МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Физический факультет

Кафедра высшей математики физического факультета

А. П. Ульянов

ГЕОМЕТРИЯ И АЛГЕБРА ДЛЯ СТУДЕНТОВ-ФИЗИКОВ

Семестр 2

Учебное пособие
по курсу линейной алгебры и геометрии

Версия от 5 июня 2020 г.

Новосибирск
2007–2020

УДК: 510
ББК: В14я73-1 + В151.54я73-1
У 517

Ульянов А. П. Геометрия и алгебра для студентов-физиков: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2020. Ч. 2. 137?? с.

ISBN 978-5-94356-571-7

Пособие содержит краткий конспект лекций, читаемых автором для студентов 1-го курса физического и геолого-геофизического факультетов НГУ в весеннем семестре 2020 года.

Наполняется постепенно...

Версия от 5 июня 2020 г.

ISBN 978-5-94356-571-7

Содержание

Глава 7. ЛИНЕЙНЫЕ ОПЕРАТОРЫ

7.1. Линейные отображения

Отображения пространств столбцов и матрицы

Возьмём любую $m \times n$ матрицу A. С помощью умножения матрицы на столбец определим отображение

$$\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^m, \quad X \mapsto AX,$$

пространств столбцов. Оно обладает свойством **линейности**:

$$\varphi_A(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 \varphi_A(X_1) + \alpha_2 \varphi_A(X_2).$$

Следовательно, для произвольных линейных комбинаций имеем

$$\varphi_A(\alpha_1 X_1 + \ldots + \alpha_k X_k) = \alpha_1 \varphi_A(X_1) + \ldots + \alpha_k \varphi_A(X_k).$$

Также можно записать это отображение через столбцы $A^{(j)}$:

$$\varphi_A([x_1, \ldots, x_n]^\top) = x_1 A^{(1)} + \ldots + x_n A^{(n)}.$$

Отсюда видно, что образы $\varphi_A(E^{(j)})$ столбцов стандартного базиса пространства $\mathbb{R}^n$ есть как раз столбцы $A^{(j)}$.

Наоборот, всякое линейное отображение $\varphi \colon \mathbb{R}^n \to \mathbb{R}^m$ задаёт $m \times n$ матрицу, столбцы которой есть образы $\varphi(E^{(j)})$.

Теорема. *Между линейными отображениями $\mathbb{R}^n$ в $\mathbb{R}^m$ и $m \times n$ матрицами имеется взаимно однозначное соответствие.*

Доказательство. Упражнение! Проверьте, что указанные переходы между матрицей и линейным отображением обратны друг другу. □

примеры

Теорема. *Композиция $\varphi \circ \psi$ линейных отображений $\psi \colon \mathbb{R}^n \to \mathbb{R}^s$ и $\varphi \colon \mathbb{R}^s \to \mathbb{R}^m$ есть линейное отображение $\mathbb{R}^n$ в $\mathbb{R}^m$. Матрица композиции $\varphi \circ \psi$ равна произведению матрицы φ на матрицу ψ.*

Доказательство. Проверяется непосредственно по определению. □

Следствие. *Умножение матриц ассоциативно.* □

Богатую специфику теория линейных отображений имеет в двух случаях. В них применяются особые термины:

(а) линейные отображения $\mathbb{R}^n$ в себя называют **линейными преобразованиями** или **операторами**;

(b) линейные отображения $\mathbb{R}^n$ в $\mathbb{R}$ называют **линейными функциями** или **функционалами**.

В этом курсе мы уделим много времени изучению операторов, но практически не коснёмся функционалов (они сыграют свою роль позже).

Выбор базиса

В чём отличие **координатного** пространства $\mathbb{R}^n$ от произвольного вещественного линейного пространства размерности n? Фактически оно сводится к наличию стандартного базиса в $\mathbb{R}^n$ и к отсутствию оного в произвольном пространстве. В самом деле, как только в линейном пространстве $\mathcal{V}$ выбран базис, каждый вектор $\mathcal{V}$ однозначно представляется набором своих координат; эти наборы можно понимать уже как элементы $\mathbb{R}^n$ и применять к любой задаче в $\mathcal{V}$ все средства, доступные при работе с $\mathbb{R}^n$.

Более формально можно сказать так: выбор базиса $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ в линейном пространстве $\mathcal{V}$ устанавливает биекцию

$$f\colon \mathbb{R}^n \to \mathcal{V}, \qquad [x_1, \dots, x_n]^\top \mapsto x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n.$$

Проверим, что она является линейным отображением:

$$\begin{aligned} f(\alpha X + \beta Y) &= f(\alpha[x_1, \dots, x_n]^\top + \beta[y_1, \dots, y_n]^\top) \\ &= f([\alpha x_1 + \beta y_1, \dots, \alpha x_n + \beta y_n]^\top) \\ &= (\alpha x_1 + \beta y_1)\mathbf{v}_1 + \dots + (\alpha x_n + \beta y_n)\mathbf{v}_n \\ &= \alpha(x_1\mathbf{v}_1 + \dots + x_n\mathbf{v}_n) + \beta(y_1\mathbf{v}_1 + \dots + y_n\mathbf{v}_n) \\ &= \alpha f(X) + \beta f(Y). \end{aligned}$$

Лемма. *Отображение, обратное к биективному линейному, также линейное.*

Доказательство. Упражнение. □

Изоморфизм

Определение. Биективное линейное отображение $f\colon \mathcal{V} \to \mathcal{W}$ вещественных линейных пространств называют **изоморфизмом**; пространства $\mathcal{V}$ и $\mathcal{W}$ называют **изоморфными** и пишут $\mathcal{V} \simeq \mathcal{W}$.

Эти понятия применимы также к линейным пространствам над любым полем $\mathbb{F}$; оно лишь должно быть общим для $\mathcal{V}$ и $\mathcal{W}$. Требование общности поля скаляров объясняется тем, что в аксиоме линейности

$$f(\alpha X + \beta Y) = \alpha f(X) + \beta f(Y)$$

для изоморфизма $f\colon \mathcal{V} \to \mathcal{W}$ на те же самые (хотя и произвольные) скаляры слева умножаются векторы в $\mathcal{V}$, а справа — в $\mathcal{W}$.

Теорема. (1) *Каждое n-мерное вещественное линейное пространство изоморфно $\mathbb{R}^n$.*

(2) *Все n-мерные вещественные линейные пространства изоморфны друг другу.*

(3) *Второе утверждение также верно для n-мерных линейных пространств над любым полем $\mathbb{F}$.*

Доказательство. Первое утверждение разобрано выше. Второе и третье следуют соответственно из диаграмм изоморфизмов

поскольку все изоморфизмы обратимы, а композиция изоморфизмов тоже изоморфизм. (Иными словами, изоморфность является отношением эквивалентности.) □

Матрица отображения в базисах

Совладав с линейными отображениями координатных пространств и изоморфизмом между произвольным n-мерным линейным пространством и координатным пространством $\mathbb{R}^n$, естественно перейти к следующему вопросу: как задать линейное отображение $\varphi\colon \mathcal{V} \to \mathcal{W}$ произвольных линейных пространств?

Выберем базисы $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathcal{V}$ и $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \subset \mathcal{W}$. Нарисуем диаграмму со всеми интересующими нас отображениями:

$$\begin{array}{ccc} X = [x_1, \ldots, x_n]^\top & \longmapsto & x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n \\ \;\;?\big\downarrow & & \big\downarrow\varphi \\ Y = [y_1, \ldots, y_m]^\top & \longmapsto & y_1\mathbf{w}_1 + \ldots + y_m\mathbf{w}_m, \end{array} \qquad \begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\;\simeq\;} & \mathcal{V} \\ ?\big\downarrow & & \big\downarrow\varphi \\ \mathbb{R}^m & \xrightarrow{\;\simeq\;} & \mathcal{W}. \end{array}$$

Красный вопросик есть композиция трёх линейных отображений; значит, это тоже линейное отображение. Оно должно задаваться какой-то $m \times n$ матрицей A. Значит, $Y = AX$, или

$$[y_1, \ldots, y_m]^\top = A \cdot [x_1, \ldots, x_n]^\top.$$

Её называют **матрицей отображения** φ в выбранных базисах. Убедитесь, что саму эту матрицу $A = [a_{ij}]$ можно найти по выражениям

$$\varphi(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \ldots + a_{mj}\mathbf{w}_m,$$

то есть в столбце j матрицы отображения записаны координаты образа базисного вектора $\mathbf{v}_j$ в базисе $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. Для $\mathcal{V} = \mathbb{R}^n$ и $\mathcal{W} = \mathbb{R}^m$ со стандартными базисами эта формула даёт ожидаемое:

$$\varphi(E^{(j)}) = a_{1j}E^{(1)} + \ldots + a_{mj}E^{(m)} = A^{(j)}.$$

Пример. Посчитаем в следующей таблице несколько примеров матриц отображений между двумерными пространствами функций.

Поле	*Базис* $\mathcal{V}$	*Базис* $\mathcal{W}$	*Отображение*	*Матрица*
$\mathbb{R}$	$\mathbf{v}_1 = 1$ $\mathbf{v}_2 = \cos t$	$\mathbf{w}_1 = 1$ $\mathbf{w}_2 = \cos^2 t$	$f(t) \mapsto f(2t)$	$\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$
$\mathbb{R}$	$\mathbf{v}_1 = \cos t$ $\mathbf{v}_2 = \sin t$	$\mathbf{w}_1 = \cos t$ $\mathbf{w}_2 = \sin t$	$f(t) \mapsto \frac{d}{dt} f(t)$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
$\mathbb{C}$	$\mathbf{v}_1 = e^{it}$ $\mathbf{v}_2 = e^{-it}$	$\mathbf{w}_1 = e^{it}$ $\mathbf{w}_2 = e^{-it}$	$f(t) \mapsto \frac{d}{dt} f(t)$	$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$
$\mathbb{C}$	$\mathbf{v}_1 = e^{it}$ $\mathbf{v}_2 = e^{-it}$	$\mathbf{w}_1 = \cos t$ $\mathbf{w}_2 = \sin t$	$f(t) \mapsto f(t)$	$\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$

Смена базиса и матрица перехода

Линейное пространство $\mathcal{V}$ имеет много базисов, поэтому необходимо уметь переводить всю информацию, добытую при работе в одном базисе $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, в её эквивалент в другом базисе $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$. Изменение координат векторов изображается диаграммой

$$\begin{array}{ccc} [x_1, \ldots, x_n]^\top & \longmapsto & x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n \\ \;?\uparrow\downarrow & & \| \\ [x'_1, \ldots, x'_n]^\top & \longmapsto & x'_1\mathbf{v}'_1 + \ldots + x'_n\mathbf{v}'_n, \end{array} \qquad \begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\sim} & \\ ?\uparrow\downarrow & & \mathcal{V}. \\ \mathbb{R}^n & \xrightarrow{\sim} & \end{array}$$

Красный вопросик есть композиция двух изоморфизмов; значит, это обратимое линейное преобразование пространства $\mathbb{R}^n$. Оно задаётся $n \times n$ **матрицей перехода** T (непременно обратимой). Таким образом, координаты векторов связаны соотношением $X = TX'$, или

$$[x_1, \ldots, x_n]^\top = T \cdot [x'_1, \ldots, x'_n]^\top,$$

а саму матрицу $T = [t_{ij}]$ можно найти по выражениям

$$\mathbf{v}'_j = t_{1j}\mathbf{v}_1 + \ldots + t_{nj}\mathbf{v}_n,$$

то есть в столбце j матрицы перехода записаны координаты «нового» базисного вектора $\mathbf{v}'_j$ в «старом» базисе $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Закон изменения матрицы отображения

Рассмотрим теперь более сложную задачу. Пусть имеется линейное отображение $\varphi\colon \mathcal{V} \to \mathcal{W}$. Мы уже выяснили, что оно задаётся матрицей, если выбраны базисы пространств $\mathcal{V}$ и $\mathcal{W}$. Что произойдёт с этой матрицей, если сменить базис $\mathcal{V}$? Базис $\mathcal{W}$? Оба базиса?

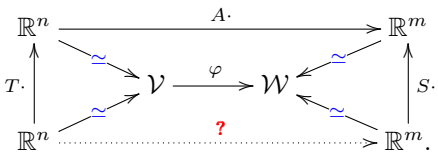
Теорема. *Если в старых базисах линейное отображение $\varphi\colon \mathcal{V} \to \mathcal{W}$ задано матрицей A, а переходы от старых базисов $\mathcal{V}$ и $\mathcal{W}$ к новым заданы соответственно матрицами T и S, то в новых базисах φ задаётся матрицей $S^{-1}AT$.*

Доказательство. Запишем отображение в двух базисах: $Y = AX$ и $Y' = A'X'$. Запишем переходы между базисами: $X = TX'$ и $Y = SY'$. Отсюда получим $Y' = S^{-1}ATX'$. Поскольку все эти равенства верны для любого исходного столбца X, они влекут искомое $A' = S^{-1}AT$. □

Эту простую выкладку можно выразить на языке диаграмм отображений. Скомбинируем использованные ранее диаграммы:

(в старых базисах)

(в новых базисах)



Красный вопросик есть композиция трёх линейных отображений координатных пространств столбцов, заданных матрицами T, A и S^{-1}. При этом S нужно обращать, потому что стрелка проходится в обратную сторону. Эта композиция — тоже линейное отображение, а матрица его есть произведение этих трёх матриц.

7.2. Структура линейного отображения

Недолгое экспериментирование с матрицами показывает, что с виду матрица $S^{-1}AT$ может разительно отличаться от матрицы A. Вывод: структура данного линейного отображения может значительно проясняться при надлежащем выборе базисов.

Элементарные матрицы

Пример. Пусть линейные пространства $\mathcal{V}$ и $\mathcal{W}$ двумерны и линейное отображение $\varphi\colon \mathcal{V} \to \mathcal{W}$ задано в базисах $(\mathbf{v}_1, \mathbf{v}_2)$ и $(\mathbf{w}_1, \mathbf{w}_2)$ матрицей

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}.$$

Изучим поведение матрицы при простых преобразованиях базисов.

