в проработке математического языка для выражения одной из самых общих естественнонаучных идей — идеи линейности. Возможно, ее важнейшим специальным случаем является принцип линейности малых приращений: почти всякий естественный процесс почти всюду в малом линеен. Этот принцип лежит в основе всего математического анализа и его приложений. Векторная алгебра трехмерного физического пространства, исторически ставшая краеугольным камнем в здании линейной алгебры, восходит к тому же источнику: после Эйнштейна мы понимаем, что и физическое пространство приближенно линейно лишь в малой окрестности наблюдателя. К счастью, эта малая окрестность довольно велика.

Физика двадцатого века резко и неожиданно расширила сферу применения идеи линейности, добавив к принципу линейности малых приращений принцип суперпозиции векторов состояний. Грубо говоря, пространство состояний любой квантовой системы является линейным пространством над полем комплексных чисел. В результате почти все конструкции комплексной линейной алгебры превратились в аппарат, используемый для формулировки фундаментальных законов природы: от теории линейной двойственности, объясняющей квантовый принцип дополнительности Бора, до теории представлений групп, объясняющей таблицу Менделеева, «зоологию» элементарных частиц и даже структуру пространства-времени.

Выбор материала для этого курса определялся желанием авторов не только изложить основы аппарата, почти завершенного к началу этого века, но и дать представление о его приложениях, обычно относимых к другим дисциплинам. Традиции преподавания способствуют рассечению живого тела математики на изолированные органы, жизнеспособность которых приходится поддерживать искусственно. Особенно это относится к «критическим периодам» в истории нашей науки, которые характерны вниманием к логической стройности и детальной проработке оснований. Последние полвека были временем теоретико-множественной перестройки языка и фундаментальных понятий; единство математики стало рассматриваться преимущественно как единство ее логических принципов. Не отказываясь от замечательных достижений этого периода, мы хотели отразить в книге и намечающиеся тенденции к синтезу математики как орудия познания внешнего мира. (К сожалению, нам пришлось оставить в стороне теорию вычислительных аспектов линейной алгебры, выросшую в самостоятельную науку.)

По этим соображениям в предлагаемую книгу, как и во «Введение в алгебру» одного из авторов, включен не только материал для лекционного курса, но и разделы для домашнего чтения, которые могут быть использованы также на семинарских занятиях. Жесткого разделения здесь не может быть. Все же в соответствии со стандартной программой лекционный курс (один семестр, по две лекции в неделю) должен включать основной материал § 1—9 части 1; § 2—8 части 2; § 1, 3, 5, 6 части 3 и § 1, 3—6 части 4. При этом под «основным материалом» мы понимаем не столько доказательства трудных теорем (которых в линейной алгебре немного), сколько систему понятий, которыми следует овладеть. Поэтому многие теоремы из этих разделов могут быть рассказаны в более простом варианте или вовсе опущены; по недостатку времени такие сокращения неизбежны. Как избежать при этом превращения лекций в унылый список определений, составляет серьезную заботу преподавателя. Мы надеемся, что остальные разделы курса помогут ему в этом.

При переиздании внесены исправления опечаток и мелких стилистических погрешностей, сообщенных нам рядом читателей. В двух местах пришлось корректировать доказательства. Многочисленные ценные замечания были сделаны сотрудниками кафедры алгебры и теории чисел Ленинградского государственного университета. Конструктивная критика рецензентов, а также высококвалифицированная, тщательная работа редактора издания В. Л. Попова во многом способствовали улучшению качества книги. Мы выражаем им всем глубокую благодарность.

Все оставшиеся недочеты в книге авторы, разумеется, относят на свой счет.

СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ

1. *Кострикин А. И.* Введение в алгебру. — М.: Наука, 1977.
2. *Ленг С.* Алгебра. — М.: Мир, 1968.
3. *Мальцев А. И.* Основы линейной алгебры. — М.: Наука, 1975.
4. *Гельфанд И. М.* Лекции по линейной алгебре. — М.: Наука, 1966.
5. *Халмош П. Р.* Конечномерные векторные пространства. — М.: Физматгиз, 1963.
6. *Артин Э.* Геометрическая алгебра. — М.: Мир, 1970.
7. *Глазман И. М., Любич Ю. И.* Конечномерный линейный анализ. — М.: Наука, 1969.
8. *Воеводин В. В.* Вычислительные основы линейной алгебры. — М.: Наука, 1977.

Часть 1. ЛИНЕЙНЫЕ ПРОСТРАНСТВА И ЛИНЕЙНЫЕ ОТОБРАЖЕНИЯ

§ 1. Линейные пространства

1. Векторы с началом в выбранной точке пространства можно умножать на числа и складывать по правилу параллелограмма. Это — классическая модель законов сложения перемещений, скоростей, сил в механике. В общем определении векторного, или линейного, пространства вещественные числа заменяются произвольным полем, а простейшие свойства сложения и умножения векторов постулируются в качестве аксиом. Никаких следов «трехмерности» физического пространства в определении не остается. Понятие размерности вводится и изучается отдельно.

Из курса аналитической геометрии на плоскости и в трехмерном пространстве известно много примеров геометрической интерпретации алгебраических соотношений между двумя или тремя переменными. Но, по выражению Н. Бурбаки, «...ограничение геометрическим языком, отвечающим пространству только трех измерений, было бы ярмом для современного математика, столь же неудобным, как то, которое мешало грекам распространить понятие числа на отношения несоизмеримых величин...».

2. Определение. *Линейным* (или *векторным*) *пространством* L *над полем* $\mathscr{K}$ *называется множество, снабженное бинарной операцией* $L \times L \to L$, *обычно обозначаемой как сложение:* $(l_1, l_2) \mapsto l_1 + l_2$, *и внешней бинарной операцией* $\mathscr{K} \times L \to L$, *обычно обозначаемой как умножение:* $(a, l) \mapsto al$, *которые удовлетворяют следующим аксиомам:*

а) *Сложение элементов* L, *или векторов, превращает* L *в коммутативную (абелеву) группу. Ее нулевой элемент обычно обозначается* 0; *элемент, обратный к* l, *обычно обозначается* $-l$.

б) *Умножение векторов на элементы поля* $\mathscr{K}$, *или скаляры, унитарно, т. е.* $1l = l$ *для всех* l, *и ассоциативно, т. е.* $a(bl) = (ab)l$ *для всех* $a, b \in \mathscr{K}$; $l \in L$.

в) *Сложение и умножение связаны законами дистрибутивности, т. е.*

$$a(l_1 + l_2) = al_1 + al_2, \qquad (a_1 + a_2)l = a_1 l + a_2 l$$

для всех $a, a_1, a_2 \in \mathscr{K}$; $l, l_1, l_2 \in L$.

3. Вот некоторые простейшие следствия определения.

а) $0l = a0 = 0$ для всех $a \in \mathscr{K}$, $l \in L$. Действительно, $0l + 0l = (0 + 0)l = 0l$, откуда $0l = 0$ по свойству сокращения в абелевой группе. Аналогично, $a0 + a0 = a(0 + 0) = a0$, т. е. $a0 = 0$.

б) $(-1)l=-l$. Действительно, $l+(-1)l=1l+(-1)l=(1+(-1))l=0l=0$, так что вектор $(-1)l$ обратен к l.

в) Если $al=0$, то либо $a=0$, либо $l=0$. В самом деле, если $a\neq 0$, то $0=a^{-1}(al)=(a^{-1}a)l=1l=l$.

г) Для любых $a_1, \ldots, a_n\in\mathcal{K}$; $l_1, \ldots, l_n\in L$ однозначно определено выражение $a_1l_1+\ldots+a_nl_n=\sum\limits_{i=1}^{n}a_il_i$: благодаря ассоциативности сложения в абелевой группе можно не расставлять скобки, указывающие порядок вычисления попарных сумм. Аналогично, однозначно определено выражение $a_1a_2\ldots a_nl$.

Выражение вида $\sum\limits_{i=1}^{n}a_il_i$ называется *линейной комбинацией* векторов $l_1, \ldots, l_n$; скаляры a_i — коэффициенты этой линейной комбинации.

Следующие примеры линейных пространств будут постоянно встречаться в дальнейшем.

4. Нульмерное пространство. Это — абелева группа $L=\{0\}$, состоящая из одного нуля. Единственно возможный закон умножения на скаляры: $a0=0$ для всех $a\in\mathcal{K}$ (убедитесь в справедливости аксиом!).

Предостережение: нульмерные пространства над *разными полями* — это *разные пространства*: задание поля $\mathcal{K}$ входит в определение линейного пространства.

5. Основное поле $\mathcal{K}$ как одномерное координатное пространство. Здесь $L=\mathcal{K}$, сложение — это сложение в $\mathcal{K}$, умножение на скаляры — это умножение в $\mathcal{K}$. Справедливость аксиом линейного пространства следует из аксиом поля.

Более общо, если имеется поле K и его подполе $\mathcal{K}$, то K можно рассматривать как линейное пространство над $\mathcal{K}$. Например, поле комплексных чисел $\mathbf{C}$ является линейным пространством над полем вещественных чисел $\mathbf{R}$, которое в свою очередь является линейным пространством над полем рациональных чисел $\mathbf{Q}$.

6. n-мерное координатное пространство. Положим $L=\mathcal{K}^n=\mathcal{K}\times\ldots\times\mathcal{K}$ (декартово произведение $n\geqslant 1$ множителей). Элементы L можно записывать в виде строк $(a_1, \ldots, a_n)$, $a_i\in\mathcal{K}$, или столбцов высоты n. Определим сложение и умножение на скаляр формулами:

$$(a_1, \ldots, a_n)+(b_1, \ldots, b_n)=(a_1+b_1, \ldots, a_n+b_n),$$
$$a(a_1, \ldots, a_n)=(aa_1, \ldots, aa_n).$$

При $n=1$ получается предыдущий пример. Одномерные пространства над $\mathcal{K}$ называют прямыми, или $\mathcal{K}$-прямыми; двумерные — $\mathcal{K}$-плоскостями.

7. Пространства функций. Пусть S — произвольное множество, $F(S)$ — множество функций на S со значениями в $\mathcal{K}$, или отображений S в $\mathcal{K}$. Как обычно, если $f\colon S\to\mathcal{K}$ — такая функция, то через $f(s)$ обозначается значение f на элементе $s\in S$.

Сложение и умножение функций на скаляр определяется поточечно:

$$(f+g)(s)=f(s)+g(s) \text{ для всех } s\in S,$$
$$(af)(s)=a(f(s)) \text{ для всех } a\in\mathcal{K};\ s\in S.$$

Если $S=\{1,\ \dots,\ n\}$, то $F(S)$ можно отождествить с $\mathcal{K}^n$: функции f ставится в соответствие «вектор» всех ее значений $(f(1),\ \dots,\ f(n))$. Правила сложения и умножения согласованы относительно такого отождествления.

Каждому элементу $s\in S$ можно поставить в соответствие важную «дельта-функцию δ_s, сосредоточенную на $\{s\}$», которая определяется так: $\delta_s(s)=1$, $\delta_s(t)=0$, если $t\neq s$. Если $S=\{1,\ \dots,\ n\}$, вместо $\delta_i(k)$ обычно пишут δ_{ik} — это *символ Кронекера*.

Если множество S конечно, то всякую функцию из $F(S)$ можно однозначно представить в виде линейной комбинации дельта-функций: $f=\sum\limits_{s\in S} f(s)\delta_s$. В самом деле, это равенство следует из совпадения значений левой и правой части в каждой точке $s\in S$. Наоборот, если $f=\sum\limits_{s\in S} a_s\delta_s$, то, беря значение в точке s, получаем $f(s)=a_s$.

Если множество S бесконечно, то этот результат неверен, точнее говоря, не может быть сформулирован в рамках наших определений: суммы бесконечного числа векторов в общем линейном пространстве не определены! Некоторые бесконечные суммы можно определить в линейных пространствах, снабженных понятием предельного перехода, или топологией (см. § 10). Такие пространства составляют основной предмет изучения в функциональном анализе.

В случае $S=\{1,\ \dots,\ n\}$ функция δ_i представлена вектором $e_i=(0,\ \dots,\ 0, 1, 0,\ \dots,\ 0)$ (единица на i-м месте, нули на остальных), а равенство $f=\sum\limits_{s\in S} f(s)\,\delta_s$ превращается в равенство

$$(a_1,\ \dots,\ a_n)=\sum_{i=1}^{n} a_ie_i.$$

8. Линейные условия и линейные подпространства. В анализе прежде всего рассматриваются вещественнозначные функции, определенные на всем $\mathbf{R}$ или интервалах $(a, b)\subset\mathbf{R}$. Для большинства приложений, однако, пространство всех таких функций слишком велико: полезно рассматривать непрерывные или дифференцируемые функции. После введения соответствующих определений обычно доказывается, что сумма непрерывных функций непрерывна и произведение непрерывной функции на скаляр непрерывно; то же для дифференцируемости.

Это означает, что только непрерывные или только дифференцируемые функции сами по себе образуют линейное пространство.

Более общо, пусть L — линейное пространство над полем $\mathcal{K}$, а $M\subset L$ — его подмножество, которое является подгруппой и которое переходит в себя при умножении на скаляры. Тогда M вместе с операциями, индуцированными операциями в L (другими словами, ограничениями на M операций, определенных в L), называется *линейным подпространством* в L, а условия, определяющие принадлежность к M общего вектора из L, называются *линейными условиями.*

Вот пример линейных условий в координатном пространстве $\mathcal{K}^n$: фиксируем скаляры $a_1, \ldots, a_n \in \mathcal{K}$ и определим $M \subset L$:

$$(x_1, \ldots, x_n) \in M \Leftrightarrow \sum_{i=1}^{n} a_i x_i = 0. \tag{1}$$

Объединение любого числа линейных условий также является линейным условием. Иными словами, пересечение любого числа линейных подпространств также является линейным подпространством (проверьте это!). Позже мы докажем, что в $\mathcal{K}^n$ любое подпространство описывается конечным числом условий вида (1).

Важный пример линейного условия дает следующая конструкция.

9. Двойственное линейное пространство. Пусть L — линейное пространство над $\mathcal{K}$. Рассмотрим сначала линейное пространство $F(L)$ *всех* функций на L со значениями в $\mathcal{K}$. Назовем теперь функцию $f \in F(L)$ *линейной* (иногда говорят «*линейный функционал*»), если она удовлетворяет условиям

$$f(l_1 + l_2) = f(l_1) + f(l_2), \quad f(al) = af(l)$$

для всех $l, l_1, l_2 \in L$, $a \in \mathcal{K}$. Индукцией по числу слагаемых отсюда получаем, что

$$f\left(\sum_{i=1}^{n} a_i l_i\right) = \sum_{i=1}^{n} a_i f(l_i).$$

Мы утверждаем, что *линейные функции образуют линейное подпространство в* $F(L)$, или «*условие линейности является линейным условием*». В самом деле, если f, f_1 и f_2 линейны, то

$$(f_1 + f_2)(l_1 + l_2) = f_1(l_1 + l_2) + f_2(l_1 + l_2) =$$
$$= f_1(l_1) + f_1(l_2) + f_2(l_1) + f_2(l_2) = (f_1 + f_2)(l_1) + (f_1 + f_2)(l_2).$$

(Здесь последовательно используются: правило сложения функций, линейность f_1 и f_2, коммутативность и ассоциативность сложения в поле и опять правило сложения функций.) Аналогично,

$$(af)(l_1 + l_2) = a[f(l_1 + l_2)] = a[f(l_1) + f(l_2)] =$$
$$= a[f(l_1)] + a[f(l_2)] = (af)(l_1) + (af)(l_2).$$

Таким образом, $f_1 + f_2$ и af также линейны.

Пространство линейных функций на линейном пространстве L называется двойственным, или сопряженным к L пространством, и обозначается L^*.

В дальнейшем мы встретимся со многими другими конструкциями линейных пространств.

10. Замечания относительно обозначений. Обозначать нуль и сложение в $\mathscr{K}$ и L одинаковыми значками не вполне последовательно, но очень удобно. Все формулы обычной школьной алгебры, которые осмысленны в этой ситуации, оказываются верными (см. образцы в п. 3).

Вот два примера, когда предпочтительнее другие обозначения.

а) Положим $L = \{x \in \mathbf{R} \mid x > 0\}$. Рассмотрим L как абелеву группу *по умножению* и введем на L умножение на скаляры из $\mathbf{R}$ по формуле $(a, x) \mapsto x^a$. Легко проверить, что все условия определения п. 2 выполнены, хотя принимают в обычной записи другой вид: нулевой вектор в L есть 1, вместо $1l = l$ мы имеем $x^1 = x$; вместо $a(bl) = (ab)l$ — тождество $(x^b)^a = x^{ba}$; вместо $(a+b)l = al + bl$ — тождество $x^{a+b} = x^a x^b$ и т. д.

б) Пусть L — векторное пространство над полем комплексных чисел $\mathbf{C}$. Определим новое векторное пространство $\bar{L}$ с той же аддитивной группой L, но другим законом умножения на скаляры:

$$(a, l) \mapsto \bar{a} l,$$

где $\bar{a}$ — комплексно сопряженное число к a. Из формул $\overline{a+b} = \bar{a} + \bar{b}$ и $\overline{ab} = \bar{a}\bar{b}$ без труда следует, что $\bar{L}$ — векторное пространство. Если в какой-то ситуации нам приходится рассматривать одновременно L и $\bar{L}$, то может оказаться удобно писать вместо $\bar{a}l$, скажем, $a * l$ или $a \circ l$.

11. Замечания о чертежах и наглядных образах. Очень многие общие понятия и теоремы линейной алгебры удобно иллюстрировать схематическими чертежами и картинками. Мы хотим сразу же предупредить читателя о некоторых опасностях таких изображений.

а) *Малая размерность.* Мы живем в трехмерном пространстве, и наши чертежи изображают обычно двух- или трехмерные образы. В линейной алгебре работают с пространствами любой конечной размерности, а в функциональном анализе — с бесконечномерными. Наша «маломерная» интуиция поддается очень серьезному развитию, но развивать ее нужно сознательно.

Простой пример: как представить себе общее расположение двух плоскостей в четырехмерном пространстве? Вообразите две пересекающиеся по прямой плоскости в $\mathbf{R}^3$, которые отрываются вдоль этой прямой всюду, кроме начала координат, расходясь в четвертое измерение.

б) *Вещественное поле.* Физическое пространство $\mathbf{R}^3$ линейно над вещественным полем. Непривычность геометрии линейного пространства над $\mathscr{K}$ может быть связана со свойствами поля $\mathscr{K}$.

Например, пусть $\mathscr{K} = \mathbf{C}$ (важнейший для квантовой механики случай). Прямая над $\mathbf{C}$ — это одномерное координатное пространство $\mathbf{C}^1$. Мы привыкли, что умножение точек прямой $\mathbf{R}^1$ на вещественное число a есть растяжение в a раз (при $a > 1$), сжатие в a^{-1} раз (при $0 < a < 1$) или их комбинация с «переворачиванием» прямой (при $a < 0$).

Но умножение на комплексное число a, действующее на $\mathbf{C}^1$, естественно представлять себе при геометрическом изображении $\mathbf{C}^1$ в виде $\mathbf{R}^2$ («плоскость Аргана» или «комплексная плоскость» — не путать с $\mathbf{C}^2$!). При этом изображении числу $z=x+iy\in\mathbf{C}^1$ отвечает точка $(x, y)\in\mathbf{R}^2$, а умножение на $a\neq 0$ соответствует растяжению в $|a|$ раз и повороту на угол $\arg a$ против часовой стрелки. Мы видим, в частности, что при $a=-1$ вещественное «переворачивание» прямой $\mathbf{R}^1$ есть ограничение на $\mathbf{R}^1$ поворота $\mathbf{C}^1$ на $180°$.

Вообще, n-мерное комплексное пространство $\mathbf{C}^n$ можно, и часто полезно, представлять себе как $2n$-мерное вещественное пространство $\mathbf{R}^{2n}$ (ср. § 12 о комплексификации и овеществлении).

Другим примером являются конечные поля $\mathcal{K}$, в частности поле из двух элементов $F_2=\{0, 1\}$, важное в теории кодирования. Здесь конечномерные координатные пространства конечны, и иногда удобно связывать с линейной геометрией над $\mathcal{K}$ дискретные образы. Например, F_2^n часто отождествляют с вершинами n-мерного единичного куба в $\mathbf{R}^n$ — множеством точек $(\varepsilon_1, \ldots, \varepsilon_n)$, где $\varepsilon_i=0$ или 1. Покоординатное сложение в F_2^n — это операции Буля: $1+0=0+1=1$; $0+0=1+1=0$. Подпространство, состоящее из точек с $\varepsilon_1+\ldots+\varepsilon_n=0$, определяет простейший код с обнаружением ошибок. Условившись, что точки $(\varepsilon_1, \ldots, \varepsilon_n)$ кодируют сообщения только при $\sum_{i=1}^{n}\varepsilon_i=0$, и приняв сигнал $(\varepsilon_1', \ldots, \varepsilon_n')$ с $\sum_{i=1}^{n}\varepsilon_i'\neq 0$, мы можем быть уверены, что помехи при передаче привели к ошибочному приему.

в) *Физическое пространство евклидово.* Это значит, что в нем определены не только сложение векторов и умножение на скаляр, но также длины векторов, углы между ними, площади и объемы некоторых фигур и т. п. Наши чертежи несут принудительную информацию об этих «метрических» свойствах, и мы их машинально воспринимаем, хотя в общей аксиоматике линейных пространств они никак не отражены. Нельзя представлять себе, что один вектор короче другого, или что пара векторов образует прямой угол, до тех пор, пока пространство не наделено специальной дополнительной структурой, скажем, абстрактным скалярным произведением. Таким структурам посвящена вторая часть книги.

УПРАЖНЕНИЯ

1. Образуют ли линейное пространство над $\mathbf{Q}$ следующие множества вещественных чисел:

а) положительные вещественные числа;
б) неотрицательные вещественные числа;
в) целые числа;
г) рациональные числа со знаменателем $\leqslant N$;
д) числа вида $a+b\pi$, где a, b — любые рациональные числа?

2. Пусть S — некоторое множество, $F(S)$ — пространство функций со значениями в поле $\mathcal{K}$. Какие из следующих условий являются линейными:

а) f обращается в нуль в данной точке S;

б) f обращается в единицу в данной точке S;

в) f обращается в нуль во всех точках подмножества $S_0 \subset S$;

г) f обращается в нуль хотя бы в одной точке подмножества $S_0 \subset S$;

д) $f(x) \to 0$ при $|x| \to \infty$;

е) $f(x) \to 1$ при $|x| \to \infty$;

ж) f имеет не более конечного числа точек разрыва

(в д) — ж) предполагаем $S = \mathbf{R}$ и $\mathscr{K} = \mathbf{R}$)?

3. Пусть L — линейное пространство непрерывных вещественных функций на отрезке $[-1, 1]$. Какие из следующих функционалов на L являются линейными:

а) $f \longmapsto \int\limits_{-1}^{1} f(x)\,dx$;

б) $f \longmapsto \int\limits_{-1}^{1} f^2(x)\,dx$;

в) $f \longmapsto f(0)$ (это — «дельта-функционал Дирака»);

г) $f \longmapsto \int\limits_{-1}^{1} f(x)\,g(x)\,dx$, где g — фиксированная непрерывная функция на $[-1, 1]$?

4. Пусть $L = \mathscr{K}^n$. Какие из следующих условий на $(x_1, \ldots, x_n) \in L$ являются линейными:

а) $\sum\limits_{i=1}^{n} a_i x_i = 1$; $a_1, \ldots, a_n \in \mathscr{K}$;

б) $\sum\limits_{i=1}^{n} x_i^2 = 0$ (разберите отдельно случаи: $\mathscr{K} = \mathbf{R}$, $\mathscr{K} = \mathbf{C}$, $\mathscr{K}$ — поле из двух элементов, или, более общо, поле характеристики два);

в) $x_3 = 2x_4$?

5. Пусть $\mathscr{K}$ — конечное поле из q элементов. Сколько элементов имеется в линейном пространстве $\mathscr{K}^n$? Сколько решений есть у уравнения $\sum\limits_{i=1}^{n} a_i x_i = 0$?

6. Пусть $\mathscr{K}^\infty$—пространство бесконечных последовательностей $(a_1, a_2, a_3 \ldots)$, $a_i \in \mathscr{K}$, с покоординатным сложением и умножением. Какие из следующих условий на векторы из $\mathscr{K}^\infty$ являются линейными:

а) только конечное число координат a_i отлично от нуля;

б) только конечное число координат a_i равно нулю;

в) среди координат a_i никакая не равна 1;

г) условие Коши: для каждого $\varepsilon > 0$ существует такое $N > 0$, что $|a_m - a_n| < \varepsilon$ при $m, n > N$;

д) условие Гильберта: ряд $\sum\limits_{n=1}^{\infty} |a_n|^2$ сходится;

е) (a_i) образуют ограниченную последовательность, т. е. существует такая константа c, зависящая от (a_i), что $|a_i| < c$ для всех i (в г) — е) предполагаем $\mathscr{K} = \mathbf{R}$ или $\mathbf{C}$)?

7. Пусть S — конечное множество. Докажите, что каждый линейный функционал на $F(S)$ однозначно определяется семейством элементов $\{a_s | s \in S\}$ поля $\mathscr{K}$: функции f ставится в соответствие скаляр $\sum\limits_{s \in S} a_s f(s)$.

Если n — число элементов в S и $a_s = 1/n$ для всех s, мы получаем функционал $f \mapsto \frac{1}{n} \sum\limits_{s \in S} f(s)$ — среднее арифметическое значений функции.

Если $\mathscr{K}=\mathbf{R}$ и. $a_s \geqslant 0$, $\sum_{s \in S} a_s = 1$, функционал $\sum_{s \in S} a_s f(s)$ называется *взвешенным средним* функции f (с весами a_s).

§ 2. Базис и размерность

1. Определение. *Семейство векторов $\{e_1, \ldots, e_n\}$ в линейном пространстве L называется (конечным) базисом L, если каждый вектор из L однозначно представляется в виде линейной комбинации $l = \sum_{i=1}^{n} a_i e_i$, $a_i \in \mathscr{K}$. Коэффициенты a_i называются координатами вектора l относительно базиса $\{e_i\}$.*

2. Примеры. а) Векторы $e_i = (0, \ldots, 1, \ldots, 0)$, $1 \leqslant i \leqslant n$, образуют базис $\mathscr{K}^n$. б) Если множество S конечно, функции $\delta_s \in F(S)$ образуют базис $F(S)$. Оба эти утверждения были проверены в § 1.

Если в L выбран базис из n векторов и каждый вектор задается своими координатами в этом базисе, то сложение и умножение на скаляр выполняются покоординатно: $\sum_{i=1}^{n} a_i e_i + \sum_{i=1}^{n} b_i e_i = \sum_{i=1}^{n} (a_i + b_i) e_i$, $a \sum_{i=1}^{n} a_i e_i = \sum_{i=1}^{n} a a_i e_i$. Поэтому выбор базиса равносилен отождествлению L с координатным векторным пространством. Вместо равенства $l = \sum_{i=1}^{n} a_i e_i$ иногда пишут $l = \vec{a}$, подразумевая под $\vec{a}$ вектор-столбец

$$[a_1, \ldots, a_n] = \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

или вектор-строку $(a_1, \ldots, a_n) = [a_1, \ldots, a_n]^t$ координат $a_1, \ldots$ $\ldots, a_n$; в этих обозначениях явное указание базиса опущено.

3. Определение. *Пространство L называется конечномерным, если оно либо нульмерно* (см. § 1, п. 4), *либо имеет конечный базис. Остальные пространства называются бесконечномерными.*

Удобно считать, что базис нульмерного пространства образует пустое множество векторов. Поскольку для нульмерных пространств все наши утверждения тривиализируются, мы обычно будем ограничиваться рассмотрением непустых базисов.

4. Теорема. *В конечномерном пространстве число элементов базиса не зависит от базиса.*

Это число называется размерностью пространства L и обозначается $\dim L$ или $\dim_{\mathscr{K}} L$. Если $\dim L = n$, пространство L называется n-мерным. В бесконечномерном случае мы пишем $\dim L = \infty$.

Доказательство. Пусть $\{e_1, \ldots, e_n\}$ — некоторый базис L. Мы докажем, что никакое семейство векторов $\{e'_1, \ldots, e'_m\}$ с $m > n$ не может служить базисом L по следующей причине: существует представление нулевого вектора $0=\sum_{i=1}^{m} x_i e'_i$, в котором не все x_i равны нулю. Поэтому 0 не однозначно представляется в виде линейной комбинации векторов $\{e'_i\}$: всегда существует тривиальное представление $0=\sum_{i=1}^{m} 0e'_i$.

Отсюда уже следует полное утверждение теоремы, поскольку этим мы проверим, что никакой базис не может содержать больше элементов, чем другой базис.

Положим $e'_k=\sum_{i=1}^{n} a_{ik}e_i$, $k=1, \ldots, m$. Для любых $x_k \in \mathcal{K}$ имеем

$$\sum_{k=1}^{m} x_k e'_k=\sum_{k=1}^{m} x_k \sum_{i=1}^{n} a_{ik} e_i=\sum_{i=1}^{n}\left(\sum_{k=1}^{m} a_{ik} x_k\right) e_i.$$

Поскольку $\{e_i\}$ образуют базис в L, нулевой вектор имеет единственное представление $\sum_{k=1}^{n} 0e_k$ в виде линейной комбинации $\{e_k\}$. Поэтому условие $\sum_{k=1}^{m} x_k e'_k=0$ равносильно системе однородных линейных уравнений относительно x_k:

$$\sum_{k=1}^{m} a_{ik} x_k=0, \quad i=1, \ldots, n.$$

Поскольку число неизвестных m больше числа уравнений n, эта система имеет ненулевое решение. Теорема доказана.

5. Замечания. а) Можно было бы рассматривать произвольные семейства векторов и называть такое семейство базисом, если любой вектор пространства однозначно представляется в виде *конечной* линейной комбинации элементов семейства. В этом смысле любое линейное пространство имеет базис, и у бесконечномерного пространства базис всегда бесконечен. Однако это понятие не слишком полезно. Как правило, бесконечномерные пространства снабжаются топологией, и определение базиса видоизменяется с учетом этой топологии и возможности определять некоторые бесконечные линейные комбинации.

б) В общих линейных пространствах базисные векторы по традиции нумеруются целыми числами от 1 до n (иногда от 0 до n), но это совершенно не обязательно. Базис $\{\delta_s\}$ в $F(S)$ естественно нумеруется элементами множества $s \in S$. Можно также считать базис L просто *подмножеством* в L, элементы которого не снабжены никакими индексами (ср. п. 20). Нумерация, или, скорее, порядок элементов базиса, существенны при использовании матричного формализма (см. § 4). В других вопросах может оказаться важной другая структура на множестве индексов базиса. Например, если S — конечная группа, то важно, как индексы s базиca $\{\delta_s\}$ перемножаются внутри S, а случайная нумерация S целыми числами может только загромоздить обозначения.

6. Примеры. а) $\mathscr{K}^n$ имеет размерность n. б) $F(S)$ имеет размерность n, равную числу элементов S, если S конечно.

Позже мы научимся вычислять размерности линейных пространств, не строя их базисов. Это очень важно, потому что многие числовые инварианты в математике определяются как размерности («числа Бетти» в топологии, индексы операторов в теории дифференциальных уравнений); базисы же соответствующих пространств могут оказаться трудно вычислимыми или не имеющими особого смысла. Но пока мы еще должны поработать с базисами.

Проверка того, что данное семейство векторов $\{e_1, \ldots, e_n\}$ в L образует базис, в соответствии с определением состоит из двух частей. Их отдельное рассмотрение приводит к следующим понятиям.

7. Определение. *Линейной оболочкой семейства векторов называется множество их всевозможных линейных комбинаций в L.*

Легко проверить, что линейная оболочка является линейным подпространством в L (см. § 1, п. 8). Линейную оболочку $\mathscr{K}e_1 + \mathscr{K}e_2 + \ldots$ также называют подпространством, *натянутым* на векторы $\{e_i\}$ или *порожденным* векторами семейства $\{e_i\}$. Ее можно определить еще как *пересечение всех линейных подпространств в L, содержащих все e_i* (докажите!). *Рангом* семейства векторов называется размерность его линейной оболочки.

Первое характеристическое свойство базиса: *его линейная оболочка совпадает со всем L.*

8. Определение. *Семейство векторов $\{e_i\}$ называется линейно независимым, если никакая нетривиальная линейная комбинация $\{e_i\}$ не равна нулю, т. е. если из $\sum_{i=1}^{n} a_i e_i = 0$ следует, что все $a_i = 0$. В противном случае оно называется линейно зависимым.*

Линейная независимость семейства $\{e_i\}$ означает, что *нулевой вектор однозначно представляется в виде линейной комбинации элементов семейства.* Тогда любой другой вектор имеет либо единственное представление, либо ни одного. Действительно, сравнивая два представления

$$l = \sum_{i=1}^{n} a_i e_i = \sum_{i=1}^{n} a_i' e_i, \text{ находим } 0 = \sum_{i=1}^{n} (a_i - a_i') e_i, \text{ откуда } a_i = a_i'.$$

Отсюда следует второе характеристическое свойство базиса: *его элементы линейно независимы.*

Объединение этих двух свойств равносильно первоначальному определению базиса.

Заметим еще, что *семейство векторов линейно независимо тогда и только тогда, когда оно образует базис своей линейной оболочки.*

Семейство $\{e_1, \ldots, e_n\}$ заведомо линейно зависимо, если среди векторов e_i есть нулевой или два одинаковых (почему?).

Более общо:

9. Лемма. а) *Семейство векторов $\{e_1, \ldots, e_n\}$ линейно зависимо тогда и только тогда, когда хотя бы один из векторов e_j является линейной комбинацией остальных.*

б) *Если семейство $\{e_1, \ldots, e_n\}$ линейно независимо, а семейство $\{e_1, \ldots, e_n, e_{n+1}\}$ линейно зависимо, то e_{n+1} является линейной комбинацией $e_1, \ldots, e_n$.*

Доказательство. а) Если $\sum_{i=1}^{n} a_i e_i = 0$ и $a_j \neq 0$, то $e_j = = \sum_{\substack{i=1 \\ i\neq j}}^{n} (-a_j^{-1} a_i) e_i$. Наоборот, если $e_j = \sum_{i \neq j} b_i e_i$, то $e_j - \sum_{i\neq j} b_i e_i = 0$.

б) Если $\sum_{i=1}^{n+1} a_i e_i = 0$ и не все a_i равны нулю, то обязательно $a_{n+1} \neq 0$, иначе мы получили бы нетривиальную линейную зависимость между $e_1, \ldots, e_n$. Поэтому $e_{n+1} = \sum_{i=1}^{n} \left(-a_{n+1}^{-1} a_i\right) e_i$. Лемма доказана.

Пусть $E = \{e_1, \ldots, e_n\}$ — некоторое конечное семейство векторов в L, $F = \{e_{i_1}, \ldots, e_{i_m}\}$ — его линейно независимое подсемейство. Назовем F *максимальным*, если каждый элемент из E линейно выражается через элементы из F.

10. Предложение. *Каждое линейно независимое подсемейство $E' \subset E$ содержится в некотором максимальном линейно независимом подсемействе $F \subset E$. Линейные оболочки F и E совпадают.*

Доказательство. Если в $E \setminus E'$ есть вектор, не представимый в виде линейной комбинации элементов E', добавим его к E'. В силу утверждения б) леммы п. 9 полученное семейство E'' будет линейно независимым. Применим то же рассуждение к E'' и т. д. Поскольку E конечно, этот процесс оборвется на максимальном семействе F. Любой элемент линейной оболочки E, очевидно, линейно выражается через векторы семейства F.

В случае $E' = \varnothing$ в качестве E'' нужно выбрать ненулевой вектор из E, если он есть; иначе F пусто.

11. Замечание. Этот результат верен и для бесконечных семейств E. Для его доказательства следует применить трансфинитную индукцию или лемму Цорна: см. пп. 18—20.

Максимальное подсемейство не обязательно единственно: рассмотрим $E = \{(1,0),\ (0,1),\ (1,1)\}$, $E' = \{(1,0)\}$ в $\mathcal{K}^2$. Тогда E' содержится в двух максимальных независимых подсемействах $\{(1,0),\ (0,1)\}$ и $\{(1,0),\ (1,1)\}$. Однако число элементов максимального подсемейства определено однозначно; оно совпадает с размерностью линейной оболочки E и называется рангом семейства E.

Часто бывает полезна следующая теорема.

12. Теорема о продолжении базиса. *Пусть $E' = \{e_1, \ldots, e_m\}$ — линейно независимое семейство векторов в конечномерном пространстве L. Тогда существует базис L, содержащий E';*

Доказательство. Выберем какой-нибудь базис $\{e_{m+1}, \ldots$ $\ldots, e_n\}$ в L и положим $E = \{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$. Обозначим через F максимальное линейно независимое подсемейство E, содержащее E'. Оно является искомым базисом.

В самом деле, нужно только проверить, что линейная оболочка F совпадает с L. Но она равна линейной оболочке E по предложению п. 10, а последняя равна L, потому что в E содержится базис пространства L.

13. Следствие (монотонность размерности). *Пусть M — линейное подпространство в L. Тогда* $\dim M \leqslant \dim L$, *и если L конечномерно, то из* $\dim M = \dim L$ *следует, что* $M = L$.

Доказательство. Если M бесконечномерно, то L также бесконечномерно. Действительно, покажем сначала, что в M можно найти сколь угодно большие независимые семейства векторов. Если семейство из n линейно независимых векторов $\{e_1, \ldots, e_n\}$ уже найдено, то его линейная оболочка $M' \subset M$ не может совпадать с M — иначе M было бы n-мерно. Поэтому в M есть вектор e_{n+1}, линейно не выражающийся через $\{e_1, \ldots, e_n\}$, и утверждение б) леммы п. 9 показывает, что семейство $\{e_1, \ldots, e_n, e_{n+1}\}$ линейно независимо. Теперь предположим, что M бесконечномерно, а L n-мерно. Тогда любые $n+1$ линейных комбинаций элементов базиса L линейно зависимы по рассуждению в доказательстве теоремы п. 4, что противоречит бесконечномерности M.

Остается разобрать случай, когда M и L конечномерны. Но тогда любой базис M по теореме п. 12 можно продолжить до базиса L, откуда и следует, что $\dim M \leqslant \dim L$.

Наконец, если $\dim M = \dim L$, то любой базис M должен быть базисом L — иначе его продолжение до базиса состояло бы из $> \dim L$ элементов, что невозможно.

14. Базисы и флаги. Один из стандартных способов изучения множеств S с алгебраическими структурами состоит в выделении в них последовательности подмножеств $S_0 \subset S_1 \subset S_2 \subset \ldots$ или $S_0 \supset S_1 \supset S_2 \supset \ldots$ так, что переход от одного подмножества к следующему устроен в каком-то смысле просто. Общее название таких последовательностей — *фильтрации* (возрастающая и убывающая соответственно). В теории линейных пространств строго возрастающая последовательность подпространств $L_0 \subset L_1 \subset \ldots \subset L_n$ пространства L называется *флагом*. (Мотивировка названия: флаг {точка 0} $\subset$ {прямая} $\subset$ {плоскость} — это «гвоздь», «древко» и «полотнище».)

Число n назовем *длиной* флага $L_0 \subset L_1 \subset \ldots \subset L_n$.

Флаг $L_0 \subset L_1 \subset \ldots \subset L_n \subset \ldots$ назовем *максимальным*, если $L_0 = \{0\}$, $\bigcup L_i = L$ и между L_i, L_{i+1} (для всех i) нельзя вставить подпространство: если $L_i \subset M \subset L_{i+1}$, то либо $L_i = M$, либо $M = L_{i+1}$.

По всякому базису $\{e_1, \ldots, e_n\}$ пространства L можно построить флаг длины n, положив $L_0 = \{0\}$, L_i — линейная оболочка $\{e_1, \ldots, e_i\}$ (при $i \geqslant 1$). Из доказательства следующей теоремы будет видно, что этот флаг максимален и что наша конструкция дает все максимальные флаги.

15. Теорема. *Размерность пространства L равна длине любого его максимального флага.*

Доказательство. Пусть $L_0 \subset L_1 \subset L_2 \subset \ldots$ — максимальный флаг в L. Для каждого $i \geqslant 1$ выберем вектор $e_i \in L_i \setminus L_{i-1}$ и покажем, что $\{e_1, \ldots, e_i\}$ образуют базис пространства L_i.

Прежде всего, линейная оболочка семейства $\{e_1, \ldots, e_{i-1}\}$ содержится в L_{i-1}, а e_i не лежит в L_{i-1}, откуда индукцией по i (с учетом $e_1 \neq 0$) следует, что $\{e_1, \ldots, e_i\}$ линейно независимы для всех i.

Теперь индукцией по i покажем, что $\{e_1, \ldots, e_i\}$ порождают L_i. Пусть это верно для $i-1$, и пусть M — линейная оболочка семейства $\{e_1, \ldots, e_i\}$. Тогда $L_{i-1} \subset M$ по индуктивному предположению и $L_{i-1} \neq M$ из-за того, что $e_i \notin L_{i-1}$. По определению максимальности флага отсюда следует, что $M = L_i$.

Теперь нетрудно завершить доказательство теоремы. Если $L_0 \subset L_1 \subset \ldots \subset L_n = L$ — конечный максимальный флаг в L, то векторы $\{e_1, \ldots, e_n\}$, $e_i \in L_i \setminus L_{i-1}$, по доказанному образуют базис в L, так что $n = \dim L$. Если в L есть бесконечный максимальный флаг, то эта конструкция дает сколь угодно большие линейно независимые семейства векторов в L, так что L бесконечномерно.

16. Дополнение. В конечномерном пространстве L любой флаг можно дополнить до максимального, и поэтому его длина всегда $\leqslant \dim L$. Действительно, будем вставлять в исходный флаг промежуточные подпространства, пока это возможно. Этот процесс не может продолжаться до бесконечности, ибо конструкция систем векторов $\{e_1, \ldots, e_i\}$, $e_i \in L_i \setminus L_{i-1}$, по *любому* флагу дает линейно независимые системы (см. начало доказательства теоремы п. 15), и потому длина флага не может превзойти $\dim L$.

17. Основной принцип работы с бесконечномерными пространствами: лемма Цорна, или трансфинитная индукция. Большинство теорем конечномерной линейной алгебры нетрудно доказать, опираясь на существование конечных базисов и теорему п. 12 о продолжении базисов: много примеров тому читатель увидит в дальнейшем. Но привычка к базисам затрудняет переход к функциональному анализу. Мы опишем сейчас теоретико-множественный принцип, который в очень многих случаях заменяет апелляцию к базисам.

Напомним (см. «Введение в алгебру», гл. 1, § 6), что *частично упорядоченным множеством* называется множество X вместе с бинарным *отношением порядка* $\leqslant$ на X, которое рефлексивно ($x \leqslant x$), транзитивно (если $x \leqslant y$, $y \leqslant z$, то $x \leqslant z$) и антисимметрично (если $x \leqslant y$ и $y \leqslant x$, то $x = y$). Вполне может оказаться, что пара элементов $x, y \in X$ не находится ни в отношении $x \leqslant y$, ни в отношении $y \leqslant x$. Если же для любой пары либо $x \leqslant y$, либо $y \leqslant x$, то множество называется *линейно упорядоченным*, или *цепью*.

Верхняя грань подмножества Y в частично упорядоченном множестве X — это любой элемент $x \in X$ такой, что $y \leqslant x$ для всех $y \in Y$. Верхняя грань подмножества может и не существовать: если $X = \mathbf{R}$ с обычным отношением $\leqslant$, а $Y = \mathbf{Z}$ (целые числа), то верхней грани у Y нет.

Наибольшим элементом частично упорядоченного множества X называется элемент $n \in X$ такой, что $x \leqslant n$ для всех $x \in X$, а *максимальным* — элемент $m \in X$, для которого из $m \leqslant x \in X$ следует $x = m$. Наибольший элемент всегда максимален, но не наоборот.

18. Пример. Типичный пример упорядоченного множества X — это множество всех подмножеств $\mathscr{P}(S)$ множества S или некоторая его часть, упорядоченное отношением $\subseteq$. Если S имеет больше двух элементов, то $\mathscr{P}(S)$ частично упорядочено, но не линейно упорядочено (почему?). Элемент $S \in \mathscr{P}(S)$ максимальный, и даже наибольший в $\mathscr{P}(S)$.

19. Лемма Цорна. *Пусть X — непустое частично упорядоченное множество, любая цепь в котором обладает верхней гранью в X. Тогда любая цепь обладает такой верхней гранью, которая является в то же время максимальным элементом в X.*

Лемму Цорна можно выводить из других, более приемлемых интуитивно аксиом теории множеств, но логически она эквивалентна так называемой «аксиоме выбора», если остальные аксиомы приняты. Поэтому удобно причислять ее к числу основных аксиом, что часто и делается.

20. Пример применения леммы Цорна: существование базиса в бесконечномерных линейных пространствах. Пусть L — линейное пространство над полем $\mathscr{K}$. Обозначим через $X \subset \mathscr{P}(L)$ *множество линейно независимых подмножеств векторов в L, упорядоченное отношением $\subseteq$.*

Иными словами, $Y \in X$, если любая конечная линейная комбинация векторов из Y, равная нулю, имеет нулевые коэффициенты. Проверим условие леммы Цорна: если S — некоторая цепь в X, то у нее есть верхняя грань в X. Действительно, положим $Z = \bigcup\limits_{Y \in S} Y$. Ясно, что $Y \subseteq Z$ для всякого $Y \in S$; кроме того, Z образует линейно независимое множество векторов, потому что любое конечное множество векторов $\{y_1, \ldots, y_n\}$ из Z содержится в некотором элементе $Y \in S$. В самом деле, пусть $y_i \in Y_i \in S$; так как S — цепь, из каждых двух элементов Y_i, $Y_j \in S$ один является подмножеством другого; выкидывая по очереди меньшие множества из таких пар, мы получим, что среди Y_i есть наибольшее множество; в нем и содержатся все $y_1, \ldots, y_n$, которые, таким образом, линейно независимы.

Применим теперь заключение леммы Цорна. Здесь достаточна только часть его: существование в X максимального элемента. Согласно определению, это такое линейно независимое множество векторов $Y \in X$, что если добавить к нему любой вектор $l \in L$, то множество $Y \cup \{l\}$ уже не будет линейно независимым. Точно такое же рассуждение, как при доказательстве утверждения б) леммы п. 9, показывает тогда, что l есть (конечная) линейная комбинация элементов Y, т. е. Y образует базис в L.

УПРАЖНЕНИЯ

1. Пусть L — пространство многочленов от x степени $\leqslant n-1$ с коэффициентами в поле $\mathscr{K}$. Проверить следующие утверждения:

а) $1, x, \ldots, x^{n-1}$ образуют базис в L. Координаты многочлена f в этом базисе — это его коэффициенты.

б) $1, x-a, (x-a)^2, \ldots, (x-a)^{n-1}$ образуют базис в L. Если $\operatorname{char}\mathscr{K} = p \geqslant n$, то координаты многочлена f в этом базисе: $\left\{f(a), f'(a), \frac{f''(a)}{2!}, \ldots \ldots, \frac{f^{(n-1)}(a)}{(n-1)!}\right\}$.

в) Пусть $a_1, \ldots, a_n \in \mathscr{K}$ — попарно различные элементы. Положим $g_i(x) = \prod_{j\neq i}(x-a_j)(a_i-a_j)^{-1}$. Многочлены $g_1(x), \ldots, g_n(x)$ образуют базис L («интерполяционный базис»). Координаты многочлена f в этом базисе: $\{f(a_1), \ldots, f(a_n)\}$.

2. Пусть L — n-мерное пространство, $f\colon L\to\mathscr{K}$ — ненулевой линейный функционал. Доказать, что $M = \{l\in L\,|\,f(l)=0\}$ является $(n-1)$-мерным подпространством в L. Доказать, что все $(n-1)$-мерные подпространства получаются таким способом.

3. Пусть L — n-мерное пространство, $M\subset L$ — m-мерное подпространство. Доказать, что существуют линейные функционалы $f_1, \ldots, f_{n-m}\in L^*$ такие, что $M = \{l\,|\,f_1(l) = \ldots = f_{n-m}(l) = 0\}$.

4. Вычислить размерности следующих пространств:

а) пространства многочленов степени $\leqslant p$ от n переменных;

б) пространства однородных многочленов (форм) степени p от n переменных;

в) пространства функций из $F(S)$, $|S|<\infty$, обращающихся в нуль во всех точках из подмножества $S_0\subset S$.

5. Пусть $\mathscr{K}$ — конечное поле характеристики p. Доказать, что число его элементов равно p^n для некоторого $n\geqslant 1$. (*Указание.* Рассмотреть $\mathscr{K}$ как линейное пространство над простым подполем, состоящим из всех «сумм единиц» в $\mathscr{K}$: $0, 1, 1+1, \ldots$)

6. Заменой понятия флага в бесконечномерном случае служит понятие цепи подпространств (упорядоченных по включению). Пользуясь леммой Цорна, доказать, что всякая цепь содержится в максимальной.

§ 3. Линейные отображения

1. Определение. *Пусть L, M — линейные пространства над полем $\mathscr{K}$. Отображение $f\colon L\to M$ называется линейным, если для всех $l, l_1, l_2\in L$, $a\in\mathscr{K}$ имеем*

$$f(al) = af(l),\ f(l_1+l_2) = f(l_1)+f(l_2).$$

Линейное отображение является гомоморфизмом аддитивных групп. В самом деле, $f(0)=0f(0)=0$ и $f(-l)=f((-1)l)= -f(l)$. Индукция по n показывает, что для любых $a_i\in\mathscr{K}$, $l_i\in L$ имеем $f\left(\sum_{i=1}^n a_il_i\right)=\sum_{i=1}^n a_if(l_i)$.

Линейные отображения $f\colon L\to L$ называются также *линейными операторами* на L.

2. Примеры. а) *Нулевое* линейное отображение $f\colon L\to M$, $f(l)=0$ для всех $l\in L$. *Тождественное* линейное отображение: $f\colon L\to L$, $f(l)=l$ для всех $l\in L$. Оно обозначается id_L или id (от английского слова «identity»). *Умножение на скаляр* $a\in\mathscr{K}$, или *гомотетия* $f\colon L\to L$, $f(l)=al$ для всех $l\in L$. При $a=0$ получается нулевой оператор, при $a=1$ — тождественный.

б) Линейные отображения $f\colon L\to\mathscr{K}$ — это линейные функции, или функционалы, на L (см. § 1, п. 9). Пусть L — пространство с базисом $\{e_1, \ldots, e_n\}$. Для любого $1\leqslant i\leqslant n$ отображение $e^i\colon L\to\mathscr{K}$, где $e^i(l)$ — i-я координата l в базисе $\{e_1, \ldots, e_n\}$, является линейным функционалом.

в) Пусть $L=\{x\in\mathbf{R}\mid x>0\}$ наделено структурой линейного пространства над $\mathbf{R}$, описанной в § 1, пример а) п. 10, $M=\mathbf{R}^1$. Отображение $\log\colon L\to M$, $x\longmapsto\log x$, $\mathbf{R}$-линейно.

г) Пусть $S\subset T$ — два множества. Отображение $F(T)\to F(S)$, которое всякой функции на T ставит в соответствие ее ограничение на S, линейно. В частности, если $S=\{s\}$, $s\in T$, $f\in F(T)$, то отображение: $f\longmapsto$(значение f в точке s) линейно.

Конструкция линейных отображений с нужными свойствами часто основывается на следующем результате.

3. Предложение. *Пусть L, M — линейные пространства над полем $\mathscr{K}$; $\{l_1, \ldots, l_n\}\subset L$ и $\{m_1, \ldots, m_n\}\subset M$ — два семейства векторов с одинаковым числом элементов. Тогда:*

а) *если линейная оболочка $\{l_1, \ldots, l_n\}$ совпадает с L, то существует не больше одного линейного отображения $f\colon L\to M$, для которого $f(l_i)=m_i$ при всех i;*

б) *если $\{l_1, \ldots, l_n\}$ к тому же линейно независимы, т. е. образуют базис L, то такое отображение существует.*

Доказательство. Пусть f, f' — пара отображений с $f(l_i)=f'(l_i)=m_i$ для всех i. Рассмотрим отображение $g=f-f'$, где $(f-f')(l)=f(l)-f'(l)$. Легко проверить, что оно линейно. Кроме того, оно переводит в нуль все l_i и потому любую линейную комбинацию векторов l_i. Значит, f и f' совпадают на каждом векторе из L, откуда $f'=f$.

Пусть теперь $\{l_1, \ldots, l_n\}$ образует базис L. Так как каждый элемент L однозначно представляется в виде $\sum\limits_{i=1}^{n} a_i l_i$, мы можем определить теоретико-множественное отображение $f\colon L\to M$ формулой

$$f\left(\sum_{i=1}^{n} a_i l_i\right)=\sum_{i=1}^{n} a_i m_i.$$

Его линейность проверяется непосредственно.

В этом доказательстве использовалась разность двух линейных отображений $L\to M$. Это частный случай следующей более общей конструкции.

4. Обозначим через $\mathscr{L}(L, M)$ множество линейных отображений из L в M. Для $f, g\in\mathscr{L}(L, M)$ и $a\in\mathscr{K}$ определим af и $f+g$

формулами
$$(af)(l)=a(f(l)),\ (f+g)(l)=f(l)+g(l)$$
для всех $l\in L$. Точно так же, как в § 1, п. 9, проверяется, что af и $f+g$ линейны, так что $\mathscr{L}(L, M)$ — линейное пространство.

5. Пусть $f\in\mathscr{L}(L,M)$ и $g\in\mathscr{L}(M,N)$. Теоретико-множественная композиция $g\circ f=gf\colon L\to N$ является линейным отображением. Действительно,
$$(gf)(l_1+l_2)=g[f(l_1+l_2)]=g[f(l_1)+f(l_2)]=g[f(l_1)]+g[f(l_2)]=\\ =gf(l_1)+gf(l_2)$$
и, аналогично, $(gf)(al)=a(gf(l))$.

Очевидно, $\mathrm{id}_M\circ f=f\circ\mathrm{id}_L=f$. Кроме того, $h(gf)=(hg)f$, когда обе части определены, так что скобки можно опустить; это общее свойство ассоциативности теоретико-множественных отображений. Наконец, композиция gf линейна по каждому из аргументов при фиксированном втором: например, $g\circ(af_1+bf_2)=a(g\circ f_1)+$ $+b(g\circ f_2)$.

6. Пусть $f\in\mathscr{L}(L,M)$ — биективное отображение. Тогда у него есть теоретико-множественное обратное отображение $f^{-1}\colon M\to L$. Мы утверждаем, что f^{-1} автоматически линейно. Для этого следует проверить, что
$$f^{-1}(m_1+m_2)=f^{-1}(m_1)+f^{-1}(m_2),\quad f^{-1}(am_1)=af^{-1}(m_1)$$
для всех $m_1, m_2\in M$; $a\in\mathscr{K}$. Поскольку f биективно, существуют и однозначно определены такие векторы $l_1, l_2\in L$, что $m_i=f(l_i)$. Написав формулы
$$f(l_1)+f(l_2)=f(l_1+l_2),\quad af(l_1)=f(al_1),$$
применив к их обеим частям f^{-1} и заменив в результате l_i на $f^{-1}(m_i)$, получим требуемое.

Биективные линейные отображения $f\colon L\to M$ называются *изоморфизмами*. Пространства L и M называются *изоморфными*, если между ними существует изоморфизм.

Следующая теорема показывает, что размерность пространства полностью определяет его с точностью до изоморфизма.

7. Теорема. *Два конечномерных пространства L и M над полем $\mathscr{K}$ изоморфны тогда и только тогда, когда у них одинаковые размерности.*

Доказательство. Изоморфизм $l\colon L\to M$ сохраняет все свойства, формулируемые в терминах линейных комбинаций. В частности, он переводит любой базис L в некоторый базис M, так что размерности L и M совпадают. (Из этого рассуждения следует также, что конечномерное пространство не может быть изоморфно бесконечномерному.)

Наоборот, пусть размерности L и M равны n. Выберем базисы $\{l_1, \ldots, l_n\}$ и $\{m_1, \ldots, m_n\}$ в L и M соответственно. Формула
$$f\left(\sum_{i=1}^{n}a_il_i\right)=\sum_{i=1}^{n}a_im_i$$

определяет линейное отображение L в M по предложению п. 3. Оно является биекцией, ибо формула

$$f^{-1}\left(\sum_{i=1}^{n} a_i m_i\right)=\sum_{i=1}^{n} a_i l_i$$

определяет обратное линейное отображение f^{-1}.

8. Предупреждение. Если даже изоморфизм между двумя линейными пространствами L, M и существует, он определен однозначно только в двух случаях:

а) $L=M=\{0\}$,

б) L и M одномерны, а $\mathscr{K}$ — поле из двух элементов (попробуйте доказать это!).

Во всех остальных случаях имеется много (если $\mathscr{K}$ бесконечно, то бесконечно много) изоморфизмов. В частности, имеется много изоморфизмов пространства L с самим собой. В силу результатов пп. 5 и 6 они образуют группу относительно теоретико-множественной композиции. Эта группа называется *полной линейной группой пространства* L. Позже мы сможем описать ее в более явном виде как группу невырожденных квадратных матриц.

Иногда бывает, что между двумя линейными пространствами определен некоторый изоморфизм, не зависящий ни от каких произвольных выборов (как выборы базисов в пространствах L и M в доказательстве теоремы п. 7). Такие изоморфизмы мы будем называть *каноническими* или *естественными* (точное определение этих терминов можно дать только на категорном языке, о котором см. § 13). Следует тщательно отличать естественные изоморфизмы от «случайных». Мы приведем два характерных примера, очень важных для понимания этого различия.

9. «Случайный» изоморфизм между пространством и двойственным к нему. Пусть L — конечномерное пространство с базисом $\{e_1, \ldots, e_n\}$. Обозначим через $e^i \in L^*$ линейный функционал

$$l \mapsto e^i(l), \text{ где } e^i(l) — i\text{-я координата вектора } l \text{ в базисе } \{e_i\}$$

(не путать с i-й степенью; в линейном пространстве она не определена). Мы утверждаем, что *функционалы* $\{e^1, \ldots, e^n\}$ *образуют базис в* L^*, *так называемый двойственный к* $\{e_1, \ldots, e_n\}$ *базис.* Равносильное описание $\{e^i\}$ такое: $e^i(e_k)=\delta_{ik}$ (символ Кронекера: 1 при $i=k$, 0 при $i \neq k$).

В самом деле, всякий линейный функционал $f: L \to \mathscr{K}$ можно представить в виде линейной комбинации $\{e^i\}$:

$$f=\sum_{i=1}^{n} f(e_i)\, e^i.$$

Действительно, значения левой и правой части совпадают на любой линейной комбинации $\sum_{k=1}^{n} a_k e_k$, потому что $e^i\left(\sum_{k=1}^{n} a_k e_k\right)=a_i$ по определению e^i.

Кроме того, $\{e^i\}$ линейно независимы: если $\sum_{i=1}^{n} a_i e^i = 0$, то для всех k, $1 \leqslant k \leqslant n$, имеем $a_k = \left(\sum_{i=1}^{n} a_i e^i\right)(e_k) = 0$.

Поэтому L и L^* имеют одинаковую размерность n и даже определен изоморфизм f: $L \to L^*$, который переводит e_i в e^i.

Однако *этот изоморфизм не каноничен*: замена базиса $\{e_1, \ldots, e_n\}$, вообще говоря, меняет его. Так, если L одномерно, то для любого ненулевого вектора $e_1 \in L$ семейство $\{e_1\}$ является базисом L. Пусть $\{e^1\}$ — двойственный базис к $\{e_1\}$, $e^1(e_1) = 1$. Тогда к базису $\{ae_1\}$, $a \in \mathcal{K} \setminus \{0\}$, двойствен базис $\{a^{-1}e^1\}$. Но линейные отображения f_1: $e_1 \mapsto e^1$ и f_2: $ae_1 \mapsto a^{-1}e^1$ различны, если только $a^2 \neq 1$.

10. Канонический изоморфизм между пространством и дважды двойственным к нему. Пусть L — линейное пространство, L^* — пространство линейных функций на нем, $L^{**} = (L^*)^*$ — пространство линейных функций на L^* — «дважды двойственное к L пространство».

Опишем каноническое отображение ε_L: $L \to L^{**}$, не зависящее ни от каких произвольных выборов. *Оно ставит в соответствие каждому вектору $l \in L$ функцию на L^*, значение которой на функционале $f \in L^*$ равно $f(l)$; в краткой записи:*

$$\varepsilon_L: \quad l \mapsto [f \mapsto f(l)].$$

Проверим следующие свойства ε_L:

а) *Для каждого $l \in L$ отображение $\varepsilon_L(l)$: $L^* \to \mathcal{K}$ линейно.* Действительно, это означает, что *выражение $f(l)$ как функция от f при фиксированном l линейно по f.* Но это следует из правил сложения функционалов и умножения их на скаляр (§ 1, п. 7).

Следовательно, ε_L действительно определяет отображение L в L^{**}, как и утверждалось.

б) *Отображение ε_L: $L \to L^{**}$ линейно.* Действительно, это означает, что выражение $f(l)$ *как функция от l при фиксированном f линейно*, — это так, ибо $f \in L^*$.

в) Если L *конечномерно, то отображение ε_L: $L \to L^{**}$ является изоморфизмом.* В самом деле, пусть $\{e_1, \ldots, e_n\}$ — базис L, $\{e^1, \ldots, e^n\}$ — двойственный базис L^*, $\{e_1', \ldots, e_n'\}$ — базис в L^{**}, двойственный к $\{e^1, \ldots, e^n\}$.

Покажем, что $\varepsilon_L(e_i) = e_i'$, откуда и будет следовать, что ε_L — изоморфизм (в этой проверке использование базиса L безобидно, ибо в определении ε_L он не участвовал!).

В самом деле, $\varepsilon_L(e_i)$ согласно определению есть функционал на L^*, значение которого на e^k равно $e^k(e_i) = \delta_{ik}$ («символ Кронекера»). Но e_i' — точно такой же функционал на L^* по определению двойственного базиса.

Заметим, что если L бесконечномерно, то ε_L: $L \to L^{**}$ остается инъективным, но перестает быть сюръективным (см. упражнение 2). В функциональном анализе вместо полного L^* обычно рассматривают только подпространство линейных функционалов L', *непрерывных в подходящей топологии* на L и $\mathscr{K}$, и тогда отображение $L \to L''$ может быть определено и иногда оказывается изоморфизмом. Такие (топологические) пространства называют *рефлексивными.* Мы доказали, что *конечномерные пространства* (без учета топологии) *рефлексивны.*

Рассмотрим теперь связь между линейными отображениями и линейными подпространствами.

11. Определение. *Пусть* $f\colon L \to M$ *— линейное отображение. Множество* $\operatorname{Ker} f = \{l \in L \mid f(l) = 0\} \subset L$ *называется ядром* f, *а множество* $\operatorname{Im} f = \{m \in M \mid \exists\, l \in L, f(l) = m\} \subset M$ *называется образом* f.

Нетрудно убедиться, что ядро f является линейным подпространством в L, а образ f — линейным подпространством в M. Проверим, например, второе утверждение. Пусть $m_1, m_2 \in \operatorname{Im} f$, $a \in \mathscr{K}$. Тогда существуют такие векторы $l_1, l_2 \in L$, что $f(l_1) = m_1$, $f(l_2) = m_2$. Значит, $m_1 + m_2 = f(l_1 + l_2)$, $am_1 = f(al_1)$. Следовательно, $m_1 + m_2 \in \operatorname{Im} f$ и $am_1 \in \operatorname{Im} f$.

Отображение f инъективно тогда и только тогда, когда $\operatorname{Ker} f = \{0\}$. В самом деле, если $f(l_1) = f(l_2)$, $l_1 \neq l_2$, то $0 \neq l_1 - l_2 \in \operatorname{Ker} f$. Наоборот, если $0 \neq l \in \operatorname{Ker} f$, то $f(l) = 0 = f(0)$.

12. Теорема. *Пусть* L *— конечномерное линейное пространство,* $f\colon L \to M$ *— линейное отображение. Тогда* $\operatorname{Ker} f$ *и* $\operatorname{Im} f$ *конечномерны и*

$$\dim \operatorname{Ker} f + \dim \operatorname{Im} f = \dim L.$$

Доказательство. Ядро f конечномерно по следствию п. 13, § 2. Выберем базис $\{e_1, \ldots, e_m\}$ в $\operatorname{Ker} f$ и продолжим его до базиса $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_{m+n}\}$ пространства L по теореме п. 12, § 2. Покажем, что векторы $f(e_{m+1}), \ldots, f(e_{m+n})$ образуют базис в $\operatorname{Im} f$. Отсюда, очевидно, будет следовать теорема.

Любой вектор из $\operatorname{Im} f$ имеет вид

$$f\left(\sum_{i=1}^{m+n} a_i e_i\right) = \sum_{i=m+1}^{m+n} a_i f(e_i).$$

Следовательно, $f(e_{m+1}), \ldots, f(e_{m+n})$ порождают $\operatorname{Im} f$.

Предположим, что $\sum\limits_{i=m+1}^{m+n} a_i f(e_i) = 0$. Тогда $f\left(\sum\limits_{i=m+}^{m+n} a_i e_i\right) = 0$. Это значит, что $\sum\limits_{i=m+1}^{m+n} a_i e_i \in \operatorname{Ker} f$, т. е. $\sum\limits_{i=m+1}^{m+n} a_i e_i = \sum\limits_{j=1}^{m} a_j e_j$. Это возможно, только если все коэффициенты равны нулю, ибо $\{e_1, \ldots, e_{m+n}\}$ — базис L. Следовательно, векторы $f(e_{m+1}), \ldots, f(e_{m+n})$ линейно независимы. Теорема доказана.

13. Следствие. *Следующие свойства* f *равносильны* (*в случае конечномерного* L):

а) f *инъективно.*

б) $\dim L = \dim \operatorname{Im} f$.

Доказательство. Согласно теореме, $\dim L = \dim \operatorname{Im} f$ тогда и только тогда, когда $\dim \operatorname{Ker} f = 0$, т. е. $\operatorname{Ker} f = \{0\}$.

УПРАЖНЕНИЯ

1. Пусть f: $\mathbf{R}^m \to \mathbf{R}^n$ — отображение, задаиное диффереицируемыми функциями, вообще говоря, нелииейиыми и переводящими нуль в нуль:

$$f(x_1, \ldots, x_m) = (\ldots, f_i(x_1, \ldots, x_m), \ldots), \; i = 1, \ldots, n,$$
$$f_i(0, \ldots, 0) = 0.$$

Поставим ему в соответствие *линейное отображение* df_0: $\mathbf{R}^m \to \mathbf{R}^n$, называемое *дифференциалом* f *в точке* 0, по формуле

$$(df_0)(e_j) = \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j}(0)\, e_i' = \left(\frac{\partial f_1}{\partial x_j}(0), \ldots, \frac{\partial f_n}{\partial x_j}(0)\right),$$

где $\{e_j\}$, $\{e_i'\}$ — стандартиые базисы $\mathbf{R}^m$ и $\mathbf{R}^n$. Показать, что если произвести замену базисов в пространствах $\mathbf{R}^m$ и $\mathbf{R}^n$ и вычислить df_0 по тем же формулам в других базисах, то новое лииейное отображеиие df_0 совпадает со старым.

2. Докажите, что пространство миогочлеиов $\mathbf{Q}[x]$ не изоморфио своему двойствеиному. (*Указание*. Сравните мощиости.)

§ 4. Матрицы

1. Цель этого параграфа — ввести язык матриц и установить основные связи его с языком линейных пространств и отображений. За дальнейшими подробностями и примерами мы отсылаем читателя к главам 2 и 3 «Введения в алгебру»; в частности, мы будем пользоваться развитой там теорией определителей, не повторяя ее. Читателю следует самостоятельно убедиться в том, что изложение в этих главах без изменений переносится с поля вещественных чисел на любое поле скаляров; исключения составляют лишь те случаи, где используются такие специфические свойства вещественных чисел, как порядок и непрерывность.

2. **Термины.** *Матрицей* A размера $m \times n$ с элементами из множества S называется семейство (a_{ik}) элементов из S, пронумерованное упорядоченными парами чисел (i, k), где $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant n$. Часто пишут $A = (a_{ik})$, $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant n$; указание размера может быть опущено.

При фиксированном i семейство $(a_{i1}, \ldots, a_{in})$ называется *i-й строкой* матрицы A. При фиксированном k семейство $(a_{1k}, \ldots, a_{mk})$ называется *k-м столбцом* матрицы A. Матрица размера $1 \times n$ называется просто строкой, а матрица размера $m \times 1$ — столбцом.

Если $m = n$, матрица A называется *квадратной* (иногда говорят «порядка n» вместо «размера $n \times n$»).

Если A — квадратная матрица порядка n, $S = \mathcal{K}$ (поле) и $a_{ik} = 0$ при $i \neq k$, матрица A называется *диагональной*; иногда ее записывают $\operatorname{diag}(a_{11}, \ldots, a_{nn})$. Вообще, элементы (a_{ii}) называются элементами *главной диагонали*. Элементы $a_{1, k+1}$; $a_{2, k+2}$; $\ldots$, где $k > 0$, образуют диагональ, стоящую выше главной, а элементы $a_{k+1, 1}$; $a_{k+2, 2}$; $\ldots$, где $k > 0$, — диагональ, стоящую ниже главной. Если $S = \mathcal{K}$ и $a_{ik} = 0$ при $k < i$, матрица называется *верхней* *треугольной*, а если $a_{ik}=0$ при $k>i$, то *нижней треугольной.* Диагональная квадратная матрица над $\mathscr{K}$, у которой все элементы на главной диагонали одинаковы, называется *скалярной.* Если эти элементы равны единице, матрица называется *единичной.* Единичная матрица порядка n обозначается E_n или просто E, если порядок ясен из контекста.

Все эти термины обязаны своим происхождением стандартной записи матрицы в виде таблицы

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Транспонированная к A матрица A^t имеет размеры $n\times m$, и ее элемент в i-й строке и k-м столбце равен a_{ki}. (Иногда используемое обозначение $A^t=(a_{ki})$ двусмысленно!)

3. Замечания. Бо́льшая часть матриц, встречающихся в теории линейных пространств над полем $\mathscr{K}$, имеет своими элементами элементы самого этого поля. Однако бывают и исключения. Например, мы будем иногда рассматривать упорядоченный базис $\{e_1, \dots, e_n\}$ пространства L, как матрицу размера $1\times n$ с элементами из этого пространства. Другой пример — *блочные матрицы*, элементами которых в свою очередь являются матрицы — блоки исходной. Именно разбиение номеров строк $[1, \dots, m]=I_1\cup \cup I_2\cup \dots \cup I_\mu$ и номеров столбцов $[1, \dots, n]=J_1\cup \dots \cup J_\nu$ на идущие подряд попарно непересекающиеся отрезки определяет разбиение матрицы A на блоки

$$\left(\begin{array}{c|c|c|c} A_{11} & A_{12} & \dots & A_{1\nu} \\ \hline \dots & \dots & \dots & \dots \\ \hline A_{\mu 1} & A_{\mu 2} & \dots & A_{\mu\nu} \end{array}\right),$$

где $A_{\alpha\beta}$ имеет своими элементами a_{ik}, $i\in I_\alpha$, $k\in J_\beta$. Если $\mu=\nu$, можно очевидным способом определить понятия блочно диагональной, блочной верхней треугольной, блочной нижней треугольной матриц. Этот же пример показывает, что не всегда удобно нумеровать столбцы и строки матрицы числами от 1 до m (или n): часто существен лишь порядок строк и столбцов.

4. Матрица линейного отображения. Пусть N и M — конечномерные линейные пространства над $\mathscr{K}$ с отмеченными базисами $\{e_1, \dots, e_n\}$ и $\{e'_1, \dots, e'_m\}$ соответственно. Рассмотрим произвольное линейное отображение $f\colon N\to M$ и поставим ему в соответствие матрицу A_f размера $m\times n$ с элементами из поля $\mathscr{K}$ следующим образом (заметьте, что размеры A_f суть размерности N, M *в обратном порядке*). Представим векторы $f(e_k)$ в виде линейных комбинаций: $f(e_k)=\sum\limits_{i=1}^{m} a_{ik}e'_i$. Тогда по определению $A_f=(a_{ik})$. Иными словами, коэффициенты этих линейных комбинаций суть последовательные *столбцы* матрицы A_f.

Матрица A_f называется матрицей линейного отображения f *относительно базисов* (или *в базисах*) $\{e_k\}$, $\{e'_i\}$.

В силу предложения п. 3, § 3, линейное отображение f однозначно определяется образами $f(e_k)$, и в качестве последних можно взять любое семейство из n векторов пространства M. Поэтому описанное соответствие устанавливает *биекцию* между множеством $\mathscr{L}(N, M)$ и множеством матриц размера $m \times n$ с элементами из $\mathscr{K}$ (или *над* $\mathscr{K}$). Эта биекция, однако, зависит от выбора базисов (см. п. 8 ниже).

Матрица A_f позволяет также описывать линейное отображение f в терминах его действия на координаты. Если вектор l представлен столбцом $\vec{x} = [x_1, \ldots, x_n]$ своих координат в базисе $\{e_1, \ldots, e_n\}$, т. е. $l = \sum_{i=1}^{n} x_i e_i$, то $f(l)$ представлен вектором-столбцом $\vec{y} = [y_1, \ldots, y_m]$, где

$$y_i = \sum_{k=1}^{n} a_{ik} x_k, \quad i = 1, \ldots, m.$$

Иными словами, $\vec{y} = A_f \cdot \vec{x}$ — обычное произведение матрицы A_f на столбец $\vec{x}$.

Когда речь идет о матрице *линейного оператора* $A = (a_{ik})$, всегда подразумевается, что в «двух экземплярах» пространства N выбирается один и тот же базис. Матрица линейного оператора квадратна. Матрица тождественного оператора единична.

Согласно п. 4, § 3, множество $\mathscr{L}(N, M)$ является в свою очередь линейным пространством над $\mathscr{K}$. При отождествлении элементов $\mathscr{L}(N, M)$ с матрицами эта структура описывается следующим образом.

5. Сложение матриц и умножение на скаляр. Пусть $A = (a_{ik})$, $B = (b_{ik})$ — две матрицы одинакового размера над полем $\mathscr{K}$, $a \in \mathscr{K}$. Положим

$$A + B = (c_{ik}), \quad \text{где} \quad c_{ik} = a_{ik} + b_{ik},$$
$$aA = (aa_{ik}).$$

Эти операции определяют на матрицах данного размера структуру линейного пространства. Легко проверить, что если $A = A_f$, $B = A_g$ (в одинаковых базисах), то

$$A_f + A_g = A_{f+g}, \quad A_{af} = aA_f,$$

так что указанное соответствие (а оно биективно) является изоморфизмом. В частности, $\dim \mathscr{L}(N, M) = \dim M \dim N$, потому что пространство матриц изоморфно $\mathscr{K}^{mn}$ (размер $m \times n$).

Композиция линейных отображений описывается в терминах умножения матриц.

6. Умножение матриц. Произведение матрицы A размера $m \times n'$ над полем $\mathscr{K}$ на матрицу B размера $n'' \times p$ над полем $\mathscr{K}$

определено тогда и только тогда, когда $n' = n'' = n$; размер AB в этом случае равен $m \times p$, и по определению

$$AB = (c_{ik}), \quad \text{где} \quad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Нетрудно проверить, что $(AB)^t = B^t A^t$.

Может случиться, что AB определена, но BA не определена (если $m \neq p$) или обе матрицы AB и BA определены, но имеют разные размеры (если $m \neq n$), или даже определены и имеют одинаковые размеры ($m = n = p$), но не совпадают. Иными словами, *умножение матриц не коммутативно*. Однако оно *ассоциативно*: *если матрицы AB и BC определены, то $(AB)C$ и $A(BC)$ определены и совпадают*. В самом деле, положим $A = (a_{ij})$, $B = (b_{jk})$, $C = (c_{kl})$. Согласованность размеров A с BC и AB с C предлагается проверить читателю. Если она уже проверена, то мы можем вычислять (il)-й элемент $(AB)C$ по формуле

$$\sum_k \left(\sum_j a_{ij} b_{jk} \right) c_{kl} = \sum_{j,\, k} (a_{ij} b_{jk})\, c_{kl},$$

а (il)-й элемент $A(BC)$ по формуле

$$\sum_j a_{ij} \left(\sum_k b_{jk} c_{kl} \right) = \sum_{j,\, k} a_{ij} (b_{jk} c_{kl}).$$

Так как умножение в $\mathscr{K}$ ассоциативно, эти элементы совпадают. Зная уже, что умножение матриц над $\mathscr{K}$ ассоциативно, мы можем убедиться, что «поблочное умножение» блочных матриц также ассоциативно (см. также упражнение 1).

Кроме того, произведение матриц линейно по каждому аргументу:

$$(aA + bB)C = aAC + bBC; \quad A(bB + cC) = bAB + cAC.$$

Важнейшее свойство умножения матриц состоит в том, что оно отвечает композиции линейных отображений. Однако целый ряд других ситуаций в линейной алгебре также удобно описывается умножением матриц: это главная причина унифицирующей роли матричного языка и некоторой самостоятельности матричной алгебры внутри линейной алгебры. Перечислим некоторые из этих ситуаций.

7. Матрица композиции линейных отображений. Пусть P, N, M — три конечномерных линейных пространства, $P \xrightarrow{g} N \xrightarrow{f} M$ — два линейных отображения. Выберем базисы $\{e_i''\}$, $\{e_k'\}$ и $\{e_m\}$ в P, N, M соответственно и обозначим через A_g, A_f, A_{fg} матрицы g, f, fg в этих базисах. Мы утверждаем, что $A_{fg} = A_f A_g$. В самом деле, пусть $A_f = (a_{jl})$, $A_g = (b_{ik})$. Имеем

$$g(e_k'') = \sum_i b_{ik} e_i',$$

$$fg(e_k'') = \sum_i b_{ik} f(e_i') = \sum_i b_{ik} \sum_j a_{ji} e_j = \sum_j \left(\sum_i a_{ji} b_{ik} \right) e_k.$$

Следовательно, (j, k)-й элемент матрицы A_{fg} равен $\sum_i a_{jl} b_{ik}$, т. е. $A_{fg} = A_f A_g$.

Согласно результатам пп. 4—6 множество линейных операторов $\mathcal{L}(L, L)$ после выбора базиса в L можно отождествить с множеством квадратных матриц $M_n(\mathcal{K})$ порядка $n = \dim L$ над полем $\mathcal{K}$. Имеющиеся в обоих множествах структуры линейных пространств и колец при этом отождествлении согласованы. Биекциям, т. е. линейным автоморфизмам $f: L \to L$, отвечают обратимые матрицы: если $f \circ f^{-1} = \mathrm{id}_L$, то $A_f A_{f^{-1}} = E_n$, так что $A_{f^{-1}} = A_f^{-1}$. Напомним, что матрица A обратима, или невырождена тогда и только тогда, когда $\det A \neq 0$.

8. а) *Действие линейного отображения в координатах.* В обозначениях п. 4 мы можем представлять векторы пространств N, M в координатах столбцами

$$\vec{x} = \begin{pmatrix} x_1 \\ \cdot \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ \cdot \\ \vdots \\ y_m \end{pmatrix}$$

и тогда действие оператора f записывается на языке матричного умножения формулой

$$\begin{pmatrix} y_1 \\ \cdot \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \cdot & \cdots & \cdot \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \cdot \\ \vdots \\ x_n \end{pmatrix},$$

или $\vec{y} = A_f \vec{x}$ (ср. п. 4). Иногда удобно писать аналогичную формулу в терминах базисов $\{e_i\}$, $\{e'_k\}$, где она принимает вид

$$f(e_1, \ldots, e_n) = (f(e_1), \ldots, f(e_n)) = (e'_1, \ldots, e'_m) A_f.$$

При этом формализм матричного умножения требует, чтобы в выражении справа векторы M умножались на скаляры *справа*, а не слева; это безобидно, мы просто будем считать, что $e'a = ae'$ для любых $e' \in M$, $a \in \mathcal{K}$.

Пользуясь такого рода записями, мы будем иногда нуждаться в проверке ассоциативности или линейности по аргументам «смешанных» произведений матриц, часть которых имеет элементы из $\mathcal{K}$, а другая часть из L, например

$$((e_1, \ldots, e_n) A) B = (e_1, \ldots, e_n)(AB)$$

или

$$(e + e'_1, \ldots, e_n + e'_n) A = (e_1, \ldots, e_n) A + (e'_1, \ldots, e'_n) A$$

и т. п. Формализм пп. 4, 5 автоматически переносится на эти случаи. То же замечание относится к блочным матрицам.

б) *Координаты вектора в измененном базисе.* Пусть в пространстве L выбраны два базиса $\{e_i\}$ и $\{e'_i\}$. Любой вектор $l \in l$

можно представить его координатами в этих базисах: $l=\sum_{i=1}^{n} x_i e_i = \sum_{k=1}^{n} x'_k e'_k$. Покажем, что существует квадратная матрица A порядка n, не зависящая от l, такая, что $\vec{x}=A\vec{x}'$.

Действительно, если $e'_k=\sum_{i=1}^{n} a_{ik}e_i$, то $A=(a_{ik})$:

$$\sum_{i=1}^{n} x_i e_i = l = \sum_{k=1}^{n} x'_k e'_k = \sum_{k=1}^{n} x'_k \left(\sum_{i=1}^{n} a_{ik} e_i\right) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik} x'_k\right) e_i.$$

Матрица A называется *матрицей перехода* (от нештрихованного базиса к штрихованному), или от штрихованных координат к нештрихованным. Заметим, что она обратима: обратная матрица есть матрица перехода от штрихованного базиса к нештрихованному.

Заметим, что формулу $\vec{x}=A\vec{x}'$ можно было прочесть также как формулу, выражающую координаты старого вектора-столбца $\vec{x}$ через координаты вектора $f(\vec{x}')$, где f — линейное отображение $L\to L$, описанное матрицей A в базисе $\{e_k\}$.

В физике эти две точки зрения называются соответственно «пассивной» и «активной». В первом случае мы описываем *одно и то же состояние системы* (вектор l) с точки зрения *разных наблюдений* (со своими системами координат). Во втором случае *наблюдатель один*, а состояние системы подвергается преобразованиям, состоящим, например, из симметрий пространства состояний этой системы.

в) *Матрица линейного отображения в измененных базисах.* В ситуации п. 4 выясним, как изменится матрица A_f линейного отображения, если перейти от базисов $\{e_k\}$, $\{e'_i\}$ к новым базисам $\{\bar{e}_k\}$, $\{\bar{e}'_i\}$ пространств N, M. Пусть B — матрица перехода от $\{e_k\}$-координат к $\{\bar{e}_k\}$-координатам, а C — матрица перехода от $\{e'_i\}$-координат к $\{\bar{e}'_i\}$-координатам. Мы утверждаем, что матрица $\bar{A}_f$ отображения f в базисах $\{\bar{e}_k\}$, $\{\bar{e}'_i\}$ равна

$$\bar{A}_f=C^{-1}A_fB.$$

В самом деле, вычисляя в базисах, имеем

$$(\bar{e}_k)\bar{A}_f=f((\bar{e}_k))=f((e_k)B)=(f(e_k))B=(e'_i)A_fB=(\bar{e}'_i)C^{-1}A_fB.$$

Рекомендуем проделать аналогичные вычисления в координатах.

Особенно важен частный случай $N=M$, $\{e_i\}=\{e'_i\}$, $\{\bar{e}_i\}=\{\bar{e}'_i\}$, $B=C$. Матрица линейного оператора f в новом базисе равна

$$\bar{A}_f=B^{-1}A_fB.$$

Отображение $M_n(\mathscr{K}) \to M_n(\mathscr{K})$: $A \mapsto B^{-1}AB$ называется *сопряжением* (посредством невырожденной матрицы B). Сопряжение является *автоморфизмом матричной алгебры* $M_n(\mathscr{K})$:

$$B^{-1}\left(\sum_{i=1}^{n} a_i A_i\right) B = \sum_{i=1}^{n} a_i B^{-1} A_i B, \quad a_i \in \mathscr{K};$$

$$B^{-1}(A_1 \ldots A_m) B = (B^{-1}A_1 B) \ldots (B^{-1}A_m B)$$

(в произведении справа внутренние сомножители B и B^{-1} попарно сокращаются, ибо стоят рядом).

Особую роль играют те функции от элементов $M_n(\mathscr{K})$, которые не меняются при замене матрицы на сопряженную, потому что с помощью этих функций можно строить *инварианты линейных операторов*: если φ — такая функция, то, полагая $\varphi(f) = \varphi(A_f)$, получим результат, зависящий лишь от f, но не от базиса, в котором пишется A_f. Вот два важных примера.

9. Определитель и след линейного оператора. Положим

$$\operatorname{Tr} f = \operatorname{Tr} A_f = \sum_{i=1}^{n} a_{ii}, \text{ где } A_f = (a_{ik})$$

(след — «trace» — матрицы A есть сумма элементов ее главной диагонали);

$$\det f = \det A_f.$$

Инвариантность определителя относительно сопряжения очевидна:

$$\det(B^{-1}AB) = (\det B)^{-1} \cdot \det A \cdot \det B = \det A.$$

Чтобы установить инвариантность следа, докажем более общий факт: *если A, B — такие матрицы, что AB и BA определены, то* $\operatorname{Tr} AB = \operatorname{Tr} BA$.

Действительно,

$$\operatorname{Tr} AB = \sum_i \sum_j a_{ij} b_{ji}, \qquad \operatorname{Tr} AB = \sum_j \sum_i b_{ji} a_{ij}.$$

Если теперь B невырождена, то, применяя доказанный факт к матрицам $B^{-1}A$ и B, получим

$$\operatorname{Tr}(B^{-1}AB) = \operatorname{Tr}(BB^{-1}A) = \operatorname{Tr} A.$$

В § 8 мы введем собственные значения матриц и операторов, симметрические функции от которых дадут другие инвариантные функции.

В заключение этого параграфа мы приведем определения, названия и стандартные обозначения для нескольких классов матриц над вещественными и комплексными числами, исключительно важных в теории групп и алгебр Ли и ее многочисленных приложениях, в частности в физике. Первый класс образуют так называемые *классические группы*: они действительно являются группами относительно матричного умножения. Второй класс образуют *алгебры Ли*: они составляют линейные пространства и устойчивы относительно операции взятия *коммутатора*: $[A, B] = AB - BA$. Параллелизм обозначений для этих классов получит некоторое объяснение в § 11 и в упражнении 8.

10. Классические группы.

а) *Полная линейная группа* $\mathrm{GL}(n, \mathscr{K})$. Она состоит из невырожденных квадратных матриц размера $n \times n$ над полем $\mathscr{K}$.

б) *Специальная линейная группа* $\mathrm{SL}(n, \mathscr{K})$. Она состоит из квадратных матриц размера $n \times n$ над полем $\mathscr{K}$ с определителем единица.

В этих двух случаях $\mathscr{K}$ может быть любым полем. Дальше мы ограничимся полями $\mathscr{K} = \mathbf{R}$ или $\mathbf{C}$, хотя существуют обобщения этих определений на другие поля.

в) *Ортогональная группа* $\mathrm{O}(n, \mathscr{K})$. Она состоит из матриц размера $n \times n$ с условием $AA^t = E_n$. Такие матрицы действительно образуют группу, ибо

$$E_n E_n^t = E_n, \quad A^{-1}(A^{-1})^t = A^{-1}(A^t)^{-1} = (A^t A)^{-1} = (E_n^t)^{-1} = E_n,$$

наконец,

$$(AB)(AB)^t = ABB^t A^t = AA^t = E_n.$$

При $\mathscr{K} = \mathbf{R}, \mathbf{C}$ эта группа называется вещественной или комплексной соответственно. Элементы группы $\mathrm{O}(n, \mathscr{K})$ называются ортогональными матрицами. Вместо $\mathrm{O}(n, \mathbf{R})$ обычно пишут $\mathrm{O}(n)$.

г) *Специальная ортогональная группа* $\mathrm{SO}(n, \mathscr{K})$. Она состоит из ортогональных матриц с определителем единица:

$$\mathrm{SO}(n, \mathscr{K}) = \mathrm{O}(n, \mathscr{K}) \cap \mathrm{SL}(n, \mathscr{K}).$$

Вместо $\mathrm{SO}(n, \mathbf{R})$ обычно пишут $\mathrm{SO}(n)$.

д) *Унитарная группа* $\mathrm{U}(n)$. Она состоит из комплексных матриц размера $n \times n$, удовлетворяющих условию $A\bar{A}^t = E_n$, где $\bar{A}$ — матрица, элементы которой комплексно сопряжены с соответствующими элементами матрицы A: если $A = (a_{ik})$, то $\bar{A} = (\bar{a}_{ik})$. Пользуясь равенством $\bar{A}\bar{B} = \overline{AB}$, нетрудно проверить, как и в случае в), что $\mathrm{U}(n)$ является группой, как в предыдущем пункте. Элементы $\mathrm{U}(n)$ называют *унитарными матрицами*.

Матрицу $\bar{A}^t$ часто называют *эрмитово сопряженной* с матрицей A; математики обычно обозначают ее A^*, а физики A^+. Заметим, что операция эрмитова сопряжения определена для комплексных матриц любых размеров.

е) *Специальная унитарная группа* $\mathrm{SU}(n)$. Она состоит из унитарных матриц с определителем единица:

$$\mathrm{SU}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n, \mathbf{C}).$$

Из определений ясно, что вещественные унитарные матрицы — это ортогональные матрицы: $\mathrm{O}(n) = \mathrm{U}(n) \cap \mathrm{GL}(n, \mathbf{R})$, $\mathrm{SO}(n) = \mathrm{U}(n) \cap \mathrm{SL}(n, \mathbf{R})$.

11. Классические алгебры Ли. (Матричной) алгеброй Ли называется любая аддитивная подгруппа квадратных матриц $M_n(\mathscr{K})$, замкнутая относительно операции коммутирования $[A, B] = AB - BA$. (Общее определение см. в упражнении 14.) Следующие множества матриц составляют классические алгебры Ли; обычно они даже образуют линейные пространства над $\mathscr{K}$ (иногда над $\mathbf{R}$, хотя $\mathscr{K} = \mathbf{C}$). Они не являются группами по умножению!

а) *Алгебра* $\mathrm{gl}(n, \mathscr{K})$. Она состоит из всех матриц $M_n(\mathscr{K})$

б) *Алгебра* $\mathrm{sl}(n, \mathscr{K})$. Она состоит из всех матриц $M_n(\mathscr{K})$ со следом нуль (иногда говорят «бесследных»). Замкнутость относительно коммутатора следует из формулы $\mathrm{Tr}[A, B] = 0$, доказанной в п. 9. Заметим, что Tr является линейной функцией на пространствах квадратных матриц и линейных операторов, так что $\mathrm{sl}(n, \mathscr{K})$ является линейным пространством над $\mathscr{K}$.

в) *Алгебра* $\mathrm{o}(n, \mathscr{K})$. Она состоит из всех матриц в $M_n(\mathscr{K})$, удовлетворяющих условию $A + A^t = 0$. Равносильное условие: $A = (a_{ik})$, где $a_{ii} = 0$ (если характеристика $\mathscr{K}$ отлична от двух), $a_{ik} = -a_{ki}$. Такие матрицы называются *антисимметричными*, или *кососимметричными*. Заметим, что $\mathrm{Tr}\, A = 0$ для всех $A \in \mathrm{o}(n, \mathscr{K})$.

Если $A^t = -A$, $B^t = -B$, то $[A, B]^t = [B^t, A^t] = [-B, -A] = -[A, B]$, так что $[A, B]$ кососимметрична. Такие матрицы образуют линейное пространство над $\mathscr{K}$.

Попутно заметим, что матрица A называется *симметричной*, если $A^t = A$. Множество таких матриц не замкнуто относительно коммутирования, но замкнуто относительно антикоммутирования $AB + BA$ или операции Йордана $\frac{1}{2}(AB + BA)$.

г) *Алгебра* $\mathrm{u}(n)$. Она состоит из комплексных матриц размера $n \times n$, удовлетворяющих условию $A + \bar{A}^t = 0$, или $a_{ik} = -\bar{a}_{ki}$. В частности, на диагонали у них стоят чисто мнимые элементы. Такие матрицы называются *эрмитово антисимметричными*, или *антиэрмитовыми*, или *косоэрмитовыми*. Они образуют линейное пространство над $\mathbf{R}$, но не над $\mathbf{C}$.

Если $A^t = -\overline{A}$, $B^t = -\overline{B}$, то

$$[A, B]^t = [B^t, A^t] = [-\overline{B}, -\overline{A}] = -\overline{[A, B]},$$

так что $\mathrm{u}(n)$ является алгеброй Ли.

Попутно заметим, что матрица A называется *эрмитово симметричной*, или просто *эрмитовой*, если $A = \bar{A}^t$, т. е. $a_{ki} = \bar{a}_{ik}$. Очевидно, вещественные эрмитовы матрицы симметричны, а антиэрмитовы — антисимметричны. В частности,

$$\mathrm{o}(n, \mathbf{R}) = \mathrm{u}(n) \cap \mathrm{sl}(n, \mathbf{R}).$$

Матрица A эрмитова, если матрица $i\bar{A}$ антиэрмитова, и наоборот.

д) *Алгебра* $\mathrm{su}(n)$ Это есть $\mathrm{u}(n) \cap \mathrm{sl}(n, \mathbf{C})$ — алгебра бесследных антиэрмитовых матриц. Они образуют $\mathbf{R}$-линейное пространство.

Во второй части книги, изучая линейные пространства, снабженные евклидовыми или эрмитовыми метриками, мы выясним геометрический смысл операторов, которые представлены матрицами из описанных классов, а также пополним наши списки.

УПРАЖНЕНИЯ

1. Сформулировать точно и доказать утверждение о том, что матрицы над полем, разбитые на блоки, можно умножать поблочно, если размеры и количество блоков согласованы:

$$(A_{ij})(B_{jk})=\left(\sum_j A_{ij}B_{jk}\right),$$

когда число столбцов в блоке A_{ij} равно числу строк в блоке B_{jk} и число блочных столбцов матрицы A равно числу блочных строк матрицы B.

2. Ввести понятие бесконечной матрицы (с бесконечным числом строк и/или столбцов). Найти условия, когда можно перемножать две такие матрицы над полем (примеры: финитные матрицы, т. е. матрицы только с конечным числом ненулевых элементов; матрицы, у которых в каждом столбце и/или каждой строке конечно число ненулевых элементов). Найти условия существования тройных произведений.

3. Доказать, что уравнение $XY-YX=E$ неразрешимо в конечных квадратных матрицах X, Y над полем нулевой характеристики. (*Указание.* Рассмотреть след обеих частей.) Найти решение этого уравнения в бесконечных матрицах. (*Указание.* Рассмотреть линейные операторы d/dx и умножения на x на пространстве всех многочленов от x и воспользоваться тем, что $\frac{d}{dx}(xf)-x\frac{d}{dx}f=$ $=f$.)

4. Описать явно классические группы и классические алгебры Ли в случаях $n=1$ и $n=2$. Построить изоморфизм групп $\mathrm{U}(1)$ и $\mathrm{SO}(2,\mathbf{R})$.

5. Следующие матрицы над $\mathbf{C}$ называются матрицами Паули:

$$\sigma_0=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

(Их ввел известный физик В. Паули, один из создателей квантовой механики, в своей теории спина электрона.) Проверить их свойства:

а) $[\sigma_a,\sigma_b]=2i\varepsilon_{abc}\sigma_c$, где $\{a,b,c\}=\{1,2,3\}$ и ε_{abc} — знак перестановки $\begin{pmatrix}1&2&3\\a&b&c\end{pmatrix}$.

б) $\sigma_a\sigma_b+\sigma_b\sigma_a=2\delta_{ab}\sigma_0$ (δ_{ab} — символ Кронекера).

в) Матрицы $i\sigma_1$, $i\sigma_2$, $i\sigma_3$ над $\mathbf{R}$ образуют базис $\mathrm{su}(2)$; над $\mathbf{C}$ — базис $\mathrm{sl}(2)$; матрицы σ_0, $i\sigma_1$, $i\sigma_2$, $i\sigma_3$ над $\mathbf{R}$ образуют базис $\mathrm{u}(2)$, над $\mathbf{C}$ — базис $\mathrm{gl}(2)$.

6. Следующие матрицы над $\mathbf{C}$ порядка 4 называются матрицами Дирака (здесь σ_a — матрицы Паули):

$$\gamma_0=\begin{pmatrix}\sigma_0&0\\0&-\sigma_0\end{pmatrix},\quad \gamma_1=\begin{pmatrix}0&\sigma_1\\-\sigma_1&0\end{pmatrix},\quad \gamma_2=\begin{pmatrix}0&\sigma_2\\-\sigma_2&0\end{pmatrix},\quad \gamma_3=\begin{pmatrix}0&\sigma_3\\-\sigma_3&0\end{pmatrix}$$

(их ввел известный физик П. А. М. Дирак, один из создателей квантовой механики, в своей теории релятивистского электрона со спином). Пользуясь результатами упражнений 5 и 1, проверить их свойства:

а) $\gamma_a\gamma_b+\gamma_b\gamma_a=2g_{ab}E_4$, где $g_{ab}=0$ при $a\neq b$, $g_{00}=1$, $g_{11}=g_{22}=$ $=g_{33}=-1$.

б) По определению, $\gamma_5=i\gamma_1\gamma_2\gamma_3\gamma_0$. Проверить, что $\gamma_5=-\begin{pmatrix}0&\sigma_0\\\sigma_0&0\end{pmatrix}$.

в) $\gamma_a\gamma_5=-\gamma_5\gamma_a$ для $a=0,1,2,3$; $\gamma_5^2=E_4$.

7. Проверить следующую таблицу размерностей классических алгебр Ли (как линейных пространств над соответствующими полями):

$\mathrm{gl}(n, \mathcal{K})$	$\mathrm{sl}(n, \mathcal{K})$	$\mathrm{o}(n, \mathcal{K})$	$\mathrm{u}(n)$	$\mathrm{su}(n)$
n^2	n^2-1	$\frac{n(n-1)}{2}$	n^2	n^2-1

8. Пусть A — квадратная матрица порядка n, ε — вещественная переменная, $\varepsilon \to 0$. Показать, что матрица $U = E + \varepsilon A$ «унитарна с точностью до ε^2» тогда и только тогда, когда A антиэрмитова:

$$U\overline{U}^t = E + O(\varepsilon^2) \Leftrightarrow A + \overline{A}^t = 0.$$

Сформулировать и доказать аналогичные утверждения для других пар классических групп и алгебр Ли.

9. Пусть $U = E + \varepsilon A$, $V = E + \varepsilon B$, где $\varepsilon \to 0$. Проверить, что

$$UVU^{-1}V^{-1} = E + \varepsilon^2 [A, B] + O(\varepsilon^3)$$

(выражение слева называется групповым коммутатором элементов U, V).

10. Ранг $\mathrm{rk}\, A$ матрицы над полем — это максимальное число ее линейно независимых столбцов. Доказать, что $\mathrm{rk}\, A_f = \dim \mathrm{Im}\, f$.

Доказать, что квадратная матрица ранга 1 представляется в виде произведения столбца на строку.

11. Пусть A, B — матрицы над полем размеров $m \times n$, $m_1 \times n_1$ и пусть зафиксированы нумерации всех mn элементов A и всех $m_1 n_1$ элементов B (например, последовательно по строкам). Тензорное произведение, или произведение Кронекера, $A \otimes B$ — это матрица размера $mn \times m_1 n_1$ с элементом $a_{ik} b_{lm}$ на месте $\alpha\beta$, где α — номер a_{ik}, β — номер b_{lm}. Проверить следующие утверждения:

а) $A \otimes B$ линейно по каждому из аргументов, когда другой фиксирован.

б) Если $m = n$, $m_1 = n_1$, то $\det(A \otimes B) = (\det A)^{m_1} (\det B)^{m}$.

12. Сколько нужно операций, чтобы перемножить две большие матрицы? В следующей серии утверждений излагается метод Штрассена, позволяющий значительно сократить их число, если матрицы действительно большие.

а) Умножение двух матриц порядка N обычным методом требует N^3 умножений и $N^2(N-1)$ сложений.

б) Имеет место следующая формула умножения при $N = 2$, обходящаяся 7 умножениями (вместо 8) за счет 18 сложений (вместо 4), коммутативность элементов не предполагается:

$$\begin{pmatrix} a & b \\ -c & -d \end{pmatrix} \begin{pmatrix} A & B \\ -C & -D \end{pmatrix} =$$

$$= \begin{pmatrix} (a+d)(A+D) - (b+d)(C+D) - d(A-C) - (a-b)D, & (a-b)D - a(D-B) \\ (d-c)A - d(A-C), & (a+d)(A+D) - (a+c)(A+B) - a(D-B) - (d-c)A \end{pmatrix}.$$

в) Применяя этот метод к матрицам порядка 2^n, разбитым на четыре $2^{n-1} \times 2^{n-1}$-блока, показать, что их можно перемножить, применив 7^n умножений и $6(7^n - 4^n)$ сложений.

г) Дополнив матрицы порядка N до ближайшего порядка 2^n нулями, показать, что для их умножения достаточно $O(N^{\log_2 7}) = O(N^{2,81})$ операций.

Не удастся ли вам придумать что-нибудь лучшее?

13. Пусть $L = M_n(\mathcal{K})$ — пространство квадратных матриц порядка n. Доказать, что для любого функционала $f \in L^*$ существует единственная матрица $A \in M_n(\mathcal{K})$ со свойством

$$f(X) = \mathrm{Tr}(AX)$$

для всех $X \in M_n(\mathcal{K})$. Вывести отсюда существование канонического изоморфизма

$$\mathcal{L}(L, L) \to [\mathcal{L}(L, L)]^*$$

для любого конечномерного пространства L.

14. $\mathscr{K}$-алгеброй Ли называется линейное пространство L над $\mathscr{K}$ вместе с бинарной операцией (коммутатор): $L \times L \to L$, обозначаемой $[\ ,\ ]$ и удовлетворяющей условиям:

а) коммутатор $[l, m]$ линеен по каждому из аргументов $l, m \in L$ при фиксированном другом аргументе;

б) $[l, m] = -[m, l]$ при всех l, m;

в) $[l_1, [l_2, l_3]] + [l_3, [l_1, l_2]] + [l_2, [l_3, l_1]] = 0$ (тождество Якоби) при всех $l_1, l_2, l_3 \in L$.

Проверить, что описанные в п. 11 классические алгебры Ли являются алгебрами Ли в смысле этого определения.

Более общо, проверить, что коммутатор $[X, Y] = XY - YX$ в любом ассоциативном кольце удовлетворяет тождеству Якоби.

§ 5. Подпространства и прямые суммы

1. В этом параграфе мы изучим некоторые геометрические свойства взаимного расположения подпространств конечномерного пространства L. Поясним первую задачу на простейшем примере. Пусть $L_1, L_1' \subset L$ — два подпространства. Естественно считать, что они одинаково расположены внутри L, если существует такой линейный автоморфизм $f: L \to L$, который переводит L_1 в L_1'. Для этого, конечно, необходимо, чтобы $\dim L_1 = \dim L_1'$, потому что f сохраняет все линейные соотношения и, значит, переводит базис L_1 в базис L_1'. Но этого и достаточно. В самом деле, выберем базисы $\{e_1, \ldots, e_m\}$ в L_1 и $\{e_1', \ldots, e_m'\}$ в L_1'. По теореме п. 12 § 2 их можно дополнить до базисов $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ и $\{e_1', \ldots, e_m', e_{m+1}', \ldots, e_n'\}$ пространства L. По предложению п. 3 § 3 существует линейное отображение $f: L \to L$, переводящее e_i в e_i' для всех i. Это отображение обратимо и переводит L_1 в L_1'. Таким образом, *все линейные подпространства одинаковой размерности одинаково расположены внутри L.*

Дальше естественно рассмотреть возможные расположения (упорядоченных) пар подпространств $L_1, L_2 \subset L$. Как выше, будем говорить, что пары (L_1, L_2) и (L_1', L_2') одинаково расположены, если существует такой линейный автоморфизм $f: L \to L$, что $f(L_1) = L_1'$ $f(L_2) = L_2'$. Снова равенства $\dim L_1 = \dim L_1'$ и $\dim L_2 = \dim L_2'$ являются необходимыми для одинаковой расположенности. Однако, вообще говоря, этих условий уже недостаточно. Действительно, если (L_1, L_2) и (L_1', L_2') одинаково расположены, то f переводит подпространство $L_1 \cap L_2$ в $L_1' \cap L_2'$, и потому необходимо также условие $\dim(L_1 \cap L_2) = \dim(L_1' \cap L_2')$. Если $\dim L_1$ и $\dim L_2$ фиксированы, но L_1 и L_2 в остальном произвольны, то $\dim(L_1 \cap L_2)$ может принимать, вообще говоря, целый ряд значений.

Чтобы выяснить, какими они могут быть, введем понятие суммы линейных подпространств.

2. **Определение.** *Пусть $L_1, \ldots, L_n \subset L$ — линейные подпространства в L. Их суммой называется множество*

$$\sum_{i=1}^{n} L_i = L_1 + \ldots + L_n = \left\{ \sum_{i=1}^{n} l_i \mid l_i \in L_i \right\}.$$

Легко убедиться, что сумма также является линейным подпространством и что эта операция сложения ассоциативна и коммутативна, так же как и операция пересечения линейных подпространств. Другое описание суммы $L_1 + \ldots + L_n$ состоит в том, что это *наименьшее подпространство в* L, *содержащее все* L_i.

Следующая теорема связывает размерности суммы двух подпространств и их пересечения:

3. Теорема. *Если* $L_1, L_2 \subset L$ *конечномерны, то* $L_1 \cap L_2$ *и* $L_1 + L_2$ *конечномерны и*

$$\dim (L_1 \cap L_2) + \dim (L_1 + L_2) = \dim L_1 + \dim L_2.$$

Доказательство. $L_1 + L_2$ является линейной оболочкой объединения базисов L_1 и L_2 и потому конечномерно; $L_1 \cap L_2$ содержится в конечномерных пространствах L_1 и L_2.

Положим $m = \dim L_1 \cap L_2$, $n = \dim L_1$, $p = \dim L_2$. Выберем базис $\{e_1, \ldots, e_m\}$ пространства $L_1 \cap L_2$. По теореме п. 12 § 2 его можно дополнить до базисов пространств L_1 и L_2: пусть это будет $\{e_1, \ldots, e_m, e'_{m+1}, \ldots, e'_n\}$ и $\{e_1, \ldots, e_m, e''_{m+1}, \ldots, e''_p\}$. Назовем такую пару базисов в L_1 и L_2 согласованной.

Мы докажем сейчас, что семейство $\{e_1, \ldots, e_m, e'_{m+1}, \ldots, e'_n, e''_{m+1}, \ldots, e''_p\}$ составляет базис пространства $L_1 + L_2$. Отсюда будет следовать утверждение теоремы:

$$\dim (L_1 + L_2) = p + n - m = \dim L_1 + \dim L_2 - \dim L_1 \cap L_2.$$

Поскольку каждый вектор из $L_1 + L_2$ есть сумма векторов из L_1 и L_2, т. е. сумма линейных комбинаций $\{e_1, \ldots, e_m, e'_{m+1}, \ldots \ldots, e'_n\}$ и $\{e_1, \ldots, e_m, e''_{m+1}, \ldots, e''_p\}$, объединение этих семейств порождает $L_1 + L_2$. Поэтому остается лишь проверить его линейную независимость.

Предположим, что существует нетривиальная линейная зависимость

$$\sum_{i=1}^{m} x_i e_i + \sum_{j=m+1}^{n} y_j e'_j + \sum_{k=m+1}^{p} z_k e''_k = 0.$$

Тогда обязательно должны существовать индексы j и k, для которых $y_j \neq 0$ и $z_k \neq 0$: иначе мы получили бы нетривиальную линейную зависимость между элементами базиса L_1 или L_2.

Следовательно, ненулевой вектор $\sum_{k=m+1}^{p} z_k e''_k \in L_2$ должен лежать также в L_1, либо он равен $-\left(\sum_{i=1}^{m} x_i e_i + \sum_{j=m+1}^{n} y_j e'_j\right)$. Значит, он лежит в $L_1 \cap L_2$ и потому представим в виде линейной комбинации векторов $\{e_1, \ldots, e_m\}$, составляющих базис $L_1 \cap L_2$. Но это представление дает нетривиальную линейную зависимость между векторами $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e''_p\}$, что противоречит их определению. Теорема доказана.

4. Следствие. *Пусть* $n_1 \leqslant n_2 \leqslant n$ — *размерности пространств* L_1, L_2 *и* L *соответственно. Тогда числа* $i = \dim L_1 \cap L_2$ *и* $s = \dim (L_1 + L_2)$ *могут принимать любые значения, подчиненные условиям* $0 \leqslant i \leqslant n_1$, $n_2 \leqslant s \leqslant n$ *и* $i + s = n_1 + n_2$.

Доказательство. Необходимость условий следует из включений $L_1 \cap L_2 \subset L_1$, $L_2 \subset L_1 + L_2 \subset L$ и из теоремы п. 3. Для доказательства достаточности выберем $s = n_1 + n_2 - i$ линейно независимых векторов в L: $\{e_1, \ldots, e_i; e'_{i+1}, \ldots, e'_{n_1}; e''_{i+1}, \ldots, e''_{n_2}\}$ и обозначим через L_1, L_2 линейные оболочки $\{e_1, \ldots, e_i; e'_{i+1}, \ldots, e'_{n_1}\}$ и $\{e_1, \ldots, e_i; e''_{i+1}, \ldots, e''_{n_2}\}$ соответственно. Как в теореме, нетрудно проверить, что $L_1 \cap L_2$ есть линейная оболочка $\{e_1, \ldots, e_i\}$.

5. Теперь мы можем устаиовить, что инварианты $n_1 = \dim L_1$, $n_2 = \dim L_2$ и $i = \dim L_1 \cap L_2$ полностью характеризуют расположение пары подпространств (L_1, L_2) в L. Для доказательства возьмем другую пару (L'_1, L'_2) с теми же инвариантами, построим согласованные пары базисов для L_1, L_2 и L'_1, L'_2, затем их объединения — базисы $L_1 + L_2$ и $L'_1 + L'_2$, как в доказательстве теоремы п. 3, наконец, продолжим эти объединения до двух базисов L. Линейный автоморфизм, переводящий первый базис во второй, устанавливает одинаковость расположения L_1, L_2 и L'_1, L'_2.

6. Общее положение. В обозначениях предыдущего пункта будем говорить, что подпространства $L_1, L_2 \subset L$ находятся *в общем положении*, если их пересечение имеет наименьшую, а сумма — наибольшую размерность, допускаемую неравенствами из следствия п. 4.

Например, две плоскости в трехмерном пространстве находятся в общем положении, если они пересекаются по прямой, а в четырехмерном пространстве, — если они пересекаются по точке.

Другой термин для того же понятия: L_1 и L_2 *пересекаются трансверсально.*

Название «общее положение» обусловлено тем, что в некотором смысле большинство пар подпространств (L_1, L_2) находится в общем положении, а другие расположения являются вырожденными. Уточнить это утверждение можно разными способами. Один из них состоит в том, чтобы описать множество пар подпространств некоторыми параметрами и проверить, что пара не находится в общем положении, только если эти параметры удовлетворяют дополнительным соотношениям, которым общие параметры не удовлетворяют.

Другой способ, который годится для $\mathcal{K} = \mathbf{R}$ и $\mathbf{C}$, состоит в следующем: выбрать в L некоторый базис, определить L_1 и L_2 двумя системами линейных уравнений и показать, что можно как угодно мало изменить коэффициенты этих уравнений («пошевелить L_1 и L_2») так, чтобы новая пара оказалась в общем положении.

Можно было бы пытаться далее рассматривать инварианты, характеризующие взаимное расположение троек, четверок и большего числа подпространств в L. Комбинаторные трудности здесь быстро растут, и для решения этой задачи нужна другая техника; кроме того, начиная с четверок, расположение перестает характеризоваться только дискретными инвариантами типа размерностей разных сумм и пересечений.

Заметим еще, что, как показывает наша «физическая» интуиция, расположение, скажем, прямой относительно плоскости характеризуется углом между ними. Но, как мы отмечали в § 1, понятие угла требует введения дополнительной структуры. В чисто линейной ситуации есть только различие между «нулевым» и «ненулевым» углом.

Теперь мы изучим один частный, но очень важный класс взаимных расположений n-ок подпространств.

7. Определение. *Пространство L является прямой суммой своих подпространств $L_1, \ldots, L_n$, если каждый вектор $l \in L$ однозначно представляется в виде $\sum_{i=1}^{n} l_i$, где $l_i \in L_i$.*

Когда условия определения выполнены, мы пишем $L = L_1 \oplus \ldots \oplus L_n$, или $L = \bigoplus_{i=1}^{n} L_i$. Например, если $\{e_1, \ldots, e_n\}$ — базис L, а $L_i = \mathscr{K} e_i$ — линейная оболочка вектора e_i, то $L = \bigoplus_{i=1}^{n} L_i$. Очевидно, если $L = \bigoplus_{i=1}^{n} L_i$, то $L = \sum_{i=1}^{n} L_i$; последнее условие является более слабым.

8. Теорема. *Пусть $L_1, \ldots, L_n \subset L$ — подпространства в L. $L = \bigoplus_{i=1}^{n} L_i$ тогда и только тогда, когда выполнено любое из следующих двух условий:*

а) $\sum_{i=1}^{n} L_i = L$ *и* $L_j \cap \left(\sum_{i \neq j} L_i\right) = \{0\}$ *для всех* $1 \leqslant j \leqslant n$;

б) $\sum_{i=1}^{n} L_i = L$ *и* $\sum_{i=1}^{n} \dim L_i = \dim L$ (*здесь предполагается, что L конечномерно*).

Доказательство. а) Однозначность представления любого вектора $l \in L$ в виде $\sum_{i=1}^{n} l_i$, $l_i \in L$, равносильна однозначности такого представления для нулевого вектора. В самом деле, если $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} l_i'$, то $0 = \sum_{i=1}^{n} (l_i - l_i')$, и наоборот. Если имеется нетривиальное представление $0 = \sum_{i=1}^{n} l_i$, в котором, скажем, $l_j \neq 0$, то $l_j = -\sum_{i \neq j} l_i \in L_j \cap \left(\sum_{i \neq j} L_i\right)$, так что условие а) нарушено. Обращая это рассуждение, получаем, что из нарушения условия а) следует неоднозначность представления нуля.

б) Если $\bigoplus_{i=1}^{n} L_i = L$, то во всяком случае

$$\sum_{i=1}^{n} L_i = L \text{ и } \sum_{i=1}^{n} \dim L_i \geqslant \dim L,$$

потому что объединение базисов L_i порождает L и, значит, содержит базис L. По теореме п. 3, примененной к L_j и $\sum_{i \neq j} L_i$, имеем

$$\dim L_j \cap \Bigl(\sum_{i \neq j} L_i\Bigr) + \dim L = \dim L_j + \dim \Bigl(\sum_{i \neq j} L_i\Bigr).$$

Но размерность пересечения слева нулевая по предыдущему утверждению. Кроме того, если сумма всех L_i прямая, то и сумма всех L_i, кроме L_j, прямая, и мы можем по индукции считать, что

$$\dim \Bigl(\sum_{i \neq j} L_i\Bigr) = \sum_{i \neq j} \dim L_i. \text{ Поэтому } \sum_{i} \dim L_i = \dim L.$$

Наоборот, если $\sum_i \dim L_i = \dim L$, то объединение базисов всех L_i состоит из $\dim L$ элементов и порождает все L, а потому является базисом в L. В самом деле, нетривиальное представление нуля $0 = \sum_{i=1}^{n} l_i$, $l_i \in L_i$, дало бы нетривиальную линейную комбинацию элементов этого базиса, равную нулю, что невозможно.

Рассмотрим теперь связь между разложениями в прямую сумму и специальными линейными операторами — проекторами.

9. Определение. *Линейный оператор $p\colon L \to L$ называется проектором, если $p^2 = p \circ p = p$.*

Прямому разложению $L = \bigoplus_{i=1}^{n} L_i$ естественно сопоставляются n проекторов, которые определяются так: для любых $l_j \in L_j$

$$p_i \Bigl(\sum_{j=1}^{n} l_j\Bigr) = l_i.$$

Поскольку любой элемент $l \in L$ однозначно представляется в виде $\sum_{j=1}^{n} l_j$, $l_j \in L_j$, отображения p_i определены корректно. Их линейность и свойство $p_i^2 = p_i$ проверяются прямо из определения. Очевидно, $L_i = \operatorname{Im} p_i$.

Сверх того, если $i \neq j$, то $p_i p_j = 0$: вектору l_i отвечает представление $l_i = \sum_{j=1}^{n} l'_j$, где $l'_j = 0$ при $i \neq j$, $l'_i = l_i$.

Наконец, $\sum_{i=1}^{n} p_i = \mathrm{id}$, ибо $\Bigl(\sum_{i=1}^{n} p_i\Bigr)\Bigl(\sum_{j=1}^{n} l_j\Bigr) = \sum_{j=1}^{n} l_j$, если $l_j \in L_j$. Наоборот, по такой системе проекторов можно определить отвечающее ей прямое разложение.

10. Теорема. *Пусть* $p_1, \ldots, p_n$: $L \to L$ — *конечное множество проекторов, удовлетворяющих условиям*

$$\sum_{i=1}^{n} p_i = \mathrm{id}, \quad p_i p_j = 0 \text{ при } i \neq j.$$

Положим $L_i = \operatorname{Im} p_i$. *Тогда* $L = \bigoplus_{i=1}^{n} L_i$.

Доказательство. Применяя оператор $\mathrm{id} = \sum_{i=1}^{n} p_i$ к любому вектору $l \in L$, получим $l = \sum_{i=1}^{n} p_i(l)$, где $p_i(l) \in L_i$. Поэтому $L = \sum_{i=1}^{n} L_i$. Для доказательства того, что эта сумма прямая, применим критерий а) из теоремы п. 8. Пусть $l \in L_j \cap \left(\sum_{i \neq j} L_i\right)$. В силу определения пространств $L_i = \operatorname{Im} p_i$ существуют такие векторы $l_1, \ldots, l_n$, что

$$l = p_j(l_j) = \sum_{i \neq j} p_i(l_i).$$

Применим к этому равенству оператор p_j и воспользуемся тем, что $p_j^2 = p_j$, $p_j p_i = 0$ при $i \neq j$. Получим

$$p_j(l_j) = \sum_{i \neq j} p_j p_i(l_i) = 0.$$

Следовательно, $l = 0$, что завершает доказательство.

11. Прямые дополнения. Если L — конечномерное пространство, то для любого подпространства $L_1 \subset L$ можно выбрать такое подпространство $L_2 \subset L$, что $L = L_1 \oplus L_2$; кроме тривиальных случаев $L_1 = \{0\}$ или $L_1 = L$ этот выбор неоднозначен. В самом деле, выбрав базис $\{e_1, \ldots, e_m\}$ в L_1 и продолжив его до базиса $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ в L, мы можем взять в качестве L_2 линейную оболочку векторов $\{e_{m+1}, \ldots, e_n\}$.

12. Внешние прямые суммы. До сих пор мы исходили из семейства подпространств $L_1, \ldots, L_n$ одного и того же пространства L. Пусть теперь $L_1, \ldots, L_n$ — пространства, не вложенные заранее в общее пространство. Определим их *внешнюю прямую сумму* L следующим образом:

а) L как множество есть $L_1 \times \ldots \times L_n$, т. е. элементы L суть семейства $(l_1, \ldots, l_n)$, где $l_i \in L_i$.

б) Сложение и умножение на скаляр производятся покоординатно:

$$(l_1, \ldots, l_n) + (l_1', \ldots, l_n') = (l_1 + l_1', \ldots, l_n + l_n'),$$
$$a(l_1, \ldots, l_n) = (al_1, \ldots, al_n).$$

Нетрудно проверить, что L удовлетворяет аксиомам линейного пространства. Отображение f_i: $L_i \to L$, $f_i(l) = (0, \ldots, 0, l, 0, \ldots, 0)$

(1 на i-м месте) является линейным вложением L_i в L, и из определений немедленно следует, что $L=\bigoplus\limits_{i=1}^{n} f_i(L_i)$. Отождествив L_i с $f_i(L_i)$, получим линейное пространство, в котором L_i содержатся и которое разлагается в прямую сумму L_i. Это оправдывает название внешней прямой суммы. Часто удобно обозначать внешнюю прямую сумму также $\bigoplus\limits_{i=1}^{n} L_i$.

13. Прямые суммы линейных отображений. Пусть $L=\bigoplus\limits_{i=1}^{n} L_i$, $M=\bigoplus\limits_{i=1}^{n} M_i$, $f\colon L\to M$ — такое линейное отображение, что $f(L_i)\subset \subset M_i$. Обозначим через f_i индуцированное линейное отображение $L_i\to M_i$. В таком случае принято писать $f=\bigoplus\limits_{i=1}^{n} f_i$. Аналогично определяется внешняя прямая сумма линейных отображений. Выбрав в L и M базисы, являющиеся объединением базисов L_i и M_i соответственно, мы получаем, что матрица f является объединением стоящих по диагонали блоков, которые представляют собой матрицы отображений f_i; на остальных местах стоят нули.

14. Ориентация вещественных линейных пространств. Пусть L— конечномерное линейное пространство над полем вещественных чисел. Два упорядоченных базиса $\{e_i\}$ и $\{e_i'\}$ в нем всегда одинаково расположены в том смысле, что имеется единственный линейный изоморфизм $f\colon L\to L$, переводящий e_i в e_i'. Поставим, однако, более тонкий вопрос: когда можно перевести базис $\{e_i\}$ в базис $\{e_i'\}$ *непрерывным движением*, или *деформацией*, т. е. найти такое семейство $f_t\colon L\to L$ линейных изоморфизмов, непрерывно зависящее от параметра $t\in[0,1]$, что $f_0=\mathrm{id}, f_1(e_i)=e_i'$ для всех i? (Непрерывно зависеть от t должны просто элементы матрицы f в каком-нибудь из базисов.) Для этого имеется очевидное необходимое условие: поскольку при изменении t определитель f_t меняется непрерывно и не проходит через нуль, знак $\det f_t$ должен совпадать со знаком $\det f_0=1$, т. е. $\det f_t>0$.

Верно и обратное утверждение: *если определитель матрицы перехода от базиса $\{e_i\}$ к базису $\{e_i'\}$ положителен, то $\{e_i\}$ можно перевести в $\{e_i'\}$ непрерывным движением.*

Это утверждение, очевидно, можно сформулировать иначе: *любую вещественную матрицу с положительным определителем можно соединить с единичной матрицей непрерывной кривой, состоящей из невырожденных матриц* (*множество вещественных невырожденных матриц с положительным определителем связно*). Чтобы перейти от языка базисов к языку матриц, достаточно работать не с парой базисов $\{e_i\}$, $\{f_1(e_i)\}$, а с матрицей перехода от первого ко второму.

Мы докажем это утверждение, разбив его на серию шагов.

а) *Пусть* $A = B_1 \ldots B_n$, *где* A, B_i — *матрицы с положительными определителями. Если все* B_j *можно соединить с* E *непрерывной кривой, то это верно и для* A.

Действительно, пусть $B_j(t)$ — такие непрерывные кривые в пространстве невырожденных матриц, что $B_j(0) = B_j$, $B_j(1) = E$. Тогда кривая $A(t) = B_1(t) \ \ldots \ B_n(t)$ соединяет A с E.

б) *Если* A *можно соединить непрерывной кривой с* B, *а* B *с* E, *то можно соединить* A *с* E.

Действительно, если $A(t)$ такова, что $A(0) = A$, $A(1) = B$, и $B(t)$ такова, что $B(0) = B$, $B(1) = E$, то кривая

$$t \mapsto \begin{cases} A(2t) & \text{при } 0 \leqslant t \leqslant 1/2, \\ B(2t-1) & \text{при } 1/2 \leqslant t \leqslant 1 \end{cases}$$

соединяет A с E. Трюк с изменением масштаба и начала отсчета t использован лишь потому, что мы условились параметризовать кривые матриц числами $t \in [0, 1]$. Очевидно, можно пользоваться любыми промежуточными интервалами параметризации, проводить последовательно все нужные деформации и менять масштаб лишь в конце. Поэтому дальше мы не будем заботиться об интервалах параметризации.

в) *Любую квадратную невырожденную матрицу* A *можно представить в виде произведения конечного числа элементарных матриц следующих типов:* $F_{s,t}$, $F_{s,t}(\lambda)$, $F_s(\lambda)$, $\lambda \in \mathbf{R}$. *Обозначим через* E_{st} *матрицу с единицей на месте* (s, t) *и нулями на остальных местах. Тогда по определению*

$$F_{s,t} = E - E_{ss} - E_{tt} + E_{st} + E_{ts};$$
$$F_{s,t}(\lambda) = E + \lambda E_{st}; \quad F_s(\lambda) = E + (\lambda - 1) E_{ss}.$$

Этот результат доказан в книге «Введение в алгебру», гл. 2, § 4, следствие из теоремы п. 5.

г) Пусть теперь матрица A представлена в виде произведения элементарных матриц. Предполагая ее определитель положительным, покажем, как соединить ее с E с помощью нескольких последовательных деформаций, пользуясь результатами шагов а) и б).

Прежде всего, $\det F_{s,t}(\lambda) = 1$ при всех λ и $F_{s,t}(0) = E$. Меняя в исходных сомножителях λ от начального значения до нуля, мы можем деформировать все такие множители в E, так что можно считать, что их нет с самого начала.

Матрицы $F_s(\lambda)$ диагональны: на месте (s, s) стоит λ, на остальных — единицы. Изменим λ от начального значения до $+1$ или -1 в соответствии со знаком начального значения. Результатом деформации будет либо единичная матрица, либо матрица линейного отображения, меняющего один из базисных векторов на противоположный и оставляющий остальные на месте.

Результатом деформации A на этом этапе будет матрица композиции двух преобразований: одно сводится к перестановке векторов базиса ($F_{s,t}$ меняет местами s-й и t-й вектор), другое меняет знаки части векторов (композиция $F_s(+1)$ и $F_t(-1)$).

Любую перестановку можно разложить в произведение попарных перестановок. Матрицу перестановки векторов базиса в плоскости $\begin{pmatrix}0&1\\1&0\end{pmatrix}$ можно соединить с $\begin{pmatrix}-1&0\\0&1\end{pmatrix}$ кривой $\begin{pmatrix}-\cos t&\sin t\\\sin t&\cos t\end{pmatrix}$, $\pi/2\geqslant t\geqslant 0$. Очевидно, разнеся элементы последней матрицы по местам (s,s), (s,t), (t,s), (t,t), получим соответствующую деформацию в любой размерности, уничтожающую множители $F_{s,t}$.

К этому моменту A превратилась в диагональную матрицу с элементами ± 1 на диагонали, причем число минус единиц четно, ибо определитель A положителен. Матрицу $\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$ можно соединить с $\begin{pmatrix}1&0\\0&1\end{pmatrix}$ кривой $\begin{pmatrix}\cos t&-\sin t\\\sin t&\cos t\end{pmatrix}$, $\pi\geqslant t\geqslant 0$. Собрав все -1 попарно и проведя такие деформации всех пар, мы завершим доказательство.

Вернемся теперь к ориентации.

Будем говорить, что базисы $\{e_i\}$, $\{e'_i\}$ *одинаково ориентированы*, если определитель матрицы перехода от одного из них к другому положителен. Ясно, что множество упорядоченных базисов L разбивается в точности на два класса, состоящих из одинаково ориентированных базисов, тогда как базисы из разных классов ориентированы по-разному (или противоположно).

Выбор одного из этих классов называется *ориентацией пространства* L.

Ориентация одномерного пространства соответствует указанию «положительного направления в нем» или полупрямой $\mathbf{R}_+e=\{ae\,|\,a>0\}$, где e — любой вектор, определяющий ориентацию.

В двумерном пространстве задание ориентации с помощью базиса $\{e_1, e_2\}$ можно представлять себе как указание «положительного направления вращения» плоскости от e_1 к e_2. Это интуитивно согласуется с тем, что базис $\{e_2, e_1\}$ задает противоположную ориентацию (определитель матрицы перехода $\begin{pmatrix}0&1\\1&0\end{pmatrix}$ равен -1) и противоположное направление вращения.

В общем случае переход от базиса $\{e_i\}$ к базису $\{e'_i\}$, состоящему из тех же векторов, но в другом порядке, сохраняет ориентацию, если перестановка четная, и меняет ее, если перестановка нечетная. Замена знака у одного из векторов e_i меняет ориентацию на противоположную.

В трехмерном физическом пространстве выбор конкретной ориентации может быть связан с физиологическими особенностями человека — асимметрией правой и левой стороны. Левая сторона — это та, где у подавляющего большинства людей находится сердце. Большой, указательный и средний пальцы левой руки, согнутые по направлению к ладони, в линейно независимом положении образуют упорядоченный базис, фиксирующий ориентацию («правило левой руки»). Вопрос о том, существуют ли чисто физические процессы, позволяющие выбрать ориентацию пространства, т. е.

«неинвариантные относительно зеркального отражения», был решен около двадцати лет назад положительно, ко всеобщему изумлению, экспериментом, установившим несохранение четности в слабых взаимодействиях.

УПРАЖНЕНИЯ

1. Пусть (L_1, L_2, L_3) — упорядоченная тройка плоскостей в $\mathcal{K}^3$, попарно различных. Доказать, что имеются два возможных типа взаимного расположения таких троек, характеризующихся тем, что $\dim L_1 \cap L_2 \cap L_3 = 0$ или 1. Какой из этих типов следует считать общим?

2. Доказать, что тройки попарно разных прямых в $\mathcal{K}^3$ все одинаково расположены, а для четверок это уже неверно.

3. Пусть $L_1 \subset L_2 \subset \ldots \subset L_n$ — флаг в конечномерном пространстве L, $m_i = \dim L_i$. Доказать, что если $L'_1 \subset \ldots \subset L'_n$ — другой флаг, $m'_i = \dim L'_i$, то автоморфизм L, переводящий первый флаг во второй, существует тогда и только тогда, когда $m_i = m'_i$ для любого i.

4. То же для прямых разложений.

5. Доказать утверждения пятого абзаца п. 6.

6. Пусть $p\colon L \to L$ — проектор. Доказать, что $L = \operatorname{Ker} p \oplus \operatorname{Im} p$. Вывести отсюда, что в подходящем базисе L любой проектор p представлен матрицей вида

$$\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right),$$

где $r = \dim \operatorname{Im} p$.

7. Пусть L — n-мерное пространство над конечным полем из q элементов

а) Вычислить количество k-мерных подпространств в L, $1 \leqslant k \leqslant n$.

б) Вычислить количество пар подпространств $L_1, L_2 \subset L$ с данными размерностями L_1, L_2 и $L_1 \cap L_2$. Убедиться, что при $q \to \infty$ доля этих пар, находящихся в общем положении, среди всех пар с данными $\dim L_1$, $\dim L_2$ стремится к 1.

§ 6. Факторпространства

1. Пусть L — линейное пространство, $M \subset L$ — его линейное подпространство, а $l \in L$ — вектор. Различные вопросы приводят к рассмотрению множеств вида

$$l + M = \{l + m \mid m \in M\},$$

«сдвигов» линейного пространства M на вектор l. Вскоре мы убедимся, что такие сдвиги не обязаны быть линейными подпространствами в L; их называют *линейными подмногообразиями.* Начнем с доказательства следующей леммы.

2. Лемма. *$l_1 + M_1 = l_2 + M_2$ тогда и только тогда, когда $M_1 = M_2 = M$ и $l_1 - l_2 \in M$. Таким образом, всякое линейное подмногообразие однозначно определяет линейное подпространство M, сдвигом которого оно является. Вектор же сдвига определяется лишь с точностью до элемента из этого подпространства.*

Доказательство. Прежде всего, пусть $l_1 - l_2 \in M$. Положим $l_1 - l_2 = m_0$. Имеем

$$l_1 + M = \{l_1 + m \mid m \in M\}, \quad l_2 + M = \{l_1 + m - m_0 \mid m \in M\}.$$

Но когда m пробегает все векторы из M, $m - m_0$ тоже пробегает все векторы из M. Поэтому $l_1 + M = l_2 + M$.

Наоборот, пусть $l_1+M_1=l_2+M_2$. Положим $m_0=l_1-l_2$. Из определения ясно, что тогда $m_0+M_1=M_2$. Так как $0\in M_2$, мы должны иметь $m_0\in M_1$. Значит, $m_0+M_1=M_1$ по рассуждению в предыдущем абзаце, так что $M_1=M_2=M$. Это завершает доказательство.

3. Определение. *Факторпространством L/M линейного пространства L по M называется множество всех линейных подмногообразий в L, являющихся сдвигами подпространства M, со следующими операциями*:

а) $(l_1+M)+(l_2+M)=(l_1+l_2)+M$,

б) $a(l_1+M)=al_1+M$ *для любых* $l_1,\ l_2\in L,\ a\in\mathscr{K}$.

Эти операции определены корректно и превращают L/M в линейное пространство над полем $\mathscr{K}$.

4. Проверка корректности определения. Она состоит из следующих шагов:

а) *Если* $l_1+M=l_1'+M$ *и* $l_2+M=l_2'+M$, *то* $l_1+l_2+M=l_1'+l_2'+M$.

В самом деле, из леммы п. 2 следует, что $l_1-l_1'=m_1\in M$ и $l_2-l_2'=m_2\in M$. Поэтому снова по лемме п. 2

$$(l_1+l_2)+M=(l_1'+l_2')+(m_1+m_2)+M=(l_1'+l_2')+M,$$

ибо $m_1+m_2\in M$.

б) *Если* $l_1+M=l_1'+M$, *то* $al_1+M=al_1'+M$.

В самом деле, снова полагая $l_1-l_1'=m\in M$, имеем $al_1-al_1'=am\in M$, и применение леммы п. 2 дает требуемое.

Таким образом, сложение и умножение на скаляр действительно однозначно определены в L/M. Остается проверить аксиомы линейного пространства. Но они сразу же следуют из соответствующих формул в L. Например, одна из формул дистрибутивности проверяется так:

$$a[(l_1+M)+(l_2+M)]=a((l_1+l_2)+M)=a(l_1+l_2)+M=$$
$$=al_1+al_2+M=(al_1+M)+(al_2+M)=a(l_1+M)+a(l_2+M).$$

Здесь последовательно используются: определение сложения в L/M, определение умножения на скаляр в L/M, дистрибутивность в L и снова определение сложения и умножения на скаляр в L/M.

5. Замечания. а) Из определения видно, что аддитивная группа L/M совпадает с факторгруппой аддитивной группы L по аддитивной группе M. В частности, подмногообразие $M\subset L$ является нулем в L/M.

б) Имеется каноническое отображение $f\colon L\to L/M$: $f(l)=l+M$. Оно сюръективно, а его слои — прообразы элементов — суть как раз подмногообразия, отвечающие этим элементам. Действительно, по лемме п. 2

$$f^{-1}(l_0+M)=\{l\in L\mid l+M=l_0+M\}=\{l\in L\mid l-l_0\in M\}=l_0+M.$$

Заметим, что в этой цепочке равенств $l_0 + M$ первый раз рассматривается как *элемент* множества L/M, а остальные — как *подмножества* в L.

Из п. 4 ясно, что f — линейное отображение, а лемма п. 2 показывает, что $\operatorname{Ker} f = M$, ибо $l_0 + M = M$ тогда и только тогда, когда $l_0 \in M$.

6. Следствие. *Если* L *конечномерно, то* $\dim L/M = \dim L - \dim M$.

Доказательство. Применить теорему п. 12, § 3 к построенному отображению $L \to L/M$.

Многие важные задачи в математике приводят к ситуации, когда пространства $M \subset L$ бесконечномерны, а факторпространство L/M конечномерно. В этом случае пользоваться следствием п. 6 нельзя, и вычисление $\dim L/M$ обычно становится нетривиальной задачей. Число $\dim L/M$ вообще называется *коразмерностью* подпространства M в L и обозначается $\operatorname{codim} M$ или $\operatorname{codim}_L M$.

7. Поставим следующую задачу: даны два отображения $f\colon L \to M$ и $g\colon L \to N$; когда существует такое отображение $h\colon M \to N$, что $g = hf$? На языке диаграмм: когда можно вложить диаграмму

$$\begin{tikzcd} & L \arrow[dl, "f"'] \arrow[dr, "g"] & \\ M & & N \end{tikzcd}$$

в коммутативный треугольник

$$\begin{tikzcd} & L \arrow[dl, "f"'] \arrow[dr, "g"] & \\ M \arrow[rr, "h"] & & N \end{tikzcd}$$

(ср. § 13 о коммутативных диаграммах). Ответ для линейных отображений дается следующим результатом.

8. Предложение. *Для существования* h *необходимо и достаточно, чтобы* $\operatorname{Ker} f \subset \operatorname{Ker} g$; *если это условие выполнено и* $\operatorname{Im} f = M$, *то* h *единствен.*

Доказательство. Если h существует, из $g = hf$ следует, что $g(l) = hf(l) = 0$, коль скоро $f(l) = 0$. Поэтому $\operatorname{Ker} f \subset \operatorname{Ker} g$.

Наоборот, пусть $\operatorname{Ker} f \subset \operatorname{Ker} g$. Построим сначала h на подпространстве $\operatorname{Im} f \subset M$. Единственная возможность состоит в том, чтобы положить $h(m) = g(l)$, если $m = f(l)$. Нужно проверить, что h определено однозначно и линейно на $\operatorname{Im} f$. Первое следует из того, что если $m = f(l_1) = f(l_2)$, то $l_1 - l_2 \in \operatorname{Ker} f \subset \operatorname{Ker} g$, откуда $g(l_1) = g(l_2)$. Второе следует автоматически из линейности f и g.

Теперь достаточно продолжить отображение h с подпространства $\operatorname{Im} f \subset M$ на все пространство M, например, выбрав базис $\operatorname{Im} f$, дополнив его до базиса M и положив h равным нулю на дополняющих векторах.

9. Пусть $g\colon L\to M$ — линейное отображение. Мы уже определили ядро и образ g; дополним это определение, положив

$$\text{кообраз } g\colon \operatorname{Coim} g = L/\operatorname{Ker} g,$$
$$\text{коядро } g\colon \operatorname{Coker} g = M/\operatorname{Im} g.$$

Имеется цепочка линейных отображений, «разбирающая g на части»,

$$\operatorname{Ker} g \xrightarrow{i} L \xrightarrow{c} \operatorname{Coim} g \xrightarrow{h} \operatorname{Im} g \xrightarrow{j} M \xrightarrow{f} \operatorname{Coker} g,$$

где все отображения, кроме h, — канонические вложения и факторизация, а h — единственное отображение, делающее коммутативной диаграмму

$$\begin{tikzcd} & L \arrow[dl, "c"'] \arrow[dr, "g"] & \\ \operatorname{Coim} g \arrow[rr, "h"] & & \operatorname{Im} g \end{tikzcd}$$

Оно определено однозначно, потому что $\operatorname{Ker} c = \operatorname{Ker} g$, и является изоморфизмом, потому что обратное отображение тоже существует и определено однозначно.

Смысл объединения этих пространств в пары (с приставкой «ко» и без нее) объясняется в теории двойственности (см. следующий параграф и упражнение 1 к нему).

10. Конечномерная альтернатива Фредгольма. Пусть $g\colon L\to M$ — линейное отображение. Число

$$\operatorname{ind} g = \dim \operatorname{Coker} g - \dim \operatorname{Ker} g$$

называется *индексом* оператора g. Из предыдущего пункта следует, что *если L и M конечномерны, то индекс g зависит только от L и M*:

$$\operatorname{ind} g = (\dim M - \dim \operatorname{Im} g) - (\dim L - \dim \operatorname{Im} g) = \dim M - \dim L.$$

В частности, если $\dim M = \dim L$, например, если g — линейный оператор на L, то $\operatorname{ind} g = 0$ для любого g. Отсюда вытекает так называемая альтернатива Фредгольма:

либо уравнение $g(x) = y$ разрешимо для всех y, и тогда уравнение $g(x) = 0$ имеет лишь нулевые решения;

либо это уравнение разрешимо не для всех y, и тогда однородное уравнение $g(x) = 0$ имеет ненулевые решения.

Точнее, *если* $\operatorname{ind} g = 0$, *то размерность пространства решений однородного уравнения равна коразмерности пространства правых частей, при которых разрешимо неоднородное уравнение.*

УПРАЖНЕНИЯ

1. Пусть M, $N \subset L$. Доказать, что следующее отображение является линейным изоморфизмом:

$$(M+N)/N \to M/M \cap N: \; m+n+N \mapsto m+M \cap N.$$

2. Пусть $L = M \oplus N$. Тогда каноническое отображение

$$M \to L/N: \; m \mapsto m+N$$

является изоморфизмом.

§ 7. Двойственность

1. В § 1 мы поставили в соответствие каждому линейному пространству L двойственное к нему пространство $L^* = \mathscr{L}(L, \mathscr{K})$, а в § 3 показали, что если $\dim L < \infty$, то $\dim L^* = \dim L$, и построили канонический изоморфизм $\varepsilon_L: L \to L^{**}$. Здесь мы продолжим описание двойственности, включив в рассмотрение линейные отображения, подпространства и факторпространства.

Теория двойственности получила свое название благодаря тому, что она выявляет ряд свойств «двусторонней симметрии» линейных пространств, довольно трудных для наглядного воображения, но совершенно фундаментальных. Достаточно сказать, что дуализм «волна — частица» в квантовой механике адекватно выражается именно на языке линейной двойственности бесконечномерных линейных пространств (точнее, соединения линейной и групповой двойственности в технике анализа Фурье).

Удобно следить за этой симметрией, несколько изменив обозначения, принятые в § 1 и 3.

2. Симметрия между L и L^*. Пусть $l \in L$, $f \in L^*$. Вместо $f(l)$ мы будем писать (f, l) (в знак аналогии со скалярным произведением — но векторов из разных пространств!). Таким образом, мы определили отображение $L^* \times L \to \mathscr{K}$. Оно линейно по каждому из двух аргументов f, l при фиксированном втором:

$$(f_1+f_2,\ l) = (f_1,\ l) + (f_2,\ l), \quad (af_1,\ l) = a(f_1,\ l),$$
$$(f,\ l_1+l_2) = (f,\ l_1) + (f,\ l_2), \quad (f,\ al_1) = a(f,\ l_1).$$

Вообще, отображения $L \times M \to \mathscr{K}$ с таким свойством называются *билинейными*, а также *спариваниями* между пространствами L и M. Введенное выше спаривание между L и L^* называется *каноническим* (ср. обсуждение этого слова в § 3, п. 8).

Отображение $\varepsilon_L: L \to L^{**}$ из § 3, п. 10, как видно из его определения, можно задать условием:

$$(\varepsilon_L(l),\ f) = (f,\ l),$$

где слева стоит символ спаривания между L^{**} и L^*, а справа — между L^* и L. Если $\dim L < \infty$, так что ε_L является изоморфизмом, и мы условимся отождествлять L^{**} и L посредством ε_L, эта формула приобретает симметричный вид $(l, f) = (f, l)$. Иными словами, мы можем также рассматривать *L как пространство, двойственное к L^**.

3. Симметрия между двойственными базисами. Пусть $\{e_1, \ldots, e_n\}$ — базис в L, $\{e^1, \ldots, e^n\}$ — двойственный базис в L^*. Согласно п. 9 § 3 он определяется формулами

$$(e^i, e_k) = \delta_{ik} = \begin{cases} 0 & \text{при } i \neq k, \\ 1 & \text{при } i = k. \end{cases}$$

Симметрия $(e^i, e_k) = (e_k, e^i)$ в соглашениях предыдущего пункта означает, что базис (e_k) двойствен к базису (e^i), если L рассматривать как пространство линейных функционалов на L^*. Таким образом, (e^i) и (e_k) образуют *двойственную пару базисов*, и это отношение симметрично.

Представим вектор $l^* \in L^*$ в виде линейной комбинации $\sum\limits_{i=1}^{n} b_i e^i$, а вектор $l \in L$ в виде $\sum\limits_{j=1}^{n} a_j e_j$. Тогда

$$(l^*, l) = \sum_{i, j=1}^{n} a_j b_i (e^i, e_j) = \sum_{i=1}^{n} a_i b_i = (a_1, \ldots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} =$$

$$= \left(\vec{a}^t\right) \vec{b} = \left(\vec{b}^t\right) \vec{a} = (b_1, \ldots, b_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = (l, l^*),$$

где $\vec{a}$, $\vec{b}$ — вектор-столбцы соответствующих коэффициентов. Эта формула совершенно аналогична формуле для скалярного произведения векторов в евклидовом пространстве, однако связывает в этой ситуации векторы из разных пространств.

4. Двойственное, или сопряженное отображение. Пусть $f: L \to M$ — линейное отображение линейных пространств. Мы покажем сейчас, что существует единственное линейное отображение $f^*: M^* \to L^*$, которое удовлетворяет условию

$$(f^*(m^*), l) = (m^*, f(l))$$

для любых векторов $m^* \in M^*$, $l \in L$.

а) *Единственность* f^*. Пусть f_1^*, f_2^* — два таких отображения. Тогда $(f_1^*(m^*), l) = (m^*, f(l)) = (f_2^*(m^*), l)$ для всех $m^* \in M^*$, $l \in L$, откуда следует, что $((f_1^* - f_2^*)(m^*), l) = 0$. Фиксируем m^* и будем менять l. Тогда элемент $(f_1^* - f_2^*)(m^*) \in L^*$ как линейный функционал на L принимает только нулевые значения и, значит, равен нулю. Поэтому $f_1^* = f_2^*$.

б) *Существование* f^*. Фиксируем $m^* \in M$ и рассмотрим выражение $(m^*, f(l))$ как функцию на L. В силу линейности f и билинейности $(,)$ эта функция линейна. Значит, она принадлежит L^*. Обозначим ее через $f^*(m^*)$. Равенства

$$f^*(m_1^* + m_2^*) = f^*(m_1^*) + f^*(m_2^*), \quad f^*(am^*) = af^*(m^*)$$

следуют из линейности $(m^*, f(l))$ по m^*. Значит, f^* — линейное отображение.

Пусть в L, M выбраны некоторые базисы, а в L^*, M^* — двойственные базисы. Пусть f в этих базисах представлено матрицей A Мы утверждаем, что f^* в двойственных базисах представлено транспонированной матрицей A^t. В самом деле, пусть B — матрица f^*. Согласно определениям и п. 3 имеем, обозначив вектор-столбцы координат m^*, l через $\vec{a}$, $\vec{b}$,

$$(m^*, f(l)) = \vec{a}^t(A\vec{b}),$$

$$(f^*(m^*), l) = (B\vec{a})^t\,\vec{b} = (\vec{a}^t B^t)\,\vec{b}.$$

Из ассоциативности умножения матриц и единственности f^* следует, что $A = B^t$, т. е. $B = A^t$.

Основные свойства сопряженных отображений собраны в следующей теореме:

5. Теорема. а) $(f+g)^* = f^* + g^*$;

б) $(af)^* = af^*$; *здесь* $f, g\colon L \to M$ *и* $a \in \mathscr{K}$;

в) $(fg)^* = g^* f^*$; *здесь* $L \xrightarrow{g} M \xrightarrow{f} N$;

г) $\mathrm{id}^* = \mathrm{id}$, $0^* = 0$;

д) *если канонически отождествить* L^{**} *с* L *и* M^{**} *с* M, *то* $f^{**}\colon L^{**} \to M^{**}$ *отождествляется с* $f\colon L \to M$.

Доказательство. Если считать, что L и M конечномерны, то проще всего проверить все эти утверждения, представив f, g матрицами в двойственных базисах и воспользовавшись простыми свойствами операции транспонирования:

$$(aA + bB)^t = aA^t + bB^t,\ (AB)^t = B^t A^t,\ E^t = E,\ 0^t = 0,\ (A^t)^t = 0.$$

Инвариантную проверку мы оставляем читателю в качестве упражнения.

6. Двойственность между подпространствами в L и в L^*. Пусть $M \subset L$ — некоторое линейное подпространство. Обозначим через $M^\perp \subset L^*$ и будем называть *ортогональным дополнением к* M множество функционалов, обращающихся в нуль на M. Иными словами,

$$m^* \in M^\perp \Leftrightarrow (m^*, m) = 0 \text{ для всех } m \in M.$$

Легко видеть, что $M^\perp$ является линейным пространством. В следующих утверждениях собраны основные свойства этой конструкции (L предполагается конечномерным).

а) *Имеется канонический изоморфизм* $L^*/M^\perp \to M^*$. Строится он так: многообразию $l^* + M^\perp$ ставится в соответствие ограничение функционала l^* на M. От выбора l^* оно не зависит, ибо ограничения функционалов из $M^\perp$ на M нулевые. Линейность этого отображения очевидна. Оно сюръективно, ибо всякий линейный функционал на M продолжается до некоторого функционала на L.

В самом деле, пусть $\{e_1, \ldots, e_m\}$ — базис в M, $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ — его продолжение до базиса L. Функционал f на M, заданный значениями $f(e_1), \ldots, f(e_n)$, продолжается на L, например, если положить $f(e_{m+1}) = \ldots = f(e_n) = 0$.

Наконец, построенное отображение $L^*/M^\perp \to M^*$ инъективно. В самом деле, у него нулевое ядро: если ограничение l^* на M равно нулю, то $l^* \in M^\perp$ и $l^* + M^\perp = M^\perp$ — нулевой элемент из $L^*/M^\perp$.

б) $\dim M + \dim M^\perp = \dim L$. Действительно, это следует из предыдущего утверждения, следствия п. 6 § 6 и того, что $\dim L^* = \dim L$, $\dim M^* = \dim M$.

в) *При каноническом отождествлении L^{**} с L пространство $(M^\perp)^\perp$ совпадает с M.*

Действительно, так как $(m^*, m) = 0$ для всех m^* и данного $m \in M$, ясно, что $M \subset (M^\perp)^\perp$. Но, кроме того, по предыдущему свойству, примененному дважды,

$$\dim (M^\perp)^\perp = \dim L - \dim M^\perp = \dim M.$$

Значит, $M = (M^\perp)^\perp$.

г) $(M_1 + M_2)^\perp = M_1^\perp \cap M_2^\perp; \quad (M_1 \cap M_2)^\perp = M_1^\perp + M_2^\perp$.

Доказательство предоставляется читателю в качестве упражнений.

УПРАЖНЕНИЯ

1. Пусть с линейным отображением $g: L \to M$ конечномерных пространств связана цепочка отображений, построенных в п. 8 § 6. Построить канонические изоморфизмы

$$\operatorname{Ker} g^* \to \operatorname{Coker} g,\ \operatorname{Coim} g^* \to \operatorname{Im} g,\ \operatorname{Im} g^* \to \operatorname{Coim} g,\ \operatorname{Coker} g^* \to \operatorname{Ker} g.$$

2. Вывести отсюда «третью теорему Фредгольма»: для того чтобы уравнение $g(x) = y$ было разрешимо (по x при данном y), необходимо и достаточно, чтобы y был ортогонален к ядру сопряженного отображения $g^*: M^* \to L^*$.

3. Последовательность линейных пространств и отображений $L \xrightarrow{f} M \xrightarrow{g} N$ называется точной в члене M, если $\operatorname{Im} f = \operatorname{Ker} g$. Проверить следующие утверждения:

а) Последовательность $0 \longrightarrow L \xrightarrow{f} M$ точна в члене L тогда и только тогда, когда f — инъекция.

б) Последовательность $M \xrightarrow{g} N \longrightarrow 0$ точна в члене N тогда и только тогда, когда g — сюръекция.

в) Последовательность конечномерных пространств $0 \longrightarrow L \xrightarrow{f} M \xrightarrow{g} N \longrightarrow 0$ точна (во всех членах) тогда и только тогда, когда точна двойственная последовательность $0 \longrightarrow N^* \xrightarrow{g^*} M^* \xrightarrow{f^*} L^* \longrightarrow 0$.

4. Мы знаем, что если отображение $f: L \to M$ в некоторых базисах представлено матрицей A, то отображение f^* в двойственных базисах представлено матрицей A^t. Вывести отсюда, что ранг матрицы совпадает с рангом транспонированной матрицы, т. е. что максимальные числа линейно независимых строк и столбцов матрицы совпадают.

§ 8. Структура линейного отображения

1. В этом параграфе мы начнем изучать следующую задачу: возможно яснее геометрически представить себе устройство линейного отображения $f: L \to M$. Ответ совсем прост, когда L и M никак

не связаны друг с другом: он дается теоремой из п. 2 этого параграфа. Гораздо интереснее и многообразнее получается картина, когда $M = L$ (этот и следующий параграфы) и $M = L^*$ (следующая часть). На матричном языке речь идет о приведении матрицы f к возможно более простой форме с помощью подходящего, специально приспособленного к структуре f, базиса. В первом случае базисы в L и M можно выбирать независимо, во втором речь идет об одном базисе в L или о базисе в L и двойственном к нему базисе L^*: меньшая свобода выбора приводит к большему разнообразию ответов.

На языке § 5 нашу задачу можно переформулировать следующим образом. Построим внешнюю прямую сумму пространств $L \oplus M$ и поставим в соответствие отображению f его график Γ_f: множество всех векторов вида $(l, f(l)) \in L \oplus M$. Легко убедиться, что Γ_f есть линейное подпространство в $L \oplus M$. Нас интересуют инварианты расположения Γ_f в $L \oplus M$. Для случая, когда базисы в L и M можно выбирать независимо, ответ дается следующей теоремой.

2. Теорема. *Пусть f: $L \to M$ — линейное отображение конечномерных пространств. Справедливы следующие утверждения:*

а) *Существуют такие прямые разложения $L = L_0 \oplus L_1$, $M = M_1 \oplus M_2$, что* $\operatorname{Ker} f = L_0$ *и f индуцирует изоморфизм L_1 с M_1.*

б) *Существуют такие базисы в L и M, что матрица f в этих базисах имеет вид (a_{ij}), где $a_{ii} = 1$ для $1 \leqslant i \leqslant r$ и $a_{ij} = 0$ для остальных i, j.*

в) *Пусть A — некоторая матрица размера $m \times n$. Тогда существуют такие невырожденные квадратные матрицы B и C размеров $m \times m$ и $n \times n$ и такое число $r \leqslant \min(m, n)$, что матрица BAC имеет вид, описанный в предыдущем пункте. Число r определено однозначно и равно рангу A.*

Доказательство. а) Положим $L_0 = \operatorname{Ker} f$, а в качестве L_1 выберем прямое дополнение к L_0: это возможно в силу п. 10 § 5. Затем положим $M_1 = \operatorname{Im} f$, а в качестве M_2 выберем прямое дополнение к M_1. Нужно лишь проверить, что f определяет изоморфизм L_1 с M_1. Отображение f: $L_1 \to M_1$ инъективно, потому что ядро f, т. е. L_0, пересекается с L_1 лишь по нулю. Оно сюръективно, потому что если $l = l_0 + l_1 \in L$, $l_0 \in L_0$, $l_1 \in L_1$, то $f(l) = f(l_1)$.

б) Положим $r = \dim L_1 = \dim M_1$ и выберем в L базис $\{e_1, \ldots, e_r, e_{r+1}, \ldots, e_n\}$, где первые r векторов образуют базис L_1, а следующие — базис L_0. Далее, векторы $e'_i = f(e_i)$, $1 \leqslant i \leqslant r$, образуют базис в $M_1 = \operatorname{Im} f$. Дополним его до базиса M векторами $\{e'_{r+1}, \ldots, e'_m\}$. Очевидно,

$$f(e_1, \ldots, e_r; e_{r+1}, \ldots, e_n) = (e'_1, \ldots, e'_r; 0, \ldots, 0) =$$

$$= (e'_1, \ldots, e'_r; e'_{r+1}, \ldots, e'_m)\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right),$$

так что матрица f в этих базисах имеет требуемый вид.

в) Построим по матрице A линейное отображение f координатных пространств $\mathscr{K}^n \to \mathscr{K}^m$ с этой матрицей, затем применим к нему утверждение б) В новых базисах матрица f будет иметь требуемый вид и выражаться через A в виде BAC, где B, C — матрицы перехода: см. п. 8 § 4. Наконец, $\operatorname{rk} A = \operatorname{rk} BAC = \operatorname{rk} f = \dim \operatorname{Im} f$. Это завершает доказательство.

Теперь перейдем к изучению линейных операторов. Начнем с введения простейшего класса: *диагонализируемых* операторов.

Назовем в общем случае подпространство $L_0 \subset L$ *инвариантным* относительно оператора f, если $f(L_0) \subset L_0$.

3. Определение. *Линейный оператор f: $L \to L$ называется диагонализируемым, если выполнено любое из двух равносильных условий:*

а) *L разлагается в прямую сумму одномерных инвариантных подпространств;*

б) *существует базис L, в котором матрица оператора f диагональна.*

Равносильность этих условий проверяется без труда. Если в базисе (e_i) матрица оператора f диагональна, то $f(e_i) = \lambda_i e_i$, так что одномерные подпространства, натянутые на e_i, инвариантны и L разлагается в их прямую сумму. Наоборот, если $L = \bigoplus L_i$ — такое разложение и e_i — любой ненулевой вектор из L_i, то $\{e_i\}$ образуют базис в L.

Диагонализируемые операторы образуют простейший и во многих отношениях самый важный класс. Например, над полем комплексных чисел, как мы убедимся, любой оператор можно делать диагональным, как угодно мало изменив его матрицу, так что оператор «в общем положении» диагонализируем.

Чтобы понять, что может помешать оператору быть диагонализируемым, введем два определения и докажем одну теорему.

4. Определение. 1) *Одномерное подпространство $L_1 \subset L$ называется собственным для оператора f, если оно инвариантно, т. е. $f(L_1) \subset L_1$. Если L_1 — такое подпространство, то f действует на нем как умножение на скаляр $\lambda \in \mathscr{K}$. Этот скаляр называется собственным значением оператора f (на L_1).*

б) *Вектор $l \in L$ называется собственным для f, если линейная оболочка $\mathscr{K}l$ является собственным подпространством. Иными словами, $l \neq 0$ и $f(l) = \lambda l$ для подходящего $\lambda \in \mathscr{K}$.*

Согласно определению п. 3, диагонализируемые операторы f допускают разложение L в прямую сумму своих собственных подпространств. Выясним, когда у f имеется хотя бы одно собственное подпространство.

5. Определение. *Пусть L — конечномерное линейное пространство, f: $L \to L$ — линейный оператор, A — его матрица в каком-нибудь базисе. Обозначим через $P(t)$ и назовем характеристическим многочленом оператора f, а также матрицы A, многочлен* $\det(tE - A)$ *с коэффициентами в поле $\mathscr{K}$* ($\det$ — *определитель*).

6. Теорема. а) *Характеристический многочлен линейного оператора f не зависит от выбора базиса, в котором представлена его матрица.*

б) *Любое собственное значение оператора является корнем $P(t)$ и любой корень $P(t)$, лежащий в $\mathscr{K}$, является собственным значением для f, отвечающим некоторому (не обязательно единственному) собственному подпространству в L.*

Доказательство. а) Согласно п. 8 § 4 матрица оператора f в другом базисе имеет вид $B^{-1}AB$. Поэтому, пользуясь мультипликативностью определителя, находим

$$\det(tE-B^{-1}AB)=\det(B^{-1}(tE-A)B)=$$
$$=(\det B)^{-1}\det(tE-A)\det B=\det(tE-A).$$

Заметим, что $P(t)=t^n-\operatorname{Tr}f\cdot t^{n-1}+\ldots+(-1)^n\det f$ (обозначения из п. 9 § 4).

б) Пусть $\lambda\in\mathscr{K}$ — корень $P(t)$. Тогда отображение $\lambda\cdot\mathrm{id}-f$ представлено вырожденной матрицей и, значит, имеет нетривиальное ядро. Пусть $l\neq 0$ — элемент из ядра; тогда $f(l)=\lambda l$, так что λ есть собственное значение для f, а l — соответствующий собственный вектор. Наоборот, если $f(l)=\lambda l$, то l лежит в ядре $\lambda\cdot\mathrm{id}-f$, так что $\det(\lambda\cdot\mathrm{id}-f)=P(\lambda)=0$.

7. Теперь мы видим, что оператор f вообще не имеет собственных значений и тем более не диагонализируем, если его характеристический многочлен $P(t)$ не имеет корней в поле $\mathscr{K}$. Это вполне может случиться над алгебраически не замкнутыми полями такими, как $\mathbf{R}$ и конечные поля. Например, пусть $A=\begin{pmatrix} a & b\\ c & d\end{pmatrix}$ — матрица с вещественными элементами. Тогда

$$\det(tE-A)=t^2-(a+d)t+(ad-bc),$$

и если $(a+d)^2-4(ad-bc)=(a-d)^2+4bc<0$, то A недиагонализируема.

Таким образом, мы впервые столкнулись здесь со случаем, когда свойства линейных отображений существенно зависят от свойств поля.

Чтобы не принимать последние во внимание как можно дольше, в следующем параграфе до п. 9 мы будем предполагать, что поле $\mathscr{K}$ является алгебраически замкнутым. Читатель, не знакомый с другими алгебраически замкнутыми полями, кроме $\mathbf{C}$, может всюду считать, что $\mathscr{K}=\mathbf{C}$. Алгебраическая замкнутость $\mathscr{K}$ равносильна любому из двух условий: а) любой многочлен от одной переменной $P(t)$ с коэффициентами в $\mathscr{K}$ имеет корень $\lambda\in\mathscr{K}$; б) любой такой многочлен $P(t)$ может быть представлен в виде $a\prod_{i=1}^{n}(t-\lambda_i)^{r_i}$, где a, $\lambda_i\in\mathscr{K}$; $\lambda_i\neq\lambda_j$ при $i\neq j$; это представление однозначно, если $P(t)\not\equiv 0$. В этом случае число r_i называется кратностью корня λ_i многочлена $P(t)$. *Множество всех корней характе-*

ристического многочлена называется спектром оператора f. Если все кратности равны 1, говорят, что f имеет *простой спектр*.

Если поле $\mathscr{K}$ алгебраически замкнуто, то согласно теореме п. 6 любой линейный оператор $f: L\to L$ имеет собственное подпространство. Однако он все равно может оказаться недиагонализируемым, ибо сумма всех собственных подпространств может оказаться меньше L, тогда как у диагонализируемого оператора она всегда равна L. Прежде чем переходить к общему случаю, разберемся с комплексными 2×2-матрицами.

8. Пример. Пусть L — двумерное комплексное пространство с базисом, оператор $f: L\to L$ представлен в этом базисе матрицей $A=\begin{pmatrix} a & b \\ c & d\end{pmatrix}$. Характеристический многочлен для f равен $t^2-(a+d)t+ +(ad-bc)$, его корни суть $\lambda_{1,2}=-\frac{a+d}{2}\pm\sqrt{\frac{(a-d)^2}{4}+bc}$. Рассмотрим отдельно следующие случаи:

а) $\lambda_1\neq\lambda_2$. Пусть e_1 — собственный вектор для λ_1, e_2 — для λ_2. Они линейно независимы, потому что если $ae_1+be_2=0$, то

$$f(ae_1+be_2)=a\lambda_1e_1+b\lambda_2e_2=0,$$

откуда $\lambda_1(ae_1+be_2)-(a\lambda_1e_1+b\lambda_2e_2)=b(\lambda_1-\lambda_2)e_2=0$, т. е. $b=0$ и аналогично $a=0$. Следовательно, в базисе $\{e_1, e_2\}$ матрица f диагональна.

б) $\lambda_1=\lambda_2=\lambda$. Здесь оператор f диагонализируем, только если он умножает на λ все векторы из L: это значит, что $\begin{pmatrix} a & b \\ c & d\end{pmatrix}=\begin{pmatrix} \lambda & 0 \\ 0 & \lambda\end{pmatrix}$, т. е. $a=d=\lambda$, $b=c=0$. Если же эти условия не выполнены, а выполнено только более слабое условие $(a-d)^2+4bc=0$, гарантирующее, что $\lambda_1=\lambda_2$, то у оператора f может быть, с точностью до пропорциональности, только один собственный вектор и f заведомо не диагонализируем.

Пример такой матрицы: $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda\end{pmatrix}$. Эта матрица называется *жордановой клеткой размера* 2×2 (*или ранга* 2).

В § 9 мы покажем, что именно такие матрицы образуют «строительные блоки» для нормальной формы общего линейного оператора над алгебраически замкнутым полем. Дадим общее определение:

9. Определение. а) *Жардановой клеткой* $J_r(\lambda)$ *размера* $r\times r$ *с собственным значением* λ *называется матрица вида*

$$J_r(\lambda)=\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & \lambda\end{pmatrix}.$$

б) *Жордановой матрицей называется матрица, состоящая из диагональных блоков* $J_{r_i}(\lambda_i)$ *и нулей вне этих блоков*:

$$J=\left(\begin{array}{c|c|c} J_{r_1}(\lambda_1) & 0 & \dots \\ \hline 0 & J_{r_2}(\lambda_2) & \dots \\ \hline \dots & \dots & \dots \end{array}\right).$$

в) *Жордановым базисом для оператора* $f\colon L \to L$ *называется такой базис пространства* L, *в котором матрица оператора* f *является жордановой, или, как говорят, имеет жорданову нормальную форму.*

г) *Приведением квадратной матрицы* A *к жордановой нормальной форме называется решение уравнения в матрицах вида* $X^{-1}AX = J$, *где* X — *(неизвестная) невырожденная матрица, а* J — *(неизвестная) жорданова матрица.*

10. Пример. Пусть $L_n(\lambda)$ — линейное пространство комплексных функций вида $e^{\lambda x}f(x)$, где $\lambda \in C$, $f(x)$ пробегает многочлены степени $\leqslant n-1$. Поскольку $\frac{d}{dx}(e^{\lambda x}f(x)) = e^{\lambda x}(\lambda f(x) + f'(x))$, дифференцирование $\frac{d}{dx}$ является линейным оператором на этом пространстве. Положим $e_{i+1} = \frac{x^i}{i!}e^{\lambda x}$ (напомним, что $0! = 1$), $i = 0, \ldots$ $\ldots, n-1$. Очевидно,

$$\frac{d}{dx}(e_{i+1}) = \frac{x^{i-1}}{(i-1)!}e^{\lambda x} + \lambda\frac{x^i}{i!}e^{\lambda x} = e_i + \lambda e_{i+1}$$

(первое слагаемое отсутствует при $i = 0$). Следовательно,

$$\frac{d}{dx}(e_1, \ldots, e_n) = (e_1, \ldots, e_n)\begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ . & . & . & . & . \\ . & . & . & . & \lambda \end{pmatrix}.$$

Таким образом, функции $\left(\frac{x^i}{i!}e^{\lambda x}\right)$ образуют жорданов базис для оператора $\frac{d}{dx}$ в нашем пространстве.

Этот пример показывает особую роль жордановых матриц в теории линейных дифференциальных уравнений (см. упражнения 1—3 к § 9).

11. Кроме уже рассмотренных геометрических соображений для нужд следующего параграфа нам понадобятся алгебраические сведения о полиномиальных функциях от оператора. Пусть $f\colon L \to L$ — фиксированный оператор.

а) Для любого многочлена $\sum_{i=0}^{n} a_i t^i = Q(t)$ с коэффициентами из поля $\mathscr{K}$ выражение $\sum_{i=0}^{n} a_i f^i$ имеет смысл в кольце $\mathscr{L}(L, L)$ эндоморфизмов пространства L; мы будем обозначать его $Q(f)$.

б) Будем говорить, что многочлен $Q(t)$ *аннулирует* оператор f, если $Q(f) = 0$. Ненулевые многочлены, аннулирующие f, существуют всегда, если L конечномерно. В самом деле, если $\dim L = n$, то $\dim \mathscr{L}(L, L) = n^2$ и операторы id, $f, \ldots, f^{n^2}$ линейно зависимы над $\mathscr{K}$. Это рассуждение показывает, что имеется аннулирующий f многочлен степени $\leqslant n^2$. На самом деле теорема Гамильтона — Кэли, которую мы докажем ниже, устанавливает существование аннулирующего многочлена степени n.

в) Рассмотрим многочлен $M(t)$ со старшим коэффициентом единица, аннулирующий f и имеющий наименьшую возможную степень. Он называется *минимальным многочленом* оператора f. Очевидно, он определен однозначно: если $M_1(t)$, $M_2(t)$ — два таких многочлена, то $M_1(t) - M_2(t)$ аннулирует f и имеет строго меньшую степень, так что $M_1(t) - M_2(t) = 0$.

г) Покажем, что любой многочлен, аннулирующий f, делится на минимальный многочлен f. Действительно, пусть $Q(f) = 0$. Разделим Q с остатком на M: $Q(t) = X(t)M(t) + R(t)$, $\deg R(t) < \deg M(t)$. Тогда $R(f) = Q(f) - X(f)M(f) = 0$, так что $R = 0$.

12. Теорема Гамильтона — Кэли. *Характеристический многочлен $P(t)$ оператора f аннулирует этот оператор.*

Доказательство. Мы будем пользоваться этой теоремой и докажем ее только для случая алгебраически замкнутого поля $\mathcal{K}$, хотя она верна и без этого ограничения.

Проведем индукцию по $\dim L$. Если L одномерно, то f есть умножение на скаляр λ, $P(t) = t - \lambda$ и $P(f) = 0$.

Пусть $\dim L = n \geqslant 2$ и теорема доказана для пространств размерности $n-1$. Выберем собственное значение λ оператора f и одномерное собственное подпространство $L_1 \subset L$, отвечающее λ. Пусть $\{e_1\}$ — базис L_1; дополним его до базиса $\{e_1, \ldots, e_n\}$ пространства L. Матрица оператора f в этом базисе имеет вид

$$\left(\begin{array}{c|c} \lambda & * \ldots * \\ \hline 0 & A \end{array}\right).$$

Поэтому $P(t) = (t-\lambda)\det(tE - A)$. Оператор f определяет линейное отображение $\bar{f}$: $L/L_1 \to L/L_1$, $\bar{f}(l + L_1) = f(l) + L_1$. Векторы $\bar{e}_i = e_i + L_1 \in L/L_1$, $i \geqslant 2$, образуют базис L/L_1, и матрица оператора $\bar{f}$ в этом базисе равна A. Поэтому $\bar{P}(t) = \det(tE - A)$ есть характеристический многочлен оператора $\bar{f}$, и по индуктивному предположению $\bar{P}(\bar{f}) = 0$. Значит, $\bar{P}(f)l \in L_1$ для любого вектора $l \in L$. Следовательно,

$$P(f)l = (f - \lambda)\bar{P}(f)l = 0,$$

ибо $f - \lambda$ переводит в нуль любой вектор из L_1. Это завершает доказательство.

13. Примеры. а) Пусть $f = \mathrm{id}_L$, $\dim L = n$. Тогда характеристический многочлен f равен $(t-1)^n$, а минимальный многочлен равен $t - 1$, так что они не совпадают при $n > 1$.

б) Пусть f представлен жордановой клеткой $J_r(\lambda)$. Характеристический многочлен оператора f равен $(t-\lambda)^r$. Чтобы вычислить минимальный многочлен, заметим, что $J_r(\lambda) = \lambda E_r + J_r(0)$.

Далее,

$$J_k(0)^k = \begin{pmatrix} 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & & 0 & 1 & \ldots & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix},$$

где единицы стоят на k-й диагонали выше главной; $J_r(0)^k = 0$ при

$k \geqslant r$. С другой стороны,

$$(J_r(\lambda) - \lambda E_r)^k = J_r(0)^k$$

при $0 \leqslant k \leqslant r-1$, а поскольку минимальный многочлен — делитель характеристического, это доказывает, что они совпадают.

УПРАЖНЕНИЯ

1. Пусть f: $L \to L$ — диагонализируемый оператор с простым спектром.

а) Доказать, что любой оператор g: $L \to L$ такой, что $gf = fg$, может быть представлен в виде многочлена от f.

б) Доказать, что размерность пространства таких операторов g равна $\dim L$.

Верны ли эти утверждения, если спектр оператора f не прост?

2. Пусть f, g: $L \to L$ — линейные операторы в пространстве размерности n над полем нулевой характеристики. Предположим, что $f^n = 0$, $\dim \operatorname{Ker} f = 1$, $gf - fg = f$. Доказать, что собственные значения g имеют вид $a, a-1, a-2, \ldots$ $\ldots, a-(n-1)$ для некоторого $a \in \mathcal{K}$.

§ 9. Жорданова нормальная форма

Основная цель этого параграфа — доказательство следующей теоремы о существовании и единственности жордановой нормальной формы для матриц и линейных операторов.

1. Теорема. *Пусть $\mathcal{K}$ — алгебраически замкнутое поле, L — конечномерное линейное пространство над $\mathcal{K}$, f: $L \to L$ — линейный оператор. Тогда:*

а) *Для оператора f существует жорданов базис, т. е. его матрица A в некотором базисе может быть приведена заменой базиса X к жордановой форме: $X^{-1}AX = J$.*

б) *Матрица J определена однозначно с точностью до перестановки входящих в нее жордановых клеток.*

2. Доказательство теоремы разбивается на ряд промежуточных шагов. Мы начнем с конструкции прямого разложения $L = \bigoplus\limits_{i=1}^{n} L_i$, где L_i — инвариантные подпространства для f, которые впоследствии будут отвечать *набору жордановых клеток для f с одним и тем же числом λ на диагонали.* Чтобы инвариантно охарактеризовать эти подпространства, вспомним, что $(J_r(\lambda) - \lambda E_r)^n = 0$. Оператор, некоторая степень которого равна нулю, принято называть *нильпотентным.* Итак, на подпространстве, отвечающем клетке $J_r(\lambda)$, оператор $f - \lambda$ нильпотентен; то же верно для его ограничения на сумму подпространств с фиксированным λ. Это мотивирует следующее определение.

3. Определение. *Вектор $l \in L$ называется корневым вектором оператора f, отвечающим $\lambda \in \mathcal{K}$, если существует такое r, что* $(f - \lambda)^r l = 0$ (здесь $f - \lambda$ обозначает оператор $f - \lambda\,\mathrm{id}$).

Очевидно, все собственные векторы корневые.

4. Предложение. *Обозначим через $L(\lambda)$ множество корневых векторов оператора f в L, отвечающих λ. Тогда $L(\lambda)$ — линейное подпространство в L и $L(\lambda) \neq \{0\}$ тогда и только тогда, когда λ — собственное значение для f.*

Доказательство. Допустим, $(f-\lambda)^{r_1}l_1=(f-\lambda)^{r_2}l_2=0$. Полагая $r=\max(r_1, r_2)$, находим $(f-\lambda)^r(l_1+l_2)=0$ и $(f-\lambda)^{r_1}(al_1)=0$. Следовательно, $L(\lambda)$ является линейным подпространством.

Если λ — собственное значение для f, то имеется собственный вектор, отвечающий λ, так что $L(\lambda)\neq\{0\}$. Наоборот, пусть $l\in L(\lambda)$, $l\neq 0$. Выберем *наименьшее* значение r, для которого $(f-\lambda)^r l=0$. Очевидно, $r\geqslant 1$. Вектор $l'=(f-\lambda)^{r-1}l$ является собственным для f с собственным значением λ: $l'\neq 0$ по выбору r и $(f-\lambda)l'=0$, откуда $f(l')=\lambda l'$.

5. Предложение. $L=\bigoplus L(\lambda_i)$, *где λ_i пробегает все собственные значения оператора f, т. е. различные корни характеристического многочлена f.*

Доказательство. Пусть $P(t)\doteq\prod_{i=1}^{s}(t-\lambda_i)^{r_i}$ — характеристический многочлен f, $\lambda_i\neq\lambda_j$ при $i\neq j$. Положим $F_i(t)=P(t)(t-\lambda_i)^{-r_i}$, $f_i=F_i(f)$, $L_i=\operatorname{Im} f_i$. Проверим следующую серию утверждений.

а) $(f-\lambda_i)^{r_i}L_i=\{0\}$, т. е. $L_i\subset L(\lambda_i)$. Действительно, $(f-\lambda_i)^{r_i}f_i=(f-\lambda_i)^{r_i}F_i(f)=P(f)=0$ по теореме Гамильтона — Кэли.

б) $L=L_1+\ldots+L_s$. Действительно, так как многочлены $F_i(t)$ в совокупности взаимно простые, существуют такие многочлены $X_i(t)$, что $\sum_{i=1}^{s}F_i(t)X_i(t)=1$. Поэтому, подставляя вместо t оператор f, имеем

$$\sum_{i=1}^{s}F_i(f)X_i(f)=\mathrm{id}.$$

Применяя это тождество к любому вектору $l\in L$, находим

$$l=\sum_{i=1}^{s}f_i(X_i(f)l)\in\sum_{i=1}^{s}L_i.$$

в) $L=L_1\oplus\ldots\oplus L_s$. Действительно, выберем $1\leqslant i\leqslant s$ и проверим, что $L_i\cap\left(\sum_{j\neq i}L_j\right)=\{0\}$. Пусть l — вектор из этого пересечения. Тогда

$$(f-\lambda_i)^{r_i}l=0,\ \text{ибо}\ l\in L_i;$$
$$F_i(f)l=\prod_{j\neq i}(f-\lambda_j)^{r_j}l=0,\ \text{ибо}\ l\in\sum_{j\neq i}L_j.$$

Так как $(t-\lambda_i)^{r_i}$ и $F_i(t)$ — взаимно простые многочлены, существуют такие многочлены $X(t)$ и $Y(t)$, что $X(t)(t-\lambda_i)^{r_i}+Y(t)\times F_i(t)=1$. Подставляя сюда f вместо t и применяя полученное операторное тождество к l, находим $X(f)(0)+Y(f)(0)=l=0$.

г) $L_i=L(\lambda_i)$. В самом деле, мы уже проверили, что $L_i\subset L(\lambda_i)$. Для доказательства обратного включения выберем вектор $l\in$

$\in L(\lambda_i)$ и представим его в виде $l = l' + l''$, $l' \in L_i$, $l'' \in \bigoplus\limits_{j \neq i} L_j$. Существует такое число r', что $(f - \lambda_i)^{r'} l'' = 0$, поскольку $l'' = l - l' \in L(\lambda_i)$. Кроме того, $F_i(f) l'' = 0$. Написав тождество $X(t)(t - \lambda_i)^{r'} + Y(t) F_i(t) = 1$, подставив в него f вместо t и применив к l'', получим $l'' = 0$, так что $l = l' \in L_i$.

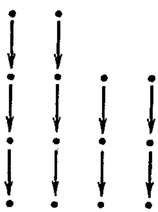
6. Следствие. *Если оператор f имеет простой спектр, то он диагонализируем.*

Доказательство. В самом деле, число разных собственных значений f тогда равно $n = \deg P(t) = \dim L$. Поэтому в разложении $L = \bigoplus\limits_{i=1}^{n} L(\lambda_i)$ все пространства $L(\lambda_i)$ одномерны, а так как каждое из них содержит собственный вектор, в базисе из этих векторов матрица оператора f становится диагональной.

Теперь мы *фиксируем* одно собственное значение λ и докажем, что ограничение f на $L(\lambda)$ обладает жордановым базисом, отвечающим этому λ. Чтобы не вводить новых обозначений, *мы будем до конца п. 7 считать, что f имеет единственное собственное значение λ и $L = L(\lambda)$*. Более того, мы можем считать даже, что $\lambda = 0$, потому что любой жорданов базис для оператора f является одновременно жордановым базисом для оператора $f + \mu$, где μ — любая константа. Тогда оператор f нильпотентен по теореме Гамильтона — Кэли: $P(t) = t^n$, $f^n = 0$, и мы докажем следующий факт:

7. Предложение. *Нильпотентный оператор f на конечномерном пространстве L имеет жорданов базис; матрица оператора f в этом базисе является объединением клеток вида $J_r(0)$.*

Доказательство. Если у нас уже есть жорданов базис в пространстве L, удобно поставить ему в соответствие диаграмму



D, подобную изображенной здесь. В этой диаграмме точки изображают элементы базиса, а стрелки описывают действие f (в общем случае действие $f - \lambda$). Элементы нижней строки оператор f переводит в нуль, т. е. в ней стоят собственные векторы оператора f, входящие в базис. Каждый столбец, таким образом, изображает базис инвариантного подпространства, отвечающего одной жордановой клетке, размер которой равен высоте этого столбца (числу точек в нем): если

$$f(e_h) = e_{h-1},\ f(e_{h-1}) = e_{h-2},\ \ldots,\ f(e_1) = 0,$$

то

$$f(e_1, \ldots, e_h) = (e_1, \ldots, e_h)\begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

Наоборот, если мы найдем в L базис, элементы которого f переводит в другие его элементы или в нуль так, что элементы этого базиса вместе с действием f можно изобразить подобной диаграммой, то он будет жордановым базисом для L.

Проведем доказательство существования индукцией по размерности L. Если $\dim L = 1$, то нильпотентный оператор f является нулевым, и любой ненулевой вектор в L образует его жорданов базис. Пусть теперь $\dim L = n > 1$, и пусть для размерностей, меньших n, существование жорданова базиса уже доказано. Обозначим через $L_0 \subset L$ подпространство собственных векторов для f, т. е. $\operatorname{Ker} f$. Так как $\dim L_0 > 0$, имеем $\dim L/L_0 < n$, а оператор $f\colon L \to L$ индуцирует оператор $\bar{f}\colon L/L_0 \to L/L_0$: $\bar{f}(l + L_0) = f(l) + L_0$. (Корректность определения оператора $\bar{f}$ и его линейность проверяются немедленно.)

По индуктивному предположению $\bar{f}$ имеет жорданов базис. Мы можем считать его непустым: иначе $L = L_0$, и любой базис L_0 будет жордановым для $\bar{f}$. Построим диаграмму $\bar{D}$ для элементов жорданова базиса оператора $\bar{f}$, в каждом ее столбце возьмем самый верхний вектор $\bar{e}_i$, $i = 1, \ldots, m$, и положим $\bar{e}_i = e_i + L_0$, $e_i \in L$. Теперь построим диаграмму D из векторов пространства L следующим образом. Для $i = 1, \ldots, m$ столбец с номером i диаграммы D будет состоять (сверху вниз) из векторов e_i, $f(e_i)$, ..., $f^{h_i - 1}(e_i)$, $f^{h_i}(e_i)$, где h_i — высота i-го столбца в диаграмме $\bar{D}$. Так как $\bar{f}^{h_i}(\bar{e}_i) = 0$, то $f^{h_i}(e_i) \in L_0$ и $f^{h_i+1}(e_i) = 0$. Выберем базис линейной оболочки векторов $f^{h_1}(e_1), \ldots, f^{h_m}(e_m)$ в L_0, дополним его до базиса L_0 и поставим дополняющие векторы в качестве дополнительных столбцов (высоты единица) в нижней строке диаграммы D; f переводит их в нуль.

Таким образом, диаграмма D из векторов пространства L вместе с действием оператора f на ее элементы имеет в точности такой вид, как требуется для жорданова базиса. Нужно только проверить, что элементы D действительно образуют базис L.

Сначала покажем, что линейная оболочка векторов из D равна L. Пусть $l \in L$, $\bar{l} = l + L_0$. По предположению $\bar{l} = \sum_{i=1}^{m} \sum_{j=0}^{h_i - 1} a_{ij} \bar{f}^j(\bar{e}_i)$. Так как L_0 f-инвариантно, отсюда следует, что

$$l - \sum_{i=1}^{m} \sum_{j=0}^{h_i - 1} \overline{a}_{ij} f^j(e_i) \in L_0.$$

Но все векторы $f^j(e_i)$, $j \leqslant h_i - 1$, лежат в строках диаграммы D, начиная со второй снизу, а подпространство L_0 порождено элементами первой строки D по построению. Поэтому l можно представить в виде линейной комбинации элементов D.

Остается проверить линейную независимость элементов D. Прежде всего, элементы нижней строки D линейно независимы. Действительно, если некоторая их нетривиальная линейная комбинация равна нулю, то она должна иметь вид $\sum_{i=1}^{m} a_i f^{h_i}(e_i) = 0$, ибо остальные элементы нижней строки дополняют базис линейной оболочки $\{f^{h_1}(e_1), \ldots, f^{h_m}(e_m)\}$ до базиса L_0. Но все $h_i \geqslant 1$, поэтому

$$f\left(\sum_{i=1}^{m} a_i f^{h_i-1}(e_i)\right) = 0,$$

так что

$$\sum_{i=1}^{m} a_i f^{h_i-1}(e_i) \in L_0 \quad \text{и} \quad \sum_{i=1}^{m} a_i \bar{f}^{h_i-1}(\bar{e}_i) = 0.$$

Из последнего же соотношения следует, что все $a_i = 0$, потому что векторы $\bar{f}^{h_i-1}(\bar{e}_i)$ составляют нижнюю строку диаграммы $\bar{D}$ и являются частью базиса L/L_0.

Наконец, покажем, что если имеется любая нетривиальная линейная комбинация векторов D, равная нулю, то из нее можно получить нетривиальную линейную зависимость между векторами нижней строки D. В самом деле, отметим самую верхнюю строку D, в которой имеются ненулевые коэффициенты этой воображаемой линейной комбинации. Пусть номер этой строки (считая снизу) равен h. Применим к этой комбинации оператор f^{h-1}. Очевидно, ее часть, отвечающая h-й строке, перейдет в нетривиальную линейную комбинацию элементов нижней строки, а остальные слагаемые обратятся в нуль. Это завершает доказательство предложения.

Теперь нам осталось проверить часть теоремы из п. 1, относящуюся к единственности.

8. Пусть фиксирован произвольный жорданов базис оператора f. Любой диагональный элемент матрицы оператора f в этом базисе, очевидно, является одним из собственных значений λ этого оператора. Рассмотрим часть базиса, отвечающего всем блокам матрицы с этим значением λ, и обозначим через L_λ его линейную оболочку. Поскольку $(J_r(\lambda) - \lambda)^r = 0$, имеем $L_\lambda \subset L(\lambda)$, где $L(\lambda)$ — корневое подпространство L. Кроме того, $L = \bigoplus L_{\lambda_i}$ по определению жорданова базиса и $L = \bigoplus L(\lambda_i)$ по предложению п. 5, где в обоих случаях λ_i пробегает все собственные значения оператора f по одному разу. Следовательно, $\dim L_{\lambda_i} = \dim L(\lambda_i)$ и $L_{\lambda_i} = L(\lambda_i)$. Значит, сумма размеров жордановых клеток, отвечающих каждому λ_i, не зависит от выбора жорданова базиса, и, более того, от выбора базиса не зависят линейные оболочки L_{λ_i}. Поэтому достаточно проверить теорему единственности для случая $L = L(\lambda)$ или даже для $L = L(0)$.

Построим диаграмму D, отвечающую данному жорданову базису $L = L(0)$. Размеры жордановых клеток — это высоты ее столбцов; если, как на чертеже, расположить столбцы в порядке убывания, то эти высоты однозначно определятся, если известны длины строк в диаграмме, начиная с нижней, в порядке убывания. Покажем, что длина нижней строки равна размерности $L_0 = \operatorname{Ker} f$. Действительно, возьмем любой собственный вектор l для f и представим его в виде линейной комбинации элементов D. В эту линейную комбинацию все векторы, находящиеся выше нижней строки, войдут с нулевыми коэффициентами. Действительно, если бы самые высокие векторы с ненулевыми коэффициентами лежали в строке с номером $h \geqslant 2$, то вектор $f^{h-1}(l) = 0$ был бы нетривиальной линейной комбинацией элементов нижней строки D (ср. конец доказательства предложения п. 7), а это противоречит линейной независимости элементов D. Значит, нижняя строка D образует базис L_0, так что ее длина равна $\dim L_0$, и потому эта длина одинакова для всех жордановых базисов. Точно так же длина второй строки не зависит от выбора базиса, так как она равна размерности $\operatorname{Ker} \bar{f}$ в L/L_0 в обозначениях предыдущего пункта. Это завершает доказательство единственности и теоремы п. 1.

9. Замечания. а) Пусть оператор f представлен матрицей A в некотором базисе, тогда задача приведения A к жордановой форме может быть решена с помощью следующих действий.

Вычислить характеристический многочлен A и его корни.

Вычислить размеры жордановых клеток, отвечающих корням λ. Для этого достаточно вычислить длины строк соответствующих диаграмм, т. е. $\dim \operatorname{Ker}(A - \lambda)$, $\dim \operatorname{Ker}(A - \lambda)^2 - \dim \operatorname{Ker}(A - \lambda)$, $\dim \operatorname{Ker}(A - \lambda)^3 - \dim \operatorname{Ker}(A - \lambda)^2, \ldots$

Построить жорданову форму J матрицы A и решить матричное уравнение $AX - XJ = 0$. Пространство решений этой линейной системы уравнений будет, вообще говоря, многомерно, и среди решений будут и вырожденные матрицы. Но по теореме существования обязательно есть невырожденные решения; можно взять любое из них.

б) Одно из важных приложений жордановой формы — вычисление функций от матриц (пока мы знакомы лишь с полиномиальными функциями). Пусть, скажем, нам нужно знать большую степень A^N матрицы A. Так как степень жордановой матрицы вычислить легко (см. § 8, п. 13), экономный способ может состоять в использовании формулы $A^N = XJ^NX^{-1}$, где $A = XJX^{-1}$: дело в том, что матрица X вычисляется раз навсегда и не зависит от N. Эту же формулу можно использовать для оценки роста элементов матрицы A^N.

в) В терминах жордановой формы легко вычислить минимальный многочлен матрицы A. В самом деле, ограничимся для простоты случаем поля нулевой характеристики. Тогда минимальный многочлен $J_r(\lambda)$ равен $(t - \lambda)^r$ (см. п. 13 § 8), минимальный многочлен блочной матрицы $(J_{r_i}(\lambda))$ равен $(t - \lambda)^{\max(r_i)}$, наконец, минимальный многочлен общей жордановой матрицы с диагональными элементами $\lambda_1, \ldots, \lambda_s$ ($\lambda_i \neq \lambda_j$ при $i \neq j$) равен $\prod_{j=1}^{s}(t-\lambda_j)^{r_j}$, где r_j — наибольший размер жордановой клетки, отвечающей λ_j.

10. Другие нормальные формы. В этом пункте мы вкратце опишем другие нормальные формы матриц, пригодные, в частности, для алгебраически незамкнутых полей.

а) *Циклические пространства и циклические клетки.* Пространство L называется *циклическим* относительно оператора f, если в L существует такой вектор l, также называемый *циклическим*, что векторы $l, f(l), \ldots, f^{n-1}(l)$ образуют базис L. Полагая $e_i = f^{n-i}(l)$, $i = 1, \ldots, n = \dim L$, имеем

$$f(e_1, \ldots, e_n) = (e_1, \ldots, e_n)\begin{pmatrix} a_{n-1} & 1 & 0 & \ldots & 0 \\ a_{n-2} & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_0 & 0 & 0 & \ldots & 0 \end{pmatrix},$$

где $a_i \in \mathcal{K}$ однозначно определяются из соотношения $f^n(l) = \sum_{i=0}^{n-1} a_i f^i(l)$. Матрица оператора f в таком базисе называется *циклической клеткой.* Наоборот, если матрица оператора f в базисе $(e_1, \ldots, e_n)$ является циклической клеткой, то вектор $l = e_n$ цикличен, и $e_i = f^{n-i}(e_n)$ (индукция вниз по i).

Покажем, что вид циклической клетки, отвечающей f, не зависит от выбора исходного циклического вектора. Для этого проверим, что первый столбец клетки состоит из коэффициентов минимального многочлена оператора f: $M(t) = t^n - \sum_{i=0}^{n-1} a_i t^i$.

В самом деле, $M(f) = 0$, потому что $M(f)[f^i(l)] = f^i[M(f)\,l] = 0$, а векторы $f^i(l)$ порождают L. С другой стороны, если $N(t)$ — многочлен степени $<n$, то $N(f) \neq 0$, потому что иначе, применив оператор $N(f) = 0$ к циклическому вектору l, мы получим нетривиальное линейное соотношение между векторами базиса $l, f(l), \ldots, f^{n-1}(l)$.

б) *Критерий цикличности пространства.* Согласно предыдущим рассмотрениям, если пространство L циклично относительно f, то его размерность n равна степени минимального многочлена оператора f и, стало быть, минимальный многочлен совпадает с характеристическим. Обратное тоже верно: если операторы $\mathrm{id}, f, \ldots, f^{n-1}$ линейно независимы, то существует такой вектор l, что векторы $l, f(l), \ldots, f^{n-1}(l)$ линейно независимы, так что L циклично. Мы не будем доказывать это утверждение.

в) *Матрица любого оператора в подходящем базисе может быть приведена к прямой сумме циклических клеток.* Доказательство можно провести аналогично доказательству теоремы о жордановой форме. Вместо множителей $(t-\lambda_i)^{r_i}$ характеристического многочлена следует рассматривать множители $p_i(t)^{r_i}$, где $p_i(t)$ — неприводимые над полем $\mathcal{K}$ делители характеристического многочлена. Теорема единственности также имеет место, если ограничиться случаем, когда минимальные многочлены всех циклических клеток неприводимы. Без этого ограничения она неверна: циклическое пространство может быть прямой суммой двух циклических подпространств, минимальные многочлены которых взаимно просты

УПРАЖНЕНИЯ

1. Пусть L — конечномерное пространство дифференцируемых функций комплексной переменной x, обладающее тем свойством, что если $f \in L$, то $\frac{df}{dx} \in L$. Доказать, что существуют такие комплексные числа $\lambda_1, \ldots, \lambda_s$ и целые числа $r_1, \ldots, r_s \geqslant 1$, что $L = \oplus L_i$, где L_i — пространство функций вида $e^{\lambda_i x} P_i(x)$ $P_i(x)$ — произвольный многочлен степени $\leqslant r_i - 1$. (*Указание.* Рассмотреть жорданов базис для оператора $\frac{d}{dx}$ на L и последовательно вычислить вид входящих в него функций, начиная с нижней строки его диаграммы.)

2. Пусть $y(x)$ — функция комплексной переменной x, удовлетворяющая дифференциальному уравнению вида

$$\frac{d^n y}{dx^n} + \sum_{i=0}^{n-1} a_i \frac{d^i y}{dx^i} = 0, \ a_i \in \mathbf{C}.$$

Обозначим через L линейное пространство функций, натянутое на $d^i y/dx^i$ для всех $i \geqslant 0$. Доказать, что оно конечномерно и оператор d/dx переводит его в себя.

3. Пользуясь результатами упражнений 1 и 2, вывести, что $y(x)$ представляется в виде $\sum e^{\lambda_i x} P_i(x)$, P_i — многочлены. Как связаны числа λ_i с видом дифференциального уравнения?

4. Пусть $J_r(\lambda)$ — жорданова клетка над $\mathbf{C}$. Доказать, что, как угодно мало изменив ее элементы, можно добиться того, что полученная матрица будет диагонализируемой. (*Указание.* Изменить элементы на диагонали, сделав их попарно разными.)

5. Перенести результат этого упражнения на произвольные матрицы над $\mathbf{C}$, воспользовавшись тем, что коэффициенты характеристического многочлена непрерывно зависят от элементов матрицы, а условие отсутствия кратных корней многочлена равносильно тому, что его дискриминант не обращается в нуль.

6. Придать точный смысл следующим утверждениям и доказать их:

а) общая 2×2-матрица над $\mathbf{C}$ диагонализируема.

б) общая 2×2-матрица с одинаковыми собственными значениями недиагонализируема.

§ 10. Нормированные линейные пространства

В этом параграфе изучаются специальные свойства линейных пространств над вещественными и комплексными числами, связанные с возможностью определить в них понятие предельного перехода и построить начала анализа. Особую роль эти свойства играют в бесконечномерном случае, так что по существу излагаемый материал является элементарным введением в функциональный анализ.

1. Определение. *Пара* (E, d), *где* E — *множество, а* d: $E \times E \to \mathbf{R}$ — *вещественнозначная функция, называется метрическим про-*

странством, если выполнены следующие условия для всех x, y, $z \in E$:

а) $d(x, y) = d(y, x)$ (*симметрия*);

б) $d(x, x) = 0$; $d(x, y) > 0$, *если* $x \neq y$ (*положительность*);

в) $d(x, z) \leqslant d(x, y) + d(y, z)$ (*неравенство треугольника*).

Функция d с такими свойствами называется метрикой, а $d(x, y)$ — расстоянием между точками x, y.

2. Примеры. а) $E = \mathbf{R}$ или $\mathbf{C}$, $d(x, y) = |x - y|$.

б) $E = \mathbf{R}^n$ или $\mathbf{C}^n$, $d(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2}$. Это так называемая естественная метрика. Во второй части мы рассмотрим ее систематически и изучим ее обобщения на произвольные основные поля в теории квадратичных форм. Другие метрики:

$$d_1(\vec{x}, \vec{y}) = \max(|x_i - y_i|), \quad d_2(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i|.$$

в) $E = C(a, b)$ — непрерывные функции на отрезке $[a, b]$. Вот три наиболее важные метрики:

$$d_1(f, g) = \max_{a \leqslant t \leqslant b} |f(t) - g(t)|,$$

$$d_2(f, g) = \int_a^b |f(t) - g(t)|\, dt,$$

$$d_3(f, g) = \left(\int_a^b |f(t) - g(t)|^2\, dt\right)^{1/2}.$$

(Проверьте аксиомы. Для d_2 в примере б) и d_3 в примере в) неравенство треугольника будет доказано в следующей части.)

г) E — любое множество, $d(x, y) = 1$ при $x \neq y$. Это — одна из *дискретных* метрик на E.

(С каждой метрикой связана некоторая топология на E, и последняя описанная метрика индуцирует дискретную топологию.)

3. Шары, ограниченность и полнота. В метрическом пространстве E с метрикой d множества

$$B(x_0, r) = \{x \in E \mid d(x_0, x) < r\},$$

$$\bar{B}(x_0, r) = \{x \in E \mid d(x_0, x) \leqslant r\},$$

$$S(x_0, r) = \{x \in E \mid d(x_0, x) = r\}$$

называются соответственно *открытым шаром, замкнутым шаром* и *сферой* с центром в точке x_0 и радиуса r. Не следует связывать с ними интуитивные представления, слишком близкие к трехмерным пространственным. Например, в примере г) п. 2 все сферы радиуса $r \neq 1$ пусты.

Подмножество $F \subset E$ называется *ограниченным*, если оно содержится в некотором шаре (конечного радиуса).

Последовательность точек $x_1, x_2, \ldots, x_n, \ldots$ в E *сходится к точке* $a \in E$, если $\lim\limits_{n\to\infty} d(x_n, a)=0$. Последовательность называется *фундаментальной* (или *последовательностью Коши*), если для всякого $\varepsilon > 0$ существует $N = N(\varepsilon)$, такое, что $d(x_m, x_n) < \varepsilon$ при $m, n > N(\varepsilon)$.

Метрическое пространство E называется *полным*, если любая последовательность Коши в нем сходится. Из полноты $\mathbf{R}$ и $\mathbf{C}$, доказываемой в анализе, следует, что пространства $\mathbf{R}^n$ и $\mathbf{C}^n$ с любой из метрик d, d_1, d_2 примера б) п. 2 полны.

4. Нормированные линейные пространства. Пусть теперь L — линейное пространство над $\mathbf{R}$ или $\mathbf{C}$. Особо важную роль играют метрики на L, которые удовлетворяют двум условиям:

а) $d(l_1, l_2) = d(l_1 + l, l_2 + l)$ для любых $l, l_1, l_2 \in L$ (инвариантность относительно сдвига);

б) $d(al_1, al_2) = |a|d(l_1, l_2)$ (умножение на скаляр a увеличивает расстояния в $|a|$ раз).

Пусть d — такая метрика. Назовем *нормой* вектора l (относительно d) и будем обозначать через $\|l\|$ число $d(l, 0)$. Из аксиом метрики (п. 2) и условий а), б) вытекают следующие свойства нормы:

$$\|0\| = 0,\ \|l\| > 0,\ \text{если } l \neq 0;$$

$$\|al\| = |a|\|l\| \text{ для всех } a \in \mathcal{K},\ l \in L;$$

$$\|l_1 + l_2\| \leqslant \|l_1\| + \|l_2\| \text{ для всех } l_1,\ l_2 \in L.$$

Первые два свойства очевидны, третье проверяется так: $\|l_1 + l_2\| = d(l_1 + l_2, 0) = d(l_1, -l_2) \leqslant d(l_1, 0) + d(0, -l_2) = \|l_1\| + \|l_2\|$.

Линейное пространство L, снабженное функцией нормы $\|\ \|: L \to \mathbf{R}$, удовлетворяющей перечисленным трем условиям, называется *нормированным*.

Наоборот, по норме восстанавливается метрика: положив $d(l_1, l_2) = \|l_1 - l_2\|$, легко проверить аксиомы метрики. Для нее $d(l, 0) = \|l\|$.

Полное нормированное линейное пространство называется банаховым пространством. Пространства $\mathbf{R}^n$ и $\mathbf{C}^n$ с любыми нормами, отвечающими метрикам из п. 2, банаховы.

Общее понятие сходимости последовательности в метрическом пространстве, данное в п. 3, специализируется на случай нормированных линейных пространств и называется *сходимостью по норме*. Линейная структура позволяет определить понятие сходимости ряда, более сильное, чем сходимость по норме его частичных сумм. Именно, ряд $\sum\limits_{i=1}^{\infty} l_i$ называется *абсолютно сходящимся*, если сходится ряд $\sum\limits_{i=1}^{\infty} \|l_i\|$.

5. Норма и выпуклость. Нетрудно описать все нормы на одномерном пространстве L: *любые две из них отличаются друг от друга умножением на положительную константу.* В самом деле, пусть $l \in L$ — ненулевой вектор, $\| \; \|_1$, $\| \; \|_2$ — две нормы. Если $\|l\|_1 = c\|l\|_2$, то $\|al\|_1 = |a| \|l\|_1 = c|a| \|l\|_2 = c\|al\|_2$ для всех $a \in \mathscr{K}$.

Будем называть *кругами* (соответственно *окружностями*) в одномерном пространстве L шары (соответственно сферы) ненулевого радиуса с центром в нуле относительно любой из норм. Как следует из предыдущего рассуждения, *множества всех кругов и окружностей в L не зависят* от *выбора исходной нормы.* Вместо задания любой нормы можно указать ее единичный круг B или единичную окружность S: S восстанавливается по B как граница B, а B восстанавливается по S как множество точек вида $\{al \mid l \in S, |a| \leqslant 1\}$. Заметим, что при $\mathscr{K} = \mathbf{R}$ круги суть отрезки с центром в нуле, а окружности — пары точек, симметричные относительно нуля.

Чтобы перенести это описание на пространства любой размерности, нам понадобится понятие выпуклости. Подмножество $E \subset L$ называется *выпуклым*, если для любых двух векторов $l_1, l_2 \in E$ и для любого числа $0 \leqslant a \leqslant 1$ вектор $al_1 + (1-a)l_2$ лежит в E. Это согласуется с обычным определением выпуклости в $\mathbf{R}^2$ и $\mathbf{R}^3$: вместе с любыми двумя точками («концами векторов l_1 и l_2») множество E должно содержать весь соединяющий их отрезок («концы векторов $al_1 + (1-a)l_2$»).

Пусть $\| \; \|$ — некоторая норма на L. Положим $B = \{l \in L \mid \|l\| \leqslant 1\}$, $S = \{l \in L \mid \|l\| = 1\}$. Ограничение $\| \; \|$ на любое линейное подпространство $L_0 \subset L$ индуцирует норму на L_0. Отсюда следует, что для любого *одномерного* подпространства $L_0 \subset L$ множество $L_0 \cap B$ является кругом в L_0, а множество $L_0 \cap S$ — окружностью в смысле данного выше определения. Кроме того, из неравенства треугольника следует, что если $l_1, l_2 \in B$, $0 \leqslant a \leqslant 1$, то

$$\|al_1 + (1-a)l_2\| \leqslant a\|l_1\| + (1-a)\|l_2\| \leqslant 1,$$

т. е. $al_1 + (1-a)l_2 \in B$, так что B — выпуклое множество.

Справедлива и обратная теорема:

6. Теорема. *Пусть $S \subset L$ — множество, удовлетворяющее двум условиям:*

а) *Пересечение $S \cap L_0$ с любым одномерным подпространством L_0 является окружностью.*

б) *Множество $B = \{al \mid |a| \leqslant 1, l \in S\}$ выпукло.*

Тогда на L существует единственная норма $\| \; \|$, для которой B является единичным шаром, а S — единичной сферой.

Доказательство. Обозначим через $\| \; \|: L \to \mathbf{R}$ функцию, которая на каждом одномерном подпространстве L_0 является нормой с единичной сферой $S \cap L_0$. Ясно, что такая функция существует и единственна, и нуждается в проверке лишь неравенство треугольника для нее. Пусть $l_1, l_2 \in L$, $\|l_1\| = N_1$, $\|l_2\| = N_2$, $N_i \neq 0$. Применим условие выпуклости B к векторам $N_1^{-1}l_1$ и $N_2^{-1}l_2 \in S$.

Получим

$$\left\| \frac{N_1}{N_1+N_2} N_1^{-1} l_1 + \frac{N_2}{N_1+N_2} N_2^{-1} l_2 \right\| \leqslant 1,$$

откуда

$$\| l_1 + l_2 \| \leqslant N_1 + N_2 = \| l_1 \| + \| l_2 \|.$$

7. Теорема. *Любые две нормы $\| \ \|_1$ и $\| \ \|_2$ на конечномерном пространстве L эквивалентны в том смысле, что существуют положительные константы $0 < c \leqslant c'$ с условием*

$$c \| l \|_2 \leqslant \| l \|_1 \leqslant c' \| l \|_2$$

для всех $l \in L$. В частности, топологии, т. е. понятия сходимости, отвечающие любым двум нормам, совпадают, и все конечномерные нормированные пространства банаховы.

Доказательство. Выберем базис в L и рассмотрим естественную норму $\| \vec{x} \| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$ относительно координат в этом базисе. Достаточно проверить, что любая норма $\| \ \|_1$ эквивалентна этой. Ее ограничение на единичную сферу нормы $\| \ \|$ является непрерывной функцией координат $\vec{x}$, принимающей лишь положительные значения (непрерывность следует из неравенства треугольника). Следовательно, эта функция отграничена от нуля константой $c > 0$ и ограничена константой $c' > 0$ по теореме Больцано — Вейерштрасса (единичная сфера S для $\| \ \|$ замкнута и ограничена). Из неравенства $c \leqslant \|l\|_1 \leqslant c'$ для всех $l \in S$ следует неравенство $c\|l\| \leqslant \|l\|_1 \leqslant c'\|l\|$ для всех $l \in L$. Поскольку L полно в топологии, отвечающей норме $\| \ \|$, и понятия сходимости для эквивалентных норм совпадают, L полно в любой норме.

8. Норма линейного оператора. Пусть L, M — нормированные линейные пространства над одним и тем же полем $\mathbf{R}$ или $\mathbf{C}$.

Рассмотрим линейное отображение $f\colon L \to M$. Оно называется *ограниченным*, если существует такое вещественное число $N \geqslant 0$, что для всех $l \in L$ выполнено неравенство $\|f(l)\| \leqslant N\|l\|$ (левая норма — в M, правая — в L). Обозначим через $\mathscr{L}^1(L, M)$ множество ограниченных линейных операторов. Для каждого $f \in \mathscr{L}^1(L, M)$ обозначим через $\|f\|$ нижнюю грань всех N, для которых выполняются неравенства $\|f(l)\| \leqslant N\|l\|$, $l \in L$.

9. Теорема. а) *$\mathscr{L}^1(L, M)$ является нормированным линейным пространством относительно функции $\|f\|$, которая называется индуцированной нормой.*

б) *Если L конечномерно, то $\mathscr{L}^1(L, M) = \mathscr{L}(L, M)$, т. е. любое линейное отображение ограничено.*

Доказательство. а) Пусть $f, g \in \mathscr{L}^1(L, M)$. Если $\|f(l)\| \leqslant N_1\|l\|$ и $\|g(l)\| \leqslant N_2\|l\|$ для всех l, то

$$\|(f+g)(l)\| \leqslant (N_1 + N_2)\| l \|, \quad \|af(l)\| \leqslant |a| N_1 \| l \|.$$

Поэтому $f + g$ и af ограничены и, более того, переходя к нижним граням, имеем

$$\|f+g\| \leqslant \|f\|+\|g\|, \ \|af\|=|a|\|f\|.$$

Если $\|f\|=0$, то для любого $\varepsilon>0$ $\|f(l)\| \leqslant \varepsilon\|l\|$. Значит, $\|f(l)\|=0$, так что $f=0$.

в) На единичной сфере в L отображение $l \mapsto \|f(l)\|$ является непрерывной функцией. Так как эта сфера ограничена и замкнута, эта функция ограничена и, более того, верхняя грань ее значений достигается. Поэтому на сфере $\|f(l)\| \leqslant N$, так что $\|f(l)\| \leqslant N\|l\|$ для всех $l \in L$.

Попутно мы обнаружили, что $\|f\|=\max\{\|f(l)\|, \ l \in$ единичная сфера в $L\}$.

10. Примеры: а) В конечномерном пространстве L последовательность векторов $l_1, \ldots, l_n, \ldots$ сходится к вектору l тогда и только тогда, когда в некотором (и потому в любом) базисе последовательность i-х координат векторов l_j сходится к i-й координате вектора l, т. е. если $f(l_1), \ldots, f(l_n), \ldots$ сходится для любого линейного функционала $f \in L^*$. Последнее условие можно перенести на бесконечномерные пространства, потребовав сходимость $f(l_i)$ лишь для ограниченных функционалов f. Это приводит, вообще говоря, к *новой топологии на* L, называемой *слабой топологией.*

б) Пусть L — пространство вещественных дифференцируемых функций на $[0, 1]$ с нормой $\|f\|=\left(\int\limits_0^1 f(t)^2\,dt\right)^{1/2}$. Тогда оператор умножения на t ограничен, ибо $\int\limits_0^1 t^2 f(t)^2\,dt \leqslant \int\limits_0^1 f(t)^2\,dt$, а оператор $\dfrac{d}{dt}$ неограничен. В самом деле, для любого целого $n \geqslant 0$ функция $\sqrt{2n+1}\,t^n$ лежит на единичной сфере, а норма ее производной равна $n\sqrt{\dfrac{2n+1}{2n-1}} \to \infty$ при $n \to \infty$.

11. Теорема. *Пусть* $L \xrightarrow{f} M \xrightarrow{g} N$ — *ограниченные линейные отображения нормированных пространств. Тогда их композиция ограничена и*

$$\|g \circ f\| \leqslant \|g\|\|f\|.$$

Доказательство. Если $\|f(l)\| \leqslant N_1\|l\|$ и $\|g(m)\| \leqslant N_2\|m\|$ для всех $l \in L$, $m \in M$, то

$$\|g \circ f(l)\| \leqslant N_2\|f(l)\| \leqslant N_2 N_1\|l\|,$$

откуда, переходя к нижним граням, получаем требуемое утверждение.

УПРАЖНЕНИЯ

1. Вычислить нормы на $\mathbf{R}^2$, для которых единичными шарами являются множества:

а) $x^2+y^2\leqslant 1$;
б) $x^2+y^2\leqslant r^2$;
в) квадрат с вершинами $(\pm 1, \pm 1)$;
г) квадрат с вершинами $(0, \pm 1)$, $(\pm 1, 0)$.

2. Пусть $f(x)\geqslant 0$ — дважды дифференцируемая вещественная функция на $[a, b]\subset\mathbf{R}$ и $f''(x)\leqslant 0$. Доказать, что множество $\{(x, y)\,|\,a\leqslant x\leqslant b,\ 0\leqslant y\leqslant f(x)\}\subset\mathbf{R}^2$ выпуклое.

3. Пользуясь результатом упражнения 2, доказать, что множество $|x|^p+|y|^p\leqslant 1$ для $p>1$ в $\mathbf{R}^2$ является единичным шаром для некоторой нормы. Вычислив эту норму, доказать неравенство Минковского:

$$(|x_1+y_1|^p+|x_2+y_2|^p)^{1/p}\leqslant(|x_1|^p+|x_2|^p)^{1/p}+(|y_1|^p+|y_2|^p)^{1/p}.$$

4. Обобщить результаты упражнеиия 3 на случай $\mathbf{R}^n$.

5. Пусть B — единичный шар некоторой нормы в L, B^* — единичный шар индуцированной нормы в $L^*=\mathscr{L}(L, \mathscr{K})$, $\mathscr{K}=\mathbf{R}$ или $\mathbf{C}$. Дать явное описание B^* и вычислить B^* для норм из упражнений 1 и 3.

§ 11. Функции линейных операторов

1. В § 8 и 9 мы определили операторы $Q(f)$, где $f\colon L\to L$ — линейный оператор, а Q — любой многочлен с коэффициентами из основного поля $\mathscr{K}$. Если $\mathscr{K}=\mathbf{R}$ или $\mathbf{C}$, пространство L нормировано, а оператор f ограничен, то $Q(f)$ можно определить для более общего класса функций Q с помощью предельного перехода.

Мы ограничимся рассмотрением голоморфных функций Q, задаваемых степенными рядами с ненулевым радиусом сходимости: $Q(t)=\sum\limits_{i=0}^{\infty}a_it^i$. Положим $Q(f)=\sum\limits_{i=0}^{\infty}a_if^i$, если этот ряд из операторов абсолютно сходится, т. е. если сходится ряд $\sum\limits_{i=0}^{\infty}a_i\|f^i\|$. (В случае $\dim L<\infty$, которым мы в основном будем заниматься здесь, $\mathscr{L}^1(L, L)=\mathscr{L}(L, L)$, и пространство всех операторов конечномерно и банахово; см. § 10, утверждение б) теоремы п. 9.)

2. Примеры. а) Пусть f — нильпотентный оператор. Тогда $\|f^i\|=0$ для достаточно больших i, и ряд $Q(f)$ всегда абсолютно сходится. На самом деле он совпадает с одной из своих частичных сумм.

б) Пусть $\|f\|<1$. Ряд $\sum\limits_{i=0}^{\infty}f^i$ абсолютно сходится и

$$(\mathrm{id}-f)\sum_{i=0}^{\infty}f^i=\left(\sum_{i=0}^{\infty}f^i\right)(\mathrm{id}-f)=\mathrm{id}.$$

Действительно,

$$(\mathrm{id}-f)\sum_{i=0}^{N}f^i=\mathrm{id}-f^{N+1}=\left(\sum_{i=0}^{N}f^i\right)(\mathrm{id}-f)$$

и переход к пределу при $N \to \infty$ дает требуемое. В частности, если $\|f\| < 1$, то оператор $\mathrm{id} - f$ обратим.

в) Назовем экспонентой ограниченного оператора f оператор

$$e^f = \exp(f) = \sum_{n=0}^{\infty} \frac{1}{n!} f^n.$$

Так как $\|f^n\| \leqslant \|f\|^n$ (см. теорему п. 11 § 10) и числовой ряд для экспоненты равномерно сходится на любом ограниченном множестве, функция $\exp(f)$ определена для любого ограниченного оператора f и непрерывна по f.

Например, ряд Тейлора $\sum_{i=0}^{\infty} \frac{(\Delta t)^i}{i!} \varphi^{(i)}(t)$ для значения $\varphi(t + \Delta t)$ можно формально записать в виде $\exp\left(\Delta t \frac{d}{dt}\right)\varphi$. Чтобы эта запись приобрела точный смысл, нужно, конечно, выбрать пространство бесконечно дифференцируемых функций φ с нормой и проверить сходимость в индуцированной норме.

Частный случай: $\exp(a\,\mathrm{id}) = e^a\,\mathrm{id}$ (a — скаляр); $\exp(\mathrm{diag}(a_1, \ldots \ldots, a_n)) = \mathrm{diag}(\exp a_1, \ldots, \exp a_n)$.

Основное свойство числовой экспоненты: $e^a e^b = e^{a+b}$, вообще говоря, *нарушается* для экспоненты операторов. Однако есть важный частный случай, когда оно выполнено:

3. Теорема. *Если операторы* $f, g: L \to L$ *коммутируют, т. е.* $fg = gf$, *то* $(\exp f)(\exp g) = \exp(f + g)$.

Доказательство. Применяя формулу бинома Ньютона и пользуясь возможностью переставлять члены абсолютно сходящегося ряда, получаем

$$(\exp f)(\exp g) = \left(\sum_{i \geqslant 0} \frac{1}{i!} f^i\right)\left(\sum_{k \geqslant 0} \frac{1}{k!} g^k\right) = \sum_{i,\,k \geqslant 0} \frac{1}{i!\,k!} f^i g^k =$$

$$= \sum_{m \geqslant 0} \sum_{i=0}^{m} \frac{1}{i!\,(m-i)!} f^i g^{m-i} = \sum_{m \geqslant 0} \frac{1}{m!} \sum_{i=0}^{m} \frac{m!}{i!\,(m-i)!} f^i g^{m-i} =$$

$$= \sum_{m \geqslant 0} \frac{1}{m!} (f+g)^m = \exp(f+g).$$

Коммутативность f и g используется в том месте, где $(f+g)^m$ разлагается по биному.

4. Следствие. *Пусть* f: $L \to L$ — *ограниченный оператор. Тогда отображение* $\mathbf{R} \to \mathscr{L}^1(L, L)$: $t \longmapsto \exp(tf)$ *является гомоморфизмом группы* $\mathbf{R}$ *в подгруппу обратимых операторов* $\mathscr{L}^1(L, L)$ *по умножению.*

Множество операторов $\{\exp tf \mid t \in R\}$ называется *однопараметрической подгруппой операторов.*

5. Спектр. Пусть f — некоторый оператор в конечномерном пространстве, $Q(t)$ — такой степенной ряд, что $Q(f)$ абсолютно сходится. Нетрудно видеть, что если $Q(t)$ — многочлен, то в жордановом базисе f матрица $Q(f)$ является верхней треугольной, и на ее диагонали стоят числа $Q(\lambda_i)$, где λ_i — собственные значения f. Применив это соображение к частичным суммам Q и перейдя к пределу, получим, что это же верно для любого ряда $Q(t)$. В частности, если $S(f)$ — спектр f, то $S(Q(f)) = Q(S(f)) = \{Q(\lambda) \mid \lambda \in S(f)\}$. Более того, если учитывать характеристические корни λ_i с их кратностью, то $Q(S(f))$ будет спектром $Q(f)$ с правильными кратностями. В частности,

$$\det(\exp f) = \prod_{i=1}^{n} \exp \lambda_i = \exp\left(\sum_{i=1}^{n} \lambda_i\right) = \exp \operatorname{Tr} f.$$

Переходя на язык матриц, мы отметим еще два простых свойства, которые доказываются таким же образом:

а) $Q(A^t) = Q(A)^t$;

б) $Q(\bar{A}) = \overline{Q(A)}$, где черта означает комплексное сопряжение; здесь предполагается, что ряд Q имеет вещественные коэффициенты.

Пользуясь этими свойствами и обозначениями § 4, докажем следующую теорему, относящуюся к теории классических групп Ли (здесь $\mathscr{K} = \mathbf{R}$ или $\mathbf{C}$).

6. Теорема. *Отображение* exp *переводит* $\mathrm{gl}(n, \mathscr{K})$, $\mathrm{sl}(n, \mathscr{K})$, $\mathrm{o}(n, \mathscr{K})$, $\mathrm{u}(n)$, $\mathrm{su}(n)$ *в* $\mathrm{GL}(n, \mathscr{K})$, $\mathrm{SL}(n, \mathscr{K})$, $\mathrm{SO}(n, \mathscr{K})$, $\mathrm{U}(n)$, $\mathrm{SU}(n)$ *соответственно.*

Доказательство. Пространство $\mathrm{gl}(n, \mathscr{K})$ переходит в $\mathrm{GL}(n, \mathscr{K})$, ибо согласно следствию п. 4 матрицы $\exp A$ обратимы. Если $\operatorname{Tr} A = 0$, то $\det \exp A = 1$, как было доказано в предыдущем пункте. Из условия $A + A^t = 0$ следует, что $(\exp A)(\exp A)^t = 1$, а из условия $A + \bar{A}^t = 0$ следует, что $\exp A\, \overline{(\exp A)}^t = 1$. Это завершает доказательство.

7. Замечание. Во всех случаях образ exp покрывает некоторую окрестность единицы соответствующей группы. Для доказательства можно определить логарифм операторов f с условием $\|f - \mathrm{id}\| < 1$ обычной формулой $\log f = \sum_{n \geqslant 0} (-1)^n \frac{(f - \mathrm{id})^n}{n}$ и показать, что $f = \exp(\log f)$.

Однако в целом отображения exp, вообще говоря, не сюръективны. Например, не существует матрицы $A \in \mathrm{sl}(2, \mathbf{C})$, для которой $\exp A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \in \mathrm{SL}(2, \mathbf{C})$. В самом деле, A не может быть диагонализируемой, иначе $\exp A$ была бы диагонализируемой. Значит, собственные значения A совпадают, а так как след A равен нулю, эти собственные значения должны быть нулевыми. Но тогда собственные значения $\exp A$ равны 1, тогда как собственные значения $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ равны -1.

§ 12. Комплексификация и овеществление

1. В § 8 и 9 мы убедились, что работа над алгебраически замкнутым полем проясняет геометрическую структуру линейных операторов и дает удобную каноническую форму матриц. Поэтому, даже работая с вещественным полем, удобно иногда пользоваться комплексными числами. В этом параграфе будут изучены две основные операции: увеличения и уменьшения поля скаляров в применении к линейным пространствам и линейным отображениям. Мы уделим больше всего внимания переходу от $\mathbf{R}$ к $\mathbf{C}$ (*комплексификация*) и от $\mathbf{C}$ к $\mathbf{R}$ (*овеществление*) и кратко коснемся более общего случая.

2. Овеществление. Пусть L — линейное пространство над $\mathbf{C}$. Забудем про возможность умножать векторы из L на все комплексные числа и оставим лишь умножение на $\mathbf{R}$. Очевидно, мы получим линейное пространство над $\mathbf{R}$, которое будем обозначать $L_{\mathbf{R}}$ и называть *овеществлением* L.

Пусть L, M — два линейных пространства над $\mathbf{C}$, $f\colon L\to M$ — линейное отображение. Очевидно, рассмотренное как отображение $L_{\mathbf{R}}\to M_{\mathbf{R}}$, оно остается линейным. Мы будем обозначать его $f_{\mathbf{R}}$ и называть овеществлением f. Ясно, что $\mathrm{id}_{\mathbf{R}}=\mathrm{id}$, $(fg)_{\mathbf{R}}=f_{\mathbf{R}}g_{\mathbf{R}}$; $(af+bg)_{\mathbf{R}}=af_{\mathbf{R}}+bg_{\mathbf{R}}$, если a, $b\in\mathbf{R}$.

3. Теорема. а) *Пусть* $\{e_1, \ldots, e_m\}$ — *базис пространства* L *над* $\mathbf{C}$. *Тогда* $\{e_1, \ldots, e_m, ie_1, \ldots, ie_m\}$ *является базисом пространства* $L_{\mathbf{R}}$ *над* $\mathbf{R}$. *В частности*, $\dim_{\mathbf{R}} L_{\mathbf{R}}=2\dim_{\mathbf{C}} L$.

б) *Пусть* $A=B+iC$ — *матрица линейного отображения* $f\colon L\to M$ *в базисах* $\{e_1, \ldots, e_m\}$ *и* $\{e'_1, \ldots, e'_n\}$ *над* $\mathbf{C}$, *где* B, C — *вещественные матрицы. Тогда матрицей линейного отображения* $f_{\mathbf{R}}\colon L_{\mathbf{R}}\to M_{\mathbf{R}}$ *в базисах* $\{e_1, \ldots, e_m, ie_1, \ldots, ie_m\}$, $\{e'_1, \ldots, e'_n, ie'_1, \ldots, ie'_n\}$ *будет*

$$\begin{pmatrix} B & -C \\ C & B \end{pmatrix}.$$

Доказательство. а) Для любого элемента $l\in L$ имеем

$$l=\sum_{k=1}^{m} a_k e_k=\sum_{k=1}^{m}(b_k+ic_k)e_k=\sum_{k=1}^{m} b_k e_k+\sum_{k=1}^{m} c_k(ie_k),$$

где b_k, c_k — вещественная и мнимая части a_k. Поэтому $\{e_k, ie_k\}$ порождают $L_{\mathbf{R}}$. Если $\sum_{k=1}^{m} b_k e_k+\sum_{k=1}^{m} c_k(ie_k)=0$, где b_k, $c_k\in\mathbf{R}$, то $b_k+ic_k=0$ в силу линейной независимости $\{e_1, \ldots, e_k\}$ над $\mathbf{C}$, откуда следует, что $b_k=c_k=0$ для всех k.