Новый базис $\mathcal{V}$	T	*Новый базис* $\mathcal{W}$	S	$S^{-1}AT$	
$\mathbf{v}_1' = \mathbf{v}_1$ $\mathbf{v}_2' = 2\mathbf{v}_2$	$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$	$\mathbf{w}_1' = \mathbf{w}_1$ $\mathbf{w}_2' = \mathbf{w}_2$	E	$\begin{bmatrix} 1 & 6 \\ 2 & 8 \end{bmatrix}$	(C1)
$\mathbf{v}_1' = \mathbf{v}_1$ $\mathbf{v}_2' = \mathbf{v}_1 + \mathbf{v}_2$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\mathbf{w}_1' = \mathbf{w}_1$ $\mathbf{w}_2' = \mathbf{w}_2$	E	$\begin{bmatrix} 1 & 4 \\ 2 & 6 \end{bmatrix}$	(C2)
$\mathbf{v}_1' = \mathbf{v}_1$ $\mathbf{v}_2' = \mathbf{v}_2$	E	$\mathbf{w}_1' = \mathbf{w}_1$ $\mathbf{w}_2' = 2\mathbf{w}_2$	$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 3 \\ 1 & 2 \end{bmatrix}$	(R1)
$\mathbf{v}_1' = \mathbf{v}_1$ $\mathbf{v}_2' = \mathbf{v}_2$	E	$\mathbf{w}_1' = \mathbf{w}_1 + \mathbf{w}_2$ $\mathbf{w}_2' = \mathbf{w}_2$	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$	(R2)

Метки справа от таблицы могут помочь понаблюдать, как элементарные преобразования строк/столбцов матрицы реализуются посредством умножения её слева/справа на квадратные матрицы специального вида. Каждая такая матрица (легко!) обратима и отличается от единичной лишь в одном элементе α, стоящем где-то на главной диагонали для типов R1 и C1 и где-то вне её для типов R2′ и C2′. Назовём такие матрицы **элементарными**.

Простейший вид матрицы отображения

Лекция 1
07.02.20

Теорема. *Для каждой $m \times n$ матрицы A найдутся такие обратимые квадратные матрицы S и T, что*

$$S^{-1}AT = \left[\begin{array}{c|c} E & 0 \\ \hline 0 & 0 \end{array}\right],$$

где (квадратный) единичный блок имеет размер $\operatorname{rk} A$.

Доказательство. Элементарными преобразованиями строк приведём матрицу A к ступенчатому виду A'; в нём $\operatorname{rk} A$ ненулевых строк, а все столбцы есть линейные комбинации главных. В ходе процесса слева накопится какое-то произведение элементарных матриц; оно и будет требуемой матрицей S^{-1}, так что $A' = S^{-1}A$.

Далее, элементарными преобразованиями столбцов A' переставим главные вперёд, а остальные занулим. Результат будет матрицей A''

заявленного в теореме вида. В ходе процесса справа накопится какое-то произведение элементарных матриц; оно и будет требуемой матрицей T, так что $A'' = A'\,T = S^{-1}A\,T$. □

Теорема. *Для каждого линейного отображения $\varphi\colon \mathcal{V} \to \mathcal{W}$ найдутся базисы пространств $\mathcal{V}$ и $\mathcal{W}$, в которых его матрица имеет вид, указанный в предыдущей теореме.*

Первое доказательство. Возьмём сперва базисы произвольно и представим φ матрицей в них. Переходы к искомым базисам указаны матрицами S и T из предыдущей теоремы. □

Второе доказательство. Воспользуемся леммой, отложенной на конец раздела. Возьмём любые базисы $\mathcal{V}_0$, $\mathcal{V}_1$ и $\mathcal{W}_2$, а базис $\mathcal{W}_1$ составим из образов векторов базиса $\mathcal{V}_1$. □

Образ, прообраз, ядро

Отображение множеств $f\colon \mathcal{X} \to \mathcal{Y}$ определяет образ в $\mathcal{Y}$ каждого $\mathcal{X}' \subseteq \mathcal{X}$ и прообраз в $\mathcal{X}$ каждого $\mathcal{Y}' \subseteq \mathcal{Y}$:

$$f(\mathcal{X}') = \{f(x) \in \mathcal{Y} \mid x \in \mathcal{X}'\}; \quad f^{-1}(\mathcal{Y}') = \{x \in \mathcal{X} \mid f(x) \in \mathcal{Y}'\}.$$

Теорема. *При линейном отображении $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ образ каждого линейного подпространства в $\mathbb{R}^n$ есть линейное подпространство в $\mathbb{R}^m$; аналогично для прообразов.*

Доказательство. В образе подпространства $\mathcal{L} \subseteq \mathbb{R}^n$ возьмём любые векторы X'_i. Тогда $X'_i = \varphi(X_i)$ для каких-то $X_i \in \mathcal{L}$, а произвольная линейная комбинация $\alpha_1 X'_1 + \ldots + \alpha_k X'_k$ равна $\varphi(\alpha_1 X_1 + \ldots + \alpha_k X_k)$ ввиду линейности φ, поэтому она также лежит в $\varphi(\mathcal{L})$.

В прообразе подпространства $\mathcal{L}' \subseteq \mathbb{R}^m$ возьмём любые векторы X_i. Тогда $\varphi(X_i) = X'_i \in \mathcal{L}'$, а образ произвольной линейной комбинации $\alpha_1 X_1 + \ldots + \alpha_k X_k$ ввиду линейности φ равен $\alpha_1 X'_1 + \ldots + \alpha_k X'_k$ и попадает в $\mathcal{L}'$, поэтому исходная комбинация лежит в $\varphi^{-1}(\mathcal{L}')$. □

Два подпространства этих типов в ходу особенно часто. **Образ** φ есть $\varphi(\mathbb{R}^n)$ — «образ всего», а **ядро** φ есть $\varphi^{-1}(\{\mathbf{0}\})$ — «прообраз ничего»:

$$\operatorname{Im}\varphi = \{Y \in \mathbb{R}^m \mid Y = \varphi(X) \text{ для какого-то } X \in \mathbb{R}^n\},$$
$$\operatorname{Ker}\varphi = \{X \in \mathbb{R}^n \mid \varphi(X) = \mathbf{0}\}.$$

При этом нетрудно заметить соответствие:

$$\begin{aligned} \text{отображение } \varphi &\leftrightarrow \text{ матрица } A, \\ \text{образ } \varphi &= \text{ пространство столбцов } A, \\ \text{ядро } \varphi &= \text{ пространство решений системы } AX = \mathbf{0}. \end{aligned}$$

Следствие. *Для всякого линейного отображения* $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$

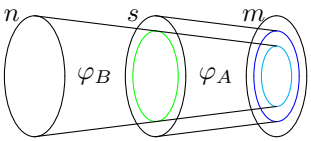
$$\dim \operatorname{Im} \varphi + \dim \operatorname{Ker} \varphi = n.$$

Доказательство. Если вдуматься, это всего лишь перевод на новый язык следствия в теории линейных систем: сумма числа главных столбцов и числа неглавных столбцов равна общему числу столбцов. □

Теорема. *Для* $m \times s$ *матрицы* A *и* $s \times n$ *матрицы* B *всегда*

$$\operatorname{rk} AB \leqslant \min\{\operatorname{rk} A, \operatorname{rk} B\}.$$

Доказательство. Вместо рангов матриц лучше докажем соответствующие неравенства для размерностей образов линейных отображений:



$$\dim \operatorname{Im}(\varphi_A \circ \varphi_B) \leqslant \dim \operatorname{Im} \varphi_A,$$
$$\dim \operatorname{Im}(\varphi_A \circ \varphi_B) \leqslant \dim \operatorname{Im} \varphi_B. \qquad \square$$

Отложенная структурная лемма

Лемма. *Для каждого линейного отображения* $\varphi\colon \mathcal{V} \to \mathcal{W}$ *имеются такие разложения в прямые суммы*

$$\mathcal{V} = \mathcal{V}_0 \oplus \mathcal{V}_1 \quad \text{и} \quad \mathcal{W} = \mathcal{W}_1 \oplus \mathcal{W}_2,$$

что $\varphi\colon \mathcal{V}_1 \to \mathcal{W}_1$ *есть изоморфизм на образ, а* $\varphi(\mathcal{V}_0) = \{\mathbf{0}\}$.

Доказательство. Положим $\mathcal{V}_0 = \operatorname{Ker} \varphi$ и возьмём любое его дополнение до $\mathcal{V}$ за $\mathcal{V}_1$. Далее положим $\mathcal{W}_1 = \varphi(\mathcal{V}_1)$ и возьмём любое его дополнение до $\mathcal{W}$ за $\mathcal{W}_2$. Тогда $\varphi\colon \mathcal{V}_1 \to \mathcal{W}_1$ инъективно, так как $\mathcal{V}_0 \cap \mathcal{V}_1 = \{\mathbf{0}\}$, и сюръективно по определению. □

Итак, структуру отображения φ можно изобразить диаграммой

$$\begin{array}{ccc} \mathcal{V}_0 & \longrightarrow & \{\mathbf{0}\} \\ \oplus & & \\ \mathcal{V}_1 & \xrightarrow{\simeq} & \mathcal{W}_1 \\ & & \oplus \\ \varnothing & \cdots\cdots\cdots\cdots & \mathcal{W}_2. \end{array}$$

7.3. Линейные операторы

Пространство линейных отображений

Будем обозначать через $\mathcal{L}(\mathcal{V}, \mathcal{W})$ множество всех линейных отображений из $\mathcal{V}$ в $\mathcal{W}$. Определим на нём операции сложения двух отображений и умножения отображения на скаляр. Результат каждой такой операции есть линейное отображение:

- $\mathbf{v} \mapsto \varphi_1(\mathbf{v}) + \varphi_2(\mathbf{v})$ задаёт сумму $\varphi_1 + \varphi_2$ отображений φ_1 и φ_2;
- $\mathbf{v} \mapsto \alpha\varphi(\mathbf{v})$ задаёт произведение $\alpha\varphi$ отображения φ на скаляр α.

Далее, можно строить линейные комбинации отображений по правилу

- $\mathbf{v} \mapsto \alpha_1\varphi_1(\mathbf{v}) + \alpha_2\varphi_2(\mathbf{v})$.

Простая рутинная проверка восьми аксиом показывает, что $\mathcal{L}(\mathcal{V}, \mathcal{W})$ есть линейное пространство (над тем же полем, что $\mathcal{V}$ и $\mathcal{W}$).

Лемма. *Размерность пространства линейных отображений есть*

$$\dim \mathcal{L}(\mathcal{V}, \mathcal{W}) = \dim \mathcal{V} \cdot \dim \mathcal{W}.$$

Доказательство. Выбор базисов в $\mathcal{V}$ и $\mathcal{W}$ задаёт изоморфизм $\mathcal{L}(\mathcal{V}, \mathcal{W})$ и пространства матриц с $\dim \mathcal{V}$ столбцами и $\dim \mathcal{W}$ строками. □

Ранг и дефект оператора

Определение. Линейное отображение линейного пространства $\mathcal{V}$ в себя называют **линейным оператором** на $\mathcal{V}$.

Специфика теории линейных операторов сравнительно с теорией линейных отображений проистекает из того, что вообще на $\mathcal{L}(\mathcal{V}, \mathcal{W})$ нет возможности рассматривать композицию отображений, а появляется она лишь в случае $\mathcal{V} = \mathcal{W}$. Часто композицию операторов называют также произведением, потому что, как и множество $n \times n$ матриц, множество $\mathcal{L}(\mathcal{V}, \mathcal{V})$ со сложением операторов как отображений и с умножением путём композиции есть кольцо.

Определение. **Рангом** линейного оператора называют размерность его образа, а **дефектом** — размерность его ядра.

Уже известное свойство линейных отображений

$$\dim \mathcal{V} = \dim \operatorname{Ker} \varphi + \dim \operatorname{Im} \varphi$$

в случае операторов звучит так: для каждого оператора A на $\mathcal{V}$,

$$\text{размерность } \mathcal{V} = \text{дефект } \mathsf{A} + \text{ранг } \mathsf{A}.$$

Однако пример оператора D в следующей таблице показывает, что равенство $\mathcal{V} = \operatorname{Ker}\mathsf{A} \oplus \operatorname{Im}\mathsf{A}$ может не выполняться.

Примеры линейных операторов

Примеры. Посмотрим на небольшой список несложных операторов. Впоследствии выяснится, что в нём собрано достаточно свойств, чтобы можно было представить любой оператор в виде некоей комбинации этих операторов или очень им аналогичных. Названия их традиционны, кроме последнего:

O — **нулевой** оператор;
E — **тождественный** оператор;
P — оператор проектирования, или **проектор**;
D — оператор **дифференцирования**.
Z_λ — оператор **масштабирования** (zoom).

$\mathcal{V}$	A	*Правило*	*Ядро*	*Образ*
Любое	O	$\mathbf{v} \mapsto \mathbf{0}$	$\mathcal{V}$	$\{\mathbf{0}\}$
Любое	E	$\mathbf{v} \mapsto \mathbf{v}$	$\{\mathbf{0}\}$	$\mathcal{V}$
Любое	Z_λ	$\mathbf{v} \mapsto \lambda\mathbf{v}$, $\lambda \neq 0$	$\{\mathbf{0}\}$	$\mathcal{V}$
$\mathcal{V}_1 \oplus \mathcal{V}_2$	P_1	$\mathbf{v}_1 + \mathbf{v}_2 \mapsto \mathbf{v}_1 = \mathbf{v}_1 + \mathbf{0}$	$\mathcal{V}_2$	$\mathcal{V}_1$
	P_2	$\mathbf{v}_1 + \mathbf{v}_2 \mapsto \mathbf{v}_2 = \mathbf{0} + \mathbf{v}_2$	$\mathcal{V}_1$	$\mathcal{V}_2$
$\langle \cos x, \sin x \rangle$	D	$f \mapsto \frac{d}{dx} f$	$\{\mathbf{0}\}$	$\mathcal{V}$
$\mathbb{R}[x]_{<n}$	D	$f \mapsto \frac{d}{dx} f$	$\mathbb{R}[x]_{\leqslant 0}$	$\mathbb{R}[x]_{<n-1}$

Вслед за умножением операторов приходит возведение в степень. Например, оператор $\mathsf{D}^2 = \mathsf{D} \circ \mathsf{D}$ действует по правилу $f \mapsto \frac{d^2}{dx^2} f$. На пространстве многочленов степени менее n оператор D обладает свойством $\mathsf{D}^n = \mathsf{O}$, называемым **нильпотентностью** (степени n).