б) Согласно определению A, имеем

$$f(e_1, \ldots, e_m)=(e'_1, \ldots, e'_n)(B+iC),$$

откуда, в силу линейности f над $\mathbf{C}$,

$$f(ie_1, \ldots, ie_m)=(e'_1, \ldots, e'_n)(-C+iB).$$

Поэтому

$$(f(e_1), \ldots, f(e_m), f(ie_1), \ldots, f(ie_m)) = \\ = (e'_1, \ldots, e'_n, ie'_1, \ldots, ie'_m)\begin{pmatrix} B & -C \\ C & B \end{pmatrix},$$

что завершает доказательство.

Следствие. *Пусть* $f\colon L \to L$ — *линейный оператор на конечномерном комплексном пространстве* L. *Тогда* $\det f_{\mathbf{R}} = |\det f|^2$.

Доказательство. Пусть f представлен матрицей $B + iC$ (B, C вещественны) в базисе $\{e_1, \ldots, e_m\}$. Тогда, применяя элементарные преобразования (прямо в блочной структуре) сначала к строкам, потом к столбцам, находим:

$$\det f_{\mathbf{R}} = \det\begin{pmatrix} B & -C \\ C & B \end{pmatrix} = \det\begin{pmatrix} B+iC & -C+iB \\ C & B \end{pmatrix} = \\ = \det\begin{pmatrix} B+iC & 0 \\ C & B-iC \end{pmatrix} = \det(B+iC)\det(B-iC) = \\ = \det f\,\overline{\det f} = |\det f|^2.$$

4. **Спуск поля скаляров: общая ситуация.** Довольно очевидно, как обобщаются определения п. 2. Пусть K — некоторое поле, $\mathcal{K}$ — его подполе, L — линейное пространство над K. Забыв про умножение векторов на все элементы поля K и оставив лишь умножение на элементы $\mathcal{K}$, мы получим линейное пространство $L_{\mathcal{K}}$ над $\mathcal{K}$. Аналогично, линейное отображение $f\colon L \to M$ над K превращается в линейное отображение $f_{\mathcal{K}}\colon L_{\mathcal{K}} \to M_{\mathcal{K}}$. Одно из названий этих операций — *спуск поля скаляров* (от K до $\mathcal{K}$). Ясно, что $\mathrm{id}_{\mathcal{K}} = \mathrm{id}$, $(fg)_{\mathcal{K}} = f_{\mathcal{K}} g_{\mathcal{K}}$, $(af + bg)_{\mathcal{K}} = af_{\mathcal{K}} + bg_{\mathcal{K}}$, если $a, b \in \mathcal{K}$. Само поле K можно также рассматривать как линейное пространство над $\mathcal{K}$. Если оно конечномерно, то размерности $\dim_K L$ и $\dim_{\mathcal{K}} L_{\mathcal{K}}$ связаны формулой

$$\dim_{\mathcal{K}} L_{\mathcal{K}} = \dim_{\mathcal{K}} K \dim_K L.$$

Для доказательства достаточно проверить, что если $\{e_1, \ldots, e_n\}$ — базис L над K, а $\{b_1, \ldots, b_m\}$ — базис K над $\mathcal{K}$, то $\{b_1e_1, \ldots, b_1e_n; \ldots; b_me_1, \ldots, b_me_n\}$ образуют базис $L_{\mathcal{K}}$ над $\mathcal{K}$.

5. **Комплексная структура на вещественном линейном пространстве.** Пусть L — комплексное линейное пространство, $L_{\mathbf{R}}$ — его овеществление. Чтобы полностью восстановить умножение на комплексные числа в $L_{\mathbf{R}}$, достаточно знать оператор $J\colon L_{\mathbf{R}} \to L_{\mathbf{R}}$ умножения на i: $J(l) = il$. Очевидно, этот оператор линеен над $\mathbf{R}$ и удовлетворяет условию $J^2 = -\mathrm{id}$; если мы знаем его, то для любого комплексного числа $a + bi$, $a, b \in \mathbf{R}$, имеем

$$(a + bi)\,l = al + bJ(l).$$

Это соображение приводит к следующему важному понятию:

6. **Определение.** *Пусть* L — *вещественное пространство. Комплексной структурой на* L *называется задание линейного оператора* $J\colon L \to L$, *удовлетворяющего условию* $J^2 = -\mathrm{id}$.

Описанная выше комплексная структура на $L_{\mathbf{R}}$ называется *канонической*. Это определение оправдывается следующей теоремой:

7. Теорема. *Пусть* (L, J) — *вещественное линейное пространство с комплексной структурой. Введем на* L *операцию умножения на комплексные числа из* $\mathbf{C}$ *по формуле*

$$(a+bi)\,l = al + bJ(l).$$

Тогда L *превратится в комплексное линейное пространство* $\tilde{L}$, *для которого* $\tilde{L}_{\mathbf{R}} = L$.

Доказательство. Обе аксиомы дистрибутивности легко проверяются, исходя из линейности J и формул сложения комплексных чисел. Проверим аксиому ассоциативности умножения:

$$(a+bi)\,[(c+di)\,l] = (a+bi)\,[cl + dJ(l)] = a[cl + dJ(l)] + \\ + bJ\,[cl + dJ(l)] = acl + adJ(l) + bcJ(l) - bdl = \\ = (ac-bd)\,l + (ad+bc)\,J(l) = [ac - bd + (ad+bc)\,i]\,l = \\ = [(a+bi)(c+di)]\,l.$$

Все остальные аксиомы выполнены по той причине, что L и $\tilde{L}$ совпадают как аддитивные группы.

8. Следствие. *Если* (L, J) — *конечномерное вещественное пространство с комплексной структурой, то* $\dim_{\mathbf{R}} L = 2n$ *четна, и матрица* J *в подходящем базисе имеет вид*

$$\begin{pmatrix} 0 & -E_n \\ E_n & 0 \end{pmatrix}.$$

Доказательство. Действительно, $\dim_{\mathbf{R}} L = 2\dim_{\mathbf{C}} \tilde{L}$ в силу теоремы п. 7 и утверждения а) теоремы п. 3 (конечномерность $\tilde{L}$ следует из того, что любой базис L над $\mathbf{R}$ порождает $\tilde{L}$ над $\mathbf{C}$). Далее, выберем базис $\{e_1, \ldots, e_n\}$ пространства $\tilde{L}$ над $\mathbf{C}$. Матрица умножения на i в этом базисе равна iE_n. Поэтому матрица оператора J в базисе $\{e_1, \ldots, e_n; ie_1, \ldots, ie_n\}$ пространства L имеет требуемый вид в силу утверждения б) теоремы п. 3.

9. Замечания. а) Пусть L — комплексное пространство, $g\colon L_{\mathbf{R}} \to L_{\mathbf{R}}$ — вещественно линейное отображение. Поставим вопрос, когда существует такое комплексно линейное отображение $f\colon L \to L$, что $g = f_{\mathbf{R}}$. Очевидно, для этого необходимо, чтобы g коммутировал с оператором J естественной комплексной структуры на $L_{\mathbf{R}}$, ибо $g(il) = g(Jl) = ig(l) = Jg(l)$ для всех $l \in L$. Это условие является также достаточным, потому что из него автоматически следует линейность g над $\mathbf{C}$:

$$g((a+bi)\,l) = ag(l) + bg(il) = ag(l) + bgJ(l) = \\ = ag(l) + bJg(l) = (a+bJ)\,g(l) = (a+bi)\,g(l).$$

б) Пусть теперь L — четномерное вещественное пространство, $f\colon L \to L$ — вещественно линейный оператор. Поставим вопрос, когда на L существует такая комплексная структура J, что f является овеществлением комплексно линейного отображения g: $\tilde{L}\to\tilde{L}$, где $\tilde{L}$ — комплексное пространство, построенное с помощью J. Вот частичный ответ, относящийся к случаю $\dim_{\mathbf{R}} L=2$: такая структура существует, если f не имеет собственных векторов в L.

В самом деле, тогда f имеет два комплексно сопряженных собственных значения $\lambda\pm i\mu$, $\lambda,\mu\in R$, $\mu\neq 0$. Положим $J=\mu^{-1}(f-\lambda\,\mathrm{id})$. По теореме Гамильтона — Кэли, $f^2-2\lambda f+(\lambda^2+\mu^2)\,\mathrm{id}=0$, откуда

$$J^2=\mu^{-2}(f^2-2\lambda f+\lambda^2\,\mathrm{id})=-\mathrm{id}.$$

Кроме того, J коммутирует с f. Это завершает доказательство.

10. Комплексификация. Теперь мы фиксируем вещественное линейное пространство L и введем комплексную структуру J на внешней прямой сумме $L\oplus L$, определив ее формулой

$$J(l_1,\ l_2)=(-l_2,\ l_1).$$

Ясно, что $J^2=-1$. Назовем *комплексификацией пространства* L комплексное пространство $\widetilde{L\oplus L}$, связанное с этой структурой. Мы будем обозначать его $L^{\mathbf{C}}$. Другие стандартные обозначения: $\mathbf{C}\underset{\mathbf{R}}{\otimes} L$ или $\mathbf{C}\oplus L$; их происхождение станет ясно после ознакомления с тензорными произведениями линейных пространств. Отождествив L с подмножеством векторов вида $(l,0)$ в $L\oplus L$ и пользуясь тем, что $i(l,0)=J(l,0)=(0,l)$, мы можем записать любой вектор из $L^{\mathbf{C}}$ в виде

$$(l_1,\ l_2)=(l_1,\ 0)+(0,\ l_2)=(l_1,\ 0)+i(l_2,\ 0)=l_1+il_2.$$

Иными словами, $L^{\mathbf{C}}=L\oplus iL$, последняя сумма является прямой над $\mathbf{R}$, но не над $\mathbf{C}$!

Любой базис L над $\mathbf{R}$ будет базисом $L^{\mathbf{C}}$ над $\mathbf{C}$, так что $\dim_{\mathbf{R}} L=\dim_{\mathbf{C}} L^{\mathbf{C}}$.

Пусть теперь f: $L\to M$ — линейное отображение линейных пространств над $\mathbf{R}$. Тогда отображение $f^{\mathbf{C}}$ (или $f\otimes\mathbf{C}$): $L^{\mathbf{C}}\to M^{\mathbf{C}}$, определенное формулой

$$f(l_1,\ l_2)=(f(l_1),\ f(l_2)),$$

линейно над $\mathbf{R}$ и перестановочно с J, ибо

$$fJ(l_1,\ l_2)=f(-l_2,\ l_1)=(-f(l_2),\ f(l_1))=Jf(l_1,\ l_2).$$

Следовательно, оно комплексно линейно. Оно называется комплексификацией отображения f. Очевидно, $\mathrm{id}^{\mathbf{C}}=\mathrm{id}$, $(af+bg)^{\mathbf{C}}=af^{\mathbf{C}}+bg^{\mathbf{C}}$; $a,b\in\mathbf{R}$; и $(fg)^{\mathbf{C}}=f^{\mathbf{C}}g^{\mathbf{C}}$. Рассматривая пару базисов L и M как базисы $L^{\mathbf{C}}$ и $M^{\mathbf{C}}$ соответственно, убеждаемся, что матрица отображения f в исходной паре базисов совпадает с матрицей отображения $f^{\mathbf{C}}$ в этой «новой» паре. В частности, (комплексные) собственные значения отображений f и $f^{\mathbf{C}}$ и их жордановы формы совпадают.

Проследим теперь, что происходит при композиции операций овеществления и комплексификации в двух возможных порядках

11. Сначала комплексификация, затем овеществление. Пусть L — вещественное пространство. Мы утверждаем, что существует естественный изоморфизм

$$(L^{\mathbf{C}})_{\mathbf{R}} \to L \oplus L.$$

Действительно, по конструкции $L^{\mathbf{C}}$ совпадает с $L \oplus L$ как вещественное пространство. Аналогично, $(f^{\mathbf{C}})_{\mathbf{R}} \to f \oplus f$ (в смысле этого отождествления) для любого вещественного линейного отображения $f: L \to M$.

Композиция в обратном порядке приводит к несколько менее очевидному ответу. Введем следующее определение.

12. Определение. *Пусть L — комплексное пространство. Сопряженным комплексным пространством $\bar{L}$ называется множество L с той же структурой аддитивной группы, но с новым умножением на скаляры из* $\mathbf{C}$, *которое мы временно обозначим $a*l$:*

$$a * l = \bar{a}l \quad \textit{для любых} \quad a \in \mathbf{C},\ l \in L.$$

Аксиомы проверяются без труда, если воспользоваться тем, что $\overline{ab} = \bar{a}\bar{b}$ и $\overline{a+b} = \bar{a} + \bar{b}$.

Аналогично, если (L, J) — вещественное пространство с комплексной структурой, оператор $-J$ также определяет комплексную структуру, которая называется *сопряженной* с исходной. В обозначениях теоремы п. 7, если $\tilde{L}$ — комплексное пространство, отвечающее (L, J), то $\bar{\tilde{L}}$ — комплексное пространство, отвечающее $(L, -J)$.

13. Сначала овеществление, потом комплексификация. Теперь мы можем для всякого комплексного линейного пространства L построить канонический комплексно линейный изоморфизм

$$f: (L_{\mathbf{R}})^{\mathbf{C}} \to L \oplus \bar{L}.$$

С этой целью заметим, что на $(L_{\mathbf{R}})^{\mathbf{C}}$ имеются два вещественно линейных оператора: оператор канонической комплексной структуры $J(l_1, l_2) = (-l_2, l_1)$ и оператор умножения на i, отвечающий исходной комплексной структуре L: $i(l_1, l_2) = (il_1, il_2)$. Так как J коммутирует с i, он комплексно линеен в этой структуре. Поскольку $J^2 = -\mathrm{id}$, его собственные значения равны $\pm i$. Введем стандартные обозначения для двух подпространств, отвечающих этим собственным значениям:

$$L^{1,0} = \{(l_1, l_2) \in (L_{\mathbf{R}})^{\mathbf{C}} \mid J(l_1, l_2) = i(l_1, l_2)\},$$
$$L^{0,1} = \{(l_1, l_2) \in (L_{\mathbf{R}})^{\mathbf{C}} \mid J(l_1, l_2) = -i(l_1, l_2)\}.$$

Оба множества $L^{1,0}$ и $L^{0,1}$ являются комплексными подпространствами в $(L_{\mathbf{R}})^{\mathbf{C}}$: ясно, что они замкнуты относительно сложения и умножения на вещественные числа, а замкнутость относительно умножения на J следует из того, что J и i коммутируют. Покажем, что $L = L^{1,0} \oplus L^{0,1}$, а также, что $L^{1,0}$ естественно изоморфно L, тогда как $L^{0,1}$ естественно изоморфно $\bar{L}$.

Из определений сразу же следует, что $L^{1,0}$ состоит из векторов вида $(l, -il)$, а $L^{0,1}$ — из векторов вида (m, im). Для данных $l_1, l_2 \in L$ уравнение $(l_1, l_2) = (l, -il) + (m, im)$ на l, m имеет единственное решение $l = \frac{l_1 + il_2}{2}$, $m = \frac{l_1 - il_2}{2}$. Следовательно, $L = L^{1,0} \oplus L^{0,1}$. Отображения $L \to L^{1,0}$: $l \mapsto (l, -il)$ и $\bar{L} \to L^{0,1}$: $l \mapsto (l, il)$ являются вещественно линейными изоморфизмами. Кроме того, они перестановочны с действием i на L, $\bar{L}$ и действием J на $L^{1,0}$, $L^{0,1}$ в силу определений. Это завершает нашу конструкцию.

14. Полулинейные отображения комплексных пространств. Пусть L, M — комплексные линейные пространства. *Полулинейным* (или *антилинейным*) отображением f: $L \to M$ называется линейное отображение f: $L \to \bar{M}$. Иными словами, f — гомоморфизм аддитивных групп, и

$$f(al) = \bar{a} f(l)$$

для всех $a \in \mathbf{C}$, $l \in L$. Особая роль полулинейных отображений станет ясна во второй части, при изучении эрмитовых комплексных пространств.

15. Подъем поля скаляров: общая ситуация. Пусть, как в п. 4, K — некоторое поле, $\mathcal{K}$ — его подполе. Тогда для любого линейного пространства L над $\mathcal{K}$ можно определить линейное пространство $K \underset{\mathcal{K}}{\otimes} L$, или L^K, над K, сохранив размерность. До введения языка тензорных произведений дать общее определение L^K затруднительно, но для практических целей достаточно следующего полуфабриката: если $\{e_1, \ldots, e_n\}$ — базис L над $\mathcal{K}$, то L^K состоит из всех формальных линейных комбинаций $\left\{\sum_{i=1}^{n} a_i e_i \mid a_i \in K\right\}$, т. е. имеет тот же базис над K. В частности, $(\mathcal{K}^n)^K = K^n$. По $\mathcal{K}$-линейному отображению f: $L \to M$ определяется K-линейное отображение f^K: $L^K \to M^K$: если f задано матрицей в некоторых базисах L и M, то f^K задается той же матрицей.

В заключение укажем одно приложение комплексификации:

16. Предложение. *Пусть f: $L \to L$ — линейный оператор в вещественном пространстве размерности $\geqslant 1$. Тогда f имеет инвариантное подпространство размерности* 1 *или* 2.

Доказательство. Если f имеет вещественное собственное значение, то подпространство, натянутое на соответствующий собственный вектор, инвариантно. В противном случае все собственные значения комплексны. Выберем одно из них $\lambda + i\mu$. Оно будет также собственным значением $f^{\mathbf{C}}$ в $L^{\mathbf{C}}$. Возьмем соответствующий собственный вектор $l_1 + il_2$ в $L^{\mathbf{C}}$, $l_1, l_2 \in L$. Согласно определениям

$$f^{\mathbf{C}}(l_1 + il_2) = f(l_1) + if(l_2) = (\lambda + i\mu)(l_1 + il_2) = (\lambda l_1 - \mu l_2) + i(\mu l_1 + \lambda l_2).$$

Следовательно, $f(l_1) = \lambda l_1 - \mu l_2$, $f(l_2) = \mu l_1 + \lambda l_2$, и линейная оболочка $\{l_1, l_2\}$ в L f-инвариантна.

§ 13. Язык категорий

1. Определение категории. Категория C состоит из следующих данных:

а) Класс (или множество) $\mathrm{Ob}\, C$, элементы которого называются *объектами* категории.

б) Класс (или множество) $\mathrm{Mor}\, C$, элементы которого называются *морфизмами* категории, или *стрелками*.

в) Для каждой упорядоченной пары объектов $X, Y \in \mathrm{Ob}\, C$ задано множество $\mathrm{Hom}_C(X, Y) \subset \mathrm{Mor}\, C$, элементы которого называются *морфизмами из* X *в* Y и обозначаются $X \to Y$ или $f\colon X \to Y$ или $X \xrightarrow{f} Y$.

г) Для каждой упорядоченной тройки объектов $X, Y, Z \in \mathrm{Ob}\, C$ задано отображение

$$\mathrm{Hom}_C(X, Y) \times \mathrm{Hom}_C(Y, Z) \to \mathrm{Hom}_C(X, Z),$$

сопоставляющее паре морфизмов (f, g) морфизм gf, или $g \circ f$, называемый их *композицией*, или *произведением*.

Эти данные должны удовлетворять следующим условиям:

д) $\mathrm{Mor}\, C$ есть несвязное объединение $\bigcup \mathrm{Hom}_C(X, Y)$ по всем упорядоченным парам $X, Y \in \mathrm{Ob}\, C$. Иными словами, для каждого морфизма f однозначно определены объекты X, Y такие, что $f \in \mathrm{Hom}\, C(X, Y)$: *начало* X и *конец* Y стрелки f.

е) Композиция морфизмов ассоциативна.

ж) Для каждого объекта X существует тождественный морфизм $\mathrm{id}_X \in \mathrm{Hom}_C(X, X)$ такой, что $\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$ всякий раз, когда эти композиции определены. Нетрудно видеть, что такой морфизм единствен: если id'_X — другой морфизм с тем же свойством, то $\mathrm{id}'_X = \mathrm{id}'_X \circ \mathrm{id}_X = \mathrm{id}_X$.

Морфизм $f\colon X \to Y$ называется *изоморфизмом*, если существует такой морфизм $g\colon Y \to X$, *что* $gf = \mathrm{id}_X$, $fg = \mathrm{id}_Y$.

2. Примеры. а) *Категория множеств* Set. Ее объекты — множества, морфизмы — отображения множеств.

б) *Категория* $\mathcal{L}in_{\mathcal{K}}$ *линейных пространств над полем* $\mathcal{K}$. Ее объекты — линейные пространства, морфизмы — линейные отображения.

в) *Категория групп.*

г) *Категория абелевых групп.*

Различия между классом и множеством обсуждаются в аксиоматической теории множеств и связаны с необходимостью избежать знаменитого парадокса Рассела. Не всякое собирание объектов воедино образует множество, ибо понятие «множество всех множеств, не содержащих самих себя в качестве элемента», противоречиво. В аксиоматике Гёделя — Бернайса такие собрания множеств называются классами. Техника теорий категорий требует собираний объектов, лежащих в опасной близости к таким парадоксальным ситуациям. Мы, однако, будем пренебрегать этими тонкостями.

3. Диаграммы. Поскольку в аксиоматике категорий ничего не говорится о теоретико-множественной структуре объектов, мы не можем в общем случае работать с «элементами» этих объектов. Все основные общекатегорные конструкции и их приложения к конкретным категориям формулируются преимущественно в терминах морфизмов и их композиции. Удобный язык для таких формулировок — это язык диаграмм. Например, вместо того чтобы говорить, что у нас имеются четыре объекта X, Y, U, V, четыре морфизма $f \in \mathrm{Hom}_C(X, Y)$, $g \in \mathrm{Hom}_C(Y, V)$, $h \in \mathrm{Hom}_C(X, U)$ и $d \in \mathrm{Hom}_C(U, V)$, причем $gf = dh$, говорят, что задан *коммутативный квадрат*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & Y \arrow[d, "g"] \\ U \arrow[r, "d"] & V \end{tikzcd}$$

«Коммутативность» здесь — это равенство $gf = dh$, которое означает, что «два пути вдоль стрелок» от X к V приводят к одному и тому же результату. Более общо, *диаграмма* — это ориентированный граф, вершины которого являются объектами C, а ребра — морфизмами, например

$$\begin{tikzcd} X \arrow[r] & Y \arrow[r] \arrow[d] & Z \arrow[d] \\ U \arrow[r] & V \arrow[r] & W \end{tikzcd}$$

Диаграмма называется *коммутативной*, если любые пути вдоль стрелок в ней с общими началом и концом отвечают одинаковым морфизмам.

В категории линейных пространств, а также в категории абелевых групп особенно важен класс диаграмм, называемых *комплексами*. *Комплекс* имеет вид последовательности объектов и стрелок

$$\ldots X \to Y \to U \to V \to \ldots,$$

конечной или бесконечной, которая удовлетворяет следующему условию: *композиция любых двух соседних стрелок является нулевым морфизмом.* Заметим, что понятие нулевого морфизма не является общекатегорным: оно специфично для линейных пространств и абелевых групп и для специального класса категорий — так называемых аддитивных категорий. Часто объекты, входящие в комплекс, и морфизмы нумеруются некоторым отрезком целых чисел:

$$\ldots \to X_{-1} \xrightarrow{f_{-1}} X_0 \xrightarrow{f_0} X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} \ldots$$

Такой комплекс линейных пространств (или абелевых групп) называется *точным в члене* X_i, если $\operatorname{Im} f_{i-1} = \operatorname{Ker} f_i$ (заметим, что в определении комплекса условие $f_i \circ f_{i-1} = 0$ означает только, что $\operatorname{Im} f_{i-1} \subset \operatorname{Ker} f_i$). Комплекс, точный во всех членах, называется *точным*, или *ацикличным*, или *точной последовательностью*.

Вот три простейших примера:

а) Последовательность $0 \to L \xrightarrow{i} M$ всегда является комплексом; она точна в члене L тогда и только тогда, когда $\operatorname{Ker} i$ — образ нулевого пространства 0. Иными словами, точность здесь означает, что i — инъекция.

б) Последовательность $M \xrightarrow{j} N \to 0$ всегда является комплексом; точность его в члене N означает, что $\operatorname{Im} j = N$, т. е. что j — сюръекция.

в) Комплекс $0 \to L \xrightarrow{i} M \xrightarrow{j} N \to 0$ точен, если i — инъекция, j — сюръекция и $\operatorname{Im} i = \operatorname{Ker} j$. Отождествив L с образом i — подпространством в M, мы можем поэтому отождествить N с факторпространством M/L, так что такие *«точные тройки»* являются категорными представителями троек $(L \subset M,\ M/L)$.

4. Естественные конструкции и функторы. В математике весьма важны конструкции, которые можно применять к объектам некоторой категории так, что при этом снова получаются объекты категории (другой или той же самой). Если эти конструкции являются однозначными (не зависят от произвольных выборов) и универсально применимыми, то часто оказывается, что они переносятся и на морфизмы. Аксиоматизация ряда примеров привела к важному понятию функтора, впрочем, естественному и с чисто категорной точки зрения

5. Определение функтора. Пусть C, D — две категории. *Функтором* F из категории C в категорию D называется задание двух отображений (обычно обозначаемых также F): $\operatorname{Ob} C \to \operatorname{Ob} D$, $\operatorname{Mor} C \to \operatorname{Mor} D$, которые удовлетворяют следующим условиям:

а) если $f \in \operatorname{Hom}_C(X, Y)$, то $F(f) \in \operatorname{Hom}_D(F(X),\ F(Y))$;

б) $F(gf) = F(g)F(f)$ всякий раз, когда композиция gf определена, и $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ для всех $X \in \operatorname{Ob} C$.

Функторы, которые мы определили, часто называют *ковариантными* функторами. Определяют также *контравариантные* функторы, «обращающие стрелки»: для них условия а) и б) заменяются условиями

а′) если $f \in \operatorname{Hom}_C(X, Y)$, то $F(f) \in \operatorname{Hom}_D(F(Y),\ F(X))$;

б′) $F(gf) = F(f)F(g)$ и $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

Можно избежать этого различения, если ввести конструкцию, ставящую в соответствие каждой категории C *дуальную категорию* C° по правилу: $\operatorname{Ob} C = \operatorname{Ob} C^\circ$, $\operatorname{Mor} C = \operatorname{Mor} C^\circ$ и $\operatorname{Hom}_C(X, Y) = \operatorname{Hom}_{C^\circ}(Y, X)$, причем композиция gf морфизмов в C отвечает композиции fg этих же морфизмов в C°, взятых в обратном порядке. Удобно обозначать через X°, f° объекты и морфизмы в C°, отвечающие объектам и морфизмам X, f в C. Тогда коммутативная диаграмма в C

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "gf"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

отвечает коммутативной диаграмме в C°:

$$\begin{tikzcd} X^\circ & Y^\circ \arrow[l, "f^\circ"'] \\ & Z^\circ \arrow[ul, "f^\circ g^\circ"] \arrow[u, "g^\circ"'] \end{tikzcd}$$

(Ковариантный) функтор $F\colon C \to D^\circ$ можно отождествить с контравариантным функтором $F\colon C \to D$ в смысле данного выше определения.

6. Примеры. а) Пусть $\mathcal{K}$ — поле, $\mathcal{L}in_{\mathcal{K}}$ — категория линейных пространств над $\mathcal{K}$, Set — категория множеств. В § 1 мы объяснили, как любому множеству $S \in \mathrm{Ob}\,\mathrm{Set}$ поставить в соответствие линейное пространство $F(S) \in \mathrm{Ob}\,\mathcal{L}in_{\mathcal{K}}$ функций на S со значениями в $\mathcal{K}$. Поскольку это естественная конструкция, следует ожидать, что она может быть продолжена до функтора. Так оно и есть. Функтор оказывается *контравариантным*: морфизму $f\colon S \to T$ он ставит в соответствие линейное отображение $F(f)\colon F(T) \to F(S)$, чаще обозначаемое f^* и называемое *подъемом*, или *обратным образом*, на функциях:

$$f^*(\varphi) = \varphi \circ f, \quad \text{где} \quad f\colon S \to T, \quad \varphi\colon T \to \mathcal{K}.$$

Иными словами, $f^*(\varphi)$ — это функция на S, значения которой постоянны вдоль «слоев» $f^{-1}(t)$ отображения f и равны $\varphi(t)$ на таком слое. Хорошее упражнение для читателя — проверить, что мы действительно построили функтор.

б) Отображение двойственности: $\mathcal{L}in_{\mathcal{K}} \to \mathcal{L}in_{\mathcal{K}}$, на объектах задаваемое формулой $L \mapsto L^* = \mathcal{L}(L, \mathcal{K})$, а на морфизмах — формулой $f \mapsto f^*$, является *контравариантным функтором* из категории $\mathcal{L}in_{\mathcal{K}}$ в себя. По существу, это доказано в § 7.

в) Операции комплексификации и овеществления, изученные в § 12, определяют функторы $\mathcal{L}in_{\mathbf{R}} \to \mathcal{L}in_{\mathbf{C}}$ и $\mathcal{L}in_{\mathbf{C}} \to \mathcal{L}in_{\mathbf{R}}$ соответственно. То же относится к более общим конструкциям подъема и спуска поля скаляров, кратко описанным в § 12.

г) Для любой категории C и любого объекта $X \in \mathrm{Ob}\,C$ определены два функтора из C в категорию множеств: ковариантный $h_X\colon C \to \mathrm{Set}$ и контравариантный $h^X\colon C \to \mathrm{Set}^\circ$.

Вот их определения: $h_X(Y) = \mathrm{Hom}_C(X, Y)$, $h_Y(f\colon Y \to Z)$ есть отображение $h_X(Y) = \mathrm{Hom}_C(X, Y) \to h_X(Z) = \mathrm{Hom}_C(X, Z)$, которое ставит в соответствие морфизму $X \to Y$ его композицию с морфизмом $f\colon Y \to Z$.

Аналогично, $h^X(Y) = \mathrm{Hom}_C(Y, X)$ и $h^X(f\colon Y \to Z)$ есть отображение $h^X(Z) = \mathrm{Hom}_C(Z, X) \to h^X(Y) = \mathrm{Hom}_C(Y, X)$, которое ставит в соответствие морфизму $Z \to X$ его композицию с морфизмом $f\colon Y \to Z$.

Проверьте, что h_X и h^X действительно являются функторами. Их называют функторами, *представляющими* объект X категории.

Заметим, что если $C = \mathcal{L}in_{\mathcal{K}}$, то h^X и h_X можно считать функторами со значениями также в $\mathcal{L}in_{\mathcal{K}}$, а не в Set.

7. Композиция функторов. Если $C_1 \xrightarrow{F} C_2 \xrightarrow{G} C_3$ — три категории и два функтора между ними, то композиция $GF\colon C_1 \to C_3$ определяется как теоретико-множественная композиция отображений на объектах и морфизмах. Тривиально проверяется, что она является функтором.

Можно ввести «категорию категорий», объектами которой являются категории, а морфизмами — функторы!

Более важной, однако, является следующая ступень этой высокой лестницы абстракций: *категория функторов*. Мы ограничимся объяснением, что такое морфизмы функторов.

8. Естественные преобразования естественных конструкций, или функторные морфизмы. Пусть F, $G\colon C \to D$ — два функтора с общими началом и концом. *Функторным морфизмом* $\varphi\colon F \to G$ называется класс морфизмов объектов $\varphi(X)\colon F(X) \to G(X)$ в категории D, по одному для каждого объекта X категории C, обладающий тем свойством, что для каждого морфизма $f\colon X \to Y$ в категории C квадрат

$$\begin{array}{ccc} F(X) & \xrightarrow{\varphi(X)} & G(X) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(Y) & \xrightarrow{\varphi(Y)} & G(Y) \end{array}$$

коммутативен. Функторный морфизм называется *изоморфизмом*, если все $\varphi(X)$ суть изоморфизмы.

9. Пример. Пусть $**\colon \mathcal{L}in_{\mathcal{K}} \to \mathcal{L}in_{\mathcal{K}}$ — функтор «двойного сопряжения»: $L \mapsto L^{**}$, $f \mapsto f^{**}$. В § 7 мы построили для каждого линейного пространства L каноническое линейное отображение $\varepsilon_L\colon L \mapsto L^{**}$. Оно определяет функторный морфизм $\varepsilon_M\colon \mathrm{Id} \to **$, где Id — тождественный функтор на $\mathcal{L}in_{\mathcal{K}}$, ставящий в соответствие каждому линейному пространству само это пространство и каждому линейному отображению — само это отображение. Действительно, согласно определению, мы должны проверить коммутативность всевозможных квадратов вида

$$\begin{array}{ccc} L & \xrightarrow{\varepsilon_L} & L^{**} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f^{**}} \\ M & \xrightarrow{\varepsilon_M} & M^{**} \end{array}$$

Для конечномерных пространств L, M это устанавливается с помощью утверждения д) теоремы п. 5 § 7. Проверку в общем случае предоставляем читателю.

УПРАЖНЕНИЯ

1. Пусть Set_0 — категория, объектами которой являются множества, а морфизмами — такие отображения множеств $f\colon S \to T$, что для любой точки $t \in T$ слой $f^{-1}(t)$ конечен. Показать, что следующие данные определяют *ковариантный* функтор $F_0\colon \mathrm{Set}_0 \to \mathcal{L}in_{\mathcal{K}}$:

а) $F_0(S) = F(S)$: функции на S со значениями в $\mathcal{K}$;

б) для любого морфизма $f\colon S \to T$ и функции $\varphi\colon S \to \mathcal{K}$ функция $F_0(f)(\varphi) = f_*(\varphi) \in F(T)$ определяется так:

$$f_*(\varphi)(t) = \sum_{s \in f^{-1}(t)} \varphi(s)$$

(«интегрирование по слоям»).

2. Доказать, что спуск поля скаляров с K до $\mathcal{K}$ (см. § 12) определяет функтор $\mathcal{L}in_{\mathcal{K}} \to \mathcal{L}in_{\mathcal{K}}$.

§ 14. Категорные свойства линейных пространств

1. В этом параграфе собраны некоторые утверждения о категории всех линейных пространств $\mathcal{L}in_{\mathcal{K}}$ или конечномерных линейных пространств $\mathcal{L}in\,f_{\mathcal{K}}$ над данным полем $\mathcal{K}$. По большей части они являются переформулировкой на категорном языке утверждений, которые мы уже доказали раньше. Их выбор обусловлен следующим своеобразным критерием: это как раз те свойства категории $\mathcal{L}in_{\mathcal{K}}$, которые *нарушаются* для таких наиболее близких категорий, как категория модулей над общими кольцами (например, над $\mathbf{Z}$, т. е. категория абелевых групп) или даже категория бесконечномерных топологических пространств. Детальное изучение этих нарушений для категории модулей составляет основной предмет *гомологической алгебры*, а в функциональном анализе часто приводит к поиску новых определений, которые позволяют восстановить «хорошие» свойства $\mathcal{L}in f_{\mathcal{K}}$ (таково понятие ядерных топологических пространств).

2. Теорема о продолжении отображений.

а) *Пусть P, M, N — линейные пространства, P конечномерно, $j\colon M \to N$ — сюръективное линейное отображение. Тогда любое линейное отображение $g\colon P \to N$ можно поднять до такого линейного отображения $h\colon P \to M$, что $g = jh$. Иными словами, диаграмму с точной строкой*

$$\begin{array}{ccccc} & & P & & \\ & & \downarrow{\scriptstyle g} & & \\ M & \xrightarrow{j} & N & \longrightarrow & 0 \end{array}$$

можно вложить в коммутативную диаграмму

$$\begin{tikzcd} & P \arrow[dl, "h"'] \arrow[d, "g"] & \\ M \arrow[r, "j"] & N \arrow[r] & 0 \end{tikzcd}$$

б) *Пусть P, L, M — линейные пространства, M конечномерно, i: $L \to M$ — инъективное линейное отображение. Тогда любое линейное отображение g: $L \to P$ можно продолжить до линейного отображения h: $M \to P$ так, что $g = hi$. Иными словами, диаграмму с точной нижней строкой*

$$\begin{tikzcd} & P & \\ M & L \arrow[l, "i"'] \arrow[u, "g"'] & 0 \arrow[l] \end{tikzcd}$$

можно вложить в коммутативную диаграмму

$$\begin{tikzcd} & P & \\ M \arrow[ur, "h"] & L \arrow[l, "i"'] \arrow[u, "g"'] & 0 \arrow[l] \end{tikzcd}$$

Доказательство. а) Выберем базис $\{e_1, \ldots, e_n\}$ в P, положим $e'_i = g(e_i) \in N$. В силу сюръективности j существуют векторы $e''_i \in M$ такие, что $j(e''_i) = e'_i$, $i = 1, \ldots, n$. По предложению п. 3 § 3, существует единственное линейное отображение h: $P \to M$ такое, что $h(e_i) = e''_i$, $i = 1, \ldots, n$. По конструкции $jh(e_i) = j(e''_i) = e'_i = g(e_i)$. Так как $\{e_i\}$ образуют базис P, имеем $jh = g$.

б) Выберем базис $\{e'_1, \ldots, e'_m\}$ пространства L и продолжим $e_k = i(e'_k)$, $1 \leqslant k \leqslant m$, до базиса $\{e_1, \ldots, e_m; e_{m+1}, \ldots, e_n\}$ пространства M. Положим $h(e_i) = g(e'_i)$ при $1 \leqslant i \leqslant m$, $h(e_j) = 0$ при $m + 1 \leqslant j \leqslant n$. Такое отображение существует по тому же предложению п. 3 § 3. Можно также прямо применить предложение п. 8 § 6. Теорема доказана.

В категории модулей объекты P, удовлетворяющие условию а) теоремы (при всех M, N). называются *проективными*, а объекты, удовлетворяющие условию б), — *инъективными*. Мы доказали, что в категории конечномерных линейных пространств все объекты проективны и инъективны.

3. Теорема о точности функтора $\mathscr{L}$. *Пусть $0 \to L \xrightarrow{i} M \xrightarrow{j} N \to$ $\to 0$ — точная тройка конечномерных линейных пространств, P — любое конечномерное пространство над тем же полем. Тогда $\mathscr{L}$*

как функтор отдельно по первому и второму аргументу индуцирует точные тройки линейных пространств:

а) $0 \to \mathscr{L}(P, L) \xrightarrow{i_1} \mathscr{L}(P, M) \xrightarrow{j_1} \mathscr{L}(P, N) \to 0,$

б) $0 \leftarrow \mathscr{L}(L, P) \xleftarrow{i_2} \mathscr{L}(M, P) \xleftarrow{j_2} \mathscr{L}(N, P) \leftarrow 0.$

Доказательство. а) Напомним (см. пример г) п. 6 § 13), что i_1 ставит в соответствие морфизму $P \to L$ его композицию с i: $L \to M$, а j_1 ставит в соответствие морфизму $P \to M$ его композицию с j: $M \to N$. Отображение i_1 инъективно, потому что i — инъекция, так что если композиция $P \to L \longrightarrow M$ нулевая, то $P \to L$ — нулевой морфизм. Отображение j_1 сюръективно в силу утверждения а) теоремы п. 2: любой морфизм g: $P \to N$ можно поднять до морфизма $P \to M$, композиция которого с j дает исходный морфизм. Композиция $j_1 i_1$ нулевая: она переводит стрелку $P \to L$ в стрелку $P \to N$, которая является композицией $P \to L \xrightarrow{i} M \xrightarrow{j} N$, но $ji = 0$.

Мы проверили, таким образом, что последовательность а) является комплексом, и остается установить ее точность в среднем члене, т. е. $\operatorname{Ker} j_1 = \operatorname{Im} i_1$. Мы уже знаем, что $\operatorname{Ker} j_1 \supset \operatorname{Im} i_1$. Для доказательства обратного включения заметим, что если стрелка $P \to M$ лежит в ядре j_1, то композиция этой стрелки с j: $M \to N$ равна нулю, а потому образ P в M лежит в ядре j. Но ядро j совпадает с образом $i(L) \subset M$ в силу точности исходной тройки. Значит, P отображается в подпространство $i(L)$, и потому стрелку $P \to M$ можно поднять до стрелки $P \to L$, композиция которой с i даст исходную стрелку. Это и означает, что последняя лежит в образе i_1.

б) Здесь рассуждения совершенно аналогичны, или, точнее, двойственны. Отображение i_2 сюръективно в силу утверждения б) теоремы п. 2. Отображение j_2 инъективно, потому что если композиция $M \xrightarrow{j} N \to P$ равна нулю, то и стрелка $N \to P$ нулевая, так как j сюръективно. Композиция $i_2 j_2$ равна нулю, ибо композиция $L \to M \xrightarrow{j} N \to P$ нулевая для любой последней стрелки. Поэтому остается доказать, что $\operatorname{Ker} i_2 \subset \operatorname{Im} j_2$ (обратное включение только что проверено). Но стрелка $M \xrightarrow{f} P$ лежит в ядре i_2, если композиция $L \xrightarrow{i} M \xrightarrow{f} P$ нулевая. Значит, $L = \operatorname{Ker} j$ лежит в ядре f. Определим отображение $\bar{f}$: $N \to P$ формулой $\bar{f}(n) = f(j^{-1}(n))$, где $j^{-1}(n) \in M$ — любой прообраз n. От выбора этого прообраза ничего не зависит, ибо $\operatorname{Ker} j \subset \operatorname{Ker} f$. Легко проверить, что $\bar{f}$ линейно и что $j_2(\bar{f}) = f$; в самом деле, $j_2(\bar{f})$ есть композиция $M \xrightarrow{j} N \xrightarrow{\bar{f}} P$, которая переводит $m \in M$ в $f\bar{j}(m) = f(j^{-1}(j(m))) = f(m)$. Теорема доказана.

4. Категорная характеризация размерности. Пусть G — некоторая абелева группа, записываемая аддитивно, χ: $\operatorname{Ob} \mathscr{L}\, in\, f_{\mathscr{K}} \to G$ —

произвольная функция, определенная на конечномерных линейных пространствах и удовлетворяющая двум условиям:

а) если L и M изоморфны, то $\chi(L)=\chi(M)$;

б) для любой точной тройки пространств $0\to L\to M\to N\to 0$ имеем $\chi(M)=\chi(L)+\chi(N)$ (такие функции называются аддитивными).

Имеет место

5. Теорема. *Для любой аддитивной функции* χ *имеем*

$$\chi(L)=\dim_{\mathscr{K}} L\cdot\chi(\mathscr{K}^1),$$

где L — произвольное конечномерное пространство.

Доказательство. Проведем индукцию по размерности L. Если L одномерно, то L изоморфно $\mathscr{K}^1$, так что $\chi(L)=\chi(\mathscr{K}^1)=\dim_{\mathscr{K}} L\cdot\chi(\mathscr{K}^1)$. Пусть теорема доказана для всех L размерности n. Если размерность L равна $n+1$, выберем одномерное подпространство $L_0\subset L$ и рассмотрим точную тройку

$$0\to L_0\xrightarrow{i} L\xrightarrow{j} L/L_0\to 0,$$

где i — вложение L_0, а $j(l)=l+L_0\in L/L_0$. В силу аддитивности χ и индуктивного предположения

$$\chi(L)=\chi(L_0)+\chi(L/L_0)=\chi(\mathscr{K}^1)+\dim_{\mathscr{K}}(L/L_0)\,\chi(\mathscr{K}^1)=$$
$$=\chi(\mathscr{K}^1)+n\chi(\mathscr{K}^1)=(n+1)\,\chi(\mathscr{K}^1)=\dim_{\mathscr{K}} L\cdot\chi(\mathscr{K}^1).$$

Теорема доказана.

Этот результат является самым началом большой алгебраической теории, которая сейчас активно развивается, — так называемой K-теории, лежащей на стыке топологии и алгебры.

УПРАЖНЕНИЯ

1. Пусть K: $0\xrightarrow{d_0} L_1\xrightarrow{d_1} L_2\xrightarrow{d_2}\ldots\xrightarrow{d_{n-1}} L_n\xrightarrow{d_n} 0$ — комплекс конечномерных линейных пространств. Факторпространство $H^i(K)=\operatorname{Ker} d_i/\operatorname{Im} d_{i-1}$ называется i-м пространством когомологий этого комплекса. Число $\chi(K)=\sum_{i=1}^{n}(-1)^i\dim L_i$ называется эйлеровой характеристикой комплекса. Доказать, что

$$\chi(K)=\sum_{i=1}^{n}(-1)^i\dim H^i(K).$$

2. «Лемма о змее». Пусть дана коммутативная диаграмма линейных пространств

$$\begin{tikzcd}
 & L \arrow[r, "d_1"] \arrow[d, "f"'] & M \arrow[r, "d_2"] \arrow[d, "g"] & N \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & L' \arrow[r, "d_1'"] & M' \arrow[r, "d_2'"] & N' &
\end{tikzcd}$$

с точными строками Показать, что существует точная последовательность пространств

$$\operatorname{Ker} f \to \operatorname{Ker} g \to \operatorname{Ker} h \xrightarrow{\delta} \operatorname{Coker} f \to \operatorname{Coker} g \to \operatorname{Coker} h,$$

в которой все стрелки, кроме δ, индуцированы d_1, d_2, d_1', d_2' соответственно, а связывающий гомоморфизм δ (называемый также кограничным оператором) определяется так: чтобы определить $\delta(n)$ для $n \in \operatorname{Ker} h$, следует найти $m \in M$ с $n = d_2(m)$, построить $g(m) \in M'$, найти $l' \in L'$ с $d_1'(l') = g(m)$ и положить $\delta(n) = l' + \operatorname{Im} f \in \operatorname{Coker} f$. В частности, следует проверить существование $\delta(n)$ и его независимость от произвола в промежуточных выборах.

3. Пусть K: $\ldots \to L_i \xrightarrow{d_i} L_{i+1} \to \ldots$ и K': $\ldots \to L_i' \xrightarrow{d_i'} L_{i+1}' \to \ldots$ — два комплекса. Морфизмом f: $K \to K'$ называется такой набор линейных отображений f_i: $L_i \to L_i'$, что все квадраты

$$\begin{tikzcd}
L_i \arrow[r, "d_i"] \arrow[d, "f_i"'] & L_{i+1} \arrow[d, "f_{i+1}"] \\
L_i' \arrow[r, "d_i'"] & L_{i+1}'
\end{tikzcd}$$

коммутативны. Показать, что комплексы и их морфизмы образуют категорию.

4. Показать, что отображение $K \to H^i(K)$ продолжается до функтора из категорий комплексов в категорию линейных пространств.

5. Пусть $0 \to K \xrightarrow{f} K' \xrightarrow{g} K'' \to 0$ — точная тройка комплексов и их морфизмов. По определению, это означает, что для каждого i тройки линейных пространств

$$0 \to L_i \xrightarrow{f_i} L_i' \xrightarrow{g_i} L_i'' \to 0$$

точны. Пусть H^i — соответствующие пространства когомологий. Пользуясь леммой о змее, построить последовательность пространств когомологий

$$\ldots \to H^i(K) \to H^i(K') \to H^i(K'') \xrightarrow{\delta} H^{i+1}(K) \to \ldots$$

и показать, что она является точной.

Часть 2. ГЕОМЕТРИЯ ПРОСТРАНСТВ СО СКАЛЯРНЫМ ПРОИЗВЕДЕНИЕМ

§ 1. О геометрии

1. Эта и следующая части нашего курса посвящены теме, которую можно назвать «линейные геометрии», и ей уместно предпослать краткое обсуждение современного смысла слов «геометрия» и «геометрический». В течение многих столетий под геометрией понималась геометрия Евклида на плоскости и в пространстве. Она продолжает составлять основное содержание обычного школьного курса, и эволюцию геометрических понятий удобно проследить на примере характерных особенностей этой, ныне весьма частной, геометрической дисциплины.

2. **«Фигуры».** Школьная геометрия начинается с изучения таких фигур на плоскости, как прямые, углы, треугольники, окружности и круги и т. п. Естественное обобщение этой ситуации состоит в выборе некоторого пространства M, «объемлющего пространства» нашей геометрии, и некоторого множества подмножеств в M — изучаемых в этом пространстве «фигур».

3. **«Движения».** Вторая существенная компонента школьной геометрии — это измерение длин и углов и выяснение соотношений между линейными и угловыми элементами различных фигур. Потребовалось длительное историческое развитие, прежде чем было осознано, что в основе этих измерений лежит существование отдельного математического объекта — группы движений евклидовой плоскости или евклидова пространства как целого, и что все метрические понятия могут быть определены в терминах этой группы. Например, расстояние между точками является единственной функцией от пары точек, инвариантной относительно группы евклидовых движений (если потребовать ее непрерывности и еще выбрать «единицу длины» — расстояние между выбранной парой точек). «Эрлангенская программа» Ф. Клейна (1872) зафиксировала понимание этого замечательного принципа, и «геометрией» надолго стало изучение пространств M, снабженных достаточно большой группой симметрий, и свойств фигур, инвариантных относительно действия этой группы, включая углы, расстояния и объемы.

4. **«Числа».** Открытием столь же фундаментальной важности (и гораздо более ранним) был декартов «метод координат» и основанная на нем аналитическая геометрия плоскости и пространства.

С современной точки зрения координаты суть некоторые функции на пространстве M (или на его подмножествах) с вещественными, комплексными или еще более общими значениями. Задание конкретных значений этих функций позволяет зафиксировать точку пространства, а задание соотношений между этими значениями определяет множество точек. Описание класса рассматриваемых в данной геометрии фигур в M можно заменить описанием класса соотношений между координатами, которые описывают интересующие нас фигуры. Поразительная гибкость и сила метода Декарта связана с тем, что функции на пространстве можно складывать и умножать, интегрировать, дифференцировать и применять другие процессы предельного перехода и в конечном счете пользоваться всей мощью математического анализа. Все общие современные геометрические дисциплины — топология, дифференциальная и комплексно аналитическая геометрия, алгебраическая геометрия — выбирают в качестве исходного определения понятие геометрического объекта как совокупности пространства M и заданной на нем совокупности F (локальных) функций.

5. «Отображения». Если (M_1, F_1) и (M_2, F_2) — два геометрических объекта описанного выше типа, то можно рассматривать отображения $M_1 \to M_2$, которые обладают тем свойством, что обратное отображение на функциях переводит элементы из F_2 в элементы из F_1. В наиболее логически завершенных схемах среди таких отображений находятся как группы симметрий Ф. Клейна, так и сами координатные функции (как отображения M в $\mathbf{R}$ или $\mathbf{C}$). Геометрические объекты образуют категорию, и ее морфизмы служат достаточно тонкой заменой симметрий даже в тех случаях, когда этих симметрий не слишком много (как у общих римановых пространств, где можно измерять длины, углы и объемы, но движений, вообще говоря, недостаточно).

6. Линейные геометрии. Теперь мы можем охарактеризовать место линейных геометрий в этой общей картине. В известном смысле слова линейные геометрии относятся к числу непосредственных потомков геометрии Евклида. Рассматриваемые в них пространства M суть либо *линейные пространства* (теперь уже над общими полями, хотя $\mathbf{R}$ или $\mathbf{C}$ по-прежнему остаются в центре внимания, особенно ввиду многочисленных приложений), либо пространства, производные от линейных: *аффинные* («линейные пространства без отмеченного начала координат») и *проективные* («аффинные пространства, пополненные бесконечно удаленными точками»). Группы симметрий суть подгруппы линейной группы, которые сохраняют фиксированное «скалярное произведение», а также их расширения сдвигами (аффинные группы) или фактор-группы по гомотетиям (проективные группы). Рассматриваемые функции линейны или близки к линейным, иногда квадратичны. Фигуры суть линейные подпространства и многообразия (обобщения прямых на евклидовой плоскости) и квадрики (обобщения окружностей). Можно представлять себе эти обобщения евклидовой геометрии как результат чисто логического анализа, и установившийся формализм линейных геометрий действительно обладает удивительной стройностью и компактностью. Но жизнеспособность этой ветви математики в значительной мере связана с ее многообразными естественнонаучными приложениями. Понятие скалярного произведения, лежащее в основе всей второй части курса, может служить для измерения углов в абстрактных евклидовых пространствах. Но математик, который не знает, что оно же измеряет вероятности (в моделях квантовой механики), скорости (в пространстве Минковского специальной теории относительности) и коэффициенты корреляции случайных величин (в теории вероятности), лишается не только общей широты кругозора, но и гибкости чисто математической интуиции. Поэтому мы сочли необходимым включить в курс сведения и об этих интерпретациях.

§ 2. Скалярные произведения

1. Полилинейные отображения. Пусть $L_1, \ldots, L_n$, M — линейные пространства над общим полем $\mathscr{K}$. *Полилинейным отображением* (при $n=2$ *билинейным*) называется отображение

$$f\colon L_1 \times \ldots \times L_n \to M,\ (l_1, \ldots, l_n) \mapsto f(l_1, \ldots, l_n) \in M,$$

которое линейно как функция любого из аргументов $l_i \in L_i$ при фиксированных остальных $l_j \in L_j$, $j = 1, \ldots, n$, $j \neq i$. Иными словами,

$$f(l_1, \ldots, l_i + l_i', l_{i+1}, \ldots, l_n) = \\ = f(l_1, \ldots, l_i, \ldots, l_n) + f(l_1, \ldots, l_i', \ldots, l_n),$$

$$f(l_1, \ldots, al_i, l_{i+1}, \ldots, l_n) = af(l_1, \ldots, l_i, \ldots, l_n)$$

для $i = 1, \ldots, n$; $a \in \mathscr{K}$. В случае $M = \mathscr{K}$ полилинейные отображения называются также *полилинейными функциями*, или *формами*.

В первой части мы уже встречались с билинейными отображениями

$$L^* \times L \to \mathscr{K}\colon (f, l) \mapsto f(l), \quad f \in L^*, \ l \in L;$$

$$\mathscr{L}(L, M) \times L \to M\colon (f, l) \mapsto f(l), \quad f \in \mathscr{L}(L, M), \ l \in L.$$

Определитель квадратной матрицы полилинеен как функция от ее строк и столбцов. Еще один пример:

$$\mathscr{K}^n \times \mathscr{K}^m \to \mathscr{K}\colon (\vec{x}, \vec{y}) \mapsto \sum_{i,j} g_{ij} x_i y_j = \vec{x}^t G \vec{y},$$

где G — любая матрица размера $n \times m$ над $\mathscr{K}$, векторы из $\mathscr{K}^n$, $\mathscr{K}^m$ представлены столбцами своих координат.

Общие полилинейные отображения мы будем изучать позже, в части, посвященной тензорной алгебре. Здесь же мы займемся важнейшим для приложений классом билинейных функций $L \times L \to \mathscr{K}$, а также, при $\mathscr{K} = \mathbf{C}$, функций $L \times \bar{L} \to \mathbf{C}$, где $\bar{L}$ — пространство, комплексно сопряженное с L (см. ч. 1, § 12). Каждая такая функция называется также *скалярным произведением* или *метрикой*, на пространстве L, и пара (L, скалярное произведение) рассматривается как единый геометрический объект. Изучаемые в этой части метрики лишь в специальных случаях являются метриками в смысле определения п. 1 § 10 ч. 1, и читатель не должен смешивать эти омонимы.

Скалярное произведение $L \times \bar{L} \to \mathbf{C}$ чаще всего рассматривают как *полуторалинейное* отображение $g: L \times L \to \mathbf{C}$, линейное по первому аргументу и полулинейное по второму: $g(al_1, bl_2) = a\bar{b}g(l_1, l_2)$.

2. Способы задания скалярного произведения. а) Пусть $g: L \times \times L \to \mathcal{K}$ (или $L \times \bar{L} \to \mathbf{C}$) — некоторое скалярное произведение на конечномерном пространстве L. Выберем базис $\{e_1, \ldots, e_n\}$ в L и определим матрицу

$$G = (g(e_i, e_j)); \quad i, j = 1, \ldots, n.$$

Она называется *матрицей Грама* базиса $\{e_1, \ldots, e_n\}$ относительно g, а также *матрицей g в базисе* $\{e_1, \ldots, e_n\}$. Задание $\{e_i\}$ и G вполне определяет g, потому что в силу свойства билинейности

$$g(\vec{x}, \vec{y}) = g\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\right) = \sum_{i, j=1}^{n} x_i y_j g(e_i, e_j) = \vec{x}^t G \vec{y}.$$

В случае полуторалинейной формы аналогичная формула приобретает вид

$$g(\vec{x}, \vec{y}) = g\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\right) = \sum_{i, j=1}^{n} x_i \bar{y}_j g(e_i, e_j) = \vec{x}^t G \bar{\vec{y}}.$$

Наоборот, если базис $\{e_1, \ldots, e_n\}$ фиксирован, а G — любая матрица размера $n \times n$ над $\mathcal{K}$, то отображение $(\vec{x}, \vec{y}) \to \vec{x}^t G \vec{y}$ (или $\vec{x}^t G \bar{\vec{y}}$ в полуторалинейном случае) определяет скалярное произведение на L с матрицей G в этом базисе, как показывают очевидные проверки. Таким образом, наша конструкция устанавливает биекцию между скалярными произведениями (билинейными или полуторалинейными) на n-мерном пространстве с базисом и матрицами размера $n \times n$.

Выясним, как меняется G при замене базиса. Пусть A — матрица перехода к штрихованному базису. В координатах: $\vec{x} = A\vec{x}'$, где $\vec{x}$ — столбец координат вектора в старом базисе, а $\vec{x}'$ — столбец его же координат в новом. Тогда в билинейном случае

$$g(\vec{x}, \vec{y}) = \vec{x}^t G \vec{y} = (A\vec{x}')^t G (A\vec{y}') = (\vec{x}')^t A^t G A \vec{y}',$$

так что матрца Грама штрихованного базиса равна $A^t G A$. Аналогично, в полуторалинейном случае она равна $A^t G \bar{A}$.

В первой части курса матрицы служили нам в основном для записи линейных отображений, и интересно выяснить, нет ли естественного линейного отображения, связанного с g и отвечающего матрице Грама G. Оно действительно существует, и его конструкция дает равносильный способ задания скалярного произведения.

б) Пусть $g\colon L\times L\to\mathcal{K}$ — скалярное произведение. Поставим в соответствие каждому вектору $l\in L$ функцию $g_l\colon L\to\mathcal{K}$, для которой

$$g_l(m)=g(l, m),\qquad m\in L.$$

Эта функция линейна по m в билинейном случае и антилинейна в полуторалинейном, т е. $g_l\in L^*$ или соответственно $g_l\in\bar{L}^*$ для каждого l. Кроме того, отображение

$$\tilde{g}\colon L\to L^* \quad \text{или} \quad L\to\bar{L}^*\colon l\mapsto g_l=\tilde{g}(l)$$

линейно, канонически построено по g и однозначно определяет g по формуле

$$g(l, m)=(\tilde{g}(l), m),$$

где внешние скобки справа обозначают каноническое билинейное отображение $L^*\times L\to\mathcal{K}$ или $\bar{L}^*\times L\to\mathbf{C}$.

Наоборот, по любому линейному отображению $\tilde{g}\colon L\to L^*$ (или $L\to\bar{L}^*$) однозначно восстанавливается билинейное отображение $g\colon L\times L\to\mathcal{K}$ (или $g\colon L\times\bar{L}\to\mathbf{C}$) по той же формуле

$$g(l, m)=(\tilde{g}(l), m).$$

Связь с предыдущей конструкцией такова: *если в L выбран базис $\{e_1, \ldots, e_n\}$ и g задается матрицей G в этом базисе, то $\tilde{g}$ задается матрицей G^t в базисах $\{e_1, \ldots, e_n\}$, $\{e^1, \ldots, e^n\}$, двойственных друг другу.*

Действительно, если $\tilde{g}$ задано матрицей G^t, то соответствующее скалярное произведение g в двойственных базисах имеет вид

$$g(\vec{x}, \vec{y})=(\tilde{g}(\vec{x}), \vec{y})=(\tilde{g}(\vec{x}))^t\vec{y}\,(\text{или }(\tilde{g}(\vec{x}))^t\bar{\vec{y}})=$$
$$=(G^t\vec{x})^t\vec{y}\,(\text{или }(G^t\vec{x})^t\bar{\vec{y}})=\vec{x}^tG\vec{y}\,(\text{или }\vec{x}^tG\bar{\vec{y}}),$$

что доказывает требуемое. Здесь мы пользовались замечанием, сделанным в § 7 ч. 1, о том, что каноническое отображение $L^*\times\times L\to\mathcal{K}$ в двойственных базисах определяется формулой $(\vec{x}, \vec{y})==x^t\vec{y}$.

3. Свойства симметрии скалярных произведений. Перестановка аргументов в билинейном скалярном произведении g определяет новое скалярное произведение g^t:

$$g^t(l, m)=g(m, l).$$

В полуторалинейном случае эта операция также меняет места «линейного» и «полулинейного» аргументов; если мы хотим, чтобы этого не произошло, то удобнее рассматривать $\bar{g}^t$:

$$\bar{g}^t(l, m)=\overline{g(m, l)}.$$

У $\bar{g}^t$ линейный аргумент будет на первом месте, если у g он был на первом месте, а полулинейный — соответственно на втором. Операция $g \mapsto g^t$ или $g \mapsto \bar{g}^t$ легко описывается на языке матриц Грама: она отвечает операции $G \mapsto G^t$ или $G \mapsto \bar{G}^t$ соответственно (предполагается, что g, g^t, $\bar{g}^t$ пишутся в одном и том же базисе L). Действительно,

$$g^t(\vec{x}, \vec{y}) = g(\vec{y}, \vec{x}) = \vec{y}^t G \vec{x} = (\vec{y}^t G \vec{x})^t = \vec{x}^t G^t \vec{y},$$

$$\bar{g}^t(\vec{x}, \vec{y}) = \overline{g(\vec{y}, \vec{x})} = \overline{\vec{y}^t G \vec{x}} = (\bar{\vec{y}}^t \bar{G} \vec{x})^t = \vec{x}^t \bar{G}^t \bar{\vec{y}}.$$

Мы будем заниматься почти исключительно скалярными произведениями, которые удовлетворяют одному из следующих условий симметрии относительно этой операции:

а) $g^t = g$. Такие скалярные произведения называются *симметричными*, а геометрия пространств с симметричным скалярным произведением называется *ортогональной геометрией*. Симметричные скалярные произведения задаются симметричными матрицами Грама G.

б) $g^t = -g$. Такие скалярные произведения называются *антисимметричными*, или *симплектическими*, а соответствующие геометрии называются *симплектическими*. Им отвечают антисимметричные матрицы Грама.

Полуторалинейный случай:

в) $\bar{g}^t = g$. Такие скалярные произведения называются *эрмитово симметричными*, или просто *эрмитовыми*, а соответствующие геометрии — *эрмитовыми*. Им отвечают эрмитовы матрицы Грама. Из условия $\bar{g}^t = g$ следует, что $\bar{g}(l, l) = g(l, l)$ для всех $l \in L$, т. е. все значения $g(l, l)$ вещественны.

Эрмитово антисимметричные скалярные произведения обычно не рассматриваются специально, ибо отображение $g \mapsto ig$ устанавливает биекцию между ними и эрмитово симметричными скалярными произведениями:

$$\bar{g}^t = g \Leftrightarrow \overline{(ig)^t} = -ig.$$

Геометрические свойства скалярных произведений, отличающихся друг от друга лишь множителем, практически одни и те же. Напротив, ортогональная геометрия во многом отличается от симплектической: редукция соотношений $g^t = g$ и $g^t = -g$ друг к другу таким простым способом невозможна.

4. Ортогональность. Пусть (L, g) — векторное пространство со скалярным произведением. Векторы $l_1, l_2 \in L$ называются *ортогональными* (относительно g), если $g(l_1, l_2) = 0$. Подпространства $L_1, L_2 \subset L$ называются ортогональными, если $g(l_1, l_2) = 0$ для всех $l_1 \in L_1$, $l_2 \in L_2$. Основная причина, по которой важны лишь скалярные произведения с одним из свойств симметрии предыдущего пункта, состоит в том, что для них *свойство ортогональности векторов или подпространств симметрично относительно этих векто-*

ров или подпространств. Действительно: если $g^t = \pm g$ или $g^t = \bar{g}$, то

$$g(l, m) = 0 \Leftrightarrow \pm g^t(l, m) = 0 \Leftrightarrow g(m, l) = 0$$

и аналогично в эрмитовом случае (по поводу обратного утверждения см. упражнение 3).

Если не оговорено обратное, в дальнейшем мы будем рассматривать только ортогональные, симплектические или эрмитовы скалярные произведения. Первое применение понятия ортогональности содержится в следующем определении.

5. Определение. а) *Ядром скалярного произведения g на пространстве L называется множество всех векторов $l \in L$, ортогональных ко всем векторам L.*

б) *g называется невырожденным, если ядро формы g тривиально, т. е. состоит только из нуля.*

Очевидно, ядро формы g совпадает с ядром линейного отображения

$$\tilde{g}\colon L \to L^* \text{ (или } L \to \overline{L}^*)$$

и потому является линейным подпространством в L. Поэтому задание *невырожденной* формы g можно заменить заданием *изоморфизма $L \to L^*$* (или $\overline{L}^*$). Так как матрицей $\tilde{g}$ служит транспонированная матрица Грама G^t базиса L, *невырожденность g равносильна невырожденности матрицы Грама* (любого базиса). В тензорной алгебре и ее приложениях к дифференциальной геометрии и физике очень широко используется то обстоятельство, что невырожденная ортогональная форма g определяет изоморфизм $L \to L^*$: оно служит основой техники «поднятия и опускания индексов».

Ранг g определяется как размерность образа $\tilde{g}$, или как ранг матрицы Грама G.

6. Задача классификации. Пусть (L_1, g_1), (L_2, g_2) — два линейных пространства со скалярными произведениями над полем $\mathcal{K}$. Назовем их *изометрией* любой линейный изоморфизм $f\colon L_1 \xrightarrow{\sim} L_2$, который сохраняет значения всех скалярных произведений, т. е.

$$g_1(l, l') = g_2(f(l), f(l')) \quad \text{для всех} \quad l, l' \in L_1.$$

Назовем такие пространства *изометричными*, если между ними существует изометрия. Очевидно, тождественное отображение является изометрией, композиция изометрий есть изометрия и линейное отображение, обратное к изометрии, есть изометрия. В следующем параграфе мы решим задачу классификации пространств с точностью до изометрии, а затем изучим группы изометрий пространства с самим собой и покажем, что среди них содержатся классические группы, описанные в § 4 ч. 1.

Классическое решение задачи классификации состоит в том, что всякое пространство со скалярным произведением разлагается в прямую сумму попарно ортогональных подпространств малой размерности (один в ортогональном и эрмитовом случае, один или два — в симплектическом). Поэтому мы закончим этот параграф непосредственным описанием таких маломерных пространств с метрикой.

7. Одномерные ортогональные пространства. Пусть $\dim L = 1$, g — ортогональное скалярное произведение на L. Возьмем любой ненулевой вектор $l \in L$. Если $g(l, l) = 0$, то $g \equiv 0$, так что g вырожденное и нулевое. Если $g(l, l) = a \neq 0$, то для любого $x \in \mathcal{K}$ значение $g(xl, xl)$ равно ax^2, так что все значения $g(l, l)$ на ненулевых векторах в L составляют в мультипликативной группе $\mathcal{K}^* = \mathcal{K} \setminus \{0\}$ поля $\mathcal{K}$ смежный класс по подгруппе, состоящей из квадратов: $\{ax^2 \mid x \in \mathcal{K}^*\} \in \mathcal{K}^*/(\mathcal{K}^*)^2$. *Этот смежный класс полностью характеризует невырожденное симметричное скалярное произведение на одномерном пространстве* L: для (L_1, g_1) и (L_2, g_2) два таких класса совпадают тогда и только тогда, когда эти пространства изометричны. В самом деле, если $g_1(l_1, l_1) = ax^2$, $g_2(l_2, l_2) = ay^2$, где $l_i \in L_i$, то отображение f: $l_1 - y^{-1}xl_2$ определяет изометрию L_1 с L_2, что доказывает достаточиость. Необходимость очевидна.

Так как $\mathbf{R}^*/(\mathbf{R}^*)^2 = \{\pm 1\}$ и $\mathbf{C}^* = (\mathbf{C}^*)^2$, мы получаем следующие важные частные случаи классификации.

Над $\mathbf{R}$ *любое одномерное ортогональное пространство изометрично одномерному координатному пространству с одним из трех скалярных произведений*: xy, $-xy$, 0.

Над $\mathbf{C}$ *любое одномерное ортогональное пространство изометрично одномерному координатному пространству с одним из двух скалярных произведений*: xy, 0.

8. Одномерные эрмитовы пространства. Здесь рассуждения аналогичны. Основное поле равно $\mathbf{C}$; вырожденность формы равносильна ее обращению в нуль. Если же форма невырождена, то множество значений $g(l, l)$ для ненулевых векторов $l \in L$ есть смежный класс подгруппы $\mathbf{R}^*_+ = \{x \in R^* \mid x > 0\}$ в группе $\mathbf{C}^*$, ибо $g(al, al) = a\bar{a}g(l, l) = |a|^2 g(l, l)$, и $|a|^2$ пробегает все значения в $\mathbf{R}^*_+$, когда $a \in \mathbf{C}^*$. Но каждое ненулевое комплексное число z однозначно представляется в виде $re^{i\varphi}$, где $r \in R^*_+$, а $e^{i\varphi}$ лежит на единичной комплексной окружности, которую мы обозначим

$$\mathbf{C}^*_1 = \{z \in \mathbf{C}^* \mid |z| = 1\}.$$

На групповом языке это определяет прямое разложение $\mathbf{C}^* = \mathbf{R}^*_+ \times \times \mathbf{C}^*_1$ и изоморфизм $\mathbf{C}^*/\mathbf{R}^*_+ \to \mathbf{C}^*_1$. Таким образом, невырожденные полуторалинейные формы классифицируются комплексными числами, по модулю равными единице. Однако мы еще не полностью учли свойства эрмитовости, которое означает, что $g(l, l) = \overline{g(l, l)}$, т. е. что значения $g(l, l)$ все вещественны. Поэтому эрмитовым формам отвечают только числа ± 1 в $\mathbf{C}^*_1$, как и в ортогональном случае над $\mathbf{R}$. Окончательный ответ:

Над $\mathbf{C}$ *любое одномерное эрмитово пространство изометрично одномерному координатному пространству с одним из трех скалярных произведений*: $x\bar{y}$, $-x\bar{y}$, 0.

Одномерные ортогональные пространства над **R** (или эрмитовы над **C**) со скалярными произведениями xy, $-xy$, 0 (или $x\bar{y}$, $-x\bar{y}$, 0) в подходящем базисе мы будем называть соответственно *положительными*, *отрицательными* и *нулевыми*. Скалярные произведения ненулевых векторов на себя в них принимают соответственно только положительные, только отрицательные или только нулевые значения.

9. Одномерные симплектические пространства. Здесь мы встречаемся с новой ситуацией: любая антисимметричная форма на одномерном пространстве над полем характеристики $\neq 2$ тождественно равна нулю, в частности, вырождена! Действительно,

$$g(l, l) = -g(l, l) \Rightarrow 2g(l, l) = 0,$$
$$g(al, bl) = abg(l, l) = 0.$$

Что касается характеристики 2, то условие антисимметрии $g(l, m) = -g(m, l)$ в этом случае равносильно условию симметрии $g(l, m) = g(m, l)$, так что над такими полями симплектическая геометрия не отличается от ортогональной. Впрочем, у ортогональной геометрии также появляются свои особенности, и мы обычно будем этот случай исключать из рассмотрения.

Ясно поэтому, что одномерные симплектические пространства не могут быть строительным материалом для конструкции общих симплектических пространств, и нужно пойти по крайней мере на шаг дальше.

10. Двумерные симплектические пространства. Пусть (L, g) — двумерное пространство с кососимметрической формой g над полем $\mathcal{K}$ характеристики $\neq 2$. Если форма g вырождена, то она автоматически нулевая. В самом деле, пусть $l \neq 0$ — такой вектор, что $g(l, m) = 0$, для всех $m \in L$. Дополним l до базиса $\{l, l'\}$ в L и учтем, что $g(l', l') = g(l, l) = 0$ по предыдущему пункту. Тогда для любых a, b, a', $b' \in \mathcal{K}$ имеем

$$g(al + a'l', bl + b'l') = \\ = abg(l, l) + ab'g(l, l') - a'bg(l, l') + bb'g(l', l') = 0.$$

Пусть теперь g ненулевая и, значит, невырожденная. Тогда существует пара векторов e_1, e_2 с $g(e_1, e_2) = a \neq 0$ и даже с $a = 1$: $g(a^{-1}e_1, e_2) = a^{-1}a = 1$.

Пусть $g(e_1, e_2) = 1$. Тогда векторы e_1, e_2 линейно независимы и, значит, образуют базис L: если, скажем, $e_1 = ae_2$, то $g(ae_2, e_2) = ag(e_2, e_2) = 0$. В координатах относительно такого базиса скалярное произведение g записывается в виде

$$g(x_1e_1 + x_2e_2, y_1e_1 + y_2e_2) = x_1y_2 - x_2y_1$$

и имеет матрицу Грама

$$G = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Окончательно, получаем:

Над полем $\mathscr{K}$ характеристики $\neq 2$ любое двумерное симплектическое пространство изометрично координатному пространству $\mathscr{K}^2$ со скалярным произведением $x_1y_2 - x_2y_1$ или 0.

УПРАЖНЕНИЯ

1. Пусть L, M — конечномерные линейные пространства над полем $\mathscr{K}$ и g: $L \times M \to \mathscr{K}$ — билинейное отображение. Назовем левым ядром g множество $L_0 = \{l \in L \mid g(l, m) = 0$ для всех $m \in M\}$, правым ядром g множество $M_0 = \{m \in M \mid g(l, m) = 0$ для всех $l \in L\}$. Докажите следующие утверждения:

а) $\dim L/L_0 = \dim M/M_0$.

б) g индуцирует билинейное отображение g': $L/L_0 \times M/M_0 \to \mathscr{K}$, $g'(l + L_0, m + M_0) = g(l, m)$, у которого левое и правое ядра нулевые.

2. Докажите, что всякое билинейное скалярное произведение g: $L \times L \to \mathscr{K}$ (над полем $\mathscr{K}$ характеристики $\neq 2$) однозначно разлагается в сумму симметричного и антисимметричного скалярного произведения.

3. Пусть g: $L \times L \to \mathscr{K}$ — такое билинейное скалярное произведение, что свойство ортогональности пары векторов симметрично: из $g(l_1, l_2) = 0$ следует, что $g(l_2, l_1) = 0$. Докажите, что тогда g либо симметрично, либо антисимметрично. (*Указания.* а) Пусть l, m, $n \in L$. Докажите, что $g(l, g(l, n)m - g(l, m)n) = 0$. Пользуясь симметрией ортогональности, выведите отсюда, что $g(l, n)g(m, l) = g(n, l)g(l, m)$. б) Положив $n = l$, выведите отсюда, что если $g(l, m) \neq g(m, l)$, то $g(l, l) = 0$. в) Покажите, что $g(n, n) = 0$ для любого вектора $n \in L$, если g несимметрично. С этой целью выберите l, m с $g(l, m) \neq g(m, l)$ и разберите отдельно случаи $g(l, n) \neq g(n, l)$, $g(l, n) = g(n, l)$. г) Покажите, что если $g(n, n) = 0$ для всех $n \in L$, то g антисимметрично.)

4. Дайте классификацию одномерных ортогональных пространств над конечным полем $\mathscr{K}$ характеристики $\neq 2$, показав, что $\mathscr{K}^*/(\mathscr{K}^*)^2$ есть циклическая группа второго порядка. (*Указание*: Покажите, что ядро гомоморфизма $\mathscr{K}^* \to \mathscr{K}^*$: $x \mapsto x^2$ имеет порядок 2, пользуясь тем, что любой многочлен над полем имеет не больше корней, чем его степень.)

5. Пусть (L, g) — линейное пространство размерности n с невырожденным скалярным произведением. Докажите, что семейство векторов $\{e_1, \ldots, e_n\}$ в L линейно независимо тогда и только тогда, когда матрица $(g(e_i, e_j))$ невырождена.

§ 3. Теоремы классификации

1. Основная цель этого параграфа — дать классификацию конечномерных ортогональных, эрмитовых и симплектических пространств с точностью до изометрии. Пусть (L, g) — такое пространство, $L_0 \subset L$ — его подпространство. Ограничение g на L_0 является скалярным произведением на L_0. Назовем L_0 *невырожденным*, если ограничение g на L_0 невырождено, и *изотропным*, если ограничение g на L_0 равно нулю. Существенно, что если даже L невырождено, ограничения g на нетривиальные подпространства могут быть вырожденными или нулевыми. Например, в симплектическом случае вырождены все одномерные подпространства, а в ортогональном пространстве $\mathbf{R}^2$ с произведением $x_1y_1 - x_2y_2$ вырождено подпространство, натянутое на вектор $(1, 1)$.

Ортогональным дополнением к подпространству $L_0 \subset L$ называется множество

$$L_0^{\perp} = \{l \in L \mid g(l_0, l) = 0 \quad \text{для всех} \quad l_0 \in L_0\}$$

(не путать с введенным в первой части ортогональным дополнением к L_0, лежащим в L^*, здесь мы им пользоваться не будем). Легко видеть, что $L_0^\perp$ является линейным подпространством в L.

2. Предложение. *Пусть* (L, g) *конечномерно.*

а) *Если подпространство* $L_0 \subset L$ *невырождено, то* $L = L_0 \oplus L_0^\perp$.

б) *Если оба подпространства* L_0 *и* $L_0^\perp$ *невырождены, то* $(L_0^\perp)^\perp = L_0$.

Доказательство. а) Пусть $\tilde{g}$: $L \to L^*$ (или $\bar{L}^*$) — отображение, ассоциированное с g, как в предыдущем параграфе. Обозначим через $\tilde{g}_0$ его ограничение на L_0, $\tilde{g}_0$: $L_0 \to L^*$ (или $\bar{L}^*$). Если L_0 невырождено, то $\operatorname{Ker} \tilde{g}_0 = 0$: иначе в L_0 есть вектор, ортогональный ко всему L и тем более к L_0. Поэтому $\dim \operatorname{Im} \tilde{g}_0 = \dim L_0$. Это означает, что когда l_0 пробегает L_0, линейные формы $g(l_0, \cdot)$ от второго аргумента из L или $\bar{L}$ пробегают $\dim L_0$-мерное пространство линейных форм на L или $\bar{L}$. Так как $L_0^\perp$ есть пересечение ядер этих форм, $\dim L_0^\perp = \dim L - \dim L_0$, т. е.

$$\dim L_0 + \dim L_0^\perp = \dim L.$$

С другой стороны, из невырожденности L_0 следует, что $L_0 \cap L_0^\perp = \{0\}$, ибо $L_0 \cap L_0^\perp$ есть ядро ограничения g на L_0. Поэтому сумма $L + L_0^\perp$ прямая; но ее размерность равна $\dim L$, так что $L_0 \oplus L_0^\perp = L$.

б) Из определений ясно, что $L_0 \subset (L_0^\perp)^\perp$. С другой стороны, если L_0, $L_0^\perp$ невырождены, то по предыдущему

$$\dim (L_0^\perp)^\perp = \dim L - \dim L_0^\perp = \dim L_0.$$

Это завершает доказательство.

3. Теорема. *Пусть* (L, g) — *конечномерное ортогональное (над полем характеристики* $\neq 2$*), эрмитово или симплектическое пространство. Тогда существует разложение* L *в прямую сумму попарно ортогональных подпространств:*

$$L = L_1 \oplus \ldots \oplus L_m,$$

одномерных в ортогональном и эрмитовом случае и одномерных вырожденных или двумерных невырожденных в симплектическом случае.

Доказательство. Проведем индукцию по размерности L. Случай $\dim L = 1$ тривиален; пусть $\dim L \geqslant 2$. Если g нулевая, доказывать нечего. Если g ненулевая, то в симплектическом случае имеется пара векторов $l_1, l_2 \in L$ с $g(l_1, l_2) \neq 0$. Согласно п. 10 предыдущего параграфа, натянутое на них подпространство L_0 невырождено. По предложению п. 2 $L = L_0 \oplus L_0^\perp$, и по индуктивному предположению мы можем далее разложить $L_0^\perp$, как сформулировано в теореме. Это даст требуемое разложение L.

В ортогональном и эрмитовом случае мы покажем, что из нетривиальности g следует существование невырожденного одномерного подпространства L_0; проверив это, мы сможем положить

$L = L_0 \oplus L_0^\perp$ и применить прежнее рассуждение, т. е. индукцию по размерности L.

В самом деле, допустим, что $g(l, l) = 0$ для всех $l \in L$, и покажем, что тогда $g \equiv 0$. Действительно, для всех $l_1, l_2 \in L$ имеем

$$0 = g(l_1 + l_2, l_1 + l_2) = g(l_1, l_1) + 2g(l_1, l_2) + g(l_2, l_2) = 2g(l_1, l_2)$$

или

$$0 = g(l_1 + l_2, l_1 + l_2) = g(l_1, l_1) + 2\operatorname{Re} g(l_1, l_2) + g(l_2, l_2) = 2\operatorname{Re} g(l_1, l_2).$$

В ортогональном случае отсюда сразу следует, что $g(l_1, l_2) = 0$. В эрмитовом мы получаем лишь, что $\operatorname{Re} g(l_1, l_2) = 0$, т. е. $g(l_1, l_2) = ia$, $a \in \mathbf{R}$. Но если $a \neq 0$, то также

$$0 = \operatorname{Re} g\left((ia)^{-1} l_1, l_2\right) = \operatorname{Re}(ia)^{-1} g(l_1, l_2) = 1.$$

Получаем противоречие.

Это завершает доказательство.

Перейдем теперь к проблеме единственности. Само по себе разложение в ортогональную прямую сумму, существование которого утверждается в теореме п. 3, далеко не единственно, кроме тривиальных случаев размерности 1 (или 2 в симплектическом случае). Над общими полями в случае ортогональной геометрии неоднозначно определяется также и набор инвариантов $a_i \in \mathcal{K}^*/\mathcal{K}^{*2}$, который характеризует ограничения g на одномерные подпространства L_i. Точный ответ на вопрос о классификации ортогональных пространств существенно зависит от свойств основного поля, и для $\mathcal{K} = \mathbf{Q}$, например, связан с такими довольно тонкими теоретико-числовыми фактами, как квадратичный закон взаимности. Поэтому в ортогональном случае мы ограничимся описанием результата для $\mathcal{K} = \mathbf{R}$ и $\mathbf{C}$ (дальнейшие подробности см. в § 14).

4. Инварианты пространств с метрикой. Пусть (L, g) — пространство со скалярным произведением. Положим $n = \dim L$, $r_0 = \dim L_0$, где L_0 — ядро формы g. Кроме того, введем два дополнительных инварианта, относящихся только к ортогональной для $\mathcal{K} = \mathbf{R}$ и эрмитовой геометриям: r_+ и r_-, числа положительных и отрицательных одномерных подпространств L_i в некотором ортогональном разложении L в прямую сумму, как в теореме п. 3.

Очевидно, $r_0 \leqslant n$ и $n = r_0 + r_+ + r_-$ для эрмитовой и ортогональной геометрии над $\mathbf{R}$. Набор (r_0, r_+, r_-) называется *сигнатурой* пространства. При $r_0 = 0$ сигнатурой называют иногда также (r_+, r_-) или $r_+ - r_-$ (считая $n = r_+ + r_-$ известным).

Теперь мы можем сформулировать теорему единственности.

5. Теорема. а) *Симплектические пространства над произвольным полем, а также ортогональные пространства над* $\mathbf{C}$ *с точностью до изометрии определяются двумя целыми числами* n, r_0, *т. е. размерностями пространства и ядра скалярного произведения.*

б) *Ортогональные пространства над* $\mathbf{R}$ *и эрмитовы пространства над* $\mathbf{C}$ *с точностью до изометрии определяются сигнатурой*

(r_0, r_+, r_-), *которая не зависит от выбора ортогонального разложения (это утверждение называется теоремой инерции).*

Доказательство. а) Пусть (L, g) — симплектическое пространство или ортогональное пространство над $\mathbf{C}$. Рассмотрим его прямое разложение $L = \bigoplus_{i=1}^{n} L_i$, как в теореме п. 3, и покажем, что r_0 совпадает с числом одномерных пространств в этом разложении, вырожденных для g. На самом деле сумма этих пространств L_0 совпадает с ядром g. Действительно, очевидно, что она содержится в этом ядре, ибо элементы L_0 ортогональны как к L_0, так и к остальным слагаемым. С другой стороны, если $L_0 = \bigoplus_{i=1}^{r_0} L_i$ и

$$l = \sum_{j=1}^{n} l_j, \quad l_j \in L_j, \ \exists j > r_0, \ l_j \neq 0,$$

то

$$g(l, l_j) = g(l_j, l_j) \neq 0$$

в ортогональном случае, и существует вектор $l_j' \in L_j$ с

$$g(l, l_j') = g(l_j, l_j') \neq 0$$

в симплектическом случае, ибо иначе ядро ограничения g на L_j было бы нетривиально, и g на L_j была бы нулевой по п. 10 § 2, вопреки тому, что $j > r_0$. Поэтому $l \notin$ (ядро g), и $L_0 =$ (ядро g). Если теперь (L, g) и (L', g') — два таких пространства с одинаковыми n и r_0, то, построив их ортогнальные прямые разложения $L = \bigoplus_{i=1}^{n} L_i$ и $L' = \bigoplus_{i=1}^{n} L_i'$, для которых (ядро g) $= \bigoplus_{i=1}^{r_0} L_i$ и (ядро g') $= \bigoplus_{i=1}^{r_0} L_i'$, мы можем определить изометрию (L, g) с (L, g') как прямую сумму изометрий $\bigoplus f_i$, $f_i \colon L_i \to L_i'$, которые существуют в силу результатов пп. 7 и 10 § 2.

б) Пусть теперь (L, g) и (L', g') — пара ортогональных пространств над $\mathbf{R}$ или эрмитовых над $\mathbf{C}$ с сигнатурами (r_0, r_+, r_-) и (r_0', r_+', r_-'), определенными с помощью некоторых ортогональных разложений $L = \bigoplus L_i$, $L' = \bigoplus L_i'$, как в теореме п. 3. Предположим, что между ними существует изометрия. Тогда прежде всего $\dim L = \dim L'$, так что $r_0 + r_+ + r_- = r_0' + r_+' + r_-'$. Далее, точно так же, как в предыдущем пункте, проверяется, что r_0 совпадает с размерностью ядра g, а r_0' — с размерностью ядра g', а эти ядра суть суммы нулевых пространств L_i и L_i' в соответствующих разложениях. Поскольку изометрия определяет линейный изоморфизм между ядрами, имеем $r_0 = r_0'$ и $r_+ + r_- = r_+' + r_-'$.

Остается проверить, что $r_+ = r_+'$, $r_- = r_-'$. Положим $L = L_0 \oplus L_+ \oplus L_-$, $L' = L_0' \oplus L_+' \oplus L_-'$, где L_0, L_+, L_- — суммы нулевых, положительных и отрицательных подпространств исходного разложения L, и соответственно для L'. Предположим, что $r_+ = \dim L_+ > r'_+ = \dim L'_+$, и придем к противоречию; возможность $r_+ < r'_+$ разбирается аналогично. Ограничим изометрию $f\colon L \to L'$ на $L_+ \subset L$. Каждый вектор $f(l)$ однозначно представляется в виде суммы

$$f(l) = f(l)_0 + f(l)_+ + f(l)_-,$$

где $f(l)_+ \in L'_+$ и т. п. Отображение $L_+ \to L'_+$, $l \mapsto f(l)_+$ линейно. Так как по предположению $\dim L_+ > \dim L'_+$, существует ненулевой вектор $l \in L_+$, для которого $f(l)_+ = 0$, так что

$$f(l) = f(l)_0 + f(l)_-.$$

Но $g(l, l) > 0$, потому что $l \in L_+$ и L_+ есть ортогональная прямая сумма положительных одномерных пространств. Так как f — изометрия, мы должны иметь также $g'(f(l), f(l)) > 0$. С другой стороны,

$$g'(f(l), f(l)) = g'(f(l)_0 + f(l)_-, \ f(l)_0 + f(l)_-) = g'(f(l)_-, f(l)_-) \leqslant 0.$$

Это противоречие завершает доказательство того, что у изометричных пространств сигнатуры, вычисленные по любым ортогональным разложениям, одинаковы.

Наоборот, если (L, g), (L', g') — два пространства с одинаковыми сигнатурами, то между подпространствами из их ортогональных разложений $L = \bigoplus L_i$ и $L' = \bigoplus L'_i$ можно установить взаимно однозначное соответствие $L_i \leftrightarrow L'_i$, сохраняющее знак ограничения g на L_i и g' на L'_i соответственно. По результатам пп. 7 и 8 § 2 существуют изометрии $f_i\colon L_i \to L'_i$, и их прямая сумма $\bigoplus f_i$ будет изометрией между L и L'.

Теперь мы выведем несколько следствий и переформулировок теорем пп. 3 и 5, которые подчеркивают разные аспекты ситуации.

6. Базисы. Пусть (L, g) — пространство со скалярным произведением. Базис $\{e_1, \ldots, e_n\}$ в L называется *ортогональным*, если $g(e_i, e_j) = 0$ для всех $i \neq j$. Из теоремы п. 5 следует, что *у любого ортогонального или эрмитова пространства имеется ортогональный базис*. Действительно, достаточно построить разложение $L = \bigoplus L_i$ на ортогональные одномерные подпространства и затем выбрать $e_i \in L_i$, $e_i \neq 0$.

Ортогональный базис $\{e_i\}$ называется *ортонормированным*, если $g(e_i, e_i) = 0$ или ± 1 для всех i. Обсуждение в конце § 2 показывает, что *у любого ортогонального пространства над* $\mathbf{R}$ *или* $\mathbf{C}$ *и у любого эрмитова пространства имеется ортонормированный базис*. Теорема п. 5 показывает, что числа элементов e ортонормированного базиса с $g(e, e) = 0$, 1 или -1 не зависят от базиса для $\mathcal{K} = \mathbf{R}$ (ортогональный случай) и $\mathcal{K} = \mathbf{C}$ (эрмитов случай). В ортогональном случае над $\mathbf{C}$ всегда можно добиться того, что

$g(e_i, e_i)=0$ или 1, и количества таких векторов в базисе не зависят от самого базиса. Матрица Грама ортонормированного базиса имеет вид

$$\left(\begin{array}{c|c|c} E_{r_+} & 0 & 0 \\ \hline 0 & -E_{r_-} & 0 \\ \hline 0 & 0 & 0 \end{array}\right)$$

(при надлежащем упорядочении). Чаще всего понятие ортонормированного базиса применяют в невырожденном случае, когда векторов e_i с $g(e_i, e_i)=0$ нет. Следующая простая, но важная формула позволяет явно написать коэффициенты разложения любого вектора $e \in L$ по ортогональному базису (в невырожденном случае):

$$e=\sum_{i=1}^{n} \frac{g(e, e_i)}{g(e_i, e_i)} e_i.$$

Действительно, скалярные произведения левой и правой части со всеми e_i совпадают, а из невырожденности следует, что если $g(e, e_i)=g(e', e_i)$ для всех i, то $e=e'$, ибо $e-e'$ лежит в ядре формы g.

В симплектическом пространстве ортогональный базис, очевидно, может существовать, только если $g=0$. Теорема п. 3 обеспечивает, однако, существование *симплектического базиса* $\{e_1, e_2, \ldots \ldots, e_r; e_{r+1}, \ldots, e_{2r}; e_{2r+1}, \ldots, e_n\}$, который характеризуется тем, что

$$g(e_i, e_{r+i})=-g(e_{r+i}, e_i)=1, \qquad i=1, \ldots, r,$$

а все остальные попарные скалярные произведения равны нулю. Действительно, следует разложить L в ортогональную прямую сумму двумерных невырожденных подпространств L_i, $1 \leqslant i \leqslant r$, и одномерных вырожденных L_j, $2r+1 \leqslant j \leqslant n$, и в качестве $\{e_i, e_{r+i}\}$ для $1 \leqslant i \leqslant r$ взять базис L_i, построенный в п. 10 § 2, а в качестве e_j для $2r+1 \leqslant j \leqslant n$ взять любой ненулевой вектор из L_j.

Матрица Грама симплектического базиса имеет вид

$$\left(\begin{array}{c|c|c} 0 & E_r & 0 \\ \hline -E_r & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right)$$

Ранг симплектической формы, по теореме п. 5, равен $2r$. В частности, *невырожденное симплектическое пространство обязательно четномерно.*

Пусть L — невырожденное симплектическое пространство, $\{e_1, \ldots, e_r, e_{r+1}, \ldots, e_{2r}\}$ — симплектический базис в нем. Пусть L_1 — линейная оболочка $\{e_1, \ldots, e_r\}$; L_2 — линейная оболочка $\{e_{r+1}, \ldots, e_{2r}\}$. Очевидно, пространства L_1 и L_2 изотропны, имеют половинную размерность и $L = L_1 \oplus L_2$. Каноническое отображение $\tilde{g}\colon L \to L^*$ определяет отображение

$$\tilde{g}_1 : L_2 \to L_1^*; \quad \tilde{g}_1(l_2)(l_1) = g(l_2, l_1).$$

Это отображение является изоморфизмом, ибо $\dim L_2 = \dim L_1 = \dim L_1^*$ и $\operatorname{Ker} \tilde{g}_1 = 0$: вектор из $\operatorname{Ker} \tilde{g}_1$ ортогонален к L_2, ибо L_2 изотропно, и к L_1 по определению, а L невырождено.

Отсюда следует, что *любое невырожденное симплектическое пространство изометрично пространству вида* $L = L_1^* \oplus L_1$ *с симплектической формой*

$$g((f, l), (f', l')) = f(l') - f'(l); \; f, f' \in L_1^*, \quad l, l' \in L_1.$$

Дальнейшие подробности см. в § 12.

7. Матрицы. Описывая скалярные произведения их матрицами Грама и переходя от случайного базиса к ортогональному или симплектическому, мы в силу результатов п. 2 § 2 и п. 6 этого параграфа получаем следующие факты:

а) Всякую квадратную симметричную матрицу G над полем $\mathcal{K}$ можно привести к диагональному виду преобразованием $G \mapsto A^t G A$, где A невырождена. При $\mathcal{K} = \mathbf{R}$ можно добиться, чтобы на диагонали стояли только $0, \pm 1$, а при $\mathcal{K} = \mathbf{C}$ — только $0, 1$; количества 0 и ± 1 (соответственно 0 и 1) будут зависеть лишь от G, но не от A.

б) Всякую квадратную антисимметричную матрицу G над полем $\mathcal{K}$ характеристики $\neq 2$ можно привести преобразованием $G \mapsto A^t G A$, где A невырождена, к виду

$$\left(\begin{array}{c|c|c} 0 & E_r & 0 \\ \hline -E_r & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

Число $2r$ равно рангу G.

в) Всякую эрмитову матрицу G над $\mathbf{C}$ можно привести к диагональному виду с числами $0, \pm 1$ на диагонали преобразованием $G \mapsto A^t G \bar{A}$, где A невырождена. Количества 0 и ± 1 зависят лишь от G.

8. Билинейные формы. Если векторы пространства (L, g) с фиксированным базисом записываются координатами в этом базисе, то выражение g через координаты является билинейной формой от $2n$ переменных, $n = \dim L$:

$$g = (x_1, \ldots, x_n; \; y_1, \ldots, y_n) = \sum_{i, j=1}^{n} g_{ij} x_i y_j = \vec{x}^t G \vec{y},$$

где G — матрица Грама базиса. Замена базиса сводится к линейному преобразованию переменных $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$ с помощью одной и той же невырожденной матрицы A в билинейном случае (или матрицы A для $\vec{x}$, $\bar{A}$ для $\vec{y}$ в полуторалинейном случае). Предыдущие результаты означают, что в зависимости от свойств симметрии матрицы G форму можно привести таким преобразованием к одному из следующих видов, называемых каноническими.

Ортогональный случай над любым полем:

$$g\left(\vec{x}, \vec{y}\right)=\sum_{i=1}^{n} a_i x_i y_i;$$

над полем $\mathbf{R}$ можно добиться того, чтобы $a_i=0, \pm 1$; над полем $\mathbf{C}$ — чтобы $a_i=0$ или 1.

Эрмитов случай (форма полуторалинейная):

$$g\left(\vec{x}, \vec{y}\right)=\sum_{i=1}^{n} a_i x_i \bar{y}_i;$$

$a_i=0$ или 1.

Симплектический случай: $n=2r+r_0$, и форма имеет вид

$$g\left(\vec{x}, \vec{y}\right)=\sum_{i=1}^{r} (x_i y_{r+i}-y_i x_{r+i}).$$

9. Квадратичные формы. *Квадратичной формой* q на пространстве L называется такое отображение $q\colon L\to\mathscr{K}$, для которого существует билинейная форма $h\colon L\times L\to\mathscr{K}$ со свойством

$$q(l)=h(l, l) \text{ для всех } l\in L.$$

Покажем, что если характеристика поля $\mathscr{K}$ не равна 2, то для всякой квадратичной формы q существует *единственная симметричная* билинейная форма g со свойством $q(l)=g(l, l)$, называемая *поляризацией* q.

Для доказательства существования положим $q(l)=h(l, l)$, где h — исходная билинейная форма, и

$$g(l, m)=\frac{1}{2}[h(l, m)+h(m, l)].$$

Очевидно, g симметрична, т. е. $g(l, m)=g(m, l)$. Кроме того,

$$g(l, l)=\frac{1}{2}[h(l, l)+h(l, l)]=q(l).$$

Билинейность g сразу же следует из билинейности h.

Для доказательства единственности заметим, что если $q(l)= g_1(l, l)=g_2(l, l)$, где g_1, g_2 симметричны и билинейны, то форма $g=g_1-g_2$ тоже симметрична и билинейна, и $g(l, l)=0$ для всех $l\in L$. Но по рассуждению в доказательстве теоремы п. 3 отсюда следует, что $g(l, m)=0$ для всех $l, m\in L$, что завершает доказательство. Заметим, что если $q(l)=g(l, l)$, g симметрична, то

$$g(l, m)=\frac{1}{2}[q(l+m)-q(l)-q(m)].$$

Мы установили, таким образом, что ортогональные геометрии (над полями характеристики $\neq 2$) можно рассматривать как геометрии пар (L, q), где $q\colon L \to \mathscr{K}$ — квадратичная форма. В координатах квадратичная форма записывается в виде

$$q(\vec{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

где матрица (a_{ij}) определяется однозначно, если она симметрична: $a_{ij} = a_{ji}$. Например,

$$q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2.$$

Теоремы классификации означают, что невырожденной линейной заменой переменных квадратичную форму можно привести к сумме квадратов с коэффициентами:

$$q(\vec{x}) = \sum_{i=1}^{n} a_i x_i^2.$$

Если $\mathscr{K} = \mathbf{R}$, можно считать, что $a_i = 0, \pm 1$; количества r_0, r_+, r_- нулей и плюс-минус единиц определены однозначно и составляют сигнатуру исходной квадратичной формы; $r_+ + r_-$ — это ее ранг. Если $\mathscr{K} = \mathbf{C}$, можно считать, что $a_i = 0, 1$; количество единиц — это ранг формы; он также определен однозначно.

§ 4. Алгоритм ортогонализации и ортогональные многочлены

В этом параграфе мы опишем классические алгоритмы для отыскания ортогональных базисов и важные примеры таких базисов в пространствах функций.

1. Приведение квадратичной формы к сумме квадратов. Пусть

$$q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j, \quad a_{ij} = a_{ji},$$

— квадратичная форма над полем $\mathscr{K}$ характеристики $\neq 2$. Следующая процедура дает удобный практический способ отыскания линейной замены переменных x_i, приводящей q к сумме квадратов (с коэффициентами).

Случай 1. *Существует ненулевой диагональный коэффициент.* Перенумеровав переменные, мы можем считать, что $a_{11} \neq 0$. Тогда

$$q(x_1, \ldots, x_n) = a_{11}x_1^2 + x_1(2a_{12}x_2 + \ldots + 2a_{1n}x_n) + q'(x_2, \ldots, x_n),$$

где q' — квадратичная форма от $\leqslant n-1$ переменных. Выделяя полный квадрат, находим

$$q(x_1, \ldots, x_n) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + q''(x_2, \ldots, x_n),$$

где q'' — новая квадратичная форма от $\leqslant n-1$ переменных. Полагая

$$y_1 = x_1 + a_{11}^{-1}(a_{12}x_2 + \ldots + a_{1n}x_n), \quad y_2 = x_2, \ \ldots, \ y_n = x_n,$$

мы получаем в новых переменных форму

$$a_{11}y_1^2 + q''(y_2, \ldots, y_n),$$

и следующий шаг алгоритма состоит в применении его к q''.

Случай 2. *Все диагональные коэффициенты равны нулю.* Если вообще $q = 0$, то делать ничего не нужно: $q = \sum_{i=1}^{n} 0 \cdot x_i^2$. Иначе, перенумеровав переменные, можно считать, что $a_{12} \neq 0$. Тогда

$$q(x_1, \ldots, x_n) =$$
$$= 2a_{12}x_1x_2 + x_1l_1(x_3, \ldots, x_n) + x_2l_2(x_3, \ldots, x_n) + q'(x_3, \ldots, x_n),$$

где l_1, l_2 — линейные формы, а q' — квадратичная. Положим

$$x_1 = y_1 + y_2, \quad x_2 = y_1 - y_2, \quad x_i = y_i, \quad i \geqslant 3.$$

В новых переменных форма q приобретает вид

$$2a_{12}(y_1^2 - y_2^2) + q''(y_1, y_2, \ldots, y_n),$$

где q'' не содержит членов с y_1^2, y_2^2. Поэтому к ней можно применить способ выделения полного квадрата и снова свести задачу к меньшему числу переменных. Последовательное применение этих шагов приведет форму к виду $\sum_{i=1}^{n} a_i z_i^2$. Окончательная линейная замена переменных будет невырожденной, так как таковы все промежуточные замены.

Последняя замена переменных $u_i = \sqrt{|a_i|}z_i$ при $a_i \neq 0$ в случае $\mathscr{K} = \mathbf{R}$ и $u_i = \sqrt{a_i}z_i$ при $a_i \neq 0$ в случае $\mathscr{K} = \mathbf{C}$ приведет форму к сумме квадратов с коэффициентами 0, ± 1 или 0, 1.

2. Алгоритм ортогонализации Грама — Шмидта. Он весьма близок к описанному в предыдущем пункте, но формулируется в более геометрических терминах. Мы будем рассматривать одновременно ортогональный и эрмитов случай.

Исходными данными являются: пространство (L, g) с ортогональной или эрмитовой метрикой, заданной в базисе $\{e_1', \ldots, e_n'\}$. Пусть L_i — подпространство, натянутое на $e_1', \ldots, e_i'$, $i = 1, \ldots, n$. Процесс ортогонализации, примененный к базису $\{e_1', \ldots, e_n'\}$, можно рассматривать как конструктивное доказательство следующего результата:

3. Предложение. *Предположим, что в описанных обозначениях все подпространства $L_1, \ldots, L_n$ невырождены. Тогда существует такой ортогональный базис $\{e_1, \ldots, e_n\}$ пространства L, что линейная оболочка $\{e_1, \ldots, e_i\}$ совпадает с L_i для всех $i = 1, \ldots, n$. Он называется результатом ортогонализации исходного базиса $\{e_1', \ldots, e_n'\}$. Каждый вектор e_i определен однозначно с точностью до умножения на ненулевой скаляр.*

Доказательство. Построим e_i индукцией по i. В качестве e_1 можно взять e_1'. Если $e_1, \ldots, e_{i-1}$ уже построены, будем искать e_i в виде

$$e_i = e_i' - \sum_{j=1}^{i-i} x_j e_j', \quad x_j \in \mathcal{K}.$$

Так как $\{e_1', \ldots, e_i'\}$ порождают L_i, а $\{e_1', \ldots, e_{i-1}'\}$ и $\{e_1, \ldots \ldots, e_{i-1}\}$ порождают L_{i-1}, любой такой вектор e_i вместе с $e_1, \ldots \ldots, e_{i-1}$ будет порождать L_i. Поэтому достаточно добиться того, чтобы e_i был ортогонален к $e_1, \ldots, e_{i-1}$, или, что то же самое, к $e_1', \ldots, e_{i-1}'$. Эти условия означают, что $g(e_i, e_k') = 0$, $k = 1, \ldots \ldots, i-1$, или

$$\sum_{j=1}^{i-1} x_j g(e_j', e_k') = g(e_i', e_k'), \qquad k = 1, \ldots, i-1.$$

Это система $i-1$ линейных уравнений для $i-1$ неизвестных x_j. Ее матрица коэффициентов есть матрица Грама базиса $\{e_1', \ldots, e_{i-1}'\}$ пространства L_{i-1}. По предположению, она невырождена, так что x_j существуют и определены однозначно. Любой ненулевой вектор $\tilde{e}_i$, ортогональный к L_{i-1}, должен быть пропорционален e_i.

Более простая и решаемая сразу система уравнений получится, если искать e_i в виде

$$e_i = e_i' - \sum_{j=1}^{i-1} y_j e_j, \quad y_j \in \mathcal{K},$$

считая $e_1, \ldots, e_{i-1}$ уже найденными. Поскольку $e_1, \ldots, e_{i-1}$ попарно ортогональны, из условий $g(e_i, e_j) = 0$, $1 \leqslant j \leqslant i-1$, находим

$$y_j = \frac{g(e_i', e_j)}{g(e_j, e_j)}, \qquad j = 1, \ldots, i-1.$$

Весь смысл этого доказательства состоит в явном выписывании систем линейных уравнений, последовательное решение которых определяет e_i. Заметим, что матрица коэффициентов первой системы суть последовательные диагональные миноры матрицы Грама исходного базиса:

$$G_i = (g(e_j', e_k')), \qquad 1 \leqslant j, \ k \leqslant i.$$

Если бы мы не стремились к алгоритмичности, проще всего было бы рассуждать так: в силу предложения п. 2 § 3 и невырожденности L_{i-1} имеем

$$L_i = L_{i-1} \oplus L_{i-1}^{\perp}; \qquad \dim L_{i-1}^{\perp} = \dim L_i - \dim L_{i-1} = 1.$$

Возьмем теперь в качестве e_i любой ненулевой вектор из $L_{i-1}^{\perp}$.

4. Замечания и следствия. а) Процесс ортогонализации Грама — Шмидта чаще всего применяется в ситуации, когда $g(l, l) > 0$

для всех $l \in L$, $l \neq 0$, т. е. к *евклидовым* и *унитарным* пространствам, которые мы подробно изучим позже. В этом случае *все подпространства* L автоматически *невырождены*, и ортогонализировать можно любой исходный базис. Форма g с таким свойством называется *положительно определенной*, и ее матрицы Грама называются *положительно определенными*.

б) В случае $\mathcal{K} = \mathbf{R}$ или $\mathbf{C}$ можно строить сразу ортонормированный базис. Для этого, отыскав вектор e_i, как в доказательстве предложения, следует тут же заменить его на $|g(e_i, e_i)|^{-1/2} e_i$ или $g(e_i, e_i)^{-1/2} e_i$ (для ортогональных пространств над $\mathbf{C}$).

в) *Любой ортогональный базис невырожденного подпространства $L_0 \subset L$ можно дополнить до ортогонального базиса всего пространства L.*

Действительно, $L = L_0 \oplus L_0^{\perp}$, и в качестве дополнения можно взять ортогональный базис $L_0^{\perp}$. Искать его можно методом Грама — Шмидта, если сначала как-нибудь дополнить базис L_0 до базиса L, позаботившись о невырожденности промежуточных подпространств.

г) Пусть $\{e'_1, \ldots, e'_n\}$ — базис (L, g), а $\{e_1, \ldots, e_n\}$ — его ортогонализация. Положим $a_i = g(e_i, e_i)$ — это единственные ненулевые элементы матрицы Грама базиса $\{e_i\}$. Будем считать, что g эрмитова или g ортогональна над $\mathbf{R}$. Тогда все числа a_i вещественны, и сигнатура g определяется количеством положительных и отрицательных чисел a_i. Покажем, как восстановить ее по минорам исходной матрицы Грама $G = \{g(e'_i, e'_k)\}$. Пусть G_i — i-й диагональный минор, т. е. матрица Грама $\{e'_1, \ldots, e'_i\}$. Если A_i — матрица перехода к базису $\{e_1, \ldots, e_i\}$, то

$$\det(g(e_k, e_j))_{1 \leqslant k,\, j \leqslant i} = a_1 \ldots a_i = \det(A_i^t G_i A_i) = \det G_i (\det A_i)^2$$

в ортогональном случае или

$$a_1 \ldots a_i = \det(A_i^t G_i \overline{A}_i) = \det G_i |\det A_i|^2$$

в эрмитовом случае. Поэтому всегда

$$\text{знак } a_1 \ldots a_i = \text{знак } \det G_i.$$

Итак, *сигнатура формы g определяется числом положительных и отрицательных элементов последовательности*

$$\det G_1, \ \frac{\det G_2}{\det G_1}, \ \ldots, \ \frac{\det G_n}{\det G_{n-1}}.$$

В частности, *форма g (и ее матрица G) положительно определена тогда и только тогда, когда все миноры* $\det G_i$ *положительны* (напомним, что G либо вещественна и симметрична, либо комплексна и эрмитово симметрична). Этот результат называется *критерием Сильвестра*.

Более общо, для невырожденной квадратичной формы над любым полем тождество

$$a_1 \ldots a_i = \det G_i (\det A_i)^2$$

показывает, что исходную *форму с симметричной матрицей G и невырожденными диагональными минорами G_i можно линейным преобразованием переменных привести к виду*

$$\sum_{i=1}^{n} \frac{\det G_i}{\det G_{i-1}} y_i^2, \ \det G_0 = 1,$$

ибо квадраты $(\det A_i)^2$, мешающие непосредственно выразить a_i через $\det G_i$, можно внести сомножителями в переменные. Этот результат называется *теоремой Якоби.*

5. Билинейные формы на пространствах функций. Рассмотрим функции f_1, f_2, заданные на отрезке (a, b) вещественной прямой (возможно, $a = -\infty$, $b = \infty$) и принимающие вещественные или комплексные значения. Пусть $G(x)$ — фиксированная функция от $x \in (a, b)$. Билинейные формы на пространствах функций в анализе часто задаются выражениями типа

$$g(f_1, f_2) = \int_a^b G(x) f_1(x) f_2(x)\, dx$$

или (полуторалинейный случай)

$$g(f_1, f_2) = \int_a^b G(x) f_1(x) \overline{f_2(x)}\, dx.$$

Разумеется, G, f_1 и f_2 должны удовлетворять каким-то условиям интегрируемости; в последующих примерах они будут выполнены автоматически.