Проекторы

Каждый проектор P обладает свойством $\mathsf{P}^2 = \mathsf{P}$, называемым **идемпотентностью**. В самом деле, если $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, как в таблице, то каждый вектор $\mathbf{v} \in \mathcal{V}$ однозначно записывается в виде $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, где $\mathbf{v}_i \in \mathcal{V}_i$. Проектор на $\mathcal{V}_i$ оставляет нетронутым одно слагаемое $\mathbf{v}_i$ и обнуляет другое (все остальные, когда прямых слагаемых $\mathcal{V}_i$ более двух). Отсюда видно не только, что $\mathsf{P}_i^2 = \mathsf{P}_i$, но и что $\mathsf{P}_i\mathsf{P}_j = \mathsf{O}$ при $i \neq j$.

Отметим дополнительно, что если $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, то $\mathsf{E} = \mathsf{P}_1 + \mathsf{P}_2$, поскольку $\mathsf{P}_1 + \mathsf{P}_2$ действует по правилу

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v} \mapsto (\mathsf{P}_1 + \mathsf{P}_2)\mathbf{v} = \mathsf{P}_1\mathbf{v} + \mathsf{P}_2\mathbf{v} = (\mathbf{v}_1 + \mathbf{0}) + (\mathbf{0} + \mathbf{v}_2) = \mathbf{v}.$$

В общей ситуации $\mathcal{V} = \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_s$ аналогично получаем

$$\mathsf{E} = \mathsf{P}_1 + \ldots + \mathsf{P}_s, \quad \mathsf{P}_i\mathsf{P}_j = \delta_{ij}\mathsf{P}_i,$$

что называют **разложением E в сумму идемпотентов**.

Инвариантные подпространства

Определение. **Инвариантным** относительно (или под действием) оператора A называют такое подпространство $\mathcal{U} \subseteq \mathcal{V}$, что

$$\mathbf{u} \in \mathcal{U} \implies \mathsf{A}\mathbf{u} \in \mathcal{U}.$$

Инвариантность также выражают краткой записью $\mathsf{A}\,\mathcal{U} \subseteq \mathcal{U}$.

Пример. Поскольку $\mathsf{D}\mathbb{R}[x]_{\leqslant k} = \mathbb{R}[x]_{<k} \subset \mathbb{R}[x]_{\leqslant k}$, находится цепочка

$$\{\mathbf{0}\} \subset \mathbb{R}[x]_{\leqslant 0} \subset \ldots \subset \mathbb{R}[x]_{\leqslant n-1}$$

подпространств, инвариантных относительно дифференцирования.

Пример. Если $\mathcal{V} = \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_s$, то $\mathsf{P}_i\mathcal{V}_j \subseteq \mathcal{V}_j$, поскольку $\mathsf{P}_i\mathcal{V}_i = \mathcal{V}_i$ и $\mathsf{P}_i\mathcal{V}_j = \{\mathbf{0}\}$ при $i \neq j$. Здесь получается, что $\mathcal{V}$ есть прямая сумма подпространств, инвариантных относительно каждого из проекторов P_i.

Лемма. *Для всякого оператора A на линейном пространстве $\mathcal{V}$ равносильны следующие условия:*

(1) *$\mathcal{V} = \mathcal{V}_1 \oplus \ldots \oplus \mathcal{V}_s$ и все $\mathcal{V}_i \neq \{\mathbf{0}\}$ инвариантны относительно A;*
(2) *в каком-то базисе $\mathcal{V}$ оператор A задаётся блочно-диагональной матрицей с k блоками по диагонали.*

Доказательство. Рассмотрим случай $s = 2$; обобщение делается легко. Составим базис $\mathcal{V}$, приписав к любому базису $\mathcal{V}_1$ любой базис $\mathcal{V}_2$, и представим матрицу оператора A в соответствующем блочном виде

$$\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right].$$

По определению инвариантности,

$$\mathsf{A}\,\mathcal{V}_1 \subseteq \mathcal{V}_1 \iff A_{21} = 0 \quad \text{и} \quad \mathsf{A}\,\mathcal{V}_2 \subseteq \mathcal{V}_2 \iff A_{12} = 0. \qquad \square$$

Эта лемма показывает ключевое значение инвариантных подпространств при изучении линейного оператора A. Разложение $\mathcal{V}$, как в первом условии, сводит задачу к изучению нескольких операторов: по одному на каждом пространстве $\mathcal{V}_i$, уже меньшем, чем $\mathcal{V}$.

Диагонализуемость

Идеальный случай — когда удаётся разложить $\mathcal{V}$ в прямую сумму $\dim \mathcal{V}$ штук одномерных инвариантных подпространств. Тогда на каждом из них A обязательно действует как Z_λ или O, а матрица оператора A принимает диагональный вид.

Определение. В таком идеальном случае оператор A называют **диагонализуемым**.

К сожалению, не все операторы диагонализуемы!

Пример. Оператор R поворота в $\mathbb{R}^3$ вокруг оси Ox на угол $\pi/2$ в стандартном базисе задаётся матрицей

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

Этот оператор не диагонализуем; лишь одно одномерное подпространство в $\mathbb{R}^3$ инвариантно относительно R — ось вращения. Плоскость, ортогональная оси вращения, является двумерным инвариантным подпространством.

Пример. Оператор $\mathsf{D} = \frac{d}{dx}$ на $\mathbb{R}[x]_{\leqslant n}$ не диагонализуем для $n \in \mathbb{Z}_{>0}$. Корень зла виден уже при $n = 1$. Матрица D в базисе $\{1, x\}$ есть

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix};$$

подпространство констант $\mathbb{R}[x]_{\leqslant 0}$ есть и ядро, и образ D, а потому никакое дополнение к нему не может быть инвариантно.

7.4. Диагонализация 1

Собственные числа и векторы

Пример (Эйлер). Когда твёрдое тело вращается без воздействия внешних сил, сохраняется его момент импульса $\mathbf{L} = \mathsf{I}\boldsymbol{\omega}$. Однако для стороннего наблюдателя переменны сомножители: псевдовектор угловой скорости $\boldsymbol{\omega}$ и тензор моментов инерции I, записанный здесь как линейный оператор, причём он невырожденный. Эйлер перешёл в систему отсчёта, вращающуюся вместе с телом. В ней оператор I постоянен, а закон сохранения выражается уравнением $\mathsf{I}\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \times \mathsf{I}\boldsymbol{\omega} = \mathbf{0}$.

Лекция 2
14.02.20

Интересно найти стабильные оси. Для них $\dot{\boldsymbol{\omega}} = \mathbf{0}$ и остающееся уравнение $\boldsymbol{\omega} \times \mathsf{I}\boldsymbol{\omega} = \mathbf{0}$ означает, что псевдовектор $\mathsf{I}\boldsymbol{\omega}$ коллинеарен $\boldsymbol{\omega}$.

Именно из этой задачи в математику пришло понятие главных осей. Хотя де Витт открыл переход к канонической системе координат для линий второго порядка гораздо раньше, обобщение на поверхности и квадратичные формы удалось, лишь когда идеи Эйлера о вращении подхватил Лагранж и потом Коши.

Продолжим теперь вводимую серию ключевых понятий. Грубо говоря, ненулевой вектор, сохраняющий направление под действием оператора A, называют его собственным вектором. Эту формулировку необходимо уточнить: образ вектора $\mathbf{v} \neq \mathbf{0}$ ему коллинеарен, то есть линейная оболочка $\langle \mathbf{v} \rangle$ является инвариантным подпространством. Однако, сразу же оказывается полезным следить и за изменением длины. Формально у нас ещё нет понятия длины, но при наличии коллинеарности достаточно следить за коэффициентом пропорциональности.

Определение. **Собственным вектором** оператора A называют такой вектор $\mathbf{v} \neq \mathbf{0}$, что $\mathsf{A}\mathbf{v} = \lambda\mathbf{v}$ для какого-то скаляра λ; последний называют **собственным числом** оператора A.

Характеристический многочлен

Следующее понятие является важным шагом на пути теоретического поиска инвариантных подпространств. На практике неизбежные ошибки вычислений обычно заставляют искать иные пути, но эти вопросы выходят за рамки нашего курса.

Определение. **Характеристическим многочленом** матрицы A называют многочлен $\chi_A(t) \leftrightharpoons \det(tE - A)$.

Лемма. *Равносильны два условия на пару матриц A и A':*

(1) *они задают один и тот же оператор (но в разных базисах);*
(2) *найдётся такая матрица S, что $A' = S^{-1}A\,S$.*

Доказательство. Это частный случай теоремы об изменении матрицы линейного отображения при смене базисов. □

Определение. Такую пару матриц называют **подобными** матрицами.

Лемма. *Характеристические многочлены подобных матриц равны.*

Доказательство. Если $A' = S^{-1}AS$, то $tE - A' = S^{-1}(tE - A)S$. Поэтому,

$$\det(tE - A') = \det S^{-1} \cdot \det(tE - A) \cdot \det S = \det(tE - A). \qquad \square$$

Определение. **Характеристическим многочленом** оператора A называют многочлен $\chi_{\mathsf{A}}(t) \leftrightharpoons \chi_A(t)$, где матрица A задаёт A в каком-то базисе.

Определение. Сумму диагональных элементов квадратной матрицы A называют её **следом** и обозначают через $\operatorname{tr} A$.

Формулы в следующих упражнениях, особенно первых двух, упрощают вычисление характеристического многочлена.

Упражнение. *Для матриц порядка 2 проверьте формулу*

$$\chi_A(t) = t^2 - (\operatorname{tr} A)t + \det A.$$

Упражнение. *Для матриц порядка 3 проверьте формулу*

$$\chi_A(t) = t^3 - (\operatorname{tr} A)t^2 + I_2 t - \det A,$$

где I_2 равно сумме $M_{11} + M_{22} + M_{33}$ миноров матрицы A.

Упражнение. *Напишите общую формулу для коэффициентов $\chi_A(t)$ через миноры матрицы.*

Характеристические корни

Пример. Все собственные векторы оператора R коллинеарны оси Ox. Его характеристический многочлен $\chi_{\mathsf{R}}(t) = (t-1)(t^2+1)$ имеет только один вещественный корень.

Собственные числа (или значения) также называют характеристическими числами или характеристическими корнями, потому что они оказываются корнями характеристического многочлена.

Лемма. *Для каждого скаляра λ равносильны следующие условия:*

(1) *λ есть собственное число оператора A;*
(2) *$\chi_{\mathsf{A}}(\lambda) = 0$;*
(3) *$\operatorname{Ker}(\mathsf{A} - \lambda\mathsf{E}) \neq \{\mathbf{0}\}$;*
(4) *$\operatorname{Ker}(\mathsf{A} - \lambda\mathsf{E})^n \neq \{\mathbf{0}\}$, где $n = \dim \mathcal{V}$.*

Доказательство. (1 $\Leftrightarrow$ 3) Тут $\mathsf{A}\mathbf{v} = \lambda\mathbf{v}$ равносильно $(\mathsf{A} - \lambda\mathsf{E})\mathbf{v} = \mathbf{0}$.

(2 $\Leftrightarrow$ 3) Положим $\mathsf{B} = \mathsf{A} - \lambda\mathsf{E}$, и соответственно $B = A - \lambda E$ для матриц в некотором базисе. По критерию невырожденности матрицы и формуле, связывающей ранг и дефект оператора,

$$\det B = 0 \iff n > \operatorname{rk} B = \dim \operatorname{Im} \mathsf{B} \iff \dim \operatorname{Ker} \mathsf{B} > 0.$$

(2 $\Leftrightarrow$ 4) Аналогично предыдущему, но для оператора $(\mathsf{A} - \lambda\mathsf{E})^n$. При этом пользуемся равенством $\det(B^n) = (\det B)^n$. $\square$

Экстремумы квадратичной формы на сфере

Здесь замечательно переплелись геометрия, алгебра и анализ.

Пример (Лагранж). Для простоты, разберём плоский случай: условные экстремумы квадратичной формы

$$f(x,y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2$$

на единичной окружности $F(x,y) = x^2 + y^2 - 1 = 0$.

Найдём подозрительные точки по методу множителей Лагранжа, излагаемому в курсе основ математического анализа. Дифференцируем функцию Лагранжа

$$L(x,y,\lambda) = f(x,y) - \lambda F(x,y)$$

и видим, что условие экстремума $dL = 0$ в данном случае выражается системой линейных уравнений, которую мы запишем в матричном виде с симметричной матрицей:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix},$$

или $(A - \lambda E)\mathbf{x} = \mathbf{0}$. Ненулевые решения есть лишь при $\det(A - \lambda E) = 0$.

Следовательно, тут множители Лагранжа оказываются собственными числами матрицы коэффициентов формы, а радиус-векторы точек экстремума — собственными векторами. Можно убедиться, что максимум отвечает (наи)большему собственному числу, а минимум (наи)меньшему. Эти выводы сохранятся и в многомерной задаче, причём позже мы их применим; собственные же векторы, соответствующие промежуточным собственным числам, окажутся обобщёнными седловыми точками, а не экстремумами.

окрасить сферу?