Функция G называется *весом* формы g. Значение

$$g(f, f) = \int_a^b G(x) f(x)^2\, dx \quad \text{или} \quad \int_a^b G(x) |f(x)|^2\, dx$$

есть *взвешенное квадратичное среднее* функции f (с весом G); если $G \geqslant 0$, его можно рассматривать как некоторую интегральную меру уклонения f от нуля. Типичная задача *аппроксимации* функции f линейными комбинациями некоторого заданного набора функций $f_1, \ldots, f_n, \ldots$ состоит в поиске таких коэффициентов $a_1, \ldots, a_n, \ldots$, которые при данном n минимизируют взвешенное среднее квадратичное функции

$$f - \sum a_i f_i.$$

Позже будет видно, что коэффициенты a_i особенно просто находятся в случае, когда $\{f_i\}$ образуют ортогональную или ортонормированную систему относительно скалярного произведения g. В этом параграфе мы ограничимся явным описанием нескольких важных ортогональных систем

6. Тригонометрические многочлены. Здесь $G = 1$, $(a, b) = (0, 2\pi)$. *Тригонометрическими многочленами* (или *многочленами*

Фурье) называются конечные линейные комбинации функций $\cos nx$, $\sin nx$ или конечные линейные комбинации функций e^{inx}, $n \in \mathbf{Z}$. Обычно первые применяются в теории вещественнозначных функций, а вторые — комплекснозначных. Поскольку $e^{inx} = \cos nx + i \sin nx$, над $\mathbf{C}$ оба пространства многочленов Фурье совпадают. Над $\mathbf{R}$ используется билинейная метрика, над $\mathbf{C}$ — полуторалинейная. Функции $\{1, \cos nx, \sin nx \mid n \geqslant 1\}$ и $\{e^{inx} \mid n \in \mathbf{Z}\}$ линейно независимы (как над $\mathbf{R}$, так и над $\mathbf{C}$). Кроме того, они образуют ортогональную систему, как следует из легко проверяемых формул:

$$\int_0^{2\pi} \cos mx \cos nx\, dx = \int_0^{2\pi} \sin mx \sin nx\, dx = \begin{cases} \pi & \text{при } m = n > 0, \\ 0 & \text{при } m \neq n, \end{cases}$$

$$\int_0^{2\pi} \cos mx \sin nx\, dx = 0$$

$$\int_0^{2\pi} e^{imx} \overline{e^{inx}}\, dx = \int_0^{2\pi} e^{i(m-n)x}\, dx = \begin{cases} 2\pi & \text{при } m = n, \\ 0 & \text{при } m \neq n. \end{cases}$$

Системы

$$\left\{ \frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos nx, \quad \frac{1}{\sqrt{\pi}} \sin nx \mid n \geqslant 1 \right\} \text{ и } \left\{ \frac{1}{\sqrt{2\pi}} e^{inx} \mid n \in \mathbf{Z} \right\}$$

поэтому ортонормированы. Скалярные произведения любой функции f на $[0, 2\pi]$ с элементами этих ортонормированных систем называются *коэффициентами Фурье* этой функции:

$$a_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\, dx,$$

$$a_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \cos nx\, dx, \qquad n \geqslant 1,$$

$$b_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \sin nx\, dx, \qquad n \geqslant 1,$$

для вещественных функций f и

$$a_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-inx}\, dx, \qquad n \in \mathbf{Z},$$

для комплексных функций. Если сама функция f является многочленом Фурье, то по формуле разложения из п. 6 § 3 имеем

$$f(x) = \frac{1}{\sqrt{2\pi}} a_0 + \frac{1}{\sqrt{\pi}} \sum_{n \geqslant 1} (a_n \cos nx + b_n \sin nx)$$

для вещественных функций f и

$$f(x)=\frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty}a_n e^{inx}$$

для комплексных функций f. Суммы справа, разумеется, конечны в рассматриваемом случае. Бесконечные ряды такой структуры называются *рядами Фурье*. Вопрос об их сходимости вообще и сходимости к той функции f, коэффициентами Фурье которой являются a_n, b_n, в частности, исследуется в одной из важнейших глав анализа.

7. Многочлены Лежандра. Здесь $G=1$, $(a,b)=(-1,1)$. Многочлены Лежандра $P_0(x)$, $P_1(x)$, $P_2(x)$, ... определяются как результат процесса ортогонализации, примененного к базису $\{1, x, x^2, \ldots,\}$ пространства вещественных многочленов. Обычно они нормируются условием $P_n(1)=1$. В такой нормировке их явный вид дается следующим результатом:

8. Предложение. $P_0(x)=1,\quad P_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n,\quad n\geqslant 1.$

Доказательство. Так как степень многочлена $(x^2-1)^n$ равна $2n$, степень $\frac{d^n}{dx^n}(x^2-1)^n$ равна n, так что $P_1, \ldots, P_i$ порождают то же пространство над $\mathbf{R}$, что и $1, x, \ldots, x^i$. Поэтому для проверки ортогональиости P_i, P_j, $i\neq j$, достаточно убедиться, что

$$\int_{-1}^{1}x^kP_n(x)\,dx=0\quad\text{при}\quad k<n.$$

Интегрируя по частям, получим

$$\int_{-1}^{1}x^k\frac{d^n}{dx^n}(x^2-1)^n\,dx=$$

$$=x^k\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\Bigg|_{-1}^{1}-k\int_{-1}^{1}x^{k-1}\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\,dx.$$

Первое слагаемое обращается в нуль, ибо $(x^2-1)^n$ в точках ±1 имеет корень кратности n, а каждое дифференцирование снижает кратность корня на единицу. Ко второму слагаемому можно применить аналогичную процедуру; после k шагов получится интеграл, пропорциональный

$$\int_{-1}^{1}\frac{d^{n-k}}{dx^{n-k}}(x^2-1)^n\,dx=\frac{d^{n-k-1}}{dx^{n-k-1}}(x^2-1)^n\Bigg|_{-1}^{1}=0.$$

Далее, по формуле Лейбница

$$\frac{d^n}{dx^n}[(x-1)^n(x+1)^n]=\sum_{k=0}^{n}\binom{n}{k}\frac{d^k}{dx^k}(x-1)^n\frac{d^{n-k}}{dx^{n-k}}(x+1)^n.$$

В точке $x=1$ не обращается в нуль только слагаемое, отвечающее $k=n$, так что

$$P_n(1)=\frac{1}{2^n n!}\binom{n}{n}\left[\frac{d^n}{dx^n}(x-1)^n\right](x+1)^n\Big|_{x=1}=\frac{1}{2^n n!}\cdot 1\cdot n!\cdot 2^n=1,$$

что завершает доказательство.

9. Многочлены Чебышева. $G=\dfrac{1}{\sqrt{1-x^2}}$, $(a,b)=(-1,1)$. Многочлены $T_n(x)$, $n\geqslant 0$, суть результат ортогонализации базиса $\{1,x,x^2,\ldots\}$. Явные формулы:

$$T_n(x)=\frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\,\frac{d^n}{dx^n}(1-x^2)^{n-1/2}=\cos(n\arccos x).$$

Нормировка:

$$\int_{-1}^{1}\frac{T_m(x)\,T_n(x)\,dx}{\sqrt{1-x^2}}=\begin{cases}0 & \text{при}\quad m\neq n,\\ \pi/2 & \text{при}\quad m=n\neq 0,\\ \pi & \text{при}\quad m=n=0.\end{cases}$$

10. Многочлены Эрмита. $G=e^{-x^2}$, $(a,b)=(-\infty,\infty)$. Многочлены $H_n(x)$ суть результат ортогонализации базиса $\{1,x,x^2,\ldots\}$. Явные формулы:

$$H_n(x)=(-1)^n e^{x^2}\frac{d^n}{dx^n}(e^{-x^2}).$$

Нормировка:

$$\int_{-\infty}^{\infty}e^{-x^2}H_m(x)\,H_n(x)\,dx=\begin{cases}0 & \text{при}\quad m\neq n,\\ 2^n n!\sqrt{\pi} & \text{при}\quad m=n.\end{cases}$$

Мы оставляем доказательства читателю в качестве упражнения.

УПРАЖНЕНИЯ

1. Доказать, что эрмитова или ортогональная форма g неотрицательно определена, т. е. $g(l,l)\geqslant 0$ для всех $l\in L$, тогда и только тогда, когда все диагональные миноры ее матрицы Грама неотрицательны.
2. Доказать утверждения п. 9 и 10 этого параграфа.

§ 5. Евклидовы пространства

1. Определение. *Евклидовым пространством называется конечномерное вещественное линейное пространство L с симметричным положительно определенным скалярным произведением.*

Мы будем писать (l, m) вместо $g(l, m)$ и $|l|$ вместо $(l, l)^{1/2}$; число $|l|$ будем называть *длиной* вектора l.

Из результатов, доказанных в § 3—4, следует, что:

а) во всяком евклидовом пространстве есть ортонормированный базис, все векторы которого имеют длину 1;

б) поэтому оно изометрично координатному евклидову пространству $\mathbb{R}^n$ $(n = \dim L)$, в котором

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i y_i, \quad |\vec{x}| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

Ключом ко многим свойствам евклидова пространства является многократно переоткрывавшееся неравенство Коши — Буняковского — Шварца:

2. Предложение. *Для любых $l_1, l_2 \in L$ имеем*

$$(l_1, l_2) \leqslant |l_1||l_2|.$$

Равенство достигается тогда и только тогда, когда векторы l_1, l_2 линейно зависимы.

Доказательство. В случае $l_1 = 0$ имеет место равенство и l_1, l_2 линейно зависимы. Будем считать, что $l_1 \neq 0$. Для любого вещественного числа t имеем

$$|tl_1 + l_2|^2 = (tl_1 + l_2, tl_1 + l_2) = t^2|l_1|^2 + 2t(l_1, l_2) + |l_2|^2 \geqslant 0$$

в силу положительной определенности скалярного произведения. Поэтому дискриминант квадратного трехчлена справа неположителен, т. е.

$$(l_1, l_2)^2 - |l_1|^2|l_2|^2 \leqslant 0.$$

Он равен нулю тогда и только тогда, когда этот трехчлен имеет вещественный корень t_0. В этом случае

$$|t_0 l_1 + l_2|^2 = 0 \Leftrightarrow l_2 = -t_0 l_1,$$

что завершает доказательство.

3. Следствие (неравенство треугольника). *Для любых $l_1, l_2, l_3 \in L$*

$$|l_1 + l_2| \leqslant |l_1| + |l_2|, \qquad |l_1 - l_3| \leqslant |l_1 - l_2| + |l_2 - l_3|.$$

Доказательство. Имеем

$$|l_1 + l_2|^2 = |l_1|^2 + 2(l_1, l_2) + |l_2|^2 \leqslant |l_1|^2 + 2|l_1||l_2| + |l_2|^2 = \\ = (|l_1| + |l_2|)^2.$$

Заменив здесь l_1 на $l_1 - l_2$ и l_2 на $l_2 - l_3$, получим второе неравенство.

4. Следствие. *Евклидова длина вектора $|l|$ является нормой на L в смысле определения в п. 4 § 10 ч. 1, а функция $d(l, m) = |l - m|$ — метрикой в смысле определения п. 1 там же.*

Доказательство. Остается проверить только, что $|al| = |a||l|$ для всех $a \in \mathbb{R}$, но

$$|al| = (al, al)^{1/2} = (a^2|l|^2)^{1/2} = |a||l|.$$

5. Углы и расстояния. Пусть $l_1, l_2 \in L$ — ненулевые векторы. В силу предложения п. 2

$$-1 \leqslant \frac{(l_1, l_2)}{|l_1||l_2|} \leqslant 1.$$

Поэтому существует единственный угол φ, $0 \leqslant \varphi \leqslant \pi$, для которого

$$\cos\varphi = \frac{(l_1, l_2)}{|l_1||l_2|}.$$

Он называется углом между векторами l_1, l_2. Поскольку скалярное произведение симметрично, это «неориентированный угол», чем и объясняется интервал его значений. В соответствии со школьной геометрией угол между ортогональными векторами равен $\pi/2$. Можно систематически развить евклидову геометрию на основе данных определений длины и угла и убедиться, что в размерностях два и три она совпадает с классической.

Например, многомерная теорема Пифагора есть тривиальное следствие определений: если векторы $l_1, \ldots, l_n$ попарно ортогональны, то

$$\left|\sum_{i=1}^{n} l_i\right|^2 = \sum_{i=1}^{n} |l_i|^2.$$

Обычная формула косинусов в геометрии плоскости, примененная к треугольнику со сторонами l_1, l_2, l_3, утверждает, что

$$|l_3|^2 = |l_1|^2 + |l_2|^2 - 2\,|l_1|\,|l_2|\cos\varphi,$$

где φ — угол между l_1 и l_2. В векторном варианте $l_3 = l_1 - l_2$, и эта формула превращается в тождество

$$|l_1 - l_2|^2 = |l_1|^2 + |l_2|^2 - 2(l_1, l_2)$$

в соответствии с нашим определением угла.

Пусть $U, V \subset L$ — два множества в евклидовом пространстве. *Расстоянием* между ними называется неотрицательное число

$$d(U, V) = \inf\{|l_1 - l_2|\,|\,l_1 \in U,\ l_2 \in V\}.$$

Рассмотрим частный случай: $U = \{l\}$ (один вектор), $V = L_0 \subset L$ — линейное подпространство. В силу предложения п. 2 § 3 имеем $L = L_0 \oplus L_0^{\perp}$ и $l = l_0 + l_0'$, где $l_0 \in L_0$, $l_0' \in L_0^{\perp}$. Векторы l_0, l_0' суть *ортогональные проекции* l на L_0, $L_0^{\perp}$ соответственно.

6. Предложение. *Расстояние от l до L_0 равно длине ортогональной проекции l на $L_0^{\perp}$.*

Доказательство. Для любого вектора $m \in L_0$ имеем

$$|l - m|^2 = |l_0 + l_0' - m|^2 = |l_0 - m|^2 + |l_0'|^2$$

в силу теоремы Пифагора, ибо векторы $l_0 - m \in L_0$ и $l_0' \in L_0^{\perp}$ ортогональны. Следовательно,

$$|l - m|^2 \geqslant |l_0'|^2,$$

и равенство достигается только в случае $m = l_0$, что доказывает требуемое.

Если в L_0 выбран ортонормированный базис $\{e_1, \ldots, e_m\}$, то проекция l на L_0 определяется формулой

$$l_0 = \sum_{i=1}^{m} (l, e_i) e_i.$$

Действительно, левая и правая части имеют одинаковые скалярные произведения со всеми e_i, поэтому их разность лежит в $L_0^\perp$. Окончательно,

$$d(l, L_0) = \left| l - \sum_{i=1}^{m} (l, e_i) e_i \right|$$

есть наименьшее значение $|l - m|$, когда m пробегает L_0. Поскольку $|l_0|^2 \leqslant |l|^2$ по той же теореме Пифагора, имеем

$$\sum_{i=1}^{m} (l, e_i)^2 \leqslant |l|^2.$$

7. Приложения к пространствам функций. Рассмотрим в качестве примера пространство непрерывных вещественных функций на $[a, b] \subset \mathbb{R}$ со скалярным произведением

$$(f, g) = \int_a^b fg\, dx.$$

Оно бесконечномерно, но все наши неравенства будут относиться к конечному числу таких функций, так что каждый раз можно будет считать, что мы работаем в конечномерном евклидовом пространстве: $\int_a^b f^2(x)\, dx \geqslant 0$ и если $\int_a^b f^2(x)\, dx = 0$, то $f(x) \equiv 0$.

Неравенство Коши — Буняковского — Шварца приобретает вид

$$\left(\int_a^b f(x) g(x)\, dx \right)^2 \leqslant \int_a^b f(x)^2\, dx \int_a^b g(x)^2\, dx.$$

Неравенство треугольника:

$$\left(\int_a^b (f(x) + g(x))^2\, dx \right)^{1/2} \leqslant \left(\int_a^b f(x)^2\, dx \right)^{1/2} + \left(\int_a^b g(x)^2\, dx \right)^{1/2}.$$

Если $(a, b) = (0, 2\pi)$ и a_i, b_i — коэффициенты Фурье функции $f(x)$, как в п. 6 § 4, то многочлен Фурье

$$f_N(x) = \frac{1}{\sqrt{2\pi}} a_0 + \frac{1}{\sqrt{\pi}} \sum_{n=1}^{N} (a_n \cos nx + b_n \sin nx)$$

является ортогональной проекцией $f(x)$ на линейную оболочку $\{1, \cos nx, \ \sin nx \mid 1 \leqslant n \leqslant N\}$. Поэтому коэффициенты Фурье $f(x)$ при каждом N минимизируют среднеквадратичное отклонение $f(x)$ от многочленов Фурье «степени» $\leqslant N$. Неравенство $|f_N|^2 \leqslant |f|^2$ приобретает вид

$$a_0^2 + \sum_{i=1}^{N} (a_i^2 + b_i^2) \leqslant \int_0^{2\pi} f(x)^2 \, dx.$$

Поскольку правая часть не зависит от N, а a_i^2, $b_i^2 \geqslant 0$, ряд

$$a_0^2 + \sum_{i=1}^{\infty} (a_i^2 + b_i^2)$$

сходится для любой непрерывной функции $f(x)$ на $[0, 2\pi]$. Можно доказать, что он сходится в точности к $\int\limits_0^{2\pi} f(x)^2 \, dx$.

Совершенно аналогичные соображения применимы к многочленам Лежандра, Чебышева и Эрмита. Мы оставляем их в качестве упражнения читателю.

8. Метод наименьших квадратов. Рассмотрим систему m линейных уравнений для n неизвестных с вещественными коэффициентами

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad i = 1, \ldots, m.$$

Предположим, что эта система «переопределена», т. е. $m > n$ и ранг матрицы коэффициентов равен n. Тогда она, вообще говоря, не имеет решений. Но можно попробовать найти такие значения неизвестных $x_1^0, \ldots, x_n^0$, чтобы суммарное среднеквадратичное отклонение левых частей от правых

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j^0 - b_i \right)^2$$

принимало наименьшее возможное значение. Эта задача имеет существенные практические приложения. Например, при геодезических работах местность разбивается на сеть треугольников, некоторые элементы которых измеряются, а другие вычисляются по формулам тригонометрии. Поскольку все измерения приближенные, рекомендуется сделать их больше, чем строго необходимо для вычисления остальных элементов, но по той же причине тогда уравнения для этих элементов почти наверняка окажутся несовместными. Метод наименьших квадратов позволяет получить «приближенное решение», более надежное из-за большего количества вложенной в систему информации.

Покажем, что наша задача может быть решена с использованием результатов п. 7. Интерпретируем столбцы матрицы коэффициентов $e_i = (a_{1i}, \ldots, a_{mi})$ и столбец свободных членов $f = (b_1, \ldots, b_m)$ как векторы координатного евклидова пространства $\mathbf{R}^m$ со стандартным скалярным произведением. Положив

$$e = \sum_{i=1}^{n} x_i e_i,$$

получим, что

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 = \left| \sum_{i=1}^{n} x_i e_i - f \right|^2.$$

Поэтому минимум среднеквадратичного отклонения достигается тогда, когда $\sum_{i=1}^{n} x_i^0 e_i$ является ортогональной проекцией f на подпространство, натянутое на e_i. Это означает, что коэффициенты x_i^0 должны находиться из системы n уравнений с n неизвестными

$$\left(\sum_{i=1}^{n} x_i^0 e_i,\ e_j \right) = (f, e_j), \quad j = 1, \ldots, n,$$

так называемой *«нормальной системы»*. Ее определитель есть определитель матрицы Грама $((e_i, e_j))$, где

$$(e_i, e_j) = \sum_{k=1}^{m} a_{ki} a_{kj}.$$

Он отличен от нуля, ибо предполагалось, что ранг исходной системы, т. е. системы векторов (e_i), равен n (см. упражнение 5 к § 2). Поэтому решение существует и единственно.

Вернемся теперь к теме «измерения в евклидовом пространстве».

9. n-мерный объем. На одномерном евклидовом пространстве простейшим фигурам — отрезкам и их конечным объединениям — можно поставить в соответствие длины и суммы длин. На евклидовой плоскости школьная геометрия учит измерять площади таких фигур, как прямоугольники, треугольники и, с некоторым трудом, круги. Обобщение этих понятий дает глубокая общая *теория меры*, естественное место которой не здесь. Мы ограничимся списком основных свойств и элементарными вычислениями, связанными со специальной мерой фигур в n-мерном евклидовом пространстве — их *n-мерным объемом.*

n-мерный объем есть функция vol^n, определенная на *некоторых* подмножествах n-мерного евклидова пространства L, называемых *измеримыми*, и принимающая неотрицательные вещественные значения или ∞ (на ограниченных измеримых множествах — только конечные значения). Совокупность измеримых множеств достаточно богата. Мы просто постулируем следующий список свойств vol^n и измеримость фигурирующих в них множеств, не доказывая существование функции с такими свойствами и не указывая естественную область ее определения.

а) Функция vol^n счетно аддитивна, т. е.

$$\mathrm{vol}^n\left(\bigcup_{i=1}^{\infty} U_i\right) = \sum_{i=1}^{\infty} \mathrm{vol}^n U_i, \text{ если } U_i \cap U_j = \varnothing \text{ при } i \neq j;$$

vol^1 (точка) $= 0$; vol^1 (отрезок) $=$ длина отрезка. Отрезок в одномерном евклидовом пространстве есть множество векторов вида $tl_1 + (1-t)l_2$, $0 \leqslant t \leqslant 1$; его длина есть $|l_1 - l_2|$.

б) Если $U \subseteq V$, то $\mathrm{vol}^n U \leqslant \mathrm{vol}^n V$.

в) Если $L = L_1 \oplus L_2$ (ортогональная прямая сумма), $\dim L_1 = m$, $\dim L_2 = n$, $U \subset L_1$, $V \subset L_2$, то для $U \times V = \{(l_1, l_2) \mid l_1 \in U,\ l_2 \in V\} \in L_1 \oplus L_2$ имеем

$$\mathrm{vol}^{m+n}(U \times V) = \mathrm{vol}^m U \cdot \mathrm{vol}^n V.$$

г) Если $f\colon L \to L$ — произвольный линейный оператор, то

$$\mathrm{vol}^n (f(U)) = |\det f|\, \mathrm{vol}^n U,\ n = \dim L.$$

Свойства, а), б) едва ли нуждаются в комментариях. Свойство в) является сильным обобщением формулы площади прямоугольника (произведение длин сторон) или объема прямого цилиндра (произведение площади основания на длину образующей). Заметим, что из свойства в) вытекает, что $(m+n)$-мерный объем ограниченного множества W в L, лежащего в подпространстве L_1 размерности $m < m+n$, равен нулю. Действительно, тогда $L = L_1 \oplus L_1^{\perp}$ и $W = V \times \{0\}$, наконец, $\mathrm{vol}^n(\{0\}) = 0$ при $n > 0$ в силу а) и в).

Смысл свойства г) менее очевиден. Оно является основным вкладом линейной алгебры в теорию евклидовых объемов и служит причиной появления якобианов в формализме многомерного интегрирования. Возможно, наиболее интуитивное объяснение его состоит в замечании, что оператор растяжения в $a \in \mathbf{R}$ раз вдоль одного из векторов ортогонального базиса должен умножать объемы на $|a|$ в силу свойств а) и в). Но любой ненулевой вектор можно дополнить до ортогонального базиса, поэтому диагонализируемый оператор f с собственными значениями $a_1, \ldots, a_n \in \mathbf{R}$ должен умножать объемы на $|a_1| \ldots |a_n| = |\det f|$. Наконец, изометрии должны сохранять объемы, и, как мы убедимся позже, любой оператор есть композиция диагонализируемого и изометрии (см. § 8, упражнение 11).

Теперь, пользуясь этими аксиомами, приведем список объемов простейших и наиболее важных n-мерных фигур.

10. Единичный куб. Это множество $\{t_1e_1 + \ldots + t_ne_n \mid 0 \leqslant t_i \leqslant 1\}$, где $\{e_1, \ldots, e_n\}$ — некоторый ортонормированный базис L. Из свойств а) и в) из п. 9 сразу следует, что его объем равен единице.

Куб со стороной $a > 0$ получится, если разрешить t_i пробегать значения $0 \leqslant t_i \leqslant a$. Так как он является образом единичного куба относительно гомотетии — умножения на a, — его объем равен a^n.

11. Параллелепипед со сторонами $\{l_1, \ldots, l_n\}$. Это множество $\{t_1l_1 + \ldots + t_nl_n \mid 0 \leqslant t_i \leqslant 1\}$. Мы покажем, что его объем равен

$\sqrt{|\det G|}$, где $G = ((l_i, l_j))$ — матрица Грама сторон. В самом деле, если $\{l_1, \ldots, l_n\}$ линейно зависимы, то соответствующий параллелепипед лежит в подпространстве размерности $< \dim L$ и его n-мерный объем равен нулю по замечанию в п. 9. В то же время матрица G вырождена.

Поэтому остается разобрать случай, когда $\{l_1, \ldots, l_n\}$ линейно независимы. Пусть $\{e_1, \ldots, e_n\}$ — ортонормированный базис в L, а f — линейное отображение $L \to L$, переводящее e_i в l_i, $i = 1, \ldots$ $\ldots, n$. Если A — матрица этого отображения в базисе $\{e_i\}$:

$$(l_1, \ldots, l_n) = (e_1, \ldots, e_n) A,$$

то матрица Грама $\{l_i\}$ равна $A^t A$, ибо матрица Грама $\{e_i\}$ единичная. Следовательно,

$$\sqrt{|\det G|} = \sqrt{|\det (A^t A)|} = |\det A|.$$

С другой стороны, $|\det A| = |\det f|$, и f переводит единичный куб в наш параллелепипед. В силу свойства г) из п. 9 объем параллелепипеда равен $|\det f|$, что завершает доказательство.

12. n-мерный шар радиуса r. Это множество векторов

$$B^n(r) = \{l \mid |l| \leqslant r\},$$

или, в ортогональных координатах,

$$B^n(r) = \left\{(x_1, \ldots, x_n) \,\middle|\, \sum_{i=1}^{n} x_i^2 \leqslant r^2\right\}.$$

Так как $B^n(r)$ получается из $B^n(1)$ растяжением в r раз, имеем

$$\operatorname{vol}^n B^n(r) = \operatorname{vol}^n B^n(1)\, r^n.$$

Константа $\operatorname{vol}^n B^n(1) = b_n$ может быть вычислена лишь аналитическими средствами. Рассекая $(n+1)$-мерный шар n-мерными линейными подмногообразиями, ортогональными к некоторому направлению, получим индуктивную формулу

$$b_{n+1} = \left[2 \int_0^1 \left(\sqrt{1 - x_{n+1}^2}\right)^n dx_{n+1}\right] b_n, \quad n \geqslant 1.$$

Разумеется, $b_1 = 2$, $b_2 = \pi$, $b_3 = \frac{4}{3}\pi$.

13. n-мерный эллипсоид с полуосями $r_1, \ldots, r_n$. Он задается в ортогональных координатах уравнениями

$$\sum_{i=1}^{n} \left(\frac{x_i}{r_i}\right)^2 \leqslant 1.$$

Поскольку он получается из $B^n(1)$ растяжениями в r_i раз вдоль i-й полуоси, его объем равен $b_n r_1 \ldots r_n$.

14. Одно свойство n-мерного объема. Оно состоит в том, что при очень больших n *«объем n-мерной фигуры сосредоточен* *вблизи ее поверхности*». Например, объем шарового кольца между сферами радиуса 1 и $1-\varepsilon$ равен $b_n[1-(1-\varepsilon)^n]$, что при фиксированном сколь угодно малом ε, но растущем n стремится к b_n. Двадцатимерный арбуз радиуса 20 см с толщиной корки 1 см чуть не на две трети состоит из корки:

$$1-\left(1-\frac{1}{20}\right)^{20}\approx 1-e^{-1}.$$

Это обстоятельство играет большую роль в статистической механике. Рассмотрим, например, простейшую модель газа в резервуаре, состоящего из n атомов, которые будем считать материальными точками массы 2 (в подходящей системе единиц). Представим мгновенное состояние газа n трехмерными векторами $(\vec{v}_1, \ldots, \vec{v}_n)$ скоростей всех молекул в физическом евклидовом пространстве, т. е. точкой $3n$-мерного координатного пространства $\mathbf{R}^{3n}$. Квадрат длины векторов в $\mathbf{R}^{3n}$ имеет прямой физический смысл *энергии* системы (суммы кинетических энергий атомов):

$$E=\sum_{i=1}^{n}|v_i|^2.$$

Для макроскопического объема газа в нормальных условиях порядок n есть 10^{23} (число Авогадро), так что состояние газа описывается точкой на сфере огромной размерности, радиус которой есть корень квадратный из энергии.

Пусть два таких резервуара соединены так, что они могут обмениваться энергией, но не атомами, и сумма их энергий $E_1+E_2=E$ остается постоянной. Тогда энергии E_1 и E_2 бо́льшую часть времени будут близки к таким, которые максимизируют «объем пространства состояний», доступный объединенной системе, т. е. произведение

$$\mathrm{vol}^{n_1} B\left(E_1^{1/2}\right) \mathrm{vol}^{n_2} B\left(E_2^{1/2}\right)$$

(мы заменили площади сфер объемами шаров, что буквально не верно, но почти не влияет на результат). Так как с ростом E_1 и убыванием E_2 $(E_1+E_2=\mathrm{const})$ первый объем невероятно быстро растет, а второй убывает, имеется резкий пик этого произведения при некоторых значениях E_1, E_2, отвечающий «наиболее вероятному» состоянию объединенной системы. Очевидно, это происходит там, где

$$\frac{d}{dE_1}\log \mathrm{vol}^{n_1} B\left(E_1^{1/2}\right)=\frac{d}{dE_2}\log \mathrm{vol}^{n_2} B\left(E_2^{1/2}\right).$$

Обратные к этим величинам суть (с точностью до пропорциональности) *температуры* резервуаров, и наиболее вероятное состояние отвечает равенству температур.

УПРАЖНЕНИЯ

1. Доказать, что угол φ наклона прямой в плоскости $\mathbf{R}^2$, проходящей из начала координат в среднеквадратичном как можно ближе к m заданным точкам (a_i, b_i), $i = 1, \ldots, m$, определяется формулой

$$\operatorname{tg} \varphi = \left(\sum_{i=1}^{m} a_i b_i \right) \Bigg/ \left(\sum_{i=1}^{m} a_i^2 \right).$$

(*Указание.* Найти наилучшее «приближенное решение» системы уравнений $a_i x = b_i$.)

2. Пусть $P_n(x)$ — n-й многочлен Лежандра. Доказать, что старший коэффициент многочлена $u_n(x) = \frac{2^n (n!)^2}{(2n)!} P_n(x)$ равен единице и что минимум интеграла $I(u) = \int\limits_{-1}^{1} u(x)^2 \, dx$ на множестве многочленов $u(x)$ степени n со старшим коэффициентом 1 достигается при $u = u_n$. (*Указание.* Разложить u по многочленам Лежандра степени $\leqslant n$.)

3. Пусть (S, μ) — пара, состоящая из конечного множества S и вещественной функции $\mu\colon S \to \mathbf{R}$, удовлетворяющей двум условиям: $\mu(s) \geqslant 0$ для всех $s \in S$ и $\sum\limits_{s \in S} \mu(s) = 1$. Рассмотрим на пространстве вещественных функций $F(S)$ на S (со значениями в $\mathbf{R}$) линейный функционал $E\colon F(S) \to \mathbf{R}$:

$$E(f) = \sum_{s \in S} \mu(s) f(s).$$

Обозначим через $F_0(S)$ ядро E.

(S, μ) называется конечным вероятностным пространством, элементы $F(S)$ — случайными величинами на нем, элементы $F_0(S)$ — нормированными случайными величинами, число $E(f)$ — математическим ожиданием величины f. Случайные величины образуют кольцо относительно обычного умножения функций.

Доказать следующие факты.

а) $F(S)$ и $F_0(S)$ имеют структуру ортогонального пространства с квадратом длины вектора f, равным $E(f^2)$. Пространство $F(S)$ евклидово тогда и только тогда, когда $\mu(s) > 0$ для всех $s \in S$.

б) Для любых случайных величин $f, g \in F(S)$ и $a, b \in \mathbf{R}$ положим

$$P(f = a) = \sum_{f(s) = a} \mu(s); \quad P(f = a;\ g = b) = \sum_{\substack{f(s) = a \\ g(s) = b}} \mu(s)$$

(«вероятности того, что f принимает значение a или $f = a$ и $g = b$ одновременно»). Назовем две случайные величины независимыми, если

$$P(f = a;\ g = b) = P(f = a)\, P(g = b)$$

при всех $a, b \in \mathbf{R}$. Доказать, что если нормированные случайные величины $f, g \in F_0(S)$ независимы, то они ортогональны.

Построить пример, показывающий, что обратное неверно.

Скалярное произведение величин $f, g \in F_0(S)$ называется их ковариацией, а косинус угла между ними — коэффициентом корреляции.

§ 6. Унитарные пространства

1. Определение. *Унитарным пространством называется комплексное линейное пространство $\mathbf{L}$ с эрмитовым положительно определенным скалярным произведением.*

Как в § 5, мы будем писать (l, m) вместо $g(l, m)$ и $|l|$ вместо $(l, l)^{1/2}$. Ниже мы убедимся, что $|l|$ является нормой на L в смысле § 10 ч. 1. Унитарные пространства, полные относительно этой нормы, называются также *гильбертовыми*. В частности, конечномерные унитарные пространства гильбертовы.

Из результатов, доказанных в § 3—4, следует, что:

а) всякое конечномерное унитарное пространство имеет ортонормированный базис, все векторы которого имеют длину 1;

б) поэтому оно изоморфно координатному унитарному пространству $\mathbf{C}^n$ $(n = \dim L)$ со скалярным произведением

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i \bar{y}_i, \quad |\vec{x}| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

Ряд свойств унитарных пространств близок к свойствам евклидовых, главным образом по следующей причине: если L — конечномерное унитарное пространство, то на его овеществлении $L_{\mathbf{R}}$ имеется (единственная) структура евклидова пространства, в которой норма $|l|$ вектора та же, что и в L. Существование видно из предыдущего абзаца: если $\{e_1, \ldots, e_n\}$ — ортонормированный базис L, а $\{e_1, ie_1, e_2, ie_2, \ldots, e_n, ie_n\}$ — соответствующий базис $L_{\mathbf{R}}$, то

$$\left| \sum_{j=1}^{n} x_j e_j \right|^2 = \sum_{j=1}^{n} |x_j|^2 = \sum_{j=1}^{n} ((\operatorname{Re} x_j)^2 + (\operatorname{Im} x_j)^2),$$

и выражение справа есть евклидов квадрат нормы вектора $\sum_{j=1}^{n} \operatorname{Re} x_j e_j + \sum_{j=1}^{n} \operatorname{Im} x_j (ie_j)$ в ортонормированном базисе $\{e_j, ie_j\}$. Единственность следует из п. 9 § 3.

Однако *скалярные произведения* в унитарном пространстве L и евклидовом $L_{\mathbf{R}}$ *не совпадают*: второе принимает только вещественные значения, а первое — комплексные. На самом деле, эрмитово скалярное произведение на комплексном пространстве приводит не только к ортогональной, но и к симплектической структуре на $L_{\mathbf{R}}$ с помощью следующей конструкции.

Временно мы возвращаемся к обозначению $g(l, m)$ для эрмитова скалярного произведения на L и положим

$$a(l, m) = \operatorname{Re} g(l, m),$$
$$b(l, m) = \operatorname{Im} g(l, m).$$

Тогда имеют место следующие факты:

2. Предложение. а) *$a(l, m)$ — симметричное, а $b(l, m)$ — антисимметричное скалярное произведение на $L_{\mathbf{R}}$; оба они инвариантны относительно умножения на i, т. е. канонической комплексной структуры на $L_{\mathbf{R}}$:*

$a(il, im) = a(l, m)$, $b(il, im) = b(l, m)$;

б) *a и b связаны следующими соотношениями:*

$$a(l, m) = b(il, m), \quad b(l, m) = -a(il, m);$$

в) *любая пара связанных соотношениями* б) *i-инвариантных форм a, b на $L_{\mathbf{R}}$, первая из которых симметрична, а вторая антисимметрична, определяет эрмитово скалярное произведение на L по формуле*

$$g(l,m)=a(l,\ m)+ib(l,m);$$

г) *форма g положительно определена тогда и только тогда, когда форма a положительно определена.*

Доказательство. Условие эрмитовой симметрии $g(l,m)=\overline{g(m,l)}$ равносильно тому, что

$$a(l,\ m)+ib(l,\ m)=a(m,\ l)-ib(m,\ l),$$

т. е. симметрии a и антисимметрии b. Условие $g(il,im)=i\bar{i}g(l,m)=g(l,m)$ равносильно i-инвариантности a и b. Условие $\mathbf{C}$-линейности g по первому аргументу означает $\mathbf{R}$-линейность и линейность относительно умножения на i, т. е.

$$a(il,\ m)+ib(il,\ m)=g(il,\ m)=ig(l,\ m)=-b(l,\ m)+ia(l,\ m),$$

откуда следуют соотношения б) и утверждение в). Наконец, $g(l,l)=a(l,l)$ в силу антисимметрии b, откуда следует г).

3. Следствие. *В прежних обозначениях, если g положительно определена и $\{e_1,\ \ldots,\ e_n\}$ — ортонормированный базис для g, то $\{e_1,\ \ldots,\ e_n,\ ie_1,\ \ldots,\ ie_n\}$ является ортонормированным базисом для a и симплектическим для b.*

Наоборот, если L — $2n$-мерное вещественное пространство с евклидовой формой a и симплектической b, а также базисом $\{e_1,\ \ldots,\ e_n,\ e_{n+1},\ \ldots,\ e_{2n}\}$, ортонормированным для a и симплектическим для b, то, введя на L комплексную структуру с помощью оператора

$$J(e_j)=e_{n+j},\ 1\leqslant j\leqslant n;\ J(e_j)=-e_{j-n},\ n+1\leqslant j\leqslant 2n,$$

и скалярное произведение $g(l,\ m)=a(l,\ m)+ib(l,\ m)$, мы получим комплексное пространство с положительно определенной эрмитовой формой, для которого $\{e_1,\ \ldots,\ e_n\}$ является ортонормированным базисом над $\mathbf{C}$.

Доказательство получается простой проверкой с помощью предложения п. 2, и мы оставляем его читателю.

Вернемся теперь к унитарным пространствам L. Комплексное неравенство Коши — Буняковского — Шварца имеет следующий вид:

4. Предложение. *Для любых $l_1,\ l_2\in L$*

$$|(l_1,\ l_2)|^2\leqslant|l_1|^2|l_2|^2,$$

причем равенство достигается тогда и только тогда, когда векторы l_1, l_2 пропорциональны.

Доказательство. Как в п. 2 § 5, для любых вещественных t имеем

$$|tl_1+l_2|^2=t^2|l_1|^2+2t\operatorname{Re}(l_1,\ l_2)+|l_2|^2\geqslant 0.$$

Случай $l_1 = 0$ тривиален Считая, что $l_1 \neq 0$, выводим отсюда, что

$$(\operatorname{Re}(l_1, l_2))^2 \leqslant |l_1|^2 |l_2|^2.$$

Но если $(l_1, l_2) = |(l_1, l_2)| e^{i\varphi}$, $\varphi \in \mathbf{R}$, то $\operatorname{Re}(e^{-i\varphi} l_1, l_2) = |(l_1, l_2)|$. Поэтому

$$|(l_1, l_2)|^2 \leqslant |e^{-i\varphi} l_1|^2 |l_2|^2 = |l_1|^2 |l_2|^2.$$

Строгое равенство здесь достигается тогда и только тогда, когда $|t_0 e^{-i\varphi} l_1 + l_2| = 0$ для подходящего $t_0 \in \mathbf{R}$, что завершает доказательство.

В точности так же, как в евклидовом случае, отсюда выводятся следствия:

5. Следствие (неравенство треугольника). *Для любых* $l_1, l_2, l_3 \in L$

$$|l_1 + l_2| \leqslant |l_1| + |l_2|,$$
$$|l_1 - l_3| \leqslant |l_1 - l_2| + |l_2 - l_3|.$$

6. Следствие. *Унитарная длина вектора* $|l|$ *является нормой на* L *в смысле определения в п.* 4 § 10 *ч.* 1.

(Здесь несколько изменяется проверка свойства $|al| = |a|\,|l|$:

$$|al| = (al, al)^{1/2} = (a\bar{a}(l, l))^{1/2} = |a|\,|l|.)$$

7. Углы. Пусть $l_1, l_2 \in L$ — ненулевые векторы. В силу предложения п. 4

$$0 \leqslant \frac{|(l_1, l_2)|}{|l_1|\,|l_2|} \leqslant 1.$$

Поэтому существует единственный угол φ, $0 \leqslant \varphi \leqslant \pi/2$, для которого

$$\cos\varphi = \frac{|(l_1, l_2)|}{|l_1|\,|l_2|}.$$

Однако в важнейших естественнонаучных моделях, использующих унитарные пространства, эта же величина $\frac{|(l_1, l_2)|}{|l_1|\,|l_2|}$ (точнее, ее квадрат) интерпретируется не как косинус угла, а как *вероятность*. Опишем вкратце постулаты квантовой механики, включающие такую трактовку.

8. Пространство состояний квантовой системы. В квантовой механике постулируется, что с такими физическими системами, как электрон, атом водорода и т. п., можно связать (неоднозначно!) математическую модель, состоящую из следующих данных.

а) Унитарное пространство $\mathscr{H}$, называемое *пространством состояний* системы. Такие пространства, рассматриваемые в стандартных учебниках, по большей части являются бесконечномерными гильбертовыми пространствами, которые реализуются как пространства функций на моделях «физического» пространства или пространства-времени. Конечномерные пространства $\mathscr{H}$

возникают, грубо говоря, как пространства *внутренних степеней свободы* системы, если она рассматривается как локализованная или если ее движением в физическом пространстве можно так или иначе пренебречь. Таково двумерное унитарное пространство «спиновых состояний» электрона, к которому мы еще вернемся.

б) Лучи, т. е. одномерные комплексные подпространства в $\mathscr{H}$, называются (чистыми) *состояниями* системы.

Вся информация о состоянии системы в фиксированный момент времени определяется заданием луча $L \subset \mathscr{H}$ или ненулевого вектора $\psi \in L$, который называется иногда *ψ-функцией*, отвечающей этому состоянию, или *вектором состояния*.

Фундаментальный постулат о том, что ψ-функции образуют комплексное линейное пространство, называется *принципом суперпозиции*, а линейная комбинация $\sum_{j=1}^{n} a_j\psi_j$, $a_j \in \mathbf{C}$, описывает *суперпозицию* состояний $\psi_1, \ldots, \psi_n$. Заметим, что, поскольку физический смысл имеют только лучи $\mathbf{C}\psi_j$, а не сами векторы ψ_j, коэффициентам a_j также нельзя приписать однозначно определенного смысла. Однако, если выбирать ψ_j нормированными, $|\psi_j|^2 = 1$, и линейно независимыми, а также нормировать $\sum_{j=1}^{n} a_j\psi_j$, то произвол в выборе вектора ψ_j в своем луче сводится к умножениям на числа $e^{i\varphi_j}$, которые называются *фазовыми множителями*; таков же будет произвол в выборе коэффициентов a_j, которые мы сможем тогда сделать вещественными и неотрицательными, что вместе с условием нормировки $\left|\sum_{j=1}^{n} a_j\psi_j\right| = 1$ позволяет определить их однозначно.

Сильно идеализированные предположения о связи этой схемы с реальностью состоят в том, что у нас имеются физические приборы A_ψ («печки»), способные приготовлять много экземпляров нашей системы в мгновенных состояниях ψ (точнее, $\mathbf{C}\psi$) для различных $\psi \in \mathscr{H}$. Сверх того, имеются физические приборы B_χ («фильтры»), на вход которых подаются системы в состоянии ψ, на выходе обнаруживаются они же в некотором (возможно, другом) состоянии χ, или же не обнаруживается ничего (система «не проходит» через фильтр B).

Второй основной (после принципа суперпозиции) постулат квантовой механики состоит в том, что:

система, приготовленная в состоянии $\psi \in \mathscr{H}$, может быть сразу же после этого обнаружена в состоянии $\chi \in \mathscr{H}$ с вероятностью

$$\frac{|(\psi, \chi)|^2}{|\psi|^2 |\chi|^2} = \cos^2\theta, \text{ где } \theta \text{ — угол между } \psi \text{ и } \chi.$$

В дальнейшем, по мере введения дополнительных геометрических понятий, мы уточним математическое описание «печек» и «фильтров». Сверх того, мы объясним, что произойдет, если приготовленную в состоянии ψ систему ввести в фильтр не сразу, а по истечении времени t: оказывается, что в промежутке состояние ψ, а вместе с ним и скалярное произведение (ψ, χ) будет меняться, и это изменение также прекрасно описывается в терминах линейной алгебры.

Если ψ, χ нормированы, то указанная выше вероятность равна $|(\psi, \chi)|^2$, а само скалярное произведение (ψ, χ), являющееся комплексным числом, называется *амплитудой вероятности* (перехода от ψ к χ). Заметим, что физики вслед за Дираком обычно рассматривают скалярные произведения, антилинейные по *первому* аргументу, и записывают наше (ψ, χ) в виде $<\chi|\psi>$, так что начальное и конечное состояние системы расположены *справа налево*. Скобки $< \ >$ по английски называются «bracket». Соответственно, Дирак называет символ $|\psi>$ «кет-вектором», а символ $<\chi|$ — соответствующим «бра-вектором». С математической точки зрения, $|\psi>$ есть элемент $\mathscr{H}$, а $<\psi|$ — соответствующий ему элемент пространства антилинейных функционалов $\overline{\mathscr{H}}^*$, и $<\chi|\psi>$ есть значение χ на ψ.

Если ψ, χ ортогональны, т. е. $(\psi, \chi) = 0$, то систему, приготовленную в состоянии ψ, нельзя будет (сразу же после приготовления) обнаружить в состоянии χ, т. е. она не пройдет через фильтр B_χ (наоборот, через фильтр B_ψ она пройдет с достоверностью). Во всех остальных случаях ненулевая вероятность перехода от ψ к χ имеется.

Элементы любого ортонормированного базиса $\{\psi_1, \ldots, \psi_n\}$ образуют набор *базисных состояний* системы. Предположим, что у нас есть фильтры $B_{\psi_1}, \ldots, B_{\psi_n}$. Многократно пропуская через них системы, приготовленные в состоянии $\psi = \sum_{i=1}^{n} a_i \psi_i$, $0 \leqslant a_i \leqslant 1$ (вектор считается нормированным), мы обнаружим ψ_i с вероятностью a_i^2. Таким образом, коэффициенты этой линейной комбинации могут быть измерены экспериментально, однако в принципиально статистическом опыте. Это одна из причин, по которым квантовомеханические измерения требуют обработки большого статистического материала. Впрочем, часто системы в состоянии ψ идут в фильтр «потоком» и на выходе вероятности a_i^2 получаются в виде интенсивностей, чего-то вроде «спектральных линий»; эти интенсивности сами по себе уже являются результатом статистического усреднения. В дальнейшем мы уточним связь этой схемы с теорией спектров линейных операторов.

9. Правила Фейнмана. Пусть в $\mathscr{H}$ выбран ортонормированный базис $\{\psi_1, \ldots, \psi_n\}$. Для любого вектора состояния $\psi \in \mathscr{H}$ имеем

$$\psi = \sum_{i=1}^{n} (\psi, \psi_i)\, \psi_i,$$

откуда

$$(\psi, \chi) = \sum_{i=1}^{n} (\psi, \psi_i)(\psi_i, \chi).$$

Аналогично, $(\psi, \psi_i) = \sum_{j=1}^{n} (\psi, \psi_j)(\psi_j, \psi_i)$: подставляя эту формулу в предыдущую, получим

$$(\psi, \chi) = \sum_{i_1, i_2=1}^{n} (\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2})(\psi_{i_2}, \chi)$$

и вообще для любого $m \geqslant 1$

$$(\psi, \chi) = \sum_{i_1, \ldots, i_m=1}^{n} (\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2}) \ldots (\psi_{i_m}, \chi).$$

Эти простые формулы линейной алгебры можно интерпретировать, по Фейнману, как законы «комплексной теории вероятностей», относящиеся к амплитудам вместо вероятностей. Именно, будем рассматривать последовательности типа $(\psi, \psi_{i_1}, \psi_{i_2}, \ldots, \psi_{i_m}, \chi)$ как «классические траектории» системы, последовательно пробегающей состояния в скобках, а число $(\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2}) \ldots (\psi_{i_m}, \chi)$ — как амплитуду вероятности перехода из ψ в χ вдоль соответствующей классической траектории. Эта амплитуда является произведением амплитуд переходов вдоль последовательных отрезков траектории.

Тогда приведенная выше формула для (ψ, χ) означает, что *эта амплитуда перехода есть сумма амплитуд перехода от ψ к χ по всевозможным классическим траекториям* («одинаковой длины»).

Бесконечномерный и более рафинированный вариант этого замечания, в котором основную роль играют пространственно-временные (или энергетически-импульсные) наблюдаемые, Р. Фейнман положил в основу своей полуэвристической техники выражения амплитуд через «континуальные интегралы по классическим траекториям». Пространство траекторий является бесконечномерным функциональным пространством, и математикам до сих пор не удалось построить общую теорию, в которой были бы оправданы все замечательные вычисления физиков.

10. Расстояния. Расстояние между подмножествами в унитарном пространстве L можно определить точно так же, как в евклидовом:

$$d(U, V) = \inf \{ |l_1 - l_2| \mid l_1 \in U, l_2 \in V \}.$$

Расстояние от вектора l до подпространства L_0 также равно длине ортогональной проекции l на $L_0^{\perp}$. Доказательство ничем не отличается от евклидова случая. В частности, если $\{e_1, \ldots, e_m\}$ — ортонормированный базис L_0, то

$$d(l, L_0) = \left| l - \sum_{i=1}^{m} (l, e_i) e_i \right|,$$

как в евклидовом случае, и

$$\left| \sum_{i=1}^{m} (l, e_i) e_i \right|^2 = \sum_{i=1}^{m} |(l, e_i)|^2 \leqslant |l|^2$$

по теореме Пифагора.

11. Приложение к пространствам функций. Как в § 5 и 4, мы можем вывести неравенства для комплекснозначных функций:

$$\left|\int_a^b f(x)\,\bar{g}(x)\,dx\right|^2 \leqslant \int_a^b |f(x)|^2\,dx \int_a^b |g(x)|^2\,dx,$$

$$\left(\int_a^b |f(x)+g(x)|^2\,dx\right)^{1/2} \leqslant \left(\int_a^b |f(x)|^2\,dx\right)^{1/2} + \left(\int_a^b |g(x)|^2\,dx\right)^{1/2}.$$

а также для их коэффициентов Фурье. Рассматривая функции на отрезке $[0, 2\pi]$ и полагая

$$a_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)\,e^{-inx}\,dx,$$

получаем, что в пространстве со скалярным произведением $\int_0^{2\pi} f(x)\,\bar{g}(x)\,dx$ сумма

$$f_N(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} a_n e^{inx}$$

является ортогональной проекцией f на пространство многочленов Фурье «степени $\leqslant N$» и минимизирует среднеквадратичное отклонение f от этого пространства. В частности,

$$\sum_{n=-N}^{N} |a_n|^2 \leqslant \int_0^{2\pi} |f(x)|^2\,dx,$$

так что ряд $\sum_{n=-\infty}^{\infty} |a_n|^2$ сходится.

§ 7. Ортогональные и унитарные операторы

1. Пусть L — линейное пространство со скалярным произведением g. Множество всех изометрий $f: L \to L$, т. е. обратимых линейных операторов с условием

$$g(f(l_1),\ f(l_2)) = g(l_1,\ l_2)$$

для всех $l_1, l_2 \in L$, очевидно, образует группу. Если L — евклидово пространство, такие операторы называются *ортогональными*, а если L унитарно, то *унитарными*. Симплектические изометрии будут рассмотрены позже.

2. Предложение. *Пусть L — конечномерное линейное пространство с невырожденным скалярным произведением (,), симметричным или эрмитовым. Для того чтобы оператор $f: L \to L$ был* *изометрией, необходимо и достаточно, чтобы выполнялось любое из следующих условий*:

а) $(f(l), f(l)) = (l, l)$ *для всех* $l \in L$ (*здесь предполагается, что характеристика поля скаляров отлична от двух*);

б) *пусть* $\{e_1, \ldots, e_n\}$ — *базис в* L *с матрицей Грама* G, A — *матрица оператора* f *в этом базисе. Тогда*

$$A^t G A = G, \quad \text{или} \quad A^t G \bar{A} = G;$$

в) f *переводит некоторый ортонормированный базис в ортонормированный базис*;

г) *если сигнатура скалярного произведения равна* (p, q), *то матрица оператора* f *в любом ортонормированном базисе* $\{e_1, \ldots, e_p, e_{p+1}, \ldots, e_{p+q}\}$ *с* $(e_i, e_i) = +1$ *при* $i \leqslant p$ *и* $(e_i, e_i) = -1$ *при* $p+1 \leqslant i \leqslant p+q$ *удовлетворяет условию*

$$A^t \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix} A = \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix}$$

или

$$A^t \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix} \bar{A} = \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix}$$

в симметричном и эрмитовом случае соответственно.

Доказательство. а) В симметричном случае это утверждение следует из п. 9 § 3: если f сохраняет квадратичную форму $(l, l) = q(l)$, то f сохраняет и ее поляризацию

$$(l, m) = \frac{1}{2}[q(l+m) - q(l) - q(m)].$$

В эрмитовом случае имеем аналогично

$$\operatorname{Re}(l, m) = \frac{1}{2}[q(l+m) - q(l) - q(m)]$$

и предложение п. 2 § 6 показывает, что (l, m) однозначно восстанавливается по $\operatorname{Re}(l, m)$ по формуле

$$(l, m) = \operatorname{Re}(l, m) - i \operatorname{Re}(il, m)$$

и потому f сохраняет (l, m).

б) Если f — изометрия, то матрицы Грама базисов $\{e_1, \ldots, e_n\}$ и $\{f(e_1), \ldots, f(e_n)\}$ совпадают. Но последняя матрица Грама равна $A^t G A$ в симметричном и $A^t G \bar{A}$ в эрмитовом случае. Наоборот, если f переводит базис $\{e_1, \ldots, e_n\}$ в $\{e'_1, \ldots, e'_n\}$ и матрицы Грама базисов $\{e_i\}$ и $\{e'_i\}$ совпадают, то f — изометрия в силу формул координатной записи скалярного произведения из п. 2 § 2.

в), г). Эти утверждения являются частными случаями предыдущих.

Из предложения п. 2 следует, что ортогональные (соответственно унитарные) операторы — это операторы, которые в одном (и потому в любом) *ортонормированном* базисе задаются *ортого-*

нальными (соответственно *унитарными*) матрицами, т. е. матрицами U, которые удовлетворяют соотношениям

$$UU^t = E_n \text{ или } U\bar{U}^t = E_n.$$

Множества таких матриц размера $n \times n$ были введены впервые в § 4 ч. 1; они обозначались $\mathrm{O}(n)$ и $\mathrm{U}(n)$ соответственно. Аналогично, матрицы изометрий в ортонормированных базисах сигнатур (p, q), удовлетворяющие условиям г) предложения 2, обозначаются $\mathrm{O}(p, q)$ и $\mathrm{U}(p, q)$; при p, $q \neq 0$ они называются иногда *псевдоортогональными* и *псевдоунитарными* соответственно. В этом параграфе мы будем заниматься только группами $\mathrm{O}(n)$ и $\mathrm{U}(n)$. Фундаментальная для физики *группа Лоренца* $\mathrm{O}(1, 3)$ будет изучена в § 10.

3. Группы $\mathrm{U}(1)$, $\mathrm{O}(1)$ и $\mathrm{O}(2)$. Из определения немедленно следует, что

$$\mathrm{U}(1) = \{a \in \mathbf{C} \mid |a| = 1\} = \{e^{i\varphi} \mid \varphi \in \mathbf{R}\},$$
$$\mathrm{O}(1) = \{\pm 1\} = \mathrm{U}(1) \cap \mathbf{R}.$$

Далее, если $U \in \mathrm{O}(n)$, то $UU^t = E_n$, откуда $(\det U)^2 = 1$ и $\det U = \pm 1$. Если $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — ортогональная матрица с определителем -1, то $\begin{pmatrix} a & b \\ -c & -d \end{pmatrix}$ — ортогональная матрица с определителем 1, принадлежащая $\mathrm{SO}(2)$. Матрицы из $\mathrm{SO}(2)$ имеют вид

$$\left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \, ad - bc = a^2 + b^2 = c^2 + d^2 = 1, \ ac + bd = 0\right\}.$$

Очевидно, любую такую матрицу можно представить в виде

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \quad \varphi \in [0, 2\pi),$$

т. е. она задает евклидов поворот на угол φ. Отображение

$$\mathrm{U}(1) \to \mathrm{SO}(2) : e^{i\varphi} \mapsto \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

является изоморфизмом. Его геометрический смысл объясняется следующим замечанием: овеществление одномерного унитарного пространства $(\mathbf{C}^1, |z|^2)$ есть двумерное евклидово пространство $(\mathbf{R}^2, x_1^2 + x_2^2)$, а овеществление унитарного преобразования $z \mapsto e^{i\varphi}z$ задается матрицей поворота на угол φ.

В § 9 мы построим значительно менее тривиальный эпиморфизм $\mathrm{SU}(2) \to \mathrm{SO}(3)$ с ядром $\{\pm 1\}$.

Повороты $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ на угол $\varphi \neq 0$, π не имеют собственных векторов в $\mathbf{R}^2$ и потому не диагонализируемы. Наоборот, все матрицы $U \in \mathrm{O}(2)$ с $\det U = -1$ диагонализируемы. Точнее говоря, характеристический многочлен матрицы

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ -\sin\varphi & -\cos\varphi \end{pmatrix}$$

равен t^2-1 и имеет корни ± 1. Легко проверить непосредственно, что соответствующие этим корням собственные подпространства ортогональны; ниже это будет доказано в гораздо большей общности. Поэтому любой оператор из $\mathrm{O}(2)$ с $\det U=-1$ является *отражением* относительно некоторой прямой: он действует тождественно на этой прямой и меняет знак векторов, ортогональных к ней.

Пользуясь этой информацией, мы можем теперь установить структуру общих ортогональных и унитарных операторов.

4. Теорема. а) *Для того чтобы оператор f в унитарном пространстве был унитарным, необходимо и достаточно, чтобы он диагонализировался в ортонормированном базисе и имел спектр, расположенный на единичной окружности в* $\mathbf{C}$.

б) *Для того чтобы оператор f в евклидовом пространстве был ортогональным, необходимо и достаточно, чтобы его матрица в подходящем ортонормированном базисе имела вид*

$$\begin{pmatrix} A(\varphi_1) & & & & & & & & & & \\ & \ddots & & & & & & & & 0 & \\ & & A(\varphi_m) & & & & & & & & \\ & & & 1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & 1 & & & & & \\ & 0 & & & & & -1 & & & & \\ & & & & & & & \ddots & & & \\ & & & & & & & & -1 & & \end{pmatrix}, \quad A(\varphi)=\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

где на пустых местах стоят нули.

в) *Собственные векторы ортогонального или унитарного оператора, отвечающие разным собственным значениям, ортогональны.*

Доказательство а) Достаточность утверждения очевидна: если $U=\mathrm{diag}(\lambda_1,\ \ldots,\ \lambda_n)$, $|\lambda_i|^2=1$, то $U\bar{U}^t=E_n$, так что U — матрица унитарного оператора. Наоборот, пусть f — унитарный оператор, λ — его собственное значение, L_λ — соответствующее собственное подпространство. По предложению п. 2 § 3 имеем $L=L_\lambda\oplus L_\lambda^\perp$. Подпространство L_λ одномерно, f-инвариантно, и ограничение f на L_λ является одномерным унитарным оператором, поэтому $\lambda\in\mathrm{U}(1)$, т. е. $|\lambda|^2=1$. Если мы покажем, что подпространство $L_\lambda^\perp$ также f-инвариантно, то индукцией по $\dim L$ отсюда можно будет вывести, что L разлагается в прямую сумму f-инвариантных попарно ортогональных одномерных подпространств, что докажет требуемое.

В самом деле, если $l_0 \in L_\lambda$, $l_0 \neq 0$ и $(l_0, l) = 0$, то

$$(l_0, f(l)) = (f(\lambda^{-1} l_0), f(l)) = (\lambda^{-1} l_0, l) = \lambda^{-1}(l_0, l) = 0,$$

так что $f(l) \in L_\lambda^\perp$.

б) В ортогональном случае рассуждения аналогичны: непосредственно проверяется достаточность условия и затем проводится индукция по $\dim L$. Случаи $\dim L = 1, 2$ разобраны в предыдущем пункте. Если $\dim L \geqslant 3$ и f имеет вещественное собственное значение λ, нужно снова положить $L = L_\lambda \oplus L_\lambda^\perp$ и рассуждать, как выше (заметим, что здесь обязательно $\lambda = \pm 1$). Наконец, если f не имеет вещественных собственных значений, то следует выбрать двумерное f-инвариантное подпространство $L_0 \subset L$, которое существует по предложению п. 16 § 12 ч. 1. На нем матрица ограничения f в любом ортонормированном базисе будет иметь вид $A(\varphi)$ в силу предыдущего пункта. Поэтому остается проверить, что подпространство $L_0^\perp$ также f-инвариантно. Действительно, если $(l_0, l) = 0$ для всех $l_0 \in L_0$, то

$$(l_0, f(l)) = (f(f^{-1}(l_0)), f(l)) = (f^{-1}(l_0), l) = 0,$$

ибо $f^{-1}(l_0) \in L_0$ для всех $l_0 \in L_0$. Это завершает доказательство.

в) Пусть $f(l_i) = \lambda_i l_i$, $i = 1, 2$. Тогда

$$(l_1, l_2) = (f(l_1), f(l_2)) = \lambda_1 \overline{\lambda_2}(l_1, l_2).$$

Так как $|\lambda_i|^2 = 1$, при $\lambda_1 \neq \lambda_2$ имеем $\lambda_1 \overline{\lambda_2} \neq 1$. Следовательно, $(l_1, l_2) = 0$. Это рассуждение применимо одновременно к унитарному и ортогональному случаю. Доказательство окончено.

5. Следствие («теорема Эйлера»). *В трехмерном евклидовом пространстве любое ортогональное отображение f, не меняющее ориентацию (т. е. элемент группы* $SO(3)$*), является вращением относительно некоторой оси.*

Доказательство. Так как характеристический многочлен f имеет степень 3, у него обязательно есть вещественный корень. Если он единственный, то он должен быть равен 1, ибо $\det f = 1$. Если есть больше одного вещественного корня, то все корни вещественные, и возможны комбинации $(1, 1, 1)$ или $(1, -1, -1)$. В любом случае собственное значение 1 имеется. Соответствующее собственное подпространство является осью вращения, а в ортогональной к нему плоскости индуцируется элемент $SO(2)$, т. е. вращение на некоторый угол.

§ 8. Самосопряженные операторы

1. В первой части мы видели, что простейший и наиболее важный класс линейных операторов образуют диагонализируемые операторы. Оказывается, что в евклидовых и унитарных пространствах совершенно особую роль играют операторы с вещественным спектром, диагонализируемые в некотором ортонормированном базисе. Иными словами, эти операторы осуществляют вещественные растяжения пространства вдоль системы попарно ортогональных направлений.

Пусть $\{e_1, \ldots, e_n\}$ — ортонормированный базис в L и $f\colon L \to L$ — оператор, для которого $f(e_i) = \lambda_i e_i$, $\lambda_i \in \mathbf{R}$, $i = 1, \ldots, n$. Нетрудно убедиться, что он обладает следующим простым свойством:

$$(f(l_1), l_2) = (l_1, f(l_2)) \text{ для всех } l_1, l_2 \in L. \tag{1}$$

Действительно,

$$\left(f\left(\sum x_i e_i\right), \sum y_j e_j\right) = \sum \lambda_i x_i y_i \text{ или } \sum \lambda_i x_i \bar{y}_i,$$
$$\left(\sum x_i e_i, f\left(\sum y_j e_j\right)\right) = \sum \lambda_i x_i y_i \text{ или } \sum \lambda_i x_i \bar{y}_i$$

(в унитарном случае вещественность λ_i использовалась во второй формуле). Операторы со свойством (1) называются *самосопряженными*, и мы установили, что *операторы с вещественным спектром, диагонализируемые в ортонормированном базисе, самосопряжены.* Вскоре мы докажем и обратное утверждение, но сначала исследуем свойство самосопряженности более систематично.

2. Сопряженные операторы в пространствах с билинейной формой. В первой части курса мы показали, что для любого линейного отображения $f\colon L \to M$ существует единственное линейное отображение $f^*\colon M^* \to L^*$, для которого

$$(f^*(m^*), l) = (m^*, f(l)),$$

где $m^* \in M^*$, $l \in L$ и где скобки означают канонические билинейные отображения $L^* \times L \to \mathscr{K}$, $M^* \times M \to \mathscr{K}$.

В частности, при $M = L$ оператору $f\colon L \to L$ отвечает оператор $f^*\colon L^* \to L^*$. Предположим теперь, что на L имеется невырожденная билинейная форма $g\colon L \times L \to \mathscr{K}$, определяющая изоморфизм $\tilde{g}\colon L \to L^*$. Тогда, отождествив L^* с L посредством $\tilde{g}^{-1}$, мы можем рассмотреть f^*, точнее $\tilde{g}^{-1} \circ f^* \circ \tilde{g}$, как оператор на L. Мы по-прежнему будем обозначать его f^* (точнее было бы писать, например, f_g^* но f^* в старом смысле в этом параграфе больше не будет фигурировать). Очевидно, новый оператор f^* однозначно определяется формулой

$$g(f^*(l), m) = g(l, f(m)).$$

Он по-прежнему называется сопряженным с f (относительно скалярного произведения g).

В полуторалинейном случае $\tilde{g}$ определяет изоморфизм L с $\bar{L}^*$, а не с L^*. Поэтому на L с помощью этого изоморфизма следует переносить оператор $\bar{f}^*\colon \bar{L}^* \to \bar{L}^*$, который определяется как $\bar{f}^*(m) = \overline{f^*(m)}$. Перенесенный оператор $\tilde{g}^{-1} \circ \bar{f}^* \circ \tilde{g}\colon L \to L$ линеен. Следовало бы обозначить его f^+, но мы сохраним более традиционное обозначение f^*. Тогда и в полуторалинейном случае будет справедлива формула

$$g(f^*(l), m) = g(l, f(m)).$$

Операция $f \mapsto f^*$ линейна, если g билинейна, и антилинейна, если g полуторалинейна.

Операторы $f\colon L \to L$ со свойством $f^* = f$ в евклидовых и конечномерных унитарных пространствах называются *самосопряженными*, в евклидовом случае — также *симметричными*, а в унитарном — *эрмитовыми*. Эта терминология объясняется следующим простым замечанием.

3. Предложение. *Если оператор $f\colon L \to L$ в ортонормированном базисе задается матрицей A, то оператор f^* задается в этом же базисе матрицей A^t (евклидов случай) или $\bar{A}^t$ (унитарный случай).*

В частности, оператор самосопряжен тогда и только тогда, когда его матрица в ортонормированном базисе симметрична или эрмитова.

Доказательство. Обозначая скалярное произведение в L скобками, а векторы — столбцами их координат в ортонормированном базисе, имеем

$$(f(\vec{x}), y) = (A\vec{x})^t y = (\vec{x}^t A^t)\vec{y} = \vec{x}^t (A^t\vec{y}) = (\vec{x}, f^*(\vec{y}))$$

(евклидов случай). Отсюда и следует, что матрица f^* равна A^t. Унитарный случай разбирается аналогично.

4. Самосопряженные операторы и скалярные произведения. Пусть L — пространство с симметричным или эрмитовым скалярным произведением $(\ ,\ )$. Для любого линейного оператора $f\colon L \to L$ мы можем определить новое скалярное произведение $(\ ,\ )_f$ на L, положив

$$(l_1, l_2)_f = (f(l_1), l_2).$$

Предположим, что L невырождено, так что мы можем пользоваться понятием сопряженного оператора. Тогда

$$(l_2, l_1)_f = (f(l_2), l_1) = (l_2, f^*(l_1)) = (f^*(l_1), l_2) = (l_1, l_2)_{f^*}$$

в евклидовом случае, и аналогично

$$\overline{(l_2, l_1)_f} = \overline{(f(l_2), l_1)} = \overline{(l_2, f^*(l_1))} = (f^*(l_1), l_2) = (l_1, l_2)_{f^*}$$

в унитарном. Следовательно, *если оператор f самосопряжен, то построенная по нему новая метрика $(l_1, l_2)_f$ будет по-прежнему симметричной или эрмитовой.* Верно и обратное, как нетрудно убедиться прямо или с помощью предложения п. 3.

Таким образом, мы установили биекцию между множествами самосопряженных операторов, с одной стороны, и симметричных скалярных произведений в пространстве, где одно невырожденное скалярное произведение задано, — с другой. В евклидовом и унитарном случае после выбора ортонормированного базиса соответствие легко описывается на матричном языке: матрица Грама $(\ ,\ )_f$ транспонирована к матрице отображения f.

Теперь мы докажем основную теорему о самосопряженных операторах, параллельную теореме п. 4 § 7 об ортогональных и унитарных операторах и тесно с ней связанную.

5. Теорема. а) *Для того чтобы оператор f в конечномерном евклидовом или унитарном пространстве был самосопряжен, необходимо и достаточно, чтобы он диагонализировался в ортонормированном базисе и имел вещественный спектр.*

б) *Собственные векторы самосопряженного оператора, отвечающие разным собственным значениям, ортогональны.*

Доказательство. а) Достаточность мы проверили в начале этого параграфа. Вещественность спектра в унитарном случае устанавливается просто: пусть λ — собственное значение оператора f, $l \in L$ — соответствующий собственный вектор. Тогда

$$\lambda(l, l) = (f(l), l) = (l, f(l)) = \bar{\lambda}(l, l),$$

откуда $\lambda = \bar{\lambda}$, ибо $(l, l) \neq 0$. Ортогональный случай сводится к унитарному следующим приемом: рассмотрим комплексифицированное пространство $L^{\mathbf{C}}$ и введем на нем полуторалинейное скалярное произведение по формуле

$$(l_1 + il_2, l_3 + il_4) = (l_1, l_3) + (l_2, l_4) + i(l_2, l_3) - i(l_1, l_4).$$

Легкая прямая проверка показывает, что $L^{\mathbf{C}}$ превращается в унитарное пространство, а $f^{\mathbf{C}}$ — в эрмитов оператор на нем. Спектр оператора $f^{\mathbf{C}}$ совпадает со спектром оператора f, ибо в любом $\mathbf{R}$-базисе L, являющемся в то же время $\mathbf{C}$-базисом $L^{\mathbf{C}}$, f и $f^{\mathbf{C}}$ задаются одинаковыми матрицами. Поэтому спектр оператора f веществен.

Дальше оба случая можно рассматривать параллельно и провести индукцию по $\dim L$. Случай $\dim L = 1$ тривиален. При $\dim L > 1$ выберем собственное значение λ и отвечающее ему собственное подпространство L_0, затем положим $L_1 = L_0^{\perp}$. По предложению п. 2 § 3 имеем $L = L_0 \oplus L_1$. Подпространство L_1 инвариантно относительно f, потому что если $l_0 \in L_0$, $l_0 \neq 0$ и $l \in L_1$, т. е. $(l_0, l) = 0$, то

$$(l_0, f(l)) = (f(l_0), l) = \lambda(l_0, l) = 0,$$

так что $f(l) \in L$. По индуктивному предположению ограничение f на L_1 диагонализируется в ортонормированном базисе L_1. Добавив к нему вектор $l_0 \in L_0$, $|l_0| = 1$, получим требуемый базис в L.

б) Пусть $f(l_1) = \lambda_1 l_1$, $f(l_2) = \lambda_2 l_2$. Тогда

$$\lambda_1(l_1, l_2) = (f(l_1), l_2) = (l_1, f(l_2)) = \lambda_2(l_1, l_2),$$

откуда следует, что если $\lambda_1 \neq \lambda_2$, то $(l_1, l_2) = 0$.

6. Следствие. *Любая вещественная симметричная или комплексная эрмитова матрица имеет вещественный спектр и диагонализируема.*

Доказательство. Построим по матрице A самосопряженный оператор в координатном пространстве $\mathbf{R}^n$ или $\mathbf{C}^n$ с канонической евклидовой или унитарной метрикой и применим теорему п. 5. Из нее видно даже больше: матрицу X такую, что $X^{-1}AX$ диагональна, можно найти в $\mathrm{O}(n)$ или в $\mathrm{U}(n)$ соответственно.

7. Следствие. *Отображение* $\exp: \mathrm{u}(n) \to \mathrm{U}(n)$ *сюръективно.*

Доказательство Алгебра Ли $\mathrm{u}(n)$ состоит из антиэрмитовых матриц (см. § 4 ч. 1), а любая антиэрмитова матрица имеет вид iA, где A — эрмитова матрица. Чтобы решить относительно A уравнение $\exp(iA) = U$, где $U \in \mathrm{U}(n)$, реализуем U как унитарный оператор f в эрмитовом координатном пространстве $\mathbf{C}^n$. После этого по теореме п. 4 § 7 найдем в $\mathbf{C}^n$ новый ортонормированный базис $\{e_1, \ldots, e_n\}$, в котором матрица оператора f имеет вид $\mathrm{diag}(e^{i\varphi_1}, \ldots, e^{i\varphi_n})$, зададим в этом базисе оператор g матрицей $\mathrm{diag}(\varphi_1, \ldots, \varphi_n)$ и обозначим через A матрицу оператора g в исходном базисе. Очевидно, $\exp(ig) = f$ и $\exp(iA) = U$.

8. Следствие. а) *Пусть g_1, g_2 — две ортогональные или эрмитовы формы в конечномерном пространстве L, и одна из них, скажем g_1, положительно определена. Тогда в пространстве L существует базис, матрица Грама которого относительно g_1 единична, а относительно g_2 диагональна и вещественна.*

б) *Пусть g_1, g_2 — две вещественные симметричные или комплексные эрмитово симметричные формы относительно переменных $x_1, \ldots, x_n$; $y_1, \ldots, y_n$, и g_1 положительно определена. Тогда с помощью невырожденной линейной замены переменных (общей для $\vec{x}$ и $\vec{y}$) эти две формы можно привести к виду*

$$g_1(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i y_i; \quad g_2(\vec{x}, \vec{y}) = \sum_{i=1}^{n} \lambda_i x_i y_i, \quad \lambda_i \in \mathbf{R},$$

или

$$g_1(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i \bar{y}_i; \quad g_2(\vec{x}, \vec{y}) = \sum_{i=1}^{n} \lambda_i x_i \bar{y}_i, \quad \lambda_i \in \mathbf{R}.$$

Доказательство. Очевидно, обе формулировки эквивалентны. Чтобы доказать их, рассмотрим (L, g_1) как ортогональное или унитарное пространство, переобозначим $g_1(l_1, l_2)$ через (l_1, l_2) и представим $g_2(l_1, l_2)$ в виде $(l_1, l_2)_f$, где $f: L \to L$ — некоторый самосопряженный оператор, как это было сделано в п. 4. После этого найдем ортонормированный базис в L, в котором f диагонализируется. По замечанию в конце п. 4 этот базис будет удовлетворять требованиям следствия (точнее, утверждения а)).

9. Ортогональные проекторы. Пусть L — линейное пространство над $\mathscr{K}$ и пусть дано его разложение в прямую сумму: $L = L_1 \oplus L_2$. Как было показано в ч. 1, оно определяет два проектора $p_i: L \to L$ таких, что $\mathrm{Im}\, p_i = L_i$, $\mathrm{id}_L = p_1 + p_2$, $p_1 p_2 = p_2 p_1 = 0$, $p_i^2 = p_i$. Собственные значения проекторов равны 0 или 1.

Если L — евклидово или унитарное пространство и $L_2 = L_1^{\perp}$, то соответствующие ортогональные проекторы диагонализируются в ортонормированном базисе L — объединении таких базисов L_1 и L_2 — и потому самосопряжены. Наоборот, *любой самосопряженный* *проектор p есть оператор ортогонального проектирования на подпространство.* Действительно, $\operatorname{Ker} p$ и $\operatorname{Im} p$ натянуты на собственные векторы p, отвечающие собственным значениям 0 и 1 соответственно, так что $\operatorname{Ker} p$ и $\operatorname{Im} p$ ортогональны по теореме п. 5 и $L = \operatorname{Ker} p \oplus \operatorname{Im} p$.

Далее, если самосопряженный оператор f диагонализируется в ортонормированном базисе $\{e_i\}$, $f(e_i) = \lambda_i e_i$, и p_i — ортогональный проектор L на подпространство, натянутое на e_i, то

$$f = \sum_{i=1}^{n} \lambda_i p_i. \tag{2}$$

Эта формула называется *спектральным разложением* оператора f.

Можно считать, что λ_i пробегает только попарно различные собственные значения, а p_i есть оператор ортогонального проектирования на полное корневое подпространство $L(\lambda_i)$; формула (2) останется верной.

Теорема п. 5 обобщается также на ограниченные по норме (и, с осложнениями, на неограниченные) самосопряженные операторы в бесконечномерных гильбертовых пространствах. Однако это обобщение требует очень нетривиального изменения некоторых основных понятий. Главные проблемы связаны со структурой спектра: в конечномерном случае λ является собственным значением f тогда и только тогда, когда оператор $\lambda \operatorname{id} - f$ необратим, тогда как в бесконечномерном случае множество точек необратимости оператора $\lambda \operatorname{id} - f$ может быть больше множества собственных значений f: для неизолированных в спектре точек λ_0 собственных векторов, вообще говоря, нет. С другой стороны, именно множество точек необратимости оператора $\lambda \operatorname{id} - f$ служит правильным обобщением спектра в бесконечномерном случае. Эта нехватка собственных векторов требует изменения многих формулировок. Основной результат является обобщением формулы (2), где, однако, суммирование заменяется интегрированием.

Мы ограничимся описанием нескольких важных принципов на примерах, где эти затруднения не возникают.

10. Формально сопряженные дифференциальные операторы. Рассмотрим какое-нибудь пространство вещественных функций на отрезке $[a, b]$ со скалярным произведением

$$(f, g) = \int_a^b f(x) g(x)\, dx.$$

Предположим, что оператор $\frac{d}{dx}$ переводит его в себя. Согласно формуле интегрирования по частям

$$\left(\frac{df}{dx}, g\right) + \left(f, \frac{dg}{dx}\right) = fg\,\Big|_a^b = f(b)g(b) - f(a)g(a).$$

Поэтому, если пространство состоит только из функций, принимающих на концах интервала одинаковые значения, то

$$\left(\frac{df}{dx}, g\right) = \left(f, -\frac{dg}{dx}\right),$$

т.е. на таком пространстве оператор $-\frac{d}{dx}$ сопряжен с оператором $\frac{d}{dx}$.

Используя формулу интегрирования по частям несколько раз или пользуясь формальным операторным соотношением $(f_1 \circ \ldots \circ f_n)^* = f_n^* \circ \ldots \circ f_1^*$, получаем, что на таких пространствах

$$\left[\sum_{i=0}^{n} a_i(x) \frac{d^i}{dx^i}\right]^* = \sum_{i=0}^{n} (-1)^i \frac{d^i}{dx^i} \circ a_i(x), \tag{3}$$

где запись $\frac{d^i}{dx^i} \circ a_i(x)$ для оператора означает, что, применяя его к функции $f(x)$, мы сначала умножаем ее на $a_i(x)$ и затем дифференцируем i раз по x. Формула (3) определяет операцию (*формального*) *сопряжения* дифференциальных операторов: $D \mapsto D^*$. Оператор D называется (формально) самосопряженным, если $D = D^*$. Слово «формальный» здесь напоминает о том, что в определении не указано явно пространство, на котором D реализуется как линейный оператор.

Если скалярное произведение определяется с помощью веса $G(x)$:

$$(f, g)_G = \int_a^b G(x) f(x) g(x)\, dx,$$

то очевидные вычисления показывают, что вместо D^* следует рассматривать оператор $G^{-1} \circ D^* \circ G$ (считая, что G не обращается в нуль); именно он является кандидатом на роль сопряженного оператора к D относительно $(f, g)_G$.

Покажем, что ортогональные системы функций, рассмотренные в § 4, состоят из собственных функций самосопряженных дифференциальных операторов.

а) *Вещественные многочлены Фурье степени* $\leqslant N$. Оператор $\frac{d^2}{dx^2}$, формально самосопряженный, переводит это пространство в себя и самосопряжен на нем. Кроме того, его собственные значения равны 0 (кратность 1) и $-1^2, -2^2, \ldots, -N^2$ (кратность 2). Соответствующие собственные векторы суть 1 и $\{\cos nx, \sin nx\}$, $1 \leqslant n \leqslant N$.

б) *Многочлены Лежандра.* Оператор

$$(x^2 - 1)\frac{d^2}{dx^2} + 2x\frac{d}{dx} = \frac{d}{dx} \circ \left[(x^2 - 1)\frac{d}{dx}\right]$$

формально самосопряжен и переводит пространство многочленов степени $\leqslant N$ в себя. Имеет место очевидное тождество

$$(x^2-1)\frac{d}{dx}(x^2-1)^n=2nx(x^2-1)^n,$$

откуда по формуле Лейбница, примененной к обеим частям,

$$\frac{d^{n+1}}{dx^{n+1}}\left[(x^2-1)\frac{d}{dx}(x^2-1)^n\right]=$$

$$=(x^2-1)\frac{d^{n+2}}{dx^{n+2}}(x^2-1)^n+2(n+1)x\frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n+$$

$$+n(n+1)\frac{d^n}{dx^n}(x^2-1)^n=2nx\frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n+2n(n+1)\frac{d^n}{dx^n}(x^2-1)^n.$$

Разделив последнее равенство на $2^n n!$ и вспомнив определение многочленов Лежандра, получим отсюда

$$\left[(x^2-1)\frac{d^2}{dx^2}+2x\frac{d}{dx}\right]P_n(x)=n(n+1)P_n(x).$$

Таким образом, *оператор* $(x^2-1)\frac{d^2}{dx^2}+2x\frac{d}{dx}$ *на пространстве многочленов степени* $\leqslant N$ *диагонализируется в ортогональном базисе из многочленов Лежандра и имеет простой вещественный спектр. Стало быть, он самосопряжен.*

Разумеется, самосопряженность на этом пространстве можно было бы проверить и непосредственным интегрированием по частям: член типа $fg\big|_{-1}^{1}$ пропадет здесь из-за множителя x^2-1 в коэффициентах оператора. Тогда из результатов этого пункта и теоремы п. 4 получается другое доказательство попарной ортогональности многочленов Лежандра.

Мы оставляем читателю часть проверок и интерпретацию в терминах линейной алгебры соответствующих фактов для многочленов Эрмита и Чебышева (помнить о весовых множителях $G(x)$!).

в) *Многочлен Эрмита* $H_n(x)=(-1)^n e^{x^2}\frac{d^n}{dx^n}(e^{-x^2})$ *есть собственный вектор с собственным значением* $-2n$ *оператора*

$$K=\frac{d^2}{dx^2}-2x\frac{d}{dx}.$$

Функция $e^{-x^2/2}H_n(x)=(-1)^n e^{x^2/2}\frac{d^n}{dx^n}(e^{-x^2})$ *является собственным вектором оператора*

$$H=\frac{d^2}{dx^2}-x^2$$

с собственным значением $-(2n+1)$.

Первое утверждение проверяется прямой индукцией по n, которую мы опускаем. Для доказательства второго утверждения рассмотрим вспомогательный оператор

$$M = \frac{d}{dx} - x.$$

Легко проверить, что

$$[H, M] = HM - MH = -2\left(\frac{d}{dx} - x\right) = -2M.$$

Отсюда следует, что если f есть собственная функция оператора H с собственным значением λ, то Mf есть собственная функция оператора H с собственным значением $\lambda - 2$:

$$HMf = [H, M]f + MHf = -2Mf + \lambda Mf = (\lambda - 2)Mf.$$

Поскольку $H(e^{-x^2/2}) = -e^{-x^2/2}$, мы получаем, что $M^n(e^{-x^2/2})$ есть собственная функция для H с собственным значением $-(2n+1)$ при всех $n \geqslant 0$. С другой стороны, прямая проверка показывает, что

$$e^{x^2/2}M(e^{-x^2/2}f(x)) = e^{x^2}\frac{d}{dx}(e^{-x^2}f(x)),$$

откуда вытекает, что

$$e^{-x^2/2}H_n(x) = (-1)^n M^n(e^{-x^2/2}),$$

что завершает доказательство второго утверждения.

г) *Многочлен Чебышева*

$$T_n(x) = \frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\,\frac{d^n}{dx^n}(1-x^2)^{n-\frac{1}{2}}$$

есть собственный вектор с собственным значением $-n^2$ *оператора*

$$(1-x^2)\frac{d^2}{dx^2} - x\frac{d}{dx}.$$

11. Нормальные операторы. Как унитарные, так и самосопряженные операторы в унитарном пространстве являются частным случаем *нормальных операторов*, которые можно описать двумя равносильными свойствами:

а) *Это операторы, диагонализируемые в ортонормированном базисе.*

б) *Это операторы, коммутирующие со своим сопряженным оператором.*

Проверим равносильность.

Если $\{e_i\}$ — ортонормированный базис с $f(e_i) = \lambda e_i$, то $f^*(e_i) = \bar{\lambda}_i e_i$, так что $[f, f^*] = 0$, и из а) следует б).

Для доказательства обратной импликации выберем собственное значение λ оператора f и положим

$$L_\lambda = \{l \in L \mid f(l) = \lambda l\}.$$

Проверим, что $f^*(L_\lambda) \subset L_\lambda$. В самом деле, если $l \in L$, то

$$f(f^*(l)) = f^*(f(l)) = f^*(\lambda l) = \lambda f^*(l),$$

поскольку $ff^* = f^*f$. Отсюда вытекает, что пространство $L_\lambda^\perp$ f-инвариантно: если $(l, l_0) = 0$ для всех $l_0 \in L$, то

$$(f(l), l_0) = (l, f^*(l_0)) = 0.$$

Такое же рассуждение показывает, что $L_\lambda^\perp$ f^*-инвариантно. Ограничения f и f^* на $L_\lambda^\perp$, очевидно, коммутируют. Применяя индукцию по размерности L, мы можем считать, что на $L_\lambda^\perp$ f диагонализируется в ортонормированном базисе. Так как то же верно для L_λ, это завершает доказательство.

УПРАЖНЕНИЯ

1. Пусть $f: L \to L$ — оператор в унитарном пространстве. Доказать, что если $|(f(l), l)| \leqslant c|l|^2$ для всех $l \in L$ и некоторого $c > 0$, то

$$|(f(l), m)| + |(l, f(m))| \leqslant 2c|l||m|$$

для всех $l, m \in L$.

2. Пусть $f: L \to L$ — самосопряженный оператор. Доказать, что

$$|(f(l), l)| \leqslant |f||l|^2$$

для всех $l \in L$, где $|f|$ — индуцированная норма f, и если $c < |f|$, то существует вектор $l \in L$ с

$$|(f(l), l)| > c|l|^2.$$

3. Самосопряженный оператор f называется неотрицательным, $f \geqslant 0$, если $(f(l), l) \geqslant 0$ для всех l. Доказать, что это условие равносильно неотрицательности всех точек спектра f.

4. Доказать, что отношение $f \geqslant g$: $f - g \geqslant 0$ является отношением порядка на множестве самосопряженных операторов.

5. Доказать, что произведение двух коммутирующих неотрицательных самосопряженных операторов неотрицательно.

6. Доказать, что из каждого неотрицательного самосопряженного оператора можно извлечь единственный неотрицательный квадратный корень.

7. Вычислить явно поправку второго приближения к собственному вектору и собственному значению оператора $H_0 + \varepsilon H_1$

8. Пусть f — самосопряженный оператор, $\omega \in \mathbf{C}$, $\operatorname{Im} \omega \neq 0$ Доказать, что оператор

$$g = (f - \bar{\omega}\,\mathrm{id})(f - \omega\,\mathrm{id})^{-1}$$

унитарен, его спектр не содержит единицы и

$$f = (\omega g - \bar{\omega}\,\mathrm{id})(g - \mathrm{id})^{-1}.$$

9. Наоборот, пусть g — унитарный оператор, спектр которого не содержит едиииицы. Доказать, что оператор

$$f = (\omega g - \bar{\omega}\,\mathrm{id})(g - \mathrm{id})^{-1}$$

самосопряжен и

$$g = (f - \bar{\omega}\,\mathrm{id})(f - \omega\,\mathrm{id})^{-1}.$$

(Описанные здесь отображения, которые связывают самосопряженные и унитарные операторы, называются преобразованиями Кэли. В одномерном случае они отвечают отображению $a \longmapsto \dfrac{a - \bar{\omega}}{a - \omega}$, которое переводит вещественную ось в единичную окружность.)

10. Пусть f: $L \to L$ — любой линейный оператор в унитарном пространстве. Доказать, что f^*f — неотрицательный самосопряженный оператор и что он положителен тогда и только тогда, когда f обратим.

11. Пусть f обратим и $r_1^2 = ff^*$, $r_2^2 = f^*f$, где r_1, r_2 — положительные самосопряженные операторы. Доказать, что

$$f = r_1 u_1 = u_2 r_2,$$

где u_1, u_2 унитарны. (Эти представления называются полярными разложениями линейного оператора f, где u_1, u_2 — соответственно правый и левый фазовые множители f. В одномерном случае получается представление ненулевых комплексных чисел в виде $re^{i\varphi}$.)

12. Доказать, что полярные разложения $f = r_1u_1 = u_2r_2$ единственны.

13. Доказать, что полярные разложения существуют также для необратимых операторов f, но однозначно определяются лишь r_1, r_2, а не унитарные сомножители.

§ 9. Самосопряженные операторы в квантовой механике

1. Мы продолжаем здесь обсуждение основных постулатов квантовой механики, начатое в п. 8 § 6.

Пусть $\mathscr{H}$ — унитарное пространство состояний некоторой квантовой системы. Для характеризации конкретных состояний в физике пользуются возможностью определить на них («измерить») значения некоторых *физических величин* таких, как энергия, спин, координата, импульс и т. п. Если единица измерения каждой такой величины, а также начало отсчета («нуль») выбраны, то возможные значения являются *вещественными числами* (это по существу определение скалярных величин), и мы всегда будем считать это условие выполненным.

Третий (после принципа суперпозиции и интерпретации скалярных произведений как амплитуд вероятности) постулат квантовой механики состоит в следующем.

Каждой скалярной физической величине, значения которой можно измерять на состояниях системы с пространством состояний $\mathscr{H}$, можно поставить в соответствие самосопряженный оператор $f: \mathscr{H} \to \mathscr{H}$ со следующими свойствами:

а) *Спектр оператора f есть полное множество значений величины, которое можно получить, производя измерения этой величины на разных состояниях системы.*

б) *Если $\psi \in \mathscr{H}$ — собственный вектор оператора f с собственным значением λ, то при измерении этой величины на состоянии ψ с достоверностью получится значение λ.*

в) *Более общо, измеряя величину f на состоянии ψ, $|\psi| = 1$, мы можем получить значение λ из спектра оператора f с вероятностью, равной квадрату нормы ортогональной проекции ψ на полное собственное подпространство $\mathscr{H}(\lambda)$, отвечающее λ.*

Так как в силу теоремы п. 4 § 8 $\mathscr{H}$ разлагается в ортогональную прямую сумму $\bigoplus_{i=1}^{m} \mathscr{H}(\lambda_i)$, $\lambda_i \neq \lambda_j$ при $i \neq j$, мы можем разложить ψ в соответствующую сумму проекций $\psi_i \in \mathscr{H}(\lambda_i)$, $i = 1, \ldots, m$. Теорема Пифагора

$$1 = |\psi|^2 = \sum_{i=1}^{m} |\psi_i|^2$$

интерпретируется тогда как утверждение о том, что, производя измерения f на любом состоянии ψ, мы с вероятностью 1 получим хоть какое-нибудь из возможных значений f.

Физические величины, о которых мы говорили выше, и соответствующие им самосопряженные операторы также называют *наблюдаемыми*. Постулат о наблюдаемых иногда трактуется более широко и считается, что любому самосопряженному оператору отвечает некоторая физическая наблюдаемая.

В бесконечномерных пространствах $\mathscr{H}$ эти постулаты несколько меняются. В частности, вместо б) и в) следует рассматривать вероятность того, что при измерении f в состоянии ψ значения попадут в некоторый интервал $(a, b) \subset \mathbf{R}$. Этому интервалу также можно поставить в соответствие подпространство $\mathscr{H}_{(a, b)} \subset \mathscr{H}$ — образ ортогонального проектора $p_{(a, b)}$ на $\bigoplus\limits_{\lambda_i \in (a, b)} \mathscr{H}(\lambda_i)$ в конечномерном случае, — и искомая вероятность равна

$$|p_{(a, b)}\psi|^2 = (\psi, p_{(a, b)}\psi).$$

Кроме того, в бесконечномерном случае операторы наблюдаемых могут оказаться определенными лишь на некотором подпространстве $\mathscr{H}_0 \subset \mathscr{H}$.

Связь этой терминологии с понятиями, введенными в п. 8 § 6, такова. *Фильтр* B_χ — это прибор, который измеряет наблюдаемую, отвечающую ортогональному проектору на подпространство, порожденное χ. Ей приписывается значение 1, если система прошла через фильтр, и 0 в противном случае. *Печка* A_ψ — это комбинация прибора, производящего систему, вообще говоря, в разных состояниях, и фильтра B_ψ, пропускающего затем лишь системы в состоянии ψ. Рецепт вычисления вероятностей, данный в п. 8 § 6, очевидно, согласуется с рецептом, данным в свойствах б), в) выше

На этом примере видно, что прибор, измеряющий некоторую наблюдаемую, скажем B_χ, в состоянии ψ, вообще говоря, *меняет* это состояние: с вероятностью $|(\psi, \chi)|^2$ он превращает его в χ, а с вероятностью $1 - |(\psi, \chi)|^2$ «уничтожает» систему. Поэтому термин «измерение» в применении к такому акту взаимодействия системы с прибором может привести к совершенно неадекватным интуитивным представлениям. Классическая физика основана на предположении о том, что акт измерения можно в принципе произвести, сколь угодно мало повлияв на состояние системы, подвергшейся измерению. Тем не менее термин «измерение» общепринят в физических текстах, и мы сочли необходимым ввести его здесь, начав ранее с менее обычных, но интуитивно более удобных «печек» и «фильтров».

2. Средние значения и принцип неопределенности. Пусть f — некоторая наблюдаемая, $\{\lambda_i\}$ — ее спектр, $\mathscr{H} = \bigoplus\limits_i \mathscr{H}(\lambda_i)$ — соответствующее ортогональное разложение. Как было сказано, на состоянии ψ, $|\psi| = 1$, f принимает значение λ_i с вероятностью $(\psi, p_i\psi)$, где p_i — ортогональный проектор на $\mathscr{H}(\lambda_i)$. Поэтому *среднее значение* $\widehat{f_\psi}$ величины f на состоянии ψ, взятое по многим измерениям, можно вычислить так:

$$\widehat{f_\psi}=\sum_i \lambda_i(\psi,\ p_i\psi)=\sum_i(\psi,\ \lambda_i p_i\psi)=(\psi,\ f(\psi))$$

(повторяем, что $|\psi|=1$).

Наша величина $(\psi, f(\chi))$ в обозначениях Дирака выглядит так: $\langle\chi|f|\psi\rangle$. Часть этого символа $f|\psi\rangle$ есть результат действия оператора f на кет-вектор $|\psi\rangle$, а $\langle\chi|f$ — результат действия сопряженного оператора на бра-вектор $\langle\chi|$.

Вернемся к средним значениям. Если операторы f, g самосопряжены, то оператор fg, вообще говоря, не является самосопряженным:

$$(fg)^*=g^*f^*=gf\neq fg,$$

если f, g не коммутируют. Однако f^2, $f-\lambda$ $(\lambda\in\mathbf{R})$ и коммутатор $\frac{1}{i}[f,\ g]=\frac{1}{i}(fg-gf)$ по-прежнему самосопряжены.

Среднее значение $[(f-\widehat{f_\psi})^2]\widehat{_\psi}$ наблюдаемой $(f-\widehat{f_\psi})^2$ в состоянии ψ есть *среднеквадратичное отклонение значений* f *от их среднего значения*, или дисперсия (разброс) значений f. Положим

$$\widehat{\Delta f_\psi}=\sqrt{[(f-\widehat{f_\psi})^2]\widehat{_\psi}}.$$

3. Предложение (принцип неопределенности Гейзенберга). *Для любых самосопряженных операторов* f, g *в унитарном пространстве*

$$\widehat{\Delta f_\psi}\cdot\widehat{\Delta g_\psi}\geqslant\frac{1}{2}|([f,\ g]\psi,\ \psi)|.$$

Доказательство. Пользуясь очевидной формулой

$$[f-\widehat{f_\psi},\ g-\widehat{g_\psi}]=[f,\ g],$$

самосопряженностью операторов f, g и неравенством Коши — Буняковского — Шварца, находим $(f_1=f-\widehat{f_\psi},\ g_1=g-\widehat{g_\psi})$

$$|([f,\ g]\psi,\ \psi)|=|((f_1g_1-g_1f_1)\psi,\ \psi)|=|(g_1\psi,\ f_1\psi)-(f_1\psi,\ g_1\psi)|=$$
$$=|2\,\mathrm{Im}\,(g_1\psi,\ f_1\psi)|\leqslant 2|(g_1\psi,\ f_1\psi)|\leqslant$$
$$\leqslant 2\sqrt{(f_1\psi,\ f_1\psi)}\sqrt{(g_1\psi,\ g_1\psi)}=2\widehat{\Delta f_\psi}\widehat{\Delta g_\psi}.$$

Это показывает, что средний разброс значений некоммутирующих наблюдаемых f, g, вообще говоря, не может быть одновременно сделан как угодно малым. Говорят еще, что некоммутирующие наблюдаемые *не измеримы одновременно*; к этой формулировке следует относиться с теми же предосторожностями, что и к термину «измерение».

Особую роль играет применение неравенства Гейзенберга к случаю *канонически сопряженных* пар наблюдаемых, которые по определению удовлетворяют соотношению $\frac{1}{i}[f, g] = \mathrm{id}$. Для них

$$\widehat{\Delta f}_\psi \cdot \widehat{\Delta g}_\psi \geqslant \frac{1}{2},$$

каково бы ни было состояние ψ. Заметим, что в конечномерных пространствах таких пар нет, ибо $\mathrm{Tr}\,[f, g] = 0$, $\mathrm{Tr}\,\mathrm{id} = \dim \mathscr{H}$. Однако в бесконечномерных пространствах они существуют. Классический пример:

$$\frac{1}{i}\left[x, \frac{1}{i}\frac{d}{dx}\right] = \mathrm{id}.$$

Эти операторы появляются в квантовых моделях физических систем, которые на классическом языке называются «частица, движущаяся в одномерном потенциальном поле».

Опишем эти и некоторые другие наблюдаемые подробнее.

4. а) *Наблюдаемая координаты.* Это оператор умножения на x в пространстве комплексных функций на $\mathbf{R}$ (или некоторых подмножествах $\mathbf{R}$) со скалярным произведением $\int f(x)\,\bar{g}(x)\,dx$. Подразумевается квантовая система: «частица, движущаяся по прямой, во внешнем поле».

б) *Наблюдаемая импульса.* Это оператор $\frac{1}{i}\frac{d}{dx}$ в аналогичных пространствах функций. (При нем обычно пишут множителем постоянную Планка $\hbar$; это относится к выбору системы единиц, на котором мы не останавливаемся.)

в) *Наблюдаемая энергии квантового осциллятора.* Это — оператор $\frac{1}{2}\left[-\frac{d^2}{dx^2} + x^2\right]$, снова в подходящих единицах.

г) *Наблюдаемая проекции спина* для системы «частица со спином 1/2». Это любой самосопряженный оператор с собственными значениями $\pm 1/2$ на двумерном унитарном пространстве. Дальнейшие подробности о нем будут даны позже.

В примерах а) — в) мы намеренно не уточняли, в каких унитарных пространствах действуют наши операторы. Они существенно бесконечномерны и строятся и изучаются средствами функционального анализа. О примере г) мы скажем кое-что еще ниже.

5. Наблюдаемая энергии и эволюция системы во времени. В описание любой квантовой системы вместе с ее пространством состояний $\mathscr{H}$ входит задание фундаментальной наблюдаемой $H: \mathscr{H} \to \mathscr{H}$, которая называется *наблюдаемой энергии*, или *оператором Гамильтона*, или *гамильтонианом*.

В ее терминах формулируется последний из основных постулатов квантовой механики.

Если в момент времени 0 система находилась в состоянии ψ и за промежуток времени t развивалась как изолированная система, в частности, над ней не производились измерения, то в мо-

мент времени t она будет находиться в состоянии $\exp(-iHt)(\psi)$, *где*

$$\exp(-iHt) = \sum_{n=0}^{\infty} \frac{(-iH)^n t^n}{n!} : \mathscr{H} \to \mathscr{H}$$

(см. § 11 ч. 1).

Оператор $\exp(-iHt) = U(t)$ унитарен. Однопараметрическая группа унитарных операторов $\{U(t) \mid t \in \mathbf{R}\}$ целиком определяет эволюцию изолированной системы.

Физическая размерность (энергия)$\times$(время) называется «действием». Многие эксперименты позволяют определить универсальную единицу действия — знаменитую постоянную Планка $\hbar = 1{,}055 \cdot 10^{-34}$ Дж$\cdot$с. В нашей формуле подразумевается, что Ht измеряется в единицах $\hbar$, и чаще ее пишут в виде $\exp\left(\frac{Ht}{i\hbar}\right)\psi$. Мы будем опускать $\hbar$ для сокращения записи.

Заметим еще, что, поскольку оператор e^{-iHt} линеен, он переводит лучи в $\mathscr{H}$ в лучи и в самом деле действует на состояния системы, а не просто на векторы ψ.

Закон эволюции можно записать в дифференциальной форме:

$$\frac{d}{dt}(e^{-iHt}\psi) = -iH(e^{-iHt}\psi),$$

или, полагая $\psi(t) = e^{-iHt}\psi$,

$$\frac{d\psi}{dt} = -iH\psi \left(\frac{1}{i\hbar} H\psi, \text{ если помнить о единицах}\right).$$

Последнее уравнение называется *уравнением Шрёдингера*. Впервые оно было написано для случая, когда ψ реализованы как функции в физическом пространстве и H представлен дифференциальным оператором относительно координат.

В следующих комментариях мы, как обычно, ограничимся в основном конечномерными пространствами состояний $\mathscr{H}$.

6. Энергетический спектр и стационарные состояния системы. *Энергетический спектр* системы — это спектр ее гамильтониана H. *Стационарные состояния* — это состояния, которые не меняются со временем. Отвечающие им лучи должны быть инвариантны относительно оператора e^{itH}, т. е. быть одномерными собственными подпространствами этого оператора. Но эти подпространства те же, что и для оператора H. Собственному значению E_j гамильтониана, или *энергетическому уровню* системы, отвечает собственное значение $e^{itE_j} = \cos tE_j + i \sin tE_j$ оператора эволюции, меняющееся со временем.

Если H имеет простой спектр, то пространство $\mathscr{H}$ снабжено каноническим ортонормированным базисом, состоящим из векторов стационарных состояний (они определены с точностью до фазовых множителей $e^{i\varphi}$). Если кратность энергетического уровня E больше единицы, этот уровень и соответствующие состояния называются *вырожденными*, а кратность E — степенью вырождения.

Все состояния, отвечающие нижнему уровню, т. е. наименьшему собственному значению H, называются *основными состояниями* системы; основное состояние единственно, если нижний уровень невырожден. Этот термин связан с представлением о том, что квантовая система никогда не может рассматриваться как полностью изолированная от внешнего мира: с некоторой вероятностью она может излучить или получить порцию энергии. В некоторых условиях гораздо вероятнее, что энергия будет потеряна, чем приобретена, и система будет иметь тенденцию «свалиться» в свое нижнее состояние и в дальнейшем в нем оставаться. Поэтому неосновные состояния называются иногда *возбужденными*.

В примере г) п. 4 был написан гамильтониан квантового осциллятора: $\frac{1}{2}\left[-\frac{d^2}{dx^2}+x^2\right]$. В разделе в) п. 10 § 8 было показано, что функции $e^{-x^2/2}H_n(x)$ образуют систему стационарных состояний гармонического осциллятора с уровнями энергии $E_n=n+\frac{1}{2}$. $n=1, 2, 3, \ldots$ (Более подробный анализ показывает, что энергия измеряется здесь в единицах $\hbar\omega$, где константа ω отвечает частоте колебаний соответствующего классического осциллятора.) Разумным образом определив унитарное пространство, в котором следует работать, можно показать, что это полная система стационарных состояний. При $n>0$ осциллятор может излучить порцию энергии $E_n-E_m=(n-m)\hbar\omega$ и перейти из состояния ψ_n в состояние ψ_m. В применении к квантовой теории электромагнитного поля об этом говорят как об «*излучении $n-m$ фотонов частоты* ω». Обратный процесс будет поглощением $n-m$ фотонов; при этом осциллятор перейдет в более высокое (возбужденное) состояние. Важно, что энергия может быть получена или передана лишь целыми кратными $\hbar\omega$.

В основном состоянии осциллятор имеет ненулевую энергию $\frac{1}{2}\hbar\omega$, которая, однако, никак не может быть передана — более низких энергетических состояний осциллятор не имеет. Электромагнитное поле в квантовых моделях рассматривается как суперпозиция бесконечно многих осцилляторов (отвечающих, в частности, разным частотам ω). В основном состоянии — вакууме — оказывается поэтому, что поле имеет бесконечную энергию, хотя с классической точки зрения оно является нулевым — раз от него нельзя отнять энергию, оно не может ни на что воздействовать! Это простейшая модель глубоких трудностей современной квантовой теории поля. Ни математический аппарат, ни физическая интерпретация квантовой теории поля не достигли какой-либо степени законченности. Это открытая и увлекательная наука.

7. Формулы теории возмущений. В аппарате квантовой механики важную роль играют ситуации, когда гамильтониан H системы может рассматриваться как сумма $H_0+\varepsilon H_1$, где H_0 — «невозмущенный» гамильтониан, а εH_1 — малая добавка, «возмущение». С физической точки зрения возмущение часто обусловливает взаимодействие системы с «внешним миром» (например, внешним магнитным полем) или компонент системы между собой (тогда H_0 отвечает идеализированному случаю системы, состоящей из свободных, невзаимодействующих компонент). С математической точки зрения такое представление оправдано, когда спектральный анализ невозмущенного гамильтониана H_0 проще, чем H, и спектральные характеристики H удобно представлять рядами по степеням ε, первые члены которых определяются через H_0. Мы ограничимся следующими наиболее употребительными формулами и качественными замечаниями к ним.

а) *Поправки первого порядка.* Пусть $H_0 e_0 = \lambda_0 e_0$, $|e_0| = 1$. Попытаемся найти собственный вектор и собственное значение $H_0 + \varepsilon H_1$, близкие к e_0 и λ_0 соответственно, с точностью до членов второго порядка малости по ε, т. е. решить уравнение

$$(H_0 + \varepsilon H_1)(e_0 + \varepsilon e_1) = (\lambda_0 + \varepsilon\lambda_1)(e_0 + \varepsilon e_1) + o(\varepsilon^2).$$

Приравнивая коэффициенты при ε, получаем

$$(H_0 - \lambda_0) e_1 = (\lambda_1 - H_1) e_0.$$

Неизвестные здесь — это число λ_1 и вектор e_1. Их можно найти по очереди с помощью следующего приема. Рассмотрим скалярное произведение обеих частей последнего равенства на e_0. Слева будет нуль в силу самосопряженности $H - \lambda_0$:

$$((H_0 - \lambda_0) e_1, e_0) = (e_1, (H_0 - \lambda_0) e_0) = 0.$$

Поэтому $((\lambda_1 - H_1) e_0, e_0) = 0$ и в силу нормированности e_0

$$\lambda_1 = (H_1 e_0, e_0).$$

Это поправка первого порядка к собственному значению λ_0: «сдвиг энергетического уровня» $\varepsilon\lambda_1$ равен $(\varepsilon H_1 e_0, e_0)$, т. е. по результатам п. 12 *совпадает со средним значением «энергии возмущения» εH_1 на состоянии e_0.*

Для определения e_1 теперь нам нужно обратить оператор $H_0 - \lambda_0$. Разумеется, он необратим, ибо λ_0 — собственное значение H_0; но *правая часть уравнения,* $(\lambda_1 - H_1) e_0$, *ортогональна к* e_0. Поэтому достаточно, чтобы $H_0 - \lambda_0$ был обратим на ортогональном дополнении к e_0, которое мы обозначим $e_0^{\perp}$. Это условие (в конечномерном случае), очевидно, равносильно тому, чтобы *кратность собственного значения λ_0 у H_0* была равна единице, т. е. *невырожденности энергетического уровня λ_0.*

Если это так, то

$$e_1 = ((H_0 - \lambda_0)|_{e_0^{\perp}})^{-1} (\lambda_1 - H_1) e_0,$$

что дает поправку первого порядка к собственному вектору.

Выберем ортонормированный базис $\{e_0 = e^{(0)}, e^{(1)}, \ldots, e^{(n)}\}$, в котором H_0 диагонален с собственными значениями $\lambda_0 = \lambda^{(0)}$,

$\lambda^{(1)}, \ldots, \lambda^{(n)}$. В базисе $\{e^{(1)}, \ldots, e^{(n)}\}$ пространства $e_0^{\perp}$ имеем

$$(\lambda_1 - H_1) e_0 = \sum_{i=1}^{n} ((\lambda_1 - H_1) e_0, e^{(i)}) e^{(i)} = - \sum_{i=1}^{n} (H_1 e_0, e^{(i)}) e^{(i)},$$

откуда

$$e_1 = \sum_{i=1}^{n} \frac{(H_1 e_0, e^{(i)})}{\lambda_0 - \lambda^{(i)}} e^{(i)}.$$

Интуитивно ясно, что эта поправка первого порядка может быть хорошим приближением, если энергия возмущения мала по сравнению с расстоянием от уровня λ_0 до соседнего: ε должно компенсировать знаменатели $\lambda_0 - \lambda^{(i)}$. Физики так обычно и считают.

б) *Поправки высших порядков.* По аналогии с разобранным случаем покажем, что когда собственное значение λ_0 невырождено, можно индуктивно найти поправку $(i+1)$-го порядка к (λ_0, e_0), считая, что поправки порядков $\leqslant i$ уже найдены. Пусть $i \geqslant 1$. Мы решаем уравнение

$$(H_0 + \varepsilon H_1) \left(\sum_{k=1}^{i+1} \varepsilon^k e_k \right) = \left(\sum_{l=1}^{i+1} \varepsilon^l \lambda_l \right) \left(\sum_{j=1}^{i+1} \varepsilon^j e_j \right) + o(\varepsilon^{i+2})$$

относительно e_{i+1}, λ_{i+1} Приравнивая коэффициенты при ε^{i+1}, получаем

$$(H_0 - \lambda_0) e_{i+1} = (\lambda_1 - H_1) e_i + \sum_{l=2}^{i} \lambda_l e_{i+1-l} + \lambda_{i+1} e_0.$$

Как выше, левая часть ортогональна e_0, откуда

$$\lambda_{i+1} = ((H_1 - \lambda_1) e_i, e_0) - \sum_{l=2}^{i} \lambda_l (e_{i+1-l}, e_0),$$

$$e_{i+1} = \left((H_0 - \lambda_0) |_{e_0^{\perp}} \right)^{-1} \left[(\lambda_1 - H_1) e_i + \sum_{l=2}^{i+1} \lambda_l e_{i+1-l} \right].$$

Таким образом, все поправки существуют и единственны.

в) *Ряды теории возмущений.* Формальные ряды по степеням ε

$$\sum_{i=0}^{\infty} \lambda_i \varepsilon^i, \quad \sum_{i=0}^{\infty} e_i \varepsilon^i,$$

где λ_i и e_i находятся по выписанным рекуррентным формулам, называются рядами теории возмущений. Можно доказать, что в конечномерном случае они сходятся при достаточно малых ε. В бесконечномерном случае они могут расходиться; тем не менее несколько первых членов часто приводят к предсказаниям, хорошо согласующимся с экспериментом. Физическая роль рядов теории возмущений в квантовой теории поля очень велика. Их математическое исследование приводит к многим интересным и важным задачам.

г) *Кратные собственные значения и геометрия.* В наших предыдущих вычислениях запрет на кратное собственное значение λ_0 проистекал из желания обратить $H_1 - \lambda_1$ на $e_0^{\perp}$ и формально выражался в появлении разностей $\lambda_0 - \lambda^{(i)}$ в знаменателях. Можно получить формулы и в общем случае, надлежащим образом изменив рассуждения, но мы ограничимся разбором геометрических эффектов кратности. Они видны на типичном случае $H_0 = \mathrm{id}$: все собственные значения равны единице. Малое изменение H_0 приводит к следующим эффектам.

Собственные значения становятся *разными*, если это изменение достаточно общее: этот эффект в физике называется «расщеплением уровней», или «снятием вырождения». Например, одна спектральная линия может расщепиться на две или больше либо при увеличении разрешения прибора, либо при помещении системы во внешнее поле. Математическая модель в обоих случаях будет состоять в учете малой поправки к H_0, ранее неучтенной (впрочем, иногда и в изменении исходного пространства состояний).

Теперь обдумаем, что может происходить с собственными векторами. В полностью вырожденном случае H_0 диагонализируется в *любом* ортобазисе. Малое изменение H_0, снимающее вырождение, означает выбор ортобазиса, вдоль осей которого происходят растяжения, и коэффициентов этих растяжений. Коэффициенты должны мало отличаться от исходного λ_0, но сами оси могут идти в любых направлениях. Таким образом, вблизи вырожденного собственного значения его собственные направления начинают зависеть от возмущения очень сильно. Два сколь угодно малых возмущения единичного оператора с простым спектром могут диагонализироваться в двух фиксированных и жестко повернутых друг от друга ортобазисах. Это показывает внутреннюю причину появления разностей $\lambda_0 - \lambda^{(i)}$ в знаменателях.

§ 10. Геометрия квадратичных форм и собственные значения самосопряженных операторов

1. Этот параграф посвящен изучению геометрии графиков квадратичных форм q на вещественном линейном пространстве, т. е. множеств вида $x_{n+1} = q(x_1, \ldots, x_n)$ в $\mathbf{R}^{n+1}$, и описанию некоторых приложений. Одна из важнейших причин, по которой эти классические результаты вызывают живой интерес и в наши дни, состоит в том, что квадратичные формы дают следующее (после линейного) приближение к любой дважды дифференцируемой функции и потому являются ключом к пониманию геометрии «искривления» любой гладкой многомерной поверхности.

В § 3 и 8 мы уже доказали общие теоремы о классификации квадратичных форм с помощью любых линейных или только ортогональных преобразований, поэтому здесь мы начнем с прояснения их геометрических следствий. Будем считать, что мы работаем в евклидовом пространстве со стандартной метрикой $\sum_{i=1}^{n} x_i^2$. Забвение евклидовой структуры означает лишь введение более грубого отношения эквивалентности между графиками. План геометрического исследования состоит в том, чтобы разобраться с малыми размерностями, где форму графика можно представить себе наглядно, и затем посмотреть на маломерные сечения многомерных графиков в разных направлениях. Читателю рекомендуется рисовать картинки, иллюстрирующие наш текст. Мы считаем ось x_{n+1} направленной вверх, а пространство $\mathbf{R}^n$ расположенным горизонтально.

2. Одномерный случай. График кривой $x_2 = \lambda x_1^2$ в $\mathbf{R}$ имеет три основных формы: «чаша» (выпуклость вниз) при $\lambda > 0$, «купол» (выпуклость вверх) при $\lambda < 0$ и горизонтальная прямая при $\lambda = 0$. Относительно линейной классификации, допускающей произвольное изменение масштаба вдоль оси x_1, эти три случая исчерпывают все возможности: можно считать, что $\lambda = \pm 1$ или 0. При ортогональной классификации λ является инвариантом: $|\lambda|$ определяет крутизну стенок чаши или купола, она тем больше, чем больше $|\lambda|$. Другая характеризация $|\lambda|$ состоит в том, что $\frac{1}{2|\lambda|}$ есть радиус кривизны графика на дне или в вершине $(0, 0)$. Действительно, уравнение окружности радиуса R, касающейся оси x_1 в начале, имеет вид $x_1^2 + (x_2 - R)^2 = R^2$, и вблизи нуля имеем $x_2 \approx \frac{x_1^2}{2R}$.

3. Двумерный случай. Чтобы разобраться в нем, перейдем к ортонормированному базису в $\mathbf{R}^2$, в котором q приводится к сумме квадратов с некоторыми коэффициентами: $q(y_1, y_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2$. Прямые, натянутые на элементы этого базиса, называются *главными осями* формы q; они, вообще говоря, повернуты относительно исходных осей. Числа λ_1 и λ_2 определены однозначно, будучи собственными значениями самосопряженного оператора A, для которого $q(\vec{x}) = \vec{x}^t A \vec{x}$ (в старых координатах). При $\lambda_1 \neq \lambda_2$ сами оси также определены однозначно, но при $\lambda_1 = \lambda_2$ их можно выбирать произвольно (лишь бы они были ортогональны). Для каждого из коэффициентов λ_1, λ_2 есть три основные возможности ($\lambda_i > 0$, $\lambda_i < 0$, $\lambda_i = 0$), но соображения симметрии позволяют ограничиться четырьмя основными случаями (из которых только первые два невырождены).

а) $x_3 = \lambda_1 y_1^2 + \lambda_2 y_2^2$, $\lambda_1, \lambda_2 > 0$. График является *эллиптическим параболоидом*, имеющим форму чаши. Прилагательное «эллиптический» объясняется тем, что проекции горизонтальных сечений $\lambda_1 y_1^2 + \lambda_2 y_2^2 = c$ при $c > 0$ суть эллипсы с полуосями $\sqrt{c\lambda_i^{-1}}$, направленными вдоль главных осей формы q (при $\lambda_1 = \lambda_2$ — окружности). (Эти проекции являются линиями уровня функции q.) Существительное «параболоид» объясняется тем, что сечения графика вертикальными плоскостями $ay_1 + by_2 = 0$ суть параболы (при $\lambda_1 = \lambda_2$ график является параболоидом вращения).

Случай $\lambda_1, \lambda_2 < 0$ — это та же чаша, но опрокинутая.

б) $x_3 = \lambda_1 y_1^2 - \lambda_2 y_2^2$, $\lambda_1, \lambda_2 > 0$. График является *гиперболическим параболоидом.* Линии уровня q суть гиперболы, непустые для всех значений x_3, так что график уходит и выше, и ниже плоскости $x_3 = 0$; сечения вертикальными плоскостями — по-прежнему параболы. Линия уровня $x_3 = 0$ — это «вырожденная гипербола», сводящаяся к своим асимптотам, двум прямым $\sqrt{\lambda_1} y_1 \pm \sqrt{\lambda_2} y_2 = 0$. Эти прямые в $\mathbf{R}^2$ называются *«асимптотическими направлениями»* формы q. Если рассматривать q как (неопределенную) метрику в $\mathbf{R}^2$, то асимптотические прямые состоят из всех векторов длины нуль. Асимптотические прямые делят $\mathbf{R}^2$ на четыре сектора. Линии уровня $q = x_3$ при $x_3 > 0$ лежат в паре противоположных секторов; когда $x_3 \to +0$ сверху, они «прижимаются» к асимптотам, превращаются в них при $x_3 = 0$ и при $x_3 < 0$, «пройдя насквозь», оказываются в другой паре противоположных секторов. Вертикальные сечения графика плоскостями, проходящими через асимптотические прямые, суть сами эти прямые, «распрямившиеся параболы».

Случай $-\lambda_1 y_1^2 + \lambda_2 y_2^2$, $\lambda_1, \lambda_2 > 0$, получается из разобранного заменой знака x_3.

в) $x_3 = \lambda y_1^2$, $\lambda > 0$. Поскольку от y_2 функция не зависит, сечения графика вертикальными плоскостями $y_2 = \text{const}$ имеют один и тот же вид: весь график заметается параболой $x_3 = \lambda y_1^2$ в плоскости (y_1, x_3) при ее движении вдоль оси y_2 и называется *параболическим цилиндром.* Линии уровня суть пары прямых $y_1 = \pm\sqrt{x_3 \lambda^{-1}}$; при $x_3 = 0$ они склеиваются в одну прямую; весь график лежит над плоскостью $x_3 = 0$.

Случай $x_3 = \lambda y_1^2$, $\lambda < 0$ получается «опрокидыванием».

г) $x_3 = 0$. Это — плоскость.

4. Общий случай. Теперь мы в состоянии понять геометрию графика $x_{n+1} = q(x_1, \ldots, x_n)$ при произвольных значениях n.

Перейдем к главным осям в $\mathbf{R}^n$, т. е. к ортонормированному базису, в котором $q(y_1, \ldots, y_n) = \sum\limits_{i=1}^{m} \lambda_i y_i^2$, $\lambda_1 \ldots \lambda_m \neq 0$. Как выше, они определяются однозначно, если $m = n$ и $\lambda_i \neq \lambda_j$ при $i \neq j$ или если $m = n - 1$ и $\lambda_i \neq \lambda_j$ при $i \neq j$. От координат $y_{m+1}, \ldots, y_n$ форма q не зависит, поэтому весь график получается из графика $\sum\limits_{i=1}^{m} \lambda_i y_i^2$ в $\mathbf{R}^{m+1}$ переносом вдоль подпространства, натянутого на $\{e_{m+1}, \ldots, e_n\}$. Иными словами, вдоль этого подпространства график «цилиндричен». Нетрудно убедиться, что оно является как раз ядром билинейной формы, полярной к q, и тривиально тогда и только тогда, когда q невырождена.

Пусть q невырождена, т. е. $m = n$. Можно считать, что $\lambda_1, \ldots, \lambda_r > 0$, $\lambda_{r+1}, \ldots, \lambda_{r+s} < 0$, т. е. (r, s) — сигнатура формы q. Если форма q положительно определена, т. е. $r = n$, $s = 0$, то график имеет вид n-мерной чаши: все его сечения вертикальными плоскостями суть параболы, а все линии уровня $q = c > 0$ суть эллипсоиды с полуосями $\sqrt{c\lambda_i^{-1}}$, направленными вдоль главных осей. Уравнение такого эллипсоида имеет вид

$$\sum_{i=1}^{n}\left(\frac{x_i}{\sqrt{c\lambda_i^{-1}}}\right)^2 = 1,$$

т. е. он получается из единичного шара растяжениями вдоль ортогональных направлений. В частности, он ограничен: целиком лежит в прямоугольном параллелепипеде $|x_i| \leqslant \sqrt{c\lambda_i^{-1}}$, $i = 1, \ldots, n$. Ниже мы убедимся, что изучение вариации длин полуосей разных сечений эллипсоида (тоже эллипсоидов) дает полезную информацию о собственных значениях самосопряженных операторов.

При $r = 0$, $s = n$ получается купол. В обоих случаях график называется (n-мерным) *эллиптическим параболоидом.*

Промежуточные случаи $rs \neq 0$ приводят к многомерным гиперболическим параболоидам разных сигнатур. Ключом к их геометрии является снова структура *конуса асимптотических направлений C* в $\mathbf{R}^n$, т. е. нулевого уровня формы $q(y_1, \ldots, y_n) = 0$.

Конусом он называется потому, что заметается своими *образующими*: прямая, содержащая один вектор из C, целиком лежит в нем. Чтобы составить себе представление о базе этого конуса, рассмотрим его пересечение, скажем, с линейным многообразием $y_n = 1$:

$$-\lambda_n^{-1}\sum_{i=1}^{n-1}\lambda_i y_i^2 = 1.$$

Видно, что база является множеством уровня квадратичной формы от $n-1$ переменной. Простейший случай получается, когда она положительно определена: тогда это множество уровня есть эллипсоид, в частности, оно ограничено, и наш конус похож на «школьные» трехмерные конусы. Этот случай отвечает сигнатуре $(n-1, 1)$ или $(1, n-1)$; при $n = 4$ пространство $(\mathbf{R}^4, q)$ есть знаменитое *пространство Минковского,* которое будет подробно изучено ниже. Для других сигнатур C устроен заметно сложнее, ибо его база «уходит на бесконечность». Сечения графика q вертикальными плоскостями, проходящими через образующие C, совпадают с этими образующими. Для любых других плоскостей получаются либо «чаши», либо «купола» — асимптотические направления разделяют эти два случая. Поэтому конус C делит пространство $\mathbf{R}^n \setminus C$ на две части, сплошь заметаемые прямыми, вдоль которых q соответственно положительна или отрицательна. Одна из этих областей называется совокупностью внутренних пол конуса C, другая — его внешностью. *Геометрический смысл сигнатуры (r, s) грубо, но наглядно можно описать следующей фразой: график формы q по r направлениям уходит вверх, а по s — вниз.*

Хотя мы работали все время с вещественными квадратичными формами, те же результаты применимы к комплексным эрмитовым формам. Действительно, овеществление $\mathbf{C}^n$ есть $\mathbf{R}^{2n}$, и овеществление эрмитовой формы $\sum_{i=1}^{n} a_{ij} x_i \bar{x}_j$, $a_{ij} = \bar{a}_{ji}$, есть вещественная квадратичная форма. При овеществлении все размерности удваиваются: в частности, комплексная сигнатура (r, s) превращается в вещественную сигнатуру $(2r, 2s)$.

Опишем теперь вкратце и без доказательств два приложения этой теории в механике и топологии.

5. Колебания. Представим себе сначала шарик, который может кататься в плоскости $\mathbf{R}^2$ под действием силы тяжести по желобу формы $x_2 = \lambda x_1^2$. Точка $(0, 0)$ во всех случаях является одним из возможных движений шарика — *положением равновесия.* При $\lambda > 0$ это положение устойчиво: небольшое начальное отклонение шарика по положению или скорости приведет к его колебаниям около дна чаши. При $\lambda < 0$ оно неустойчиво: шарик свалится вдоль одной из двух ветвей параболы. При $\lambda = 0$ оно безразлично относительно отклонений по положению, но не по скорости: шарик может оставаться в любой точке прямой $x_2 = 0$ либо равномерно двигаться в любую сторону с начальным импульсом.

Оказывается, что математическое описание большого класса механических систем вблизи их положений равновесия хорошо моделируется качественно многомерным обобщением этой картинки: движением шарика вблизи начала координат по многомерной поверхности $x_{n+1} = q(x_1, \ldots, x_n)$ под действием силы тяжести. Если q положительно определена, любое «малое» движение будет близко к суперпозиции малых колебаний вдоль главных осей формы q. Вдоль нулевого пространства формы возможен уход на бесконечность с постоянной скоростью. Вдоль направлений, где q отрицательна, возможно сваливание вниз. Наличие как нулевого пространства, так и отрицательной компоненты сигнатуры свидетельствует о неустойчивости положения равновесия и сомнительности приближения «малых колебаний». Важно, однако, что когда это равновесие устойчиво, малые изменения формы чаши, по которой катается шарик (или, более технически, потенциала нашей системы), не нарушают этой устойчивости.

Чтобы понять это, вернемся к сделанному в начале параграфа замечанию о приближенном представлении любой (скажем, трижды дифференцируемой) вещественной функции $f(x_1, \ldots, x_n)$. Вблизи нуля она имеет вид

$$f(x_1, \ldots, x_n) = f(0, \ldots, 0) + \sum_{i=1}^{n} a_i x_i + \sum_{i,\,j=1}^{n} b_{ij} x_i x_j + o\left(\sum_{i=1}^{n} |x_i|^2\right),$$

где

$$a_i = \frac{\partial f}{\partial x_i}(0, \ldots, 0), \quad b_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(0, \ldots, 0).$$

Вычтя из f ее значение в нуле и линейную часть, получаем, что остаток квадратичен с точностью до членов более высокого порядка малости. Это вычитание означает, что мы рассматриваем отклонение графика f от касательной гиперплоскости к этому графику в нуле. Обозначив эту касательную плоскость через $\mathbf{R}^n$, обнаруживаем, что *поведение f вблизи нуля определяется квадратичной формой с матрицей* $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(0, \ldots, 0)\right)$ *по крайней мере, когда эта форма невырождена*, — иначе нужно учитывать члены более высокого порядка малости. (Например, график $x_2 = x_1^3$ слева уходит вниз, а справа — вверх; графики квадратичных функций так себя не ведут Двумерный график $x_3 = x_1^2 + x_2^3$ — это «обезьянье седло», в одном криволинейном секторе $x_1^2 + x_2^3 < 0$ уходящее вниз — «для хвоста».)

Точка, в которой дифференциал $df = \sum\limits_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$ обращается в нуль $\left(\text{т. е. } \frac{\partial f}{\partial x_i} = 0 \text{ для всех } i = 1, \ldots, n\right)$, называется *критической точкой* функции f (в наших примерах это было начало координат). Она называется *невырожденной*, если в ней квадратичная форма $\sum\limits_{i,\,j=1}^{n} \frac{\partial f}{\partial x_i \partial x_j} \Delta x_i \Delta x_j$ невырождена. Предшествующее обсуждение можно резюмировать в одной фразе: *вблизи невырожденной критической точки график функции расположен относительно касательной гиперплоскости, как график ее квадратичной части.* После этого можно доказать, что малое изменение функции (вместе с ее первыми и вторыми производными) может лишь слегка сдвинуть положение невырожденной критической точки, но не меняет сигнатуры соответствующей квадратичной формы и потому общего поведения графика (в малом).

Можно также доказать, что вблизи невырожденной критической точки можно сделать такую гладкую и гладко обратимую (хотя, вообще говоря, нелинейную) замену координат $y_i = y_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$, что в новых координатах f будет задаваться *в точности* квадратичной функцией:

$$f(y_1, \ldots, y_n) = f(0, \ldots, 0) + \sum_{i,\,j=1}^{n} b_{ij} y_i y_j.$$

Строгое изложение теории малых колебаний читатель сможет найти в книге В. И. Арнольда «Математические методы классической механики» (М.: Наука, 1974, гл. 5).

6. Теория Морса. Представим себе в $(n+1)$-мерном евклидовом пространстве $\mathbf{R}^{n+1}$ n-мерную гладкую ограниченную гиперповерхность V, вроде яйца или баранки (тора) в $\mathbf{R}^3$. Рассмотрим сечения V гиперплоскостями $x_{n+1} = \text{const}$. Предположим, что имеется только конечное число значений $c_1, \ldots, c_m$ таких, что гиперплоскости $x_{n+1}=c_i$ касаются V и притом в единственной точке $v_i \in V$. Вблизи этих точек касания V можно приблизить графиком квадратичной формы $x_{n+1}=c_i+q_i(x_1-x_1(v_i), \ldots$ $\ldots, x_n-x_n(v_i))$, если только V находится в достаточно общем положении (например, бублик не должен лежать горизонтально). Оказывается, что *важнейшие топологические свойства* V, — в частности, так называемый гомотопический тип V — *вполне определяются набором сигнатур форм* q_i, т. е. указанием того, по скольким направлениям V вблизи v_i уходит вниз и по скольким — вверх. Самое замечательное то, что, хотя информация о сигнатурах q_i чисто локальна, восстанавливаемый по ней гомотопический тип V есть глобальная характеристика формы V. Например, если имеются только две критические точки c_1 и c_2 с сигнатурами $(n, 0)$ и $(0, n)$, то V топологически устроена как n-мерная сфера.

Подробности читатель сможет найти в книге Дж. Милнора «Теория Морса» (М.: Мир, 1965).

7. Самосопряженные операторы и многомерные квадрики. Пусть теперь L — конечномерное евклидово или унитарное пространство, $f: L \to L$ — самосопряженный оператор. Нас интересуют свойства его спектра. Расположим собственные значения f в порядке убывания с учетом кратностей: $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ и выберем соответствующий ортонормированный базис $\{e_1, e_2, \ldots, e_n\}$. Вернемся к точке зрения п. 4 § 8, согласно которой задание f равносильно заданию новой симметричной или эрмитовой формы $(f(l_1), l_2)$ или же квадратичной формы $q_f(l)=(f(l), l)$ (в унитарном случае она квадратична на овеществленном пространстве). В базисе $\{e_1, \ldots, e_n\}$ она приобретает вид

$$q_f(x_1, \ldots, x_n)=\sum_{i=1}^{n} \lambda_i x_i^2 \text{ или } \sum_{i=1}^{n} \lambda_i |x_i|^2$$

и, таким образом, направления $\mathbf{R}e_i$ (или $\mathbf{C}e_i$) суть *главные оси* q_f.

Простейшее экстремальное свойство собственных значений λ_i выражается следующим фактом.

8. Предложение. *Пусть $S=\{l \in L \mid |l|=1\}$ — единичная сфера пространства L. Тогда*

$$\lambda_1=\max_{l \in S} q_f(l), \quad \lambda_n=\min_{l \in S} q_f(l).$$

Доказательство. Поскольку $|x_i|^2 \geqslant 0$ и $\lambda_1 \geqslant \ldots \geqslant \lambda_n$, очевидно

$$\lambda_n\left(\sum_{i=1}^{n}|x_i|^2\right) \leqslant \sum_{i=1}^{n} \lambda_i |x_i|^2 \leqslant \lambda_1\left(\sum_{i=1}^{n}|x_i|^2\right).$$

На единичной сфере левая часть есть λ_n, а правая λ_1. Эти значения достигаются на векторах $(0, \ldots, 0, 1)$ и $(1, 0, \ldots, 0)$ соответственно (координаты берутся в базисе $\{e_1, \ldots, e_n\}$, диагонализирующем f).

9. Следствие. *Пусть L_k^- — линейная оболочка семейства $\{e_1, \ldots, e_k\}$, L_k^+ — линейная оболочка семейства $\{e_k, \ldots, e_n\}$. Тогда*

$$\lambda_k = \max\{q_f(l) \mid l \in S \cap L_k^+\} = \min\{q_f(l) \mid l \in S \cap L_k^-\}.$$

Доказательство. Действительно, в очевидных координатах ограничение q_f на L_k^+ имеет вид $\sum_{i=k}^{n} \lambda_i |x_i|^2$, а ограничение на L_k^- — вид $\sum_{i=1}^{k} \lambda_i |x_i|^2$.

Следующее важное усиление этого результата, в котором вместо L_k^+ рассматриваются *любые* линейные подпространства в L коразмерности $k-1$, называется *теоремой Фишера — Куранта*. Она дает «минимаксную» характеристику собственных значений дифференциальных операторов.

10. Теорема. *Для любого подпространства $L' \subset L$ коразмерности $k-1$ справедливы неравенства:*

$$\lambda_k \leqslant \max\{q_f(l) \mid l \in S \cap L'\}, \quad \lambda_{n-k+1} \geqslant \min\{q_f(l) \mid l \in S \cap L'\}.$$

Эти оценки точны для некоторых L' (например, L_k^+ и L_{n-k+1}^- соответственно), так что

$$\lambda_k = \min_{L'} \max\{q_f(l) \mid l \in S \cap L'\}, \quad \lambda_{n-k+1} = \max_{L'} \min\{q_f(l) \mid l \in S \cap L'\}.$$

Доказательство. Поскольку

$$\dim L' + \dim L_k^- = (n-k+1) + k = n+1,$$

а $\dim(L' + L_k^-) \leqslant \dim L = n$, из теоремы п. 3 § 5 ч. 1 следует, что $\dim(L' \cap L_k^-) \geqslant 1$. Возьмем вектор $l_0 \in L' \cap L_k^- \cap S$. Согласно следствию 9 $\lambda_k = \min\{q_f(l) \mid l \in S \cap L_k^-\}$, так что $\lambda_k \leqslant q_f(l_0)$ и тем более $\lambda_k \leqslant \max\{q_f(l) \mid l \in S \cap L'\}$. Второе неравенство теоремы проще всего получить, применив первое неравенство к оператору $-f$ и заметив, что знаки и порядок собственных значений при этом обращаются.

11. Следствие. *Пусть $\dim L/L_0 = 1$ и p — оператор ортогонального проектирования $L \to L_0$. Обозначим через $\lambda_1' \geqslant \lambda_2' \geqslant \ldots \geqslant \lambda_{n-1}'$ собственные значения самосопряженного оператора pf: $L_0 \to L_0$. Тогда*

$$\lambda_1 \geqslant \lambda_1' \geqslant \lambda_2 \geqslant \lambda_2' \geqslant \ldots \geqslant \lambda_{n-1}' \geqslant \lambda_n,$$

т. е. собственные значения операторов f и pf перемежаются.

Доказательство. Ограничение формы q_f на L_0 совпадает с q_{pf}: $(f(l), l) = (pf(l), l)$, если $l \in L_0$. Поэтому

$$\lambda_k' = \max\{q_{pf}(l) \mid l \in S \cap L'\} = \max\{q_f(l) \mid l \in S \cap L'\}$$

для подходящего подпространства $L' \subset L_0$, имеющего коразмерность $k-1$ в L_0. Значит, в L оно имеет коразмерность k, откуда $\lambda_{k+1} \leqslant \lambda'_k$. Записав это неравенство для $-f$ вместо f, получим $-\lambda_k \leqslant -\lambda'_k$, т. е. $\lambda'_k \leqslant \lambda_k$. Это завершает доказательство.

Мы предоставляем читателю возможность убедиться в том, что следствие п. 11 имеет следующий простой геометрический смысл. Будем считать, что $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0$ и вместо функции $q_f(l)$ на S рассмотрим эллипсоид ε: $q_f(l)=1$. Тогда его сечение ε_0 подпространством L_0 также представляет собой эллипсоид, длины полуосей которого перемежаются с длинами полуосей эллипсоида ε. Вообразите себе, например, эллипсоид ε в $\mathbf{R}^3$ и его сечение плоскостью ε_0. Большая полуось ε_0 не превосходит большой полуоси ε («очевидно»), но не меньше, чем средняя полуось ε. Малая полуось ε_0 не меньше малой полуоси ε («очевидно»), но не больше средней полуоси ε. Контрольный вопрос: как получить в сечении окружность?

§ 11. Трехмерное евклидово пространство

1. Трехмерное евклидово пространство $\mathscr{E}$ является основной моделью физического пространства Ньютона и Галилея. Четырехмерное пространство Минковского $\mathscr{M}$, снабженное симметричной метрикой сигнатуры $(r_+, r_-) = (1, 3)$, является моделью пространства-времени релятивистской физики. Уже поэтому они заслуживают более пристального изучения. С математической точки зрения они также имеют особые свойства, существенные для понимания устройства мира, в котором мы живем: связь вращений в $\mathscr{E}$ с кватернионами и существование векторного произведения; геометрия векторов нулевой длины в $\mathscr{M}$.

Эти специальные свойства весьма удобно излагать на языке связи геометрии $\mathscr{E}$ и $\mathscr{M}$ с геометрией *вспомогательного двумерного унитарного пространства* $\mathscr{H}$, называемого пространством спиноров. Эта связь имеет также глубокий физический смысл, ставший ясным лишь после появления квантовой механики. Мы избрали именно такое изложение.

2. Итак, фиксируем двумерное унитарное пространство $\mathscr{H}$. Обозначим через $\mathscr{E}$ *вещественное линейное пространство самосопряженных операторов в* $\mathscr{H}$ *с нулевым следом*. Каждый оператор $f \in \mathscr{E}$ имеет два вещественных собственных значения; они отличаются только знаком, ибо след, равный их сумме, обращается в нуль. Положим

$$|f| = \sqrt{|\det f|} \text{ — положительное собственное значение } f.$$

3. Предложение. *$\mathscr{E}$ с нормой $|\ |$ является трехмерным евклидовым пространством.*

Доказательство. В ортонормированном базисе $\mathscr{H}$ операторы f представлены эрмитовыми матрицами вида

$$\begin{pmatrix} a & b \\ \bar{b} & -a \end{pmatrix}, \quad a \in \mathbf{R}, \ b \in \mathbf{C},$$

т. е. линейными комбинациями

$$\operatorname{Re} b \cdot \sigma_1 + \operatorname{Im} b \cdot \sigma_2 + a\sigma_3,$$

где σ_1, σ_2, σ_3 — матрицы Паули (см. упражнение 5 к § 4 ч. 1):

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Так как σ_1, σ_2, σ_3 линейно независимы над $\mathbf{R}$, $\dim_{\mathbf{R}} \mathscr{E} = 3$.

Положим теперь

$$(f, g) = \frac{1}{2} \operatorname{Tr}(fg).$$

Это билинейное симметричное скалярное произведение, и если собственные значения f равны $\pm\lambda$, то

$$|f|^2 = \frac{1}{2} \operatorname{Tr}(f^2) = \frac{1}{2}(\lambda^2 + \lambda^2) = |\det f|.$$

Очевидно, $\lambda^2 = 0$ тогда и только тогда, когда $f = 0$. Это завершает доказательство.

Назовем *направлением в* $\mathscr{E}$ множество векторов вида

$$\mathbf{R}_+ f = \{af \mid a > 0\},$$

где f — ненулевой вектор из $\mathscr{E}$. Иными словами, направление — это полупрямая в $\mathscr{E}$. Направление, противоположное к $\mathbf{R}_+ f$, — это $\mathbf{R}_+(-f)$.

4. Предложение. *Имеется взаимно однозначное соответствие между направлениями в* $\mathscr{E}$ *и разложениями* $\mathscr{H}$ *в прямую сумму двух ортогональных одномерных подпространств* $\mathscr{H}_+ \oplus \mathscr{H}_-$. *Именно, направлению* $\mathbf{R}_+ f$ *отвечают* $\mathscr{H}_+$ — *собственное подпространство* $\mathscr{H}$ *для положительного собственного значения* f, $\mathscr{H}_-$ *то же для отрицательного собственного значения.*

Доказательство. $\mathscr{H}_+$ и $\mathscr{H}_-$ ортогональны по теореме п. 4 § 7. Замена f на af, $a > 0$, не меняет $\mathscr{H}_+$ и $\mathscr{H}_-$. Наоборот, если ортогональное разложение $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$ задано, то множество операторов $f \in \mathscr{E}$, растягивающих $\mathscr{H}$ в $\lambda > 0$ раз вдоль $\mathscr{H}_+$ и в $-\lambda < 0$ раз вдоль $\mathscr{H}_-$, образует направление в $\mathscr{E}$.

5. Физическая интерпретация. Отождествим $\mathscr{E}$ с физическим пространством, например, посредством выбора ортогональных координат в $\mathscr{E}$ и в пространстве. Отождествим $\mathscr{H}$ с пространством внутренних состояний квантовой системы «частица со спином $1/2$, локализованная вблизи начала координат» (например, электрон). Выбрав направление $\mathbf{R}_+ f \subset \mathscr{E}$, включим магнитное поле в физическом пространстве вдоль этого направления. В этом поле система будет иметь два стационарных состояния, которые как раз и суть $\mathscr{H}_+$ и $\mathscr{H}_-$.

Если направление $\mathbf{R}_+ f$ отвечает, скажем, верхней вертикальной полуоси избранной координатной системы в физическом пространстве («ось z»), то состояние $\mathscr{H}_+$ называется состоянием «с проекцией спина $+1/2$ на ось z» (или «спин вверх»), а $\mathscr{H}_-$ — соответственно состоянием «с проекцией спина $-1/2$» (или «спин вниз»).

Такая традиционная терминология является реликтом доквантовых представлений о том, что наблюдаемая спина отвечает классической наблюдаемой «момент количества движения» — характеристике внутреннего вращения системы и потому может быть сама представлена вектором в $\mathscr{E}$, который поэтому имеет проекции на оси координат в $\mathscr{E}$. Это совершенно неверно: состояния системы суть лучи в $\mathscr{H}$, а не векторы в $\mathscr{E}$. Расхождение с классикой становится еще более очевидным при рассмотрении систем со спином $s/2$, $s > 1$, для которых $\dim \mathscr{H} = s+1$. Точное утверждение дается именно предложением п. 4.

Мы дали идеализированное описание классического эксперимента Штерна — Герлаха (1922). Вместо электронов в нем использовались ионы серебра, проходившие между полюсами электромагнита. Из-за неоднородности магнитного поля ионы, вышедшие в состояниях, близких к $\mathscr{H}_+$ и $\mathscr{H}_-$ соответственно, пространственно разделялись на два пучка, что и позволило макроскопически отождествить эти состояния. Серебро испарялось в электрической печке, а магнитное поле между полюсами играло роль объединения двух фильтров, пропускающих раздельно состояния $\mathscr{H}_+$ и $\mathscr{H}_-$.

Продолжим теперь изучение евклидова пространства $\mathscr{E}$.

5. **Предложение.** $(f, g) = 0$ *тогда и только тогда, когда* $fg + gf = 0$.

Доказательство. Имеем

$$(f,\ g) = \frac{1}{2}\operatorname{Tr}(fg) = \frac{1}{4}\operatorname{Tr}(fg + gf) = \frac{1}{4}\operatorname{Tr}[(f+g)^2 - f^2 - g^2].$$

Но f^2 имеет единственное собственное значение $|f|^2$, поэтому все квадраты операторов из $\mathscr{E}$ являются скалярными, значит и $fg + gf$ — скалярный оператор, и он равен нулю тогда и только тогда, когда его след равен нулю.

6. **Ортонормированные базисы в $\mathscr{E}$.** Из доказательства предложения п. 5 ясно, что операторы $\{e_1, e_2, e_3\}$ образуют ортонормированный базис тогда и только тогда, когда

$$e_1^2 = e_2^2 = e_3^2 = \mathrm{id};\ e_ie_j + e_je_i = 0,\ i \neq j.$$

В частности, если в $\mathscr{H}$ выбран ортонормированный базис, то операторы, заданные в нем матрицами Паули $\sigma_1, \sigma_2, \sigma_3$, образуют ортонормированный базис в $\mathscr{E}$:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ \sigma_i\sigma_j + \sigma_j\sigma_i = 0,\ i \neq j.$$

Теперь мы можем объяснить математический смысл матриц Паули, доказав обратное утверждение.

7. **Предложение.** *Для каждого ортонормированного базиса* $\{e_1, e_2, e_3\}$ *пространства* $\mathscr{E}$ *существует ортонормированный базис* $\{h_1, h_2\}$ *пространства* $\mathscr{H}$, *обладающий тем свойством, что*

$$A_{e_1} = \sigma_1,\ A_{e_2} = \sigma_2 \text{ или } -\sigma_2,\ A_{e_3} = \sigma_3,$$

где A_e — *матрица оператора* e *в базисе* $\{h_1, h_2\}$. *Он определен с точностью до умножения на комплексное число, по модулю равное единице.*

Доказательство. Собственные значения e_i суть ± 1. Пусть $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$, где e_3 действует на $\mathscr{H}_+$ тождественно, а на $\mathscr{H}_-$ — изменением знака. Выберем сначала векторы $h'_1 \in \mathscr{H}_+$, $h'_2 \in \mathscr{H}_-$, $|h'_1| = |h'_2| = 1$. Они определены с точностью до умножения на $e^{i\varphi_1}$, $e^{i\varphi_2}$; матрица e_3 в базисе $\{h'_1, h'_2\}$ есть σ_3.

Далее,

$$e_1(h'_1) = e_1 e_3(h'_1) = -e_3 e_1(h'_1),$$

так что $e_1(h'_1)$ есть ненулевой собственный вектор для e_3 с собственным значением -1. Поэтому $e_1(h'_1) = \alpha h'_2$. Аналогично $e_1(h'_2) = \beta h'_1$. Матрица e_1 в базисе $\{h'_1, h'_2\}$ эрмитова, поэтому $\alpha = \bar{\beta}$. Наконец, $e_1^2 = \mathrm{id}$, поэтому $\alpha\beta = 1 = |\alpha|^2 = |\beta|^2$. Заменив $\{h'_1, h'_2\}$ на $\{h_1, h_2\} = \{xh'_1, yh'_2\}$, где $|x| = |y| = 1$, чтобы превратить матрицу e_1 в новом базисе в σ_1, получим

$$e_1(h_1) = xe_1(h'_1) = x\alpha h'_2 = \alpha xy^{-1}h_2,$$

$$e_1(h_2) = ye_1(h'_2) = y\beta h'_1 = \beta yx^{-1}h_1.$$

Поэтому x, y должны удовлетворять еще условию $xy^{-1} = \alpha^{-1}$; тогда автоматически $\alpha xy^{-1} = \beta yx^{-1} = 1$. Можно положить, например, $x = 1$, $y = \alpha$.

Итак, в базисе $\{h_1, h_2\}$ имеем $A_{e_3} = \sigma_3, A_{e_1} = \sigma_1$, и этот базис определен с точностью до умножения на скаляр, по модулю равный единице. Те же рассуждения, что для e_1, показывают, что в таком базисе A_{e_2} имеет вид $\begin{pmatrix} 0 & \gamma \\ \bar{\gamma} & 0 \end{pmatrix}$, где $|\gamma|^2 = 1$. Кроме того, условие ортогональности $e_1e_2 + e_2e_1 = 0$ дает

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & \gamma \\ \bar{\gamma} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \gamma \\ \bar{\gamma} & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 0,$$

т. е. $\gamma + \bar{\gamma} = 0$, откуда $\gamma = i$, либо $\gamma = -i$. Поэтому $A_{e_2} = \sigma_2$ или $A_{e_2} = -\sigma_2$.

8. Следствие. *Пространство $\mathscr{E}$ снабжено отмеченной ориентацией: ортонормированный базис $\{e_1, e_2, e_3\}$ принадлежит к классу, отвечающему этой ориентации, тогда и только тогда, когда существует ортонормированный базис $\{h_1, h_2\}$ в $\mathscr{H}$, в котором $A_{e_a} = \sigma_a$, $a = 1, 2, 3$.*

Доказательство. Мы должны проверить, что если $\{e_a\}$ в базисе $\{h_b\}$ и $\{e'_a\}$ в базизе $\{h'_b\}$ задаются в точности матрицами Паули, то определитель матрицы перехода от $\{e_a\}$ к $\{e'_a\}$ положителен, или что имеется непрерывное движение, переводящее $\{e_a\}$ в $\{e'_a\}$. Мы построим такое движение, показав, что $\{h_b\}$ переводится в $\{h'_b\}$ унитарным непрерывным движением: существует такая система унитарных операторов f_t: $\mathscr{H} \to \mathscr{H}$, зависящая от параметра $t \in [0, 1]$, что $f_0 = \mathrm{id}$, $f_1(h_b) = h'_b$ и $\{f_t(h_1), f_t(h_2)\}$ образуют ортонормированный базис $\mathscr{H}$ для всех t. Тогда, обозначив через $\{g_t(e_1), g_t(e_2), g_t(e_3)\}$ ортонормированный базис $\mathscr{E}$, задающийся матрицами Паули в базисе $\{f_t(h_1), f_t(h_2)\}$, мы построим нужное нам движение в $\mathscr{E}$.

Пусть $\{h_1', h_2'\} = \{h_1, h_2\}U$. Поскольку оба базиса ортонормированы, матрица перехода U должна быть унитарна. По следствию п. 7 § 8 ее можно представить в виде $\exp(iA)$, где A — эрмитова матрица. Тогда для всех $t \in A$ матрица tA эрмитова, а оператор $\exp(itA)$ унитарен, и мы можем положить

$$f_t\{h_1, h_2\} = \{h_1, h_2\}\exp(itA),\ 0 \leqslant t \leqslant 1.$$

Это завершает доказательство.

Операторы $\sigma_1/2$, $\sigma_2/2$, $\sigma_3/2$ в $\mathscr{H}$ называются *наблюдаемыми проекций спина* на соответствующие оси в $\mathscr{E}$: эту терминологию объясняет квантовомеханическая интерпретация из п. 5. Множитель $^1/_2$ введен для того, чтобы их собственные значения были равны $\pm^1/_2$.

9. Векторное произведение. Пусть $\{e_1, e_2, e_3\}$ — ортонормированный базис в $\mathscr{E}$, принадлежащий отмеченной ориентации. Векторное произведение в $\mathscr{E}$ определяется классической формулой

$$(x_1e_1 + x_2e_2 + x_3e_3) \times (y_1e_1 + y_2e_2 + y_3e_3) =$$
$$= (x_2y_3 - x_3y_2)e_1 + (x_3y_1 - x_1y_3)e_2 + (x_1y_2 - x_2y_1)e_3.$$

Замена базиса на другой, ориентированный так же, не меняет векторное произведение; если же новый базис ориентирован противоположно, то у него меняется знак.

Инвариантную конструкцию векторного произведения нетрудно дать в наших терминах. Вспомним, что *антиэрмитовы* операторы в $\mathscr{H}$ с нулевым следом образуют алгебру Ли $\mathrm{su}(2)$ (см. § 4 ч. 1). Пространство $\mathscr{E}$ можно отождествить с этой алгеброй Ли, разделив каждый оператор из $\mathscr{E}$ на i. Поэтому на $\frac{1}{i}\mathscr{E}$ имеется структура алгебры Ли. Имеем

$$\left[\frac{1}{i}\sigma_a, \frac{1}{i}\sigma_b\right] = 2\varepsilon_{abc}\frac{1}{i}\sigma_c,$$

где $\varepsilon_{123} = 1$ и ε_{abc} кососимметричен по всем индексам, или

$$[\sigma_a, \sigma_b] = 2i\varepsilon_{abc}\sigma_c.$$

Поэтому

$$\left[\sum_{a=1}^{3} x_a\sigma_a, \sum_{b=1}^{3} y_b\sigma_b\right] = 2i\left(\sum_{a=1}^{3} x_a\sigma_a\right) \times \left(\sum_{b=1}^{3} x_b\sigma_b\right),$$

так что *векторное произведение с точностью до тривиального множителя есть просто коммутатор операторов.* Это позволяет без вычислений установить классические тождества

$$\vec{x} \times \vec{y} = -\vec{y} \times \vec{x};$$

$$\vec{x} \times (\vec{y} \times \vec{z}) + \vec{z} \times (\vec{x} \times \vec{y}) + \vec{y} \times (\vec{z} \times \vec{x}) = 0.$$

Есть еще один способ ввести векторное произведение, одновременно связав его со скалярным произведением и кватернионами.

10. Кватернионы. Как и коммутирование, умножение операторов из $\mathcal{E}$, вообще говоря, выводит нас за пределы $\mathcal{E}$: одновременно нарушается эрмитовость и условие обращения следа в нуль. На самом деле произведение операторов из $\mathcal{E}$ лежит в $\mathbf{R}\mathrm{id}+i\mathcal{E}$, причем «вещественная часть» есть как раз скалярное произведение, а «мнимая» — векторное. Действительно,

$$\sigma_a\sigma_b = i\varepsilon_{abc}\sigma_c \text{ при } a\neq b,\ \{a, b, c\}=\{1, 2, 3\},$$

$$\sigma_a^2=\sigma_0=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad a=1, 2, 3,$$

так что

$$\left(\sum_{a=1}^{3} x_a\sigma_a\right)\left(\sum_{b=1}^{3} y_b\sigma_b\right)=\left(\sum_{a=1}^{3} x_a y_a\right)\sigma_0+i\left(\sum_{a=1}^{3} x_a\sigma_a\right)\times\left(\sum_{b=1}^{3} y_b\sigma_b\right),$$

или, как пишут физики,

$$(\vec{x}\cdot\vec{\sigma})(\vec{y}\cdot\vec{\sigma})=(\vec{x}\cdot\vec{y})\,\sigma_0+i(\vec{x}\times\vec{y})\cdot\vec{\sigma},\quad \vec{\sigma}=(\sigma_1, \sigma_2, \sigma_3).$$

Отсюда видно, что вещественное пространство операторов $\mathbf{R}\mathrm{id}+i\mathcal{E}$ замкнуто относительно умножения. Его базис составляют в классических обозначениях элементы

$$\mathbf{1}=\sigma_0,\ \mathbf{i}=-i\sigma_1,\ \mathbf{j}=-i\sigma_2,\ \mathbf{k}=-i\sigma_3$$

с таблицей умножения

$$\mathbf{i}^2=\mathbf{j}^2=\mathbf{k}^2=-\mathbf{1},\quad \mathbf{ij}=-\mathbf{ji}=\mathbf{k},$$

$$\mathbf{ki}=-\mathbf{ik}=\mathbf{j};\quad \mathbf{jk}=-\mathbf{kj}=\mathbf{i},$$

Иными словами, мы получаем *тело кватернионов* в одном из традиционных матричных представлений (ср. «Введение в алгебру», гл. 9, § 4).

11. Гомоморфизм SU(2)→SO(3). Фиксируем ортонормированный базис $\{h_1, h_2\}$ в $\mathcal{H}$ и соответствующий ему ортонормированный базис $\{e_1, e_2, e_3\}$ в $\mathcal{E}$, для которого $A_{e_i}=\sigma_i$. Любой унитарный оператор $U\colon \mathcal{H}\to\mathcal{H}$ переводит $\{h_1, h_2\}$ в $\{h_1', h_2'\}$, этому последнему базису отвечает базис $\{e_1', e_2', e_3'\}$, и имеется ортогональный оператор $s(U)\colon \mathcal{E}\to\mathcal{E}$, который переводит $\{e_i\}$ в $\{e_i'\}$. По следствию п. 8 $s(U)\in \mathrm{SO}(3)$, ибо определитель $s(U)$ положителен.

Реализовав $\mathcal{E}$ матрицами в базисе $\{h_1, h_2\}$, мы можем представить действие $s(U)$ на $\mathcal{E}$ простой формулой:

$$s(U)(A)=UAU^{-1}$$

для любых $A\in\mathcal{E}$. Действительно, это частный случай общей формулы замены матрицы оператора при замене базиса. Мы можем теперь доказать следующий важный результат,

12. Теорема. *Отображение s, ограниченное на* SU(2), *определяет сюръективный гомоморфизм групп* SU(2)→SO(3) *с ядром* $\{\pm E_2\}$.

Доказательство Из формулы $s(U)(A)=UAU^{-1}$ сразу видно, что $s(E)=\mathrm{id}$ и $s(UV)=s(U)s(V)$, так что s является гомоморфизмом групп. Его сюръективность проверяется так.

Выберем элемент $g\in SO(3)$ и пусть g переводит базис $(\sigma_1, \sigma_2, \sigma_3)$ в $\mathscr{E}$ в новый базис $(\sigma_1', \sigma_2', \sigma_3')$. Построим по нему базис $\{h_1', h_2'\}$ в $\mathscr{H}$, в котором операторы σ_i' задаются матрицами σ_i. По предложению п. 7 $\{h_1', h_2'\}$ существует с точностью до того, что матрица σ_2', возможно, равна $-\sigma_2$, а не σ_2. На самом деле эта возможность исключена по следствию п. 8, ибо $g\in SO(3)$ сохраняет ориентацию $\mathscr{E}$. Оператор U, переводящий $\{h_1, h_2\}$ в $\{h_1', h_2'\}$, удовлетворяет условию $s(U)=g$. Правда, он может принадлежать лишь U(2), а не SU(2). Если $\det U=e^{i\varphi}$, то $e^{-i\varphi/2}U\in SU(2)$. Матрица $e^{-i\varphi/2}U$ переводит $\{h_1, h_2\}$ в $\{e^{-i\varphi/2}h_1', e^{-i\varphi/2}h_2'\}$, а этому базису в $\mathscr{H}$ по-прежнему отвечает базис $\{\sigma_1', \sigma_2', \sigma_3'\}$ в $\mathscr{E}$. Следовательно, также $s(e^{-i\varphi/2}U)=g$, и мы получаем, что s: SU(2)→SO(3) сюръективен.

Ядро гомоморфизма s: U(2)→SO(3) состоит только из скалярных операторов $\{e^{i\varphi}\,\mathrm{id}\}$: это следует из предложения п. 7, согласно которому базис $\{h_1', h_2'\}$ восстанавливается по $\{e_1', e_2', e_3'\}$ как раз с точностью до умножения на $e^{i\varphi}$. Пересечение группы $\{e^{i\varphi}\,\mathrm{id}\}$ с SU(2) равно в точности $\{\pm\mathrm{id}\}$, что и завершает доказательство.

Смысл построенного гомоморфизма выясняется в топологии: группа SU(2) *односвязна*, т. е. любую замкнутую кривую на ней можно непрерывным движением стянуть в точку, тогда как для SO(3) это неверно. Таким образом, SU(2) является универсальным накрытием группы SO(3).

Мы воспользуемся доказанной теоремой для того, чтобы разобраться в структуре группы SO(3), играя на том, что SU(2) устроена проще. Здесь уместно процитировать Р. Фейнмана:

«Не правда ли, странно, что, живя в трех измерениях, мы все же с трудом воспринимаем, что произойдет, если сперва повернуться так, а потом еще как-нибудь. Вероятно, если бы мы были птицами или рыбами и если бы мы на собственном опыте знали, что бывает, когда все время крутишь разные сальто в пространстве, нам было бы легче воспринимать подобные вещи». (Фейнман Р., Лейтон Р. Сэндс М. Фейнмановские лекции по физике, вып. 8, гл. 4. — М.: Мир, 1978, с. 101).

13. Структура SU(2). Прежде всего, элементы SU(2) суть 2×2-матрицы с комплексными элементами, для которых $U^t=\bar{U}$ и $\det U=1$. Отсюда сразу же следует, что

$$SU(2)=\left\{\begin{pmatrix} a & b \\ -\bar{b} & \bar{a}\end{pmatrix}\middle|\, |a|^2+|b|^2=1\right\}.$$

Множество пар $\{(a, b) \mid |a|^2 + |b|^2 = 1\}$ в $\mathbf{C}^2$ превращается в сферу единичного радиуса в овеществлении $\mathbf{C}^2$, т. е. $\mathbf{R}^4$:

$$(\operatorname{Re} a)^2 + (\operatorname{Im} a)^2 + (\operatorname{Re} b)^2 + (\operatorname{Im} b)^2 = 1.$$

Итак, группа $SU(2)$ *топологически устроена как трехмерная сфера в четырехмерном евклидовом пространстве.*

Теперь напишем некоторую систему образующих группы $SU(2)$, вдохновляясь следствием п. 7 § 8, согласно которому отображение $\exp\colon \mathfrak{u}(2) \to U(2)$ сюръективно. Непосредственное вычисление экспоненты от трех образующих пространства $\mathfrak{su}(2)$ дает:

$$\exp\left(\frac{1}{2}\, it\sigma_1\right) = \begin{pmatrix} \cos\frac{t}{2} & i\sin\frac{t}{2} \\ i\sin\frac{t}{2} & \cos\frac{t}{2} \end{pmatrix},$$

$$\exp\left(\frac{1}{2}\, it\sigma_2\right) = \begin{pmatrix} \cos\frac{t}{2} & \sin\frac{t}{2} \\ -\sin\frac{t}{2} & \cos\frac{t}{2} \end{pmatrix},$$

$$\exp\left(\frac{1}{2}\, it\sigma_3\right) = \begin{pmatrix} e^{it/2} & 0 \\ 0 & e^{-it/2} \end{pmatrix}.$$

Любой элемент $\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \in SU(2)$, для которого $ab \neq 0$, можно представить в виде

$$\exp\left(\frac{1}{2}\, i\varphi\sigma_3\right)\exp\left(\frac{1}{2}\, i\theta\sigma_1\right)\exp\left(\frac{1}{2}\, i\psi\sigma_3\right) =$$

$$= \begin{pmatrix} \cos\frac{\theta}{2}\, e^{i\frac{\varphi+\psi}{2}} & i\sin\frac{\theta}{2}\, e^{i\frac{\varphi-\psi}{2}} \\ i\sin\frac{\theta}{2}\, e^{i\frac{\psi-\varphi}{2}} & \cos\frac{\theta}{2}\, e^{-i\frac{\varphi+\psi}{2}} \end{pmatrix},$$

где $0 \leqslant \varphi < 2\pi$, $0 < \theta < \pi$, $-2\pi \leqslant \psi < 2\pi$. Для этого достаточно положить $|a| = \cos\frac{\theta}{2}$, $\arg a = \frac{\varphi+\psi}{2}$, $\arg b = \frac{\varphi-\psi+\pi}{2}$. (Элементы $SU(2)$ с $b = 0$, очевидно, имеют вид $\exp\left(\frac{1}{2}\, i\varphi\sigma_3\right)$; мы оставляем читателю возможность разобраться с элементами, для которых $a = 0$.)

Углы φ, θ, ψ называются *углами Эйлера* в группе $SU(2)$.

14. Структура SO(3). Мы отождествили $SU(2)$ топологически с трехмерной сферой. При гомоморфизме $s\colon SU(2) \to SO(3)$ в одну точку $SO(3)$ переходят пары элементов $\pm U \in SU(2)$. На сфере они образуют концы одного из диаметров. Поэтому $SO(3)$ топологически есть *результат склеивания трехмерной сферы по парам противоположных точек.* С другой стороны, пары противоположных точек сферы находятся во взаимно однозначном соответствии с прямыми в четырехмерном вещественном пространстве, соединяющими точки пары. Множество таких прямых называется *трех-*

мерным вещественным проективным пространством и обозначается иногда RP^3; позже мы изучим проективные пространства подробнее. Таким образом, SO(3) топологически эквивалентна RP^3.

Посмотрим теперь, во что гомоморфизм s переводит образующие SU(2), описанные в предыдущем пункте. В стандартном базисе $\{\sigma_1, \sigma_2, \sigma_3\}$ пространства $\mathcal{E}$ имеем

$$\exp\left(\frac{1}{2}\, it\sigma_1\right)\sigma_1 \exp\left(-\frac{1}{2}\, it\sigma_1\right) = \sigma_1,$$

$$\exp\left(\frac{1}{2}\, it\sigma_1\right)\sigma_2 \exp\left(-\frac{1}{2}\, it\sigma_1\right) = (\cos t)\,\sigma_2 - (\sin t)\,\sigma_3,$$

$$\exp\left(\frac{1}{2}\, it\sigma_1\right)\sigma_3 \exp\left(-\frac{1}{2}\, it\sigma_1\right) = (\sin t)\,\sigma_2 + (\cos t)\,\sigma_3.$$

Поэтому

$$s\left(\exp\left(\frac{1}{2}\, it\sigma_1\right)\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

является вращением $\mathcal{E}$ на угол t вокруг оси $\mathbf{R}\sigma_1$. Совершенно аналогично проверяется, что $s\left(\exp\left(\frac{1}{2}\, it\sigma_k\right)\right)$ есть вращение $\mathcal{E}$ на угол t вокруг оси σ_k также для $k = 2, 3$. В частности, любое вращение из SO(3) разлагается в произведение трех вращений относительно σ_3, σ_1, σ_3 на углы Эйлера ψ, θ, φ, причем ψ можно считать меняющимся от 0 до 2π.

УПРАЖНЕНИЯ

1. Доказать тождества

$$(\vec{x}\times\vec{y},\ \vec{z}) = (\vec{x},\ \vec{y}\times\vec{z}),$$

$$(\vec{x}\times\vec{y})\times\vec{z} - \vec{x}\times(\vec{y}\times\vec{z}) = (\vec{x},\ \vec{y})\,\vec{z} - \vec{x}\,(\vec{y},\ \vec{z}).$$

(*Указание.* Воспользоваться ассоциативностью умножения в алгебре кватернионов.)

2. В трехмерном евклидовом пространстве выделены две оси z и z', образующие между собой угол φ. Пучок электронов с проекцией спина $+1/2$ на ось z подается на фильтр, пропускающий лишь электроны с проекцией спина $+1/2$ на ось z'. Показать, что доля прошедших через него электронов будет равна $\cos^2\frac{\varphi}{2}$.

§ 12. Пространство Минковского

1. *Пространством Минковского* $\mathcal{M}$ называется четырехмерное вещественное линейное пространство с невырожденной симметричной метрикой сигнатуры (1, 3) (иногда работают с сигнатурой (3, 1)). Прежде чем приступить к математическому изучению этого пространства, укажем основные принципы его физической интерпретации, лежащие в основе *специальной теории относительности* Эйнштейна.

а) *Точки.* Точка (или вектор) пространства $\mathscr{M}$ есть идеализация физического события, локализованного в пространстве и времени, типа «вспышки», «излучения фотона атомом», «столкновения двух элементарных частиц» и т. п. Начало координат $\mathscr{M}$ следует представлять себе как событие, происходящее «здесь и сейчас» для некоторого наблюдателя; оно фиксирует одновременно начало отсчета времени и начало отсчета пространственных координат.

б) *Единицы измерения.* В классической физике длины и времена измеряются в разных единицах. Поскольку $\mathscr{M}$ есть модель пространства-времени, в специальной теории относительности должен быть способ пересчета пространственных единиц во временные и наоборот. Принятый способ эквивалентен принципу «постоянства скорости света c»: он состоит в том, что выбранной единице времени t_0 ставится в соответствие единица длины $l_0 = ct_0$ — расстояние, проходимое светом за время t_0 (например, «световая секунда»). Одна из единиц l_0 или t_0 считается далее выбранной раз навсегда; после того как вторая зафиксирована условием $l_0 = ct_0$, скорость света в этих единицах становится равной 1.

в) *Пространственно-временной интервал.* Если $l_1, l_2 \in \mathscr{M}$ — две точки пространства Минковского, скалярное произведение $(l_1 - l_2, l_1 - l_2)$ называется квадратом пространственно-временно́го интервала между ними. Этот квадрат может быть положительным, нулевым или отрицательным; в физических терминах соответственно *времениподобным*, *светоподобным* или *пространственноподобным*. (Некоторое объяснение этих терминов будет дано ниже.) Если $l_2 = 0$, эти же термины применяются к вектору l_1 в зависимости от знака (l_1, l_1).

г) *Мировые линии инерциальных наблюдателей.* Если на прямой $L \subset \mathscr{M}$ хоть один вектор времениподобен, то и все векторы времениподобны. Такие прямые называются *мировыми линиями инерциальных наблюдателей.* Хорошим приближением к отрезку такой линии может служить множество событий, происходящих в космическом корабле, который движется свободно (с выключенными двигателями) вдали от небесных тел (учет их тяготения требует изменения математической схемы описания пространства-времени и перехода к «искривленным» моделям общей теории относительности). Заметим, что мы ввели пока в рассмотрение только мировые линии, исходящие из начала координат. Инерциальный наблюдатель, не бывший «здесь и сейчас», движется по некоторому сдвигу $l + L$ времениподобной прямой L. Пусть l_1, l_2 — две точки на мировой линии инерциального наблюдателя. Тогда $(l_1 - l_2, l_1 - l_2) > 0$, и интервал $|l_1 - l_2| = (l_1 - l_2, l_1 - l_2)^{1/2}$ есть *собственное время этого наблюдателя, протекшее между событиями l_1, l_2 и измеренное по показаниям движущихся вместе с ним часов.* Мировая линия инерциального наблюдателя есть его собственная «река времени».

Физический факт *направленности* времени (из прошлого в будущее) математически выражается заданием *ориентации* каждой времениподобной прямой, так что длина $|l|$ времениподобного вектора может быть снабжена знаком, отличающим векторы, направленные в будущее и в прошлое. Ниже мы увидим, что имеет смысл представление о согласованности этих ориентаций, т. е. о существовании общего направления времени — но не самих времен! — для разных инерциальных наблюдателей.

д) *Физическое пространство инерциального наблюдателя.* Линейное подмногообразие

$$\mathscr{E}_l = l + L^\perp \subset \mathscr{M}$$

интерпретируется как множество точек «мгновенного физического пространства» для инерциального наблюдателя, находящегося в точке l своей мировой линии L. Ортогональное дополнение берется, разумеется, относительно метрики Минковского в $\mathscr{M}$. Нетрудно убедиться, что $\mathscr{M} = L \oplus L^\perp$ и что на $L^\perp$ индуцируется структура трехмерного евклидова пространства (только с отрицательно определенной метрикой вместо обычной положительно определенной). Все события, отвечающие точкам $L^\perp$, интерпретируются наблюдателем как происходящие «сейчас»; для другого наблюдателя они не будут одновременными, ибо $L_1^\perp \neq L_2^\perp$ при $L_1 \neq L_2$.

е) *Инерциальные системы координат.* Пусть L — времениподобная прямая с ориентацией, e_0 — положительно ориентированный вектор на ней длины единица, $\{e_1, e_2, e_3\}$ — ортонормированный базис в $L^\perp$: $(e_i, e_i) = -1$ для $i = 1, 2, 3$. Система координат в $\mathscr{M}$, отвечающая базису $\{e_0, \ldots, e_3\}$, называется *инерциальной системой.* В ней

$$\left(\sum_{i=0}^{3} x_i e_i, \ \sum_{i=0}^{3} y_i e_i\right) = x_0 y_0 - \sum_{i=1}^{3} x_i y_i .$$

Поскольку $x_0 = ct_0$, где t_0 — собственное время, пространственно-временной интервал от начала до точки $\sum\limits_{i=0}^{3} x_i e_i$ равен $\left(c^2 t_0^2 - \right.$ $\left. - \sum\limits_{i=1}^{3} x_i^2\right)^{1/2}$. Каждая инерциальная система координат в $\mathscr{M}$ определяет отождествление $\mathscr{M}$ с координатным пространством Минковского $\left(\mathbf{R}^4, \ x_0^2 - \sum\limits_{i=1}^{3} x_i^2\right)$. *Изометрии* $\mathscr{M}$ (или координатного пространства) образуют *группу Лоренца*; изометрии, сохраняющие ориентацию во времени, — ее *ортохронную подгруппу.*

ж) *Световой конус.* Множество точек $l \in \mathscr{M}$ с $(l, l) = 0$ называется *световым конусом* C (начала координат). В любой инерциальной системе координат C задается уравнением

$$x_0^2 = \sum_{i=1}^{3} x_i^2.$$

При $x_0 > 0$ точка (x_0, x_1, x_2, x_3) на световом конусе отделена от положения наблюдателя $(x_0, 0, 0, 0)$ пространственноподобным интервалом с квадратом $-\sum_{i=1}^{3} x_i^2 = -x_0^2$, т. е. находится на расстоянии, которое за время x_0 пройдет квант света, выпущенный из начала координат в начальный момент времени. (При $x_0 < 0$ множество таких точек отвечает вспышкам, которые произошли в момент собственного времени x_0 и могли наблюдаться в точке начала отсчета: «приходящее излучение».) Соответственно «нулевые прямые», целиком лежащие в C, — это мировые линии частиц, испущенных из начала координат и летящих со скоростью света, например, фотонов. Читатель может увидеть базу «приходящей полы» светового конуса, выглянув в окно, — это небесная сфера.

Прямые в $\mathscr{M}$, состоящие из векторов с отрицательным квадратом длины, не имеют физической интерпретации. Они должны были бы отвечать мировым линиям частиц, летящих быстрее света, — гипотетических «тахионов», не обнаруженных экспериментально.

Перейдем теперь к математическому изучению $\mathscr{M}$.

2. *Реализация $\mathscr{M}$ как пространства метрик.* Как в § 9, фиксируем двумерное комплексное пространство $\mathscr{H}$ и рассмотрим на нем множество $\mathscr{M}$ *эрмитово симметричных скалярных произведений.* Оно является вещественным линейным пространством. Если выбран базис $\{h_1, h_2\}$ в $\mathscr{H}$, то матрицы Грама этих метрик будут всевозможными эрмитовыми 2×2-матрицами. Поставим в соответствие метрике $l \in \mathscr{M}$ определитель ее матрицы Грама G, который будем обозначать $\det l$. Переход к базису $\{h_1', h_2'\} = \{h_2, h_2\} V$ приведет к замене G на $G' = V^t G \bar{V}$, и $\det G' = |\det V|^2 \det G$. В частности, если $V \in \mathrm{SL}(2, \mathbf{C})$, то $\det G = \det G'$. Поэтому вычисление $\det l$ в любом из базисов $\mathscr{H}$, лежащих в одном классе относительно действия $\mathrm{SL}(2, \mathbf{C})$, приведет к одному и тому же результату. Впредь мы *фиксируем такой класс базисов $\mathscr{H}$*, и все $\det$ будем вычислять относительно него. Замена класса только умножает $\det$ на положительный скаляр.

3. Предложение. а) *$\mathscr{M}$ является четырехмерным вещественным пространством.*

б) *На $\mathscr{M}$ имеется единственная симметричная метрика (l, m), для которой $(l, l) = \det l$. Ее сигнатура равна $(1, 3)$, так что $\mathscr{M}$ представляет собой пространство Минковского.*

Доказательство. а) Пространство эрмитовых 2×2-матриц имеет базис $\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E_2$; $\sigma_1, \sigma_2, \sigma_3$, где σ_i, $i \geqslant 1$, — матрицы Паули. Поэтому $\dim \mathscr{M} = 4$.

б) Покажем, что в матричной реализации $\mathscr{M}$ функция $\det l$ представляет собой квадратичную форму, поляризация которой имеет вид

$$(l, m) = \frac{1}{2} (\operatorname{Tr} l \operatorname{Tr} m - \operatorname{Tr} lm),$$

явно симметричный и билинейный. В самом деле, если λ, μ — собственные значения l, то $\det l = \lambda\mu$, $\operatorname{Tr} l = \lambda + \mu$, $\operatorname{Tr} l^2 = \lambda^2 + \mu^2$, так что

$$\lambda\mu = \det l = \frac{1}{2}((\lambda+\mu)^2 - \lambda^2 - \mu^2) = \frac{1}{2}((\mathrm{Tr}\, l)^2 - \mathrm{Tr}\, l^2) = (l, l).$$

Теперь очевидно, что $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ является ортонормированным базисом $\mathcal{M}$ с матрицей Грама $\mathrm{diag}(1, -1, -1, -1)$, так что сигнатура нашей метрики равна $(1, 3)$. Это завершает доказательство.

4. Следствие. *Пусть $L \subset \mathcal{M}$ — времениподобная прямая. Тогда $L^\perp$ с метрикой $-(l, m)$ является трехмерным евклидовым пространством, и $\mathcal{M} = L \oplus L^\perp$.*

Доказательство. Утверждение $\mathcal{M} = L \oplus L^\perp$ следует из предложения п. 2 § 3, ибо времениподобные прямые, очевидно, невырождены. Так как сигнатура метрики Минковского на $\mathcal{M}$ есть $(1, 3)$, а на $L^\perp$ — $(1, 0)$, на $L^\perp$ она должна быть $(0, 3)$, что завершает доказательство.

Перейдем теперь к изучению геометрического смысла скалярных произведений. Неопределенность метрики Минковского приводит к замечательным отличиям от евклидовой ситуации, которые имеют важный физический смысл. Самые яркие факты связаны с тем, что неравенство Коши — Буняковского — Шварца для времениподобных векторов оказывается обращенным в другую сторону.

5. Предложение. *Пусть $(l_1, l_1) > 0$, $(l_2, l_2) > 0$, $l_i \in \mathcal{M}$. Тогда*

$$(l_1, l_2)^2 \geqslant (l_1, l_1)(l_2, l_2).$$

Равенство достигается тогда и только тогда, когда l_1, l_2 линейно зависимы.

Доказательство. Прежде всего проверим, что квадратный трехчлен $(tl_1 + l_2, tl_1 + l_2)$ всегда имеет вещественный корень t_0. В матричной реализации $\mathcal{M}$ условие $(l_2, l_2) > 0$ означает, что $\det l_2 > 0$, т. е. что l_2 имеет вещественные характеристические корни одного знака, скажем, ε_2 $(+1$ или $-1)$. Аналогично, пусть ε_1 — знак собственных значений l_1. Тогда при $t \to -(\varepsilon_1\varepsilon_2)\infty$ матрица $tl_1 + l_2$ имеет собственные значения, примерно пропорциональные собственным значениям l_1 (ибо $l_1 + t^{-1}l_2$ стремится к l_1), и их знак будет $-\varepsilon_2$, а при $t = 0$ матрица $0l_1 + l_2 = l_2$ имеет собственные значения знака ε_2. Следовательно, при изменении t от 0 до $-(\varepsilon_1\varepsilon_2)\infty$ собственные значения $tl_1 + l_2$ проходят через нуль, и $\det(tl_1 + l_2)$ обращается в нуль. Значит, дискриминант этого трехчлена неотрицателен, так что

$$(l_1, l_2)^2 \geqslant (l_1, l_1)(l_2, l_2).$$

Если он равен нулю, то некоторое значение $t_0 \in \mathbf{R}$ является двукратным корнем, и матрица $t_0 l_1 + l_2$, имея два нулевых собственных значения и будучи диагонализируемой (она эрмитова!), равна нулю. Поэтому l_1 и l_2 линейно зависимы.

6. Следствие («неравенство треугольника в обратную сторону»). *Если l_1, l_2 времениподобны и $(l_1, l_2) \geqslant 0$, то $l_1 + l_2$ времениподобен и*

$$|l_1 + l_2| \geqslant |l_1| + |l_2|$$

(*где* $|l|=(l, l)^{1/2}$), *и равенство достигается тогда и только тогда, когда* l_1, l_2 *линейно зависимы.*

Доказательство.

$$|l_1+l_2|^2=|l_1|^2+2(l_1, l_2)+|l_2|^2 \geqslant$$
$$\geqslant |l_1|^2+2|l_1|\,|l_2|+|l_2|^2=(|l_1|+|l_2|)^2.$$

Равенство достигается лишь при $(l_1, l_2)=|l_1|\,|l_2|$.

Дадим теперь физические интерпретации этих фактов.

7. «Парадокс близнецов». Времениподобные векторы l_1, l_2 с $(l_1, l_2)\geqslant 0$ назовем *одинаково временно ориентированными.* Из предложения п. 5 видно, что для них $(l_1, l_2)>0$. Вообразим двух близнецов-наблюдателей: один инерциален и движется по своей мировой линии от точки 0 до точки l_1+l_2, другой доходит до той же точки от начала отсчета, двигаясь сначала инерциально от 0 до l_1 и затем от l_1 до l_1+l_2: вблизи нуля и вблизи l_1 он включает двигатели своего космического корабля, чтобы сначала улететь от брата, а затем снова вернуться к нему. Согласно следствию п. 6 собственное время, протекшее для путешествующего брата, будет строго меньше времени, протекшего по часам домоседа.

8. Множитель Лоренца. Если l_1 и l_2 времениподобны и одинаково временно ориентированы, то по предложению п. 5 $\frac{(l_1, l_2)}{|l_1|\,|l_2|}\geqslant 1$, и мы не можем интерпретировать эту величину как косинус угла. Чтобы понять, что она собой представляет, снова прибегнем к физической интерпретации.

Пусть $|l_1|=1$, $|l_2|=1$; в частности, инерциальный наблюдатель l_1 прожил единицу собственного времени с момента начала отсчета. В точке l_1 физическое пространство одновременных событий для него есть $l_1+(\mathbf{R}l_1)^{\perp}$. Мировая линия наблюдателя $\mathbf{R}l_2$ пересекает это пространство в точке xl_2, где x находится из условия

$$(xl_2-l_1,\ l_1)=0,$$

т. е. $x=(l_1, l_2)^{-1}$. Расстояние от l_1 до xl_2 пространственноподобно: для наблюдателя $\mathbf{R}l_1$ — это расстояние, на которое $\mathbf{R}l_2$ удалился от него за единицу времени, т. е. относительная скорость $\mathbf{R}l_2$. Она равна (учесть, что у метрики в $(\mathbf{R}l_1)^{\perp}$ следует изменить знак!)

$$v=[-(xl_2-l_1,\ xl_2-l_1)]^{1/2}=[-(xl_2-l_1,\ xl_2)]^{1/2}=$$
$$=[-x^2(l_2, l_2)+x(l_1, l_2)]^{1/2}=[-(l_1, l_2)^{-2}+1]^{1/2},$$

откуда

$$(l_1,\ l_2)=\frac{1}{\sqrt{1-v^2}}.$$

Это знаменитый *множитель Лоренца*; часто его пишут в виде $\frac{1}{\sqrt{1-v^2/c^2}}$, явно указывая, что скорости измеряются по отношению к скорости света. В частности,

$$x=\frac{1}{(l_1,\ l_2)}=\sqrt{1-v^2},$$

т. е. в момент собственного времени единица для первого наблюдателя второй наблюдатель находится в его физическом пространстве, когда часы второго наблюдателя показывают $\sqrt{1-v^2}$. Это — количественное выражение эффекта «сокращения времени» для движущегося наблюдателя, качественно описанное в предыдущем пункте.

9. Евклидовы углы. В пространстве $(\mathbf{R}l_0)^\perp$, где l_0 — времениподобный вектор, геометрия евклидова, и там скалярное произведение имеет обычный смысл. Пусть l_1, l_2 — еще два времениподобных вектора с той же ориентацией. Мы можем спроектировать их на $(\mathbf{R}l_0)^\perp$ и вычислить косинус угла между проекциями. Предоставляем читателю самостоятельно убедиться в том, что для наблюдателя $\mathbf{R}l_0$ это — угол между направлениями отлета от него наблюдателей $\mathbf{R}l_1$ и $\mathbf{R}l_2$ в его физическом пространстве. Абсолютного значения этот угол не имеет; другой наблюдатель $\mathbf{R}l'_0$ увидит его другим.

10. Четыре ориентации пространства Минковского. Пусть $\{e_i\}$, $\{e'_i\}$, $i=0, \ldots, 3$, — два ортонормированных базиса в $\mathcal{M}$: $(e_0, e_0)=(e'_0, e'_0)=1$, $(e_i, e_i)=(e'_i, e'_i)=-1$ при $i=1, \ldots, 3$. По аналогии с прежними определениями назовем их *одинаково ориентированными*, если один переводится в другой непрерывной системой *изометрий* $f_t\colon \mathcal{M}\to\mathcal{M}$, $0\leqslant t\leqslant 1$, $f_0=\mathrm{id}$, $f_1(e_i)=e'_i$. Два условия одинаковой ориентированности, очевидно, необходимы:

а) $(e_0, e'_0)>0$. Действительно, $(e_0, f_t(e_0))^2\geqslant 1$ по предложению п. 5, так что знак $(e_0, f_t(e_0))$ не может меняться при изменении t, а $(e_0, f_0(e_0))=1$. Выше мы назвали e_0 и e'_0 с таким свойством одинаково временно ориентированными.

б) Определитель отображения ортогональной проекции $\sum\limits_{i=1}^{3}\mathbf{R}e_i\to\sum\limits_{i=1}^{3}\mathbf{R}e'_i$, записанного в базисах $\{e_i\}$ или $\{e'_i\}$, положителен.

Действительно, проекция $\sum\limits_{i=1}^{3}\mathbf{R}e_i\to\sum\limits_{i=1}^{3}\mathbf{R}f_t(e_i)$ невырождена ни при каком значении t: иначе пространственноподобный вектор из $\sum\limits_{i=1}^{3}\mathbf{R}e_i$ был бы ортогонален $\sum\limits_{i=1}^{3}\mathbf{R}f_t(e'_i)=(f_t(e'_0))^\perp$, т. е. пропорционален $f_t(e'_0)$ — времениподобному вектору; это невозможно. Значит, определители этих проекций при всех t имеют одинаковый знак, а при $t=0$ он положителен.

Можно сказать, что пары базисов со свойством б) *одинаково пространственно ориентированы.*

Наоборот, если два ортонормированных базиса в $\mathcal{M}$ имеют одинаковую пространственную и временну́ю ориентацию, то они одинаково ориентированы, т. е. переводятся друг в друга непрерывной системой изометрий f_t. Чтобы построить ее, положим прежде всего $f_t(e_0)=\dfrac{te_0'+(1-t)\,e_0}{|te_0'+(1-t)\,e_0|}$. Из условия $(e_0, e_0')\geqslant 1$ следует, что $f_t(e_0)$ времениподобен и имеет квадрат длины единица при всех $0\leqslant t\leqslant 1$. Далее, в качестве $f_t(e_1, e_2, e_3)$ выберем ортонормированный базис в $f_t(e_0)^{\perp}$, получающийся из проекции $\{e_1, e_2, e_3\}$ на $f_t(e_0)^{\perp}$ процессом ортогонализации Грама — Шмидта; очевидно, он непрерывно зависит от t. Ясно, что $f_1(e_0)=e_0'$, а $\{f_1(e_1), f_1(e_2), f_1(e_3)\}$ и $\{e_1', e_2', e_3'\}$ суть одинаково ориентированные ортонормированные базисы в $(e_0')^{\perp}$. Их можно перевести друг в друга непрерывным семейством чисто евклидовых вращений $(e_0')^{\perp}$, оставляющих e_0' неподвижным. Это завершает доказательство.

Обозначим через Λ *группу Лоренца*, т. е. группу изометрий пространства $\mathscr{M}$, или $\mathrm{O}(1,3)$. Пусть далее $\Lambda_+^{\uparrow}$ — подгруппа Λ, сохраняющая ориентацию некоторого ортонормированного базиса; $\Lambda_-^{\uparrow}$ — подмножество Λ, меняющее его пространственную, но не временну́ю ориентацию; $\Lambda_+^{\downarrow}$ — подмножество Λ, меняющее его временну́ю, но не пространственную ориентацию; $\Lambda_-^{\downarrow}$ — подмножество Λ, меняющее его временну́ю и пространственную ориентации. Нетрудно убедиться, что от выбора исходного базиса эти подмножества не зависят. Мы доказали следующий результат:

11. Теорема. *Группа Лоренца Λ состоит из четырех связных компонент*: $\Lambda=\Lambda_+^{\uparrow}\cup\Lambda_-^{\uparrow}\cup\Lambda_+^{\downarrow}\cup\Lambda_-^{\downarrow}$.

Тождественное отображение лежит, очевидно, в $\Lambda_+^{\uparrow}$. Аналогом теоремы п. 12 § 11 является следующий результат.

12. Теорема. *Реализуем $\mathscr{M}$ как пространство матриц Грама эрмитовых метрик в $\mathscr{H}$ в базисе $\{h_1, h_2\}$. Для любой матрицы $V\in\mathrm{SL}(2,\mathbf{C})$ поставим в соответствие матрице $l\in\mathscr{M}$ новую матрицу*

$$s(V)\,l=V^t l\overline{V}.$$

Отображение s определяет сюръективный гомоморфизм $\mathrm{SL}(2,\mathbf{C})$ на $\Lambda_+^{\uparrow}$ с ядром $\{\pm E_2\}$.

Доказательство. Очевидно, что $s(V)\,l$ линейно по l и сохраняет квадраты длин: $\det(V^t l\overline{V})=\det l$. Поэтому $s(V)\in\Lambda$. Так как группа $\mathrm{SL}(2,\mathbf{C})$ связна, любой ее элемент можно непрерывно деформировать в единичный, оставаясь внутри $\mathrm{SL}(2,\mathbf{C})$, — преобразование Лоренца $s(V)$ можно непрерывно деформировать в тождественное, так что $s(V)\in\Lambda_+^{\uparrow}$. Поскольку $s(\mathrm{id})=\mathrm{id}$ и $s(V_1V_2)=s(V_1)s(V_2)$, s является гомоморфизмом групп. Если $V^t l\overline{V}=l$ для всех $l\in\mathscr{M}$, то, в частности, $V^t\sigma_i\overline{V}=\sigma_i$, где $\sigma_0=E_2$, σ_1, σ_2, σ_3 — матрицы Паули. Условие $V^t\overline{V}=E_2$ означает, что V унитарна; после этого условия $V^t\sigma_i\overline{V}=V^t\sigma_i(V^t)^{-1}=\sigma_i$ означает, что $V=\pm E_2$: это было доказано в п. 12 § 11. Таким образом, $\operatorname{Ker} s=\{\pm E_2\}$.

Осталось установить, что s сюръективен. Пусть $f\colon \mathscr{M}\to\mathscr{M}$ — преобразование Лоренца из $\Lambda_+^\uparrow$, переводящее ортонормированный базис $\{e_i\}$ в $\{e_i'\}$. Метрики на $\mathscr{H}$, отвечающие e_0 и e_0', определены, ибо собственные значения как e_0, так и e_0' имеют одинаковый знак, потому что $\det e_0 = \det e_0' = 1$. Из $(e_0, e_0') > 0$ следует, что эти метрики одновременно положительно или отрицательно определены. Действительно, выше мы убедились, что соединяющий их отрезок $te_0' + (1 - te_0)$, $0 \leqslant t \leqslant 1$, целиком состоит из времениподобных векторов. Отсюда уже вытекает существование такой матрицы $V \in \mathrm{SL}(2, \mathbf{C})$, что $s(V)$ переводит e_0 в e_0', т. е. $e_0' = V^t e_0 V$, где e_0 и e_0' отождествлены с их матрицами Грама. Действительно, V — это матрица изометрии $(\mathscr{H}, e_0)$ с $(\mathscr{H}, e_0')$; априори ее определитель может быть равен -1, но это противоречило бы возможности соединить V с E_2 в $\mathrm{SL}(2, \mathbf{C})$ с помощью деформации V_q, где $e_0' = (V_q)^t f_q(e_0) V_q$, f_q — соответствующая деформация в $\Lambda_+^\uparrow$.

Итак, $s(V)$ переводит e_0 в e_0'. Дальше остается показать, что евклидов поворот $\{s(V)e_1, s(V)e_2, s(V)e_3\}$ в $\{e_1', e_2', e_3'\}$ можно осуществить с помощью $s(U)$, где $U \in \mathrm{SL}(2, \mathbf{C})$ и $s(U)$ оставляет e_0 на месте. Можно считать, что e_0' представлен матрицей σ_0 в базисе $\{h_1, h_2\}$. Тогда мы должны выбрать U унитарной с условием $U(s(V)e_i)U^{-1} = e_i'$ для $i = 1, 2, 3$. Это можно сделать по теореме п. 12 § 11, ибо базисы $\{s(V)e_i\}$ и $\{e_i'\}$, $i = 1, 2, 3$, в $(e_0')^\perp$ ортонормированы и одинаково ориентированы. Доказательство окончено.

13. Евклидовы повороты и бусты. Пусть e_0, e_0' — два одинаково временно ориентированных времениподобных вектора длины единица, L_0, L_0' — ортогональные дополнения к ним. Имеется стандартное преобразование Лоренца из $\Lambda_+^\uparrow$, переводящее e_0 в e_0', которое в физической литературе называется *бустом.* При $e_0 = e_0'$ это — тождественное преобразование. При $e_0 \neq e_0'$ оно определяется так: рассмотрим плоскость $(L_0 \cap L_0')^\perp$. Она содержит e_0 и e_0'. Сигнатура метрики Минковского на ней равна $(1, 1)$. Поэтому существует пара единичных пространственноподобных векторов $e_1, e_1' \in (L_0 \cap L_0')^\perp$, ортогональных к e_0 и e_0' соответственно. Буст оставляет на месте все векторы из $L_0 \cap L_0'$ и переводит e_0 в e_0', e_1 в e_1' соответственно. Чтобы вычислить элементы матрицы перехода $\{e_0, e_1\}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \{e_0', e_1'\}$, заметим прежде всего, что $a = (e_0, e_0') = \dfrac{1}{\sqrt{1 - v^2}}$, где v — скорость относительного удаления инерциальных наблюдателей, отвечающих e_0 и e_0'. Далее, матрицы Грама $\{e_0, e_1\}$ и $\{e_0', e_1'\}$ суть $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, поэтому

$$a^2 - b^2 = 1, \quad ac - bd = 0, \quad c^2 - d^2 = -1.$$

Из первого уравнения, зная a, находим $b = \frac{v}{\sqrt{1-v^2}}$. Добавляя сюда условие, что определитель буста $ad - bc$ равен единице, получаем $d = a$, $c = b$. Окончательно, матрица буста в базисе $\{e_0, e_1, e_2, e_3\}$, где $\{e_2, e_3\}$ — ортонормированный базис $(L_0 \cap L_0')^\perp$, имеет вид

$$\begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & \frac{v}{\sqrt{1-v^2}} & 0 & 0 \\ \frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

или в терминах пространственно-временны́х координат

$$x_0 = \frac{x_0' + vx_1'}{\sqrt{1-v^2}}, \quad x_1 = \frac{vx_0' + x_1'}{\sqrt{1-v^2}}, \quad x_2 = x_2', \quad x_3 = x_3'.$$

Стоящую в левом верхнем углу матрицу можно записать также как матрицу «гиперболического поворота»

$$\begin{pmatrix} \operatorname{ch}\theta & \operatorname{sh}\theta \\ \operatorname{sh}\theta & \operatorname{ch}\theta \end{pmatrix},$$

найдя θ из условий

$$\operatorname{ch}\theta = \frac{e^\theta + e^{-\theta}}{2} = \frac{1}{\sqrt{1-v^2}}, \quad \operatorname{sh}\theta = \frac{e^\theta - e^{-\theta}}{2} = \frac{v}{\sqrt{1-v^2}}.$$

Если исходить из двух одинаково ориентированных ортонормированных базисов $\{e_0, e_1, e_2, e_3\}$ и $\{e_0', e_1', e_2', e_3'\}$, то преобразование Лоренца, переводящее один в другой, можно представить в виде произведения буста, переводящего e_0 в e_0', и затем евклидова поворота в $(e_0')^\perp$, который переводит образ базиса $\{e_1, e_2, e_3\}$ после буста в базис $\{e_1', e_2', e_3'\}$, оставляя $e_0^\perp$ на месте.

14. Пространственные и временны́е отражения. Любое трехмерное подпространство $L \subset \mathcal{M}$, на котором метрика Минковского (анти)евклидова (т. е. прямая $L^\perp$ времениподобна), определяет преобразование Лоренца, тождественное на L и меняющее знак на $L^\perp$. Все такие операторы называются *отражениями времени*.

Любое трехмерное подпространство $L \subset \mathcal{M}$, на котором метрика Минковского имеет сигнатуру $(1, 2)$ (т. е. прямая $L^\perp$ пространственноподобна), также определяет преобразование Лоренца, тождественное на L и меняющее знак на $L^\perp$. Все такие операторы называются *пространственными отражениями*.

Если фиксировать какое-нибудь отражение времени T и пространства P, то все элементы из $\Lambda_+^\downarrow$, $\Lambda_-^\uparrow$, $\Lambda_-^\downarrow$ будут получаться из элементов $\Lambda_+^\uparrow$ умножением на T, P, PT соответственно.

§ 13. Симплектические пространства

1. В этом параграфе мы будем рассматривать конечномерные линейные пространства L над полем $\mathcal{K}$ характеристики $\neq 2$, снабженные *невырожденным* кососимметрическим скалярным произведением $[\ ,\ ]$: $L \times L \to \mathcal{K}$, и называть их *симплектическими пространствами*. Напомним свойства симплектических пространств, которые уже были установлены ранее, в § 3.

Размерность симплектического пространства всегда четна. Если она равна $2r$, то в пространстве существует симплектический базис $\{e_1, \ldots, e_r; e_{r+1}, \ldots, e_{2r}\}$, т. е. базис с матрицей Грама вида

$$\begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix}.$$

В частности, все симплектические пространства одинаковой размерности над общим полем скаляров изометричны.

Подпространство $L_1 \subset L$ называется *изотропным*, если ограничение скалярного произведения $[\ ,\ ]$ на него тождественно равно нулю. Все одномерные подпространства изотропны.

2. Предложение. *Пусть L — симплектическое пространство размерности $2r$, $L_1 \subset L$ — изотропное подпространство размерности r_1. Тогда $r_1 \leqslant r$, и если $r_1 < r$, то L_1 содержится в изотропном подпространстве максимальной возможной размерности r.*

Доказательство. Поскольку форма $[\ ,\ ]$ невырождена, она определяет изоморфизм $L \to L^*$, при котором вектору $l \in L$ ставится в соответствие линейный функционал $l' \mapsto [l, l']$. Отсюда следует, что для любого подпространства $L_1 \subset L$ имеем $\dim L_1^{\perp} = \dim L - \dim L_1$ (ср. § 7 ч. 1). Если к тому же L_1 изотропно, то $L_1 \subset L_1^{\perp}$, откуда $r_1 = \dim L_1 \leqslant \dim L_1^{\perp} = \dim L - \dim L_1 = 2r - r_1$, так что $r_1 \leqslant r$.

Рассмотрим теперь ограничение формы $[\ ,\ ]$ на $L_1^{\perp}$. Во всем пространстве L ортогональное дополнение к $L_1^{\perp}$ имеет размерность $\dim L - \dim L_1^{\perp} = \dim L_1$ по предыдущему рассуждению. С другой стороны, L_1 лежит в этом ортогональном дополнении и потому совпадает с ним. Значит, L_1 есть в точности ядро ограничения $[\ ,\ ]$ на $L_1^{\perp}$. Но в $L_1^{\perp}$ имеется симплектический базис в том его варианте, который рассматривался в § 3, где допускались вырожденные пространства:

$$\{e_1, \ldots, e_{r-r_1}; e_{r-r_1+1}, \ldots, e_{2(r-r_1)}; e_{2(r-r_1)+1}, \ldots, e_{2r-r_1}\},$$

с матрицей Грама

$$\left(\begin{array}{c|c|c} 0 & E_{r-r_1} & 0 \\ \hline -E_{r-r_1} & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

Размер единичной клетки есть $\frac{1}{2}(\dim L_1^{\perp} - \dim L_1) = r - r_1$. Векторы $e_{2(r-r_1)+1}, \ldots, e_{2r-r_1}$ порождают ядро формы на $L_1^{\perp}$, т. е. L_1;

добавив к ним, например, $e_1, \ldots, e_{r-r_1}$, получим r-мерное изотропное подпространство, содержащее L_1.

3. Предложение. *Пусть L — симплектическое пространство размерности $2r$, $L_1 \subset L$ — изотропное подпространство размерности r. Тогда существует другое изотропное подпространство $L_2 \subset L$ размерности r такое, что $L = L_1 \oplus L_2$, и скалярное произведение индуцирует изоморфизм $L_2 \to L_1^*$.*

Доказательство. Мы докажем несколько более сильный результат, полезный в приложениях, а именно установим существование подпространства L_2 среди конечного числа изотропных подпространств, связанных с фиксированным симплектическим базисом $\{e_1, \ldots, e_r; e_{r+1}, \ldots, e_{2r}\}$ в L.

Именно, пусть дано разбиение $\{1, \ldots, r\} = I \cup J$ на два непересекающихся подмножества. Тогда r векторов $\{e_i, e_{r+j} \mid i \in I,\ j \in J\}$ порождают r-мерное изотропное подпространство в L, называемое координатным (относительно выбранного базиса). Очевидно, их имеется 2^r. Покажем, что L_2 можно найти среди координатных подпространств.

Пусть M натянуто на $\{e_1, \ldots, e_r\}$ и $\dim(L_1 \cap M) = s$, $0 \leqslant s \leqslant r$. Существует такое подмножество $I \subset \{1, \ldots, r\}$ из $r - s$ элементов, что $L_1 \cap M$ трансверсально к N, натянутому на $\{e_i \mid i \in I\}$, т. е. $L_1 \cap M \cap N = \{0\}$. Действительно, множество $\{$базис $L_1 \cap M\} \cup \{e_1, \ldots, e_r\}$ порождает M, поэтому базис $L_1 \cap M$ можно дополнить до базиса M с помощью $r - s$ векторов из $\{e_1, \ldots, e_r\}$ по предложению п. 10 § 2 ч. 1. Номера этих векторов образуют искомое I, ибо $L_1 \cap M + N = M$, так что $L_1 \cap M \cap N = \{0\}$.

Положим теперь $J = \{1, \ldots, r\} \setminus I$ и покажем, что изотропное подпространство L_2, натянутое на $\{e_i, e_{r+j} \mid i \in I,\ j \in J\}$, является прямым дополнением к L_1. Достаточно проверить, что $L_1 \cap L_2 = \{0\}$. Действительно, из доказательства предложения п. 2 следует, что $L_1^\perp = L_1$, $L_2^\perp = L_2$. Но $L_1 \cap M$ содержится в L_1, N содержится в L_2, так что сумма $M = L_1 \cap M + N$ ортогональна к $L_1 \cap L_2$. Но M изотропно размерности r, поэтому $M^\perp = M$, и $L_1 \cap L_2 \subset M$. Значит, окончательно

$$L_1 \cap L_2 = (L_1 \cap M) \cap (L_2 \cap M) = (L_1 \cap M) \cap N = \{0\}.$$

Линейное отображение $L_2 \to L_1^*$ ставит в соответствие вектору $l \in L_2$ линейную форму $m \mapsto [l, m]$ на L_1. Оно является изоморфизмом, ибо $\dim L_2 = \dim L_1^* = r$, а его ядро содержится в ядре формы $[\ ,\ ]$, которая, по предположению, невырождена. Это завершает доказательство.

4. Следствие. *Любые пары взаимно дополнительных изотропных подпространств в L одинаково расположены: если $L = L_1 \oplus L_2 = L_1' \oplus L_2'$, то существует изометрия $f\colon L \to L$ такая, что $f(L_1) = L_1'$, $f(L_2) = L_2'$.*

Доказательство. Выберем базис $\{e_1, \ldots, e_r\}$ в L_1 и двойственный к нему базис $\{e_{r+1}, \ldots, e_{2r}\}$ в L_2 относительно описанного выше отождествления $L_1 \to L_2$. Очевидно, $\{e_1, \ldots, e_{2r}\}$ есть симплектический базис в L. Аналогично построим симплектический базис $\{e'_1, \ldots, e'_{2r}\}$ по разложению $L'_1 \oplus L'_2$. Линейное отображение $f: e_i \mapsto e'_i$, $i = 1, \ldots, 2r$, очевидно, является требуемой изометрией.

Из этого следствия и предложений пп. 2, 3 следует, что любые изотропные подпространства одинаковой размерности в L переводятся одно в другое подходящей изометрией.

5. Симплектическая группа. Множество всех изометрий $f: L \to L$ симплектического пространства образует группу. Множество матриц, представляющих эту группу в симплектическом базисе $\{e_1, \ldots, e_{2r}\}$, называется *симплектической группой* и обозначается $\mathrm{Sp}(2r, \mathcal{K})$, если $\dim L = 2r$. Условие $A \in \mathrm{Sp}(2r, \mathcal{K})$ равносильно тому, что матрица Грама базиса $\{e_1, \ldots, e_{2r}\}A$ совпадает с $I_{2r} = \begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix}$, т. е. что $A^t I_{2r} A = I_{2r}$, так что $\det A = \pm 1$; ниже мы докажем, что $\det A = 1$ (см. п. 11). Поскольку $I_{2r}^2 = -E_{2r}$, это условие можно записать также в виде $A = -I_{2r}(A^t)^{-1} I_{2r}$. Отсюда вытекает

6. Предложение. *Характеристический многочлен* $P(t) = \det(tE_{2r} - A)$ *симплектической матрицы* A *возвратен, т. е.* $P(t) = t^{2r} P(t^{-1})$.

Доказательство. Имеем, пользуясь тем, что $\det A = 1$,

$$\det(tE_{2r} - A) = \det(tE_{2r} + I_{2r}(A^t)^{-1} I_{2r}) = \det(tE_{2r} - (A^t)^{-1}) =$$
$$= \det(tA^t - E_{2r}) = t^{2r} \det(t^{-1}E_{2r} - A^t) = t^{2r} \det(t^{-1}E_{2r} - A).$$

7. Следствие. *Если* $\mathcal{K} = \mathbf{R}$ *и* A — *симплектическая матрица, то вместе с каждым собственным значением* λ *у* A *есть собственные значения* λ^{-1}, $\bar{\lambda}$ *и* $\bar{\lambda}^{-1}$.

Доказательство. Поскольку A невырождена, $\lambda \neq 0$ и $P(\lambda^{-1}) = \lambda^{-2r} P(\lambda) = 0$. Поскольку коэффициенты P вещественны, $P(\bar{\lambda}) = \overline{P(\lambda)} = 0$.

Комплексное сопряжение есть симметрия относительно вещественной оси, а отображение $\lambda \mapsto \bar{\lambda}^{-1}$ — симметрия относительно единичной окружности. Значит, комплексные собственные значения A появляются четверками, симметричными одновременно относительно вещественной оси и единичной окружности, а вещественные собственные значения — парами.

8. Пфаффиан. Пусть $\mathcal{K}^{2r}$ — координатное пространство, A — невырожденная кососимметрическая матрица порядка $2r$ над $\mathcal{K}$. Скалярное произведение $[\vec{x}, \vec{y}] = \vec{x}^t A \vec{y}$ в $\mathcal{K}^{2r}$ невырождено и кососимметрично. Переходя от исходного базиса к симплектическому, получаем, что для матрицы A найдется такая невырожденная матрица B, что

$$A = B^t \begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix} B,$$

откуда $\det A = (\det B)^2$. Итак, определитель каждой кососимметрической матрицы является точным квадратом. Это наводит на мысль попытаться извлечь из определителя квадратный корень, который был бы *универсальным многочленом* от элементов A. Это действительно возможно.

9. Теорема. *Существует единственный многочлен с целыми коэффициентами* $\mathrm{Pf}\, A$ *от элементов кососимметрической матрицы* A *такой, что* $\det A = (\mathrm{Pf}\, A)^2$ *и* $\mathrm{Pf}\begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix} = 1$. *Этот многочлен называется* ***пфаффианом*** *и обладает следующим свойством*:

$$\mathrm{Pf}\,(B^t AB) = \det B \cdot \mathrm{Pf}\, A$$

для любой матрицы B. *(В случае* $\mathrm{char}\,\mathscr{K} \neq 0$ *коэффициенты* Pf *«целы» в том смысле, что лежат в простом подполе поля* $\mathscr{K}$, *т. е. являются суммами единиц.)*

Доказательство. Рассмотрим $r(2r-1)$ независимых переменных над полем $\mathscr{K}$: $\{a_{ij} \mid 1 \leqslant i < j \leqslant 2r\}$. Обозначим через K поле рациональных функций (отношений многочленов) от a_{ij} с коэффициентами из простого подполя поля $\mathscr{K}$. Положим $A = (a_{ij})$, где $a_{ij} = -a_{ji}$ при $i > j$, $a_{ii} = 0$, и введем на координатном пространстве K^{2r} невырожденное кососимметрическое скалярное произведение $\vec{x}^t A \vec{y}$. Перейдя к симплектическому базису с помощью некоторой матрицы B, получим, как выше, $\det A = (\det B)^2$. Априори $\det B$ является лишь рациональной функцией от a_{ij} с коэффициентами из $\mathbf{Q}$ или простого поля конечной характеристики. Но так как $\det A$ — многочлен с целыми коэффициентами, квадратный корень из него также должен иметь целые коэффициенты (здесь мы пользуемся теоремой об однозначном разложении на множители в кольце многочленов $\mathbf{Z}[a_{ij}]$ или $\mathbf{F}_p[a_{ij}]$). Знак $\sqrt{\det A}$, очевидно, однозначно фиксируется требованием, чтобы значение $\sqrt{\det I_{2r}}$ было равно единице.

Последнее равенство из формулировки теоремы устанавливается так. Прежде всего, $B^t AB$ кососимметрична вместе с A, так что

$$\mathrm{Pf}^2\,(B^t AB) = \det\,(B^t AB) = (\det B)^2 \det A = (\det B)^2\, \mathrm{Pf}^2 A.$$

Поэтому

$$\mathrm{Pf}\,(B^t AB) = \pm \det B\, \mathrm{Pf}\, A.$$

Чтобы установить знак, достаточно выяснить его в случае $A = I_{2r}$, $B = E_{2r}$, где он, очевидно, положителен.

10. Примеры.

$$\mathrm{Pf}\begin{pmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{pmatrix} = a_{12};$$

$$\mathrm{Pf}\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix} = -a_{12}a_{34} + a_{13}a_{24} - a_{14}a_{23}.$$

11. Следствие. *Определитель любой симплектической матрицы равен единице.*

Доказательство. Из условия $A^t I_{2r} A = I_{2r}$ и теоремы п. 9 следует

$$1 = \mathrm{Pf}\, I_{2r} = \mathrm{Pf}\,(A^t I_{2r} A) = \det A\, \mathrm{Pf}\, I_{2r},$$

что доказывает требуемое.

Мы пользовались этим фактом при доказательстве предложения п. 6.

12. Связь ортогональной, унитарной и симплектической групп. Пусть $\mathbf{R}^{2r}$ — координатное пространство с двумя скалярными произведениями: евклидовым $(\ ,\ )$ и симплектическим $[\ ,\ ]$:

$$(\vec{x}, \vec{y}) = \vec{x}^t \vec{y};$$
$$[\vec{x}, \vec{y}] = \vec{x}^t I_{2r} \vec{y} = (\vec{x}, I_{2r} \vec{y}).$$

Поскольку $I_{2r}^2 = -E_{2r}$, оператор I_{2r} определяет на $\mathbf{R}^{2r}$ комплексную структуру (см. § 12 ч. 1) с комплексным базисом $\{e_j + ie_{r+j} \mid j = 1, \ldots, r\}$, относительно которой имеется эрмитово скалярное произведение

$$\langle \vec{x}, \vec{y} \rangle = (\vec{x}, \vec{y}) - i\, [\vec{x}, \vec{y}]$$

(см. предложение п. 2 § 6).

В терминах этих структур имеем

$$\mathrm{U}(r) = \mathrm{O}(2r) \cap \mathrm{Sp}(2r) = \mathrm{GL}(r, \mathbf{C}) \cap \mathrm{Sp}(2r) = \mathrm{GL}(r, \mathbf{C}) \cap \mathrm{O}(2r).$$

Мы оставляем читателю проверку в качестве упражнения.

§ 14. Теорема Витта и группа Витта

1. В этом параграфе мы изложим результаты Витта, относящиеся к теории конечномерных ортогональных пространств над произвольными полями. Они уточняют теорему классификации из § 3 и могут рассматриваться как далеко идущее обобщение теоремы инерции и понятия о сигнатуре. Начнем с некоторых определений. Как обычно, считаем характеристику поля скаляров не равной двум.

Гиперболической плоскостью называется двумерное пространство L с невырожденным симметричным скалярным произведением $(\ ,\ )$, имеющее ненулевой изотропный вектор.

Гиперболическим пространством называется пространство, разлагающееся в прямую сумму попарно ортогональных гиперболических плоскостей.

Анизотропным пространством называется пространство, не имеющее (ненулевых) изотропных векторов.

Над вещественным полем анизотропные пространства L имеют сигнатуру $(n, 0)$ или $(0, n)$, где $n = \dim L$. Мы сейчас покажем, что гиперболические пространства суть обобщения пространств с сигнатурой (m, m).

2. Лемма. *У гиперболической плоскости L всегда существуют базисы $\{e'_1, e'_2\}$ с матрицей Грама $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ и $\{e_1, e_2\}$ с матрицей Грама $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.*

Доказательство. Пусть $l \in L$, $(l, l) = 0$. Если $l_1 \in L$ не пропорционален l, то $(l_1, l) \neq 0$, ибо L невырождена. Можно считать, что $(l_1, l) = 1$. Положим $e_1 = l$, $e_2 = l_1 - \frac{(l_1, l_1)}{2} l$. Тогда $(e_1, e_1) = (e_2, e_2) = 0$, $(e_1, e_2) = 1$. Положим $e'_1 = \frac{e_1 + e_2}{2}$, $e'_2 = \frac{e_1 - e_2}{2}$. Тогда $(e'_1, e'_1) = 1$, $(e'_2, e'_2) = -1$, $(e'_1, e'_2) = 0$. Лемма доказана

Базис $\{e_1, e_2\}$ мы будем называть *гиперболическим*. Аналогично, в общем гиперболическом пространстве мы будем называть гиперболическим базис с матрицей Грама, состоящей из диагональных блоков $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

3. Лемма. *Пусть $L_0 \subset L$ — изотропное подпространство в невырожденном ортогональном пространстве L, $\{e_1, \ldots, e_m\}$ — базис в L_0. Тогда существуют такие векторы $e'_1, \ldots, e'_m \in L$, что $\{e_1, e'_1, \ldots, e_m, e'_m\}$ образуют гиперболический базис своей линейной оболочки.*

Доказательство. Пусть L_1 — линейная оболочка $\{e_2, \ldots, e_m\}$. Так как L_1 строго меньше L_0, то $L_1^\perp$ строго больше $L_0^\perp$ в силу невырожденности L. Пусть $e''_1 \in L_1^\perp \setminus L_0^\perp$. Тогда $(e''_1, e_i) = 0$ при $i \geqslant 2$, но $(e''_1, e_1) \neq 0$. Можно считать, что $(e''_1, e_1) = 1$, так что e''_1 не пропорционален e_1. Как в доказательстве леммы п. 2, положим $e'_1 = e''_1 - \frac{(e''_1, e''_1)}{2} e_1$. Тогда $\{e_1, e'_1\}$ образуют гиперболический базис своей линейной оболочки. Ортогональное дополнение к ней невырождено и содержит изотропное подпространство, натянутое на $\{e_2, \ldots, e_m\}$. К этой паре можно применить аналогичное рассуждение, и индукция по m дает требуемое.

4. Теорема (Витт). *Пусть L — невырожденное конечномерное ортогональное пространство, $L', L'' \subset L$ — два его изометричных подпространства. Тогда любая изометрия $f'\colon L' \to L''$ может быть продолжена до изометрии $f\colon L \to L$, совпадающей с f' на L'.*

Доказательство. Разберем последовательно несколько случаев.

а) *$L' = L''$ и оба пространства невырождены. Тогда* $L = L' \oplus (L')^\perp$ и можно положить $f = f' \oplus \mathrm{id}_{(L')^\perp}$.

б) *$L' \neq L''$, $\dim L' = \dim L'' = 1$, и оба пространства невырождены.* Изометрию $f'\colon L' \to L''$ можно продолжить до изометрии $f''\colon L' + L'' \to L' + L''$, положив $f''(l) = f'(l)$ для $l \in L'$, $f''(l) = (f')^{-1}(l)$ для $l \in L''$. Если $L' + L''$ невырождено, то f'' продолжается до f по предыдущему случаю. Если $L' + L''$ вырождено, то ядро скалярного произведения на $L' + L''$ одномерно. Пусть e_1 порождает это ядро, e_2 порождает L'. В ортогональном дополнении к e_2 в L найдем такой вектор e_1', что базис $\{e_1, e_1'\}$ порождаемой этими векторами плоскости гиперболичен. Это возможно по лемме п. 3. Покажем, что подпространство L_0, натянутое на $\{e_1, e_1', e_2\}$, невырождено, и изометрия f': $L' \to L''$ продолжается до изометрии f: $L_0 \to L_0$. После этого можно будет применить случай а).

Невырожденность следует из того, что $(e_2, e_2) \neq 0$, а матрица Грама векторов $\{e_1, e_1', e_2\}$ имеет вид

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & (e_2, e_2) \end{pmatrix}.$$

Для продолжения f' заметим сначала, что ортогональное дополнение к $f'(e_2)$ в L_0 двумерно, невырождено и содержит изотропный вектор e_1. Поэтому оно является гиперболической плоскостью, так же как и ортогональное дополнение к e_2 в L_0. По лемме п. 2, существует изометрия второй плоскости с первой. Ее прямая сумма с f' является искомым продолжением.

в) $\dim L' = \dim L'' > 1$ *и* L', L'' *невырождены.* Проведем индукцию по $\dim L'$. Так как в L' имеется ортогональный базис, существует разложение $L' = L_1' \oplus L_2'$ в ортогональную прямую сумму подпространств ненулевой размерности. Так как f' — изометрия, то $L'' = L_1'' \oplus L_2''$, где $L_i'' = f'(L_i')$, и эта сумма ортогональна. По индуктивному предположению ограничение f' на L_1' продолжается до изометрии f_1': $L \to L$. Она переводит $(L_1')^\perp \supset L_2'$ в $(L_1'')^\perp \supset L_2''$. Снова по индуктивному предположению существует изометрия $(L_1')^\perp$ с $(L_1'')^\perp$, которая на L_2' совпадает с ограничением f'. Дополнив его ограничением f' на L_1', получим требуемое.

г) L' *вырождено.* Мы сведем этот последний случай к уже разобранному. Пусть $L_0' \subset L'$ — ядро ограничения метрики на L'. Выбрав ортонормированный базис в L', мы можем построить прямое разложение $L' = L_1' \oplus L_0'$, где L_1' невырождено. В ортогональном дополнении к L_1' внутри L мы можем найти подпространство $\overline{L}_0'$ такое, что сумма $\overline{L}_0' \oplus L_1' \oplus L_0$ прямая и пространство $\overline{L}_0' \oplus L_0'$ гиперболично, как в лемме п. 3; в частности, $\overline{L}_0' \oplus L_1' \oplus L_0'$ невырождено. Аналогично построим $\overline{L}_0'' \oplus L_1'' \oplus L_0''$, исходя из пространства L''. Очевидно, изометрия f': $L' \to L''$ продолжается до изометрий этих прямых сумм, ибо все гиперболические пространства одинаковой размерности изометричны. Возможность дальнейшего продолжения этой изометрии следует теперь из случая в). Теорема доказана.

5. Следствие. *Пусть* L_1, L_2 — *изометричные невырожденные пространства и* L_1', L_2' — *их изометричные подпространства. Тогда ортогональные дополнения* $(L_1')^\perp$, $(L_2')^\perp$ *к ним изометричны.*

6. Следствие. *Пусть L — невырожденное ортогональное пространство. Тогда любое изотропное подпространство L содержится в максимальном изотропном подпространстве, и для двух максимальных изотропных подпространств L', L'' существует изометрия $f\colon L \to L$, переводящая L' в L''.*

Доказательство. Первое утверждение тривиально. Для доказательства второго допустим, что $\dim L' \leqslant \dim L''$. Любая линейная инъекция $f'\colon L' \to L''$ является изометрией L' с $\operatorname{Im} f'$. Поэтому она продолжается до изометрии $f\colon L \to L$. Тогда $L' \subset f^{-1}(L'')$ и $f^{-1}(L'')$ изотропно. Так как L' максимально, $\dim L' = \dim f^{-1}(L'') = \dim L''$.

7. Следствие. *Для любого ортогонального пространства L существует ортогональное прямое разложение $L_0 \oplus L_h \oplus L_d$, где L_0 изотропно, L_h гиперболично и L_d анизотропно. Для любых двух таких разложений существует изометрия $f\colon L \to L$, переводящая одно из них в другое.*

Доказательство. Возьмем в качестве L_0 ядро скалярного произведения. Разложим L в прямую сумму $L_0 \oplus L_1$. В L_1 возьмем максимальное изотропное подпространство и вложим его в гиперболическое подпространство удвоенной размерности L_h, как в лемме п. 3. В качестве L_d возьмем ортогональное дополнение к L_h в L_1. Оно не содержит изотропных векторов, иначе такой вектор можно было бы добавить к исходному изотропному подпространству, которое не было бы максимальным. Это доказывает существование разложения требуемого вида.

Наоборот, в любом таком разложении $L_0 \oplus L_h \oplus L_d$ пространство L_0 является ядром. Далее, максимальное изотропное подпространство в L_h одновременно максимально изотропно в $L_h \oplus L_d$, поэтому размерность L_h определена однозначно. Значит, для двух разложений $L_0 \oplus L_h \oplus L_d$ и $L_0 \oplus L_h' \oplus L_d'$ существует изометрия, переводящая L_0 в L_0, L_h в L_h'. Она дополняется изометрией L_d в L_d' по теореме Витта, что завершает доказательство. Назовем L_d *анизотропной частью* пространства L; она определена с точностью до изометрии.

Это следствие есть обобщение принципа инерции на произвольные поля скаляров, сводящее классификацию ортогональных пространств к классификации анизотропных пространств или, на языке квадратичных форм, к классификации форм, *не представляющих нуля*, для которых из $q(l) = 0$ следует, что $l = 0$.

8. Группа Витта. Пусть $\mathcal{K}$ — поле скаляров. Обозначим через $W(\mathcal{K})$ множество классов анизотропных ортогональных пространств над $\mathcal{K}$ (с точностью до изометрии), дополненное классом нулевого пространства. Введем на $W(\mathcal{K})$ следующую операцию сложения: если L_1, L_2 — два анизотропных пространства, $[L_1]$, $[L_2]$ — их классы в $W(\mathcal{K})$, то $[L_1] + [L_2]$ — класс анизотропной части $L_1 \oplus L_2$ (справа стоит ортогональная внешняя прямая сумма).

Нетрудно убедиться, что определение корректно. Далее, эта операция сложения ассоциативна, и класс нулевого пространства служит нулем в $W(\mathcal{K})$. Более того, имеет место

9. Теорема. а) *$W(\mathscr{K})$ с введенной операцией сложения является абелевой группой, называемой группой Витта поля $\mathscr{K}$.*

б) *Пусть L_a означает одномерное координатное пространство над $\mathscr{K}$ со скалярным произведением axy, $a \in \mathscr{K} \setminus \{0\}$. Тогда $[L_a]$ зависит только от смежного класса $a(\mathscr{K}^*)^2$, и элементы $[L_a]$ составляют систему образующих группы $W(\mathscr{K})$.*

Доказательство. Нам осталось убедиться, что у каждого элемента $W(\mathscr{K})$ существует обратный. Действительно, пусть L — анизотропное пространство с метрикой, которая в ортогональном базисе $\{e_1, \ldots, e_n\}$ задана формой $\sum_{i=1}^{n} a_i x_i^2$. Обозначим через $\bar{L}$ пространство L с метрикой $-\sum_{i=1}^{n} a_i x_i^2$ и покажем, что $L \oplus \bar{L}$ гиперболично, так что $[L] + [\bar{L}] = [0]$ в $W(\mathscr{K})$. Действительно, метрика в $L \oplus \bar{L}$ задана формой $\sum_{i=1}^{n} a_i (x_i^2 - y_i^2)$. Но плоскость с метрикой $a(x^2 - y^2)$, очевидно, гиперболична, ибо форма невырождена, а вектор $(1, 1)$ изотропен. То, что $[L_a]$ зависит лишь от $a(\mathscr{K}^*)^2$, было проверено в п. 7 § 2. Кроме того, каждое n-мерное ортогональное пространство разлагается в прямую ортогональную сумму одномерных пространств вида L_a. Это завершает доказательство.