Примеры вычислений

15.02.20

Пример. Попробуем диагонализовать матрицу

$$A = \begin{bmatrix} -4 & -3 \\ 6 & 5 \end{bmatrix}.$$

Её характеристический полином равен $t^2 - t - 2 = (t+1)(t-2)$ и корни видны. Затем нужно решить две системы линейных уравнений с вырожденными матрицами $A + E$ и $A - 2E$:

$$\begin{bmatrix} -3 & -3 \\ 6 & 6 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \qquad \begin{bmatrix} -6 & -3 \\ 6 & 3 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Первый вектор можно взять равным $[1,-1]^\top$, а второй равным $[1,-2]^\top$. Они будут столбцами матрицы перехода

$$T = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}.$$

Получаем разложение $A = TDT^{-1}$ с диагональной матрицей

$$D = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} = \operatorname{diag}(-1,2).$$

Порядки столбцов в T и чисел в D должны быть согласованы.

Для проверки проще всего смотреть на равенство $AT = TD$.

Пример. Попробуем диагонализовать матрицу

$$A = \begin{bmatrix} 3 & 2 & -2 \\ -2 & -1 & 4 \\ -1 & -1 & 4 \end{bmatrix}.$$

Её характеристический многочлен найдём через миноры, пользуясь упражнением выше; он равен $t^3-6t^2+11t-6$. Корни можно подобрать: скажем, подобрав $\lambda_1 = 1$, мы поделим на $t-1$ и сведём уравнение к квадратному. Находим $\lambda_2 = 2$ и $\lambda_3 = 3$. Затем нужно решить три однородные системы линейных уравнений с матрицами $A - \lambda_i E$:

$$\begin{bmatrix} 2 & 2 & -2 \\ -2 & -2 & 4 \\ -1 & -1 & 3 \end{bmatrix}; \qquad \begin{bmatrix} 1 & 2 & -2 \\ -2 & -3 & 4 \\ -1 & -1 & 2 \end{bmatrix}; \qquad \begin{bmatrix} 0 & 2 & -2 \\ -2 & -4 & 4 \\ -1 & -1 & 1 \end{bmatrix}.$$

Легко убедиться, что тут $\operatorname{rk}(A-\lambda_i E) = 2$ для каждого из собственных чисел. Поэтому для каждого λ_i должен найтись собственный вектор, причём ровно один с точностью до коллинеарности. Иногда решения сразу видны, и такие случаи полезно привыкнуть замечать. Например, для первой матрицы подходит $[1,-1,0]^\top$, для второй $[2,0,1]^\top$, а для третьей $[0,1,1]^\top$. Получаем разложение $A = TDT^{-1}$ с матрицами

$$T = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \operatorname{diag}(1,2,3).$$

Пример. Попробуем диагонализовать матрицу

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & -1 \\ 4 & -2 & -1 \end{bmatrix}.$$

Здесь корни $\lambda_1 = 1$ кратности два и $\lambda_2 = -2$. Поскольку $\operatorname{rk}(A-E) = 1$, для λ_1 нужно взять два независимых собственных вектора. Одним из

решений задачи диагонализации здесь является пара матриц

$$T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 2 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \operatorname{diag}(1, 1, -2).$$

7.5. Диагонализация 2

Собственные и корневые подпространства

Определение. Если оператор A имеет λ собственным числом, то

$$\mathcal{V}^\lambda \leftharpoons \operatorname{Ker}(\mathsf{A} - \lambda\mathsf{E}) \quad \text{и} \quad \mathcal{V}(\lambda) \leftharpoons \operatorname{Ker}(\mathsf{A} - \lambda\mathsf{E})^n$$

соответственно называют **собственным** и **корневым подпространствами**, принадлежащими λ. Оба они инвариантны под действием оператора A, ибо это частный случай утверждений следующей леммы.

Лемма. *Если $\mathsf{A}_1\mathsf{A}_2 = \mathsf{A}_2\mathsf{A}_1$, то под действием A_1 инвариантны:*

(1) *ядро и образ A_2;*
(2) *каждое собственное подпространство для A_2.*

Упражнение. *Докажите лемму и примените её к $\mathcal{V}^\lambda$ и $\mathcal{V}(\lambda)$.*

Пример. Когда есть разложение $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, есть и проекторы P_k на слагаемые. Оба $\mathcal{V}_k$ являются собственными подпространствами для проектора P_1, принадлежащими собственным числам 1 и 0 соответственно. Также и для P_2, но собственные числа — наоборот. Позже мы увидим, что других собственных чисел у проекторов не бывает.

Пример. Оператор $\mathsf{D} = \frac{d}{dx}$ на пространстве многочленов $\mathbb{R}[x]_{<n}$ нильпотентен, поэтому всё пространство является корневым подпространством, принадлежащим корню $\lambda = 0$. При этом собственное подпространство одномерно (константы).

Сумма собственных подпространств

Теорема. *Для всякого оператора A:*

(1) *собственные векторы, принадлежащие различным собственным числам, линейно независимы;*
(2) *сумма всех собственных подпространств — прямая.*

Эти два утверждения подразумевают одно и то же; хотя первая формулировка традиционна, вторая — точнее и лаконичнее.

14.02.20

Доказательство. (1) Применим индукцию по количеству векторов. Для одного вектора доказывать нечего: $\mathbf{v} \neq \mathbf{0}$. Предположим, что искомая линейная независимость установлена для наборов из $k-1$ собственных векторов, и возьмём собственные векторы $\mathbf{v}_1, \ldots, \mathbf{v}_k$, принадлежащие различным собственным числам $\lambda_1, \ldots, \lambda_k$.

Составим линейную комбинацию $\mathbf{v} = \alpha_1\mathbf{v}_1 + \ldots + \alpha_k\mathbf{v}_k$ и изучим вектор $\mathbf{u} = \mathsf{A}\mathbf{v} - \lambda_k\mathbf{v}$. Видим, что

$$\mathbf{u} = \alpha_1(\lambda_1 - \lambda_k)\mathbf{v}_1 + \ldots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{v}_{k-1} + \alpha_k(\lambda_k - \lambda_k)\mathbf{v}_k.$$

Значит, для $\mathbf{v} = \mathbf{0}$ получаем запись вектора $\mathbf{u} = \mathbf{0}$ в виде линейной комбинации из $\langle \mathbf{v}_1, \ldots, \mathbf{v}_{k-1} \rangle$. По предположению индукции, все её коэффициенты $\alpha_i(\lambda_i - \lambda_k)$ должны быть нулевыми. Тогда и все α_i равны нулю, кроме исчезнувшего ранее α_k, но затем и последнее вынуждено быть нулём, поскольку $\mathbf{v} = \mathbf{0}$.

(2) Тут практически такое же рассуждение, как для (1), и даже в чём-то проще, если достигнут достаточный уровень восприятия подпространств. Адаптация сводится к аналогии между линейной независимостью векторов и прямотой суммы подпространств, отмеченной, как только прямые суммы появились в курсе. Среди равносильных определяющих свойств прямой суммы там же указана однозначность записи нулевого вектора; именно это свойство пригождается здесь.

Перенумеруем все собственные числа: $\lambda_1, \ldots, \lambda_s$. Для $1 \leqslant k \leqslant s$ положим $\Sigma_k = \mathcal{V}^{\lambda_1} + \ldots + \mathcal{V}^{\lambda_k}$ и покажем, что

$$\text{сумма } \Sigma_{k-1} \text{ прямая } \implies \text{сумма } \Sigma_k \text{ прямая.}$$

Взяв любой вектор $\mathbf{v} = \mathbf{v}_1 + \ldots + \mathbf{v}_k$, где $\mathbf{v}_i \in \mathcal{V}^{\lambda_i}$, видим, что

$$\mathsf{A}\mathbf{v} - \lambda_k\mathbf{v} = (\lambda_1 - \lambda_k)\mathbf{v}_1 + \ldots + (\lambda_{k-1} - \lambda_k)\mathbf{v}_{k-1} + (\lambda_k - \lambda_k)\mathbf{v}_k$$

лежит в Σ_{k-1}. Если сумма Σ_{k-1} прямая, то выводим

$$\begin{aligned}
\mathbf{v} = \mathbf{0} &\implies \mathsf{A}\mathbf{v} - \lambda_k\mathbf{v} = \mathbf{0} \\
&\implies (\lambda_i - \lambda_k)\mathbf{v}_i = \mathbf{0} \quad \text{для } 1 \leqslant i \leqslant k-1 \\
&\implies \mathbf{v}_i = \mathbf{0} \quad \text{для } 1 \leqslant i \leqslant k-1 \\
&\implies \mathbf{v}_k = \mathbf{0},
\end{aligned}$$

так что сумма Σ_k тоже прямая. □

Кратности корня

Напомним, что мы работаем с линейными операторами на линейном пространстве $\mathcal{V}$ над основным полем скаляров $\mathbb{F}$. Это означает, что коэффициенты всех линейных комбинаций берутся из $\mathbb{F}$. Чаще

пространства вещественные и $\mathbb{F} = \mathbb{R}$, но бывают и комплексные, и тогда $\mathbb{F} = \mathbb{C}$. Другие поля нас здесь не интересуют, поэтому нет нужды углубляться в само понятие поля. Однако в вещественном случае приходится дополнительно страдать из-за того, что собственные числа могут оказаться комплексными, то есть не лежать в $\mathbb{F}$.

Определение. С каждым характеристическим корнем λ данного оператора A связаны три натуральных числа:

- $m_{\text{г}}(\lambda) = \dim \mathcal{V}^\lambda$ есть **геометрическая кратность** λ;
- $m_{\text{а}}(\lambda) = \dim \mathcal{V}(\lambda)$ есть **алгебраическая кратность** λ;
- $n(\lambda)$ есть кратность λ, как корня многочлена $\chi_{\mathsf{A}}(t)$.

Если корень λ характеристического многочлена $\chi_{\mathsf{A}}(t)$ лежит в поле $\mathbb{F}$, то $m_{\text{а}}(\lambda) = n(\lambda)$. Это свойство объясняет термин «алгебраическая кратность». Оно будет частью теоремы о корневом разложении, а пока останется недоказанным, а также неиспользованным.

Пример. Для оператора $\mathsf{D} = \frac{d}{dx}$ на $\mathbb{R}[x]_{<n}$ получаем значения

$$m_{\text{г}}(0) = \dim \operatorname{Ker} \mathsf{D} = 1 \quad \text{и} \quad m_{\text{а}}(0) = \dim \operatorname{Ker} \mathsf{D}^n = n.$$

Для всякого оператора B и каждого натурального числа k имеется включение $\operatorname{Ker} \mathsf{B}^k \subseteq \operatorname{Ker} \mathsf{B}^{k+1}$; поэтому $\mathcal{V}^\lambda \subseteq \mathcal{V}(\lambda)$ и $m_{\text{г}}(\lambda) \leqslant m_{\text{а}}(\lambda)$.

Лемма. *Для всякого многочлена в $\mathbb{F}[t]$ равносильны утверждения:*

(1) *сумма кратностей его корней в поле $\mathbb{F}$ равна его степени;*
(2) *он разлагается в $\mathbb{F}[t]$ в произведение линейных множителей.*

Доказательство. Повторяем теорему Безу. □

Определение. Такой многочлен называют **вполне разложимым над** $\mathbb{F}$.

Критерий диагонализуемости

Определение. **Спектром** оператора A назовём множество $\operatorname{Spec} \mathsf{A}$ всех его характеристических корней вместе с геометрическими кратностями.

Теорема (диагонализация)**.** *Для всякого оператора A на линейном пространстве $\mathcal{V}$ над полем $\mathbb{F}$ равносильны утверждения:*

(1) *оператор A диагонализуем;*
(2) *характеристический многочлен $\chi_{\mathsf{A}}(t)$ вполне разложим над $\mathbb{F}$ и $m_{\text{г}}(\lambda) = m_{\text{а}}(\lambda) = n(\lambda)$ для всех $\lambda \in \operatorname{Spec} \mathsf{A}$;*
(3) *прямая сумма всех собственных подпространств оператора A равна всему пространству $\mathcal{V}$:*

$$\mathcal{V} = \bigoplus_{\lambda \in \operatorname{Spec} \mathsf{A}} \mathcal{V}^\lambda.$$

Определение. **Сужением** или **ограничением** оператора A на инвариантное подпространство $\mathcal{U} \subseteq \mathcal{V}$ называют оператор $\mathsf{A}|_{\mathcal{U}} : \mathcal{U} \to \mathcal{U}$, $\mathbf{u} \mapsto \mathsf{A}\mathbf{u}$.

Следствие. *Если* $\operatorname{Spec}\mathsf{A}$ *состоит из* $n = \dim \mathcal{V}$ *различных точек в поле* $\mathbb{F}$*, то* A *диагонализуем.*

Доказательство. (1 ⇒ 2) Диагональный вид матрицы оператора сразу даёт разложение $\chi_{\mathsf{A}}(t)$ в произведение линейных множителей. Все три кратности корня λ равны числу появлений λ на диагонали.

(2 ⇒ 3) Размерность прямой суммы всех подпространств $\mathcal{V}^\lambda$ есть сумма кратностей всех корней $\chi_{\mathsf{A}}(t)$ и равна $\deg \chi_{\mathsf{A}}(t) = \dim \mathcal{V}$.

(3 ⇒ 1) Разложение пространства $\mathcal{V}$ в прямую сумму инвариантных подпространств $\mathcal{V}^\lambda$ позволяет представить оператор A блочно-диагональной матрицей. Каждый блок задаёт сужение A на одно из $\mathcal{V}^\lambda$, а оно совпадает с масштабированием Z_λ, матрица которого в любом базисе $\mathcal{V}^\lambda$ есть λE. Блочно-диагональная матрица на самом деле оказывается диагональной. □

Формулировка корневого разложения

Второе условие в предыдущей теореме показывает, что только две причины могут препятствовать диагонализуемости оператора A:

(1) не все многочлены обязаны быть вполне разложимы над исходным полем $\mathbb{F}$;
(2) не все геометрические кратности обязаны равняться соответствующим алгебраическим.

Простые примеры необязательности мы уже видели: оператор поворота для (1) и оператор дифференцирования для (2).

На практике наиболее важны линейные операторы на вещественных и комплексных пространствах. Первое препятствие полностью исчезает только для комплексных, благодаря основной теореме алгебры многочленов. Обойти второе препятствие нельзя; поэтому приходится строить более сложную теорию, чем приведение к диагональному виду, основанное на собственных подпространствах.

Теорема (корневое разложение)**.** *Для всякого оператора* A *на линейном пространстве* $\mathcal{V}$ *над полем* $\mathbb{F}$ *равносильны утверждения:*

(1) *характеристический многочлен* $\chi_{\mathsf{A}}(t)$ *вполне разложим над* $\mathbb{F}$*;*
(2) *прямая сумма всех корневых подпространств равна всему* $\mathcal{V}$*.*

7.6. Функции матриц

Мотивация

Функции матриц находятся на стыке предметов, составляющих основу общематематического образования при подготовке по различным научно-техническим направлениям. Прежде всего это алгебра как источник самих объектов и дифференциальные уравнения как сфера применения матричной экспоненты, но не следует забывать и о разделах математического анализа. Студенты сталкиваются с этой темой на первом курсе; то, что они в ней зачастую получают и из неё выносят, оставляет желать много лучшего. В частности, у многих остаётся ложное представление, что для вычисления функции матрицы необходима жорданова форма, причём воспоминания о последней отрывочны и неприятны.

Ввиду неоднозначности толкования, уточним, что функцией матрицы мы здесь называем выражение $f(A)$, полученное «подстановкой» матрицы A в аргумент функции $x \mapsto f(x)$ изучаемого в одномерном анализе типа, например, экспоненты, причём случай 1×1 матриц должен совпадать с привычным числовым. Есть несколько внешне сильно отличающихся подходов к определению функций матриц, каждый со своими достоинствами и недостатками. При соблюдении естественных условий все подходы приводят к одинаковым значениям $f(A)$.

Важнейшей из функций матрицы является матричная экспонента. Глянем кратко на быстрый выход на неё и связь с материалом предыдущего раздела. Отталкиваться будем от знания решений дифференциального уравнения $y' = ay$ с постоянным коэффициентом a: общее решение даёт формула $y(t) = y(0)e^{at}$.

14.02.20

Перейдём к аналогичной задаче для двух неизвестных функций. Вместо одного дифференциального уравнения теперь система

$$\begin{cases} y_1' = a_{11}y_1 + a_{12}y_2, \\ y_2' = a_{21}y_1 + a_{22}y_2, \end{cases}$$

которую можно переписать в матричном виде $\mathbf{y}' = A\mathbf{y}$. С одной стороны, возникает идея искать решение в виде экспоненты: $\mathbf{y}(t) = \mathbf{v}e^{\lambda t}$. Подставив это в систему, видим $\mathbf{v}\lambda e^{\lambda t} = A\mathbf{v}e^{\lambda t}$. Сократив экспоненты, приходим к уравнению $A\mathbf{v} = \lambda\mathbf{v}$, то есть к задаче на собственные числа и векторы. С другой стороны, соблазнительно заполучить ответ в виде $\mathbf{y}(t) = e^{At}\mathbf{y}(0)$. Нужно лишь придать смысл выражению e^{At}! Рецепт, конечно, будет включать в себя собственные числа и векторы.

Полиномы и суммы рядов

Опираясь на исходно имеющиеся операции сложения и умножения, мы однозначно определим $p(A)$ для любого полинома $p(x)$ непосредственной подстановкой квадратной матрицы A вместо x. Естественно, нас сразу же интересуют полиномы, дающие $p(A)=0$. В таком случае говорят, что A **аннулирует** p.

Пример. Матрица $\operatorname{diag}(1,2,3)$ аннулирует полином $(x-1)(x-2)(x-3)$.

Аналогично поступают с операторами.

Пример. Всякий проектор P аннулирует полином x^2-x.

Пример. Поворот плоскости на угол $\pi/2$ аннулирует полином x^2+1.

Теорема (Cayley; Hamilton; Frobenius, 1878)**.** *Каждый оператор аннулирует свой характеристический многочлен:* $\chi_{\mathsf{A}}(\mathsf{A})=\mathsf{O}$.

Весь дальнейший материал о функциях матриц опирается на эту теорему. Сперва мы обсудим, почему это так, и лишь затем перейдём к её доказательству. По аналогии с полиномами сделаем подстановку и в функцию, заданную как сумма ряда:

$$f(x)=\sum_{k\geqslant 0} c_k x^k \quad\Rightarrow\quad f(A)=\sum_{k\geqslant 0} c_k A^k. \tag{1}$$

Возникает вопрос сходимости, но ответ на него несложен: матричный ряд гарантированно сходится, если норма матрицы A меньше радиуса сходимости первоначального ряда для $f(x)$. При этом отвлекаться на понятие нормы матрицы и обоснования сейчас не стоит.

Применить это определение для конкретного вычисления $f(A)$ удаётся лишь в простейших случаях, когда структура матрицы позволяет свернуть ряды покомпонентно.

Упражнение. *Найдите* $\exp(At)$ *для матриц*

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Теорема Гамильтона — Кэли утверждает, что каждая матрица является «корнем» своего характеристического уравнения. Тем самым, для матриц порядка n степень A^n представляется линейной комбинацией низших степеней, иными словами, лежит в их линейной оболочке. Тогда каждая высшая степень тоже лежит в этой оболочке. Поэтому представление $f(A)$ суммой ряда можно заменить на представление полиномом от A степени не выше $n-1$, коэффициенты которого зависят как от f, так и от A, а точнее, от характеристического многочлена.

Доказательство теоремы Гамильтона — Кэли

Доказательство. Вспомним транспонированную матрицу алгебраических дополнений $A^\vee$. Её основное свойство, $A^\vee A = E \det A$, верно не только в случае матриц над полем, но и тогда, когда элементы матриц живут в любом *коммутативном* кольце, в том числе кольце многочленов $\mathbb{F}[t]$; тогда мы просто лишаемся возможности делить, а также сокращать множители. Поэтому, представив наш оператор $n \times n$ матрицей A, рассмотрим тождество

$$(tE - A)^\vee (tE - A) = E \det(tE - A) = (a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n)E$$

в матрицах многочленов от t. Или — что практически то же самое, но будет удобнее — в многочленах от t с матричными коэффициентами.

Умножим справа это тождество на матричный многочлен

$$R(t^{-1}) = E + t^{-1}A + t^{-2}A^2 + \ldots + t^{-n}A^n$$

и сравним коэффициенты при t^0 в обеих частях полученного произведения. В правой части этот коэффициент равен матрице

$$a_0 E + a_1 A + a_2 A^2 + \ldots + a_n A^n = \chi_A(A).$$

В левой же части он равен нулевой матрице, ибо

$$(tE - A)\, R(t^{-1}) = tE - t^{-n} A^{n+1},$$

но в $(tE - A)^\vee$ не входят ни t^{-1}, ни t^n, и шансов найти t^0 нет. □

Фробениус доказал эту теорему, не зная, что ранее Кэли и Гамильтон проверили утверждение для матриц размеров 2×2, 3×3 и 4×4. Узнав о том, он назвал её теоремой Гамильтона — Кэли; так её называют и поныне.

Диагонализация и жорданова форма

Лекция 3
21.02.20

Естественно отметить лёгкость вычисления функций от диагональной матрицы,

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \quad \Rightarrow \quad f(D) = \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n)),$$

видимую непосредственно из определения рядом (1). Наблюдение

$$f(TDT^{-1}) = Tf(D)T^{-1}$$

связи значений функции на подобных матрицах позволяет быстро распространить эту лёгкость на диагонализуемые матрицы. Когда в курсе только что показано, как разложить данную матрицу A в произведение $A = TDT^{-1}$, это представляется замечательным методом.

Проблема в том, что работает он не для всех матриц, а модификация, делающая его универсальным (на первый взгляд), не столь безобидна и гораздо труднее мотивируема, особенно если, придерживаясь печального традиционного разделения на предметы, не упоминать тут же системы линейных дифференциальных уравнений.

Для матрицы, не подобной никакой диагональной, ищут блочно-диагональную матрицу J наиболее простого вида, называемую жордановой формой. Каждый диагональный блок порядка выше первого — двухдиагональная ленточная матрица с собственным числом на главной диагонали и единицами на соседней, называемая жордановой клеткой. Заполненные единицами диагонали обычно располагают по одну сторону, чтобы J была треугольной матрицей. Собственные числа в разных клетках могут совпадать, поэтому точный вид жордановой формы матрицы не задаётся одним лишь её спектром. Например, 3×3 матрица с единственным собственным числом λ подобна ровно одной из трёх жордановых форм

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \qquad \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \qquad \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}.$$

Для полного рецепта вычисления по формуле

$$f(TJT^{-1}) = Tf(J)T^{-1}$$

не хватает способа отыскания матрицы перехода T, указывающего заодно точный вид жордановой формы. Во многих традиционных курсах линейной алгебры обучение ему поглощает значительное учебное время, сравнительно мало прибавляя для понимания.

Главным потребителем этих достижений является курс дифференциальных уравнений, где нужна матричная экспонента. Эта функция хорошая: она зависит от матрицы непрерывно и даже бесконечно гладко. Однако рецепт вычисления её, да и любой другой тоже, через жорданову форму обесценивается тем фактом, что жорданова форма и, ещё более, матрица перехода к ней терпят разрывы.

Матрицы второго порядка

Основная идея другого метода — идея Сильвестра — в том, что значения функции матрицы интерполируют её значения на спектре этой матрицы. Именно эту идею и следует запомнить. Вместе с ней подход к интерполяции через определители приносит элегантный способ вычисления функций матрицы, включая матричную экспоненту. 22.02.20

Начнём с определителя Вандермонда, дающего интерполяцию по двум различным точкам. Подставим другие буквы и раскроем:

$$\begin{vmatrix} 1 & \lambda & \lambda^2 \\ 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \end{vmatrix} = 0 \quad \Leftrightarrow \quad (\lambda_1 - \lambda_2)(\lambda - \lambda_1)(\lambda - \lambda_2) = 0.$$

Поскольку $\lambda_1 \neq \lambda_2$ предполагается, сократим первую скобку. Две оставшихся равны характеристическому полиному $\det(A - \lambda E)$ всякой 2×2 матрицы A со спектром $\{\lambda_1, \lambda_2\}$. Значит, теорема Гамильтона — Кэли влечёт равенство

$$\begin{vmatrix} E & A & A^2 \\ 1 & \lambda_1 & \lambda_1^2 \\ 1 & \lambda_2 & \lambda_2^2 \end{vmatrix} = 0.$$

Это символический определитель, который нужно понимать как утверждение о линейной зависимости столбцов. Обозначим их через $C^{(0)}$, $C^{(1)}$ и $C^{(2)}$ соответственно степеням. При этом каждая матрица в верхней строке состоит из четырёх элементов, так что «расшифрованные» столбцы содержат уже по шесть элементов. Крайний справа столбец лежит в линейной оболочке двух первых: $C^{(2)} \in \langle C^{(0)}, C^{(1)} \rangle$.

Домножим строки определителя на элементы второго столбца:

$$\begin{vmatrix} A & A^2 & A^3 \\ \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ \lambda_2 & \lambda_2^2 & \lambda_2^3 \end{vmatrix} = 0.$$

Опять крайний справа столбец лежит в линейной оболочке двух первых: $C^{(3)} \in \langle C^{(1)}, C^{(2)} \rangle$. Следовательно, $C^{(3)} \in \langle C^{(0)}, C^{(1)} \rangle$. И так далее без ограничений, то есть $C^{(k)} \in \langle C^{(0)}, C^{(1)} \rangle$ для всех $k \in \mathbb{Z}_{\geqslant 0}$.

Затем можно ввести новую переменную t, домножить $C^{(k)}$ на t^k и просуммировать ряд. Получим формулу для вычисления функции от 2×2 матрицы с различными собственными числами:

$$\begin{vmatrix} E & A & f(At) \\ 1 & \lambda_1 & f(\lambda_1 t) \\ 1 & \lambda_2 & f(\lambda_2 t) \end{vmatrix} = 0.$$

В раскрытом виде будет

$$f(At) = f(\lambda_1 t) \cdot \frac{A - \lambda_2 E}{\lambda_1 - \lambda_2} + f(\lambda_2 t) \cdot \frac{A - \lambda_1 E}{\lambda_2 - \lambda_1}.$$

Наконец, для случая $\lambda_1 = \lambda_2$ предельный переход при $\lambda_2 \to \lambda_1$ даёт

$$f(At) = f(\lambda_1 t) \cdot E + t f'(\lambda_1 t) \cdot (A - \lambda_1 E).$$

Формула Сильвестра

Переход к произвольной размерности не потребует новых идей.

Возьмём квадратную матрицу A с различными собственными числами λ_k кратности 1. В этом случае говорят, что матрица имеет **простой спектр**. Формула Сильвестра выражает значение $f(A)$ хорошей функции на такой матрице как линейную комбинацию

$$f(A) = \sum f(\lambda_k) A_k, \text{ где } A_k = \prod_{i \neq k} \frac{A - \lambda_i E}{\lambda_k - \lambda_i}.$$

Функция хорошая, если она определена на спектре A. Матрицы A_k называют ковариантами Фробениуса. В них можно распознать слагаемые интерполяционной формулы Лагранжа с подстановкой λ_k в x_k и матрицы A в переменную x. Поскольку мы ранее переписали формулу Лагранжа через определитель, поступим так и с матричной версией, ограничившись трёхмерным случаем:

$$\begin{vmatrix} E & A & A^2 & f(A) \\ 1 & \lambda_0 & \lambda_0^2 & f(\lambda_0) \\ 1 & \lambda_1 & \lambda_1^2 & f(\lambda_1) \\ 1 & \lambda_2 & \lambda_2^2 & f(\lambda_2) \end{vmatrix} = 0.$$

В частности, для матричной экспоненты получаем простую формулу

$$\begin{vmatrix} E & A & A^2 & \exp(At) \\ 1 & \lambda_0 & \lambda_0^2 & \exp(\lambda_0 t) \\ 1 & \lambda_1 & \lambda_1^2 & \exp(\lambda_1 t) \\ 1 & \lambda_2 & \lambda_2^2 & \exp(\lambda_2 t) \end{vmatrix} = 0, \tag{2}$$

пригодную для непосредственных вычислений.