§ 15. Алгебры Клиффорда

1. *Алгеброй* над полем $\mathscr{K}$ мы будем называть ассоциативное кольцо с единицей A, содержащее поле $\mathscr{K}$ и такое, что $\mathscr{K}$ лежит в центре A, т. е. коммутирует со всеми элементами A. В частности, A является $\mathscr{K}$-линейным пространством.

Рассмотрим конечномерное ортогональное пространство L с метрикой g. В этом параграфе мы построим такую алгебру $C(L)$ и $\mathscr{K}$-линейное вложение ρ: $L \to C(L)$, что для любого элемента $l \in L$ будет выполнено соотношение

$$\rho(l)^2 = g(l, l) \cdot 1,$$

т. е. *скалярный квадрат* каждого вектора из L будет реализован как его *квадрат* в смысле умножения в $C(L)$. Кроме того, элементы $\rho(L)$ будут *мультипликативными образующими* $C(L)$, т. е. любой элемент из $C(L)$ окажется представимым в виде линейной комбинации (некоммутативных) одночленов от элементов $\rho(L)$. Алгебра $C(L)$ (вместе с отображением ρ) с такими свойствами будет называться *алгеброй Клиффорда* пространства L.

2. Теорема. а) *Для всякого конечномерного ортогонального пространства L алгебра Клиффорда $C(L)$ существует и имеет размерность 2^n над $\mathscr{K}$, где $n = \dim L$.*

б) *Пусть σ: $L \to D$ — любое $\mathscr{K}$-линейное отображение L в $\mathscr{K}$-алгебру D, для которого $\sigma(l)^2 = g(l, l) \cdot 1$ для всех $l \in L$. Тогда существует единственный гомоморфизм $\mathscr{K}$-алгебр τ: $C(L) \to D$*

такой, что $\sigma = \tau \circ \rho$. В частности, $C(L)$ определена однозначно с точностью до изоморфизма.

Доказательство. а) Выберем в L ортогональный базис $\{e_1, \ldots, e_n\}$, $(e_i, e_i) = a_i$. По определению, в $C(L)$ должны выполняться соотношения

$$\rho(e_i)^2 = a_i, \quad \rho(e_i)\rho(e_j) = -\rho(e_j)\rho(e_i), \qquad i \neq j.$$

Второе из них следует из того, что $[\rho(e_i + e_j)]^2 = \rho(e_i)^2 + \rho(e_i)\rho(e_j) + \rho(e_j)\rho(e_i) + \rho(e_j)^2 = \rho(e_i)^2 + \rho(e_j)^2$. Разложив элементы $l_1, \ldots, l_m \in L$ по базису $\{e_i\}$ и пользуясь тем, что умножение в L $\mathscr{K}$-линейно по каждому из сомножителей (это следует из того, что $\mathscr{K}$ лежит в центре), мы можем представить любое произведение $\rho(l_1) \ldots \rho(l_m)$ в виде линейной комбинации одночленов относительно $\rho(e_i)$. Заменив $\rho(e_i)^2$ на a_i и $\rho(e_i)\rho(e_j)$ при $i > j$ на $-\rho(e_j)\rho(e_i)$, мы можем привести любой одночлен к виду $a\rho(e_{i_1}) \ldots \ldots \rho(e_{i_m})$, где $a \in \mathscr{K}$, $i_1 < i_2 < \ldots < i_m$. Дальнейших соотношений между такими выражениями не видно; одночленов $\rho(e_{i_1}) \ldots \ldots \rho(e_{i_m})$ имеется 2^m (включая тривиальный одночлен 1 при $m = 0$).

План доказательства состоит в том, чтобы сделать строгими эти наводящие соображения, действуя более формально.

С этой целью для каждого подмножества $S \subset \{1, \ldots, n\}$ введем символ e_S (который впоследствии окажется равным $\rho(e_{i_1}) \ldots \ldots \rho(e_{i_m})$, если $S = \{i_1, \ldots, i_m\}$, $i_1 < \ldots < i_m$); положим также $e_\varnothing = 1$ ($\varnothing$ — пустое подмножество). Обозначим через $C(L)$ $\mathscr{K}$-линейное пространство с базисом $\{e_S\}$. Определим умножение в $C(L)$ следующим образом. Если $1 \leqslant s, t \leqslant n$, положим

$$(s, t) = \begin{cases} 1 & \text{при } s \leqslant t, \\ -1 & \text{при } s > t. \end{cases}$$

Для двух подмножеств $S, T \subset \{1, \ldots, n\}$ положим

$$a(S, T) = \prod_{\substack{s \in S \\ t \in T}} (s, t) \prod_{t \in S \cap T} a_t,$$

где, напомним, $a_i = g(e_i, e_i)$. Пустые произведения считаются равными единице. Наконец, произведение линейных комбинаций $\sum a_S e_S$, $\sum b_T e_T \in C(L)$; $a_S, b_T \in \mathscr{K}$, определим формулой

$$\left(\sum a_S e_S\right)\left(\sum b_T e_T\right) = \sum a_S b_T\, a(S, T)\, e_{S \nabla T},$$

где $S \nabla T = (S \cup T) \setminus (S \cap T)$ — симметрическая разность множеств S, T. Все аксиомы $\mathscr{K}$-алгебры проверяются тривиально, за исключением ассоциативности. Ассоциативность достаточно проверить на элементах базиса, т. е. установить тождество

$$(e_S e_T) e_R = e_S (e_T e_R).$$

Поскольку $e_S e_T = a(S,\ T) e_{S \nabla T}$, имеем

$$(e_S e_T) e_R = a(S,\ T) a(S \nabla T,\ R) e_{(S \nabla T) \nabla R},$$
$$e_S (e_T e_R) = a(S,\ T \nabla R) a(T,\ R) e_{S \nabla (T \nabla R)}.$$

Нетрудно проверить, что

$$(S \nabla T) \nabla R = S \nabla (T \nabla R) =$$
$$= \{(S \cup T \cup R) \setminus [(S \cap T) \cup (S \cap R) \cup (T \cap R)]\} \cup (S \cap T \cap R).$$

Поэтому остается убедиться лишь в совпадении скалярных коэффициентов. Часть $a(S, T) a(S \nabla T, R)$, относящаяся к знакам, имеет вид

$$\prod_{\substack{s \in S \\ t \in T}} (s,\ t) \prod_{\substack{u \in S \nabla T \\ r \in R}} (u,\ r).$$

Заставив во втором произведении u пробегать сначала все элементы S, а затем все элементы T (при фиксированном r), мы введем сомножители $(u, r)^2$, $u \in S \cap T$, равные единице, так что этот «знак» можно записать в симметрическом по S, T, R виде

$$\prod_{\substack{s \in S \\ t \in T}} (s,\ t) \prod_{\substack{u \in S \\ r \in R}} (u,\ r) \prod_{\substack{u \in T \\ r \in R}} (u,\ r).$$

Аналогично с тем же результатом преобразуется знак, относящийся к $a(S, T \nabla R) a(T, R)$. Остается разобрать множители, в которые входят скалярные квадраты a_i. Для $a(S, T) a(S \nabla T, R)$ они имеют вид

$$\prod_{i \in S \cap T} a_i \prod_{j \in (S \nabla T) \cap R} a_j.$$

Но $(S \nabla T) \cap R = (S \cap R) \nabla (T \cap R)$, а $S \cap T$ с этим множеством не пересекается, и $(S \cap T) \cup [(S \cap R) \nabla (T \cap R)]$ состоит из тех элементов $S \cup T \cup R$, которые содержатся более чем в одном из этих трех множеств. Поэтому наш множитель симметрично зависит от S, T, R. Аналогично с тем же результатом вычисляется нужная нам часть коэффициента $a(S, T \nabla R) a(T, R)$. Это завершает доказательство ассоциативности алгебры $C(L)$.

Определим, наконец, $\mathscr{K}$-линейное отображение ρ: $L \to C(L)$ условием $\rho(e_i) = e_{\{i\}}$. Согласно формулам умножения $e_\varnothing$ является единицей в $C(L)$, и

$$\rho(e_i) \rho(e_j) = e_{\{i\}} e_{\{j\}} = \begin{cases} a_i e_\varnothing & \text{при } i = j, \\ -e_{\{j\}} e_{\{i\}} & \text{при } i \neq j. \end{cases}$$

Поэтому $(\rho, C(L))$ есть алгебра Клиффорда для L.

б) Последнее утверждение доказывается формально. Пусть σ: $L \to D$ — $\mathscr{K}$-линейное отображение с $\sigma(l)^2 = g(l,\ l) \cdot 1$. Существует единственное $\mathscr{K}$-линейное отображение τ: $C(L) \to D$, которое на элементах базиса e_S определено формулой

$$\tau(e_{\{i_1 \dots i_m\}}) = \sigma(e_{i_1}) \dots \sigma(e_{i_m}),$$
$$\tau(e_\varnothing) = 1_D.$$

Для него $\tau \circ \rho = \sigma$, ибо это так на элементах базиса L. Наконец, τ является гомоморфизмом алгебр. Действительно,

$$\tau(e_S e_T) = \tau(a(S, T) e_{S \nabla T}) = a(S, T) \tau(e_{S \nabla T}),$$

и нетрудно проверить, что $\tau(e_S)\tau(e_T)$ можно привести к тому же виду, пользуясь соотношениями

$$\sigma(e_i)^2 = a_i, \ \sigma(e_i)\sigma(e_j) = -\sigma(e_j)\sigma(e_i) \text{ при } i \neq j.$$

Это завершает доказательство.

3. Примеры. а) Пусть L — двумерная вещественная плоскость с метрикой $-(x^2+y^2)$. Алгебра Клиффорда $C(L)$ имеет базис $(1, e_1, e_2, e_1e_2)$ с мультипликативными соотношениями

$$e_1^2 = e_2^2 = -1, \quad e_1 e_2 = -e_2 e_1.$$

Нетрудно убедиться, что отображение $C(L) \to H$: $1 \mapsto 1$, $e_1 \mapsto \mathbf{i}$, $e_2 \mapsto \mathbf{j}$, $e_1e_2 \mapsto \mathbf{k}$ определяет изоморфизм $C(L)$ с алгеброй кватернионов H.

б) Пусть L — линейное пространство с *нулевой* метрикой. Алгебра $C(L)$ порождена образующими $\{e_1, \ldots, e_n\}$ с соотношениями

$$e_i^2 = 0, \ e_i e_j = -e_j e_i \text{ при } i \neq j.$$

Она называется *внешней алгеброй*, или *алгеброй Грассмана*, линейного пространства L. Мы еще вернемся к ней в части 4.

в) Пусть $L = \mathcal{M}^{\mathbf{C}}$ — комплексифицированное пространство Минковского с метрикой $x_0^2 - \sum_{i=1}^{3} x_i^2$ относительно ортонормированного базиса $\{e_i\}$ в $\mathcal{M}$, являющегося одновременно базисом $\mathcal{M}^{\mathbf{C}}$. Покажем, что алгебра Клиффорда $C(\mathcal{M}^{\mathbf{C}})$ изоморфна алгебре комплексных 4×4-матриц. С этой целью рассмотрим матрицы Дирака, записываемые в блочном виде:

$$\gamma_0 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \quad \gamma_j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \qquad j = 1, 2, 3.$$

Пользуясь свойствами матриц Паули σ_j, нетрудно убедиться, что γ_j удовлетворяют тем же соотношениям в алгебре матриц $M_4(\mathbf{C})$ что и $\rho(e_j)$ в алгебре $C(\mathcal{M}^{\mathbf{C}})$:

$$\gamma_0^2 = -\gamma_1^2 = -\gamma_2^2 = -\gamma_3^2 = 1$$

(т. е. E_4); $\gamma_i\gamma_j + \gamma_j\gamma_i = 0$ при $i \neq j$. Значит, $\mathbf{C}$-линейное отображение σ: $\mathcal{M}^{\mathbf{C}} \to M_4(\mathbf{C})$ индуцирует гомоморфизм алгебр τ: $C(\mathcal{M}^{\mathbf{C}}) \to M_4(\mathbf{C})$, для которого $\tau \cdot \rho(e_i) = \gamma_i$. Непосредственным вычислением можно проверить, что отображение τ сюръективно, а так как обе $\mathbf{C}$-алгебры $C(\mathcal{M}^{\mathbf{C}})$ и $M_4(\mathbf{C})$ шестнадцатимерны, τ является изоморфизмом.

Часть 3. АФФИННАЯ И ПРОЕКТИВНАЯ ГЕОМЕТРИЯ

§ 1. Аффинные пространства, аффинные отображения и аффинные координаты

1. Определение. *Аффинным пространством над полем $\mathcal{K}$ называется тройка $(A, L, +)$, состоящая из линейного пространства L над полем $\mathcal{K}$, множества A, элементы которого называются точками, и внешней бинарной операции $A \times L \to A$: $(a, l) \mapsto a + l$, удовлетворяющей следующим аксиомам:*

а) $(a + l) + m = a + (l + m)$ *для всех* $a \in A$; $l, m \in L$;

б) $a + 0 = a$ *для всех* $a \in A$;

в) *для любых двух точек* $a, b \in A$ *существует единственный вектор* $l \in L$ *со свойством* $b = a + l$.

2. Пример. Тройка $(L, L, +)$, где L — линейное пространство, а $+$ совпадает со сложением в L, является аффинным пространством. Удобно говорить, что она задает аффинную структуру линейного пространства L. Этот пример типичен; позже мы увидим, что всякое аффинное пространство изоморфно такому.

3. Термины. Мы часто будем называть аффинным пространством пару (A, L) или даже просто A, опуская указания на $+$. Линейное пространство L называется ассоциированным с аффинным пространством A. Отображение $A \to A$: $a \mapsto a + l$ называется *сдвигом* на вектор l; удобно иметь для него специальное обозначение t_l. Мы пишем $a - l$ вместо $t_{-l}(a)$ или $a + (-l)$.

4. Предложение. *Отображение $l \mapsto t_l$ определяет инъективный гомоморфизм аддитивной группы пространства L в группу перестановок точек аффинного пространства A, т. е. эффективное действие L на A. Это действие транзитивно, т. е. для любой пары точек, $a, b \in A$ существует $l \in L$ с $t_l(a) = b$.*

Наоборот, задание транзитивного эффективного действия аддитивной группы L на множестве A определяет на A структуру аффинного пространства с ассоциированным пространством L.

Доказательство. Из аксиом а) и б) следует, что при любых $l \in L$ и $a \in A$ уравнение $t_l(x) = a$ имеет решение $x = a + (-l)$, так что все t_l сюръективны. Если $t_l(a) = t_l(b)$, то, найдя по аксиоме в) такой вектор $m \in L$, что $b = a + m$, получаем $a + l = (a + m) + l = (a + l) + m$. Но $a + l = (a + l) + 0$, поэтому из условия единственности аксиомы в) следует, что $m = 0$, так что $a = b$. Поэтому все t_l инъективны.

Аксиома а) означает, что $t_m \circ t_l = t_{l+m}$, а аксиома б) — что $t_0 = \mathrm{id}_A$. Поэтому отображение $l \mapsto t_l$ является гомоморфизмом аддитивной группы L в группу биекций A с самим собой. Его ядро равно нулю в силу аксиом б) и в).

Наоборот, пусть $L \times A \to A$: $(l, a) \mapsto a + l$ — эффективное транзитивное действие L на A. Тогда аксиомы а) и б) получаются прямо из определения действия, а аксиома в) — из соединения свойств эффективности и транзитивности.

5. Замечание. По поводу действия групп (не обязательно абелевых) на множествах см. § 2 главы 7 «Введения в алгебру». Множество, на котором группа действует транзитивно и эффективно, называется *главным однородным пространством* над этой группой.

Отметим, что в аксиомах аффинного пространства не фигурирует явно структура умножения на скаляры в L. Она появляется лишь в определении аффинных отображений и затем барицентрических комбинаций точек A. Но прежде несколько слов о формализме.

6. Правила вычислений. Тот единственный вектор $l \in L$, для которого $b = a + l$, удобно обозначать $b - a$. Эта операция «внешнего вычитания» $A \times A \to L$: $(b, a) \mapsto b - a$ обладает следующими свойствами:

а) $(c - b) + (b - a) = c - a$ для всех a, b, $c \in A$; сложение слева — это сложение в L.

Действительно, пусть $c = b + l$, $b = a + m$; тогда $c = a + (l + m)$, так что $c - a = l + m = (c - b) + (b - a)$.

б) $a - a = 0$ для всех $a \in A$.

в) $(a + l) - (b + m) = (a - b) + (l - m)$ для всех $a, b \in A$, $l, m \in L$.

В самом деле, достаточно проверить, что $(b + m) + (a - b) + (l - m) = a + l$, или $b + (a - b) = a$, а это — определение $a - b$.

Вообще, употребление знаков $\pm$ для различных операций $L \times L \to L$, $A \times L \to L$, $A \times A \to L$ подчиняется следующим формальным правилам. Выражение $\pm a_1 \pm a_2 \pm \ldots \pm a_m + l_1 + \ldots + l_n$ для $a_j \in A$, $l_k \in L$ имеет смысл, если либо m четно и все $\pm a_j$ можно объединить в пары вида $a_i - a_j$, либо m нечетно, и все точки можно объединить в такие пары, кроме одной, входящей со знаком $+$. В первом случае вся сумма лежит в L, во втором — в A. Кроме того, она зависит от своих слагаемых коммутативно и ассоциативно: например, $a_1 - a_2 + l$ можно вычислять как $(a_1 - a_2) + l$ или $(a_1 + l) - a_2$ или $a_1 - (a_2 - l)$; мы позволим себе писать $a + l$, так же как $l + a$.

Доказывать это правило в общем виде мы не станем. Всякий раз, когда мы будем им пользоваться, читатель без труда разобьет нужную выкладку на серию элементарных шагов, каждый из которых сведется к применению одной аксиомы или формулы а) — в) начала этого пункта.

Заметим, что *сумма* $a + b$, где $a, b \in A$, вообще говоря, *смысла не имеет*, так же как и *выражение* xa, где $x \in \mathcal{K}$ (исключение:

$A = L$). Тем не менее ниже мы придадим однозначный смысл, например, выражению $\frac{1}{2}a + \frac{1}{2}b$ (но не его слагаемым!).

Интуитивно аффинное пространство $(A, L, +)$ следует представлять себе как линейное пространство L с «забытым» началом координат 0. Оставлена лишь операция сдвига на векторы L, суммирования сдвигов и умножения вектора сдвига на скаляр.

7. Аффинные отображения. Пусть (A_1, L_1), (A_2, L_2) — два аффинных пространства над одним и тем же полем $\mathcal{K}$. *Аффинно линейным*, или просто *аффинным, отображением* первого во второе называется пара (f, Df), где f: $A_1 \to A_2$, Df: $L_1 \to L_2$, удовлетворяющая следующим условиям:

а) Df — линейное отображение.

б) Для любых $a_1, a_2 \in A$ имеем

$$f(a_1) - f(a_2) = Df(a_1 - a_2).$$

(Оба выражения лежат в L_2.)

Df (или $D(f)$) называется *линейной частью* аффинного отображения f. Поскольку $a_1 - a_2$ пробегает все векторы L_1, когда $a_1, a_2 \in A_1$, линейная часть Df определяется по f однозначно. Это позволяет обозначать аффинные отображения просто f: $A_1 \to A_2$.

8. Примеры. а) Любое линейное отображение f: $L_1 \to L_2$ индуцирует аффинно линейное отображение пространств $(L_1, L_1, +) \to (L_2, L_2, +)$. Для него $Df = f$.

б) Любой сдвиг t_l: $A \to A$ аффинно линеен и $D(t_l) = \mathrm{id}_L$. Действительно,

$$t_l(a_1) - t_l(a_2) = (a_1 + l) - (a_2 + l) = a_1 - a_2.$$

в) Если f: $A_1 \to A_2$ — аффинно линейное отображение и $l \in L_2$, то отображение $t_l \circ f$: $A_1 \to A_2$ аффинно линейно, и $D(t_l \circ f) = D(f)$. В самом деле,

$$t_l \circ f(a_1) - t_l \circ f(a_2) = (f(a_1) + l) - (f(a_2) + l) = f(a_1) - f(a_2) = Df(a_1 - a_2).$$

г) *Аффинно линейная функция* f: $A \to \mathcal{K}$ определяется как аффинно линейное отображение A в $(\mathcal{K}^1, \mathcal{K}^1, +)$, где $\mathcal{K}^1$ — одномерное координатное пространство. Таким образом, f принимает значения в $\mathcal{K}$, а Df есть линейный функционал на L. Любая постоянная функция f аффинно линейна: $Df = 0$.

9. Теорема. а) *Аффинные пространства вместе с аффинными отображениями образуют категорию.*

б) *Отображение, ставящее в соответствие аффинному пространству (A, L) линейное пространство L, а аффинному отображению f: $(A_1, L_1) \to (A_2, L_2)$ линейное отображение Df: $L_1 \to L_2$, является функтором из категории аффинных пространств в категорию линейных пространств.*

Доказательство. Справедливость общекатегорных аксиом (см. § 13 ч. 1) вытекает из следующих фактов.

Тождественное отображение id: $A \to A$ аффинно. Действительно, $a_1 - a_2 = \mathrm{id}_L(a_1 - a_2)$. В частности, $D(\mathrm{id}_A) = \mathrm{id}_L$.

Композиция аффинных отображений $A \xrightarrow{f_1} A \xrightarrow{f_2} A$ является аффинным отображением.

В самом деле, пусть $a,\ b \in A$. Тогда $f_1(a) - f_1(b) = Df_1(a-b)$ и далее

$$f_2f_1(a) - f_2f_1(b) = Df_2[f_1(a) - f_1(b)] = Df_2 \circ Df_1(a-b).$$

Мы доказали требуемое и заодно получили, что $D(f_2f_1) = Df_2 \circ Df_1$. Вместе с формулой $D(\mathrm{id}_A) = \mathrm{id}_L$ это доказывает утверждение б) теоремы.

Следующий важный результат характеризует изоморфизмы в нашей категории.

10. Предложение. *Следующие три свойства аффинного отображения $f\colon A_1 \to A_2$ равносильны:*

а) *f — изоморфизм;*

б) *Df — изоморфизм;*

в) *f — биекция в теоретико-множественном смысле.*

Доказательство. Согласно общекатегорному определению $f\colon A_1 \to A_2$ есть изоморфизм тогда и только тогда, когда имеется такое аффинное отображение $g\colon A_2 \to A_1$, что $gf = \mathrm{id}_{A_1}$, $fg = \mathrm{id}_{A_2}$. Если оно существует, то $D(fg) = \mathrm{id}_{L_2} = D(f)D(g)$ и $D(gf) = \mathrm{id}_{L_1} = D(g)D(f)$, откуда следует, что $D(f)$ — изоморфизм.

Покажем теперь, что Df — изоморфизм тогда и только тогда, когда f — биекция. Фиксируем точку $a_1 \in A_1$ и положим $a_2 = f(a_1)$. Любой элемент A_i однозначно представляется в виде $a_i + l_i$, $l_i \in L_i$, $i = 1, 2$. Из основного тождества

$$f(a_1 + l_1) - f(a_1) = Df[(a_1 + l_1) - a_1] = Df(l_1)$$

следует, что $f(a_1 + l_1) = a_2 + Df(l_1)$. Следовательно, f — биекция тогда и только тогда, когда $Df(l_1)$ при $l_1 \in L_1$ пробегает все элементы L_2 по одному разу, т. е. Df является биекцией. Но линейное отображение есть биекция тогда и только тогда, когда оно обратимо, т. е. является изоморфизмом.

Наконец, покажем, что биективное аффинное отображение является аффинным изоморфизмом. Для этого следует проверить, что обратное к f теоретико-множественное отображение аффинно. Но в обозначениях предыдущего абзаца это отображение определяется формулой

$$f^{-1}(a_2 + l_2) = a_1 + (Df)^{-1}(l_2), \quad l_2 \in L_2.$$

Поэтому

$$f^{-1}(a_2 + l_2) - f^{-1}(a_2 + l_2') = (Df)^{-1}(l_2) - (Df)^{-1}(l_2') = (Df)^{-1}(l_2 - l_2')$$

в силу линейности $(Df)^{-1}$. Итак, f^{-1} аффинно и $D(f^{-1}) = D(f)^{-1}$.

Окончательно, мы установили импликации а) $\Rightarrow$ б) $\Leftrightarrow$ в) $\Rightarrow$ а), откуда и следует предложение.

Конструкция конкретных аффинных отображений часто основывается на следующем результате.

11. Предложение. *Пусть* (A_1, L_1), (A_2, L_2) — *два аффинных пространства. Для любой пары точек* $a_1 \in A_1$, $a_2 \in A_2$ *и любого линейного отображения* $g: L_1 \to L_2$ *существует единственное аффинное отображение* $f: A_1 \to A_2$ *с* $f(a_1) = a_2$ *и* $Df = g$.

В самом деле, положим

$$f(a_1 + l_1) = a_2 + g(l_1)$$

для $l_1 \in L_1$. Поскольку любая точка A_1 однозначно представляется в виде $a_1 + l_1$, эта формула определяет теоретико-множественное отображение $f: A_1 \to A_2$. Оно аффинное, $f(a_1) = a_2$ и $Df = g$, потому что

$$f(a_1 + l_1) - f(a_1 + l_1') = g(l_1) - g(l_1') = g(l_1 - l_1') = \\ = g[(a_1 + l_1) - (a_1 + l_1')].$$

Это доказывает существование f. Наоборот, если f — отображение с требуемыми свойствами, то

$$f(a_1 + l) - f(a_1) = g(l),$$

откуда $f(a_1 + l) = a_2 + g(l)$ для всех $l \in L$.

12. Важный частный случай предложения п. 11 получается, если применить его к (A, L), (L, L), $a \in A$, $0 \in L$ и $g = \mathrm{id}_L$. Мы получим, что *для любой точки* $a \in A$ *существует единственный аффинный изоморфизм* $f: A \to L$, *переводящий эту точку в начало координат, с тождественной линейной частью.* Это и есть точный смысл представления о том, что аффинное пространство есть «линейное пространство с забытым началом координат».

В частности, аффинные пространства изоморфны тогда и только тогда, когда изоморфны ассоциированные линейные пространства. Последние классифицируются своей размерностью, и мы можем назвать размерностью аффинного пространства размерность соответствующего линейного пространства.

13. Следствие. *Пусть* $f_1, f_2: A_1 \to A_2$ — *два аффинных отображения. Их линейные части совпадают тогда и только тогда, когда* f_2 *есть композиция* f_1 *со сдвигом на некоторый вектор из* L_2, *который определяется однозначно.*

Доказательство. Достаточность условия была проверена в примере в) п. 8. Для доказательства необходимости выберем любую точку $a \in A_1$ и положим $f_2' = t_{f_2(a) - f_1(a)} \circ f_1$. Очевидно, $f_2'(a) = f_2(a)$ и $D(f_2') = D(f_2)$. По предложению п. 11 $f_2' = f_2$. Наоборот, если $f_2 = t_l \circ f_1$, то $l = f_2(a) - f_1(a)$; этот вектор не зависит от $a \in A$ из-за совпадения линейных частей f_1, f_2.

14. Аффинные координаты. а) *Система аффинных координат* в аффинном пространстве (A, L) есть пара, состоящая из точки $a_0 \in A$ (начала координат) и базиса $\{e_1, \ldots, e_n\}$ ассоциированного линейного пространства L. *Координаты* точки $a \in A$ в этой системе образуют вектор $(x_1, \ldots, x_n) \in \mathcal{K}^n$, однозначно определяемый условием $a = a_0 + \sum_{i=1}^{n} x_i e_i$.

Иначе говоря, отождествим A с L посредством отображения с тождественной линейной частью, переводящего a_0 в 0, и возьмем координаты образа точки a в базисе $\{e_1, \ldots, e_n\}$: это и будут $x_1, \ldots, x_n$.

Пусть в пространствах A_1, A_2 выбраны системы координат, отождествляющие их с $\mathscr{K}^m$, $\mathscr{K}^n$ соответственно. Тогда любое аффинно линейное отображение $f\colon A_1 \to A_2$ можно записать в виде

$$f(\vec{x}) = B\vec{x} + \vec{y},$$

где B — матрица отображения Df в соответствующих базисах L_1, L_2, а $\vec{y}$ — координаты вектора $f(a_0') - a_0''$ в базисе L_2; a_0' — начало координат в A_1, a_0'' — начало координат в A_2. Действительно, отображение $\vec{x} \longmapsto B\vec{x} + \vec{y}$ аффинно линейно, переводит a_0' в $f(a_0')$ и имеет ту же линейную часть, что и f.

б) Другой вариант данного определения системы координат состоит в том, чтобы заменить векторы $\{e_1, \ldots, e_n\}$ точками $\{a_0 + e_1, \ldots, a_0 + e_n\}$ в A. Положим $a_i = a_0 + e_i$, $i = 1, \ldots, n$. Координаты точки $a \in A$ находятся тогда из представления $a = a_0 + \sum\limits_{i=1}^{n} x_i(a_i - a_0)$. Возникает соблазн «привести подобные члены» и написать выражение справа в виде $\left(1 - \sum\limits_{i=1}^{n} x_i\right)a_0 + \sum\limits_{i=1}^{n} x_i a_i$. Отдельные члены в этой сумме не имеют смысла! Тем не менее оказывается, что суммы такого вида можно рассматривать, и они весьма полезны.

15. Предложение. *Пусть $a_0, \ldots, a_s$ — любые точки аффинного пространства A. Для любых $y_0, \ldots, y_s \in \mathscr{K}$ с условием $\sum\limits_{i=0}^{s} y_i = 1$ определим формальную сумму $\sum\limits_{i=0}^{s} y_i a_i$ выражением вида*

$$\sum_{i=0}^{s} y_i a_i = a + \sum_{i=0}^{s} y_i(a_i - a),$$

где a — любая точка A. Утверждается, что выражение справа не зависит от a. Поэтому точка $\sum\limits_{i=0}^{s} y_i a_i$ определена корректно. Она называется барицентрической комбинацией точек $a_0, \ldots, a_s$ с коэффициентами $y_0, \ldots, y_s$.

Доказательство. Заменим точку a на точку $a + l$, $l \in L$. Получим

$$a + l + \sum_{i=0}^{s} y_i(a_i - a - l) = a + \sum_{i=0}^{s} y_i(a_i - a),$$

ибо $\left(1-\sum_{i=0}^{s} y_i\right)l=0$. Мы пользовались здесь правилами, сформулированными в п. 6. Читателю будет полезно провести эту выкладку подробно.

16. Следствие. *Система* $\{a_0;\ a_1-a_0, \ldots, a_n-a_0\}$, *состоящая из точки* $a_0 \in A$ *и векторов* a_i-a_0 *в* L, *образует систему аффинных координат в* A *тогда и только тогда, когда любая точка* A *однозначно представима в виде барицентрической комбинации* $\sum_{i=0}^{n} x_i a_i$, $x_i \in \mathscr{K}$, $\sum_{i=0}^{n} x_i = 1$.

Когда это условие выполнено, система точек $\{a_0, \ldots, a_n\}$ называется *барицентрической системой координат* в A, а числа $x_0, \ldots$ $\ldots, x_n$ — *барицентрическими координатами* точки $\sum_{i=0}^{n} x_i a_i$.

Доказательство. Все непосредственно следует из определений, если вычислять $\sum_{i=0}^{n} x_i a_i$ по формуле $a_0+\sum_{i=1}^{n} x_i(a_i-a_0)$. Действительно, так как любая точка A однозначно представляется в виде a_0+l, $l \in L$, система $\{a_0, a_1-a_0, \ldots, a_n-a_0\}$ является аффинной системой координат в A тогда и только тогда, когда всякий вектор $l \in L$ однозначно представляется в виде линейной комбинации $\sum_{i=1}^{n} x_i(a_i-a_0)$, т. е. если $\{a_1-a_0, \ldots, a_n-a_0\}$ образуют базис L. По координатам $x_1, \ldots, x_n$ вектора l барицентрические координаты точки a_0+l восстанавливаются однозначно в виде $1-\sum_{i=1}^{n} x_i = x_0, x_1, \ldots, x_n$.

17. С барицентрическими комбинациями можно во многом обращаться так же, как с обычными линейными комбинациями в линейном пространстве. Например, слагаемые с нулевыми коэффициентами можно выбрасывать. Наиболее полезное замечание состоит в том, что барицентрическая комбинация нескольких барицентрических комбинаций точек $a_0, \ldots, a_s$ в свою очередь является барицентрической комбинацией этих точек, коэффициенты которой можно вычислять по ожидаемому формальному правилу:

$$x_1 \sum_{i=0}^{s} y_{i1}a_i + x_2 \sum_{i=0}^{s} y_{i2}a_i + \ldots + x_m \sum_{i=0}^{s} y_{im}a_i = \sum_{i=0}^{s}\left(\sum_{k=1}^{m} x_k y_{ik}\right)a_i.$$

Действительно,

$$\sum_{i=0}^{s}\sum_{k=1}^{m} x_k y_{ik} = \sum_{k=1}^{m} x_k \sum_{i=0}^{s} y_{ik} = \sum_{k=1}^{m} x_k = 1,$$

так что последняя комбинация барицентрична. Вычисляя левую и правую части этого равенства по правилу, сформулированному в предложении п. 15, с помощью одной и той же точки $a \in A$ и применяя формализм п. 6, легко получим, что они совпадают.

Наконец, барицентрические комбинации ведут себя как линейные комбинации относительно аффинных отображений.

18. Предложение. а) *Пусть* $f\colon A_1\to A_2$ — *аффинное отображение и* $a_0, \ldots, a_s \in A_1$. *Тогда*

$$f\left(\sum_{i=0}^{s} x_i a_i\right)=\sum_{i=0}^{s} x_i f(a_i),$$

если $\sum_{i=0}^{s} x_i=1$.

б) *Пусть* $a_0, \ldots, a_n$ *задают барицентрическую систему координат в* A_1. *Тогда для любых точек* $b_0, \ldots, b_n \in A_2$ *существует единственное аффинное отображение* f, *переводящее* a_i *в* b_i, $i=1, \ldots, n$.

Доказательство. Выбрав $a \in A_1$, получим

$$f\left(\sum_{i=0}^{s} x_i a_i\right)=f\left(a+\sum_{i=0}^{s} x_i (a_i-a)\right)=f(a)+Df\left(\sum_{i=0}^{s} x_i (a_i-a)\right)=$$

$$=f(a)+\sum_{i=0}^{s} x_i Df(a_i-a)=f(a)+\sum_{i=0}^{s} x_i (f(a_i)-f(a))=\sum_{i=0}^{s} x_i f(a_i)$$

по предложению п. 15, что доказывает утверждение а).

Если $a_0, \ldots, a_n$ образуют барицентрическую систему координат в A_1, то по следствию п. 16 всякая точка A представляется единственной барицентрической комбинацией $\sum_{i=0}^{n} x_i a_i$. Определим тогда теоретико-множественное отображение $f\colon A_1\to A_2$ формулой $f\left(\sum_{i=0}^{n} x_i a_i\right)=\sum_{i=0}^{n} x_i b_i$. В силу а) это единственное возможное определение, и нужно лишь проверить, что f — аффинное отображение. Действительно, вычисляя, как в предложении п. 15, получаем

$$f\left(\sum_{i=0}^{n} x_i a_i\right)-f\left(\sum_{i=0}^{n} y_i a_i\right)=\sum_{i=0}^{n} x_i b_i-\sum_{i=0}^{n} y_i b_i=b_0+\sum_{i=0}^{n} x_i (b_i-b_0)-$$

$$-\left[b_0+\sum_{i=0}^{n} y_i (b_i-b_0)\right]=\sum_{i=0}^{n} (x_i-y_i)(b_i-b_0)=Df\left(\sum_{i=0}^{n} x_i a_i-\sum_{i=0}^{n} y_i a_i\right),$$

где $Df\colon L_1\to L_2$ — линейное отображение, переводящее a_i-a_0 в b_i-b_0 для всех $i=1, \ldots, n$. Оно существует, ибо $a_1-a_0, \ldots, a_n-a_0$ по предположению образуют базис L_1.

19. Замечания. В аффинном пространстве $\mathbf{R}^n$ барицентрическая комбинация $\sum_{i=1}^{m} \frac{1}{m} a_i$ представляет положение «центра масс» системы единичных масс, помещенных в точках a_i. Этим объясняется терминология. Если $a_i=(0, \ldots, 1, \ldots, 0)$ (единица на i-м месте), то множество точек с барицентрическими координатами $x_1, \ldots, x_n$, $0\leqslant x_i\leqslant 1$, составляет пересечение линейного многообразия $\sum_{i=1}^{n} x_i = 1$ с положительным октантом (точнее, «2^n-тантом»).

В топологии это множество называется *стандартным* $(n-1)$*-мерным симплексом.* Одномерный симплекс — это отрезок прямой, двумерный — треугольник, трехмерный — тетраэдр. Вообще, множество $\left\{\sum_{i=1}^{n} x_i a_i \middle| \sum_{i=1}^{n} x_i = 1, \quad 0 \leqslant x_i \leqslant 1\right\}$ есть *замкнутый симплекс* с вершинами $a_1, \ldots, a_n$ в вещественном аффинном пространстве. Он называется вырожденным, если векторы $a_2 - a_1, \ldots, a_n - a_1$ линейно зависимы.

§ 2. Аффинные группы

1. Пусть A — аффинное пространство над полем $\mathcal{K}$. Множество аффинных биективных отображений $f\colon A \to A$ в силу предложения п. 10 § 1 образует группу, которую мы будем называть аффинной группой и обозначать $\operatorname{Aff} A$.

Ее отображение $D\colon \operatorname{Aff} A \to \mathrm{GL}(L)$, где $\mathrm{GL}(L)$ — группа линейных автоморфизмов ассоциированного векторного пространства, является гомоморфизмом. Он сюръективен по предложению п. 11 § 1 и имеет своим ядром группу сдвигов $\{t_l \mid l \in L\}$ по следствию п. 13 § 1. Эта группа сдвигов изоморфна аддитивной группе пространства L по предложению п. 4 § 1. Таким образом, $\operatorname{Aff} A$ есть расширение группы $\mathrm{GL}(L)$ с помощью аддитивной группы L, которая является нормальным делителем в $\operatorname{Aff} A$.

Это расширение является полупрямым произведением $\mathrm{GL}(L)$ и L. Чтобы убедиться в этом, фиксируем любую точку $a \in A$ и рассмотрим подгруппу $G_a \subset \operatorname{Aff} A$, состоящую из отображений, оставляющих a на месте. По предложению п. 11 § 1 каждый элемент $f \in G_a$ однозначно определяется своей линейной частью Df, и Df можно выбирать как угодно. Следовательно, D индуцирует изоморфизм G_a с $\mathrm{GL}(L)$. Для любого отображения $f \in \operatorname{Aff} A$ можно найти единственное отображение $f_a \in G_a$ с той же линейной частью, и $f = t_l \circ f_a$ для подходящего $l \in L$ по следствию п. 13 § 1. Фиксировав a, будем записывать $t_l \circ f_a$ в виде пары $[g; l]$, где $g = Df = Df_a \in \mathrm{GL}(L)$. Правила умножения в группе $\operatorname{Aff} A$ на языке таких пар имеют следующий вид.

2. Предложение. *Имеем*

$$[g_1; l_1][g_2; l_2] = [g_1 g_2; g_1(l_2) + l_1],$$
$$[g; l]^{-1} = [g^{-1}; -g^{-1}(l)].$$

Доказательство. Согласно определениям $[g; l]$ переводит точку $a + m \in A$ в $a + g(m) + l$, откуда

$$[g_1; l_1][g_2; l_2](a+m) = [g_1; l_1](a + g_1(m) + l_1) =$$
$$= a + g_1(g_2(m) + l_2) + l_1 = a + g_1 g_2(m) + g_1(l_2) + l_1 =$$
$$= [g_1 g_2; g_1(l_2) + l_1](a+m),$$

что доказывает первую формулу. Вычисляя с ее помощью произведение $[g; l][g^{-1}; -g^{-1}(l)]$, получаем $[\mathrm{id}_L; 0]$; а эта пара представляет тождественный элемент $\mathrm{Aff}\, A$. Это завершает доказательство предложения и показывает, что $\mathrm{Aff}\, A$ — полупрямое произведение.

3. Пусть теперь $G \subset \mathrm{GL}(L)$ — некоторая подгруппа. Множество всех элементов $f \in \mathrm{Aff}\, A$, линейные части которых принадлежат G, очевидно, образуют подгруппу в $\mathrm{Aff}\, A$ — прообраз G относительно канонического гомоморфизма $\mathrm{Aff}\, A \to \mathrm{GL}(L)$. Мы будем называть ее *аффинным расширением группы* G.

Особенно важен случай, когда ассоциированное с A линейное пространство снабжено дополнительной структурой — скалярным произведением, а G представляет собой соответствующую группу изометрий. Так строятся две важные в приложениях группы: *группа движений* аффинного евклидова пространства ($G = \mathrm{O}(n)$) и *группа Пуанкаре* (L — пространство Минковского, G — группа Лоренца). Изучим подробнее группу движений.

4. Определение. а) *Аффинным евклидовым пространством называется пара, состоящая из аффинного конечномерного пространства A над полем вещественных чисел и метрики d на нем* (*в смысле определения п.* 1 § 10 *ч.* 1), *которая обладает следующим свойством: для любых точек $a, b \in A$ расстояние $d(a, b)$ зависит только от $a - b \in L$ и совпадает с длиной вектора $a - b$ в подходящей евклидовой метрике пространства L (не зависящей от a, b).*

б) *Движением аффинного евклидова пространства A называется произвольное отображение $f\colon A \to A$, сохраняющее расстояния: $d(f(a), f(b)) = d(a, b)$ для всех $a, b \in A$.*

5. Теорема. *Движения аффинного евклидова пространства A образуют группу, совпадающую с аффинным расширением группы ортогональных изометрий $\mathrm{O}(L)$ ассоциированного с A евклидова пространства L.*

Доказательство. Проверим сначала, что любое аффинное отображение $f\colon A \to A$ с $Df \in \mathrm{O}(L)$ является движением. В самом деле, согласно определениям

$$d(f(a), f(b)) = |f(a) - f(b)| = |Df(a - b)| = |a - b| = d(a, b);$$

в третьем равенстве мы воспользовались тем, что $Df \in \mathrm{O}(L)$.

Основная работа связана с доказательством обратного утверждения.

Прежде всего, очевидно, что композиция движений есть движение. Далее, мы уже установили, что сдвиги являются движениями. Пусть $a \in A$ — произвольная фиксированная точка, f — движение. Положим $g = t_{a-f(a)} \circ f$. Это движение, оставляющее точку a на месте. Достаточно доказать, что оно аффинное и что $Dg \in \mathrm{O}(L)$. Отождествим A с L, как в п. 12 § 1, с помощью отображения с тождественной линейной частью, переводящего a в $0 \in L$. Тогда g превратится в отображение $g\colon L \to L$ со свойствами $g(0) = 0$ и $|g(l) - g(m)| = |l - m|$ для всех $l, m \in L$, и достаточно установить, что такое отображение лежит в $\mathrm{O}(L)$.

Проверим прежде всего, что g сохраняет скалярные произведения. Действительно, для любых $l, m \in L$

$$|l|^2 - 2(l, m) + |m|^2 = |l - m|^2 = |g(l) - g(m)|^2 = \\ = |g(l)|^2 - 2(g(l), g(m)) + |g(m)|^2,$$

откуда следует требуемое, ибо $|g(l)|^2 = |l|^2$, $|g(m)|^2 = |m|^2$. Теперь покажем, что g аддитивно: $g(l+m) = g(l) + g(m)$. Положив $l + m = n$ и воспользовавшись предыдущим свойством, имеем

$$0 = |n - l - m|^2 = |n|^2 + |l|^2 + |m|^2 - 2(n, l) - 2(n, m) + 2(l, m) = \\ = |g(n)|^2 + |g(l)|^2 + |g(m)|^2 - 2(g(n), g(l)) - 2(g(n), g(m)) + \\ + 2(g(l), g(m)) = |g(n) - g(l) - g(m)|^2,$$

откуда $g(n) = g(l) + g(m)$.

Наконец, покажем, что $g(xl) = xg(l)$ для всех $x \in \mathbf{R}$, $l \in L$. Полагая $m = xl$, имеем

$$0 = |m - xl|^2 = |m|^2 - 2x(m, l) + x^2|l|^2 = \\ = |g(m)|^2 - 2x(g(m), g(l)) + x^2|g(l)|^2 = |g(m) - xg(l)|^2.$$

Итак, g — линейное отображение, сохраняющее скалярные произведения, т. е. $g \in \mathrm{O}(L)$. Теорема доказана.

6. Теорема. *Пусть f: $A \to A$ — движение евклидова аффинного пространства с линейной частью Df. Тогда существует такой вектор $l \in L$, что $Df(l) = l$ и $f = t_l \circ g$, где g: $A \to A$ — движение, имеющее неподвижную точку $a \in A$.*

Доказательство. Прежде всего, выясним геометрический смысл этого утверждения. Отождествив A с L посредством аффинного отображения с тождественной линейной частью, которое переводит a в 0, мы получаем, что f является композицией ортогонального преобразования g и сдвига на вектор l, неподвижный относительно g (ибо $Df = Dg$). Иными словами, это «винтовое движение», если $\det g = 1$, или винтовое движение, скомбинированное с отражением, если $\det g = -1$. В самом деле, g вполне определяется своим ограничением g_0 на $l^\perp$: $g = g_0 \oplus \mathrm{id}_{\mathbf{R}l}$, так что g есть вращение вокруг оси $\mathbf{R}l$ (возможно, с отражением).

Приступим теперь к доказательству. Положим $L_2 = \mathrm{Ker}(Df - \mathrm{id}_L)$, $L_1 = L_2^\perp$. Имеем $L = L_1 \oplus L_2$; L_2 состоит из Df-инвариантных векторов, пространство L_1 инвариантно относительно $Df - \mathrm{id}_L$ (ибо Df ортогонален), и ограничение $Df - \mathrm{id}_L$ на L_1 обратимо.

Выберем сначала произвольную точку $a' \in A$ и положим $g' = t_{a' - f(a')} \circ f$. Очевидно, $g'(a') = a'$. Положим $f(a') - a' = l_1 + l_2$, где $l_1 \in L_1$, $l_2 \in L_2$, тогда $f = t_{l_2} \circ t_{l_1} \circ g'$ и $Df(l_2) = l_2$ по определению. Покажем, что $g = t_{l_1} \circ g'$ имеет неподвижную точку $a = a' + m$ для некоторого $m \in L_1$. Имеем

$$t_{l_1} \circ g'(a' + m) = g'(a' + m) + l_1 = a' + Df(m) + l_1.$$

Правая часть равна $a' + m$ тогда и только тогда, когда

$(Df-\mathrm{id}_L)m+l_1=0$. Но, как мы уже отмечали, на L_1 оператор $Df-\mathrm{id}_L$ обратим и $l_1\in L_1$. Поэтому m существует. Мы получили требуемое разложение $f=t_{l_2}\circ g$ и завершили доказательство.

Движения f со свойством $\det Df=1$ называются иногда *собственными движениями*, а остальные (с $\det Df=-1$) — несобственными. Представим более наглядно информацию о движениях аффинных евклидовых пространств размерности $n\leqslant 3$, содержащуюся в теореме п. 6. В следующем пункте сохранены обозначения этой теоремы.

7. Примеры. а) $n=1$. Поскольку $\mathrm{O}(1)=\{\pm 1\}$, собственные движения состоят только из сдвигов. Если f несобственное, то $Df=-1$, и из $Df(l)=l$ следует, что $l=0$. Поэтому всякое несобственное движение прямой имеет неподвижную точку и, стало быть, является отражением относительно этой точки.

б) $n=2$. Собственное движение f с $Df=\mathrm{id}$ является сдвигом; если $Df\neq\mathrm{id}$ и $\det Df=1$, то Df, будучи вращением, не имеет неподвижных векторов, так что снова $l=0$ и f имеет неподвижную точку, относительно которой f является вращением.

Если f — несобственное движение, то Df есть отражение плоскости относительно прямой, а f есть комбинация такого отражения и сдвига вдоль этой прямой. Значит, если несобственное движение плоскости имеет неподвижную точку, то оно имеет целую прямую неподвижных точек и представляет собой отражение относительно этой прямой.

в) $n=3$. Если $\det Df=1$, то Df всегда имеет собственное значение единица и неподвижный вектор. Поэтому все собственные движения трехмерного евклидова пространства являются винтовыми движениями вдоль некоторой оси (включая сдвиги, т. е. вырожденные винтовые движения с нулевым поворотом). Это — так называемая *теорема Шаля.*

Если движение $f=t_lg$ несобственное и $l\neq 0$, то ограничение g на плоскость, ортогональную к l и проходящую через неподвижную точку a, есть несобственное движение этой плоскости. Поэтому оно является отражением относительно прямой в этой плоскости. Обозначим через P плоскость, натянутую на l и на эту прямую. Тогда t_lg есть комбинация отражения относительно плоскости P и сдвига на вектор l, лежащий в P.

Наконец, если $l=0$, т. е. f несобственное и имеет неподвижную точку, то, отождествляя ее с нулем в L, а f с Df и пользуясь существованием у f собственной прямой L_0 с собственным значением минус единица, получаем геометрическое описание f как композиции вращения в $L_0^{\perp}$ и отражения относительно $L_0^{\perp}$.

Пользуясь полярным разложением линейных операторов, мы можем также разобраться в геометрической структуре любого обратимого аффинного преобразования евклидова аффинного пространства.

8. Теорема. *Всякое аффинное преобразование n-мерного евклидова пространства f может быть представлено в виде композиции трех отображений:* а) *n растяжений (с положительными коэффи-*

циентами) вдоль n попарно ортогональных осей, проходящих через некоторую точку $a_0 \in A$; б) *движения, оставляющего неподвижным точку* a_0; в) *сдвига.*

Доказательство. Заменив f его композицией с подходящим сдвигом, как в доказательстве теоремы п. 5, мы можем считать, что уже f имеет неподвижную точку a_0. Отождествив A с L и a_0 с нулем, мы можем разложить $f = Df$ в композицию положительно определенного симметрического оператора и ортогонального оператора. Приведя первый из них к главным осям и перенеся эти оси в A, получим требуемое.

9. Заметим в заключение, что в этом параграфе мы широко пользовались линейными подмногообразиями в A (прямыми, плоскостями), определяя их конструктивно как прообразы линейных пространств в L при разных отождествлениях A с L, зависящих от выбора начала координат. Следующий параграф посвящен более систематическому исследованию связанных с этим понятий.

§ 3. Аффинные подпространства

1. Определение. *Пусть* (A, L) — *некоторое аффинное пространство. Подмножество* $B \subset A$ *называется аффинным подпространством в* A, *если оно пусто или если множество*

$$M = \{b_1 - b_2 \in L \mid b_1,\ b_2 \in B\} \subset L$$

является линейным подпространством в L *и* $t_m(B) \subset B$ *для всех* $m \in M$.

2. Замечания. а) Если выполнены требования определения и B непусто, то пара (B, M) образует аффинное пространство, что оправдывает терминологию (подразумевается, что сдвиг B посредством вектора из M получается ограничением на B этого же сдвига на всем A). В самом деле, просмотр условий в определении п. 1 § 1 сразу же показывает, что они выполнены для (B, M). В частности, выбрав любую точку $b \in B$, получаем $B = \{b + m \mid m \in M\}$.

б) Будем называть линейное подпространство $M = \{b_1 - b_2 \mid b_1,\ b_2 \in B\}$ *направляющим* для аффинного подпространства B. Размерностью B называется размерность M. Очевидно, из $B_1 \subset B_2$ следует, что $M_1 \subset M_2$ и, значит, $\dim B_1 \leqslant \dim B_2$. Назовем два аффинных подпространства одинаковой размерности с общим направляющим пространством *параллельными.*

3. Предложение. *Аффинные подпространства одинаковой размерности* B_1, $B_2 \subset A$ *параллельны тогда и только тогда, когда существует такой вектор* $l \in L$, *что* $B_2 = t_l(B_1)$. *Любые два вектора с таким свойством отличаются на вектор из направляющего пространства для* B_1 *и* B_2.

Доказательство. Если $B_2 = t_l(B_1)$ и M_2, M_1 — направляющие B_2 и B_1 соответственно, то

$$M_2 = \{a - b \mid a,\ b \in B_2\} = \{(a' + l) - (b' + l) \mid a',\ b' \in B_1\} = M_1,$$

так что B_1 и B_2 параллельны.

Наоборот, пусть M — общее направляющее для B_1 и B_2. Выберем точки $b_1 \in B$ и $b_2 \in B_2$. Имеем $B_1 = \{b_1 + l \mid l \in M\}$, $B_2 = \{b_2 + l \mid l \in M\}$, откуда $B_2 = t_{b_2 - b_1}(B_1)$. Наконец, легко видеть, что $t_{l_1}(B_1) = t_{l_2}(B_2)$ тогда и только тогда, когда $l_1 - l_2 \in M$.

4. Следствие. *Аффинные подпространства в L (с аффинной структурой) — это линейные подмногообразия L в смысле определения п. 1 § 6 ч. 1, т. е. сдвиги линейных подпространств.*

5. Следствие. *Параллельные аффинные подпространства одинаковой размерности либо не пересекаются, либо совпадают.*

Доказательство. Если $b \in B_1 \cap B_2$, то по предыдущему $B_1 = \{b + m \mid m \in M\} = B_2$, где M — общее направляющее B_1 и B_2.

6. Аффинные подпространства B_1 и B_2 не обязательно одинаковой размерности называются *параллельными*, если одно из их направляющих содержится в другом. Слегка изменяя предыдущие доказательства, легко получить следующие факты. Пусть B_1 и B_2 параллельны и $\dim B_1 \leqslant \dim B_2$. Тогда существует такой вектор $l \in L$, что $t_l(B_1) \subset B_2$, и два вектора с этим свойством отличаются на элемент из M_1. Кроме того, либо B_1 и B_2 не пересекаются, либо B_1 содержится в B_2.

7. Предложение. *Пусть (B_1, M_1), (B_2, M_2) — два аффинных подпространства в A. Тогда $B_1 \cap B_2$ либо пусто, либо является аффинным подпространством с направляющим $M_1 \cap M_2$.*

Доказательство. Пусть $B_1 \cap B_2$ непусто и $b \in B_1 \cap B_2$. Тогда $B_1 = \{b + l_1 \mid l \in M_1\}$, $B_2 = \{b + l_2 \mid l \in M_2\}$, откуда $B_1 \cap B_2 = \{b + l \mid l \in M_1 \cap M_2\}$, что доказывает требуемое. (Следствие п. 5, очевидно, вытекает отсюда.)

8. Аффинные оболочки. Пусть $S \subset A$ — некоторое множество точек в аффинном пространстве A. Наименьшее аффинное подпространство, содержащее S, называется *аффинной оболочкой* S. Оно существует и совпадает с пересечением всех аффинных подпространств, содержащих S. Мы можем описать аффинную оболочку в терминах барицентрических линейных комбинаций (предложение п. 11 § 1).

9. Предложение. *Аффинная оболочка множества S совпадает с множеством барицентрических комбинаций элементов из S:*

$$\tilde{S} = \left\{ \sum_{i=1}^{n} x_i s_i \,\middle|\, \sum_{i=1}^{n} x_i = 1 \right\},$$

где $\{s_1, \ldots, s_n\} \subset S$ пробегает всевозможные конечные подмножества S.

Доказательство. Покажем прежде всего, что барицентрические комбинации образуют аффинное подпространство в A. В самом деле, обозначим через $M \subset L$ линейное подпространство, натянутое на всевозможные векторы $s - t$; $s, t \in S$. Любые две барицентрические комбинации точек S можно представить в виде $\sum_{i=1}^{n} x_i s_i$, $\sum_{i=1}^{n} y_i s_i$ с одним и тем же множеством $\{s_1, \ldots, s_n\}$, взяв объединение двух исходных множеств и положив лишние коэффициенты равными нулю. Поскольку $\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0$, разность этих комбинаций можно представить в виде

$$\sum_{i=1}^{n} (x_i - y_i)(s_i - s_1)$$

и потому она лежит в M. Наоборот, любой элемент из M вида $\sum_{i=1}^{n} x_i (s_i - t_i)$ есть разность точек $\sum_{i=1}^{n} x_i s_i + \left(1 - \sum_{i=1}^{n} x_i\right) s_1$ и $\sum_{i=1}^{n} x_i t_i + \left(1 - \sum_{i=1}^{n} x_i\right) s_1$ из $\bar{S}$. Поэтому $M = \{b_1 - b_2 | b_1, b_2 \in S\}$. Это же соображение показывает, что $t_m(\bar{S}) \subset \bar{S}$ для всех $m \in M$. Следовательно, $\bar{S}$ является аффинным подпространством с направляющим пространством M. Ясно, что $S \subset \bar{S}$.

Наоборот, пусть $B \supset S$ — любое аффинное подпространство, $\{s_1, \ldots, s_n\} \subset S$. Тогда для любых $x_1, \ldots, x_n \in \mathcal{K}$, $\sum_{i=1}^{n} x_i = 1$, имеем

$$\sum_{i=1}^{n} x_i s_i = s_1 + \sum_{i=1}^{n} x_i (s_i - s_1).$$

Поскольку $s_1, \ldots, s_n \in B$, вектор $\sum_{i=1}^{n} x_i (s_i - s_1)$ лежит в направляющем пространстве B и потому сдвиг s_1 на него лежит в B. Значит, $\bar{S} \subset B$ и $\bar{S}$ действительно является наименьшим аффинным подпространством, содержащим S.

10. Предложение. *Пусть f: $A_1 \to A_2$ — аффинное отображение двух аффинных пространств; $B_1 \subset A_1$ и $B_2 \subset A_2$ — аффинные подпространства. Тогда $f(B_1) \subset A_2$ и $f^{-1}(B_2) \subset A_1$ являются аффинными подпространствами.*

Доказательство. Пусть $B_1 = \{b + l | l \in M_1\}$, где M_1 — направляющее пространство для B_1. Тогда $f(B_1) = \{f(b) + Df(l) | l \in M_1\} = \{f(b) + l' | l' \in \operatorname{Im} Df\}$. Следовательно, $f(B_1)$ — аффинное подпространство с направляющим пространством $\operatorname{Im} Df$.

В частности, $f(A_1)$ есть аффинное подпространство в A_2, $B_2 \cap f(A_1)$ есть аффинное подпространство и $f^{-1}(B_2) = f^{-1}(B_2 \cap f(A_1))$ в силу общих теоретико-множественных определений. Заменив A_2 на $f(A_1)$ и B_2 на $B_2 \cap f(A_1)$, мы можем ограничиться случаем, когда f сюръективно. Пусть M_2 — направляющее пространство для B_2. Тогда $B_2 = \{b + m | m \in M_2\}$ и $f^{-1}(B_2) = \{b' + m' | f(b') = b, Df(m') \in M_2\}$, Справа можно ограничиться одним значением $b' \in f^{-1}(b)$: остальные получатся из него сдвигами на $\operatorname{Ker} Df$. Отсюда следует, что $f^{-1}(B_2)$ имеет вид $\{b' + m | m \in Df^{-1}(M_2)\}$ и потому является аффинным подпространством с направляющим подпространством $(Df)^{-1}(M_2)$.

11. Следствие. *Множество уровня любой аффинно линейной функции является аффинным подпространством.*

Доказательство. В самом деле, множества уровня аффинно линейной функции $f: A \to \mathcal{K}^1$ суть прообразы точек в $\mathcal{K}^1$. Но любая точка в аффинном пространстве является аффинным подпространством (с направляющим $\{0\}$).

12. Предложение. *Пусть $f_1, \ldots, f_n$ — аффинно линейные функции на аффинном пространстве A. Тогда множество $\{a \in A \mid f_1(a_1) = \ldots = f_n(a_n) = 0\}$ является аффинным подпространством в A. Если A конечномерно, то всякое его аффинное подпространство имеет такой вид.*

Доказательство. Указанное множество является конечным пересечением множеств уровня аффинно линейных функций. Поэтому оно аффинное в силу следствия п. 11 и предложения п. 7. Наоборот, пусть $B \subset A$ — аффинное подпространство в конечномерном аффинном пространстве A, $M \subset L$ — соответствующие линейные пространства. Если B пусто, его можно задать уравнением $f = 0$, где f — постоянная функция на A с ненулевым значением (очевидно, любая такая функция аффинно линейна, $Df = 0$). Иначе, пусть $g_1 = \ldots = g_n = 0$ — система линейных уравнений на L, задающая M; в качестве $g_1, \ldots, g_n$ можно взять, например, базис подпространства $M^{\perp} \subset L^*$. Выберем точку $b \in B$ и построим аффинно линейные функции $f_i: A \to \mathcal{K}^1$ с условиями $f_i(b) = 0$, $Df_i = g_i$, $i = 1, \ldots, n$. Очевидно, $f_i(b + l) = g_i(l)$. Поэтому точка $b + l \in A$ обращает в нуль все функции f_i тогда и только тогда, когда $l \in M$, т. е. тогда и только тогда, когда $b + l \in B$. Это завершает доказательство.

13. Назовем *конфигурацией* в аффинном пространстве A конечную упорядоченную систему аффинных подпространств $\{B_1, \ldots, B_n\}$. Две конфигурации $\{B_1, \ldots, B_n\}$ и $\{B_1', \ldots, B_n'\}$ назовем *аффинно конгруэнтными*, если существует такой аффинный автоморфизм $f \in \operatorname{Aff} A$, что $f(B_i) = B_i'$, $i = 1, \ldots, n$. Возможны варианты этого понятия, когда f разрешается выбирать лишь из некоторой подгруппы $\operatorname{Aff} A$, например, группы движений, когда A евклидово. В последнем случае будем называть конфигурации *метрически конгруэнтными*. Важные понятия и результаты аффинной геометрии связаны с отысканием инвариантов конфигураций относительно отношения конгруэнтности. Заметим, что оно является аффинным вариантом понятия «одинаковой расположенности», которое мы изучали в § 5 ч. 1.

Докажем несколько основных результатов о конгруэнтности.

Пусть A — аффинное пространство размерности n. В соответствии с результатами пп. 9—11 § 1 назовем конфигурацию $\{a_0, \ldots, a_n\}$ из $n + 1$ точки в A *координатной*, если ее аффинная оболочка совпадает с A.

14. Предложение. а) *Любые две координатные конфигурации конгруэнтны и переводятся друг в друга единственным отображением $f \in \operatorname{Aff} A$.*

б) *Координатные конфигурации* $\{a_0, \ldots, a_n\}$ *и* $\{a'_0, \ldots, a'_n\}$ *в евклидовом пространстве* A *метрически конгруэнтны тогда и только тогда, когда* $d(a_i, a_j) = d(a'_i, a'_j)$ *для любых* $i, j \in 1, \ldots, n$.

Доказательство. а) Положим $e_i = a_i - a_0$, $e'_i = a'_i - a'_0$. Системы $\{e_i\}$ и $\{e'_i\}$ образуют базисы в L. Пусть g: $L \to L$ — линейное отображение, переводящее e_i в e'_i. Построим аффинное отображение f: $A \to A$ со свойством $Df = g$ и $f(a_0) = a'_0$. Оно существует по утверждению п. 11 § 1 и лежит в $\mathrm{Aff}\, A$, ибо g обратимо. Кроме того,

$$f(a_i) = f(a_0) + g(a_i - a_0) = a'_0 + e'_i = a'_0 + (a'_i - a'_0) = a'_i$$

для всех $i = 1, \ldots, n$. Эта же формула показывает, что f единственная, ибо Df должно переводить e_i в e'_i и $f(a_0) = a'_0$.

б) В силу доказанного выше достаточно проверить, что f является движением тогда и только тогда, когда $d(a_i, a_j) = d(a'_i, a'_j)$ для всех i, j. В самом деле, $d(a_i, a_j) = |a_i - a_j| = |e_i - e_j|$, где $e_0 = a_0 - a_0 = 0$, и аналогично $d(a'_i, a'_j) = |e'_i - e'_j|$. Если f — движение, то Df ортогонально и сохраняет длины векторов, так что условие необходимо. Наоборот, пусть оно выполнено. Тогда $|e_i| = |e'_i|$ для всех $i = 1, \ldots, n$ и далее из равенств $|e_i - e_j|^2 = |e'_i - e'_j|^2$ получаем, что $(e_i, e_j) = (e'_i, e'_j)$ для всех i, j. Значит, матрицы Грама базисов $\{e_i\}$ и $\{e'_i\}$ совпадают. Но тогда отображение g, переводящее $\{e_i\}$ в $\{e'_i\}$, является изометрией, так что f является движением. Доказательство окончено.

Рассмотрим теперь конфигурации (b, B), состоящие из точки и аффинного подпространства. В евклидовом случае назовем расстоянием от b до B число

$$d(b, B) = \inf\{\|l\| \mid b + l \in B\}.$$

15. Предложение. а) *Конфигурации* (b, B) *и* (b', B') *аффинно конгруэнтны тогда и только тогда, когда* $\dim B = \dim B'$ *и либо одновременно* $b \notin B$, $b' \notin B'$, *либо одновременно* $b \in B$, $b' \in B'$.

б) *Конфигурации* (b, B) *и* (b', B') *метрически конгруэнтны тогда и только тогда, когда* $\dim B = \dim B'$ *и* $d(b, B) = d(b', B')$.

Доказательство. а) Сформулированные условия, очевидно, необходимы. Пусть они выполнены. Обозначим через M, M' направляющие B, B' соответственно и выберем линейный автоморфизм g: $L \to L$, для которого $g(M) = M'$. Если $b \in B$ и $b' \in B'$, построим аффинное отображение f: $A \to A$ с условиями $Df = g$ и $f(b) = b'$. Очевидно, $f(b + l) = b' + g(l)$, так что $f(B) = B'$.

Если $b \notin B$ и $b' \notin B'$, наложим на g дополнительные условия. Выберем по точке $a \in B$, $a' \in B'$ и потребуем, чтобы g переводил вектор $b - a$ в вектор $b' - a'$. Оба вектора ненулевые и лежат вне M, M' соответственно, поэтому стандартная конструкция, исходящая из базисов L вида {базис M, $b - a$, дополнение} и {базис M', $b'-a'$, дополнение}, показывает существование g. После этого снова построим аффинное отображение f: $A \to A$ с $Df=g$ и $f(b)=b'$. Проверим, что $f(B)=B$. В самом деле, прежде всего, $f(a)=a'$, потому что

$$f(a)=f(b-(b-a))=f(b)-g(b-a)=b'-(b'-a')=a'.$$

Далее, $f(a+l)=f(a)+g(l)$, и условие $l \in M$ равносильно условию $g(l) \in M'$, так что $f(B)=B'$.

б) Необходимость условия снова очевидна. Для доказательства достаточности подчиним выборы, сделанные в предыдущем рассуждении, дополнительным требованиям. Прежде всего, отождествим A с L, выбрав начало координат в B. Тогда B отождествится с M, b станет некоторым вектором в L. Пусть a — ортогональная проекция b на M. В линейном варианте мы уже знаем, что $d(b, B)=|b-a|$. Аналогично определим точку a' на M' или в нашем отождествлении на B'. В качестве g возьмем изометрию L, переводящую M в M' и b в b'. Она существует: дополним ортонормированные базисы в M и M' соответственно до ортонормированных базисов в L, содержащих $(b-a)/|b-a|$ и $(b'-a')/$ $/|b-a|$, и определим g как изометрию, переводящую первый базис во второй. После этого аффинное отображение f: $A \to A$ с $Df=g$ и $f(b)=b'$ будет движением, переводящим (b, B) в (b', B').

16. Рассмотрим, наконец, конфигурации, состоящие из двух подпространств B_1, B_2. Полная классификация их с точностью до аффинной конгруэнтности может быть проведена с помощью соответствующего результата для линейных подпространств, доказанного в п. 5 § 5 ч. 1. Полная метрическая классификация довольно громоздка: она требует рассмотрения расстояния между B_1 и B_2 и серии углов. Мы ограничимся обсуждением единственного метрического инварианта — расстояния, которое, как обычно, определим формулой

$$d(B_1, B_2)=\inf\{|b_1-b_2| \mid b_1 \in B_1,\ b_2 \in B_2\}.$$

Назовем *общим перпендикуляром к* B_1, B_2 такую пару точек $b_1 \in B_1$, $b_2 \in B_2$, что вектор b_1-b_2 ортогонален к направляющим B_1 и B_2. (Точнее было бы называть общим перпендикуляром отрезок $\{tb_1+(1-t)b_2 \mid 0 \leqslant t \leqslant 1\}$.)

17. Предложение. а) *Общий перпендикуляр к B_1 и B_2 всегда существует. Множество общих перпендикуляров биективно пересечению направляющих B_1 и B_2.*

б) *Расстояние между B_1 и B_2 равно длине любого общего перпендикуляра $|b_1-b_2|$ к ним.*

Доказательство. а) Пусть M_1, M_2 — направляющие B_1 и B_2 и пусть $b_1' \in B_1$, $b_2' \in B_2$. Спроектируем вектор $b_1'-b_2'$ ортогонально на M_1+M_2 и представим проекцию в виде m_1+m_2, $m_i \in$ $\in M_i$. Положим $b_1=b_1'-m_1$, $b_2=b_2'+m_2$. Очевидно, $b_i \in B_i$ и

$$b_1-b_2=b_1'-b_2'-(m_1+m_2) \in (M_1+M_2)^{\perp}.$$

Значит, $\{b_1, b_2\}$ есть общий перпендикуляр к B_1, B_2.

Пусть $\{b_1, b_2\}$ и $\{b'_1, b'_2\}$ — два общих перпендикуляра. Тогда $b_1 - b'_1 \in M_1$, $b_2 - b'_2 \in M_2$ и, кроме того,

$$b_1 - b_2 \in (M_1 + M_2)^{\perp},\ b'_1 - b'_2 \in (M_1 + M_2)^{\perp}.$$

Значит, разность $(b_1 - b'_1) - (b_2 - b'_2)$, лежит одновременно в $M_1 + M_2$ и $(M_1 + M_2)^{\perp}$. Поэтому она равна нулю. Следовательно, $b_1 - b'_1 = b_2 - b'_2 \in M_1 \cap M_2$. Наоборот, если $\{b_1, b_2\}$ — фиксированный общий перпендикуляр и $m \in M_1 \cap M_2$, то $\{b_1 + m, b_2 + m\}$ тоже является общим перпендикуляром. Это завершает доказательство первой части предложения.

б) Пусть $\{b_1, b_2\}$ — общий перпендикуляр к B_1, B_2 и $b'_1 \in B_1$, $b'_2 \in B_2$ — любая другая пара точек. Достаточно доказать, что $|b_1 - b_2| \leqslant |b'_1 - b'_2|$. В самом деле,

$$b'_1 - b'_2 = (b_1 - b_2) + (b'_1 - b_1) + (b_2 - b'_2).$$

Но $(b'_1 - b_1) + (b_2 - b'_2) \in M_1 + M_2$, а вектор $b_1 - b_2$ ортогонален $M_1 + M_2$. Значит, по теореме Пифагора

$$|b'_1 - b'_2|^2 = |b_1 - b_2|^2 + |b'_1 - b_1 + b_2 - b'_2|^2 \geqslant |b_1 - b_2|^2,$$

что завершает доказательство.

Установим в заключение один полезный результат, характеризующий аффинные подпространства.

18. Предложение. *Подмножество $S \subset A$ является аффинным подпространством тогда и только тогда, когда вместе с любыми двумя точками s, $t \in S$ оно содержит всю прямую, проходящую через эти точки, т. е. их аффинную оболочку.*

Доказательство. Прямая, проходящая через точки s, $t \in S$, — это множество $\{xs + (1 - x)t \mid x \in \mathscr{K}\}$. Поэтому необходимость условия следует из предложения п. 9. Наоборот, пусть оно выполнено. Поскольку в силу того же предложения аффинная оболочка S состоит из всевозможных барицентрических комбинаций точек S, мы должны проверить, что такие комбинации $\sum_{i=1}^{n} x_i s_i$ лежат в S. Проведем индукцию по n. При $n = 1$, 2 результат очевиден. Пусть $n > 2$ и для меньших значений n результат доказан. Представим $\sum_{i=1}^{n} x_i s_i$ в виде

$$y_1 \sum_{i=1}^{n-2} \frac{x_i}{y_1} s_i + y_2 \sum_{i=n-1}^{n} \frac{x_i}{y_2} s_i,$$

где $y_1 = \sum_{i=1}^{n-2} x_i$, $y_2 = x_{n-1} + x_n$ (мы можем считать, что обе эти суммы не равны нулю, иначе $\sum_{i=1}^{n} x_i s_i \in S$ по индуктивному предположению). Очевидно,

$$\sum_{i=1}^{n-2}\frac{x_i}{y_1}=\sum_{i=n-1}^{n}\frac{x_i}{y_2}=y_1+y_2=1.$$

Значит, $\sum_{i=1}^{n-2}\frac{x_i}{y_1}s_i$ и $\sum_{i=n-1}^{n}\frac{x_i}{y_2}s_i$ лежат в S, и потому их барицентрическая комбинация с коэффициентами y_1, y_2 лежит в S. Это завершает доказательство.

УПРАЖНЕНИЯ

1. Назовем i-й медианой системы точек $a_1, \ldots, a_n \in A$ отрезок, соединяющий точку a_i с центром масс остальных точек $\{a_j/j \neq i\}$. Доказать, что все медианы пересекаются в одной точке — центре масс $a_1, \ldots, a_n$.

2. Угол между двумя прямыми в евклидовом аффинном пространстве A — это угол между их направляющими. Доказать, что две конфигурации из двух прямых в A метрически конгруэнтны тогда и только тогда, когда углы и расстояния между прямыми в обеих конфигурациях совпадают.

3. Угол между прямой и аффинным подпространством размерности $\geqslant 1$ — это угол между направляющей прямой и ее проекцией на направляющую подпространства. Пользуясь этим определением, обобщить результат упражнения 2 на конфигурации, состоящие из прямой и подпространства.

§ 4. Выпуклые многогранники и линейное программирование

1. Постановка задачи. Основная задача линейного программирования ставится следующим образом. Дано конечномерное аффинное пространство A над полем вещественных чисел $\mathbf{R}$ и $m+1$ аффинно линейных функций $f_1, \ldots, f_m$; $f: A \to \mathbf{R}$. Требуется отыскать точку (или точки) $a \in A$, удовлетворяющие условиям $f_1(a) \geqslant 0, \ldots$ $\ldots, f_m(a) \geqslant 0$, для которых функция f принимает наибольшее возможное значение при этих ограничениях.

Вариант, в котором некоторые из неравенств направлены в обратную сторону, $f_i(a) \leqslant 0$, и/или требуется отыскать точки, в которых f принимает наименьшее возможное значение, сводится к предыдущему случаю заменой знака соответствующих функций. Условие $f_i(a)=0$ равносильно совокупности условий $f_i(a) \geqslant 0$ и $-f_i(a) \geqslant 0$. Все функции f_i можно считать непостоянными.

2. Мотивировка. Рассмотрим следующую математическую модель производства. Пусть имеется предприятие, использующее m видов различных ресурсов и производящее n видов различных продуктов. Ресурсы и продукты измеряются в своих единицах неотрицательными вещественными числами (случай, когда это целые числа, например, количество штук автомобилей, мы не рассматриваем; при больших объемах производства и потребления ресурсов он хорошо аппроксимируется «непрерывной» моделью).

План производства — это вектор $(x_1, \ldots, x_n) \in \mathbf{R}^n$, указывающий количество x_j j-го продукта, которое необходимо произвести. Принимается следующая линейная модель потребления ресурсов:

если на производство единицы j-го продукта расходуется количество a_{ij} единиц i-го ресурса, то для выполнения плана $(x_1, \ldots, x_n)$ требуется $\sum_{j=1}^{n} a_{ij}x_j$ единиц i-го ресурса. Ресурсы, отпускаемые предприятию, определяются вектором $(b_1, \ldots, b_m)$: дается b_i единиц i-го ресурса. Следовательно, план $(x_1, \ldots, x_n)$ выполним, только если выполняется система ограничений

$$f_i(x_1, \ldots, x_n) = b_i - \sum_{j=1}^{n} a_{ij}x_j \geqslant 0, \qquad i = 1, \ldots, m.$$

Мы будем всегда считать, что эти неравенства совместны.

Предположим, что предприятие реализует выпущенную им продукцию по цене c_i за единицу i-го продукта. Тогда прибыль от реализации произведенного продукта будет равна

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} c_i x_i.$$

План производства $(x_1, \ldots, x_n)$ называется оптимальным по прибыли, если $f(x)$ достигает наибольшего возможного значения при ограничениях $f_j \geqslant 0$, $x_i \geqslant 0$ $(i = 1, \ldots, n)$ (последнее условие означает, что предприятие не добывает производимых им продуктов на стороне — для продажи или для запчастей).

Мы видим, что задача составления оптимального плана является частным случаем задачи, сформулированной в п. 1.

Разумеется, практические приложения линейного программирования связаны с разработкой конкретных алгоритмов отыскания оптимального плана, которые можно применять вручную или на ЭВМ. Мы ограничимся в этом параграфе изложением геометрических аспектов задачи, лежащих, конечно, в основе всех алгоритмов.

3. Основные геометрические понятия. Фиксируем конечномерное аффинное пространство A над полем $\mathbf{R}$. Буквы f с индексами будут обозначать аффинно линейные функции на A.

Полупространством называется множество точек вида $\{a \in A \mid f(a) \geqslant 0\}$, где f — непостоянная аффинно линейная функция. *Многогранником* называется пересечение конечного числа полупространств.

Напомним, что подмножество $S \subset A$ выпуклое, если из $a_1, a_2 \in S$ и $0 \leqslant x \leqslant 1$ следует, что $xa_1 + (1-x)a_2 \in S$. Поскольку $f(xa_1 + (1-x)a_2) = xf(a_1) + (1-x)f(a_2)$, все полупространства выпуклы. Так как пересечение любого семейства выпуклых множеств выпуклое, все многогранники выпуклые. Мы будем говорить, что любая точка $xa_1 + (1-x)a_2$, $0 < x < 1$, является *внутренней* точкой отрезка с концами a_1 и a_2.

Пусть S — выпуклое множество. Выпуклое подмножество $T \subset S$ называется *гранью* S, если любой отрезок с концами в S, некоторая внутренняя точка которого лежит в T, целиком лежит в T.

Все множество S является своею гранью. Грань S, состоящая из одной точки, называется *вершиной* S. Читателю следует представить себе куб, октаэдр и многогранный угол в трехмерном пространстве, чтобы иметь наглядную картину основной ситуации, важной для линейного программирования. Грани этих фигур в смысле нашего определения — это грани, ребра и вершины школьной геометрии плюс сама фигура. Вершины шара — это все точки его поверхности.

Важнейший результат этого параграфа будет состоять в том, что максимум аффинно линейной функции на ограниченном многограннике (в приложениях этот случай наиболее распространен) достигается на одной из его вершин; последних конечное число. Но прежде нам придется разобраться в структуре многогранников и их граней подробнее.

4. Лемма. *Пересечение семейства граней и грань грани выпуклого множества S является гранью S.*

Доказательство. а) Пусть $T=\bigcap T_i$, T_i — грани S. Любой отрезок с концами в S, внутренняя точка которого принадлежит T_i, целиком лежит в T_i. Значит, если его внутренняя точка лежит в T, то он лежит в T.

б) Пусть $T_1 \subset T \subset S$, T — грань S. Любой отрезок с концами в S, внутреняя точка которого лежит в T_1, целиком лежит в T, ибо T — грань S, значит, его концы лежат в T и потому он целиком лежит в T_1, ибо T_1 — грань T.

5. Лемма. *Пусть S — многогранник, заданный неравенствами $f_i \geqslant 0$, $i=1, \ldots, m$. Тогда для любого i многогранник $S_i = S \cap \{a \mid f_i(a)=0\}$ либо пуст, либо является гранью S.*

Доказательство. Пусть S_i непуст, $a_1, a_2 \in S$ и внутренняя точка отрезка $xa_1+(1-x)a_2$ лежит в S_i. Функция $f_i(xa_1+(1-x)a_2)$, $0 \leqslant x \leqslant 1$, линейна по x, обращается в нуль для некоторого $0<x_0<1$ и, кроме того, неотрицательна при $x=0$ и $x=1$. Поэтому она тождественно равна нулю, так что весь отрезок лежит в S_i.

6. Лемма. *Непостоянная аффинно линейная функция f на многограннике $S=\{a \mid f_i(a) \geqslant 0\}$ не может принимать максимальное значение в точке $a \in S$, для которой все $f_i(a)>0$.*

Доказательство. Так как f непостоянна, $Df \neq 0$. Выберем в векторном пространстве L, ассоциированном с A, вектор $l \in L$, для которого $Df(l) \neq 0$. Можно считать, что $Df(l)>0$, изменив знак l в случае нужды. Если число $\varepsilon>0$ достаточно мало и $a \in S$, то $f_i(a+\varepsilon l)>0$ для всех $i=1, \ldots, m$: достаточно взять $\varepsilon<\min\limits_i \frac{f_i(a)}{|Df_i(l)|}$. Поэтому $a+\varepsilon l \in S$ для таких ε. Но $f(a+\varepsilon l)= f(a)+\varepsilon Df(l)>f(a)$, так что $f(a)$ не является максимальным значением f.

Теперь мы можем доказать наш основной результат.

7. Теорема. *Предположим, что аффинно линейная функция f ограничена сверху на многограннике S. Тогда она принимает свое максимальное значение во всех точках некоторой грани S, являю-*

щейся также многогранником. Если S ограничен, f принимает свое максимальное значение в некоторой вершине S.

Доказательство. Проведем индукцию по размерности A. Случай $\dim A = 0$ очевиден. Пусть $\dim A = n$ и для меньших размерностей теорема доказана. Пусть S задан системой неравенств $f_1 \geqslant 0, \ldots, f_m \geqslant 0$. Так как множество S замкнуто, ограниченная сверху функция f на нем принимает максимальное значение в некоторой точке a. Если $f_1(a) > 0, \ldots, f_m(a) > 0$, то по лемме п. 6 f может быть только константой; в частности свое единственное значение она принимает на всем S. Иначе $f_i(a) = 0$ для некоторого i. Это значит, что f принимает максимальное значение в точке непустого многогранника S_i, который является гранью S и лежит в аффинном подпространстве $\{a \mid f_i(a) = 0\}$ размерности $n-1$, ибо f_i непостоянна. По индуктивному предположению максимальное значение ограничения f на S_i принимается во всех точках некоторой многогранной грани S_i. По лемме п. 4 она же будет гранью S. Она будет многогранником, ибо к неравенствам, определяющим ее в S_i, с левыми частями, продолженными на все A, следует добавить равенство $f_i = 0$.

Теперь индукцией по размерности аффинной оболочки S покажем, что у любого ограниченного многогранника обязательно есть вершина. В самом деле, для размерности нуль это очевидно. Пусть размерность больше нуля. Мы можем считать, что аффинная оболочка S есть все A. Возьмем любую непостоянную аффинно линейную функцию на A. Она должна принимать на S максимальное значение, ибо S ограничен и замкнут. Стало быть, у S есть непустая грань, во всех точках которой это значение принимается. Она является ограниченным многогранником, аффинная оболочка которого имеет строго меньшую размерность. По индуктивному предположению у нее есть вершина, являющаяся также вершиной S по лемме п. 6.

Окончательно, пусть S ограничен и T — многогранная грань S, на которой исходная функция f принимает свое максимальное значение. Тогда любая вершина T, существование которой доказано, является искомой вершиной S.

§ 5. Аффинные квадратичные функции и квадрики

1. Определение. *Квадратичной функцией Q на аффинном пространстве (A, L) над полем $\mathscr{K}$ называется отображение $Q\colon A \to \mathscr{K}$, для которого существуют такие точка $a_0 \in A$, квадратичная форма $q\colon L \to \mathscr{K}$, линейная форма $l\colon L \to \mathscr{K}$ и константа $c \in \mathscr{K}$, что*

$$Q(a) = q(a - a_0) + l(a - a_0) + c$$

для всех $a \in A$.

Форма q называется квадратичной частью Q, а l — линейной частью Q относительно точки a_0. Очевидно, $c = Q(a_0)$. Покажем прежде всего, что от выбора точки a_0 квадратичность Q не зависит. Точнее, пусть g — симметричная билинейная форма на L, являющаяся поляризацией q. Мы, как обычно, считаем, что характеристика $\mathscr{K}$ отлична от двух.

2. Предложение. *Если* $Q(a)=q(a-a_0)+l(a-a_0)+c$, *то для любой точки* $a_0'\in A$ *имеем*

$$Q(a)=q(a-a_0')+l'(a-a_0')+c',$$

где

$$l'(m)=l(m)+2g(m,\ a_0'-a_0),\quad c'=Q(a_0').$$

Таким образом, переход к другой точкв меняет линейную часть Q *и константу.*

Доказательство. В самом деле,

$$q(a-a_0)=q((a-a_0')+(a_0'-a_0))=$$
$$=q(a-a_0')+2g(a-a_0',\ a_0'-a_0)+q(a_0'-a_0),$$
$$l(a-a_0)=l((a-a_0')+(a_0'-a_0))=l(a-a_0')+l(a_0'-a_0),$$

что доказывает требуемое.

3. Назовем точку a_0 *центральной* для квадратичной функции Q, если линейная часть Q относительно a_0 равна нулю. Объяснение этого термина состоит в замечании, что точка a_0 центральна тогда и только тогда, когда $Q(a)=Q(a_0-(a-a_0))$ для всех a: действительно, разность левой и правой части в общем случае равна $2l(a-a_0)$, ибо $q(a-a_0)=q(-(a-a_0))$. Геометрически это значит, что после отождествления A с L, при котором a_0 переходит в начало координат, функция Q становится симметричной относительно отражения $m\mapsto -m$.

Назовем *центром* функции Q множество ее центральных точек.

4. Теорема. а) *Если квадратичная часть* q *функции* Q *невырождена, то центр* Q *состоит из единственной точки.*

б) *Если* q *вырождена, то центр* Q *либо пуст, либо является аффинным подпространством в* A *размерности* $\dim A-\operatorname{rk} q$ ($\operatorname{rk} q$ — *это ранг* q), *направляющее подпространство которого совпадает с ядром* q.

Доказательство. Начнем с любой точки $a_0\in A$ и представим Q в виде $q(a-a_0)+l(a-a_0)+c$. Согласно предложению п. 2 точка $a_0'\in A$ будет центральной для Q тогда и только тогда, когда выполнены условия

$$l(m)=-2g(m,\ a_0'-a_0)$$

для всех $m\in L$. Когда a_0' пробегает все точки A, вектор $a_0'-a_0$ пробегает все элементы L, и линейная функция от $m\in L$ вида $-2g(m,\ a_0'-a_0)$ пробегает все элементы L^*, лежащие в образе канонического отображения $\tilde{g}\colon L\to L^*$, связанного с формой g.

Если q невырождена, то $\tilde{g}$ — изоморфизм. В частности, для функционала $-l/2\in L^*$ имеется единственный вектор $a_0'-a_0\in L$ со свойством $g(\cdot, a_0' - a_0) = -\frac{1}{2} l(\cdot)$. Точка a_0' в этом случае и является единственной центральной точкой Q.

Если q вырождена, то возможны два случая. Либо $-l/2$ не лежит в образе $\tilde{g}$; тогда центральных точек нет. Либо $-l/2$ лежит в образе $\tilde{g}$. Тогда для любых двух точек a_0', a_0'' с условием

$$g(\cdot, a_0' - a_0) = g(\cdot, a_0'' - a_0) = -\frac{1}{2} l(\cdot)$$

имеем $a_0' - a_0'' \in \operatorname{Ker} \tilde{g}$, и наоборот, если $g(\cdot, a_0' - a_0) = -\frac{1}{2} l(\cdot)$ и $a_0'' \in a_0' + \operatorname{Ker} \tilde{g}$, то

$$g(\cdot, a_0'' - a_0) = -\frac{1}{2} l(\cdot).$$

Таким образом, центр является аффинным подпространством, а $\operatorname{Ker} \tilde{g}$, т. е. ядро q, — его направляющим. Это завершает доказательство.

Теперь мы можем доказать теорему о приведении квадратичной функции Q к каноническому виду в подходящей аффинной системе координат $\{a_0, e_1, \ldots, e_n\}$, $\{e_i\}$ — базис L, $a_0 \in A$. Напомним, что точка $a \in A$ в ней представлена вектором $(x_1, \ldots, x_n)$, если $a = a_0 + \sum_{i=1}^{n} x_i e_i$.

5. Теорема. *Пусть Q — квадратичная функция на аффинном пространстве A. Тогда существует такая аффинная система координат в A, в которой Q принимает один из следующих видов.*

а) *Если q невырождена, то* $Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i x_i^2 + c$: λ_i, $c \in \mathcal{K}$.

б) *Если q вырождена ранга r, но центр Q непуст, то*

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{r} \lambda_i x_i^2 + c; \quad \lambda_i, c \in \mathcal{K}.$$

в) *Если q вырождена ранга r и центр Q пуст, то*

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{r} \lambda_i x_i^2 + x_{r+1}.$$

Доказательство. Если q невырождена, выберем в качестве a_0 центральную точку Q. Тогда $Q(a) = q(a - a_0) + c$. В качестве $e_1, \ldots, e_n$ выберем базис в L, в котором q приводится к сумме квадратов с коэффициентами. Тот же прием приводит к цели всегда, если центр непуст.

Если центр Q пуст, начнем с произвольной точки a_0 и базиса $\{e_1, \ldots, e_n\}$, в котором квадратичная часть Q имеет вид $\sum_{i=1}^{r} \lambda_i x_i^2$.

Пусть линейная часть имеет вид $l=\sum_{i=1}^{n}\mu_i x_i$. Мы утверждаем, что $\mu_j\neq 0$ для некоторого $j>r$. Действительно, иначе $l=\sum_{i=1}\mu_i x_i$, и тогда Q можно представить в виде

$$\sum_{i=1}^{r}\lambda_i x_i^2+\sum_{i=1}^{r}\mu_i x_i+c=\sum_{i=1}^{r}\lambda_i\left(x_i+\frac{\mu_i}{2\lambda_i}\right)^2+c'.$$

Следовательно, точка $a_0-\sum_{i=1}^{r}\frac{\mu_i}{2\lambda_i}e_i$ будет центральной для Q, что противоречит предположению о пустоте центра.

Но если $\mu_j>0$ для некоторого $j>r$, то система функционалов $\{e^1, \ldots, e^r, l\}$ в L^* линейно независима. Мы можем дополнить ее до базиса в L^* и в двойственном базисе L получить для Q выражение вида $\sum_{i=1}^{r}\lambda_i x_i^2+x_{r+1}+c$, где x_{r+1} как функция на L есть просто l. Теперь ясно, что имеется точка, в которой Q обращается в нуль, например, $x_1=\ldots=x_r=0$, $x_{r+1}=-c$, $x_{r+2}=\ldots=x_n=0$ в этой системе координат. Начав построение с этой точки, мы получим представление Q в виде $\sum_{i=1}^{r}\lambda_i x_i^2+x_{r+1}$.

6. Дополнения. а) Вопрос о единственности канонического вида сводится к уже решенной задаче о квадратичных формах. Если q невырождена и в некоторой системе координат имеет вид $\sum_{i=1}^{n}\lambda_i x_i^2+c$, то точка $(0, \ldots, 0)$ является центром и потому определена однозначно, константа c определена однозначно как значение Q в центре, а произвол в выборе осей и коэффициентов тот же, что для квадратичных форм. В частности, над $\mathbf{R}$ можно считать, что $\lambda_i=\pm 1$, и полным инвариантом является сигнатура. Над $\mathbf{C}$ можно считать, что все $\lambda_i=1$.

В вырожденном случае с непустым центром начало координат можно выбирать в центре как угодно, но константа c все равно определяется однозначно, ибо значение Q во всех точках центра постоянно: если a, a_0 лежат в центре, то $l(a-a_0)=0$ и $q(a-a_0)=0$, ибо $a-a_0$ лежит в ядре q. К квадратичной части применимы прежние замечания.

Наконец, в вырожденном случае с пустым центром начало координат можно брать в любой точке, где Q обращается в нуль; к квадратичной части применимы прежние замечания.

б) Если A — аффинное евклидово пространство, то Q приводится к каноническому виду в ортонормированном базисе. Числа $\lambda_1, \ldots, \lambda_n$ определены однозначно. Произвол в выборе центра тот же, что и в аффинном случае, произвол в выборе осей тот же, что для квадратичных форм в линейном евклидовом пространстве.

6. Аффинные квадрики. Аффинной квадрикой называется множество $\{a \in A \mid Q(a) = 0\}$, где Q — некоторая квадратичная функция на A. Взгляд на канонические формы Q показывает, что к проблеме исследования типов квадрик применимы все результаты § 10 ч. 2.

Рассмотрим вопрос о единственности функции Q, задающей данную аффинную квадрику над полем $\mathbf{R}$. Прежде всего, квадрика может быть аффинным подпространством в A (возможно, пустым): уравнение $\sum_{i=1}^{r} x_i^2 = 0$ равносильно системе уравнений $x_1 = \ldots$ $\ldots = x_r = 0$. При $r > 1$ имеется много непропорциональных друг другу квадратичных функций, задающих ту же квадрику, например $\sum_{i=1}^{r} \lambda_i x_i^2 = 0$ с любыми $\lambda_i > 0$. Покажем, что для остальных квадрик ответ проще:

7. Предложение. *Пусть аффинная квадрика X, не являющаяся аффинным подпространством, задается уравнениями $Q_1 = 0$ и $Q_2 = 0$, где Q_1, Q_2 — квадратичные функции. Тогда $Q_1 = \lambda Q_2$ для подходящего скаляра $\lambda \in \mathbf{R}$.*

Доказательство. Прежде всего, X не сводится к одной точке. В силу предложения п. 18 § 3 имеются две точки a_1, $a_2 \in X$, аффинная оболочка которых (прямая) не лежит в X целиком.

Пусть a_1, $a_2 \in X$ и прямая, проходящая через точки a_1, a_2, не лежит в X целиком. Введем в A систему координат $\{a_1; e_1, \ldots, e_n\}$, где $e_n = a_2 - a_1$. Запишем функцию Q_1 в этой системе координат

$$Q_1(x_1, \ldots, x_n) = \lambda x_n^2 + l_1'(x_1, \ldots, x_{n-1}) x_n + l_1''(x_1, \ldots, x_{n-1}),$$

где l_1', l_1'' — аффинно линейные функции, т. е. многочлены степени $\leqslant 1$ от $x_1, \ldots, x_{n-1}$. Так как прямая, проходящая через точки $a_1 = (0, \ldots, 0)$ и $a_2 = (0, \ldots, 0, 1)$, не содержится в X целиком, то $\lambda \neq 0$ и $l_1'(0)^2 - 4\lambda l_1''(0) > 0$. Разделив Q_1 на λ, можно считать, что $\lambda = 1$. Аналогично, можно считать, что
как квадратный трехчлен от x_n:

$$Q_2(x_1, \ldots, x_n) = x_n^2 + l_2'(x_1, \ldots, x_{n-1}) x_n + l_2''(x_1, \ldots, x_{n-1})$$

и $l_2'(0)^2 - 4l_2''(0) > 0$. Мы знаем теперь, что Q_1 и Q_2 имеют одинаковое множество вещественных корней, и хотим доказать, что $Q_1 = Q_2$.

Фиксируем вектор $(c_1, \ldots, c_{n-1}) \in \mathbf{R}^{n-1}$ и рассмотрим векторы $(tc_1, \ldots, tc_{n-1})$, $t \in \mathbf{R}$. При малых по модулю значениях t дискриминанты по x_n трехчленов $Q_1(tc_1, \ldots, tc_{n-1}, x_n)$ и $Q_2(tc_1, \ldots, tc_{n-1}, x_n)$ остаются положительными, и вещественные корни их, отвечающие точкам пересечения одной и той же прямой с X, совпадают. Значит, $l_1' = l_2'$ и $l_1'' = l_2''$ в таких точках $(tc_1, \ldots, tc_{n-1})$. Поэтому $l_1' \equiv l_2'$ и $l_1'' \equiv l_2''$, ибо аффинно линейные функции, совпадающие на открытом множестве, совпадают. Действительно, их разность обращается в нуль в окрестности начала координат и потому множество ее корней не может быть собственным линейным подпространством. Это завершает доказательство.

§ 6. Проективные пространства

1. Аффинные пространства получаются из линейных «забвением начала координат». Проективные пространства можно строить из линейных по меньшей мере двумя способами.

а) Добавить к аффинному пространству «бесконечно удаленные точки».

б) Реализовать проективное пространство как множество прямых в линейном пространстве.

Мы выберем в качестве основного определения б): оно яснее показывает однородность проективного пространства.

2. Определение. *Пусть L — линейное пространство над полем $\mathcal{K}$. Множество $P(L)$ прямых (т. е. одномерных линейных подпространств) в L называется проективным пространством, ассоциированным с L, а сами прямые в L называются точками $P(L)$.*

Число $\dim L-1$ называется размерностью $P(L)$ и обозначается $\dim P(L)$. Одномерные и двумерные проективные пространства называются соответственно проективной прямой или проективной плоскостью. Проективное пространство размерности n над полем $\mathcal{K}$ обозначается также $\mathcal{K}P^n$ или $P^n(\mathcal{K})$, или просто P^n. Смысл соглашения $\dim P(L)=\dim L-1$ станет сейчас ясен.

3. Однородные координаты. Выберем базис $\{e_0, \ldots, e_n\}$ в пространстве L. Каждая точка $p\in P(L)$ однозначно определяется любым ненулевым вектором на соответствующей прямой в L. Координаты $x_0, \ldots, x_n$ этого вектора называются *однородными координатами точки p*. Они определены с точностью до умножения на ненулевой скаляр: точка $(\lambda x_0, \ldots, \lambda x_n)$ лежит на той же прямой p и все точки прямой получаются таким образом. Поэтому вектор однородных координат точки p по традиции обозначается $(x_0: x_1: \ldots :x_n)$.

Таким образом, *координатное n-мерное проективное пространство $P(\mathcal{K}^{n+1})$* есть множество орбит мультипликативной группы $\mathcal{K}^*=\mathcal{K}\setminus\{0\}$, действующей на множестве ненулевых векторов $\mathcal{K}^{n+1}\setminus\{0\}$ по правилу $\lambda(x_0,\ldots,x_n)=(\lambda x_0,\ldots,\lambda x_n)$; $(x_0: x_1: \ldots :x_n)$ есть символ соответствующей орбиты.

Пользуясь однородными координатами, можно хорошо представить себе структуру P^n как множества несколькими разными способами.

а) *Аффинное покрытие P^n.* Положим

$$U_i=\{(x_0: \ldots :x_n)\mid x_i\neq 0\},\qquad i=0, \ldots, n.$$

Очевидно, $P^n=\bigcup\limits_{i=0}^{n} U_i$. В классе векторов проективных координат любой точки $p\in U_i$ имеется единственный вектор с i-й координатой, равной 1: $(x_0: \ldots :x_i: \ldots :x_n) = (x_0/x_i: \ldots :1: \ldots :x_n/x_i)$. Опуская эту единицу, получаем, что U_i биективно множеству $\mathcal{K}^n$, которое мы можем интерпретировать как n-мерное линейное или аффинное координатное пространство. Заметим, однако, что пока у нас нет никаких оснований считать, что на U_i имеется какая-то естественная не зависящая от выбора координат линейная или аффинная структура. Позже мы покажем, что инвариантно можно ввести на U_i лишь целый класс аффинных структур, связанных, впрочем, каноническими изоморфизмами, так что геометрия аффинных конфигураций в любой из них будет одна и та же.

Назовем множество $U_i \cong \mathcal{K}^n$ *i-й аффинной картой* P^n (в данной системе координат). Точки $\left(y_1^{(i)}, \ldots, y_n^{(i)}\right) \in U_i$ и $\left(y_1^{(j)}, \ldots, y_n^{(j)}\right) \in U_j$ при $i \neq j$ отвечают одной и той же точке P^n, лежащей на пересечении $U_i \cap U_j$, тогда и только тогда, когда, вставив 1 на i-е место в векторе $\left(y_1^{(i)}, \ldots, y_n^{(i)}\right)$ и на j-е место в $\left(y_1^{(j)}, \ldots, y_n^{(j)}\right)$, мы получим пропорциональные векторы.

В частности, $P^1 = U_0 \cup U_1$, $U_0 \cong U_1 \cong \mathcal{K}$; точка $y \in U_0$ отвечает точке $1/y \in U_1$ при $y \neq 0$; точка $y = 0$ из U_0 не лежит в U_1, а точка $1/y = 0$ из U_1 не лежит в U_0. Естественно считать, что P^1 получается из $U_0 \cong \mathcal{K}$ добавлением одной точки с координатой $y = \infty$. Обобщая эту конструкцию, получаем

б) *Клеточное разбиение* P^n. Положим

$$V_i = \{(x_0: \ldots :x_n) \mid x_j = 0 \text{ при } j < i, \ x_i \neq 0\}.$$

Очевидно, $V_0 = U_0$ и $P^n = \bigcup_{i=0}^{n} V_i$, но на этот раз все V_i попарно не пересекаются. В классе проективных координат любой точки $p \in V_i$ имеется единственный представитель с единицей на i-м месте; опуская эту единицу и предшествующие нули, мы получаем биекцию V_i с $\mathcal{K}^{n-i}$. Окончательно

$$P^n \cong \mathcal{K}^n \cup \mathcal{K}^{n-1} \cup \mathcal{K}^{n-2} \cup \ldots \cup \mathcal{K}^0 \cong \mathcal{K}^n \cup P^{n-1}.$$

Иными словами, P^n получается добавлением к $U_0 \cong \mathcal{K}^n$ бесконечно удаленного $(n-1)$-мерного проективного пространства, состоящего из точек $(0: x_1: \ldots :x_n)$; в свою очередь, оно получается из аффинного подпространства V_1 добавлением бесконечно удаленного (относительно V_1) проективного пространства P^{n-2} и т. д.

в) *Проективные пространства и сферы*. В случае $\mathcal{K} = \mathbf{R}$ или $\mathbf{C}$ есть удобный способ нормировки однородных координат в P^n, не требующий выбора ненулевой координаты и деления на нее. Именно, любую точку P^n можно представить координатами $(x_0: \ldots$ $\ldots: x_n)$ с условием $\sum_{i=0}^{n} |x_i|^2 = 1$, т. е. точкой на n-мерной (при $\mathcal{K} = \mathbf{R}$) или $(2n+1)$-мерной (при $\mathcal{K} = \mathbf{C}$) евклидовой сфере. Степень оставшейся неоднозначности такова: точка $(\lambda x_0: \ldots :\lambda x_n)$ по-прежнему лежит на единичной сфере тогда и только тогда, когда $|\lambda| = 1$, т. е. $\lambda = \pm 1$ при $\mathcal{K} = \mathbf{R}$, $\lambda = e^{i\varphi}$, $0 \leqslant \varphi < 2\pi$ при $\mathcal{K} = \mathbf{C}$.

Иными словами, n-мерное вещественное проективное пространство $\mathbf{R}P^n$ получается из n-мерной сферы S^n отождествлением пар ее диаметрально противоположных точек. В частности, $\mathbf{R}P^1$ устроена как окружность, а $\mathbf{R}P^2$ — как лист Мёбиуса, к которому по его границе приклеен круг (рис. 1, 2).

Сложнее «увидеть» $\mathbf{C}P^n$: в одну точку $\mathbf{C}P^n$ склеивается целый большой круг сферы S^{2n+1}, состоящий из точек $(x_0e^{i\varphi}, \ldots, x_ne^{i\varphi})$ с переменным φ. Из описания $\mathbf{C}P^1$ в случае б) в качестве $\mathbf{C}\cup\{\infty\}$ ясно, что $\mathbf{C}P^1$ можно представлять себе как двумерную сферу Римана, в которой ∞ представлена северным полюсом, как при сте-

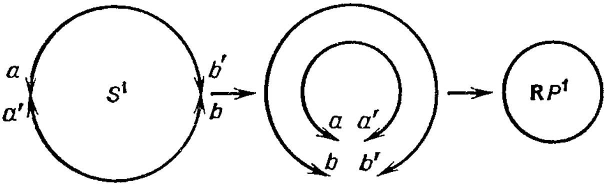


Рис. 1

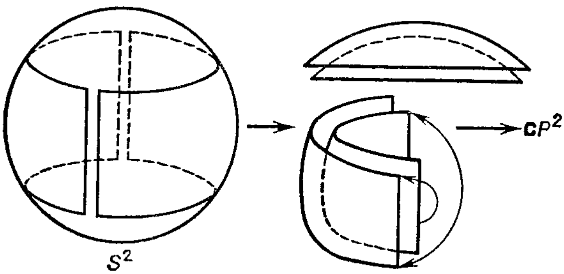


Рис. 2

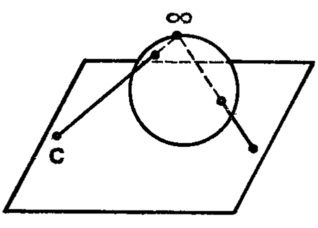


Рис. 3

реографической проекции (рис. 3). Поэтому наше новое представление $\mathbf{C}P^1$ в виде факторпространства S^3 дает замечательное отображение $S^3 \to S^2$, слои которого являются окружностями S^1. Оно называется *отображением Хопфа*.

В описании этого пункта мы совсем забыли о линейной структуре, исходной для $\mathbf{R}P^n$ и $\mathbf{C}P^n$, зато нам стали ясно видны топологические свойства этих пространств, в первую очередь их компактность. (Строго говоря, в определении P^n никакая топология не фигурировала; удобнее всего вводить ее именно с помощью отображений сфер, условившись, что открытые множества в $\mathbf{R}P^n$ и $\mathbf{C}P^n$ это те, прообразы которых в S^n и S^{2n+1} открыты.) Впредь мы не будем пользоваться топологией и вернемся к изучению линейной геометрии проективных пространств. Не будет, однако, преувеличением сказать, что важность $\mathbf{R}P^n$ и $\mathbf{C}P^n$ в значительной мере объясняется тем, что это естественные компактификации $\mathbf{R}^n$ и $\mathbf{C}^n$, позволяющие распространить основные черты линейной структуры на бесконечность. Даже над абстрактным полем $\mathscr{K}$, не несущим никакой топологии, эта «компактность» проективных пространств появляется в массе алгебраических вариантов. Типичный пример: на аффинной плоскости две разные прямые, вообще говоря, пересекаются в одной точке, но могут быть и параллельны. Это означает, что точка их пересечения «ушла в бесконечность», и при переходе в проективную плоскость она благополучно обнаруживается: любые две проективные прямые на плоскости пересекаются.

Вернемся теперь к систематическому изучению геометрии P^n

4. Проективные подпространства. Пусть $M \subset L$ — любое линейное подпространство в L. Тогда $P(M) \subset P(L)$, ибо каждая прямая лежащая в M, является в то же время прямой, лежащей в L.

Множества вида $P(M)$ называются *проективными подпространствами в* $P(L)$. Очевидно, $P(M_1 \cap M_2) = P(M_1) \cap P(M_2)$, и то же верно для пересечения любого семейства. Следовательно, семейство проективных подпространств замкнуто относительно пересечений. Поэтому в множестве проективных подпространств $P(L)$, содержащих данное множество $S \subset P(L)$, имеется наименьшее — пересечение всех таких подпространств. Оно называется *проективной оболочкой* S, обозначается $\bar{S}$ и совпадает с $P(M)$, где M — линейная оболочка всех прямых, отвечающих точкам $s \in S$, в L

При переходе от пар $L \subset M$ к парам $P(L) \subset P(M)$ размерности уменьшаются на единицу, так что *коразмерность* $\dim L - \dim M$ *совпадает* с коразмерностью $\dim P(L) - \dim P(M)$. Далее, как мы уже отмечали, $P(M_1 \cap M_2) = P(M_1) \cap P(M_2)$, а $P(M_1 + M_2)$ совпадает с проективной оболочкой $\overline{P(M_1) \cup P(M_2)}$.

Пользуясь этими замечаниями, мы можем написать проективный вариант теоремы п. 3 § 5 ч. 1. Заметим лишь, что в соответствии с определением в п. 2 размерность пустого проективного пространства следует считать равной -1: этот случай вполне реален, ибо непустые подпространства могут иметь пустое пересечение.

5. Теорема. *Пусть P_1, P_2 — два конечномерных проективных подпространства в проективном пространстве P. Тогда*

$$\dim P_1 \cap P_2 + \dim \overline{P_1 \cup P_2} = \dim P_1 + \dim P_2.$$

6. Примеры. а) P_1, P_2 — две разные точки. Тогда $\dim P_1 \cap P_2 = -1$, $\dim P_1 = \dim P_2 = 0$, откуда $\dim \overline{P_1 \cup P_2} = 1$, т. е. проективной оболочкой двух точек является прямая. Согласно определению проективной оболочки, она является единственной проективной прямой, проходящей через две точки.

б) Допустим, что $\dim P_1 + \dim P_2 \geqslant \dim P$. Тогда, поскольку $\dim \overline{P_1 \cup P_2} \leqslant \dim P$, имеем $\dim P_1 \cap P_2 \leqslant 0$. Иными словами, два проективных подпространства, сумма размерностей которых больше или равна размерности объемлющего пространства, имеют непустое пересечение. В частности, в проективной плоскости нет «параллельных» прямых: любые две прямые пересекаются либо в одной точке, либо в двух и тогда (в силу примера а)) совпадают. Аналогично, две проективных плоскости в трехмерном проективном пространстве обязательно пересекаются по прямой или совпадают. Проективная плоскость и прямая в трехмерном пространстве пересекаются по точке или прямая лежит в плоскости.

в) Условие $P_1 \cap P_2 = \varnothing$ в случае $P_i = P(M_i)$ означает, что $M_1 \cap M_2 = \{0\}$, т. е. что сумма $M_1 + M_2$ *прямая.*

7. Задание проективных подпространств уравнениями. Линейная функция $f\colon L \to \mathscr{K}$ на линейном пространстве L *не определяет* никакую функцию на $P(L)$ (кроме случая $f \equiv 0$), ибо всегда есть прямая в L, на которой эта функция непостоянна, и нет возможности фиксировать ее значение в соответствующей точке $P(L)$. Но *уравнение* $f = 0$ определяет линейное подпространство в L и потому проективное подпространство в $P(L)$. Если L конечномерно, то любое подпространство в L и потому любое подпространство в $P(L)$ можно задать системой уравнений

$$f_1 = \ldots = f_m = 0.$$

В однородных координатах P^n этот эффект проявляется так: система линейных однородных уравнений

$$\sum_{j=0}^{n} a_{ij} x_j = 0, \qquad i = 1, \ldots, m,$$

задает проективное подпространство в P^n, состоящее из точек, однородные координаты которых $(x_0\colon \ldots\colon x_n)$ удовлетворяют этой системе. Умножение всех координат на λ не нарушает обращения в нуль левых частей.

8. Аффинные подпространства и гиперплоскости. Пусть $M \subset L$ — линейное подпространство коразмерности единица. Тогда $P(M) \subset \subset P(L)$ имеет коразмерность единица, и мы будем называть такие подпространства *гиперплоскостями.*

Мы покажем сейчас, как ввести на дополнении A_M к гиперплоскости $P(M)$ структуру аффинного пространства $(A_M, M, +)$. Выберем в L линейное многообразие $M' = m' + M$, не проходящее через начало координат. Оно имеет естественную аффинную структуру: сдвиг на $m \in M$ в M' индуцирован сдвигом на m в L, т. е. состоит в прибавлении m.

С другой стороны, A_M и M' находятся в биективном соответствии: точка A_M есть прямая, не лежащая в M, и она пересекается с M' в единственной точке, которую и поставим в соответствие исходной точке A_M. Так получаются все точки по одному разу. С помощью этого биективного соответствия аффинную структуру на M' можно перенести на A_M. Однако выбор M' не однозначен, и это приводит к неоднозначности аффинной структуры A_M. Чтобы сравнить две такие структуры, покажем, что тождественное теоретико-множественное отображение A_M в себя является аффинным изоморфизмом этих двух структур.

9. Предложение. *Пусть $(A_M, M, +')$ и $(A_M, M, +'')$ — две аффинные структуры на A_M, построенные с помощью описанной конструкции. Тогда тождественное отображение A_M в себя явля-*

ется аффинным изоморфизмом, линейная часть которого есть некоторая гомотетия M.

Доказательство. Пусть две структуры отвечают подмногообразиям $m'+M$ и $m''+M$. Классы $m'+M$ и $m''+M$ в одномерном факторпространстве L/M пропорциональны. Поэтому можно считать, что $m''=am'$, $a=\mathcal{K}$. Умножение на a в L переводит $m'+M$ в $m''+M$ и индуцирует тождественное отображение $P(L)$ в себя и потому A_M в себя. С другой стороны, сдвиг на вектор $m\in M$ в $m'+M$ при гомотетии переходит в сдвиг на вектор $am\in M$ в $m''+M$. Это и доказывает требуемое.

10. Следствие. *Множество аффинных подпространств в A_M с их отношениями инцидентности, а также множества аффинных отображений A_M в другие аффинные пространства не зависят от произвола в выборе аффинной структуры A_M.*

Это оправдывает возможность рассматривать дополнение к любой гиперплоскости в проективном пространстве просто как аффинное пространство без дальнейших уточнений.

Посмотрим теперь, как выглядит проективное пространство $P(M)$ «с точки зрения» аффинного пространства A_M.

11. Предложение. *Точки $P(M)$ находятся в биективном соответствии с классами параллельных прямых в A_M. Иными словами, каждая точка $P(M)$ есть «направление ухода на бесконечность» в A_M.*

Доказательство. Отождествим A_M с $m'+M$. Класс параллельных прямых в $m'+M$ определяется своей направляющей в M, т. е. точкой в $P(M)$, и это соответствие биективно.

12. На самом деле можно сказать больше: каждая прямая l в A_M однозначно определяет содержащую ее прямую в $P(L)$ — а именно, ее проективную оболочку $\bar{l}$. Проективная оболочка получается добавлением к l единственной точки, которая как раз лежит в $P(M)$ и является «бесконечно удаленной точкой» этой прямой. Весь класс параллельных прямых в A_M имеет общую бесконечно удаленную точку в $P(M)$. При отождествлении A_M с $m'+M$ оболочка $\bar{l}$ отвечает всем прямым плоскости в L, проходящей через l и направляющую l, а бесконечно удаленная точка $\bar{l}$ — это сама направляющая.

Более общо, пусть $A\subset A_M$ — любое аффинное подпространство. Тогда его проективная оболочка $\bar{A}$ в $P(L)$ обладает следующими свойствами:

а) $\bar{A}\setminus A\subset P(M)$: добавляются лишь точки на бесконечности.

б) $\dim A=\dim\bar{A}$.

в) $\bar{A}\setminus A$ есть проективное подпространство в $P(M)$ размерности $\dim A-1$. (Поэтому $\bar{A}$ называют также проективным замыканием A.)

Отождествление A_M с $m'+M$ сводит проверку этих свойств к прямому применению определений. Действительно, $\bar{A}$ состоит из прямых, лежащих в линейной оболочке $A\subset m'+M$. Эта линейная оболочка натянута на направляющую L_0 подпространства A и любой вектор из A. Поэтому ее размерность равна $\dim L_0+1=$

$= \dim A + 1$, значит, $\dim A = \dim \bar{A}$. Все прямые в этой линейной оболочке пересекаются с $m' + M$, т. е. отвечают точкам A, за исключением прямых, лежащих в направляющей L_0. Последние лежат в $P(M)$ и образуют проективное пространство размерности $\dim L_0 - 1 = \dim A - 1$.

§ 7. Проективная двойственность и проективные квадрики

1. Пусть L — линейное пространство над полем $\mathscr{K}$, L^* — двойственное к нему пространство линейных функционалов на L. *Проективное пространство $P(L^*)$ называется двойственным к проективному пространству $P(L)$.*

Каждая точка $P(L^*)$ есть прямая $\{\lambda f\}$ в пространстве линейных функционалов на L. Гиперплоскость $f = 0$ в $P(L)$ не зависит от выбора функционала f на этой прямой и однозначно определяет всю прямую. Поэтому можно сказать, что *точками двойственного проективного пространства являются гиперплоскости исходного проективного пространства.*

Если в L и L^* выбраны двойственные базисы и соответствующие системы однородных координат в $P(L)$ и $P(L^*)$, это соответствие приобретает простой вид: гиперплоскости с уравнением

$$\sum_{i=0}^{n} a_i x_i = 0$$

в $P(L)$ отвечает точка с однородными координатами $(a_0 : \ldots : a_n)$ в $P(L^*)$. Канонический изоморфизм $L \to L^{**}$ показывает симметрию отношения двойственности между двумя проективными пространствами. Более общо, переводя результаты § 7 ч. 1 на проективный язык, мы получим следующее соответствие двойственности между системами проективных подпространств в $P(L)$ и $P(L^*)$ (мы считаем дальше, что L конечномерно).

а) Подпространству $P(M) \subset P(L)$ отвечает двойственное к нему подпространство $P(M^\perp) \subset P(L^*)$. При этом

$$\dim P(M) + \dim P(M^\perp) = \dim P(L) - 1.$$

б) Пересечению проективных подпространств отвечает проективная оболочка двойственных к ним, а проективной оболочке — пересечение. В частности, отношение инцидентности двух подпространств (т. е. включение одного в другое) переходит в отношение инцидентности.

Это позволяет сформулировать следующий *принцип проективной двойственности*, являющийся, собственно говоря, *метаматематическим*, поскольку он представляет собой утверждение *о языке* проективной геометрии.

2. Принцип проективной двойственности. *Предположим, что мы доказали теорему о конфигурациях проективных подпространств в проективных пространствах, в формулировке которой фигурируют лишь свойства размерности, инцидентности, пересечения и взятия проективной оболочки. Тогда двойственное утверждение, в котором* *все термины заменены на двойственные к ним по правилам предыдущего пункта, также является теоремой проективной геометрии.*

Простой пример: к теореме «две разные плоскости в трехмерном проективном пространстве пересекаются по одной прямой» двойственна теорема «через две разные точки в трехмерном проективном пространстве проходит одна прямая». (В § 9 мы познакомимся с гораздо более содержательными теоремами о проективных конфигурациях.)

3. Проективная двойственность и квадрики. Если линейное пространство L снабжено изоморфизмом $L \to L^*$, то $P(L^*)$ можно отождествить с $P(L)$, и отображение двойственности между проективными подпространствами $P(L)$ и $P(L^*)$ превратится в *отображение двойственности между подпространствами в* $P(L)$.

Задание изоморфизма $L \to L^*$ равносильно заданию невырожденного скалярного произведения $g\colon L \times L \to \mathscr{K}$. Рассмотрим подробнее геометрию проективной двойственности, отвечающую случаю, когда скалярное произведение g *симметрично*. Как обычно, будем считать, что характеристика поля $\mathscr{K}$ отлична от двух. Тогда g однозначно восстанавливается по квадратичной форме $q(l) = g(l, l)$.

Уравнение $q(l) = 0$ определяет квадрику Q_0 в L. Ее образ в $P(L)$ мы также будем называть квадрикой, а в применении к теории двойственности — *полярной квадрикой*. Заметим, что Q_0 есть конус с центром в начале координат: если $l \in Q_0$, то вся прямая $\mathscr{K}l$ лежит в Q_0. Отождествляя $P(L)$ с бесконечно удаленными точками L, мы можем отождествить Q с базой конуса Q_0.

Согласно общей теории g и q определяют отображение двойственности множества проективных подпространств $P(L)$ в себя; гиперплоскость в $P(L)$, двойственная точке $p \in P(L)$, называется *полярной к* p (относительно q или Q). Чтобы разобраться в геометрическом устройстве этого отображения, выведем сначала уравнение полярной гиперплоскости в однородных координатах. Мы можем работать сначала в L. Пусть уравнение Q_0 имеет вид

$$q(x_0, \ldots, x_n) \equiv \sum_{i,j=0}^{n} a_{ij}x_ix_j = 0, \qquad a_{ij} = a_{ji}.$$

Точке $(x_0^0, \ldots, x_n^0)$ в L при изоморфизме $L \to L^*$, связанном с q, отвечает линейная функция $\sum_{i,j=0}^{n} a_{ij}x_i^0x_j$ от $(x_0, \ldots, x_n) \in L$. Поэтому уравнение полярной гиперплоскости имеет вид

$$\sum_{i,j=0}^{n} a_{ij}x_i^0x_j = 0.$$

В частности, *если* $(x_0^0 : \ldots : x_n^0) \in Q$, *то полярная гиперплоскость к данной точке содержит эту точку.* Более того, в этом случае ее уравнение можно переписать в виде

$$\sum_{j=1}^{n} \frac{\partial q}{\partial x_j}(x_0^0, \ldots, x_n^0)(x_j - x_j^0) = \sum_{i,\,j=0}^{n} a_{ij} x_i^0 (x_j - x_j^0) = 0.$$

В элементарной аналитической геометрии (над $\mathbf{R}$) такое уравнение определяет *касательную гиперплоскость* к Q_0 в ее точке $(x_0^0, \ldots, x_n^0)$. Это мотивирует общее определение:

4. Определение. *Касательной гиперплоскостью к невырожденной квадрике $Q \subset P(L)$ в точке $p \in Q$ называется гиперплоскость, полярная к p относительно квадратичной формы q, задающей Q.*

Пользуясь общими свойствами проективной двойственности, мы можем теперь немедленно восстановить геометрически значительную часть отображения двойственности и получить серию красивых и неочевидных геометрических теорем, образцы которых мы приведем. Ниже Q — (невырожденная) квадрика в P^2 или P^3.

а) Пусть $Q \subset P^2$, p_1, p_2 — две точки на квадрике, p_3 — точка пересечения касательных к Q в p_1 и p_2. Согласно общему принципу двойственности точка p_3 отвечает тогда прямой $\overline{p_1 p_2}$, проходящей через p_1 и p_2, т. е. проективной оболочке p_1 и p_2.

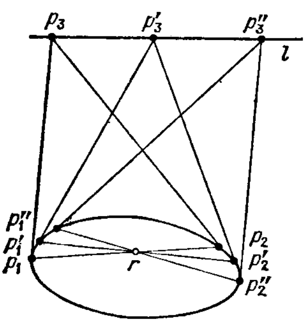


Рис. 4

Заставим точку p_3 меняться вдоль прямой l; проведем из каждой точки прямой две касательные к Q и соединим пары точек касания. Тогда все получающиеся «хорды» Q пересекутся в одной точке r, которая отвечает l в силу двойственности. Еще раз заметим, что для доказательства мы не нуждаемся ни в каких вычислениях: это следует просто из того, что по общему принципу двойственности проективная оболочка точек p_3, p_3', p_3'', ... полярна к пересечению двойственных к ним прямых, которые и суть соответствующие хорды.

Один момент, однако, заслуживает специального упоминания. Попарные пересечения касательных к точкам Q могут не заметать всю плоскость. Например для эллипса в $\mathbf{R}P^2$, как на рис. 4 (у нас нарисован, конечно, лишь кусочек аффинной карты в $\mathbf{R}P^2$), мы получим лишь *внешность* эллипса. Как же узнать, какие прямые отвечают внутренним точкам эллипса? Рис. 4 подсказывает ответ: в силу симметрии двойственности следует провести через внутреннюю точку r пучок хорд к Q, затем построить точки пересечения касательных к Q в противоположных концах этих хорд; они и заметут двойственную к точке r прямую l.

Однако, таким образом, описание отображения двойственности становится неоднородным. У нас оказываются два рецепта для построения прямой l, полярной к точке r.

1) Если точка r лежит вне эллипса Q (или на нем), проведите две касательные из r к Q (или одну) и соедините точки касания прямой l (или возьмите касательную l).

2) Если точка r лежит внутри эллипса Q, проведите все прямые через r, постройте точки пересечения касательных к двум точкам пересечения прямых через r с Q. Их геометрическое место и будет прямой, двойственной к r.

Оказывается, все дело в том, что основное поле $\mathbf{R}$ здесь не является алгебраически замкнутым. Если бы мы работали в $\mathbf{C}P^2$, *годились бы оба рецепта,* и притом *для всех* точек $r \in \mathbf{C}P^2$. Вещественная прямая l, лежащая целиком вне вещественного эллипса Q, на самом деле все равно *пересекается* с ним, *но в двух комплексно сопряженных точках,* и две комплексно сопряженные касательные к Q в этих точках пересекаются уже в *вещественной* точке r, лежащей внутри Q. Из вещественной точки r, лежащей внутри Q, все равно можно провести *две комплексно сопряженные касательные* к Q, через точки касания которых проходит вещественная прямая — это и есть l.

В этом смысле вещественная проективная геометрия $\mathbf{R}P^2$ является лишь кусочком геометрии $\mathbf{C}P^2$, и по-настоящему простая и симметричная теория двойственности имеет место в $\mathbf{C}P^2$, а $\mathbf{R}P^2$ отражает лишь ее вещественную часть.

Классическая проективная геометрия была в значительной мере посвящена выяснению деталей этого красивого мира конфигураций, состоящих из квадрик, хорд и касательных и «невидимых» комплексных точек касания и пересечения. На самом деле вся квадрика может не иметь вещественных точек, как например $x_0^2+x_1^2+x_2^2=0$. Тем не менее видимая часть двойственности разыгрывается на $\mathbf{R}P^2$.

б) Дадим еще одну иллюстрацию в трехмерном случае: рассмотрим проективную невырожденную квадрику Q в трехмерном проективном пространстве и проведем из точки r вне Q касательные плоскости к Q. Тогда все точки касания лежат в одной плоскости, а именно в плоскости, двойственной к r. Причина снова та же: пересечению касательных плоскостей двойственна проективная оболочка точек касания, и если все касательные плоскости пересекаются в точности по r (это нужно и можно доказать в случае, когда Q имеет достаточно много точек), то эта проективная оболочка должна быть двумерна.

Комментарии по поводу комплексных точек касания и пересечения те же, что и в двумерном случае.

Строгое определение вместилища недостающих точек проективного пространства и квадрики в случае $\mathscr{K}=\mathbf{R}$ опирается на понятие комплексификации (см. § 12 ч. 1).

5. а) Комплексификацией проективного пространства $P(L)$ над $\mathbf{R}$ называется проективное пространство $P(L^{\mathbf{C}})$ над $\mathbf{C}$. Каноническое вложение $L \subset L^{\mathbf{C}}$ позволяет сопоставить каждой $\mathbf{R}$-прямой в L ее комплексификацию — $\mathbf{C}$-прямую в $L^{\mathbf{C}}$, что определяет

вложение $P(L) \subset P(L^{\mathbf{C}})$. Точки $P(L^{\mathbf{C}})$ суть «комплексные точки» вещественного проективного пространства $P(L)$.

б) Изоморфизм $L \to L^*$, определяемый скалярным произведением g на L, индуцирует комплексифицированный изоморфизм $L^{\mathbf{C}} \to (L^{\mathbf{C}})^*$. Он определяется симметричным скалярным произведением $g^{\mathbf{C}}$ на $L^{\mathbf{C}}$, которое задает нам квадрику $Q^{\mathbf{C}}$ и проективную двойственность в $P(L^{\mathbf{C}})$. На $L^{\mathbf{C}}$ и $P(L^{\mathbf{C}})$ действует операция комплексного сопряжения, индуцированная антилинейным изоморфизмом $L^{\mathbf{C}} \to \overline{L}^{\mathbf{C}}$, тождественным на $L \subset L^{\mathbf{C}}$. Точки $P(L)$ — это точки $P(L^{\mathbf{C}})$, инвариантные относительно комплексного сопряжения; они называются вещественными. Более общо, проективные подпространства в $P(L^{\mathbf{C}})$, переводящиеся в себя при комплексном сопряжении, находятся в биекции с проективными подпространствами в $P(L)$. Назовем такие подпространства вещественными. Тогда два отображения, устанавливающие взаимнообратные биекции, можно описать так:

(вещественное проективное подпространство в $P(L^{\mathbf{C}})$) $\to$ (множество его вещественных точек в $P(L)$);

(проективное подпространство в $P(L)$) $\to$ (его комплексификация в $P(L^{\mathbf{C}})$).

в) Отображение двойственности в $P(L)$, определенное с помощью g, получается из отображения двойственности в $P(L^{\mathbf{C}})$, определенного с помощью $g^{\mathbf{C}}$, посредством ограничения последнего на систему вещественных подпространств в $P(L^{\mathbf{C}})$, отождествленную с системой подпространств в $P(L)$, как в разделе б).

Проверки всех этих утверждений, если учесть результаты § 12 ч. 1, проводятся непосредственно, а в вещественной системе координат $L^{\mathbf{C}}$, пришедшей из L, совсем тавтологичны. Единственная содержательная сторона ситуации, проиллюстрированная выше на примерах, состоит в возможности проявления вещественных точек на невидимых комплексных конфигурациях вроде лежащей внутри эллипса точки пересечения двух невещественных касательных к двум комплексно сопряженным точкам этого эллипса.

В случае основного поля $\mathscr{K}$, отличного от $\mathbf{R}$, нужно воспользоваться общим функтором расширения основного поля (например, до алгебраического замыкания $\mathscr{K}$) вместо комплексификации. Ситуация, однако, несколько усложняется тем, что вместо одного отображения комплексного сопряжения придется привлекать всю группу Галуа для выделения объектов, определенных над исходным полем (вещественных в случае $\mathscr{K} = \mathbf{R}$).

§ 8. Проективные группы и проекции

1. Пусть L, M — два линейных пространства, $f\colon L \to M$ — линейное отображение. Если $\operatorname{Ker} f = \{0\}$, то f переводит любую прямую из L в однозначно определенную прямую в M и, значит, индуцирует отображение $P(f)\colon P(L) \to P(M)$, называемое проективизацией f. В частности, если f — изоморфизм, $P(f)$ называется проективным изоморфизмом. При $\operatorname{Ker} f \neq \{0\}$ положение дел сложнее: прямые, лежащие в $\operatorname{Ker} f$, т. е. составляющие проективное подпространство $P(\operatorname{Ker} f) \subset P(L)$, переходят в нуль, который не определяет никакой точки в $P(M)$. Поэтому проективизация $P(f)$ определена лишь на дополнении $U_f = P(L) \setminus P(\operatorname{Ker} f)$. Оба этих случая важны, но ведут в разных направлениях, и мы исследуем их отдельно. Наиболее существенные геометрические черты ситуации выявляются уже при $L = M$.

2. Проективная группа. Пусть $M = L$, f пробегает группу линейных автоморфизмов пространства L. Следующие утверждения очевидны:

а) $P(\mathrm{id}_L) = \mathrm{id}_{P(L)}$;

б) $P(fg) = P(f)P(g)$.

В частности, все отображения $P(f)$ биективны и $P(f^{-1}) = P(f)^{-1}$. Поэтому $P(f)$ пробегает группу отображений $P(L)$ в себя, которая называется *проективной группой* пространства $P(L)$ и обозначается $\mathrm{PGL}(L)$, отображение $P\colon \mathrm{GL}(L) \to \mathrm{PGL}(L)$, $f \mapsto P(f)$ является сюръективным гомоморфизмом групп.

Каждое отображение $P(f)$ переводит проективные подпространства $P(L)$ в проективные подпространства, сохраняя размерность и все отношения инцидентности.

Вместо $\mathrm{PGL}(\mathcal{K}^{n+1})$ пишут $\mathrm{PGL}(n)$.

3. Предложение. *Ядро канонического отображения* $P\colon \mathrm{GL}(L) \to \mathrm{PGL}(L)$ *состоит в точности из гомотетий. Поэтому* $\mathrm{PGL}(L)$ *изоморфна факторгруппе* $\mathrm{GL}(L)/\mathcal{K}^*$, *где* $\mathcal{K}^* = \{a\,\mathrm{id}_L \mid a \in \mathcal{K} \setminus \{0\}\}$.

Доказательство. По определению $\operatorname{Ker} P = \{f \in \mathrm{GL}(L) \mid P(f) = \mathrm{id}_{P(L)}\}$. Любая гомотетия переводит каждую прямую из L в себя, поэтому $\mathcal{K}^* \subset \operatorname{Ker} P$. Наоборот, всякий элемент $\operatorname{Ker} P$ переводит любую прямую в себя и потому диагонализируем в любом базисе L. Но тогда все его собственные значения должны совпадать. В самом деле, пусть $f(e_1) = \lambda_1 e_1$, $f(e_2) = \lambda_2 e_2$, где e_1, e_2 линейно независимы. Тогда из условия $f(e_1 + e_2) = \mu(e_1 + e_2) = \lambda_1 e_1 + \lambda_2 e_2$ следует, что $\lambda_1 = \mu = \lambda_2$. Значит, f — гомотетия, что доказывает требуемое.

4. Отображения $P(f)$ в координатах. Если линейное отображение $f\colon L \to L$ в координатах задается матрицей A:

$$f(x_0, \ldots, x_n) = A \cdot [x_0, \ldots, x_n]$$

(произведение матрицы A на столбец $[x_0, \ldots, x_n]$), то $P(f)$ в соответствующих однородных координатах задается той же матрицей A или любой пропорциональной ей:

$$P(f)(x_0 : \ldots : x_n) \subset (\lambda A) \cdot [x_0, \ldots, x_n], \qquad \lambda \in \mathcal{K}^*.$$

Если ограничиться рассмотрением точек с $x_0 \neq 0$, проективные координаты которых можно выбирать в виде $(1 : y_1 : \ldots : y_n)$, и так же записывать координаты образа точки, мы придем к дробнолинейным формулам:

$$P(f)(1:y_1:\ldots:y_n)=(1:y_1':\ldots:y_n')=$$

$$=\left(a_{00}+\sum_{i=1}^{n}a_{i0}y_i : a_{01}+\sum_{i=1}^{n}a_{i1}y_i:\ldots:a_{0n}+\sum_{i=1}^{n}a_{in}y_i\right)=$$

$$=\left(1:\frac{a_{01}+\sum\limits_{i=1}^{n}a_{i1}y_i}{a_{00}+\sum\limits_{i=1}^{n}a_{i0}y_i}:\ldots:\frac{a_{0n}+\sum\limits_{i=1}^{n}a_{in}y_i}{a_{00}+\sum\limits_{i=1}^{n}a_{i0}y_i}\right).$$

(Аналогичный вид, разумеется, имеет $P(f)$ на множестве точек, где $x_i \neq 0$, i любое.) Эти выражения теряют смысл там, где знаменатель обращается в нуль, т. е. в тех точках дополнения к гиперплоскости $x_0=0$, которые $P(f)$ переводит в эту гиперплоскость. Если таких точек нет, то в терминах аффинных координат $(y_1, \ldots, y_n)$ на $P(L)\setminus\{x_0=0\}$ мы получаем аффинное отображение. Инвариантное объяснение этого дает следующий результат.

5. **Предложение.** *Пусть $M\subset L$ — подпространство коразмерности единица, $P(M)\subset P(L)$ — соответствующая гиперплоскость, A_M — дополнение к ней с аффинной структурой, описанной в* § 6. *Поставим в соответствие любому проективному автоморфизму $P(f)$: $P(L)\to P(L)$ с условием $f(M)\subset M$ его ограничение на A_M. Получим изоморфизм подгруппы* $\mathrm{PGL}(L)$, *переводящей $P(M)$ в себя, с* $\mathrm{Aff}\,A_M$. *Линейная часть ограничения $P(f)$ на A_M пропорциональна ограничению f на M.*

Доказательство. Введем аффинную структуру на A_M, отождествив A_M с линейным многообразием $m'+M\subset L$: каждой точке A_M ставится в соответствие пересечение соответствующей прямой в L с $m'+M$. Если $f(M)\subset M$, то в классе f с одним и тем же $P(f)$ можно выбрать единственное отображение f_0, для которого $f_0(m'+M)=m'+M$. Ограничения всех таких отображений f_0 образуют группу аффинных преобразований $m'+M$, поскольку $m'+M$ есть аффинное подпространство в L с его аффинной структурой, а f_0: $L\to L$ линейно и потому аффинно. Линейная часть такого f_0, очевидно, совпадает с ограничением f_0 на M. Для всякой линейной части можно найти соответствующее f_0, и при фиксированной линейной части можно найти f_0, переводящее любую точку $m'+M$ в любую другую: чтобы увидеть это, достаточно выбрать базис в L, состоящий из базиса в M и вектора m', после чего воспользоваться формулами п. 3. Наконец, если f тождественно действует на $m'+M$ и M, то $P(f)=\mathrm{id}_{P(L)}$, ибо f переводит каждую прямую в L в себя. Это завершает доказательство.

6. **Действие проективной группы на проективных конфигурациях.** Назовем проективной конфигурацией конечную упорядоченную систему проективных подпространств в $P(L)$. Будем говорить, что две конфигурации проективно конгруэнтны тогда и только тогда, когда одну можно перевести в другую проективным преоб-

разованием $P(L)$ в себя. Очевидно, для этого необходимо и достаточно, чтобы соответствующие конфигурации линейных подпространств в L были одинаково расположены в смысле § 5 ч. 1. Поэтому мы можем сразу же перевести доказанные там результаты на проективный язык и получить следующие факты.

а) Группа $PGL(L)$ транзитивно действует на множестве проективных подпространств фиксированной размерности в $P(L)$, т. е. все такие подпространства конгруэнтны (см. п. 1 § 5 ч. 1).

б) Группа $PGL(L)$ транзитивно действует на множестве упорядоченных пар проективных подпространств в $P(L)$ с фиксированными размерностями членов пары и их пересечений, т. е. все такие пары конгруэнтны (см. п. 5 § 5 ч. 1).

в) Группа $PGL(L)$ транзитивно действует на множестве упорядоченных n-ок проективных подпространств $(P_1, \ldots, P_n)$ с фиксированными размерностями $\dim P_i$, которые обладают следующим свойством: для каждого i подпространство P_i не пересекается с проективной оболочкой $(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$, т. е. наименьшим проективным подпространством, содержащим эту систему.

Действительно, пусть $P_i = P(L_i)$, $L_i \subset L$. Проективная оболочка $(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$, как нетрудно убедиться, совпадает с $P(L_1 + \ldots + L_{i-1} + L_{i+1} + \ldots + L_n)$, а условие пустоты ее пересечения с $P(L_i)$ означает, что $L_i \cap \sum\limits_{j \neq i}^{n} L_j = \{0\}$. В силу условия а) теоремы п. 8 § 5 ч. 1 сумма $L_1 \oplus \ldots \oplus L_n$ прямая, и $GL(L)$ транзитивно действует на таких n-ках подпространств (выбрать базис в L, дополняющий объединение базисов всех L_i, и воспользоваться тем, что $GL(L)$ транзитивна на базисах L).

В качестве частного случая ($\dim P_i = 0$ для всех i) получаем следующий результат: все наборы n точек в $P(L)$, обладающие тем свойством, что никакая точка не лежит в проективной оболочке остальных, проективно конгруэнтны.

г) Группа $PGL(L)$ транзитивно действует на множестве проективных флагов $P_1 \subset P_2 \subset \ldots \subset P_n$ в $P(L)$ фиксированной длины n с фиксированными размерностями $\dim P_i$.

Действительно, любой такой флаг является образом флага $L_1 \subset \subset L_2 \subset \ldots$ в L; выберем базис в L, первые $\dim P_i + 1$ элементов которого порождают подпространство L_i для каждого i, и снова воспользуемся транзитивностью действия $GL(L)$ на базисах.

Кроме этих результатов, являющихся прямым следствием соответствующих теорем для линейных пространств, разберем один интересный новый случай, в котором впервые появляется нетривиальный инвариант относительно проективной конгруэнтности: классическое «двойное отношение четверки точек на проективной прямой». Бóльшую часть рассуждений можно провести в случае произвольной размерности, и мы начнем с общего определения.

7. Определение. *Система точек $p_1, \ldots, p_N$ в n-мерном проективном пространстве P находится в общем положении, если для всех $m \leqslant \min\{N, n+1\}$ и всех подмножеств $S \subset \{1, \ldots, N\}$ мощности*

m проективная оболочка точек $\{p_i | i \in S\}$ имеет размерность $m-1$.

Нас особенно будут интересовать случаи $N = n+1$, $n+2$, $n+3$.

а) *$n+1$ точек в общем положении.* Поскольку никакая точка системы не лежит в проективной оболочке остальных (иначе проективная оболочка всей системы имела бы размерность $n-1$, а не n), такие конфигурации уже были рассмотрены в разделе в) п. 6; в частности, проективная группа на них транзитивна. Сейчас мы хотим обратить внимание на то, что проективное преобразование, переводящее одну систему $n+1$ точек в общем положении в другую, не определено однозначно.

Действительно, если $e_1, \ldots, e_{n+1}$ — ненулевые векторы, лежащие в $p_1, \ldots, p_{n+1}$ соответственно, то $\{e_1, \ldots, e_{n+1}\}$ есть базис L (где $P = P(L)$) и группа проективных преобразований, оставляющих на месте все точки p_i, состоит в точности из преобразований вида $P(f)$, где f диагональны в базисе $\{e_1, \ldots, e_{n+1}\}$. Эта оставшаяся степень свободы позволяет доказать транзитивность действия $\mathrm{PGL}(L)$ на системах $n+2$ точек в общем положении.

б) *$n+2$ точки в общем положении.* Если точки $\{p_1, \ldots, p_{n+2}\}$ находятся в общем положении, то точки $\{p_1, \ldots, p_{n+1}\}$ также находятся в общем положении. Как в предыдущем абзаце, выберем базис $\{e_1, \ldots, e_{n+1}\}$, $e_i \in p_i$. Он определяет систему однородных координат в P. Пусть $(x_1: \ldots : x_{n+1})$ — координаты точки p_{n+2} в этом базисе. Ни одна из координат x_i не равна нулю, иначе вектор $(x_1, \ldots, x_{n+1})$ на прямой p_{n+2} линейно выражался бы через векторы e_j, $1 \leqslant j \leqslant n+1$, $j \neq i$, откуда следует, что проективная оболочка $n+1$ точки $\{p_i | i \neq j\}$ имела бы размерность $n-1$, а не n. Но преобразование $P(f)$ с $f = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n+1})$ (в базисе $\{e_1, \ldots, e_{n+1}\}$) переводит $(x_1: \ldots : x_{n+1})$ в точку $(\lambda_1 x_1: \ldots : \lambda_{n+1} x_{n+1})$, а $p_1, \ldots, p_{n+1}$ оставляет на месте. Отсюда следует, что любую точку $(x_1: \ldots : x_{n+1})$ (все $x_i \neq 0$) можно перевести в любую другую $(y_1: \ldots : y_{n+1})$ (все $y_i \neq 0$) единственным проективным преобразованием, оставляющим $p_1, \ldots, p_n$ на месте.

Итак, мы установили, что все упорядоченные системы $n+2$ точек в общем положении в P, где $\dim P = n$, конгруэнтны и, более того, образуют главное однородное пространство над группой $\mathrm{PGL}(L)$.

Принимая «пассивную» точку зрения вместо «активной», мы можем сказать, что для любой упорядоченной системы точек $\{p_1, \ldots, p_{n+2}\}$ существует единственная система однородных координат в P, в которой координаты $p_1, \ldots, p_{n+2}$ имеют следующий вид:

$$p_1 = (1:0:\ldots:0),\ p_2 = (0:1:0:\ldots:0),\ \ldots,\ p_{n+1} = (0:\ldots:0:1),$$
$$p_{n+2} = (1:\ldots:1).$$

Можно назвать эту систему приспособленной к $\{p_1, \ldots, p_{n+2}\}$.

в) *$n+3$ точки в общем положении.* Такие конфигурации уже не все конгруэнтны: если $\{p_1, \ldots, p_{n+3}\}$ и $\{p'_1, \ldots, p'_{n+3}\}$ даны, мы можем найти единственное проективное отображение, переводящее p_i в p'_i для всех $1 \leqslant i \leqslant n+2$, но p_{n+3} попадает или не попадает в p'_{n+3} в зависимости от ситуации.

Нетрудно описать проективные инварианты системы из $n+3$ точек. Выберем систему однородных координат в P, в которой первые $n+2$ точки имеют координаты, описанные в случае б). В ней точка p_{n+3} имеет координаты $(x_1: \ldots : x_{n+1})$, определенные однозначно с точностью до пропорциональности. Любой проективный автоморфизм P, примененный одновременно к конфигурации $\{p_1, \ldots, p_{n+3}\}$ и к приспособленной к ней системе координат, переведет эту конфигурацию в другую, а систему координат — в приспособленную систему координат новой конфигурации. Поэтому координаты $(x_1: \ldots : x_{n+1})$ последней точки останутся теми же самыми.

Все предыдущие рассуждения с очевидными видоизменениями переносятся также на случай, когда у нас имеются два n-мерных проективных пространства P и P', конфигурации $\{p_1, \ldots, p_N\} \subset P$ и $\{p'_1, \ldots, p'_N\} \subset P'$, и мы интересуемся проективными изоморфизмами $P \to P'$, переводящими первую конфигурацию во вторую. Резюмируем результаты обсуждения в следующей теореме:

8. Теорема. а) *Пусть P, P' — n-мерные проективные пространства, $\{p_1, \ldots, p_{n+2}\} \subset P$ и $\{p'_1, \ldots, p'_{n+2}\} \subset P'$ — две системы точек в общем положении. Тогда существует единственный проективный изоморфизм $P \to P'$, переводящий первую конфигурацию во вторую.*

б) *Аналогичный результат верен для систем $n+3$ точек в общем положении тогда и только тогда, когда координаты $(n+3)$-й точки в системе, приспособленной к первым $n+2$ точкам, для обеих конфигураций совпадают (конечно, с точностью до скалярного множителя).*

9. Двойное отношение. Применим теорему п. 8 к случаю $n=1$. Мы получим, прежде всего, что если на двух проективных прямых заданы упорядоченные тройки попарно разных точек (это и есть здесь условие общности положения), то существует единственный проективный изоморфизм прямых, переводящий одну тройку в другую.

Далее, пусть задана четверка попарно разных точек $\{p_1, p_2, p_3, p_4\} \subset P^1$ с координатами $(1:0)$, $(0:1)$, $(1:1)$ и $(x_1:x_2)$ в приспособленной системе. Тогда $x_2 \neq 0$. Положим

$$[p_2, p_3, p_1, p_4] = x_1 x_2^{-1}.$$

Это число называется *двойным отношением* четверки точек $\{p_i\}$. Необычный порядок объясняется желанием сохранить согласованность с классическим определением: в аффинной карте, где $p_2 = \infty$, $p_3 = 0$, $p_1 = 1$, координаты точек в квадратных скобках располагаются так: $[0, 1, \infty, x]$, где x и есть двойное отношение этой четверки.

Сам термин «двойное отношение» происходит из следующей явной формулы для вычисления инварианта $[x_1, x_2, x_3, x_4]$, где $x_i \in \mathcal{K}$

на сей раз понимаются как координаты точек p_i в произвольной аффинной карте P^1. Согласно результатам п. 4 группа $\mathrm{PGL}(1)$ в этой карте представлена дробно-линейными преобразованиями вида $x \longmapsto \frac{ax+b}{cx+d}$, $ad-bc \neq 0$. Такое преобразование, переводящее (x_1, x_2, x_3) в $(0, 1, \infty)$, имеет вид

$$x \longmapsto \frac{x_1 - x}{x_3 - x} : \frac{x_1 - x_2}{x_3 - x_2}.$$

Подставляя сюда $x = x_4$, находим

$$[x_1, x_2, x_3, x_4] = \frac{x_1 - x_4}{x_3 - x_4} : \frac{x_1 - x_2}{x_3 - x_2}.$$

Еще одна классическая конструкция, связанная с утверждением а) теоремы п. 8 для $n = 1$, описывает представление симметрической группы S_3 дробно-линейными преобразованиями. Согласно этой теореме любая перестановка $\{p_1, p_2, p_3\} \longmapsto \{p_{\sigma(1)}, p_{\sigma(2)}, p_{\sigma(3)}\}$ трех точек на проективной прямой индуцирована единственным проективным преобразованием этой прямой.

В аффинной карте, где $\{p_1, p_2, p_3\} = \{0, 1, \infty\}$, эти проективные преобразования представлены дробно-линейными преобразованиями

$$x \longmapsto \left\{x, \frac{1}{x}, 1 - x, \frac{1}{1-x}, \frac{x-1}{x}, \frac{x}{x-1}\right\}.$$

Перейдем теперь к изучению проекций.

10. Пусть линейное пространство L представлено в виде прямой суммы двух своих подпространств размерности $\geqslant 1$: $L = L_1 \oplus L_2$. Положим $P = P(L)$, $P_i = P(L_i)$.

Как было показано в п. 1, линейная проекция $f\colon L \to L_2$, $f(l_1 + l_2) = l_2$, $l_i \in L_i$, индуцирует отображение

$$P(f)\colon P \setminus P_1 \to P_2,$$

которое мы будем называть *проекцией из центра* P_1 *на* P_2. Чтобы описать всю ситуацию в чисто проективных терминах, заметим следующее.

а) $\dim P_1 + \dim P_2 = \dim P - 1$ и $P_1 \cap P_2 = \varnothing$. Наоборот, любая конфигурация (P_1, P_2) с такими свойствами происходит из единственного прямого разложения $L = L_1 \oplus L_2$.

б) Если $a \in P_2$, то $P(f)a = a$; если $a \in P \setminus (P_1 \cup P_2)$, то $P(f)a$ определяется как точка пересечения с P_2 единственной проективной прямой в P, пересекающейся с P_1 и P_2 и проходящей через a.

Действительно, случай $a \in P_2$ очевиден. Если $a \notin P_1 \cup P_2$, то на языке пространства L нужный нам результат формулируется так: через любую прямую $L_0 \subset L$, не лежащую в L_1 и L_2, проходит единственная плоскость, пересекающаяся с L_1 и L_2 по прямым, и ее пересечение с L_2 совпадает с ее проекцией на L_2. В самом деле, одна плоскость с этим свойством есть: она натянута на проекции L_0 на L_1 и L_2 соответственно. Существование двух таких плоскостей влекло бы существование двух разных разложений ненулевого вектора $l_0 \in L_0$ в сумму двух векторов из L_1 и L_2 соответственно, что невозможно, ибо $L = L_1 \oplus L_2$.

Поскольку описанная проективная конструкция отображения $P \to P \setminus P_1$ поднимается до линейной, мы сразу же получаем, что для любого подпространства $P' \subset P \setminus P_1$ ограничение проекции $P' \to P_1$ является проективным отображением, т. е. имеет вид $P(g)$, где g — некоторое линейное отображение соответствующих векторных пространств.

В важном частном случае, когда P_1 — точка, P_2 — гиперплоскость, отображение проекции из центра P_1 на P_2 переводит точку a в ее образ на P_2, видимый наблюдателем из P_1. Поэтому отношение между некоторой фигурой и ее проекцией в таком случае называют еще перспективным. Интуитивно менее очевидна, например, проекция из прямой на прямую в P^3 (рис. 5, 6).

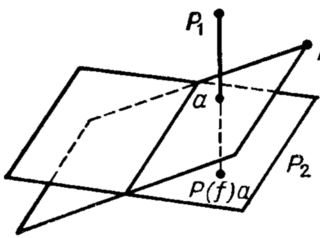


Рис. 5

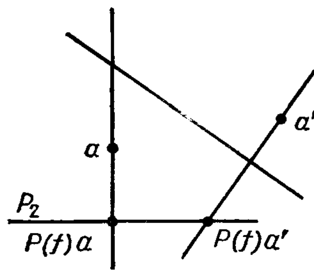


Рис. 6

Важное свойство проекций, которое следует иметь в виду, состоит в следующем: *если $P' \subset P \setminus P_1$, то проекция из центра P_1 определяет проективный изоморфизм P' и его образа в P_2*. Действительно, на языке линейных пространств это означает, что проекция f: $L_1 \oplus L_2 \to L_2$ индуцирует изоморфизм M с $f(M)$, где $M \subset L$ — любое подпространство с $L_1 \cap M = \{0\}$. Это так, ибо $L_1 = \operatorname{Ker} f$.

11. Поведение проекции вблизи центра. Ограничимся дальше рассмотрением проекций из точки $p_1 = a \in P$ и попытаемся понять, что происходит с точками, находящимися вблизи центра. В случае $\mathscr{K} = \mathbf{R}$ и $\mathbf{C}$, когда действительно можно говорить о близости точек, картина такова: в точке a нарушается непрерывность проекции, ибо точки b, как угодно близкие к a, но подходящие к a «с разных сторон», проектируются в далеко отстоящие друг от друга точки p_2. Именно это свойство проекции лежит в основе ее приложений к разного типа вопросам о «разрешении особенностей». Если в P лежит некая «фигура» (алгебраическое многообразие, векторное поле), имеющая вблизи точки a необычное строение, то, проектируя ее из точки a, мы можем растянуть окрестность этой точки и увидеть, что в ней происходит. в увеличенном масштабе, причем коэффициент увеличения при приближении к a безгранично растет.

Хотя эти приложения относятся к существенно нелинейным ситуациям (тривиализируясь в линейных моделях), стоит разобраться в структуре проекции вблизи ее центра несколько подробнее, насколько это можно сделать, оставаясь в рамках линейной геометрии.

12. Введем в $P(L)$ проективную систему координат, в которой центром проекции является точка $(0, \ldots, 0, 1)$, а $P_2 = P^{n-1}$ состоит из точек $(x_0 : x_1 : \ldots : x_{n-1} : 0)$; чтобы добиться этого, следует выбрать в L базис, являющийся объединением базисов в L_1 и L_2 (поскольку центр — точка, $\dim L_1 = 1$).

Нетрудно видеть, что тогда точка $(x_0 : \ldots : x_n)$ проектируется в $(x_0 : \ldots : x_{n-1} : 0)$. Дополнение A к P_2 снабжено аффинной системой координат

$$(y_0, \ldots, y_{n-1}) = (x_0/x_n, \ldots, x_{n-1}/x_n)$$

с началом O в центре проекции. Рассмотрим прямое произведение $A \times P_2 = A \times P^{n-1}$ и в нем график Γ_0 отображения проекции, которое, напомним, определено только на $A \setminus \{0\}$. Этот график состоит из пар точек с координатами

$$((x_0/x_n, \ldots, x_{n-1}/x_n), (x_0 : \ldots : x_{n-1})),$$

где не все $x_0, \ldots, x_{n-1}$ равны нулю одновременно. Увеличим график Γ_0, добавив к нему над точкой $O \in A$ множество $\{0\} \times P^{n-1} \subset A \times P^{n-1}$,

$$\Gamma = \Gamma_0 \cup \{0\} \times P^{n-1},$$

в соответствии с геометрической интуицией, согласно которой при проекции из 0 центр «переходит во все пространство P^{n-1}».

Множество Γ обладает рядом хороших свойств.

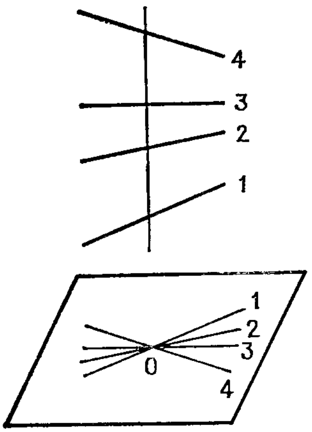


Рис. 7

а) Γ состоит в точности из пар точек

$$((y_0, \ldots, y_{n-1}), (x_0 : \ldots : x_{n-1})),$$

удовлетворяющих системе алгебраических уравнений

$$y_i x_j - y_j x_i = 0; \qquad i, j = 0, \ldots, n-1.$$

Действительно, эти уравнения означают, что все миноры матрицы $\begin{pmatrix} x_0 & \ldots & x_{n-1} \\ y_0 & \ldots & y_{n-1} \end{pmatrix}$ равны нулю, так что ее ранг равен единице (ибо первая строчка ненулевая) и, значит, вторая строка пропорциональна первой. Если коэффициент пропорциональности не равен нулю, мы получаем точку из Γ_0, а если равен, то из $\{0\} \times P^{n-1}$.

б) Отображение $\Gamma \to A$: $((y_0, \ldots, y_{n-1}), (x_0 : \ldots : x_{n-1})) \to (y_0, \ldots \ldots, y_{n-1})$ является биекцией всюду, кроме слоя над точкой $\{0\}$.

Иными словами, Г получается из A «вклеиванием» целого проективного пространства P^{n-1} вместо одной точки. Говорят, что Г получается из A «раздутием» (blowing up) точки, или σ-процессом с центром в точке O. Прообраз в Г каждой прямой в A, проходящей через точку O, пересекает вклеенное проективное пространство P^{n-1} также по одной точке, но своей для каждой прямой. В случае $\mathcal{K} = \mathbf{R}$, $n = 2$ можно представлять себе аффинную карту вклеенного проективного пространства P_1 как ось винта мясорубки Г, делающего полоборота на протяжении своей бесконечной длины (рис. 7).

Реальное применение проекции к исследованию особенности в точке $O \in A$ связано с переносом интересующей нас фигуры с A на Г и рассмотрению геометрии ее прообраза вблизи вклеенного пространства P^{n-1}. При этом, например, прообраз алгебраического многообразия будет алгебраичен благодаря тому, что Г задается алгебраическими уравнениями.

§ 9. Конфигурации Дезарга и Паппа и классическая проективная геометрия

1. Классическая синтетическая проективная геометрия была в значительной мере посвящена изучению семейства подпространств в проективном пространстве с отношением инцидентности; свойства этого отношения можно положить в основу аксиоматики и прийти затем к современному определению пространств $P(L)$ и поля скаляров $\mathcal{K}$ так, что L и $\mathcal{K}$ появятся как производные структуры. В таком построении большую роль играют две конфигурации — Дезарга и Паппа. Мы введем и изучим их в рамках наших определений и затем вкратце опишем их роль в синтетической теории.

2. Конфигурация Дезарга. Пусть S — семейство точек в проективном пространстве. Символом $\bar{S}$ мы будем обозначать его проективную оболочку. Рассмотрим в трехмерном проективном пространстве упорядоченную шестерку точек $(p_1, p_2, p_3; q_1, q_2, q_3)$. Предполагается, что точки попарно разные и что $\overline{p_1p_2p_3}$ и $\overline{q_1q_2q_3}$ суть плоскости. Далее, пусть прямые $\overline{p_1q_1}$, $\overline{p_2q_2}$ и $\overline{p_3q_3}$ пересекаются в одной точке r, отличной от p_i и q_j. Иными словами, «треугольники» $p_1p_2p_3$ и $q_1q_2q_3$ «перспективны», и каждый из них есть проекция другого из центра r, если они лежат в разных плоскостях. Тогда для любой пары различных индексов $\{i, j\} \subset \{1, 2, 3\}$ прямые $\overline{p_ip_j}$ и $\overline{q_iq_j}$ не совпадают, иначе мы имели бы $p_i = q_i$, ибо p_i и q_i суть точки пересечения этих прямых с прямой $\overline{p_iq_i}$. Кроме того, прямые $\overline{p_ip_j}$ и $\overline{q_iq_j}$ лежат в общей плоскости $\overline{rp_ip_j}$. Поэтому они пересекаются в точке, которую мы обозначим s_k, где $\{i, j, k\} = \{1, 2, 3\}$: это точка пересечения продолжений пары соответствующих сторон треугольников $p_1p_2p_3$ и $q_1q_2q_3$.

Теорема Дезарга, которую мы докажем в следующем пункте, утверждает, что три точки s_1, s_2, s_3 лежат на одной прямой. Конфигурация, состоящая из десяти точек p_i, q_j, s_k, r и десяти соединяющих их прямых, показанных на рис. 8, называется *конфигурацией Дезарга.* Каждая ее прямая содержит ровно три ее точки, и через каждую ее точку проходят ровно три ее прямые. Читателю предлагается самостоятельно убедиться в том, что она по существу симметрична (в том смысле, что группа перестановок ее точек и прямых, сохраняющая отношения инцидентности, транзитивна как на точках, так и на прямых).

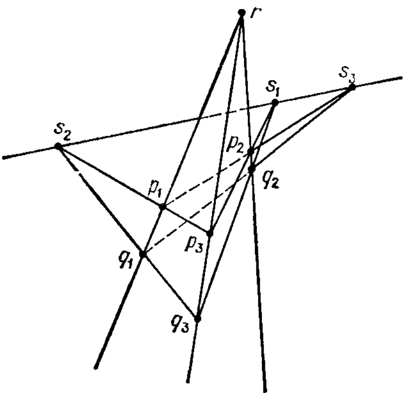


Рис. 8

3. Теорема Дезарга. *В описанных выше условиях точки* s_1, s_2, s_3 *лежат на одной прямой.*

Доказательство. Мы разберем два случая в зависимости от того, совпадают плоскости $p_1p_2p_3$ и $q_1q_2q_3$ или нет.

а) $p_1p_2p_3 \neq q_1q_2q_3$ («пространственная теорема Дезарга». В этом случае плоскости $\overline{p_1p_2p_3}$ и $\overline{q_1q_2q_3}$ пересекаются по прямой, и нетрудно убедиться, что s_1, s_2, s_3 лежат на ней. Действительно, точка s_1, например, лежит на прямых $\overline{p_2p_3}$ и $\overline{q_2q_3}$, которые в свою очередь лежат в плоскостях $\overline{p_1p_2p_3}$ и $\overline{q_1q_2q_3}$ и, значит, в их пересечении.

б) $\overline{p_1p_2p_3} = \overline{q_1q_2q_3}$ («плоская теорема Дезарга»). В этом случае выберем в пространстве точку r', не лежащую в плоскости $\overline{p_1p_2p_3}$, и соединим ее прямыми с точками r, p_i, p_j. В плоскости $\overline{r'p_1q_1}$ лежит прямая $\overline{p_1q_1}$ и, значит, точка r. Проведем в ней через r прямую, не проходящую через точки r' и p_1, и обозначим ее пересечения с прямыми $\overline{r'p_1}$, $\overline{r'q_1}$, через p_1', q_1' соответственно. Тройки (p_1', p_2, p_3) и (q_1', q_2, q_3) лежат уже в разных плоскостях — иначе содержащая их общая плоскость содержала бы прямые $\overline{p_2p_3}$ и $\overline{q_2q_3}$ и потому совпадала бы с $\overline{p_1p_2p_3}$, но это невозможно, ибо p_1', q_1' в этой начальной плоскости не лежат. Кроме того, прямые $\overline{p_1'q_1'}$, $\overline{p_2q_2}$ и $\overline{p_3q_3}$ проходят через точку r. В силу пространственной теоремы Дезарга точки $\overline{p_1'p_2} \cap \overline{q_1'q_2}$, $\overline{p_1'p_3} \cap \overline{q_1'q_3}$ и $\overline{p_2p_3} \cap \overline{q_2q_3}$ лежат на одной прямой. Но если спроектировать эти точки из r' на плоскость $\overline{p_1p_2p_3}$, то получатся как раз s_3, s_2, s_1 соответственно, потому что r' проектирует (p_1', p_2, p_3) в (p_1, p_2, p_3) и (q_1', q_2, q_3) в (q_1, q_2, q_3) и, значит, стороны каждого из этих треугольников в соответствующие стороны исходных треугольников.

Это завершает доказательство.

4. Конфигурация Паппа. Рассмотрим в проективной плоскости две разные прямые и две тройки лежащих на них попарно разных точек p_1, p_2, p_3 и q_1, q_2, q_3. Для любой пары различных индексов

$\{i, j\} \subset \{1, 2, 3\}$ построим точку $s_k = \overline{p_i q_j} \cap \overline{q_i p_j}$, где $\{i, j, k\} = \{1, 2, 3\}$.

5. Теорема Паппа. *Точки s_1, s_2, s_3 лежат на одной прямой.*

Доказательство. Проведем прямую через точки s_3, s_2 и обозначим через s_4 ее пересечение с прямой $\overline{p_1 q_1}$. Наша цель состоит в доказательстве того, что s_1 лежит на ней.

Построим два проективных отображения f_1, f_2: $\overline{p_1 p_2 p_3} \to \overline{q_1 q_2 q_3}$. Первое из них, f_1, будет композицией проекции $\overline{p_1 p_2 p_3}$ на $\overline{s_2 s_3}$ из точки q_1 с проекцией $\overline{s_3 s_2}$ на $\overline{q_1 q_2 q_3}$ из точки p_1. Очевидно, $f_1(p_i) = q_i$ для всех $i = 1, 2, 3$, и, кроме того, $f_1(t_1) = t_2$, где $t_1 = \overline{p_1 p_2 p_3} \cap \overline{q_1 q_2 q_3}$, $t_2 = \overline{s_4 s_3 s_2} \cap \overline{q_1 q_2 q_3}$ (рис. 9).

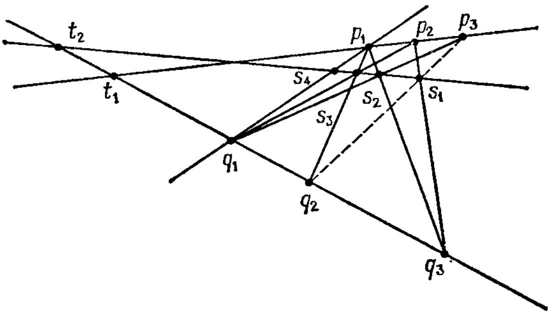


Рис. 9

Второе из них, f_2, будет композицией проекции $\overline{p_1 p_2 p_3}$ на $\overline{s_3 s_2}$ из точки q_2 с проекцией $\overline{s_3 s_2}$ на $\overline{q_1 q_2 q_3}$ из точки p_2. Эта композиция переводит p_1 в q_1, p_2 в q_2 и t_1 в t_2.

Поскольку f_1 и f_2 одинаково действуют на тройках точек (t_1, p_1, p_2), они должны совпадать. В частности, $f_1(p_3) = f_2(p_3)$. Но $f_1(p_3) = q_3$. Значит, $f_2(p_3) = q_3$. Это утверждение геометрически означает следующее: если обозначить через s_1' пересечение $\overline{q_2 p_3} \cap \overline{s_3 s_2}$, то прямая $\overline{p_2 q_3}$ проходит через s_1'. Но тогда $s_1' = \overline{q_2 p_3} \cap \overline{p_2 q_3} = s_1$. Значит, s_1 лежит на $\overline{s_2 s_3}$, что и требовалось доказать.

6. Классические аксиомы трехмерного проективного пространства и проективной плоскости. Классическое трехмерное проективное пространство определяется как множество, элементы которого называются точками, снабженное двумя системами подмножеств, элементы которых называются соответственно прямыми и плоскостями. При этом должны выполняться следующие аксиомы.

T_1. Две разные точки принадлежат единственной прямой.

T_2. Три разные точки, не лежащие на одной прямой, принадлежат единственной плоскости.

T_3. Прямая и плоскость имеют общую точку.

T_4. Пересечение двух плоскостей содержит прямую.

T_5. Существуют четыре точки, не лежащие в одной плоскости и такие, что любые три из них не лежат на одной прямой.

T_6. Каждая прямая состоит не менее чем из трех точек.

Классическая проективная плоскость определяется как множество, элементы которого называются точками, снабженное системой подмножеств, элементы которой называются прямыми. При этом должны выполняться следующие аксиомы.

Π_1. Две разные точки принадлежат единственной прямой.
Π_2. Пересечение двух прямых непусто.
Π_3. Существуют три точки, не лежащие на одной прямой.
Π_4. Каждая прямая состоит не менее чем из трех точек.

Множества $P(L)$, где L — линейное пространство над полем $\mathscr{K}$ размерности 4 или 3, вместе с системами проективных плоскостей и прямых в них, как они были введены выше, удовлетворяют аксиомам T_1—T_6 и Π_1—Π_4 соответственно, что немедленно следует из стандартных свойств линейных пространств, доказанных в ч. 1. Однако не всякое классическое проективное пространство или плоскость изоморфно (в очевидном смысле слова) одному из наших пространств $P(L)$. Следующая фундаментальная конструкция дает много новых примеров.

7. Линейные и проективные пространства над телами. *Телом* (или не обязательно коммутативным полем) называется ассоциативное кольцо K, множество ненулевых элементов которого образуют группу по умножению (не обязательно коммутативную). Все поля являются телами, но обратное неверно. Например, кольцо классических кватернионов является телом, но не полем.

Аддитивная группа L вместе с бинарным законом умножения $K\times L\to L$: $(a, l)\mapsto al$ называется (левым) *линейным пространством над телом* K, если выполнены условия определения п. 2 § 1 ч. 1. Значительная часть теории линейных пространств над полями почти без изменений переносится на линейные пространства над телами. В частности, это относится к теории размерности и базиса и теории подпространств, включая теорему о размерности пересечения. Это позволяет построить по каждому телу K и линейному пространству L над ним проективное пространство $P(L)$, состоящее из прямых в L, и систему его проективных подпространств $P(M)$, где $M\subset L$ пробегает линейные подпространства разных размерностей. Когда $\dim_K L=4$ или 3, эти объекты удовлетворяют всем аксиомам T_1—T_6 и Π_1—Π_4 соответственно.

8. Роль теоремы Дезарга. Оказывается, однако, что существуют классические проективные плоскости, не изоморфные даже никакой плоскости вида $P(L)$, где L — трехмерное линейное пространство над каким-нибудь телом. Причина этого состоит в том, что в проективных плоскостях вида $P(L)$ теорема Дезарга по-прежнему верна, тогда как существуют недезарговы плоскости, где она не выполняется. Сформулируем без доказательства следующий результат:

9. Теорема. *Три свойства классической проективной плоскости равносильны:*

а) *В ней выполняется плоская теорема Дезарга.*

б) *Ее можно вложить в классическое проективное пространство.*

в) *Существует линейное трехмерное пространство L над некоторым телом K, определенным однозначно с точностью до изоморфизма, такое, что наша плоскость изоморфна $P(L)$.*

Импликация б)$\Rightarrow$а) устанавливается прямой проверкой того, что доказательство пространственной теоремы Дезарга использует лишь аксиомы T_1—T_6. Импликация в)$\Rightarrow$б) следует из того, что L можно вложить в четырехмерное линейное пространство над тем же телом.

Наконец, импликация а)$\Rightarrow$в), являющаяся самым тонким моментом доказательства, устанавливается прямой конструкцией тела по дезарговой проективной плоскости. Именно, сначала с помощью геометрической конструкции проекций из центра вводится понятие проективного отображения проективных прямых в плоскости. Далее, доказывается, что для двух упорядоченных троек точек, лежащих на двух прямых, существует единственное проективное отображение одной прямой в другую. Наконец, фиксируется прямая D с тройкой точек p_0, p_1, p_2, множество K определяется как $D\setminus\{p_2\}$ с нулем p_0 и единицей p_1, и законы сложения и умножения в K вводятся геометрически с помощью проективных преобразований. В проверках аксиом тела существенно используется теорема Дезарга, в этом контексте возникающая как *аксиома Дезарга* Π_5.

10. Роль теоремы Паппа. Даже в дезарговых плоскостях теорема Паппа может не выполняться. Назвав соответствующее утверждение *аксиомой Паппа* Π_6, мы можем сформулировать следующую теорему, которую мы также приведем без доказательства:

11. Теорема. а) *Если в классической проективной плоскости выполнена аксиома Паппа, то плоскость является дезарговой.*

б) *Дезаргова классическая плоскость удовлетворяет аксиоме Паппа тогда и только тогда, когда связанное с ней тело коммутативно, т. е. эта плоскость изоморфна $P(L)$, где L — трехмерное линейное пространство над полем.*

Дальнейшие подробности и опущенные нами доказательства читатель может найти в книге: Х а р т с х о р н Р. Основы проективной геометрии. — М.: Мир, 1970.

§ 10. Кэлерова метрика

1. Если L — унитарное линейное пространство над $\mathbf{C}$, то на проективном пространстве $P(L)$ можно ввести специальную метрику, называемую кэлеровой в честь открывшего ее важные обобщения Э. Кэлера. Сама эта метрика была введена в прошлом веке Фубини и Штуди. Она играет особенно важную роль в комплексной алгебраической геометрии и неявно также в квантовой механике, потому что такие пространства $P(L)$, как было объяснено в ч. 2, являются пространствами состояний квантовомеханических систем.

Эта метрика инвариантна относительно тех проективных преобразований $P(L)$, которые имеют вид $P(f)$, где f — унитарное отображение L в себя. Вводится она следующим образом. Пусть $p_1, p_2 \in P(L)$. Точкам p_1, p_2 отвечают два больших круга на единичной сфере $S \subset L$, как показано в разделе в) п. 3 § 6. Тогда кэлерово расстояние $d(p_1, p_2)$ равно расстоянию между этими кругами в евклидовой сферической метрике S, т. е. длине кратчайшей дуги большого круга на S, соединяющей две точки на прообразах p_1 и p_2.

Основная цель этого параграфа — доказательство следующих двух формул для $d(p_1, p_2)$.

2. Теорема. а) *Пусть $l_1, l_2 \in L$, $|l_1| = |l_2| = 1$; $p_1, p_2 \in P(L)$ — прямые $\mathbf{C}l_1$ и $\mathbf{C}l_2$. Тогда*

$$d(p_1, p_2) = \arccos |(l_1, l_2)|,$$

где (l_1, l_2) — скалярное произведение в L.

б) *Пусть в L выбран ортонормированный базис, относительно которого в $P(L)$ определена однородная система координат. Пусть две близкие точки $p_1, p_2 \in P(L)$ заданы своими координатами $(y_1, \ldots, y_n)$ и $(y_1 + dy_1, \ldots, y_n + dy_n)$ в аффинной карте U_0 (см. раздел а) п. 3 § 6). Тогда квадрат расстояния между ними с точностью до третьего порядка малости по dy_i равен*

$$\frac{\sum_{i=1}^{n} |dy_i|^2}{1 + \sum_{i=1}^{n} |y_i|^2} - \frac{\left|\sum_{i=1}^{n} y_i \overline{dy_i}\right|^2}{\left(1 + \sum_{i=1}^{n} |y_i|^2\right)^2}.$$

Доказательство. а) В евклидовом пространстве расстояние между двумя точками на единичной сфере равно длине той дуги соединяющего их большого круга, которая лежит между 0 и π, т. е. евклидову углу, или арккосинусу евклидова скалярного произведения радиусов. Евклидова структура на L, отвечающая исходной унитарной структуре, задается скалярным произведением $\operatorname{Re}(l_1, l_2)$. Поскольку мы должны найти минимальное расстояние между точками больших кругов $(e^{i\varphi} l_1)$, $(e^{i\psi} l_2)$, а арккосинус — убывающая функция, нужно подобрать φ и ψ так, чтобы при данных l_1, l_2 величина $\operatorname{Re}(e^{i\varphi} l_1, e^{i\psi} l_2)$ приняла наибольшее возможное значение. Но она не превосходит $|(l_1, l_2)|$ и при подходящих φ, ψ достигает этого значения: если $\varphi = -\arg(l_1, l_2)$, то $(e^{i\varphi} l_1, l_2) = |(l_1, l_2)|$. Поэтому окончательно

$$d(p_1, p_2) = \arccos |(l_1, l_2)|.$$

б) Положим

$$R = \left(1 + \sum_{i=1}^{n} |y_i|^2\right)^{1/2}, \quad R + dR = \left(1 + \sum_{i=1}^{n} |y_i + dy_i|^2\right)^{1/2}.$$

Тогда прообразами точек $(y_1, \ldots, y_n)$ и $(y_1 + dy_1, \ldots, y_n + dy_n)$ на S будут точки

$$l_1 = \left(\frac{1}{R}, \frac{y_1}{R}, \ldots, \frac{y_n}{R}\right), \quad l_2 = \left(\frac{1}{R + dR}, \frac{y_1 + dy_1}{R + dR}, \ldots, \frac{y_n + dy_n}{R + dR}\right).$$

Поэтому

$$(l_1, l_2) = \frac{1}{R(R+dR)}\left(1+\sum_{i=1}^{n} y_i(\bar{y}_i+\overline{dy}_i)\right) = \frac{R^2+\sum\limits_{i=1}^{n} y_i\overline{dy}_i}{R(R+dR)}$$

и

$$|(l_1, l_2)|^2 = \frac{R^4+R^2\sum\limits_{i=1}^{n}(y_i\overline{dy}_i+\bar{y}_i dy_i)+\left|\sum\limits_{i=1}^{n} y_i\overline{dy}_i\right|^2}{R^2(R+dR)^2}.$$

Далее,

$$(R+dR)^2 = 1+\sum_{i=1}^{n}|y_i+dy_i|^2 = R^2+\sum_{i=1}^{n}(y_i\overline{dy}_i+\bar{y}_i dy_i+|dy_i|^2).$$

Поэтому с точностью до третьего порядка малости по dy_i

$$|(l_1, l_2)|^2 = 1-\frac{\sum\limits_{i=1}^{n}|dy_i|^2}{R^2}+\frac{\left|\sum\limits_{i=1}^{n} y_i\overline{dy}_i\right|^2}{R^4}+\ldots$$

С другой стороны, если $\varphi = \arccos|(l_1, l_2)|$, то с точностью до φ^4 при малых φ имеем

$$|(l_1, l_2)|^2 = (\cos\varphi)^2 = \left(1-\frac{\varphi^2}{2}+\ldots\right)^2 = 1-\varphi^2+\ldots$$

Сравнение этих формул завершает доказательство.

§ 11. Алгебраические многообразия и многочлены Гильберта

1. Пусть $P(L)$ — n-мерное проективное пространство над полем $\mathscr{K}$ с фиксированной системой однородных координат. Мы уже многократно встречались с проективными подпространствами в $P(L)$ и квадриками, которые определяются соответственно системами уравнений

$$\sum_{i=0}^{n} a_{ik}x_i = 0, \qquad k = 1, \ldots, m,$$

или

$$\sum_{i, j=0}^{n} a_{ij}x_i x_j = 0, \qquad a_{ij} = a_{ji}.$$

Более общо, рассмотрим произвольный *однородный многочлен*, или *форму*, степени $m \geqslant 1$:

$$F(x_0, \ldots, x_n) = \sum_{i_0+\ldots+i_n=m} a_{i_0\ldots i_n} x_0^{i_0} \ldots x_n^{i_n}.$$

Хотя она не определяет функции на $P(L)$, множество точек с однородными координатами $(x_0: \ldots : x_n)$, для которых $F = 0$, определено однозначно. Оно называется *алгебраической гиперповерхностью* (степени m), заданной уравнением $F=0$.

Более общо, множество точек в $P(L)$, удовлетворяющих системе уравнений

$$F_1=F_2=\ldots=F_k=0,$$

где F_i — формы (возможно, разных степеней), называется *алгебраическим многообразием*, определенным этой системой уравнений.

Изучение алгебраических многообразий в проективном пространстве составляет одну из основных целей алгебраической геометрии. Разумеется, общее алгебраическое многообразие является существенно нелинейным объектом, поэтому, как и в других геометрических дисциплинах, важное место в технике алгебраической геометрии занимают методы линеаризации нелинейных задач.

В этом параграфе мы введем один такой метод, восходящий к Гильберту и дающий с минимумом предварительной подготовки важную информацию об алгебраическом многообразии $V\subset P(L)$. Его идея состоит в том, чтобы поставить в соответствие алгебраическому многообразию V счетную серию линейных пространств $\{I_m(V)\}$ и изучить их размерность как функцию от m. Именно, пусть $I_m(V)$ — пространство форм степени m, обращающихся в нуль на V.

Покажем, что существует многочлен $Q_V(m)$ с рациональными коэффициентами такой, что $\dim I_m(V)=Q_V(m)$ для всех достаточно больших m. Коэффициенты многочлена Q_V являются важнейшими инвариантами V. На самом деле мы установим заметно более общий результат, но для его формулировки и доказательства нам придется ввести несколько новых понятий.

2. Градуированные линейные пространства. Фиксируем раз навсегда основное поле скаляров $\mathcal{K}$. *Градуированным линейным пространством над* $\mathcal{K}$ будем называть линейное пространство L вместе с фиксированным его разложением в прямую сумму подпространств: $L=\bigoplus\limits_{i=0}^{\infty}L_i$. Эта сумма бесконечна, но каждый отдельный элемент $l\in L$ однозначно представляется в виде конечной суммы: $l=\sum\limits_{i=0}^{\infty}l_i,\quad l_i\in L_i$, в том смысле, что все l_i, кроме конечного числа их, равны 0. Вектор l_i называется *однородной компонентой* l *степени* i; если $l\in L_i$, то l называется *однородным элементом степени* i.

Пример: кольцо многочленов $A^{(n)}$ от независимых переменных $x_0,\ldots,x_n$ разлагается как линейное пространство в прямую сумму $\bigoplus\limits_{i=0}^{\infty}A_i^{(n)}$, где $A_i^{(n)}$ состоит из однородных многочленов степени i. Заметим, что если интерпретировать x_i как координатные функции на линейном пространстве L, а элементы $A^{(n)}$ как полиномиальные функции на этом пространстве, то линейные обратимые замены координат сохраняют однородность и степень.

Другой пример: $I = \bigoplus_{m=0}^{\infty} I_m(V)$, где V — некоторое алгебраическое многообразие. Очевидно, $I \subset A^{(n)}$ и $I_m(V) = A_m^{(n)} \cap I$.

Более общо, *градуированное подпространство* M градуированного пространства $L = \bigoplus_{i=0}^{\infty} L_i$ — это линейное подпространство, обладающее следующим свойством: $M = \bigoplus_{i=0}^{\infty} (M \cap L_i)$. Очевидное равносильное условие: все однородные компоненты любого элемента M сами являются элементами M.

Если $M \subset L$ — пара, состоящая из градуированного пространства и его градуированного подпространства, то факторпространство L/M также обладает естественной градуировкой. Именно, рассмотрим естественное линейное отображение

$$\bigoplus_{i=0}^{\infty} L_i/M_i \to L/M\colon \sum_{i=0}^{\infty} (l_i + M_i) \mapsto \left(\sum_{i=0}^{\infty} l_i\right) + M$$

(суммы справа конечны). Оно сюръективно, потому что любой элемент $\sum_{i=0}^{\infty} l_i$, $l_i \in L$, есть образ элемента $\sum_{i=0}^{\infty} (l_i + M_i)$. Оно инъективно, ибо если $\sum_{i=0}^{\infty} l_i + M = M$, то $\sum_{i=0}^{\infty} l_i \in M$ и $l \in M$ в силу однородности M. Поэтому это отображение — изоморфизм, и мы можем определить градуировку L/M, положив $(L/M)_i = L_i/M_i$.

Семейство градуированных подпространств в L замкнуто относительно пересечений и сумм, и все обычные изоморфизмы линейной алгебры имеют очевидные градуированные варианты.

3. Градуированные кольца. Пусть A — градуированное линейное пространство над $\mathscr{K}$, являющееся в то же время коммутативной $\mathscr{K}$-алгеброй с единицей, умножение в которой подчинено условию

$$A_i A_j \subset A_{i+j}.$$

Тогда A называется *градуированным кольцом* (точнее, градуированной $\mathscr{K}$-алгеброй). Так как $\mathscr{K} A_i \subset A_i$, имеем $\mathscr{K} \subset A_0$. Важнейший пример — кольца многочленов $A^{(n)}$; в них, конечно, $A_0^{(n)} = \mathscr{K}$.

4. Градуированные идеалы. *Идеалом* I в произвольном коммутативном кольце A называется подмножество, образующее аддитивную подгруппу A и замкнутое относительно умножения на элементы A: *если* $f \in I$ и $a \in A$, то $af \in I$. *Градуированным идеалом в градуированном кольце* A называется идеал, который как $\mathscr{K}$-подпространство A градуирован, т. е. $I = \bigoplus_{m=0}^{\infty} I_m$, $I_m = I \cap A_m$.

Основной пример: идеалы $I_m(V)$ алгебраических многообразий в кольцах многочленов $A^{(n)}$. Стандартная конструкция идеалов такова: пусть $S \subset A$ — любое подмножество элементов. Тогда множество всех конечных линейных комбинаций $\left\{\sum_{s_i \in S} a_i s_i \mid a_i \in A\right\}$ является идеалом в A, *порожденным множеством* S. Множество S называется *системой образующих* этого идеала. Если идеал имеет конечное число образующих, то он называется *конечно порожденным.* В градуированном случае достаточно рассматривать множества S, состоящие только из однородных элементов; порожденные ими идеалы тогда автоматически градуированы. Действительно, однородная компонента степени j любой линейной комбинации $\sum a_i s_i$ также будет линейной комбинацией $\sum a_i^{(k_i)} s_i$, где $a_i^{(k_i)}$ — однородная компонента a_i степени $k_i = j - \deg s_i$ ($\deg s_i$ — степень s_i). Поэтому она лежит в идеале, порожденном S. Если градуированный идеал конечно порожден, то у него есть конечная система однородных образующих: она состоит из однородных компонент элементов исходной системы.

Для упрощения доказательств следует обобщить понятие градуированного идеала и рассмотреть также градуированные модули. Это — последнее из списка нужных нам понятий.

5. Градуированные модули. *Модуль* M над коммутативным кольцом A, или A-модуль, — это аддитивная группа, снабженная операцией $A \times M \to M$: $(a, m) \mapsto am$, которая ассоциативна ($(ab)m = a(bm)$ для всех $a, b \in A$, $m \in M$) и дистрибутивна по обоим аргументам:

$$(a+b)m = am + bm, \quad a(m+n) = am + an.$$

Кроме того, мы требуем, чтобы $1m = m$ для всех $m \in M$, где 1 — единица в A.

Если A — поле, то M — просто линейное пространство над A; можно сказать, что понятие модуля является обобщением понятия линейного пространства на случай, когда скаляры образуют лишь кольцо (см. Введение в алгебру, гл. 9, § 3).

Если A — градуированная $\mathscr{K}$-алгебра, то *градуированным A-модулем* M мы назовем A-модуль, являющийся градуированным линейным пространством над $\mathscr{K}$, $M = \bigoplus_{i=0}^{\infty} M_i$, и такой, что

$$A_i M_j \subset M_{i+j}$$

для всех $i, j \geqslant 0$. Примеры:

а) A является градуированным A-модулем.

б) Любой градуированный идеал в A является градуированным A-модулем.

Если M — градуированный A-модуль, то любое градуированное подпространство $N \subset M$, замкнутое относительно умножения на элементы A, само является градуированным A-модулем — подмодулем M. По любой системе однородных элементов $S \subset M$ можно построить порожденный ею градуированный подмодуль, состоящий из всех конечных линейных комбинаций $\sum a_i s_i$, $a_i \in A$, $s_i \in S$. Если он совпадает с M, то S называется однородной системой образующих M. Модуль, имеющий конечную систему образующих, называется конечно порожденным. Если градуированный модуль имеет какую-нибудь конечную систему образующих, то он имеет и конечную систему однородных образующих: однородные компоненты элементов исходной системы.

Рассмотрение всевозможных модулей, а не только идеалов, в нашей задаче дает большую свободу действий. Умножение на элементы $a \in A$ в фактормодуле вводится формулой

$$a(m+N) = am + N.$$

В градуированном случае градуировка на M/N определяется прежней формулой $(M/N)_i = M_i/N_i$. Корректность определения проверяется тривиально.

Элементы теории прямых сумм, подмодулей и фактормодулей формально не отличаются от соответствующих результатов для линейных пространств.

Теперь мы можем приступить к доказательству основных результатов этого параграфа.

6. Теорема. *Пусть M — произвольный конечно порожденный модуль над кольцом многочленов $A^{(n)} = \mathcal{K}[x_0, \ldots, x_n]$ от конечного числа переменных. Тогда любой подмодуль $N \subset M$ конечно порожден.*

Доказательство. Мы разобьем его на несколько шагов. Стандартная терминология: *модуль, каждый подмодуль которого конечно порожден*, называется *нётеровым* (в честь Эмми Нётер).

а) *Модуль M нётеров тогда и только тогда, когда любая бесконечная цепочка возрастающих подмодулей $M_1 \subset M_2 \subset \ldots$ в M стабилизируется: существует такое a_0, что $M_a = M_{a+1}$ для всех $a \geqslant a_0$.*

В самом деле, пусть M нётеров. Положим $N = \bigcup_{i=1}^{\infty} M_i$. Пусть $n_1, \ldots, n_k$ — конечная система образующих в N. Для каждого $1 \leqslant j \leqslant k$ существует $i(j)$ такой, что $n_j \in M_{i(j)}$. Положим $a_0 = \max\{i(j) \mid j = 1, \ldots, k\}$. Тогда M_a содержит $n_1, \ldots, n_k$ для всех $a \geqslant a_0$ и потому $M_a = N$.

Наоборот, пусть любая возрастающая цепочка подмодулей в M обрывается. Будем строить систему образующих подмодуля $N \subset M$ индуктивно: в качестве $n_1 \in N$ возьмем любой элемент; если $n_1, \ldots, n_i \in N$ уже построены, обозначим через $M_i \subset N$ порожденный ими подмодуль и при $N \neq M_i$ выберем n_{i+1} из $N \setminus M_i$. Этот процесс оборвется через конечное число шагов, иначе цепочка $M_1 \subset \ldots \subset M_i \subset \ldots$ не стабилизировалась бы.

б) *Если подмодуль $N \subset M$ нётеров и фактормодуль M/N нётеров, то M нётеров; верно и обратное.*

Действительно, пусть $M_1 \subset M_2 \subset \ldots$ — цепочка подмодулей в M. Пусть a_0 таково, что обе цепочки $M_1 \cap N \subset M_2 \cap N \subset \ldots$ и $(M_1+N)/N \subset (M_2+N)/N \subset \ldots$ стабилизируются при $a \geqslant a_0$. Тогда и цепочка $M_1 \subset M_2 \subset \ldots$ стабилизируется при $a \geqslant a_0$.

Обратное утверждение очевидно.

в) *Прямая сумма конечного числа нётеровых модулей нётерова.*

Действительно, пусть $M = \bigoplus\limits_{i=0}^{n} M_i$, M_i нётеровы. Проведем индукцию по n. Случай $n=1$ очевиден. При $n \geqslant 2$ модуль M содержит подмодуль, изоморфный M_n, с фактором, изоморфным $\bigoplus\limits_{i=0}^{n-1} M_i$. Оба этих модуля нётеровы, так что M нётеров в силу б).

г) *Кольцо $A^{(n)}$ нётерово как модуль над самим собой. Иными словами, любой идеал в $A^{(n)}$ конечно порожден.*

Это — основной частный случай теоремы, установленный впервые Гильбертом. Доказывается он индукцией по n. Случай $n=-1$, т. е $A^{(-1)} = \mathscr{K}$, очевиден. В самом деле, любой идеал I в поле $\mathscr{K}$ совпадает либо с $\{0\}$, либо с $\mathscr{K}$: если $a \in I$, $a \neq 0$, то $b = (ba^{-1})a \in I$ для всех $b \in \mathscr{K}$. Индуктивный шаг основан на рассмотрении $A^{(n)}$ как $A^{(n-1)}[x_n]$. Пусть $I^{(n)} \subset A^{(n)}$ — идеал. Представим каждый элемент из $I^{(n)}$ как многочлен по степеням x_n с коэффициентами из $A^{(n-1)}$. Множество всех старших коэффициентов таких многочленов есть идеал $I^{(n-1)}$ в $A^{(n-1)}$. По предположению индукции он имеет конечное число образующих $\varphi_1, \ldots, \varphi_m$. К каждой образующей φ_i подберем элемент $f_i = \varphi_i x_n^{d_i} + \ldots$ из $I^{(n)}$, где многоточием обозначены члены низших степеней по x_n. Положим $d = \max\limits_{1 \leqslant i \leqslant m} \{d_i\}$. Многочлены $f_1, \ldots, f_m$ порождают в $A^{(n)}$ некоторый идеал $I \subset I^{(n)}$.

Пусть теперь $f = \varphi x^s +$ (члены низших степеней) — любой элемент из $I^{(n)}$. По определению $\varphi \in I^{(n-1)}$, так что $\varphi = \alpha_1\varphi_1 + \ldots + \alpha_m\varphi_m$. Если $s \geqslant d$, то многочлен $f - \sum \alpha_i f_i x^{s-d_i}$ принадлежит $I^{(n)}$ и его степень $< s$. Действуя аналогичным образом, получим в результате выражение $f = g + h$, где $h \in I$, а g — многочлен из $I^{(n)}$ степени, меньшей d.

Все многочлены из $I^{(n)}$ степени $< d$ образуют подмодуль J в $A^{(n-1)}$-модуле, порожденном конечной системой $\{1, x_n, \ldots, x_n^{d-1}\}$. В соответствии с предположением индукции о нётеровости $A^{(n-1)}$ и с утверждением в) подмодуль J конечно порожден.

Мы доказали, что $I^{(n)} = I + J$ — сумма двух конечно порожденных модулей. Поэтому идеал $I^{(n)}$ конечно порожден.

Теперь мы можем без труда завершить доказательство теоремы. Пусть модуль M над $A^{(n)}$ имеет конечное число образующих $m_1, \ldots, m_k$. Тогда имеется сюръективный гомоморфизм $A^{(n)}$-модулей

$$\underbrace{A^{(n)} \oplus \ldots \oplus A^{(n)}}_{k \text{ раз}} \to M: (f_1, \ldots, f_k) \mapsto \sum_{i=1}^{k} f_i m_i.$$

Модуль $A^{(n)} \oplus \ldots \oplus A^{(n)}$ нётеров в силу г) и в). Следовательно (см. б)), его фактормодуль M нётеров.

7. Следствие. *Любая бесконечная система уравнений $F_i = 0$, $i \in S$, где F_i — многочлены в $A^{(n)}$, эквивалентна некоторой своей конечной подсистеме.*

Доказательство. Пусть I — идеал, порожденный всеми F_i. Он имеет конечную систему образующих $\{G_j\}$. Рассмотрим такое конечное подмножество $S_0 \subset S$, что все G_j линейно выражаются через F_i, $i \in S_0$. Тогда система уравнений $F_i = 0$, $i \in S_0$, эквивалентна исходной, т. е. имеет то же самое множество решений.

8. Многочлены Гильберта и ряд Пуанкаре градуированного модуля. Пусть теперь M — градуированный конечно порожденный модуль над кольцом $A^{(n)}$. Тогда все линейные пространства $M_k \subset M$ конечномерны над $\mathcal{K}$. Действительно, если $\{a_i\}$ — однородный $\mathcal{K}$-базис $A_0^{(n)} + \ldots + A_k^{(n)}$ и $\{m_j\}$ — конечная система образующих M, то M_k как линейное пространство порождено конечным числом элементов $a_i m_j$ с $\deg a_i + \deg m_j = k$.

Положим $d_k(M) = \dim M_k$. Формальный степенной ряд от переменной t

$$H_M(t) = \sum_{k=0}^{\infty} d_k(M) t^k$$

называется *рядом Пуанкаре* модуля M.

9. Теорема. а) *В условиях предыдущего пункта существует такой многочлен $f(t)$ с целыми коэффициентами, что*

$$H_M(t) = \frac{f(t)}{(1-t)^{n+1}}.$$

б) *В тех же условиях существует такой многочлен $P(k)$ с рациональными коэффициентами и такое число N, что*

$$d_k(M) = P(k) \text{ для всех } k \geqslant N.$$

Доказательство. Выведем сначала второе утверждение из первого. Положим $f(t) = \sum_{i=0}^{N} a_i t^i$ и приравняем коэффициенты при t^k в правой и левой частях тождества

$$H_M(t) = f(t)(1-t)^{-(n+1)}.$$

Получим, учитывая, что $(1-t)^{-(n+1)} = \sum_{k=0}^{\infty} \binom{n+k}{n} t^k$:

$$d_k(M) = \sum_{i=0}^{\min(k,N)} a_i \binom{n+k-i}{n}.$$

При $k \geqslant N$ выражение справа есть многочлен от k с рациональными коэффициентами.

Теперь индукцией по n докажем утверждение а). Удобно положить $A^{(-1)} = \mathscr{K} = A_0^{(-1)}$; $A_i^{(-1)} = \{0\}$ для $i \geqslant 1$. Конечно порожденный градуированный модуль над $A^{(-1)}$ — это просто конечномерное векторное пространство над $\mathscr{K}$, представленное в виде прямой суммы $\sum_{k=1}^{N} M_k$. Его ряд Пуанкаре есть многочлен $\sum_{k=0}^{N} \dim M_k t^k$, так что результат тривиально верен.

Пусть теперь он доказан для $A^{(n-1)}$, $n \geqslant 0$; установим его для $A^{(n)}$. Пусть M — конечно порожденный градуированный модуль над $A^{(n)}$. Положим

$$K = \{m \in M \mid x_n m = 0\}, \quad C = M / x_n M.$$

Очевидно, K и $x_n M$ суть градуированные подмодули M; поэтому C также имеет структуру градуированного $A^{(n)}$-модуля. Но умножение на x_n аннулирует как K, так и C. Поэтому, если мы рассмотрим K и C как модули над подкольцом $A^{(n-1)} = \mathscr{K}[x_0, \ldots, x_{n-1}] \subset \subset A^{(n)} = \mathscr{K}[x_0, \ldots, x_n]$, то любая система образующих для них над $A^{(n)}$ будет в то же время системой образующих над $A^{(n-1)}$. По теореме п. 6 K конечно порожден над $A^{(n)}$ как подмодуль конечно порожденного модуля. С другой стороны, C конечно порожден над $A^{(n)}$, потому что если $m_1, \ldots, m_p$ порождают M, то $m_1 + x_n M, \ldots \ldots, m_p + x_n M$ порождают C. Следовательно, K и C конечно порождены над $A^{(n-1)}$, и к ним применимо индуктивное предположение. Из точных последовательностей линейных пространств над $\mathscr{K}$:

$$0 \longrightarrow K_m \longrightarrow M_m \xrightarrow{x_n} M_{m+1} \longrightarrow C_{m+1} \longrightarrow 0, \quad m \geqslant 0,$$

следует, что

$$\dim M_{m+1} - \dim M_m = \dim C_{m+1} - \dim K_m.$$

Умножив это равенство на t^{m+1} и просуммировав по m от 0 до ∞, получим

$$H_M(t) - \dim M_0 - t H_M(t) = H_C(t) - \dim C_0 - t H_K(t),$$

или по индуктивному предположению для K и C

$$(1 - t) H_M(t) = \dim M_0 - \dim C_0 + \frac{f_C(t)}{(1-t)^n} - \frac{t f_K(t)}{(1-t)^n},$$

где $f_C(t)$ и $f_K(t)$ — многочлены с целыми коэффициентами. Очевидно, отсюда следует требуемое.

10. Размерность и степень алгебраического многообразия. Пусть теперь $V \subset P^n$ — некоторое алгебраическое многообразие, которому отвечает идеал $I(V)$. Рассмотрим многочлен Гильберта $P_V(k)$ фактормодуля $A^{(n)}/I(V)$:

$$P_V(k) = \dim A_k^{(n)}/I_k(V) \text{ для всех } k \geqslant k_0.$$

Нетрудно видеть, что $P_{P^n}(k) = \binom{n+k}{n}$, так что $\deg P_{P^n}(k) = n$. Поэтому $\deg P_V \leqslant n$. Число $d = \deg P_V$ называется *размер-*

ностью многообразия V. Представим старший член $P_V(k)$ в виде $e\frac{k^d}{d!}$. Можно доказать, что e — целое число, которое называется *степенью* многообразия V. Размерность и степень — важнейшие характеристики «величины» алгебраического многообразия. Можно дать их чисто геометрическое определение: если поле $\mathscr{K}$ алгебраически замкнуто, то d-мерное многообразие степени e пересекается с «достаточно общим» проективным пространством $P^{n-d} \subset \subset P^n$ дополнительной размерности в точности по e разным точкам. Мы не будем доказывать эту теорему.

Заметим в заключение, что после открытия Гильберта около полувека оставался нерешенным вопрос, как следует интерпретировать значения многочлена Гильберта $P_V(k)$ для тех целых значений k, при которых $P_V(k) \neq \dim l_k(V)$ (в частности отрицательных k). Он был решен лишь в пятидесятых годах с созданием теории когомологий когерентных пучков, когда выяснилось, что при любом k значение $P_V(k)$ есть альтернированная сумма размерностей некоторых пространств когомологий многообразия V. Аналогичная интерпретация была дана многочленам Гильберта любых конечно порожденных градуированных модулей.

УПРАЖНЕНИЯ

1. Доказать, что многочлен Гильберта проективного пространства P^n не зависит от размерности m проективного пространства P^m, в которое P^n вложено: $P^n \subset P^m$.

2. Вычислить многочлен Гильберта модуля $A^{(n)}/FA^{(n)}$, где F — форма степени e.

Часть 4. ПОЛИЛИНЕЙНАЯ АЛГЕБРА

§ 1. Тензорное произведение линейных пространств

1. Последняя часть нашей книги посвящена систематическому изучению полилинейных конструкций линейной алгебры. Основой алгебраического аппарата служит понятие тензорного произведения, которое вводится в этом параграфе и подробно изучается дальше. К сожалению, главные приложения этого формализма лежат за пределами собственно линейной алгебры: они относятся к дифференциальной геометрии, теории представлений групп и квантовой механике. Мы лишь вкратце коснемся их в последних параграфах.

2. Конструкция. Рассмотрим конечное семейство векторных пространств $L_1, \ldots, L_p$ над одним и тем же полем скаляров $\mathcal{K}$. Напомним, что отображение $L_1 \times \ldots \times L_p \to L$, где L — еще одно пространство над $\mathcal{K}$, называется полилинейным, если оно линейно по каждому из аргументов $l_i \in L_i$, $i = 1, \ldots, p$, при фиксированных остальных.

Наша ближайшая цель — построить *универсальное* полилинейное отображение пространств $L_1, \ldots, L_p$. Его образ будет называться тензорным произведением этих пространств. Точный смысл утверждения об универсальности объяснен ниже, в формулировке теоремы п. 3. Конструкция состоит из трех шагов.

а) *Пространство* $\mathcal{M}$. Это множество всех финитных функций на $L_1 \times \ldots \times L_p$ со значениями в $\mathcal{K}$, т. е. теоретико-множественных отображений $L_1 \times \ldots \times L_p \to \mathcal{K}$, равных нулю во всех точках множества $L_1 \times \ldots \times L_p$, кроме конечного числа. Оно образует линейное пространство над $\mathcal{K}$ с обычными операциями поточечного сложения и умножения на скаляр. Его базисом являются дельта-функции $\delta(l_1, \ldots, l_p)$, равные 1 в единственной точке $(l_1, \ldots, l_p) \in L_1 \times \ldots \times L_p$ и нулю в остальных. Опуская знак δ, мы можем считать, что $\mathcal{M}$ состоит из формальных конечных линейных комбинаций семейств $(l_1, \ldots, l_p) \in L_1 \times \ldots \times L_p$:

$$\mathcal{M} = \{\sum a_{l_1 \ldots l_p}(l_1, \ldots, l_p) \mid a_{l_1 \ldots l_p} \in \mathcal{K}\}.$$

Заметим, что если поле $\mathcal{K}$ бесконечно и хотя бы одно из пространств L_i не нульмерно, то $\mathcal{M}$ — бесконечномерное пространство.

б) *Подпространство* $\mathcal{M}_0$. По определению оно порождено всеми векторами из $\mathcal{M}$ вида

$$(l_1, \ldots, l'_j + l''_j, \ldots, l_p) - (l_1, \ldots, l'_j, \ldots, l_p) - (l_1, \ldots, l''_j, \ldots, l_p),$$
$$(l_1, \ldots, al_j, \ldots, l_p) - a(l_1, \ldots, l_j, \ldots, l_p), \qquad a \in \mathcal{K}.$$

в) *Тензорное произведение* $L_1 \otimes \ldots \otimes L_p$. По определению

$$L_1 \otimes \ldots \otimes L_p = \mathcal{M}/\mathcal{M}_0,$$
$$l_1 \otimes \ldots \otimes l_p = (l_1, \ldots, l_p) + \mathcal{M}_0 \in L_1 \otimes \ldots \otimes L_p,$$
$$t\colon L_1 \times \ldots \times L_p \to L_1 \otimes \ldots \otimes L_p, \quad t(l_1, \ldots, l_p) = l_1 \otimes \ldots \otimes l_p.$$

Здесь $\mathcal{M}/\mathcal{M}_0$ — факторпространство в обычном смысле слова. Элементы $L_1 \otimes \ldots \otimes L_p$ называются тензорами, $l_1 \otimes \ldots \otimes l_p$ — разложимыми тензорами. Поскольку семейства $(l_1, \ldots, l_p)$ составляют базис $\mathcal{M}$, разложимые тензоры $l_1 \otimes \ldots \otimes l_p$ порождают все тензорное произведение $L_1 \otimes \ldots \otimes L_p$, но отнюдь не являются базисом: между ними есть много линейных зависимостей.

Основное свойство тензорных произведений описано в следующей теореме:

3. Теорема. а) *Каноническое отображение*

$$t\colon L_1 \times \ldots \times L_p \to L_1 \otimes \ldots \otimes L_p, \ (l_1, \ldots, l_p) \mapsto l_1 \otimes \ldots \otimes l_p,$$

является полилинейным.

б) *Полилинейное отображение t универсально в следующем смысле слова: для любого линейного пространства M над полем $\mathcal{K}$ и любого полилинейного отображения $s\colon L_1 \times \ldots \times L_p \to M$ существует единственное линейное отображение $f\colon L_1 \otimes \ldots \otimes L_n \to M$ такое, что $s = f \circ t$.* Мы будем кратко говорить, что s *проводится через* f.

Доказательство. а) Мы должны проверить следующие формулы:

$$l_1 \otimes \ldots \otimes (l'_j + l''_j) \otimes \ldots \otimes l_p =$$
$$= l_1 \otimes \ldots \otimes l'_j \otimes \ldots \otimes l_p + l_1 \otimes \ldots \otimes l''_j \otimes \ldots \otimes l_r,$$
$$l_1 \otimes \ldots \otimes (al_j) \otimes \ldots \otimes l_p = a(l_1 \otimes \ldots \otimes l_j \otimes \ldots \otimes l_p),$$

т. е., например, для первой формулы

$$(l_1, \ldots, l'_j + l''_j, \ldots, l_p) + \mathcal{M}_0 = [(l_1, \ldots, l'_j, \ldots, l_p) + \mathcal{M}_0] +$$
$$+ [(l_1, \ldots, l''_j, \ldots, l_p) + \mathcal{M}_0].$$

Вспоминая определение факторпространства (п. 2 и 3 § 6 ч. 1), и систему образующих подпространства $\mathcal{M}_0$, описанную на шаге б) в п. 2 этого параграфа, немедленно получаем эти равенства из определений.

б) Если f вообще существует, то условие $s = f \circ t$ однозначно определяет значения f на разложимых тензорах:

$$f(l_1 \otimes \ldots \otimes l_p) = f \circ t(l_1, \ldots, l_p) = s(l_1, \ldots, l_p).$$

Поскольку последние порождают $L_1 \otimes \ldots \otimes L_p$, отображение f единственно.

Для доказательства существования f рассмотрим линейное отображение $g\colon \mathscr{M} \to M$, которое на базисных элементах $\mathscr{M}$ определено формулой

$$g(l_1, \ldots, l_p) = s(l_1, \ldots, l_p),$$

т. е.

$$g\left(\sum a_{l_1 \ldots l_p}(l_1, \ldots, l_p)\right) = \sum a_{l_1 \ldots l_p} s(l_1, \ldots, l_p).$$

Нетрудно убедиться, что $\mathscr{M}_0 \subset \operatorname{Ker} g$. Действительно,

$$g[(l_1, \ldots, l'_j + l''_j, \ldots, l_p) - (l_1, \ldots, l'_j, \ldots, l_p) - (l_1, \ldots, l''_j, \ldots, l_p)] =$$
$$= s(l_1, \ldots, l'_j + l''_j, \ldots, l_p) - s(l_1, \ldots, l'_j, \ldots, l_p) -$$
$$- s(l_1, \ldots, l''_j, \ldots, l_p) = 0$$

в силу полилинейности s. Аналогично проверяется, что g аннулирует образующие $\mathscr{M}_0$ второго типа, связанные с умножением одной из компонент на скаляр.

Отсюда вытекает, что g индуцирует линейное отображение

$$f\colon \mathscr{M}/\mathscr{M}_0 = L_1 \otimes \ldots \otimes L_p \to M$$

(см. предложение п. 8 § 6 ч. 1), для которого

$$f(l_1 \otimes \ldots \otimes l_p) = s(l_1, \ldots, l_p).$$

Это завершает доказательство.

Теперь мы приведем несколько непосредственных следствий этой теоремы и первые приложения нашей конструкции.

4. Пусть $\mathscr{L}(L_1, \ldots, L_p; M)$ — множество полилинейных отображений $L_1 \times \ldots \times L_p$ в M. В теореме п. 3 мы построили отображение множеств

$$\mathscr{L}(L_1, \ldots, L_p; M) \to \mathscr{L}(L_1 \otimes \ldots \otimes L_p; M),$$

ставящее в соответствие полилинейному отображению s линейное отображение f со свойством $s = f \circ t$. Но левая и правая части являются линейными пространствами над $\mathscr{K}$ (как пространства функций со значениями в векторном пространстве M: сложение и умножение на скаляр производится поточечно). Из конструкции очевидно, что наше отображение является линейным. Более того, оно является изоморфизмом линейных пространств. Действительно, сюръективность следует из того, что для любого линейного отображения $f\colon L_1 \otimes \ldots \otimes L_p \to M$ отображение $s = f \circ t$ полилинейно в силу утверждения а) теоремы п. 3. Инъективность следует из того, что если $s \neq 0$, то $f \circ t \neq 0$ и потому $f \neq 0$. Окончательно, мы получили каноническое отождествление линейных пространств

$$\mathscr{L}(L_1, \ldots, L_p; M) = \mathscr{L}(L_1 \otimes \ldots \otimes L_p; M).$$

Таким образом, конструкция тензорного произведения пространств сводит изучение полилинейных отображений к изучению линейных отображений путем введения новой операции на категории линейных пространств.

5. Размерность и базисы. а) *Если хоть одно из пространств $L_1, \ldots, L_p$ нулевое, то их тензорное произведение является нулевым.*

В самом деле, пусть, скажем, $L_j = 0$. Любое полилинейное отображение $f\colon L_1 \times \ldots \times L_p \to M$ при фиксированных $l_i \in L_i$, $i \neq j$, линейно на L_j; но единственное линейное отображение нулевого пространства само нулевое. Значит, $f = 0$ при всех значениях аргументов. В частности, универсальное полилинейное отображение $t\colon L_1 \times \ldots \times L_p \to L_1 \otimes \ldots \otimes L_p$ является нулевым. Но его образ порождает все тензорное произведение. Поэтому последнее нульмерно.

б) $\dim(L_1 \otimes \ldots \otimes L_p) = \dim L_1 \ldots \dim L_p$.

Если хоть одно из пространств нулевое, это следует из предыдущего результата. В противном случае будем рассуждать так: размерность $L_1 \otimes \ldots \otimes L_p$ совпадает с размерностью двойственного пространства $\mathscr{L}(L_1 \otimes \ldots \otimes L_p, \mathscr{K})$. В п. 4 мы отождествили его с пространством полилинейных отображений $\mathscr{L}(L_1 \times \ldots \times L_p, \mathscr{K})$. Выберем в пространствах L_i базисы $\{e_1^{(i)}, \ldots, e_{n_i}^{(i)}\}$. Всякому полилинейному отображению

$$f\colon L_1 \times \ldots \times L_p \to \mathscr{K}$$

поставим в соответствие набор из $n_1 \ldots n_p$ скаляров

$$f\left(e_{i_1}^{(1)}, \ldots, e_{i_p}^{(p)}\right), \qquad 1 \leqslant i_j \leqslant n_j,\ 1 \leqslant j \leqslant p.$$

В силу свойства полилинейности этот набор однозначно определяет f:

$$f\left(\sum_{i_1=1}^{n_1} x_{i_1}^{(1)} e_{i_1}^{(1)}, \ldots, \sum_{i_p=1}^{p} x_{i_p}^{(p)} e_{i_p}^{(p)}\right) = \sum_{(i_p \ldots i_p)} x_{i_1}^{(1)} \ldots x_{i_p}^{(p)} f\left(e_{i_1}^{(1)}, \ldots, e_{i_p}^{(p)}\right).$$

Кроме того, он может быть произвольным: правая часть последней формулы определяет полилинейное отображение векторов $\left(\vec{x}^{(1)}, \ldots \ldots, \vec{x}^{(p)}\right)$ при любых значениях коэффициентов. Это означает, что пространство полилинейных отображений $L_1 \times \ldots \times L_p \to \mathscr{K}$ имеет размерность $n_1 \ldots n_p = \dim L_1 \ldots \dim L_p$, что завершает доказательство.

в) *Тензорный базис $L_1 \otimes \ldots \otimes L_p$.* Предшествующее рассуждение позволяет установить также, что тензорные произведения $\left\{e_{i_1}^{(1)} \otimes \ldots \otimes e_{i_p}^{(p)}\right\}$ образуют базис пространства $L_1 \otimes \ldots \otimes L_p$ (считаем, что все пространства L_i имеют размерность $\geqslant 1$ и, для простоты, конечномерны). В самом деле, эти тензорные произведения порождают $L_1 \otimes \ldots \otimes L_p$, ибо через них линейно выражаются все разложимые тензоры. Кроме того, их количество в точности равно размерности $L_1 \otimes \ldots \otimes L_p$.

6. Тензорные произведения пространств функций. Пусть $S_1, \ldots$ $\ldots, S_p$ — конечные множества, $F(S_i)$ — пространство функций на S_i со значениями в $\mathcal{K}$. Тогда имеется каноническое отождествление

$$F(S_1 \times \ldots \times S_p) = F(S_1) \otimes \ldots \otimes F(S_p),$$

которое ставит в соответствие функции $\delta_{(s_1, \ldots, s_p)}$ элемент $\delta_{s_1} \otimes \ldots$ $\ldots \otimes \delta_{s_p}$ (см. п. 7 § 1 ч. 1). Поскольку оба этих семейства образуют базис своих пространств, это действительно изоморфизм. Если $f_i \in F(S_i)$, то

$$f_1 \otimes \ldots \otimes f_p = \Big(\sum_{s_1 \in S_1} f_1(s_1)\,\delta_{s_1}\Big) \otimes \ldots \otimes \Big(\sum_{s_p \in S_p} f_p(s_p)\,\delta_{s_p}\Big)$$

перейдет при этом изоморфизме в функцию

$$(s_1, \ldots, s_p) \mapsto f_1(s_1) \ldots f_p(s_p),$$

т. е. разложимые тензоры отвечают «разделяющимся переменным».

Если $S_1 = \ldots = S_p = S$, то тензорное произведение функций на S соответствует обычному произведению их значений «в независимых точках S».

Именно в таком контексте тензорные произведения чаще всего появляются в функциональном анализе и физике. Однако алгебраическое определение тензорного произведения подвергается в функциональном анализе существенным изменениям, связанным с учетом топологии пространств; в частности, его обычно приходится пополнять по разным топологиям.

7. Подъем поля скаляров. Пусть L — линейное пространство над полем $\mathbf{R}$, $L^{\mathbf{C}}$ — его комплексификация (см. § 12 ч. 1). Поскольку поле $\mathbf{C}$ можно рассматривать как линейное пространство над $\mathbf{R}$ (с базисом $1, i$), мы можем построить линейное пространство $\mathbf{C} \otimes L$, порожденное базисом $1 \otimes e_1, \ldots, 1 \otimes e_n, i \otimes e_1, \ldots$ $\ldots, i \otimes e_n$, над $\mathbf{R}$, где $\{e_1, \ldots, e_n\}$ — базис L. Ясно, что $\mathbf{R}$ — линейное отображение

$$\mathbf{C} \otimes L \to L^{\mathbf{C}}: 1 \otimes e_j \mapsto e_j, \quad i \otimes e_j \mapsto ie_j$$

определяет изоморфизм $\mathbf{C} \otimes L$ с $L^{\mathbf{C}}$.

Более общо, пусть $\mathcal{K} \subset K$ — поле и его подполе, L — линейное пространство над $\mathcal{K}$. Рассмотрев сначала K как линейное пространство над $\mathcal{K}$, построим тензорное произведение $K \otimes L$. После этого введем на нем структуру линейного пространства над K, определив умножение на скаляры $a \in K$ формулой

$$a(b \otimes l) = ab \otimes l; \quad a, b \in K, \quad l \in L.$$

Чтобы проверить корректность этого определения, построим пространство $\mathcal{M}$, свободно порожденное элементами $K \times L$, и его подпространство $\mathcal{M}_0$, как в п. 2, так что $K \otimes L = \mathcal{M}/\mathcal{M}_0$. Определим умножение на скаляры из K в $\mathcal{M}$, положив на базисных элементах

$$a(b, l) = (ab, l); \quad a, b \in K, \quad l \in L,$$

и распространив это правило на остальные элементы $K \times L$ по $\mathscr{K}$-линейности. Непосредственная проверка показывает, что $\mathscr{M}$ превращается в K-линейное пространство, а $\mathscr{M}_0$ — в его подпространство, так что $\mathscr{M}/\mathscr{M}_0 = K \otimes L$ также становится линейным пространством над K. Это и есть общая конструкция подъема поля скаляров, упомянутая в п. 15 § 11 ч. 1.

Важный частный случай: при $K = \mathscr{K}$ линейное пространство $\mathscr{K} \otimes L$ над $\mathscr{K}$ канонически изоморфно L. Этот изоморфизм переводит $a \otimes l$ в al.

§ 2. Канонические изоморфизмы и линейные отображения тензорных произведений

1. Тензорное умножение обладает некоторыми алгебраическими свойствами операций, называемых умножениями в других контекстах, например, ассоциативностью. Однако в формулировке этих свойств имеется своя специфика из-за того, что тензорное умножение есть операция над объектами категории. Например, пространства $(L_1 \otimes L_2) \otimes L_3$ и $L_1 \otimes (L_2 \otimes L_3)$ *не совпадают*, как явствует из сравнения их конструкции: они лишь связаны *канонически определенным изоморфизмом.*

В этом параграфе мы опишем ряд таких «элементарных» изоморфизмов, очень полезных при работе с тензорными произведениями. Предупредим читателя, однако, что мы вынуждены будем ограничиться лишь введением в теорию канонических изоморфизмов. Главный вопрос, систематическое исследование которого мы опустим, состоит в их совместности. Предположим, например, что у нас есть два естественных изоморфизма между некоторыми тензорными произведениями, по-разному скомпонованных из нескольких «элементарных» естественных изоморфизмов. Обязательно ли эти изоморфизмы совпадут? Можно проводить непосредственную проверку в каждом конкретном частном случае или попытаться построить общую теорию, которая оказывается довольно громоздкой.

Аналогичные задачи возникают в связи с естественными отображениями, которые не являются изоморфизмами, например такими, как симметризация или свертка.

2. Ассоциативность. Пусть $L_1, \ldots, L_p$ — линейные пространства над $\mathscr{K}$. Мы хотим построить канонические изоморфизмы между пространствами вида $(L_1 \otimes L_2)(\ldots \otimes L_p)$, получающимися в результате тензорного умножения $L_1, \ldots, L_p$ группами в разном порядке, который устанавливается расстановкой скобок. Самый удобный способ, автоматически обеспечивающий совместимость, состоит в том, чтобы построить для каждой расстановки скобок линейное отображение $L_1 \otimes \ldots \otimes L_p \to (L_1 \otimes L_2)(\ldots \otimes L_p)$ с помощью универсального свойства из утверждения б) п. 3 § 1 и проверить, что оно является изоморфизмом.

Мы подробно рассмотрим конструкцию $L_1 \otimes L_2 \otimes L_3 \to (L_1 \otimes L_2) \otimes \otimes L_3$; общий случай совершенно аналогичен.

Отображение $L_1 \times L_2 \to L_1 \otimes L_2$: $(l_1, l_2) \longmapsto l_1 \otimes l_2$ билинейно. Поэтому отображение $L_1 \times L_2 \times L_3 \to (L_1 \otimes L_2) \otimes L_3$: $(l_1, l_2, l_3) \longmapsto$ $\longmapsto (l_1 \otimes l_2) \otimes l_3$ трилинейно. Значит, его можно провести через единственное линейное отображение $L_1 \otimes L_2 \otimes L_3 \to (L_1 \otimes L_2) \otimes L_3$. По самой конструкции последнее отображение переводит $l_1 \otimes l_2 \otimes l_3$ в $(l_1 \otimes l_2) \otimes l_3$. Выбрав в пространствах L_1, L_2, L_3 базисы и воспользовавшись результатом п. 5 § 1, получаем, что это отображение переводит базис в базис и поэтому является изоморфизмом.

Окончательно: *произведение* $(L_1 \otimes L_2)(\ldots \otimes L_p)$ *с любой расстановкой скобок можно отождествить с* $L_1 \otimes L_2 \otimes \ldots \otimes L_p$, *просто опустив все скобки*; *на элементах* $(l_1 \otimes l_2)(\ldots \otimes l_p)$ *это отождествление действует по тому же правилу.* Поэтому мы позволим себе писать $(l_1 \otimes l_2) \otimes l_3 = l_1 \otimes l_2 \otimes l_3 = l_1 \otimes (l_2 \otimes l_3)$ и т. п.

3. Коммутативность. Пусть σ — любая перестановка чисел $1, \ldots, p$. Определим систему изоморфизмов

$$f_\sigma\colon L_1 \otimes \ldots \otimes L_p \to L_{\sigma(1)} \otimes \ldots \otimes L_{\sigma(p)}$$

со свойством $f_{\sigma\tau} = f_\sigma \circ f_\tau$ для любых σ, τ. Для этого заметим, что отображение

$$L_1 \times \ldots \times L_p \to L_{\sigma(1)} \otimes \ldots \otimes L_{\sigma(p)}\colon (l_1, \ldots, l_p) \longmapsto l_{\sigma(1)} \otimes \ldots \otimes l_{\sigma(p)}$$

полилинейно. Поэтому оно проводится через отображение f_σ: $L_1 \otimes \ldots \otimes L_p \to L_{\sigma(1)} \otimes \ldots \otimes L_{\sigma(p)}$ в силу утверждения б) теоремы п. 3 § 1. На произведениях векторов оно действует очевидным образом, переставляя сомножители, и рассмотрение его действия на тензорном произведении базисов $L_1, \ldots, L_p$ показывает, что это изоморфизм. Свойство $f_{\sigma\tau} = f_\sigma \circ f_\tau$ очевидно.

Таким образом, мы определили действие симметрической группы S_p на $L_1 \otimes \ldots \otimes L_p$. В случае, когда все пространства L_i разные, можно пользоваться изоморфизмами f_σ для однозначного отождествления $L_1 \otimes \ldots \otimes L_p$ с $L_{\sigma(1)} \otimes \ldots \otimes L_{\sigma(p)}$. В этом смысле тензорное умножение коммутативно. Однако записывать это отождествление в виде равенства, не указывая явно f_σ (как мы делали для ассоциативности), опасно, если среди пространств $L_1, \ldots, L_p$ есть одинаковые.

4. Двойственность. Имеется канонический изоморфизм

$$L_1^* \otimes \ldots \otimes L_p^* \to (L_1 \otimes \ldots \otimes L_p)^*.$$

Чтобы построить его, поставим в соответствие каждому элементу $(f_1, \ldots, f_p) \in L_1^* \times \ldots \times L_p^*$ полилинейную функцию $f_1(l_1) \ldots f_p(l_p)$ от $(l_1, \ldots, l_p) \in L_1 \times \ldots \times L_p$. В силу утверждения б) теоремы п. 3 это отображение проводится через отображение $L_1^* \otimes \ldots \otimes L_p^* \to$ (пространство полилинейных функций на $L_1 \times \ldots \times L_p$). Последнее пространство в силу конструкции п. 4 § 1 отождествляется с пространством $\mathscr{L}(L_1 \otimes \ldots \otimes L_p, \mathscr{K}) = (L_1 \otimes \ldots \otimes L_p)^*$. Таким образом, мы построили искомое отображение. Чтобы показать, что оно является изоморфизмом, заметим, что размерности пространств

$(L_1 \otimes \ldots \otimes L_p)^*$ и $L_1^* \otimes \ldots \otimes L_p^*$ совпадают (мы ограничиваемся конечномерными L_i). Поэтому достаточно проверить, что наше отображение сюръективно. Но в его образе содержатся функции $f_1(l_1) \ldots f_p(l_p)$, где f_i пробегают некоторый базис L_i^*, и, как в п. 5 § 1, нетрудно убедиться, что они образуют базис $\mathscr{L}(L_1, \ldots, L_p; \mathscr{K}) = (L_1 \otimes \ldots \otimes L_p)^*$.