Упражнение. *Раскрыв определитель, выразите явно* $\exp(At)$ *для матриц с заданными простыми спектрами:*

$$\{1, -1\};\ \{i, -i\};\ \{\alpha + i\beta, \alpha - i\beta\};\ \{1, 0, -1\};\ \{1, i, -i\};\ \{1, -1, i, -i\}.$$

Упражнение. *В третьем случае предыдущего упражнения перейдите к пределу при* $\beta \to 0$.

Формула Бухгейма

Если спектр не простой, то есть матрица имеет кратные собственные числа, то для вычисления хороших функций от неё вместо интерполяции Лагранжа нужно использовать интерполяцию Эрмита. Ожида-

емо, от функции теперь требуется существование всех используемых производных. Скажем, для матриц со спектром $\lambda_1, \lambda_2, \lambda_2$ адаптируем определитель из первого примера на интерполяцию Эрмита и видим

$$\begin{vmatrix} E & A & A^2 & \exp(At) \\ 1 & \lambda_1 & \lambda_1^2 & \exp(\lambda_1 t) \\ 1 & \lambda_2 & \lambda_2^2 & \exp(\lambda_2 t) \\ 0 & 1 & 2\lambda_2 & t\exp(\lambda_2 t) \end{vmatrix} = 0. \tag{3}$$

Здесь нужно обратить внимание на появление множителя t перед экспонентой в последней строке. Источник его в том, что при переходе от (2) к конфлюэнтному пределу (3) берётся производная от $\exp(\lambda t)$ по λ, ведь в наших новых переменных именно λ переняла роль буквы x.

Вычищение определителя Вандермонда

Как мы увидели выше, тактика воздержания от раскрытия определителя позволяет выписать простое уравнение для функции матрицы с известным спектром. Больше того, она оказывается выгодна и при практических вычислениях, особенно с вырожденным спектром. В случае простого спектра формула Сильвестра в раскрытом виде весьма эффективна, хотя комплексно-сопряжённые пары корней несколько портят картину.

Огромное преимущество свёрнутой общей формулы заключается, естественно, в возможности упрощения числового определителя — алгебраического дополнения искомой $f(A)$. Поскольку он не вырожден, его всегда можно свести к единичной матрице путём элементарных преобразований строк и/или столбцов окаймлённой матрицы, после чего ответ считывается непосредственно из верхней строки (полиномов от A) и последнего столбца (функций на спектре).

Пример. Найдём $\exp(At)$ для матрицы A со спектром $\{i, -i, i, -i\}$. Уже в начале преобразований полезно положить $c = \cos t$ и $s = \sin t$. Запишем исходный определитель интерполяции Эрмита:

$$\begin{vmatrix} E & A & A^2 & A^3 & X \\ 1 & i & -1 & -i & e^{it} \\ 0 & 1 & 2i & -3 & te^{it} \\ 1 & -i & -1 & i & e^{-it} \\ 0 & 1 & -2i & -3 & te^{-it} \end{vmatrix} = 0.$$

Опустим, то есть оставим в качестве упражнения, элементарные преобразования окаймлённой матрицы и приведём только финальный результат её вычищения:

$$\begin{vmatrix} E & A & A^2+E & A^3+A & X \\ 1 & 0 & 0 & 0 & c \\ 0 & 1 & 0 & 0 & s \\ 0 & 0 & 1 & 0 & ts/2 \\ 0 & 0 & 0 & 1 & (s-tc)/2 \end{vmatrix} = 0.$$

Отсюда считываем ответ

$$X = e^{At} = E\cos t + A\sin t + \tfrac{1}{2}(A^2+E)t\sin t + \tfrac{1}{2}(A^3+A)(\sin t - t\cos t).$$

Если при вычищении ограничиться преобразованиями строк, то ответ запишется в виде линейной комбинации степеней A^k. Наоборот, вычищая по столбцам, сразу получим линейную комбинацию значений функции и её производных.

Упражнение. *Найдите* $\exp(At)$ *для матрицы* A *с единственным собственным числом* λ*, причём соберите вместе одинаковые степени* t *вне экспонент. Размер* n *матрицы любой, но начните с* $n = 3$*.*

Обращение матрицы Вандермонда

При вычищении определителя могут появляться неудобные длинные выражения в строке матриц и в столбце функций. Избежать их можно, решив уравнение в общем виде, причём это возможно и без полного раскрытия. Источником аналогии тут служит вычисление

$$\begin{vmatrix} a & x \\ v & f \end{vmatrix} = af - xv = 0 \quad \Rightarrow \quad x = av^{-1}f.$$

Порядок сомножителей важен, ибо у нас вместо a формальная строка из степеней матрицы A, вместо f столбец значений функции f на спектре, а вместо v матрица Вандермонда с учётом конфлюэнтных модификаций, если они необходимы. Поразмыслив над процессом вычищения определителя, приходим к искомой матричной формуле.

Теорема. *Решением уравнения*

$$\begin{vmatrix} A^\bullet & X \\ V & F_\bullet \end{vmatrix} = 0$$

с невырожденной матрицей V *является* $X = A^\bullet V^{-1} F_\bullet$*.*

Таким образом, суть вычисления любой достаточно хорошей функции от A — обращение матрицы V, от функции никак не зависящей и определяемой только спектром A. Остаётся понять, почему из нескольких возможных порядков сомножителей работает именно указанный.

Доказательство. Столбцы нашей разноцветно окаймлённой матрицы линейно зависимы, причём после раскрытия определителя коэффициент $\det V$ при X ненулевой именно благодаря конфлюэнтным модификациям. Значит, $X = A^\bullet U_\bullet = \sum A^k u_k$ и надо найти скалярные функции u_k. Однако $F_\bullet = VU_\bullet$, ибо столбец функций должен быть линейной комбинацией столбцов матрицы Вандермонда с теми же коэффициентами. Отсюда находим $U_\bullet = V^{-1}F_\bullet$. □

Здесь можно пытаться продвинуться дальше. Если вести речь о больших матрицах, то полезно знать, что особая структура матрицы Вандермонда упрощает и ускоряет её обращение; есть даже готовые формулы, покрывающие и конфлюэнтный случай. Однако алгоритм нестабилен вблизи сливающихся собственных чисел.

Для функций матрицы известны принципиально другие и весьма красивые формулы и, что вовсе не одно и то же, методы практического вычисления. Однако они менее элементарны и опираются на понятия, недоступные на первом курсе.

Глава 8. ЭВКЛИДОВЫ И ЭРМИТОВЫ ПРОСТРАНСТВА

8.1. Стандартное эвклидово пространство

Стратегия: от линейной алгебры к линейной геометрии

Лекция 4 28.02.20

Перечислим сперва пространства, уже немного изученные в этом курсе, начиная с осени.

1. Трёхмерное пространство $\mathbb{R}^3$ вместе с его геометрией. Важнейшее понятие тут — скалярное произведение векторов, дающее удобные формулы для вычисления расстояний и углов.
2. Пространство столбцов $\mathbb{R}^n$. Всё, что мы прежде делали со столбцами, относилось к линейной алгебре, но без скалярного произведения не содержало почти никакой геометрии.
3. Абстрактное, данное набором аксиом, понятие вещественного линейного пространства. Приводились примеры: пространства многочленов или иных функций, пространства матриц. Опять же, никакой геометрии.

Теперь наша цель — многомерная линейная геометрия, что можно расшифровать как изучение пространств со скалярным произведением. Новые структуры будем вводить постепенно.

1. Стандартное скалярное произведение на $\mathbb{R}^n$.
2. Стандартное комплексное скалярное произведение на $\mathbb{C}^n$.
3. Абстрактное, данное набором аксиом, понятие вещественного или комплексного скалярного произведения. Нестандартное произведение бывает полезным и на пространстве столбцов, а особенно на его подпространствах. Будут также примеры скалярных произведений на пространствах функций и матриц.

Познакомившись с основными геометрическими свойствами конечномерных пространств со скалярным произведением, мы займёмся линейными операторами на них. А в следующем году вы увидите, как большинство изученного здесь развивается в бесконечномерных пространствах функций в курсе основ функционального анализа.

Ортонормированные базисы $\mathbb{R}^n$ и ортогональные матрицы

В привычном трёхмерном пространстве скалярное произведение векторов удобно вычислять по их координатам в стандартном базисе: для $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$ и $\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + y_3\mathbf{e}_3$ находим

$$\mathbf{x} \cdot \mathbf{y} = X^\top Y = x_1y_1 + x_2y_2 + x_3y_3.$$

Введём по аналогии на пространстве столбцов $\mathbb{R}^n$ стандартное скалярное произведение

$$(X, Y) = X^\top Y = x_1 y_1 + \ldots + x_n y_n.$$

Обозначение с точкой не используется; обычно пишут круглые или угловые скобки.

Определение. Базис $\{X_1, \ldots, X_n\}$ называют **ортонормированным**, или кратко **ОНБ**, если все векторы в нём ортогональны друг другу и имеют единичную длину:

$$(X_i, X_j) = \delta_{ij} = \begin{cases} 1 \text{ при } i = j, \\ 0 \text{ при } i \neq j. \end{cases}$$

Матрица перехода от одного ОНБ к другому не может быть произвольной невырожденной матрицей. Запишем векторы произвольного базиса $\mathcal{B}$ в матрицу S по столбцам; она по определению есть матрица перехода к $\mathcal{B}$ от стандартного базиса $\{E^{(1)}, \ldots, E^{(n)}\}$, являющегося ОНБ. Условие ортонормированности базиса $\mathcal{B}$ равносильно матричному равенству $S^\top S = E$, или $S^\top = S^{-1}$, что даёт также $SS^\top = E$. Такие же равенства выполнены для матрицы S^{-1} обратного перехода. Далее, переход между любыми двумя ОНБ можно совершить в два шага, пройдя через стандартный базис, так что матрица перехода будет произведением двух матриц с этим свойством, но

$$S^\top S = E \quad \text{и} \quad T^\top T = E \quad \Longrightarrow \quad (ST)^\top (ST) = T^\top S^\top S\, T = E.$$

По той же причине (упражнение) получается, что $(SX, SY) = (X, Y)$ для всех столбцов $X, Y \in \mathbb{R}^n$. Поэтому оператор, заданный матрицей со свойством $S^\top S = E$, сохраняет все углы, длины и вообще всё выразимое через скалярное произведение. Итак, это условие несомненно выделяет очень важный тип матриц.

Теорема. *Равносильны следующие свойства вещественной квадратной матрицы S:*

(1) *она является матрицей перехода между ОНБ;*
(2) *её столбцы образуют ОНБ;*
(3) *её строки образуют ОНБ;*
(4) *умножение любого вектора на S сохраняет его длину;*
(5) *умножение любой пары векторов на S сохраняет угол между ними;*
(6) $S^\top S = E = SS^\top$;
(7) $S^\top = S^{-1}$.

Доказательство. Упражнение, развлечение (намёки уже даны). □

Определение. Матрица с такими свойствами называется **ортогональной**.

Следствие. *Определитель всякой ортогональной матрицы равен* ± 1.

Доказательство. Поскольку $(\det S)^2 = \det(S^\top S) = \det E = 1$. □

Строение маломерных ортогональных матриц

Теорема. *Всякая ортогональная* 2×2 *или* 3×3 *матрица* S *задаёт либо поворот, либо композицию поворота и отражения.*

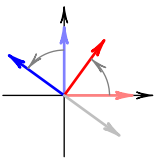
Тождественное преобразование мы посчитаем поворотом.

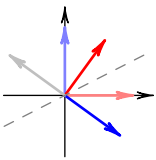
Доказательство. (2×2) Из условия $S^\top S = E$ легко найти конкретный вид матрицы S:

$$S = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \quad \text{или} \quad S = \begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}.$$

Эти два случая соответствуют двум возможностям для $\det S$.

Первый вариант давно знаком: это поворот на угол φ. Во втором варианте образ первого орта тот же, а образ второго орта меняется на противоположный. Значит, это композиция поворота и отражения. Однако полезно увидеть, что второй вариант также является просто отражением относительно прямой с уклоном $\operatorname{tg}\frac{\varphi}{2}$.





(3×3) Характеристический многочлен матрицы S кубический, а потому имеет корень $\lambda \in \mathbb{R}$. Для принадлежащего ему собственного вектора $\mathbf{v}$ заданного матрицей S оператора S находим

$$0 \neq (\mathbf{v},\mathbf{v}) = (\mathsf{S}\mathbf{v},\mathsf{S}\mathbf{v}) = (\lambda\mathbf{v},\lambda\mathbf{v}) = \lambda^2(\mathbf{v},\mathbf{v}) \implies \lambda = \pm 1.$$

Проверим, что ортогональная к $\mathbf{v}$ плоскость Π инвариантна:

$$\begin{aligned} \mathbf{w} \in \Pi &\implies 0 = (\mathbf{v},\mathbf{w}) = (\mathsf{S}\mathbf{v},\mathsf{S}\mathbf{w}) = (\lambda\mathbf{v},\mathsf{S}\mathbf{w}) \\ &\implies 0 = (\mathbf{v},\mathsf{S}\mathbf{w}) \\ &\implies \mathsf{S}\mathbf{w} \in \Pi. \end{aligned}$$

Если остальные корни — комплексно сопряжённая пара, то в Π происходит поворот, а прямая $\langle\mathbf{v}\rangle$ есть ось вращения. В случае $\lambda = -1$ происходит также отражение $\mathbb{R}^3$ относительно Π. Случай полностью

вещественного спектра, помимо операторов E и $-\mathsf{E}$, даёт отражения и повороты на угол π. □

Упражнение 1. *Убедитесь, что матрица*

$$S = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ \sin\varphi & -\cos\varphi & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

задаёт поворот на угол π, и найдите его ось.