Отождествление $(L_1 \otimes \ldots \otimes L_p)^*$ с $L_1^* \otimes \ldots \otimes L_p^*$ с помощью описанного изоморфизма обычно безобидно.

5. **Изоморфизм $\mathscr{L}(L, M)$ с $L^* \otimes M$.** Рассмотрим билинейное отображение

$$L^* \times M \to \mathscr{L}(L, M)\colon (f, m) \mapsto [l \mapsto f(l)\, m],$$

где $f \in L^*$, $l \in L$, $m \in M$. Как билинейность выражения $f(l)m$ по f и m, так и его линейность по l очевидны. В силу многократно использованного свойства универсальности ему отвечает линейное отображение

$$L^* \otimes M \to \mathscr{L}(L, M).$$

Выберем в L, M базисы $\{l_1, \ldots, l_a\}$, $\{m_1, \ldots, m_b\}$ и в L^* — двойственный базис $\{l^1, \ldots, l^a\}$. Элемент $l^i \otimes m_j$ тензорного произведения базисов L^*, M переходит в линейное отображение, которое переводит вектор $l_k \in L$ в $l^i(l_k)\, m_j = \delta_{ik} m_j$. Матрица этого линейного отображения размера $b \times a$ имеет единицу на месте (ji) и нули на остальных местах. Поскольку такие матрицы образуют базис $\mathscr{L}(L, M)$, построенное отображение переводит базис в базис и является изоморфизмом.

Рассмотрим важный частный случай: $L = M$. Здесь

$$\mathscr{L}(L, L) = L^* \otimes L.$$

Пространство эндоморфизмов $\mathscr{L}(L, L)$ содержит выделенный элемент: тождественное отображение id_L. Его образ в $L^* \otimes L$, как видно из предыдущих рассуждений, равен

$$\sum_{k=1}^{a} l^k \otimes l_k,$$

где $\{l_k\}$, $\{l^k\}$ — пара двойственных базисов в L и L^*. Таким образом, формула для этого элемента имеет одинаковый вид во всех парах двойственных базисов.

Кроме того, пространство эндоморфизмов $\mathscr{L}(L, L)$ снабжено каноническим линейным функционалом следа $\mathrm{Tr}\colon \mathscr{L}(L, L) \to \mathscr{K}$. Из прежних рассуждений следует, что след отображения, которое поставлено в соответствие элементу $l^i \otimes l_j$, равен δ_{ij} (посмотрите на матрицу), так что общему элементу тензорного произведения $L^* \otimes L$ ставится в соответствие число

$$\sum_{i, j=1}^{a} a_i^j l^i \otimes l_j \mapsto \sum_{i=1}^{a} a_i^i.$$

Этот линейный функционал $L^* \otimes L \to \mathscr{K}$ называется *сверткой*. Позже мы дадим определение свертки в более общем контексте.

6. Изоморфизм $\mathscr{L}(L\otimes M, N)$ с $\mathscr{L}(L, \mathscr{L}(M, N))$. Пространство $\mathscr{L}(L\otimes M, N)$ изоморфно пространству билинейных отображений $L\times M\to N$. Каждое такое билинейное отображение f: $(l, m)\to \to f(l, m)$ при фиксированном первом аргументе l представляет собой линейное отображение $M\to N$; от l это отображение зависит линейно. Таким образом, получаем каноническое линейное отображение

$$\mathscr{L}(L\otimes M,\ N)=\mathscr{L}(L,\ M;\ N)\to\mathscr{L}(L,\ \mathscr{L}(M,\ N)).$$

Рассуждение с базисами в L, M, N, аналогичное проведенному в предыдущем пункте, показывает, что оно является изоморфизмом (как всюду, пространства предполагаются конечномерными).

(Это отождествление является важным примером общекатегорного понятия «функторов, сопряженных по Кану».)

7. Тензорное произведение линейных отображений. Пусть $L_1, \ldots, L_p$; $M_1, \ldots, M_p$ — два семейства линейных пространств, f_i: $L_i\to M_i$ — линейные отображения. Тогда можно построить линейное отображение

$$f_1\otimes\ldots\otimes f_p\colon\ L_1\otimes\ldots\otimes L_p\to M_1\otimes\ldots\otimes M_p,$$

называемое тензорным произведением f_i и однозначно характеризуемое простым свойством

$$(f_1\otimes\ldots\otimes f_p)(l_1\otimes\ldots\otimes l_p)=f_1(l_1)\otimes\ldots\otimes f_p(l_p)$$

для всех $l_i\in L_i$. Его существование доказывается все тем же стандартным применением утверждения б) теоремы п. 3 § 1, если заметить, что отображение

$$L_1\times\ldots\times L_p\to M_1\otimes\ldots\otimes M_p\colon\ (l_1,\ \ldots,\ l_p)\mapsto f_1(l_1)\otimes\ldots\otimes f_p(l_p)$$

полилинейно.

Если все f_i суть изоморфизмы, то и $f_1\otimes\ldots\otimes f_p$ является изоморфизмом.

8. Свертка и подъем индексов. С помощью этой конструкции мы можем дать общее определение свертки «по паре или нескольким парам индексов». Пусть у нас имеется тензорное произведение $L_1\otimes\ldots\otimes L_p$, причем для некоторых двух индексов $i, j\in\{1,\ldots,p\}$ имеем $L_i=L^*$, $L_j=L$. *Свертка по индексам i, j* есть линейное отображение

$$L_1\otimes\ldots\otimes L_p\to\bigotimes_{\substack{k=1\\ k\neq i,\,j}}^{p} L_k,$$

которое получается как композиция следующих линейных отображений:

а) Отображение f_σ, где σ — перестановка индексов $\{1, \ldots, p\}$, переводящая i в 1, j в 2 и сохраняющая порядок остальных индексов:

$$f_\sigma\colon L_1 \otimes \ldots \otimes L_p \to L_i \otimes L_j \otimes \left(\bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k \right) = L^* \otimes L \otimes \left(\bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k \right).$$

б) Свертка первых двух множителей, тензорно умноженная на тождественное отображение остальных:

$$L^* \otimes L \otimes \left(\bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k \right) \to \mathcal{K} \otimes \left(\bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k \right).$$

в) Отождествление

$$\mathcal{K} \otimes \left(\bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k \right) \rightleftarrows \bigotimes_{\substack{k=1 \\ k \ne i,\, j}}^{p} L_k.$$

Если имеется несколько пар индексов $(i_1, j_1), \ldots, (i_r, j_r)$ таких, что $L_{i_k} = M_k^*$, $L_{j_k} = M_k$, то эту конструкцию можно повторить несколько раз в применении ко всем парам последовательно. Результирующее линейное отображение называется сверткой по этим парам индексов. Оно зависит от самих пар, но не от порядка проведения сверток по ним. Может оказаться, что $\{1, \ldots, p\} = \{i_1, j_1, \ldots, i_r, j_r\}$. Тогда получится *полная свертка.*

Снова рассмотрим тензорное произведение $L_1 \otimes \ldots \otimes L_p$ и предположим, что для i-го пространства задан изоморфизм $g\colon L_i \to L_i^*$ (в приложениях он чаще всего строится с помощью невырожденной симметричной билинейной формы на L_i). Тогда линейное отображение

$$\mathrm{id} \otimes \ldots \otimes g \otimes \ldots \otimes \mathrm{id}\colon L_1 \otimes \ldots \otimes L_i \otimes \ldots \otimes L_p \to$$
$$\to L_1 \otimes \ldots \otimes L_i^* \otimes \ldots \otimes L_p$$

называется «опусканием i-го индекса», обратное к нему — «подъемом i-го индекса». Объяснение термина будет дано в следующем параграфе.

Обе конструкции, свертки и подъема/опускания индекса, чаще всего применяются в случае $L_i = L$ или L^*, когда на L задана ортогональная структура. Возникает масса линейных отображений, связывающих пространства $L^{*\otimes p} \otimes L^{\otimes q}$, которые строятся как композиции поднятия и опускания индексов и сверток. Эти отображения играют большую роль в римановой геометрии, где с их помощью (и с помощью аналитических операций типа дифференцирования) строятся важнейшие дифференциально-геометрические инварианты.

9. Тензорное умножение как точный функтор. Фиксируем линейное пространство M и рассмотрим отображение категории конечномерных линейных пространств в себя: $L \mapsto L \otimes M$ на объектах,

$f \mapsto f \otimes \mathrm{id}_M$ на морфизмах. Из определений легко усмотреть, что $\mathrm{id}_L \mapsto \mathrm{id}_{L\otimes M}$ и

$$f \circ g \mapsto f \circ g \otimes \mathrm{id}_M = (f \otimes \mathrm{id}_M) \circ (g \otimes \mathrm{id}_M).$$

Поэтому данное отображение является *функтором*, который называется функтором тензорного умножения на M.

Покажем, что если последовательность $0 \longrightarrow L_1 \xrightarrow{f} L \xrightarrow{g} L_2 \longrightarrow 0$ точна, то и последовательность

$$0 \to L_1 \otimes M \xrightarrow{f \otimes \mathrm{id}_M} L \otimes M \xrightarrow{g \otimes \mathrm{id}_M} L_2 \otimes M \to 0$$

точна. Это свойство называется *точностью функтора тензорного умножения*. Как и точность функтора $\mathscr{L}$, оно *нарушается* в категориях модулей, и это нарушение служит важным объектом изучения в гомологической алгебре: ср. обсуждение в § 14 ч. 1.

Проще всего проверить точность, выбрав в L_1, L, L_2 базисы, приспособленные к f, g таким образом, что $\{e_1, \ldots, e_a\}$ — базис L_1, $\{f(e_1), \ldots, f(e_a);\ e'_{a+1}, \ldots, e'_{a+b}\}$ — базис L; $\{g(e'_{a+1}), \ldots, g(e'_{a+b})\}$ — базис L_2. Выбрав еще базис $\{e''_1, \ldots, e''_c\}$ пространства M, получим, что тензорные произведения базисов

$$\{e_i \otimes e''_j\},\ \{f(e_i) \otimes e''_j,\ e'_k \otimes e''_j\},\ \{g(e'_k) \otimes e''_j\}$$

приспособлены к $f \otimes \mathrm{id}_M$, $g \otimes \mathrm{id}_M$ в том же смысле слова.

§ 3. Тензорная алгебра линейного пространства

1. Пусть L — некоторое конечномерное линейное пространство над полем $\mathscr{K}$. Любой элемент тензорного произведения

$$T_p^q(L) = \underbrace{L^* \otimes \ldots \otimes L^*}_{p} \otimes \underbrace{L \otimes \ldots \otimes L}_{q}$$

называется *тензором на L типа* (p, q) *и валентности* (или *ранга*) $p+q$. Говорят также, что он является смешанным тензором, p раз *ковариантным* и q раз *контравариантным*. Первые две части книги фактически были посвящены изучению следующих тензоров малого ранга.

а) Удобно положить $T_0^0(L) = \mathscr{K}$, т. е. называть скаляры тензорами ранга 0.

б) $T_1^0(L) = L^*$, т. е. тензоры типа $(1, 0)$ суть линейные функционалы на L. Тензоры типа $(0, 1)$ суть просто векторы из L.

в) $T_1^1(L) = L^* \otimes L$. В п. 5 § 2 мы отождествили $L^* \otimes L$ с пространством $\mathscr{L}(L, L)$. Следовательно, тензоры типа $(1, 1)$ «суть» линейные операторы на L.

г) $T_2^0(L) = L^* \otimes L^*$. В п. 4 § 2 мы отождествили $L^* \otimes L^*$ с $(L \otimes L)^*$, или с билинейными отображениями $L \times L \to \mathscr{K}$. Таким образом, тензоры типа $(2, 0)$ «суть» скалярные произведения на L. В п. 5 § 2 мы отождествили $L^* \otimes L^*$ с $\mathscr{L}(l^{**}, L^*) \simeq \mathscr{L}(L, L^*)$. При этом отождествлении скалярному произведению на L ставится в соответствие линейное отображение $L \to L^*$, которое отвечает рассмотрению этого скалярного произведения как функции от одного из своих аргументов при фиксированном втором. Таким образом, наши тензорные конструкции § 2 обобщают конструкции части 2.

д) Приведем еще один пример: структурный тензор $\mathcal{K}$-алгебры. Здесь мы будем понимать под $\mathcal{K}$-алгеброй линейное пространство L вместе с билинейной операцией умножения $L \times L \to L$: $(l, m) \to \to lm$, не обязательно коммутативной или даже ассоциативной, так что, например, алгебры Ли подходят под это определение.

Согласно утверждению б) теоремы п. 3 § 1 умножение можно определить также как линейное отображение $L \otimes L \to L$. В п. 5 § 2 мы отождествили пространство $\mathcal{L}(L \otimes L, L)$ с $(L \otimes L)^* \otimes L$ или, пользуясь еще п. 4 § 2 и ассоциативностью, с $L^* \otimes L^* \otimes L$. Следовательно, задать на пространстве L структуру $\mathcal{K}$-алгебры — это все равно, что задать на нем тензор типа $(2, 1)$, называемый *структурным тензором* этой алгебры.

2. Тензорное умножение. В соответствии с п. 4 § 2 мы можем отождествить пространство $T_p^q(L)$ с $\left(L^{\otimes p} \otimes (L^*)^{\otimes q}\right)^*$ и затем с пространством полилинейных отображений

$$f\colon \underbrace{L \times \ldots \times L}_{p} \times \underbrace{L^* \times \ldots \times L^*}_{q} \to \mathcal{K}.$$

Два таких полилинейных отображения типа (p, q) и (p', q') можно тензорно перемножить, получив в результате полилинейное отображение типа $(p + p', q + q')$:

$$(f \otimes g)(l_1, \ldots, l_p; l_1', \ldots, l_{p'}'; l_1^*, \ldots, l_q^*; l_1'^*, \ldots, l_{q'}'^*) =$$
$$= f(l_1, \ldots, l_p; l_1'^*, \ldots, l_q'^*)\, g(l_1', \ldots, l_{p'}'; l_1'^*, \ldots, l_{q'}'^*)$$

где $l_i, l_j' \in L$, $l_i^*, l_i'^* \in L^*$. Из этого определения сразу же видна билинейность тензорного умножения по его аргументам:

$$(af_1 + bf_2) \otimes g = a(f_1 \otimes g) + b(f_2 \otimes g);$$
$$f \otimes (ag_1 + bg_2) = a(f \otimes g_1) + b(f \otimes g_2),$$

а также его ассоциативность:

$$(f \otimes g) \otimes h = f \otimes (g \otimes h).$$

Однако оно не коммутативно: $f \otimes g$, вообще говоря, не совпадает с $g \otimes f$.

Если не переходить к интерпретации тензоров как полилинейных отображений, то тензорное умножение можно определить с помощью операций перестановки п. 3 § 2, с учетом ассоциативности, как отображение

$$f_\sigma\colon \underbrace{L^* \otimes \ldots \otimes L^*}_{p} \otimes \underbrace{L \otimes \ldots \otimes L}_{q} \otimes \underbrace{L^* \otimes \ldots \otimes L^*}_{p'} \otimes \underbrace{L \otimes \ldots \otimes L}_{q'} \to$$
$$\to \underbrace{L^* \otimes \ldots \otimes L^*}_{p+p'} \otimes \underbrace{L \otimes \ldots \otimes L}_{q+q'},$$

где σ переставляет третью группу из p' индексов на места после первой группы из p индексов, сохраняя их относительный порядок и относительный порядок остальных индексов. В этом варианте билинейность тензорного умножения столь же очевидна, а его ассоциативность превращается в некоторое тождество между подстановками, которое читателю легче увидеть самому, чем следить за длинными, но банальными объяснениями.

3. Тензорная алгебра пространства L. Положим

$$T(L)=\bigoplus_{p,\,q=1}^{\infty} T_p^q(L)$$

(прямая сумма линейных пространств). Это бесконечномерное пространство вместе с операцией тензорного умножения в нем, определенной в предыдущем пункте, называется тензорной алгеброй пространства L.

Заметим, что над полем комплексных чисел бывает важно рассматривать расширенную тензорную алгебру, являющуюся прямой суммой пространств $L^{\otimes p}\otimes L^{*\otimes q}\otimes \overline{L}^{\otimes p'}\otimes \overline{L}^{*\otimes q'}$. Например, полуторалинейная форма на L как тензор лежит в $L^*\otimes\bar{L}^*$. По недостатку места мы не будем систематически изучать эту конструкцию.

§ 4. Классические обозначения

1. В классическом тензорном анализе тензорный формализм описывается в координатных обозначениях. Ими и сейчас широко пользуются в физической и геометрической литературе, и этому языку следует отдать должное: он компактен и гибок. В этом параграфе мы введем его и покажем, как выражаются на нем различные конструкции, описанные выше.

2. Базисы и координаты. Пусть L — конечномерное линейное пространство. Выберем в нем базис $\{e_1, \ldots, e_n\}$ и будем задавать векторы L их координатами $(a^1, \ldots, a^n)$ в этом базисе: $\sum\limits_{i=1}^{n} a^i e_i$.

В L^* выберем двойственный базис $\{e^1, \ldots, e^n\}$, $(e^i, e_j)=\delta_j^i=0$ при $i\neq j$; 1 при $i=j$, и векторы из L^* будем задавать координатами $(b_1, \ldots, b_n)$: $\sum\limits_{j=1}^{n} b_j e^j$. Расположение индексов в обоих случаях выбирается так, чтобы в суммировании участвовали пары одинаковых индексов, один из которых находится наверху, другой внизу.

В $L^{*\otimes p}\otimes L^{\otimes q}$ построим тензорное произведение рассмотренных базисов

$$\left\{e^{i_1}\otimes\ldots\otimes e^{i_p}\otimes e_{j_1}\otimes\ldots\otimes e_{j_q}\,\middle|\, 1\leqslant i_k\leqslant n,\ 1\leqslant j_l\leqslant n\right\}.$$

Любой тензор $T \in T_p^q(L)$ задается в нем своими координатами $T_{i_1 \dots i_p}^{j_1 \dots j_q}$:

$$T = \sum T_{i_1 \dots i_p}^{j_1 \dots j_q} e^{i_1} \otimes \dots \otimes e^{i_p} \otimes e_{j_1} \otimes \dots \otimes e_{j_q}.$$

Заметим, что суммирование здесь снова распространено на пары одинаковых индексов, один из которых верхний, другой — нижний. Это настолько характерная черта классического формализма, что зачастую принимается соглашение опускать знак суммы во всех случаях, когда подразумевается такое суммирование.

В частности, при этом соглашении векторы из L записываются в виде $a^j e_j$, а функционалы в виде $b_i e^i$. Скалярное произведение между L^* и L, т. е. значение функционала $b_i e^i$ на векторе $a^j e_j$, запишется $a^i b_i$ или $b_i a^i$.

Более того, можно пойти еще дальше по этому пути экономии и не писать сами векторы e_i и e^i. Тогда элементы L записываются в виде a^j, элементы L^* — в виде b_i, а общий тензор $T \in T_p^q(L)$ — в виде $T_{i_1 \dots i_p}^{j_1 \dots j_q}$. Иными словами, в классической записи тензора T

явно указаны: координаты, или *компоненты* T в тензорном базисе $L^{*\otimes p} \otimes L^{\otimes q}$, пронумерованные как элементы тензорного базиса; номера являются сложными индексами; ковариантная часть индекса $(i_1, \dots, i_p)$ пишется внизу, а контравариантная $(j_1, \dots, j_q)$ — наверху;

подразумевается: выбор исходного базиса $\{e_1, \dots, e_n\}$ в L, по которому строится двойственный базис $\{e^1, \dots, e^n\}$ в L^* и затем тензорные базисы во всех пространствах $T_p^q(L)$.

Иногда удобно рассматривать тензоры, лежащие в пространствах, где сомножители L и L^* расположены в ином порядке, чем принятый нами, например, $L \otimes L^*$ вместо $L^* \otimes L$, или $L \otimes L^* \otimes \otimes L \otimes L$. Указание на это делается с помощью «блочного расположения» сложных индексов у координат тензора. Например, тензор $T \in L \otimes L^*$ можно задавать компонентами, которые обозначаются T_i^j, а $T \in L \otimes L^* \otimes L \otimes L$ — компонентами $T_{i_1}{}^{j_1}{}_{i_2 i_3}$.

3. Некоторые важные тензоры. К ним относятся:

а) *Метрический тензор* g_{ij}. Согласно нашим обозначениям он лежит в $T_2^0(L)$ и в силу г) п. 1 § 3 может представлять скалярное произведение на L. Его значение на паре векторов a^i, b^j равно $\sum g_{ij} a^i b^j$ или просто $g_{ij} a^i b^j$. Таким образом, компоненты метрического тензора — это элементы матрицы Грама исходного базиса L относительно соответствующего скалярного произведения.

б) *Матрица* A_j^i. Это — элемент $T_1^1(L)$, т. е., в силу в) п. 1 § 3, линейное отображение L в себя. Оно переводит вектор a^j в вектор с i-й координатой $\sum A_j^i a^j$ или просто $A_j^i a^j$. Тензор валентности $p+q$ можно представлять себе в виде «$p+q$-мерной матрицы», обычные матрицы плоские.

в) *Тензор Кронекера* δ^i_j. Это элемент $T^1_1(L)$, представляющий тождественное отображение L в себя.

г) *Структурный тензор алгебры.* Согласно д) п. 1 § 3 он лежит в $T^1_2(L)$ и потому записывается покомпонентно в виде γ^k_{ij}. Он задает билинейное умножение в L по формуле

$$a^i \cdot b^j = c^k = \gamma^k_{ij} a^i b^j.$$

Полная запись: $\left(\sum_i a^i e_i\right)\left(\sum_j b^j e_j\right) = \sum_k \left(\sum_{i,j} \gamma^k_{ij} a^i b^j\right) e_k.$

4. Преобразование компонент тензора при замене базиса в L. Пусть A^i_j — матрица замены базиса в L: $e'_k = A^i_k e_i$; B^i_j — матрица перехода от базиса $\{e^k\}$, двойственного $\{e_k\}$, к базису $\{e'^k\}$, двойственному $\{e'_k\}$. Нетрудно убедиться, что $B = (A^t)^{-1}$: эту матрицу называют *контраградиентной* к A.

Координаты a'^j в базисе $\{e'_j\}$ вектора, первоначально заданного координатами a^i в базисе $\{e_i\}$, будут $B^j_k a^k$.

Аналогично, координаты b_i в базисе $\{e'^i\}$ функционала (или «ковектора»), первоначально заданного координатами b_i в базисе $\{e^i\}$, будут $A^k_j b_k$.

Чтобы найти координаты $T'^{j_1 \dots j_q}_{i_1 \dots i_p}$ в штрихованном тензорном базисе тензора, первоначально заданного координатами $T^{j_1 \dots j_q}_{i_1 \dots i_p}$, достаточно теперь заметить, что они преобразуются так же, как координаты тензорного произведения q-векторов и p-ковекторов, т. е.

$$T'^{j_1 \dots j_q}_{i_1 \dots i_p} = A^{l_1}_{i_1} \dots A^{l_p}_{i_p} B^{j_1}_{k_1} \dots B^{j_q}_{k_q} T^{k_1 \dots k_q}_{l_1 \dots l_p}.$$

Не забывайте, что справа подразумевается суммирование по парам одинаковых индексов.

В классическом изложении эту формулу кладут в основу *определения* тензоров.

Именно, *тензором типа* (p, q) *на* n-*мерном пространстве называют отображение* T, *которое каждому базису* L *ставит в соответствие семейство из* n^{p+q} *компонент — скаляров* $T^{j_1 \dots j_q}_{i_1 \dots i_p}$, *причем такое, что при замене базиса посредством матрицы* A *замена компонент тензора происходит по выписанным выше формулам.*

5. Тензорные конструкции в координатах.

а) *Линейные комбинации тензоров одинакового типа.* Здесь формулы очевидны:

$$(aT + bT')^{j_1 \dots j_q}_{i_1 \dots i_p} = aT^{j_1 \dots j_q}_{i_1 \dots i_p} + bT'^{j_1 \dots j_q}_{i_1 \dots i_p}.$$

б) *Тензорное умножение.* Согласно определению в п. 2 § 3

$$(T \otimes T')^{j_1 \dots j_q j'_1 \dots j'_{q'}}_{i_1 \dots i_p i'_1 \dots i'_{p'}} = T^{j_1 \dots j_q}_{i_1 \dots i_p} T'^{j'_1 \dots j'_{q'}}_{i'_1 \dots i'_{p'}}.$$

В частности, разложимый тензор имеет компоненты $T_{i_1} \dots T_{i_p} T^{j_1} \dots$ $\dots T^{j_q}$.

в) *Перестановки.* Пусть σ — перестановка индексов $1, \dots, p$, τ — перестановка индексов $1, \dots, q$; $f_{\sigma, \tau}: T_p^q(L) \to T_p^q(L)$ — линейное отображение, отвечающее этим перестановкам, как в п. 3 § 2, Тогда для любого $T \in T_p^q(L)$ имеем

$$[f_{\sigma, \tau}(T)]_{i_1 \dots i_p}^{j_1 \dots j_q} = T_{i_{\sigma^{-1}(1)} \dots i_{\sigma^{-1}(p)}}^{j_{\tau^{-1}(1)} \dots j_{\tau^{-1}(q)}}.$$

г) *Свертка.* Пусть $a \in \{1, \dots, p\}$, $b \in \{1, \dots, q\}$. Как в п. 8 § 2, имеется отображение $T_p^q(L) \to T_{p-1}^{q-1}(L)$, «уничтожающее» a-й множитель L^* и b-й множитель L с помощью отображения свертки $L^* \otimes L \to \mathcal{K}$, которое является обычным скалярным произведением векторов и функционалов: $(b_i) \otimes (a^j) \mapsto b_i a^i$. Поэтому, обозначая через T' тензор T, свернутый по паре индексов (a-й нижний, b-й верхний), получаем

$$T'^{j_1 \dots j_{b-1} j_{b+1} \dots j_q}_{i_1 \dots i_{a-1} i_{a+1} \dots i_p} = T^{j_1 \dots j_{b-1} k j_{b+1} \dots j_q}_{i_1 \dots i_{a-1} k i_{a+1} \dots i_p}$$

(справа суммирование по k). Итерируя эту конструкцию, получим определение свертки по нескольким парам индексов.

Мы уже убедились, что многие формулы тензорной алгебры пишутся в терминах тензорного умножения и последующей свертки по одной или нескольким парам индексов. Повторим их для закрепления:

$$g_{ij} a^i b^j \text{ — свертка } ((g_{ij}) \otimes (a^k) \otimes (b^l)).$$

Скалярное произведение:

$$b_i a^i \text{ — свертка } ((b_i) \otimes (a^j)).$$

Координаты тензора в новом базисе, или, с «активной точки зрения», образ тензора при линейном преобразовании основного пространства:

$$T'^{j_1 \dots j_q}_{i_1 \dots i_p} \text{ — свертка } ((A \otimes \dots \otimes A) \otimes (B \otimes \dots \otimes B) \otimes T).$$

Умножение в алгебре:

$$a^i \cdot b^j \text{ — свертка } ((\text{структурный тензор}) \otimes (a^i) \otimes (b^j)).$$

Еще один пример — умножение матриц:

$$(A_j^i)(B_k^j) = (A_j^i B_k^j) \text{ — свертка } (A_j^i \otimes B_k^l).$$

Еще раз напомним, что для полного определения свертки следует указать, по каким индексам она производится; в приведенных примерах это либо очевидно, либо ясно из приведенных ранее полных формул.

В общем, можно сказать, что операция свертки в классическом языке тензорной алгебры играет такую же унифицирующую роль, как операция умножения матриц в языке линейной алгебры. В § 4 ч. 1 мы подчеркивали, что теоретико-множественные операции разной природы единообразно описываются с помощью матричного умножения. К тензорной алгебре и свертке, скомбинированной с тензорным умножением, это замечание применимо в еще большей мере.

д) *Подъем и опускание индексов.* Согласно определению в разделе в) п. 8 § 2 подъем a-го индекса и опускание b-го индекса — это линейные отображения

$$T_p^q(L) \to T_{p-1}^{q+1}(L), \quad T_p^q(L) \to T_{p+q}^{q-1}(L),$$

которые индуцируются некоторыми изоморфизмами $g\colon L^* \to L$ или $g^{-1}\colon L \to L^*$: следует «заменить» в произведении $L^{*\otimes p} \otimes L^{\otimes q}$ a-й множитель L^* на L или соответственно b-й множитель L на L^*.

В соответствии с соглашениями в конце п. 2 этого параграфа компоненты полученных тензоров следует записывать в виде

$$T_{i_1 \dots i_{a-1}}{}^{i_a}{}_{i_{a+1} \dots i_p}{}^{j_1 \dots j_q}, \quad T_{i_1 \dots i_p}{}^{j_1 \dots j_{b-1}}{}_{j_b}{}^{j_{b+1} \dots j_q}.$$

Если условиться применять после отображения подъема (спуска) индекса отображение перестановки, перегоняющее новый сомножитель L направо, а L^* налево, пока он не станет соседствовать со старыми сомножителями, то можно сохранить прежний вид записи компонент.

Как мы уже упоминали, изоморфизмы $g\colon L^* \to L$ и $g^{-1}\colon L \to L^*$ в приложениях чаще всего происходят из симметричной невырожденной билинейной формы g_{ij} на L. Поскольку она сама является тензором, операции подъема и опускания индексов можно применять и к ней. Опишем этот формализм подробнее.

Форма g_{ij} ставит в соответствие вектору a^i линейный функционал

$$b^j \longmapsto \sum g_{ij} a^i b^j.$$

Координаты этого функционала в двойственном базисе L^* суть $g_{ij}a^i$ (суммирование по i), или ввиду симметрии $g_{ij}a^j$. Иными словами, *опускание (единственного) верхнего индекса тензора* a^i *с помощью метрического тензора* g_{ij} приводит к тензору

$$a_i = g_{ij} a^j.$$

Отсюда сразу же получается общая формула для опускания любого числа индексов у разложимого тензора и затем по линейности у любого тензора:

$$T_{i_1 \dots i_p j_1 \dots j_r}{}^{j_{r+1} \dots j_q} = g_{j_1 j_1'} \dots g_{j_r j_r'} T_{i_1 \dots i_p}{}^{j_1' \dots j_r' j_{r+1} \dots j_q}.$$

В частности, мы можем воспользоваться ею для вычисления тензора g^{ij}, получающегося из g_{ij} подъемом индексов. Действительно,

$$g_{ij} = g_{ik} g_{jl} g^{kl}.$$

Прочтем здесь правую часть как формулу для (i, j)-го элемента матрицы, получающейся умножением матрицы (g_{ik}) на матрицу $\left(\sum_i g_{jl}g^{kl}\right)$. Так как слева также стоит матрица (g_{ik}), очевидно,

$$g_{jl}g^{kl}=\delta_j^k,$$

т. е. *матрица* (g^{kl}) *обратна к матрице* (g_{ij}) (*учесть симметричность*). Это же вычисление показывает, что g_i^j есть тензор Кронекера.

Поэтому общая формула для подъема индексов имеет вид

$$T_{i_1\ldots i_b}{}^{i_{b+1}\ldots i_p j_1\ldots j_q}=g^{i_{b+1}i'_{b+1}}\ldots g^{i_p i'_p}T_{i_1\ldots i_p}{}^{j_1\ldots j_q}.$$

Если мы хотим опустить (соответственно поднять) другие наборы индексов, формулы очевидным образом видоизменяются.

§ 5. Симметричные тензоры

1. Пусть L — фиксированное линейное пространство и $T_0^q(L)=L^{\otimes q}$, $q\geqslant 1$. В п. 3 § 2 мы показали, что каждой перестановке σ из группы S_q перестановок чисел $1, \ldots, q$ можно поставить в соответствие линейное отображение $f_\sigma: T_0^q(L)\to T_0^q(L)$, которое действует на разложимых тензорах по формуле

$$f_\sigma(l_1\otimes\ldots\otimes l_q)=l_{\sigma(1)}\otimes\ldots\otimes l_{\sigma(q)}.$$

Назовем тензор $T\in T_0^q(L)$ *симметричным*, если $f_\sigma(T)=T$ для всех $\sigma\in S_q$. Очевидно, симметричные тензоры образуют линейное подпространство в $T_0^q(L)$. Все скаляры удобно считать симметричными тензорами. При отождествлении из г) п. 1 § 3 симметричные тензоры из $T_0^2(L^*)$ отвечают симметричным билинейным формам на L.

Обозначим через $S^q(L)$ подпространство симметричных тензоров в $T_0^q(L)$. Мы построим сейчас проектор $S: T_0^q(L)\to T_0^q(L)$, образ которого будет совпадать с $S^q(L)$, предполагая, что характеристика основного поля равна нулю или хотя бы не делит $q!$. Он называется отображением симметризации. В классических обозначениях вместо $S(T)$ пишут $T^{(i_1\ldots i_q)}$.

2. Предложение. Положим

$$S=\frac{1}{q!}\sum_{\sigma\in S_q}f_\sigma: T_0^q(L)\to T_0^q(L).$$

Тогда $S^2=S$ *и* $\operatorname{Im}S=S^q(L)$.

Доказательство. Очевидно, результат симметризации всякого тензора симметричен, так что $\operatorname{Im}S\subset S^q(L)$. Наоборот, на симметричных тензорах симметризация является тождественной операцией, так что если $T\in S^q(L)$, то $T=S(T)$. Это показывает одновременно, что $\operatorname{Im}S=S^q(L)$ и что $S^2=S$.

3. Пусть $\{e_1, \ldots, e_n\}$ — базис пространства L. Тогда разложимые тензоры $e_{i_1} \otimes \ldots \otimes e_{i_q}$ образуют базис $T_0^q(L)$, а их симметризации $S(e_{i_1} \otimes \ldots \otimes e_{i_q})$ порождают $S^q(L)$. Введем обозначение

$$S(e_{i_1} \otimes \ldots \otimes e_{i_q}) = e_{i_1} \ldots e_{i_q}.$$

Формальное произведение $e_{i_1} \ldots e_{i_q}$ не меняется при перестановке индексов, и можно условиться выбирать в качестве канонической записи таких симметричных тензоров запись $e_1^{a_1} \ldots e_n^{a_n}$, где $a_i \geqslant 0$, $a_1 + \ldots + a_n = q$; здесь число a_i показывает, сколько раз вектор e_i фигурирует в $e_{i_1} \otimes \ldots \otimes e_{i_q}$.

4. **Предложение.** *Тензоры* $e_1^{a_1} \ldots e_n^{a_n} \in S^q(L)$, $a_1 + \ldots + a_n = q$, *образуют базис в пространстве* $S^q(L)$, *которое, таким образом, можно отождествить с пространством однородных многочленов степени* q *от элементов базиса* L.

Доказательство. Мы должны лишь проверить, что тензоры $e_1^{a_1} \ldots e_n^{a_n}$ линейно независимы в $T_0^q(L)$. Если

$$\sum c_{a_1 \ldots a_n} e_1^{a_1} \ldots e_n^{a_n} = 0,$$

то

$$S\bigl(\sum c_{a_1 \ldots a_n} \underbrace{e_1 \otimes \ldots \otimes e_1}_{a_1} \otimes \ldots \otimes \underbrace{e_n \otimes \ldots \otimes e_n}_{a_n}\bigr) = 0.$$

Собирая в левой части подобные члены, нетрудно убедиться, что коэффициентами при элементах тензорного базиса пространства $T_0^q(L)$ окажутся скаляры $c_{a_1 \ldots a_n}$, умноженные на целые числа, состоящие из произведений простых чисел $\leqslant q$. Поскольку характеристика $\mathscr{K}$ по предположению больше $q!$, из равенства нулю этих коэффициентов следует, что все $c_{a_1 \ldots a_n}$ равны нулю.

5. **Следствие.** $\dim S^q(L) = \binom{n+q-1}{q}$.

6. Положим $S(L) = \bigoplus_{q=1}^{\infty} S^q(L)$. В силу предложения п. 4 $S(L)$ можно отождествить с пространством всех многочленов от элементов базиса L. На этом пространстве имеется структура алгебры, умножением в которой является обычное умножение многочленов. Однако сразу, возможно, не ясно, не зависит ли это умножение от выбора исходного базиса. Поэтому мы введем его инвариантно. Поскольку ниже приходится рассматривать все $S^q(L)$ одновременно, мы считаем, что характеристика $\mathscr{K}$ равна нулю.

7. **Предложение.** *Введем на пространстве* $S(L)$ *билинейное умножение по формуле*

$$T_1 T_2 = S(T_1 \otimes T_2), \quad f \in S^p(L), \quad g \in S^q(L).$$

Оно превращает $S(L)$ *в коммутативную ассоциативную алгебру над полем* $\mathscr{K}$. *В представлении симметричных тензоров в виде* *многочленов от элементов базиса L это умножение совпадает с умножением многочленов.*

Доказательство. Проверим сначала, что для любых тензоров $T_1 \in T_0^p(L)$, $T_2 \in T_0^q(L)$ имеет место формула

$$S(S(T_1) \otimes T_2) = S(T_1 \otimes S(T_2)) = S(T_1 \otimes T_2).$$

В самом деле,

$$S(T_1) \otimes T_2 = \frac{1}{p!} \sum_{\sigma \in S_p} f_\sigma(T_1) \otimes T_2,$$

откуда

$$S(S(T_1) \otimes T_2) = \frac{1}{p!} \sum_{\sigma \in S_p} S(f_\sigma(T_1) \otimes T_2).$$

Но $S(f_\sigma(T_1) \otimes T_2) = S(T_1 \otimes T_2)$ для любых $\sigma \in S_p$. Это очевидно для разложимых тензоров T_1, T_2 и следует для остальных по линейности. Поэтому сумма справа состоит из $p!$ слагаемых $S(T_1 \otimes T_2)$, так что

$$S(S(T_1) \otimes T_2) = S(T_1 \otimes T_2).$$

Аналогично устанавливается второе равенство.

Отсюда легко вывести, что на симметричных тензорах операция $(T_1, T_2) \mapsto S(T_1 \otimes T_2) = T_1 T_2$ ассоциативна. Действительно,

$$(T_1 T_2) T_3 = S(S(T_1 \otimes T_2) \otimes T_3) = S(T_1 \otimes T_2 \otimes T_3)$$

и аналогично

$$T_1(T_2 T_3) = S(T_1 \otimes S(T_2 \otimes T_3)) = S(T_1 \otimes T_2 \otimes T_3).$$

Кроме того, она коммутативна: формула $S(T_1 \otimes T_2) = S(T_2 \otimes T_1)$ очевидна для разложимых тензоров и следует для остальных по линейности.

Из этих утверждений следует, что

$$\left(e_1^{a_1} \ldots e_n^{a_n}\right)\left(e_1^{b_1} \ldots e_n^{b_n}\right) = e_1^{a_1+b_1} \ldots e_n^{a_n+b_n},$$

что завершает доказательство.

8. Построенная выше алгебра $S(L)$ называется *симметрической алгеброй* пространства L.

Элементы алгебры $S(L^*)$ можно рассматривать как полиномиальные функции на пространстве L со значениями в поле $\mathscr{K}$: элементу $f \in L^*$ ставится в соответствие он сам как функционал на L, а произведению элементов в $S(L^*)$ и их линейной комбинации — произведение и линейная комбинация соответствующих функций. Не вполне очевидно, что разные элементы $S(L^*)$ различаются также как функции на L. Мы оставляем этот вопрос читателю в качестве упражнения. Для симметрических алгебр над конечными полями, которые мы введем ниже, это уже не так: например, функция $x^p - x$ тождественно равна нулю в поле $\mathscr{K}$ из p элементов.

9. Второе определение симметрической алгебры. В принятом нами определении симметрической алгебры с помощью оператора S используется деление на факториалы. Это невозможно над полями конечной характеристики и в теории модулей над кольцами, где формализм тензорной алгебры также существует и весьма полезен. Поэтому мы вкратце опишем другое определение симметрической алгебры пространства L, в котором она реализуется не как подпространство, а как факторпространство $T_0(L) = \bigoplus\limits_{p=0}^{\infty} T_0^p(L)$.

Для этого рассмотрим *двусторонний идеал* I в тензорной алгебре $T_0(L)$, порожденный всеми элементами вида

$$T - f_\sigma(T),\ T \in T_0^p(L),\ \sigma \in S_p, \qquad p = 1, 2, 3, \ldots$$

Он состоит из всевозможных сумм таких тензоров, слева и справа тензорно умноженных на любые элементы $T_0(L)$. Нетрудно видеть, что $I = \bigoplus\limits_{i=1}^{\infty} I^p$, где $I^p = I \cap T_0^p(L)$, т. е. этот идеал градуирован.

Положим

$$\tilde{S}(L) = T_0(L)/I$$

как факторпространство. То же рассуждение, что в § 11 ч. 3, показывает, что

$$\tilde{S}(L) = \bigoplus_{p=0}^{\infty} \tilde{S}^p(L),\ \tilde{S}^p(L) = T_0^p(L)/I^p.$$

Благодаря тому, что I — идеал, в $\tilde{S}(L)$ можно ввести умножение по формуле

$$(T_1 + I)(T_2 + I) = T_1 \otimes T_2 + I.$$

Оно билинейно и ассоциативно, так как это верно для тензорного умножения. Кроме того, оно коммутативно, ибо если T_1, T_2 разложимы, то $T_2 \otimes T_1 = f_\sigma(T_1 \otimes T_2)$ для подходящей перестановки σ и, значит, $T_1 \otimes T_2 - T_2 \otimes T_1 \in I$. Таким образом, $\tilde{S}(L)$ есть коммутативная ассоциативная $\mathcal{K}$-алгебра. Можно показать, что естественное отображение $L \to \tilde{S}(L)$: $l \mapsto l + I$ является вложением и что в терминах любого базиса пространства L элементы $\tilde{S}(L)$ однозначно представляются как многочлены от этого базиса. Элементы $\tilde{S}^p(L)$ отвечают однородным многочленам степени p.

Если характеристика $\mathcal{K}$ равна нулю, то сквозное отображение

$$S(L) \to T_0(L) \to \tilde{S}(L)$$

является изоморфизмом алгебр, сохраняющим градуировку. Поскольку $\tilde{S}(L)$ существует в более общей ситуации, для алгебраических нужд симметрическую алгебру удобно вводить именно таким способом.

§ 6. Кососимметричные тензоры и внешняя алгебра линейного пространства

1. В той же ситуации, что и в п. 1 § 5, назовем тензор $T \in T_0^q(L)$ *кососимметричным* (или *антисимметричным*), если $f_\sigma(T) = \varepsilon(\sigma) T$, где $\varepsilon(\sigma)$ — знак перестановки σ, для всех $\sigma \in S_q$. Очевидно, кососимметричные тензоры образуют линейное подпространство в $T_0^q(L)$. Все скаляры удобно считать одновременно кососимметричными и симметричными тензорами. При отождествлении из г) п. 1 § 3 кососимметричные тензоры из $T_0^2(L^*)$ отвечают симплектическим билинейным формам на L.

Обозначим через $\Lambda^q(L)$ подпространство кососимметрических тензоров в $T_0^q(L)$. По аналогии с § 5 построим линейный проектор $A: T_0^q(L) \to T_0^q(L)$, образ которого совпадает с $\Lambda^q(L)$. Как и там, мы предполагаем пока, что характеристика поля скаляров не делит $q!$. Проектор A будет называться *антисимметризацией* или *альтернированием*. В классических обозначениях вместо $A(T)$ пишут $T^{[i_1 \dots i_q]}$.

2. Предложение. *Положим*

$$A = \frac{1}{q!} \sum_{\sigma \in S_q} \varepsilon(\sigma) f_\sigma : T_0^q(L) \to T_0^q(L).$$

Тогда $A^2 = A$ *и* $\operatorname{Im} A = \Lambda^q(L)$.

Доказательство. Прежде всего проверим, что результат альтернирования всякого тензора кососимметричен. Действительно, поскольку f_σ и $\varepsilon(\sigma)$ мультипликативны по σ и $\varepsilon(\sigma)^2 = 1$, имеем

$$f_\sigma(AT) = f_\sigma\left(\frac{1}{q!} \sum_{\tau \in S_q} \varepsilon(\tau) f_\tau(T)\right) = \frac{1}{q!} \sum_{\tau \in S_q} \varepsilon(\tau) f_{\sigma\tau}(T) =$$

$$= \varepsilon(\sigma) \frac{1}{q!} \sum_{\tau \in S_q} \varepsilon(\sigma\tau) f_{\sigma\tau}(T) = \varepsilon(\sigma) AT.$$

Далее, A является проектором, потому что

$$A^2 = \frac{1}{(q!)^2} \sum_{\sigma, \tau \in S_q} \varepsilon(\sigma\tau) f_{\sigma\tau} = \frac{1}{q!} \sum_{\rho \in S_q} \varepsilon(\rho) f_\rho = A.$$

Действительно, любой элемент $\rho \in S_q$ ровно $q!$ способами можно представить в виде произведения $\sigma\tau$: σ выберем любым, τ находим из равенства $\tau = \sigma^{-1}\rho$.

Отсюда, как в предложении п. 1 § 5, следует, что $\operatorname{Im} A = \Lambda^q(L)$.

3. Пусть $\{e_1, \dots, e_n\}$ — базис пространства L. Тогда разложимые тензоры $e_{i_1} \otimes \dots \otimes e_{i_q}$ образуют базис $T_0^q(L)$, а их антисимметризации $A(e_{i_1} \otimes \dots \otimes e_{i_q})$ порождают $\Lambda^q(L)$. Введем обозначение

$$A(e_{i_1} \otimes \dots \otimes e_{i_q}) = e_{i_1} \wedge \dots \wedge e_{i_q}$$

(значок $\wedge$ называется символом «внешнего умножения»).

Заметим теперь, что в отличие от симметрического случая перестановка любых двух векторов в $e_{i_1}\wedge\ldots\wedge e_{i_q}$ меняет знак этого произведения, ибо этот тензор антисимметричен. Отсюда следуют два вывода:

а) $e_{i_1}\wedge\ldots\wedge e_{i_q}=0$, *если* $i_a=i_b$ *для некоторых* a, b, *при условии, что* $\operatorname{char}\mathcal{K}\neq 2$.

б) *Пространство* $\Lambda^q(L)$ *порождено тензорами вида* $e_{i_1}\wedge\ldots\wedge e_{i_q}$, *где* $1\leqslant i_1<i_2<\ldots<i_q\leqslant n$. Отсюда, в частности, сразу же следует, что $\Lambda^m(L)=0$ при $m>n=\dim L$.

Следующий результат параллелен предложению п. 4 § 5.

4. Предложение. *Тензоры* $e_{i_1}\wedge\ldots\wedge e_{i_q}\in\Lambda^q(L)$ *при* $q\leqslant n$, $1\leqslant i_1<i_2<\ldots<i_q\leqslant n$ *образуют базис пространства* $\Lambda^q(L)$.

Доказательство. Нужно только проверить, что эти тензоры линейно независимы в $T_0^q(L)$. Если

$$\sum c_{i_1\ldots i_q}e_{i_1}\wedge\ldots\wedge e_{i_q}=0,$$

то

$$A\left(\sum e_{i_1\ldots i_q}e_{i_1}\otimes\ldots\otimes e_{i_p}\right)=0.$$

Но так как среди индексов $i_1,\ldots,i_q$ нет одинаковых и они расположены в порядке возрастания, в результате их перестановок мы получим в сумме слева линейную комбинацию различных элементов тензорного базиса $T_0^q(L)$ с коэффициентами вида $\pm\frac{1}{q!}c_{i_1\ldots i_q}$. Эта сумма может быть равна нулю, только если все $c_{i_1\ldots i_{q_n}}$ нулевые.

5. Следствие. $\dim\Lambda^q(L)=\binom{n}{q}$, $\dim\bigoplus\limits_{q=0}^{n}\Lambda^q(L)=2^n$.

6. Положим $\Lambda(L)=\bigoplus\limits_{q=0}^{n}\Lambda^q(L)$. По аналогии с симметрическим случаем введем на пространстве антисимметрических тензоров операцию внешнего умножения и покажем, что она превращает $\Lambda(L)$ в ассоциативную алгебру, называемую *внешней алгеброй*, или *алгеброй Грассмана*, пространства L.

7. Предложение. *Билинейная операция*

$$T_1\wedge T_2=A(T_1\otimes T_2);\ T_1\in\Lambda^p(L),\ T_2\in\Lambda^q(L),$$

на $\Lambda(L)$ *ассоциативна*, $T_1\wedge T_2\notin L^{p+q}(L)$, *и* $T_2\wedge T_1=(-1)^{pq}T_1\wedge T_2$ *(это свойство иногда называют косокоммутативностью).*

В частности, подпространство $\Lambda^+(L)=\bigoplus\limits_{q=0}^{[n/2]}\Lambda^{2q}(L)$ *является центральной подалгеброй* $\Lambda(L)$.

Доказательство. По аналогии с симметричным случаем проверим сначала, что для всех $T_1\in T_0^p(L)$, $T_2\in T_0^q(L)$ имеют место формулы

$$A(A(T_1)\otimes T_2)=A(T_1\otimes A(T_2))=A(T_1\otimes T_2).$$

В самом деле,

$$A(T_1)\otimes T_2=\frac{1}{p!}\sum_{\sigma\in S_p}\varepsilon(\sigma)f_\sigma(T_1)\otimes T_2,$$

откуда

$$A(A(T_1)\otimes T_2)=\sum_{\sigma\in S_p}\varepsilon(\sigma)A(f_\sigma(T_1)\otimes T_2).$$

Рассмотрим вложение $S_p\to S_{p+q}$, $\sigma\mapsto\tilde\sigma$, где

$$\tilde\sigma(i)=\begin{cases}\sigma(i) & \text{при } 1\leqslant i\leqslant p,\\ i & \text{при } i>p.\end{cases}$$

Очевидно, $f_\sigma(T_1)\otimes T_2=f_{\tilde\sigma}(T_1\otimes T_2)$, кроме того, A и $f_{\tilde\sigma}$ коммутирует, так что

$$Af_{\tilde\sigma}(T_1\otimes T_2)=f_{\tilde\sigma}A(T_1\otimes T_2)=\varepsilon(\tilde\sigma)A(T_1\otimes T_2)=\varepsilon(\sigma)A(T_1\otimes T_2).$$

Поэтому

$$A(A(T_1)\otimes T_2)=\frac{1}{p!}\sum_{\sigma\in S_p}\varepsilon^2(\sigma)A(T_1\otimes T_2)=A(T_1\otimes T_2).$$

Аналогично доказывается второе равенство. Теперь ассоциативность внешнего умножения можно проверить так же, как в симметричном случае:

$$(T_1\wedge T_2)\wedge T_3=A(A(T_1\otimes T_2)\otimes T_3)=A(T_1\otimes T_2\otimes T_3),$$
$$T_1\wedge(T_2\wedge T_3)=A(T_1\otimes A(T_2\otimes T_3))=A(T_1\otimes T_2\otimes T_3).$$

Равенство $A(T_1\otimes T_2)=(-1)^{pq}A(T_2\otimes T_1)$ при $T_1\in T_0^p(L)$, $T_2\in T_0^q(L)$ следует из того, что $T_2\otimes T_1=f_\sigma(T_1\otimes T_2)$, где σ — перестановка, являющаяся произведением pq транспозиций: сомножители T_2 следует по очереди переводить левее T_1, меняя их местами с левыми соседями из T_1.

8. Второе определение внешней алгебры. Как и в симметрическом случае, принятое нами определение внешней алгебры страдает тем недостатком, что оно требует деления на факториалы. Второе определение, избавленное от этого недостатка и реализующее $\Lambda(L)$ как факторпространство, а не подпространство $T_0(L)$, строится в полной аналогии с симметричным случаем.

Рассмотрим двусторонний идеал J в алгебре $T_0(L)$, порожденный всеми элементами вида

$$T-\varepsilon(\sigma)f_\sigma(T),\quad T\in T_0^p(L),\quad \sigma\in S_p,\qquad p=1,\ 2,\ 3,\ \ldots$$

Нетрудно убедиться, что $J=\bigoplus_{p=0}^{\infty}J^p$, где $J^p=J\cap T_0^p(L)$, т. е. это градуированный идеал. Положим $\tilde{\Lambda}(L)=T_0(L)/J$ как фактор-пространство. Тогда

$$\tilde{\Lambda}(L)=\bigoplus_{p=0}^{\infty}\tilde{\Lambda}^p(L),\quad \tilde{\Lambda}^p(L)=T_0^p(L)/J^p.$$

Поскольку J — идеал, в $\tilde{\Lambda}(L)$ можно ввести умножение по формуле

$$(T_1+J)\wedge(T_2+J)=T_1\otimes T_2+J.$$

Оно билинейно и ассоциативно, так как это верно для тензорного умножения. Кроме того, оно кососкоммутативно, ибо для $T_1\in T_0^p(L)$, $T_2\in T_0^q(L)$ имеем $T_1\otimes T_2-(-1)^{pq}T_2\otimes T_1\in J$.

Нетрудно убедиться, что *построенная таким образом алгебра изоморфна алгебре Клиффорда пространства L с нулевым скалярным произведением, введенной в* § 15 *ч.* 2. Действительно, отображение $\sigma\colon L\to\tilde{\Lambda}(L)$, $\sigma(l)=l+J$, удовлетворяет условию $\sigma(l)^2=\sigma(l)\wedge\sigma(l)=0$ для всех l, ибо $\sigma(l)\wedge\sigma(l)=-\sigma(l)\wedge\sigma(l)$. Поэтому по теореме п. 2 § 15 ч. 2, существует единый гомоморфизм $\mathscr{K}$-алгебр $C(L)\to\tilde{\Lambda}(L)$ такой, что σ совпадает с композицией $L\xrightarrow{\rho}C(L)\to\tilde{\Lambda}(L)$, где ρ — каноническое отображение. Поскольку L порождает $T_0(L)$ как алгебру, $\sigma(L)$ порождает $\tilde{\Lambda}(L)$, так что $C(L)\to\tilde{\Lambda}(L)$ сюръективен. Мы знаем, что $\dim C(L)=2^n$. Поэтому для проверки того, что это изоморфизм, достаточно убедиться, что $\dim\tilde{\Lambda}(L)=2^n$. Это можно сделать, установив, что базис $\tilde{\Lambda}^q(L)$ образует элементы вида $e_{i_1}\wedge\ldots\wedge e_{i_q}$, $1\leqslant i_1<\ldots<i_q\leqslant n$, где $\{e_1,\ldots,e_n\}$ — базис L. Эту проверку мы опустим.

Как и в симметричном случае, если характеристика $\mathscr{K}$ равна нулю, сквозное отображение

$$\Lambda(L)\to T_0(L)\to\tilde{\Lambda}(L)$$

также является изоморфизмом алгебр, сохраняющим градуировку.

Поскольку $\tilde{\Lambda}(L)$ определена в более общей ситуации, для алгебраических нужд внешнюю алгебру вводят именно таким способом. В приложениях к дифференциальной геометрии или анализу, где $\mathscr{K}=\mathbf{R}$ или $\mathbf{C}$, можно пользоваться нашим исходным определением.

9. Внешнее умножение и определители. Пусть L — n-мерное пространство. Согласно следствию п. 5 пространство $\Lambda^n(L)$ одномерно: это максимальная ненулевая внешняя степень L.

Согласно п. 7 § 2 любой эндоморфизм $f\colon L\to L$ индуцирует эндоморфизмы тензорных степеней

$$f^{\otimes p}=\underbrace{f\otimes\ldots\otimes f}_{p}\colon T_0^p(L)\to T_0^p(L).$$

Легко убедиться, что $f^{\otimes p}$ коммутирует с оператором альтернирования A и потому переводит $\Lambda^p(L)$ в $\Lambda^p(L)$. Ограничение $f^{\otimes p}$ на

$\Lambda^p(L)$ естественно обозначить $f^{\wedge p}$. В частности, при $p=n$ отображение $f^{\wedge p}: \Lambda^n(L) \to \Lambda^n(L)$ должно быть умножением на скаляр $d(f)$, ибо $\Lambda^n(L)$ одномерно.

10. Теорема. *В описанных обозначениях* $d(f) = \det f$.

Доказательство. Выберем базис $e_1, \ldots, e_n$ пространства L и зададим f матрицей в этом базисе:

$$f(e_j) = \sum_{i=1}^{n} a_j^i e_i.$$

Внешнее произведение $e_1 \wedge \ldots \wedge e_n$ является базисом $\Lambda^n(L)$, и число $d(f)$ находится из равенства

$$f^{\wedge n}(e_1 \wedge \ldots \wedge e_n) = d(f)\, e_1 \wedge \ldots \wedge e_n.$$

Но

$$f^{\wedge n}(e_1 \wedge \ldots \wedge e_n) = A(f(e_1) \otimes \ldots \otimes f(e_n)) = f(e_1) \wedge \ldots \wedge f(e_n) = \\ = \left(\sum_{i_1=1}^{n} a_1^{i_1} e_{i_1}\right) \wedge \ldots \wedge \left(\sum_{i_n=1}^{n} a_n^{i_n} e_{i_n}\right).$$

Согласно таблице умножения во внешней алгебре

$$a_1^{i_1} e_{i_1} \wedge a_2^{i_2} e_{i_2} \wedge \ldots \wedge a_n^{i_n} e_{i_n} = \\ = \begin{cases} \varepsilon(\sigma) a_1^{i_1} \ldots a_n^{i_n} e_1 \wedge \ldots \wedge e_n, & \text{если } \{i_1, \ldots, i_n\} = \{1, \ldots, n\}, \\ 0 & \text{в остальных случаях,} \end{cases}$$

где σ — перестановка, переводящая i_k в k, $1 \leqslant k \leqslant n$. Поэтому полная сумма коэффициентов $\varepsilon(\sigma) a_1^{i_1} \ldots a_n^{i_n}$ совпадает со стандартной формулой для определителя $\det(a_j^i)$, что завершает доказательство.

11. Следствие. *Векторы $e_1', \ldots, e_n' \in L$ линейно зависимы тогда и только тогда, когда $e_1' \wedge \ldots \wedge e_n' = 0$.*

Действительно, пусть $f: L \to L$ — эндоморфизм, переводящий e_i в e_i', где $\{e_1, \ldots, e_n\}$ — некоторый базис L. Тогда линейная зависимость $\{e_i'\}$ равносильна тому, что $\det f = 0$, т. е. $e_1' \wedge \ldots \wedge e_n' = 0$.

12. Разложимые p-векторы. Элементы $T \in \Lambda^p(L)$ принято называть *p-векторами*. Назовем p-вектор T разложимым, если существуют такие векторы $e_1, \ldots, e_p \in L$, что $T = e_1 \wedge \ldots \wedge e_p$. Для любого p-вектора T назовем его *аннулятором* множество

$$\operatorname{Ann} T = \{e \in L \mid e \wedge T = 0\}.$$

Очевидно, $\operatorname{Ann} T$ является подпространством в L.

13. Теорема. *Пусть T_1, T_2 — разложимые p-вектор и q-вектор соответственно, L_1, L_2 — их аннуляторы. Тогда* а) $L_1 \supset L_2$ *в том и только в том случае, когда* T_1 *делится на* T_2, *т. е.* $T_1 = T \wedge T_2$ *для некоторого* $T \in \Lambda^{p-q}(L)$.

б) $L_1 \cap L_2 = \{0\}$ *в том и только в том случае, когда* $T_1 \wedge T_2 \neq 0$.

в) *Если* $L_1 \cap L_2 = \{0\}$, *то* $L_1 + L_2 = \mathrm{Ann}(T_1 \wedge T_2)$.

Доказательство. а) Если $x \wedge T_2 = 0$, то $x \wedge (T \wedge T_2) = \pm T \wedge (x \wedge T_2) = 0$, так что из делимости T_1 на T_2 следует, что $L_2 \subset L_1$.

Для доказательства обратного утверждения вычислим аннулятор p-вектора $e_1 \wedge \ldots \wedge e_p$. Если $e_1, \ldots, e_p$ линейно зависимы, то один из векторов e_i, скажем e_1, линейно выражается через остальные, и тогда

$$e_1 \wedge \ldots \wedge e_p = \left(\sum_{i=2}^{p} a^i e_i\right) \wedge e_2 \wedge \ldots \wedge e_n = 0.$$

Будем считать, что $e_1 \wedge \ldots \wedge e_p$ отличен от нуля, и покажем, что тогда $\mathrm{Ann}(e_1 \wedge \ldots \wedge e_p)$ совпадает с линейной оболочкой векторов $e_1, \ldots, e_p$. Ясно, что эта линейная оболочка содержится в аннуляторе, ибо

$$e_j \wedge (e_1 \wedge \ldots \wedge e_p) = \\ = \pm (e_j \wedge e_j) \wedge (e_1 \wedge \ldots \wedge e_{j-1} \wedge e_{j+1} \wedge \ldots \wedge e_p) = 0.$$

Дополним линейно независимую систему векторов $\{e_1, \ldots, e_p\}$ до базиса $\{e_1, \ldots, e_p, e_{p+1}, \ldots, e_n\}$ пространства L и покажем, что если $\sum_{i=1}^{n} a^i e_i \in \mathrm{Ann}(e_1 \wedge \ldots \wedge e_p)$, то $a^i = 0$ при $i > p$. В самом деле,

$$\left(\sum_{i=1}^{n} a^i e_i\right) \wedge (e_1 \wedge \ldots \wedge e_p) = \sum_{i=p+1}^{n} a^i e_i \wedge e_2 \wedge \ldots \wedge e_p,$$

и $(p+1)$-векторы $e_i \wedge e_1 \wedge \ldots \wedge e_p$, $p+1 \leqslant i \leqslant n$, линейно независимы.

Пусть теперь $L_1 \supset L_2$, $T_1 = e_1 \wedge \ldots \wedge e_p$, $T_2 \leqslant e_1' \wedge \ldots \wedge e_q'$. Так как линейная оболочка $\{e_1', \ldots, e_q'\}$ содержится в линейной оболочке $\{e_1, \ldots, e_p\}$, мы можем выбрать в ней базис вида $\{e_1', \ldots, e_q', e_{q+1}', \ldots, e_p'\}$ и выразить e_j линейно через этот базис. Для T_1 получится выражение $a e_1' \wedge \ldots \wedge e_q' \wedge e_{q+1}' \wedge \ldots \wedge e_p'$, где a — определитель перехода от штрихованного базиса к нештрихованному. Поэтому T_1 делится на T_2.

б), в). Если $(e_1 \wedge \ldots \wedge e_p) \wedge (e_1' \wedge \ldots \wedge e_q') \neq 0$, то векторы $\{e_1, \ldots, e_p, e_1', \ldots, e_q'\}$ линейно независимы. Следовательно, линейные оболочки $\{e_1, \ldots, e_p\}$ и $\{e_1', \ldots, e_q'\}$, т. е. аннуляторы T_1 и T_2 пересекаются лишь по нулю. Это рассуждение, очевидно, обратимо. Характеризация аннулятора разложимого p-вектора, данная в доказательстве утверждения а), доказывает последнее утверждение теоремы.

14. Следствие. *Рассмотрим отображение*

$$\mathrm{Ann}\colon \begin{pmatrix}\text{разложимые ненулевые } p\text{-векторы} \\ \text{с точностью до умножения на скаляр}\end{pmatrix} \to$$
$$\to (p\text{-мерные подпространства в } L).$$

Оно является биекцией.

Доказательство. Ясно, что если два ненулевых разложимых вектора пропорциональны, то их аннуляторы совпадают. Поэтому описанное отображение определено корректно. Любое p-мерное подпространство $L_1 \subset L$ лежит в образе отображения, ибо если $\{e_1, \ldots, e_p\}$ — базис L_1, то $L_1 = \mathrm{Ann}(e_1 \wedge \ldots \wedge e_p)$. Наконец, это отображение инъективно в силу утверждения а) теоремы п. 13: если $\mathrm{Ann}\, T_1 = \mathrm{Ann}\, T_2$, то $T_1 = T \wedge T_2$ и T является O-вектором, т. е. скаляром.

15. Многообразия Грассмана. *Многообразием Грассмана*, или *грассманианом* $\mathrm{Gr}(p, L)$, называется множество всех p-мерных линейных подпространств пространства L. В случае $p = 1$ получается подробно изученное нами проективное пространство $P(L)$. Следствие п. 14 позволяет нам для любого p реализовать $\mathrm{Gr}(p, L)$ как подмножество в проективном пространстве $P(\Lambda^p(L))$.

В самом деле, отображение, обратное к Ann, дает вложение

$$\mathrm{Ann}^{-1}\colon \mathrm{Gr}(p, L) \to P(\Lambda^p(L)).$$

Выпишем его в более явном виде. Выберем базис $\{e_1, \ldots, e_n\}$ в L и рассмотрим линейную оболочку p векторов

$$\sum_{i=1}^{n} a_j^i e_i; \qquad j = 1, \ldots, p.$$

Базис $\Lambda^p(L)$ образуют $\binom{n}{p}$ p-векторов $\{e_{i_1} \wedge \ldots \wedge e_{i_p} \mid 1 \leqslant i_1 < \ldots < i_p \leqslant n\}$. Отображение Ann^{-1} ставит в соответствие нашей линейной оболочке прямую в $\Lambda^p(L)$, порожденную p-вектором

$$\left(\sum_{i_1=1}^{n} a_1^{i_1} e_{i_1}\right) \wedge \ldots \wedge \left(\sum_{i_p=1}^{n} a_p^{i_p} e_{i_p}\right).$$

Однородными координатами соответствующей точки в $P(\Lambda^p(L))$ являются коэффициенты разложения этого p-вектора по $\{e_{i_1} \wedge \ldots \wedge e_{i_p}\}$:

$$\bigwedge_{j=1}^{p} \left(\sum_{j=1}^{n} a_j^i e_i\right) = \sum_{1 \leqslant i_1 < \ldots < i_p \leqslant n} \Delta^{i_1 \ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

В точности такое же вычисление, как в доказательстве теоремы п. 10, показывает, что $\Delta^{i_1 \ldots i_p}$ совпадает с минором матрицы (a_j^i), образованным строками с номерами $i_1, \ldots, i_p$. Хоть один из этих миноров отличен от нуля в точности тогда, когда ранг матрицы (a_j^i) имеет наибольшее возможное значение p, т. е. когда линейная оболочка наших p-векторов действительно p-мерна.

Вектор $(\ldots : \Delta^{i_1 \ldots i_p} : \ldots)$ называется вектором *грассмановых координат* p-мерного подпространства, натянутого на

$$\sum_{i=1}^{n} a_j^i e_i, \qquad j = 1, \ldots, p.$$

Из этой конструкции ясно, что для характеризации образа $\mathrm{Gr}(p, L)$ в $P(\Lambda^p(L))$ нам нужно иметь критерии разложимости p-векторов. Поэтому мы займемся сейчас этой задачей.

16. Теорема. а) *Ненулевой p-вектор T разложим тогда и только тогда, когда* $\dim \operatorname{Ann} T = p$; *для остальных ненулевых p-векторов* $\dim \operatorname{Ann} T < p$.

б) *Выберем базис $\{e_1, \ldots, e_n\}$ в пространстве L и представим любой p-вектор T коэффициентами его разложения по базису $\{e_{i_1} \wedge \ldots \wedge e_{i_p} \mid 1 \leqslant i_1 < i_2 < \ldots < i_p \leqslant n\}$ в* $\Lambda^p(L)$:

$$T = \sum T^{i_1 \ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

Тогда существует такая система полиномиальных уравнений от $T^{i_1 \ldots i_p}$ с целыми коэффициентами, зависящая только от n и p, что разложимость T равносильна тому, что $\{T^{i_1 \ldots i_p}\}$ есть решение этой системы.

Доказательство. То, что $\dim \operatorname{Ann} T = p$ для разложимых p-векторов, мы знаем из доказательства теоремы п. 10.

Пусть $\dim \operatorname{Ann} T = r$ и $\operatorname{Ann} T$ порождено векторами $e_1, \ldots, e_r$. Дополним их до базиса $\{e_1, \ldots, e_n\}$ в L и положим

$$T = \sum T^{i_1 \ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

Условие $e_i \wedge T = 0$ для всех $i = 1, \ldots, r$ означает, что $T^{i_1 \ldots i_p} = 0$, если только $\{1, \ldots, r\} \not\subset \{i_1, \ldots, i_p\}$. Отсюда сразу же следует, что если $T \neq 0$, то $r \leqslant p$ и что T делится на $e_1 \wedge \ldots \wedge e_r$. Поэтому при $r = p$ p-вектор T пропорционален $e_1 \wedge \ldots \wedge e_r$ и, значит, разложим.

б) Воспользовавшись этим критерием, мы можем теперь записать условие разложимости T в виде требования, чтобы следующая линейная система уравнений относительно неизвестных $x^1, \ldots$ $\ldots, x^n \in \mathscr{K}$ имела p-мерное пространство решений:

$$\left(\sum_{i=1}^{n} x^i e_i\right) \wedge \left(\sum T^{i_1 \ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}\right) = 0.$$

В ней n неизвестных и $\binom{n}{p+1}$ уравнений. Ее матрица состоит из целочисленных линейных комбинаций $T^{i_1 \ldots i_p}$. Ранг этой матрицы всегда $\geqslant n - p$, ибо $\dim \operatorname{Ann} T \leqslant p$. Поэтому условие разложимости равносильно тому, чтобы ранг был $\leqslant n - p$, т. е. обращению в нуль всех ее миноров $(n - p + 1)$-го порядка. Это и есть искомая система уравнений на грассмановы координаты разложимого тензора.

Рассмотрим несколько примеров и частных случаев.

17. Предложение. *Любой $(n-1)$-вектор T разложим.*

Доказательство. Очевидно, $x \wedge T = f(x) e_1 \wedge \ldots \wedge e_n$, где $f(x)$ — линейная функция на L; $\{e_1, \ldots, e_n\}$ — фиксированный базис L. Значит, $\dim \operatorname{Ann} T = \dim \operatorname{Ker} f \geqslant n-1$. Но если $T \neq 0$, то $f \neq 0$, так что $\dim \operatorname{Ker} f = n-1$. В силу утверждения а) теоремы п. 16 T разложим.

В терминах грассмановых многообразий это означает, что имеется биекция

$$(\text{гиперплоскости в } L) \to P(\Lambda^{n-1} L).$$

Но гиперплоскости в L — это точки $P(L^*)$. Поэтому

$$P(L^*) \simeq P(\Lambda^{n-1}(L))$$

(канонический изоморфизм). Ниже мы обобщим этот результат.

18. Предложение. *Ненулевой бивектор $T \in \Lambda^2(L)$ разложим тогда и только тогда, когда $T \wedge T = 0$.*

Доказательство. Необходимость очевидна. Для доказательства достаточности проведем индукцию по n, начиная с тривиального случая $n = 2$. Пусть $\{e_1, \ldots, e_{n+1}\}$ — базис L. Разложив T по $e_i \wedge e_j$, мы можем представить T в виде $T = e_{n+1} \wedge T_1 + T_2$, где T_1 и T_2 разлагаются по e_i, $e_i \wedge e_j$, $1 \leqslant i, j \leqslant n$. Из условия $T \wedge T = 0$ следует, что

$$T_2 \wedge T_2 + 2e_{n+1} \wedge T_1 \wedge T_2 = 0,$$

ибо $(e_{n+1} \wedge T_1) \wedge (e_{n+1} \wedge T_1) = 0$ и T_2 лежит в центре $\Lambda(L)$. Но $T_2 \wedge T_2$ не может содержать членов с e_{n+1}, поэтому

$$T_2 \wedge T_2 = e_{n+1} \wedge T_1 \wedge T_2 = 0.$$

Поскольку $T_2 \wedge T_2 = 0$, по индуктивному предположению T_2 разложим. Так как $T_1 \wedge T_2$ не содержит членов с e_{n+1}, имеем $T_1 \wedge T_2 = 0$. Значит, T_1 лежит в двумерном аннуляторе T_2, и $T_2 = T_1' \wedge T_1$. Поэтому

$$T = e_{n+1} \wedge T_1 + T_1' \wedge T_1 = (e_{n+1} + T_1') \wedge T_1,$$

что и завершает доказательство.

Этот результат снова дает информацию о грассмановых многообразиях, на этот раз о $\mathrm{Gr}(2, L)$:

19. Следствие. *Каноническое отображение $\mathrm{Gr}(2, L) \to P(\Lambda^2(L))$ отождествляет при $n \geqslant 3$ грассманиан плоскостей в L с пересечением квадрик в $P(\Lambda^2(L))$.*

Доказательство. Плоскости в L отвечают прямым разложимых 2-векторов $\Lambda^2(L)$. Условие разложимости 2-вектора $\sum_{i_1 < i_2} T^{i_1 i_2} e_{i_1} \wedge e_{i_2}$ согласно предложению п. 18 имеет вид

$$\Big(\sum_{i_1 < i_2} T^{i_1 i_2} e_{i_1} \wedge e_{i_2}\Big) \wedge \Big(\sum_{j_1 < j_2} T^{j_1 j_2} e_{j_1} \wedge e_{j_2}\Big) = 0,$$

т. е.

$$\sum T^{i_1 i_2} T^{j_1 j_2} \varepsilon(i_1, i_2, j_1, j_2) = 0,$$

где каждая сумма слева отвечает одной четверке индексов $1 \leqslant k_1 < k_2 < k_3 < k_4 \leqslant n$ и $\varepsilon(i_1, i_2, j_1, j_2)$ есть знак перестановки множества $\{i_1, i_2, j_1, j_2\} = \{k_1, k_2, k_3, k_4\}$, размещающей эту четверку в порядке возрастания.

В частности, при $n = 4$ получается одно уравнение:

$$T^{12}T^{34} - T^{13}T^{24} + T^{14}T^{23} = 0.$$

Иными словами, $\mathrm{Gr}(2, \mathcal{K}^4)$ есть четырехмерная квадрика в $P(\Lambda^2(\mathcal{K}^4)) = P(\mathcal{K}^6)$. Она называется квадрикой Плюккера.

20. Внешнее умножение и двойственность. Пусть $\dim L = n$. Согласно следствию п. 5 и известной симметрии биномиальных коэффициентов

$$\dim \Lambda^p(L) = \binom{n}{p} = \binom{n}{n-p} = \dim \Lambda^{n-p}(L)$$

для всех $1 \leqslant p \leqslant n$. Это наводит на мысль, что между $\Lambda^p(L)$ и $\Lambda^{n-p}(L)$ должен существовать либо канонический изоморфизм, либо каноническая двойственность. С точностью до небольшой детали верно второе.

Рассмотрим операцию внешнего умножения

$$\Lambda^p(L) \times \Lambda^{n-p}(L) \to \Lambda^n(L) : (T_1, T_2) \to T_1 \wedge T_2.$$

Поскольку она билинейна, она определяет линейное отображение

$$\Lambda^p(L) \to \mathcal{L}(\Lambda^{n-p}L, \Lambda^n L) \cong (\Lambda^{n-p}(L))^* \otimes \Lambda^n L$$

(последний изоморфизм — частный случай описанного в п. 5 § 2). Ядро этого отображения нулевое. Действительно, пусть $\{e_1, \ldots \ldots, e_n\}$ — базис L. Положим $T_1 \wedge T_2 = (T_1, T_2)\ e_1 \wedge \ldots \wedge e_n$, где $T_1 \in \Lambda^p(L)$, $T_2 \in \Lambda^{n-p}(L)$. Очевидно, (T_1, T_2) — билинейное скалярное произведение между $\Lambda^p(L)$ и $\Lambda^{n-p}(L)$. Построим в $\Lambda^p(L)$ и $\Lambda^{n-p}(L)$ базисы из разложимых p-векторов и $(n-p)$-векторов $\{e_{i_1} \wedge \ldots \wedge e_{i_p}\}$, $\{e_{j_1} \wedge \ldots \wedge e_{j_{n-p}}\}$, $1 \leqslant i_1 < \ldots < i_p \leqslant n$, $1 \leqslant j_1 < \ldots < j_{n-p} \leqslant n$. Отождествим $\Lambda^p(L)$ и $\Lambda^{n-p}(L)$ с помощью линейного отображения, которое ставит в соответствие p-вектору $e_{i_1} \wedge \ldots \wedge e_{i_p}$ $(n-p)$-вектор $e_{j_1} \wedge \ldots \wedge e_{j_{n-p}}$, для которого $\{i_1, \ldots, i_p, j_1, \ldots, j_{n-p}\} = \{1, \ldots, n\}$. Тогда (T_1, T_2) будет скалярным произведением на $\Lambda^p(L)$ с диагональной матрицей Грама вида $\mathrm{diag}(\pm 1, \ldots, \pm 1)$. Оно невырождено, в частности, его левое ядро равно нулю.

Итак, мы построили канонические изоморфизмы

$$\Lambda^p(L) \to (\Lambda^{n-p}(L))^* \otimes \Lambda^n(L).$$

При $p = n-1$ получаем $\Lambda^{n-1}(L) \to L^* \otimes \Lambda^n(L)$, что и объясняет изоморфизм $P(\Lambda^{n-1}(L)) \to P(L^*)$ из п. 18: тензорное умножение L^* на одномерное пространство $\Lambda^n(L)$ «не меняет» множество прямых.

В следующем параграфе мы продолжим изучение связи внешнего умножения с двойственностью, введя в рассмотрение внешнюю алгебру $\Lambda(L^*)$.

§ 7. Внешние формы

1. Пусть L — конечномерное линейное пространство над полем $\mathscr{K}$, L^* — двойственное к нему пространство.

Элементы p-й внешней степени $\Lambda^p(L^*)$ называются *внешними p-формами на пространстве L*. В частности, внешние 1-формы — это просто линейные функционалы на L. Для произвольного p можно установить два варианта этого результата:

2. Теорема. *Пространство $\Lambda^p(L^*)$ канонически изоморфно*:

а) $(\Lambda^p(L))^*$, *т. е. пространству линейных функционалов на p-векторах*;

б) *пространству кососимметрических p-линейных отображений $F: \underbrace{L\times\ldots\times L}_{p}\to\mathscr{K}$, т. е. отображений со свойством*

$$F(l_{\sigma(1)}, \ldots, l_{\sigma(p)}) = \varepsilon(\sigma) F(l_1, \ldots, l_p)$$

для всех $\sigma\in S_p$.

Доказательство. Согласно принятому нами определению

$$\Lambda^p(L^*)\subset L^*\otimes\ldots\otimes L^* = T_0^p(L^*).$$

В п. 4 § 2 мы отождествили $T_0^p(L^*)$ с пространством всех p-линейных функций на L^*. При этом отождествлении внешние формы становятся кососимметрическими p-линейными отображениями L. В самом деле, достаточно проверить это на разложимых формах. Для них имеем

$$(f_1\wedge\ldots\wedge f_p)(l_1, \ldots, l_p) = A(f_1\otimes\ldots\otimes f_p)(l_1, \ldots, l_p) =$$
$$= \frac{1}{p!}\sum_{\tau\in S_p}\varepsilon(\tau) f_{\tau(1)}(l_1)\ldots f_{\tau(p)}(l_p).$$

Поэтому

$$(f_1\wedge\ldots\wedge f_p)(l_{\sigma(1)}, \ldots, l_{\sigma(p)}) = \frac{1}{p!}\sum_{\tau\in S_p}\varepsilon(\tau) f_{\tau(1)}(l_{\sigma(1)})\ldots f_{\tau(p)}(l_{\sigma(p)}) =$$
$$= \frac{1}{p!}\sum_{\tau\in S_p}\varepsilon(\tau\sigma) f_{\tau\sigma(1)}(l_{\sigma(1)})\ldots f_{\tau\sigma(p)}(l_{\sigma(p)}) =$$
$$= \varepsilon(\sigma)(f_1\wedge\ldots\wedge f_p)(l_1, \ldots, l_p)$$

Поэтому мы построили линейное вложение $\Lambda^p(L^*)\to$(кососимметрические p-линейные формы на L). Чтобы проверить, что оно является изоморфизмом, достаточно установить совпадение размерности правой части с $\dim\Lambda^p(L^*) = \binom{n}{p}$. Но если в L выбран базис $\{e_1, \ldots, e_n\}$, то любая кососимметрическая p-линейная форма F на L однозначно определяется своими значениями $F(e_{i_1}, \ldots, e_{i_p})$, $1\leqslant i_1 < \ldots < i_p\leqslant n$, и их можно выбирать любыми. Поэтому размерность пространства таких форм равна $\binom{n}{p}$. Это доказывает утверждение б) теоремы.

Для доказательства утверждения а) отождествим $\underbrace{L^* \otimes \ldots \otimes L^*}_{p}$ с $(\underbrace{L \otimes \ldots \otimes L}_{p})^*$ снова с помощью конструкции п. 4 § 2 и ограничим каждый элемент $\Lambda^p(L^*)$ (как линейную функцию на $L \otimes \ldots \otimes L$) на подпространство p-векторов $\Lambda^p(L)$. Мы получим линейное отображение $\Lambda^p(L^*) \to (\Lambda^p(L))^*$. Поскольку размерности пространства слева и справа совпадают, достаточно проверить, что оно сюръективно. Пространство линейных функционалов на $\Lambda^p(L)$ порождено функционалами вида $\frac{1}{p!}\delta_I$, где

$$I = \{i_1, \ldots, i_p\} \subset \{1, \ldots, n\},\ \delta_I(e_{i_1} \wedge \ldots \wedge e_{i_p}) = 1,$$
$$\delta_I(e_{j_1} \wedge \ldots \wedge e_{j_p}) = 0,$$

если $\{j_1, \ldots, j_p\} \neq I$. Мы утверждаем, что такой функционал является образом p-формы $e^{i_1} \wedge \ldots \wedge e^{i_p} \in \Lambda^p(L^*)$, где, как обычно, $\{e^i\}$ означает базис, двойственный к $\{e_i\}$. Действительно, значение $e^{i_1} \wedge \ldots \wedge e^{i_p}$ на $e_{j_1} \wedge \ldots \wedge e_{j_p}$ равно

$$A(e^{i_1} \otimes \ldots \otimes e^{i_p})(A(e_{j_1} \otimes \ldots \otimes e_{j_p})) =$$
$$= \frac{1}{(p!)^2} \sum_{\sigma, \tau \in S_p} \varepsilon(\sigma\tau) e^{i_{\sigma(1)}}(e_{j_{\tau(1)}}) \ldots e^{i_{\sigma(p)}}(e_{j_{\tau(p)}}).$$

Справа отличны от нуля лишь те слагаемые, для которых $i_{\sigma(1)} = j_{\tau(1)}, \ldots, i_{\sigma(p)} = j_{\tau(p)}$, так что вся сумма равна нулю, если $\{i_1, \ldots, i_p\} \neq \{j_1, \ldots, j_p\}$. Если же эти множества совпадают, то вся сумма равна

$$\frac{1}{(p!)^2} \sum_{\sigma \in S_p} \varepsilon(\sigma)^2 = \frac{1}{p!},$$

когда $\{i_1, \ldots, i_p\} = \{j_1, \ldots, j_p\}$ упорядочены по возрастанию. Поэтому $e^{i_1} \wedge \ldots \wedge e^{i_p}$ как функционал на $\Lambda^p(L)$ равен $\frac{1}{p!}\delta_I$, что завершает доказательство.

3. Замечания. а) Принятый нами способ отождествления $\Lambda^p(L^*)$ с $(\Lambda^p(L))^*$ отвечает билинейному отображению

$$\Lambda^p(L^*) \times \Lambda^p(L) \to \mathscr{K},$$

которое на произвольных парах разложимых p-векторов можно представить в виде

$$(f^1 \wedge \ldots \wedge f^p, l_1 \wedge \ldots \wedge l_p) = \frac{1}{p!} \det(f^i(l_j)),$$

$f^i \in L^*$, $l_j \in L$. Действительно, обе стороны полилинейны и кососимметричны в отдельности по f^i и l_j; кроме того, они совпадают для

$$(f^1, \ldots, f^p) = (e^{i_1}, \ldots, e^{i_p}),\ (l_1, \ldots, l_p) = (e_{j_1}, \ldots, e_{j_p}),$$

как было проверено в предыдущем доказательстве.

Иногда в этом скалярном произведении отбрасывают множитель $\frac{1}{p!}$.

б) Одно из отождествлений, установленных в теореме, иногда принимают в качестве *определения* внешней степени. Например, $\Lambda^p(L)$ часто, особенно в дифференциальной геометрии, вводят как пространство кососимметрических p-линейных функционалов на L^*. Общность этой конструкции — промежуточная между общностью первого и второго определений внешней степени из § 6: поскольку она не требует деления на факториалы, она годится для линейных пространств над полями конечной характеристики, а также для свободных модулей над коммутативными кольцами. Но при переходе к общим модулям лишняя дуализация мешает, и второе определение становится предпочтительным.

Результат, аналогичный теореме п. 2, справедлив также для симметрических степеней, и наше предшествующее замечание относится и к ним. В частности, $S^p(L)$ можно определить как пространство симметричных p-линейных функционалов на L^* для пространств над любыми полями и свободных модулей над коммутативными кольцами. В самом общем случае, однако, правильное определение $S^p(L)$ — это определение из п. 9, § 5.

4. Внутреннее произведение. *Внутренним произведением* называется билинейное отображение

$$L \times \Lambda^p(L^*) \to \Lambda^{p-1}(L^*): \ (l, F) \mapsto i(l)F,$$

которое определяется следующим образом. Рассмотрим $F \in \Lambda^p(L^*)$ как кососимметричную p-линейную форму на L, и аналогично $i(l)F$. Тогда по определению

$$i(l)F(l_1, \ldots, l_{p-1}) = F(l, l_1, \ldots, l_{p-1}).$$

Очевидно, правая часть $(p-1)$-линейна и кососимметрична как функция от $l_1, \ldots, l_{p-1}$, а также билинейна как функция от F, l, так что определение корректно. При $p=0$ удобно считать, что $i(l)F=0$. Вместо $i(l)F$ пишут также $l \lrcorner F$.

§ 8. Тензорные поля

1. В этом параграфе мы кратко опишем типичные дифференциально-геометрические ситуации, в которых используется тензорная алгебра.

Рассмотрим некоторую область $U \subset \mathbf{R}^n$ в координатном вещественном пространстве и кольцо C бесконечно дифференцируемых функций с вещественными значениями на U. В частности, координатные функции x^i, $i=1, \ldots, n$, принадлежат C.

2. Определение. *Касательным вектором X_a к U в точке $a \in U$ называется любое линейное отображение X_a: $C \to \mathbf{R}$, удовлетворяющее условиям:*

$X_a f = 0$, если f постоянна в некоторой окрестности a;

$$X_a(fg) = X_a f \cdot g(a) + f(a) \cdot X_a g.$$

Если X_a, Y_a — два касательных вектора в точке a, то любая их вещественная линейная комбинация также является касательным вектором в этой точке:

$$(cX_a + dY_a)(fg) = cX_a(fg) + dY_a(fg) =$$
$$= cX_af \cdot g(a) + cf(a)X_ag + dY_af \cdot g(a) + df(a)Y_ag =$$
$$= (cX_a + dY_a)f \cdot g(a) + f(a)(cX_a + dY_a)g.$$

Поэтому касательные векторы образуют линейное пространство, которое обозначается T_a и называется касательным пространством к U в точке a. Значение X_af называется производной функции f по направлению вектора X_a. Можно доказать, что пространство T_a n-мерно.

3. Определение. *Векторным полем в области U называется такое семейство касательных векторов $X = \{X_a \in T_a | a \in U\}$, что для любой функции $f \in C$ функция на U*

$$a \mapsto X_af$$

также принадлежит C.

Обозначим эту функцию через Xf. Очевидно, касательное поле определяет линейное отображение $X: C \to C$, нулевое на постоянных функциях и такое, что $X(fg) = Xf \cdot g + f \cdot Xg$ для всех $f, g \in C$. Такие отображения называются *дифференцированиями* кольца C в себя.

Наоборот, каждому дифференцированию $X: C \to C$ и точке $a \in U$ отвечает касательный вектор X_a в этой точке: $X_af = (Xf)(a)$. Это устанавливает биекцию между векторными полями на U и дифференцированиями кольца C.

Сумма векторных полей $X + Y$, определенная формулой $(X + Y)f = Xf + Yf$ для всех $f \in C$, является векторным полем. Произведение fX, определяемое формулой $(fX)g = f(Xg)$, где X — векторное поле, $f, g \in C$, является векторным полем. В частности, любая линейная комбинация векторных полей $\sum_{i=1}^{m} f^iX_i$, $f^i \in C$, является векторным полем.

4. Пример. Пусть $\frac{\partial}{\partial x^i}$, $i = 1, \ldots, n$, — классические операторы частных производных. Все они являются векторными полями на U. Имеет место следующий фундаментальный результат, который мы приведем без доказательства:

5. Теорема. *Всякое векторное поле X в связной области $U \subset \mathbf{R}^n$ однозначно представляется в виде $\sum_{i=1}^{n} f^i \frac{\partial}{\partial x^i}$, где $x^1, \ldots, x^n$ — координатные функции на $\mathbf{R}^n$.*

6. На алгебраическом языке это означает, что множество всех векторных полей T в связной области U является свободным модулем ранга n над коммутативным ассоциативным кольцом C бесконечно дифференцируемых функций на U.

Свободные модули конечного ранга над коммутативными кольцами образуют категорию, по своим свойствам чрезвычайно близкую к категории конечномерных пространств над полем. Для них, в частности, проходит вся теория двойственности и все конструкции тензорной алгебры из этой части курса.

Другой вариант, не требующий переноса тензорной алгебры на кольца и модули, но взамен предполагающий развитие некоторой геометрической техники, состоит в том, чтобы рассматривать каждое векторное поле X как семейство векторов $\{X_a | a \in U\}$, лежащих в семействе конечномерных пространств $\{T_a\}$. Тогда все нужные нам операции тензорной алгебры можно строить «поточечно», определив, скажем, $X \otimes Y$ как $\{X_a \otimes Y_a | a \in U\}$.

Оба варианта построения тензорной алгебры совершенно эквивалентны; в приводимых ниже определениях мы будем исходить из первого.

7. Обозначим через T^* C-модуль C-линейных отображений

$$T^* = \mathscr{L}_C(T, C).$$

Он состоит из отображений ω: $S \to C$ со свойством

$$\omega\left(\sum_{i=1}^{m} f^i X_i\right) = \sum_{i=1}^{m} f^i \omega(X_i)$$

для всех $X_i \in T$, $f^i \in C$. Сложение и умножение на элементы производится по стандартным формулам. C-модуль T^* часто обозначается Ω^1 или $\Omega^1(U)$ и называется модулем (дифференциальных) 1-*форм* в области U.

Каждая функция $f \in C$ определяет элемент $df \in \Omega^1$ по формуле

$$(df)(X) = Xf, \quad X \in T.$$

Он называется *дифференциалом функции* f. В частности, мы можем построить дифференциалы координатных функций $dx^1, \ldots, dx^n \in \Omega^1$. Из теоремы п. 5 легко следует

8. **Предложение.** *Любая 1-форма* $\omega \in \Omega^1$ *однозначно представляется в виде линейной комбинации* $\sum_{i=1}^{n} f_i\, dx^i$.

9. Элементы тензорного произведения C-модулей $\underbrace{T^* \otimes \ldots \otimes T^*}_{p} \otimes \underbrace{T \otimes \ldots \otimes T}_{q}$ называются *тензорными полями* типа (p, q), или p раз ковариантными и q раз контравариантными тензорными полями в области U. В дифференциальной геометрии, впрочем, слово «поля» часто опускают и называют тензорные поля просто тензорами.

Из теоремы п. 5 и предложения п. 8 следует, что всякий тензор типа (p, q) однозначно задается своими компонентами $T_{i_1 \ldots i_p}^{j_1 \ldots j_q}$ по формуле

$$T = \sum T_{i_1 \ldots i_p}^{j_1 \ldots j_q}\, dx^{i_1} \otimes \ldots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_q}},$$

где i_k, j_l независимо пробегают значения от 1 до n. В классических обозначениях опускаются все символы в правой части, кроме компонент $T^{j_1 \dots j_q}_{i_1 \dots i_p}$: этот знак и служит обозначением тензора. Подчеркнем еще раз, что здесь $T^{j_1 \dots j_q}_{i_1 \dots i_p}$ суть не числа, а вещественные бесконечно дифференцируемые функции на U.

10. Замена координат. Первый вклад анализа в изучение тензорных полей состоит в возможности делать нелинейные замены координатных функций в U: переходить от $x^1, \dots, x^n$ к $y^1, \dots, y^n$, где $y^i = y^i(x^1, \dots, x^n)$ — бесконечно дифференцируемые функции такие, что обратные функции $x^j = x^j(y^1, \dots, y^n)$ определены и бесконечно дифференцируемы. Дело в том, что компоненты векторных полей и 1-форм при этом все равно преобразуются по классическим формулам линейно, только с коэффициентами, изменяющимися от точки к точке: согласно формуле дифференцирования сложной функции имеем

$$\frac{\partial}{\partial x^j} = \frac{\partial y^k}{\partial x^j} \frac{\partial}{\partial y^k}$$

(справа подразумевается суммирование по k), а также

$$dx^j = \frac{\partial x^j}{\partial y^k} dy^k$$

(то же соглашение). Поэтому тензор $T^{j_1 \dots j_q}_{i_1 \dots i_p}$ в новых координатах имеет компоненты

$$(T')^{j'_1 \dots j'_q}_{i'_1 \dots i'_q} = \frac{\partial x^{i_1}}{\partial y^{i'_1}} \dots \frac{\partial x^{i_p}}{\partial y^{i'_q}} \frac{\partial y^{j'_1}}{\partial x^{j_1}} \dots \frac{\partial y^{j'_q}}{\partial x^{j'_q}} T^{j_1 \dots j_p}_{i_1 \dots i_p}$$

(то же соглашение с суммированием справа).

Все алгебраические конструкции и языковые соглашения § 4 можно теперь перенести на тензорные поля.

Мы закончим этот параграф несколькими примерами тензорных полей, играющих особенно важную роль в геометрии и физике.

11. Метрический тензор. Этот тензор, обозначаемый g_{ij} или, в более полной записи, $\sum_{i,j=1}^{n} g_{ij}\, dx^i \otimes dx^j$, предполагается симметричным и невырожденным в каждой точке $a \in U$, т. е. $\det(g_{ij}(a)) \neq \neq 0$. Он задает ортогональную структуру в каждом касательном пространстве T_a, и пары (U, g_{ij}) (а также обобщения на случай многообразий, «склеенных» из нескольких областей U) составляют основной объект изучения римановой геометрии, а в случае $n = 4$ и метрики сигнатуры (1, 3) — общей теории относительности.

Метрика используется для измерения длин дифференцируемых кривых $\{x^1(t), \dots, x^n(t) \mid t_0 \leqslant t \leqslant t_1\}$: длина задается формулой

$\int\limits_{t_0}^{t_1} \sqrt{g_{ij}(x^k(t)) \frac{\partial x^i}{\partial t} \frac{\partial x^j}{\partial t}}\, dt$, а также для поднятия и опускания индексов тензорных полей.

12. Внешние формы и форма объема. Элементы из $\Lambda^p(\Omega^1)$, т. е. кососимметрические тензоры типа $(p, 0)$, называются внешними p-формами в U, а внешние n-формы называются *формами объема.* Это название объясняется возможностью определить «криволинейные интегралы» $\int\limits_V f(x_1, \ldots, x_n)\, dx^1 \wedge \ldots \wedge dx^n$ по любой подобласти U, обладающие свойствами меры. В случае $f = 1$ значение такого интеграла есть евклидов объем области V, свойства которого мы описали в § 5 ч. 2.

При $p < n$ можно определить интеграл от любой формы $\omega \in \Lambda^p(\Omega^1)$ по «p-мерным дифференцируемым гиперповерхностям» в U. Все модули внешних форм связаны замечательными операторами «внешнего дифференциала» d^p: $\Lambda^p \to \Lambda^{p+1}$, который в координатах задается формулой

$$d^p\left(\sum f_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}\right) = \sum \frac{\partial f}{\partial x^{i_{p+1}}} dx^{i_{p+1}} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Эти операторы удовлетворяют условию $d^{p+1} \circ d^p = 0$ и входят в формулировку обобщенной теоремы Стокса, связывающей интеграл по p-мерной гиперповерхности с границей V с интегралом по ее границе ∂V:

$$\int\limits_{V^p} d\omega^{p-1} = \int\limits_{\partial V^p} \omega^{p-1}.$$

Особую роль играют внешние 2-формы ω^2, удовлетворяющие условию $d\omega^2 = 0$. В их терминах инвариантно формулируется аппарат гамильтоновой техники.

§ 9. Тензорные произведения в квантовой механике

1. Объединение систем. Роль тензорных произведений в квантовой механике объясняется следующим фундаментальным положением, которое продолжает серию постулатов, сформулированных в п. 8 § 6 и пп. 1—6 § 9 ч. 2.

Пусть $\mathscr{H}_1, \ldots, \mathscr{H}_n$ — пространства состояний нескольких квантовых систем. Тогда пространство состояний системы, получающейся в результате их объединения, является некоторым подпространством $\mathscr{H} \subset \mathscr{H}_1 \otimes \ldots \otimes \mathscr{H}_n$.

Строго говоря, в бесконечномерном случае вместо тензорного произведения справа должно стоять пополненное тензорное произведение гильбертовых пространств, но мы пренебрежем этой тонкостью, работая, как обычно, с конечномерными модулями.

Какое именно подпространство в $\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$ отвечает объединенной системе, приходится решать на основе дальнейших правил, к которым мы обратимся ниже. Здесь же мы рассмотрим случай $\mathcal{H} = \mathcal{H}_1 = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$ и попытаемся объяснить, как уже первый постулат квантовой механики — *принцип суперпозиции* — приводит к совершенно неклассическим связям между системами. Для этого яснее представим себе, каковы могут быть состояния объединенной системы. Пусть $\psi_i \in \mathcal{H}_i$ — некоторые состояния подсистем. Тогда разложимый тензор $\psi_1 \otimes \ldots \otimes \psi_n$ является одним из возможных состояний объединенной системы, и мы можем считать, что оно отвечает случаю, когда каждая из подсистем находится в своем состоянии ψ_i. Но такие разложимые состояния далеко не исчерпывают всех векторов в $\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$: допустимы их произвольные линейные комбинации. Когда объединенная система находится в одном из таких неразложимых состояний, представление о ее подсистемах теряет смысл, ибо они и их состояния не могут быть однозначно выделены. Иными словами, в подавляющем большинстве состояний объединенной системы подсистемы существуют лишь «виртуально».

Важно подчеркнуть, что этот вывод никак не использует представлений о взаимодействии подсистем в классическом смысле слова, подразумевающем обмен энергией между ними. Эйнштейн, Розен и Подольский предложили мысленный эксперимент, в котором две подсистемы объединенной системы после ее распада оказываются сильно разделены пространственно, и наблюдение над одной подсистемой позволяет мгновенно «перевести в определенное состояние» вторую подсистему, хотя классическое взаимодействие между ними требует конечного времени. Это следствие постулатов суперпозиции и тензорного произведения резко противоречит классической интуиции. Тем не менее их принятие привело к огромному количеству теоретических схем, правильно объясняющих действительность, и приходится доверять им и вырабатывать новую интуицию.

Заметим попутно, что описание взаимодействия требует введения гамильтониана объединенной системы. В простейшем случае он имеет «свободный» вид

$$H_1 \otimes \mathrm{id} \otimes \ldots \otimes \mathrm{id} + \mathrm{id} \otimes H_2 \otimes \ldots \otimes \mathrm{id} + \ldots + \mathrm{id} \otimes \ldots \otimes H_n,$$

где $H_i\colon \mathcal{H}_i \to \mathcal{H}_i$ — гамильтониан i-й системы, id — тождественные отображения. В этом случае говорят, что системы не взаимодействуют. Некоторое объяснение этому состоит в замечании, что если объединенная система имеет такой гамильтониан и в начальный момент времени находится в разложимом состоянии $\psi_1 \otimes \ldots \otimes \psi_n$, то в любой момент времени t она будет находиться в разложимом состоянии $e^{-itH_1}(\psi_1) \otimes \ldots \otimes e^{-itH_n}(\psi_n)$, т. е. ее подсистемы будут развиваться независимо друг от друга. В общем случае гамильтониан представляет собой сумму свободной части и оператора, который отвечает за взаимодействие.

2. Неразличимость. Имеются два фундаментальных случая, когда пространство состояний объединенной системы не совпадает с полным пространством $\mathscr{H}_1 \otimes \ldots \otimes \mathscr{H}_n$. В обоих случаях объединяемые системы тождественны, или неразличимы, скажем, являются элементарными частицами одного типа; в частности, $\mathscr{H}_1 = \ldots \ldots = \mathscr{H}_n = \mathscr{H}$.

а) *Бозоны.* По определению, система с пространством состояний $\mathscr{H}$ называется бозоном, если пространство состояний объединения n систем есть n-я симметрическая степень $S^n(\mathscr{H})$.

Согласно эксперименту бозонами являются фотоны и альфа-частицы (ядра гелия).

б) *Фермионы.* По определению, система с пространством состояний $\mathscr{H}$ называется фермионом, если пространство состояний объединения n таких систем есть n-я внешняя степень $\Lambda^n(\mathscr{H})$.

Согласно эксперименту фермионами являются электроны, протоны, нейтроны.

3. Числа заполнения и принцип Паули. Пусть $\{\psi_1, \ldots, \psi_m\}$ — базис пространства состояний бозонной или фермионной системы. Тогда элементы симметризованного (или антисимметризованного) тензорного базиса в $S^n(\mathscr{H})$ (или $\Lambda^n(\mathscr{H})$) физики записывают в виде

$$|a_1, \ldots, a_m\rangle = \begin{cases} S(\underbrace{\psi_1 \otimes \ldots \otimes \psi_1}_{a_1} \otimes \ldots \otimes \underbrace{\psi_m \otimes \ldots \otimes \psi_m}_{a_m}) & \text{в } S^n(\mathscr{H}), \\ A(\underbrace{\psi_1 \otimes \ldots \otimes \psi_1}_{a_1} \otimes \ldots \otimes \underbrace{\psi_m \otimes \ldots \otimes \psi_m}_{a_m}) & \text{в } \Lambda^n(\mathscr{H}). \end{cases}$$

В обоих случаях $a_1 + \ldots + a_m = n$, но для бозонов числа a_i могут принимать любые целые неотрицательные значения, а для фермионов — только 0 или 1: иначе соответствующие антисимметризации равны нулю и не определяют квантовое состояние.

Числа a_i называются «*числами заполнения*» соответствующего состояния. Подразумевается, что в состоянии $|a_1, \ldots, a_m\rangle$ объединенной системы можно условно считать, что a_i подсистем находятся в состоянии ψ_i. Поскольку, однако ни в фермионном, ни в бозонном случае объединенная система вообще не может находиться в состоянии, описываемом разложимым тензором $\psi_1 \otimes \ldots \ldots \otimes \psi_m$, кроме случая, когда все ψ_i одинаковы (для бозонов), это означает, что даже в базисных состояниях $|a_1, \ldots, a_m\rangle$ нельзя сказать, «которая» из подсистем находится, скажем, в состоянии ψ_i. Подсистемы являются неразличимыми.

Условие $a_i = 0$ или 1 в фермионном случае интерпретируется как утверждение о том, что две подсистемы не могут находиться в одинаковом состоянии. Это знаменитый «принцип запрета» Паули.

Когда число n очень велико, ряд физически важных утверждений о пространствах $S^n(\mathscr{H})$ и $\Lambda^n(\mathscr{H})$ делается в вероятностных терминах, скажем, в терминах доли состояний $|a_1, \ldots, a_n\rangle$ с теми или иными условиями относительно чисел заполнения. Поэтому часто говорят, что бозоны и фермионы подчиняются разным *статистикам* — соответственно Бозе — Эйнштейна или Ферми.

4. Случай переменного числа частиц. В процессе эволюции квантовой системы составляющие ее «элементарные подсистемы», или частицы, могут рождаться или уничтожаться. Для описания таких эффектов в бозонном и фермионном случае используются соответственно пространства состояния $\bigoplus_{i=1}^{\infty} S^i(\mathscr{H})$ (точнее, пополнение этого пространства) или $\bigoplus_{i=0}^{\infty} \Lambda^i(\mathscr{H})$, т. е. полная симметрическая или внешняя алгебра одночастичного пространства $\mathscr{H}$.

Оператор, умножающий векторы из $S^n(\mathscr{H})$ (соответственно, из $\Lambda^n(\mathscr{H})$) на n ($n = 0, 1, 2, 3, \ldots$), называется *оператором числа частиц*. Его ядро — подпространство $C = S^0(\mathscr{H})$ или $\Lambda^0(\mathscr{H})$ — называется *вакуумным состоянием*: в нем частиц нет.

Совершенно фундаментальную роль играют также специальные операторы *рождения* и *уничтожения частиц*. Оператор $a^-(\psi_0)$ уничтожения бозона в состоянии $\psi_0 \in \mathscr{H}$ действует на состояние $S(\psi_1 \otimes \ldots \otimes \psi_n)$ по формуле

$$a^-(\psi_0) S(\psi_1 \otimes \ldots \otimes \psi_n) = \sqrt{n+1} \cdot \frac{1}{n!} \sum_{\sigma \in S_n} (\psi_0, \psi_{\sigma(1)}) \otimes \psi_{\sigma(2)} \otimes \ldots \otimes \psi_{\sigma(n)},$$

где $(\psi_0, \psi_{\sigma(1)})$ — скалярное произведение в $\mathscr{H}$. Оператор $a^+(\psi_0)$ рождения бозона в состоянии $\psi_0 \in \mathscr{H}$ определяется как сопряженный к $a^-(\psi_0)$ в смысле эрмитовой геометрии. Аналогичные формулы можно написать в фермионном случае. Роль этих стандартных операторов тензорной алгебры объясняется тем, что в их терминах удобно записывать операторы важных наблюдаемых, в первую очередь гамильтонианы.

УПРАЖНЕНИЯ

В следующей серии упражнений изложены основные факты теории тензорного ранга, важной для оценок сложности вычислений. Основы этой теории заложил Ф. Штрассен.

1. Пусть $L_1, \ldots, L_n$ — конечномерные линейные пространства над полем $\mathscr{K}$, $t \in L_1 \otimes \ldots \otimes L_n$, $t \neq 0$. Рангом $\operatorname{rk} t$ тензора t называется такое наименьшее число r, что для подходящих векторов $l_i^{(j)} \in L_i$, $j = 1, \ldots, r$,

$$t = \sum_{j=1}^{r} l_1^{(j)} \otimes \ldots \otimes l_n^{(j)}.$$

Очевидно, при $n = 1$ имеем $\operatorname{rk} t = 1$ для любого $t \neq 0$.

Пусть $t \in L_1^* \otimes L_2 = \mathscr{L}(L_1, L_2)$ (см. п. 5 § 2). Доказать, что

$$\operatorname{rk} t = \dim \operatorname{Im} t, \qquad t\colon L_1 \to L_2.$$

Вывести отсюда следующие факты:

а) при $n = 2$ ранг t остается инвариантным при расширении основного поля;

б) при $n=2$ множество $\{t \mid \operatorname{rk} t \leqslant r\}$ задается конечной системой уравнений $P_{j,r}(t^{i_1 \ldots i_n})=0$, где $P_{j,r}$ — многочлены от координат.

Оба этих факта перестают быть верными для случая $n=3$, который представляет основной интерес в теории сложности вычислений; см. ниже упражнения 4 — 9.

2. Пусть $L = \bigoplus_{i,j=1}^{2} \mathbf{C} a_{ij}$ — пространство комплексных матриц размера 2×2. Доказать, что

$$\operatorname{rk}\left(\sum_{i,j,k=1}^{2} a_{ij} \otimes a_{jk} \otimes a_{ki}\right) = 7.$$

Указание. Воспользоваться упражнением 12 к § 4 ч. 1. Это же указание относится к следующей задаче.

3. Доказать, что

$$\operatorname{rk}\left(\sum_{i,j,k=1}^{N} a_{ij} \otimes a_{jk} \otimes a_{ki}\right) \leqslant c N^{\log_2 7}$$

для подходящей константы c. (Здесь $L = \bigoplus_{i,j=1}^{N} \mathbf{C} a_{ij}$; интересующий нас тензор есть $\operatorname{tr} A \otimes A \otimes A$, $A = (a_{ij})$ — общая матрица N-го порядка.)

4. Пусть L — некоторая конечномерная $\mathcal{K}$-алгебра, $L \underset{\mathcal{K}}{\otimes} L \to L$: $a \otimes b \longmapsto ab$ — ее закон умножения. Рассмотрим этот закон как тензор $t \in L^* \otimes L^* \otimes L$. (Его координаты — структурные константы алгебры.) Вычислить $\operatorname{rk} t$ для случая $\mathcal{K} = \mathbf{R}$, $L = \mathbf{C}$.

5. В обозначениях предыдущей задачи пусть $L = \mathcal{K}^n$, умножение покоординатное:

$$(a_1, \ldots, a_n)(b_1, \ldots, b_n) = (a_1 b_1, \ldots, a_n b_n).$$

Вычислить $\operatorname{rk} t$.

6. Пользуясь результатами упражнений 4 и 5, убедитесь, что ранг тензора структурных констант алгебры $\mathbf{C}$ над $\mathbf{R}$ падает при расширении основного поля до $\mathbf{C}$.

Указание. $\mathbf{C} \underset{\mathbf{R}}{\otimes} \mathbf{C}$ изоморфна $\mathbf{C}^2$ как $\mathbf{C}$-алгебра.

7. Пусть $L = \mathbf{C} e_1 \oplus \mathbf{C} e_2$. Доказать, что тензор

$$t = e_1 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_2 + e_2 \otimes e_1 \otimes e_2$$

имеет ранг 3.

8. Доказать, что тензор t из предыдущего упражнения является пределом некоторой последовательности тензоров ранга 2.

Указание.

$$t + \varepsilon e_2 \otimes e_2 \otimes e_2 = \frac{1}{\varepsilon}\left[e_1 \otimes e_1 \otimes (-e_2 + \varepsilon e_1) + (e_1 + \varepsilon e_2) \otimes (e_1 + \varepsilon e_2) \otimes e_2\right].$$

9. Вывести из упражнений 7 и 8, что множество тензоров ранга $\leqslant 2$ в $L \otimes L \otimes L$ не задается системой уравнений вида

$$P_j(t^{i_1 i_2 i_3}) = 0,$$

где P_j — многочлены.

10. Назовем предельным рангом $\operatorname{brk}(t)$ тензора t такое наименьшее число s, что t можно представить в виде предела последовательности тензоров ранга $\leqslant s$. Доказать, что для общей матрицы A порядка 3×3

$$\operatorname{brk}(\operatorname{tr} A \otimes A \otimes A) \leqslant 21.$$

11. Чему равны

$$\mathrm{rk}\,(\mathrm{tr}\, A \otimes A \otimes A),\ \mathrm{brk}\,(\mathrm{tr}\, A \otimes A \otimes A),$$

где A — общая матрица порядка $N \times N$? (К моменту, когда пишутся эти строки, ответ не известен даже для $N = 3$.)

12. Пусть L — n-мерное линейное пространство над полем $\mathscr{K}$, $M \subset \Lambda^2(L)$ — произвольное подпространство. Предположим, что для каждого $v \in L$, $v \neq 0$, существует $w \in L$ с $0 \neq v \wedge w \in M$. Доказать, что в L можно выбрать такой базис $e_1, \ldots, e_n$, что

$$M + \Lambda^2(L_i) = \Lambda^2(L), \qquad 1 \leqslant i \leqslant n,$$

где $L_i = \bigoplus\limits_{j \neq i} \mathscr{K} e_j$.

В случае поля $\mathscr{K} = \mathbf{F}_p$ из p элементов известно весьма непростое комбинаторное доказательство этого результата (Vaughan-Lee M. R. — J. Algebra, 1974, **32**, p. 278—285), допускающего теоретико-групповую интерпретацию. Желательно найти более прямой подход.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