Упражнение 2. *Каждая ортогональная 3×3 матрица задаёт композицию отражений в количестве от нуля до трёх.*

Диагонализация симметричных матриц

Как вы помните, матрицу A называют симметричной, если $A^\top = A$. Это выглядит слегка похожим на условие ортогональности и равносильно тому, что $(AX, Y) = (X, AY)$ для всех $X, Y \in \mathbb{R}^n$, поскольку последнее равенство разворачивается в $X^\top A^\top Y = X^\top A Y$. Симметричные матрицы очень часто возникают в приложениях и обладают многими хорошими свойствами по сравнению с общими квадратными.

Теорема. *Собственные векторы, отвечающие различным собственным числам вещественной симметричной матрицы, ортогональны.*

Доказательство. Проверим, что собственные векторы $\mathbf{v}_i$, принадлежащие различным собственным числам λ_i заданного матрицей A оператора, ортогональны друг другу:

$$\begin{aligned} 0 &= (\mathsf{A}\mathbf{v}_1, \mathbf{v}_2) - (\mathbf{v}_1, \mathsf{A}\mathbf{v}_2) \\ &= (\lambda_1 \mathbf{v}_1, \mathbf{v}_2) - (\mathbf{v}_1, \lambda_2 \mathbf{v}_2) \\ &= (\lambda_1 - \lambda_2)(\mathbf{v}_1, \mathbf{v}_2) \implies (\mathbf{v}_1, \mathbf{v}_2) = 0. \end{aligned}$$

Заметим некоторое сходство этого рассуждения с доказательством более слабого свойства линейной независимости собственных векторов в предыдущей главе. При этом тут доказательство существенно проще. Можно расценить это как демонстрацию силы, приносимой скалярным произведением, но ведь и на матрицы наложено условие. □

Теорема. *Для всякой вещественной симметричной матрицы A:*

(1) *все характеристические корни вещественны;*
(2) *найдётся ОНБ из собственных векторов;*
(3) *найдутся такая диагональная матрица D и такая ортогональная матрица S, что $D = S^\top A S = S^{-1} A S$.*

Здесь совпали два закона преобразования матриц при смене базиса, встречавшихся в разных главах курса: по закону $S^\top A\,S$ преобразуются матрицы квадратичной формы, а по закону $S^{-1}A\,S$ — матрицы оператора. Поэтому понятия и результаты теории операторов оказываются приложимы к приведению поверхностей второго порядка к каноническому виду (не только в $\mathbb{R}^3$, но и в $\mathbb{R}^n$). Исторически было наоборот: первыми появились квадратичные формы, а операторы позже.

Маломерное доказательство. Формулировка тут дана очень ёмкая, особенно (2), и нужно установить целую цепочку свойств. В малых размерностях это легко делается кустарно.

(2×2) Характеристический многочлен матрицы A квадратный, а его дискриминант Δ, равный $(a_{11}-a_{22})^2+4a_{12}^2$, неотрицателен; более того, если A не скалярная матрица λE, то $\Delta>0$, характеристические корни различны, и у каждого есть свой собственный вектор.

(3×3) Характеристический многочлен матрицы A кубический, а потому имеется корень $\lambda\in\mathbb{R}$ и принадлежащий ему собственный вектор $\mathbf{v}$ заданного матрицей A оператора A. Как и в случае ортогональной матрицы, ортогональная к $\mathbf{v}$ плоскость Π инвариантна:

$$\begin{aligned}\mathbf{w}\in\Pi &\implies 0=(\lambda\mathbf{v},\mathbf{w})=(\mathsf{A}\mathbf{v},\mathbf{w})=(\mathbf{v},\mathsf{A}\mathbf{w})\\ &\implies \mathsf{A}\mathbf{w}\in\Pi.\end{aligned}$$

В плоскости Π задача свелась к уже разобранному 2-мерному случаю, ибо сужение оператора A на Π унаследует его симметричность. □

Общее доказательство мы получим в следующем разделе. Для диагонализации оператора прямая сумма его собственных подпространств обязана давать всё пространство, поэтому нам останется найти достаточное количество собственных векторов, что удобнее в комплексной обстановке. Также надо выбрать ортогональный базис в каждом собственном подпространстве. Это достижимо вообще в любом подпространстве, как мы сейчас установим без опоры на диагонализацию.

Метод ортогонализации Грама — Шмидта

Выберем ненулевой вектор $\mathbf{v}$. Как и в обычном трёхмерном пространстве, можно разложить любой вектор на две составляющие: первая коллинеарна $\mathbf{v}$ и называется проекцией на $\mathbf{v}$, а вторая ортогональна $\mathbf{v}$. Формула для нахождения проекции выводится также и остаётся прежней:

$$\mathrm{pr}_{\mathbf{v}}\,\mathbf{x}=\frac{(\mathbf{v},\mathbf{x})}{(\mathbf{v},\mathbf{v})}\mathbf{v};$$

однако впредь мы привыкаем ставить $\mathbf{x}$ на второе место в числителе, поскольку именно так позже получится в комплексном случае.

Теорема. *Для всякого списка векторов $\mathcal{X}$ найдутся ортогональный и ортонормированный базисы линейной оболочки $\langle\mathcal{X}\rangle$.*

Простейший алгоритм построения ортогонального базиса называют **методом ортогонализации Грама — Шмидта**, хотя им пользовались гораздо раньше них (как минимум Лаплас раньше на век). Этот метод безразличен к способу вычисления скалярных произведений и потому работает без изменений в любом пространстве, включая и комплексные. То же самое относится ко всему дальнейшему «ортогональному» материалу, излагаемому далее до введения комплексных скалярных произведений.

Доказательство. Достаточно рассмотреть линейно независимые списки. На практике линейная зависимость между векторами $\{\mathbf{x}_1,\dots,\mathbf{x}_m\}$, если она есть, всегда обнаруживается в ходе вычислений. Строим ортогональный базис $\{\mathbf{h}_1,\dots,\mathbf{h}_s\}$ последовательно: на каждом шаге цикла вычитанием проекций находим новый вектор, а именно

$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{x}_1,\\
\mathbf{h}_2 &= \mathbf{x}_2 - \operatorname{pr}_{\mathbf{h}_1}\mathbf{x}_2,\\
\mathbf{h}_3 &= \mathbf{x}_3 - \operatorname{pr}_{\mathbf{h}_1}\mathbf{x}_3 - \operatorname{pr}_{\mathbf{h}_2}\mathbf{x}_3,\\
&\dots\\
\mathbf{h}_k &= \mathbf{x}_k - \sum_{1\leqslant j<k} \operatorname{pr}_{\mathbf{h}_j}\mathbf{x}_k.
\end{aligned}$$

При этом, если предыдущие $\mathbf{h}_i$ ортогональны между собой, то и $\mathbf{h}_k$ ортогонален им: умножая скалярно на $\mathbf{h}_i$, для всех $i<k$ имеем

$$(\mathbf{h}_i,\mathbf{h}_k) = (\mathbf{h}_i,\mathbf{x}_k) - \sum_{1\leqslant j<k} (\mathbf{h}_i,\operatorname{pr}_{\mathbf{h}_j}\mathbf{x}_k) = (\mathbf{h}_i,\mathbf{x}_k) - (\mathbf{h}_i,\operatorname{pr}_{\mathbf{h}_i}\mathbf{x}_k) = 0.$$

Внезапное равенство $\mathbf{h}_k=\mathbf{0}$ означает, что $\mathbf{x}_k$ лежит в линейной оболочке $\langle\mathbf{x}_1,\dots,\mathbf{x}_{k-1}\rangle$ предыдущих; в таком случае мы выбрасываем оказавшийся лишним вектор $\mathbf{x}_k$, а остаток списка $\mathcal{X}$, который ещё не был задействован, перенумеровываем ради удобства. Затем нормирование даёт единичный вектор $\mathbf{f}_k=\mathbf{h}_k/\|\mathbf{h}_k\|$.

В описанном методе всегда $\mathbf{h}_k$ и $\mathbf{f}_k$ лежат в линейной оболочке $\langle\mathbf{x}_1,\dots,\mathbf{x}_k\rangle$; поэтому, при линейной независимости $\mathcal{X}$, матрицы перехода от $\mathcal{X}$ к построенным базисам верхнетреугольны. □

Следствие. *Всякая невырожденная матрица A представляется произведением QR, где Q ортогональна, а R треугольна.*

Доказательство. Столбцы любой невырожденной матрицы A образуют базис пространства $\mathbb{R}^n$. Ортогонализация его методом Грама — Шмидта даст столбцы, которые мы соберём в ортогональную матрицу Q. Скалярные произведения столбцов Q и столбцов A составят матрицу $R = Q^\top A$, и она же в силу равенства $Q = AR^{-1}$ будет матрицей перехода от конечного базиса к исходному. □

Разложение вектора по ортонормированной системе

Упражнение 3 (теорема Пифагора)**.** *Если в списке $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ все векторы ортогональны друг другу, то*

$$\Big\| \sum_{1\leqslant i\leqslant k} \mathbf{x}_i \Big\|^2 = \sum_{1\leqslant i\leqslant k} \|\mathbf{x}_i\|^2.$$

Лемма. *Для любого ОНБ $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ и любого вектора $\mathbf{x}$ выполнено равенство*

$$\mathbf{x} = \sum_{1\leqslant i\leqslant n} \mathrm{pr}_{\mathbf{e}_i} \mathbf{x} = \sum_{1\leqslant i\leqslant n} (\mathbf{e}_i, \mathbf{x})\mathbf{e}_i.$$

Доказательство. Запишем $\mathbf{x} = \sum x_j \mathbf{e}_j$ и найдём неизвестные коэффициенты, скалярно умножая обе части на $\mathbf{e}_i$. □

Величины $(\mathbf{e}_i, \mathbf{x})$ называются **ковекторными** координатами вектора $\mathbf{x}$. Когда базис не ортонормирован, векторные и ковекторные координаты одного и того же вектора обычно отличаются, но мы не будем сейчас углубляться в этот вопрос. Лемма утверждает, что в ОНБ привычные векторные и непривычные ковекторные координаты совпадают.

задания

Упражнение 4 (равенство Парсеваля)**.** *Для любого ОНБ $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ и любого вектора $\mathbf{x}$ верно*

$$\|\mathbf{x}\|^2 = \sum_{1\leqslant i\leqslant n} |(\mathbf{e}_i, \mathbf{x})|^2.$$

Теорема (неравенство Бесселя)**.** *Для любого ортонормированного списка векторов $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ эвклидова пространства:*

(1) *неравенство*

$$\|\mathbf{x}\|^2 \geqslant \sum_{1\leqslant i\leqslant k} \|\mathrm{pr}_{\mathbf{e}_i} \mathbf{x}\|^2 = \sum_{1\leqslant i\leqslant k} |(\mathbf{e}_i, \mathbf{x})|^2$$

выполнено для каждого вектора $\mathbf{x}$.

(2) *Равенство достигается* $\iff$ *вектор* $\mathbf{x}$ *лежит в линейной оболочке* $\langle \mathbf{e}_1, \dots, \mathbf{e}_k \rangle$.
(3) *Равенство достигается для всех векторов пространства* $\iff$ *список* $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ *является базисом всего пространства.*

Доказательство. Если это необходимо, дополним исходный список до ОНБ $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Полагая для краткости $x_i = (\mathbf{e}_i, \mathbf{x})$, запишем равенство Парсеваля и отделим последние слагаемые:

$$\|\mathbf{x}\|^2 = \sum_{1 \leqslant i \leqslant n} |x_i|^2 = \sum_{1 \leqslant i \leqslant k} |x_i|^2 + \sum_{k < i \leqslant n} |x_i|^2 \geqslant \sum_{1 \leqslant i \leqslant k} |x_i|^2.$$

Равенство достигается $\iff$ выделенная вторая сумма равна нулю. Оно достигается для всех векторов $\iff$ нет вектора, не являющегося линейной комбинацией векторов исходного списка. □

Ортогональное дополнение к подпространству

Договоримся для удобства записывать условие ортогональности произвольных подмножеств $\mathcal{X}$ и $\mathcal{Y}$ пространства в краткой форме:

$$(\mathcal{X}, \mathcal{Y}) = 0 \iff (\mathbf{x}, \mathbf{y}) = 0 \text{ для всех } \mathbf{x} \in \mathcal{X} \text{ и всех } \mathbf{y} \in \mathcal{Y},$$

и опускать фигурные скобки, когда одно из подмножеств содержит единственный вектор, иными словами, при $\mathcal{X} = \{\mathbf{x}\}$.

Упражнение 5. *Проверьте, что* $(\mathbf{x}, \mathcal{Y}) = 0 \Leftrightarrow (\mathbf{x}, \langle \mathcal{Y} \rangle) = 0$ *для всяких* $\mathbf{x} \in \mathcal{V}$ *и* $\mathcal{Y} \subseteq \mathcal{V}$, *то есть подмножество в таком условии всегда можно заменить его линейной оболочкой.*

Аналогично, $(\mathcal{X}, \mathcal{Y}) = 0 \Leftrightarrow (\langle \mathcal{X} \rangle, \langle \mathcal{Y} \rangle) = 0$ *для всяких* $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{V}$.

В линейном пространстве $\mathcal{V}$ почти все подпространства имеют много дополнений. Способа выбрать среди них «лучшее» в общем случае нет. А вот при наличии в $\mathcal{V}$ скалярного произведения такой способ есть: каждое подпространство $\mathcal{L}$ имеет **ортогональное дополнение**

$$\mathcal{L}^\perp = \{\mathbf{x} \in \mathcal{V} \mid (\mathbf{x}, \mathcal{L}) = 0\}.$$

Теорема. *В любом пространстве* $\mathcal{V}$ *со скалярным произведением, для каждого подпространства* $\mathcal{L}$*:*

(1) *ортогональное дополнение* $\mathcal{L}^\perp$ *является подпространством;*
(2) $\mathcal{V} = \mathcal{L} \oplus \mathcal{L}^\perp$, *чем и оправдан термин для* $\mathcal{L}^\perp$*;*
(3) $(\mathcal{L}^\perp)^\perp = \mathcal{L}$.

Доказательство. (1) Подставим в определение линейную комбинацию векторов из $\mathcal{L}^\perp$.

(2) Выберем в $\mathcal{L}$ любой базис, дополним его до базиса всего $\mathcal{V}$ и ортогонализуем. Полученный ОНБ можно разбить на две части $\mathcal{X}$ и $\mathcal{Y}$ так, что $\mathcal{X}$ будет базисом $\mathcal{L}$. Тогда $\langle\mathcal{Y}\rangle \subseteq \mathcal{L}^\perp$, ибо $(\mathcal{X},\mathcal{Y}) = 0$. Возьмём вектор $\mathbf{x}+\mathbf{y}$ вне $\langle\mathcal{Y}\rangle$, так что $\mathbf{x} \neq \mathbf{0}$. Тогда $(\mathbf{x}, \mathbf{x}+\mathbf{y}) = (\mathbf{x},\mathbf{x}) \neq 0$, поэтому $\mathbf{x}+\mathbf{y} \notin \mathcal{L}^\perp$. Значит, $\langle\mathcal{Y}\rangle \supseteq \mathcal{L}^\perp$, а потому $\langle\mathcal{Y}\rangle = \mathcal{L}^\perp$. Однако по построению $\mathcal{V} = \langle\mathcal{X}\cup\mathcal{Y}\rangle = \langle\mathcal{X}\rangle \oplus \langle\mathcal{Y}\rangle$.

(3) Следует из описания $\mathcal{L}^\perp$ в предыдущем абзаце. □

8.2. Унитарная триангуляция

Стандартное эрмитово пространство

Лекция 5
06.03.20

Благодаря своим особым свойствам, вещественные ортогональные и симметричные матрицы применяются очень широко. Практически такими же свойствами обладают их комплексные обобщения, вводимые в этом разделе; все определения отличаются от вещественного случая лишь наличием комплексного сопряжения в нужных местах. Главное из этих мест — естественное обобщение транспонирования. Операция **эрмитова сопряжения** $A \mapsto \bar{A}^\top$ выполняется комплексным сопряжением всех элементов матрицы и транспонированием её, а порядок этих двух шагов не важен. Несколько обозначений в ходу для матрицы $\bar{A}^\top$: в математике обычно A^*, а физики пишут $A^\dagger$ или A^+. При чтении визуально качественного материала лишь значок кинжала ни с чем нельзя спутать, но с доски хуже. Отметим, что всегда $(A^\dagger)^\dagger = A$.

Оснастим линейное пространство $\mathbb{C}^n$ комплексных столбцов скалярным произведением:

x_i^*?

$$(X,Y) = X^\dagger Y = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n.$$

Тогда для каждого комплексного вектора X однозначно определена вещественная неотрицательная величина

$$\|X\| = \sqrt{(X,X)} = \sqrt{|x_1|^2 + \ldots + |x_n|^2},$$

имеющая смысл **длины** этого вектора. Упустив комплексное сопряжение при первой наивной попытке ввести скалярное произведение, мы лишились бы понятия длины, *вещественной и неотрицательной*. Вместо линейности скалярного произведения по первому аргументу, здесь мы наблюдаем свойство

$$(\alpha X, Y) = \bar{\alpha}(X,Y) \text{ для всех } \alpha \in \mathbb{C}.$$

Комплексное линейное пространство со скалярным произведением называют **эрмитовым** либо **унитарным пространством**. Там, как и в ве-

щественном эвклидовом пространстве, имеются ортонормированные базисы.

Ортонормированные базисы $\mathbb{C}^n$ и унитарные матрицы

Лемма. *Равносильны свойства комплексной матрицы, аналогичные определяющим свойствам вещественной ортогональной матрицы, с заменой транспонирования на эрмитово сопряжение.*

Доказательство. Упражнение, $U^\dagger U = E$. □

Определение. Матрица с такими свойствами называется **унитарной**.

Пример. Диагональная матрица является унитарной, если и только если все её диагональные элементы по модулю равны 1. Например, в размерности 3 она имеет вид

$$\begin{bmatrix} e^{i\varphi_1} & 0 & 0 \\ 0 & e^{i\varphi_2} & 0 \\ 0 & 0 & e^{i\varphi_3} \end{bmatrix}, \quad \varphi_k \in \mathbb{R}.$$

Следствие. *Произведение унитарных матриц унитарно.*

Следствие. *Определитель всякой унитарной матрицы по модулю равен 1.*

Доказательство. Обозначим $\det U = z$. Тогда $\det U^\dagger = \bar{z}$ и $U^\dagger U = E$ влечёт $|z|^2 = \bar{z}z = 1$. □

Теорема. *Всякая невырожденная комплексная матрица представляется произведением QR, где Q унитарна, а R треугольна.*

Доказательство. Аналогично вещественному QR-разложению. □

Унитарная триангуляция Шура

Теорема (Schur)**.** *Для всякой комплексной матрицы A найдутся унитарная матрица U и треугольная матрица T такие, что $U^\dagger A\, U = T$.*

Иными словами, всякая комплексная матрица унитарно подобна треугольной матрице. Это означает, что всякий линейный оператор можно представить треугольной матрицей в подходящем ОНБ. Если в теореме Шура взять вещественную матрицу A с полностью вещественным спектром, то всё вычисление становится вещественным, а матрица перехода U будет ортогональна. Без вещественности спектра достичь триангуляции ортогональным переходом не удастся, ибо на диагонали в T оказываются собственные числа A.

Доказательство. Возьмём $\lambda_1 \in \operatorname{Spec} \mathsf{A}$. Соответствующий собственный вектор $\mathbf{v}_1$ дополним произвольно до базиса $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ всего пространства $\mathbb{C}^n$. В нём матрица оператора A принимает блочный верхнетреугольный вид:

$$S^{-1}A\,S = B = \left[\begin{array}{c|c} \lambda_1 & * \\ \hline 0 & B_1 \end{array}\right].$$

Воспользуемся QR-разложением $S = Q_1 R$ с верхнетреугольной матрицей R. Заметим, что $Q_1^\dagger A\,Q_1 = Q_1^{-1} A\,Q_1 = RS^{-1}A\,SR^{-1} = RBR^{-1}$ также имеет блочный верхнетреугольный вид

$$\left[\begin{array}{c|c} \lambda_1 & * \\ \hline 0 & A_1 \end{array}\right].$$

Повторяя трюк для блока A_1, найдём унитарную матрицу Q_2, дающую

$$Q_2^\dagger Q_1^\dagger A Q_1 Q_2 = \left[\begin{array}{c|c|c} \lambda_1 & * & * \\ \hline 0 & \lambda_2 & * \\ \hline 0 & 0 & A_2 \end{array}\right].$$

Продолжая триангуляцию, накопим произведение $U = Q_1 Q_2 \dots Q_{n-1}$ унитарных матриц перехода, а оно всегда унитарно. □

В приведённом рассуждении $\operatorname{Spec} A_1 \subset \operatorname{Spec} A$ ввиду устроенного подобия, и в итоге диагональ матрицы T состоит из собственных чисел матрицы A с правильными кратностями.

Упражнение 6. *Для треугольной матрицы T докажите равенство*

$$\prod_{\lambda \in \operatorname{Spec} T} (T - \lambda E) = 0.$$

Упражнение 7. *С помощью предыдущего упражнения и теоремы Шура докажите теорему Гамильтона — Кэли.*

Эрмитовы матрицы и спектральная теорема

Определение. **Эрмитовой** (или **самосопряжённой**) называется комплексная матрица $A = A^\dagger$.

Упражнение 8. *Если матрица A эрмитова, то квадратичная форма $(X, AX) = X^\dagger AX$ принимает вещественные значения для всех комплексных векторов X.*

Теорема. *Всякая эрмитова матрица A обладает полностью вещественным спектром и ОНБ из собственных векторов.*

Доказательство. По теореме Шура найдём U и $T = U^\dagger A\, U$. Тогда

$$T^\dagger = U^\dagger A^\dagger U = U^\dagger A\, U = T.$$

Треугольная матрица T ещё и эрмитова! Такой может быть только вещественная диагональная матрица. Спектры подобных матриц A и T одинаковы, а искомый собственный ОНБ для матрицы A составляют столбцы матрицы U. □

Полезно также провести независимое рассуждение, привлекающее сведения из многомерного анализа.

25.05.20

Второе доказательство. Для собственного числа λ и соответствующего ему собственного вектора X эрмитовой матрицы A напишем

$$\bar\lambda X^\dagger X = (\lambda X)^\dagger X = (AX)^\dagger X = X^\dagger A^\dagger X = X^\dagger (AX) = \lambda X^\dagger X.$$

Отсюда $\bar\lambda = \lambda$, ибо $X^\dagger X \neq 0$. Для эрмитовой матрицы с простым спектром, то есть с различными собственными значениями, сразу получаем достаточное количество собственных векторов, а их ортогональность установлена ещё в предыдущем разделе.

Всякая матрица с кратными собственными значениями является пределом матриц без оных: $A_\varepsilon \to A$ при $\varepsilon \to 0$, где в данном случае все A_ε эрмитовы с простым спектром. Поэтому они унитарно диагонализуемы, то есть $A_\varepsilon = U_\varepsilon D_\varepsilon U_\varepsilon^\dagger$ с унитарными матрицами U_ε и диагональными матрицами D_ε. В пределе при $\varepsilon \to 0$ получим искомую диагонализацию $A = UDU^\dagger$, причём U унитарна в силу компактности (замкнутости и ограниченности) множества унитарных матриц фиксированного размера.

Ключ к успеху в том, что при предельном переходе столбцы матрицы U_ε остаются ортонормированными. Без этого условия в пределе возможно вырождение, с чем напрямую связана вся теория жордановой формы, в этом курсе еле затронутая. □

В частности, здесь доказаны спектральные свойства вещественных симметричных матриц, анонсированные ранее.

Следствие. *Всякая симметричная вещественная матрица A обладает полностью вещественным спектром и ОНБ из вещественных собственных векторов.*

Доказательство. Такая матрица является эрмитовой, поэтому вещественность спектра и наличие собственного ОНБ следуют. Кроме того, для вещественной матрицы с полностью вещественным спектром решения линейных уравнений $AX = \lambda X$ тоже вещественны. □

Другие версии спектральной теоремы

Определение. **Косоэрмитовой** (или **кососамосопряжённой**) называется комплексная матрица $A = -A^\dagger$.

Теорема. *Всякая косоэрмитова матрица A обладает полностью мнимым спектром и ОНБ из собственных векторов.*

Полная мнимость означает, что $\mathfrak{Re}\,\lambda = 0$ для всех $\lambda \in \operatorname{Spec} A$, так что $\lambda = 0$ допускается и, более того, в нечётных размерностях всегда присутствует.

Доказательство. Как и выше, получим из теоремы Шура, что равенство $A^\dagger = -A$ влечёт $T^\dagger = -T$. Тогда T диагональна и её диагональ полностью мнимая. □

Теорема. *Всякая унитарная матрица A обладает спектром из чисел, равных 1 по модулю, и ОНБ из собственных векторов.*

Доказательство. Как и выше, получим из теоремы Шура, что равенство $A^\dagger = A^{-1}$ влечёт $T^\dagger = T^{-1}$, и тогда T диагональна. □

Упражнение 9. *Сформулируйте и докажите спектральную теорему для матриц со свойством $A^\dagger A = AA^\dagger$, которое покрывает все три только что рассмотренных случая.*

Разложение Холецкого

Среди эрмитовых матриц, и симметричных в частности, выделяются положительно определённые. Они уже появлялись у нас в конце первого семестра при изучении квадратичных форм. В приложениях положительная определённость обычно следует из требований задачи, скажем, из физических соображений. Случается это и в статистике; пример показан в следующей главе. 25.05.20

Упражнение 10. *Матрицы $A^\dagger A$ и $AA^\dagger$ положительно определённые эрмитовы для всякой невырожденной матрицы A.*

Крайне полезным свойством положительно определённых эрмитовых матриц является возможность представить всякую такую матри-

цу P в виде произведения $P = LL^\dagger$, где матрица L нижняя треугольная, а потому $L^\dagger$ верхняя треугольная. Доказательство несложно построить на основе материала этой главы, и оно отнесено в задания.

Вычислительные алгоритмы для этого разложения весьма эффективны. В свою очередь, при его наличии легко и устойчиво (в техническом, вычислительном смысле) решаются системы линейных уравнений $PX = B$: достаточно решить две системы $LY = B$ и $L^\dagger X = Y$ с треугольными матрицами, что выполняется простой подстановкой уже найденных значений в следующее уравнение системы. Также этим методом быстро находится обратная матрица, причём он применим для любой невырожденной матрицы A ввиду формулы $A^{-1} = A^\dagger(AA^\dagger)^{-1}$, в правой части которой обращается положительно определённая эрмитова матрица.

Упомянем ещё модификацию разложения Холецкого. Это представление $A = LDL^\dagger$, где в середине появляется диагональная матрица, а на диагонали матрицы L остаются сплошь единицы (**унитреугольная** матрица). Оно существует не только для положительно определённых A. Входить в детали здесь не будем; отметим лишь, что ещё в первом семестре у нас была теорема Якоби об определении сигнатуры вещественной квадратичной формы по главным минорам её матрицы. Так вот в её доказательстве, фактически, кропотливо строится искомая матрица L, поскольку каждый переход там унитреуголен.

8.3. Общие эвклидовы и эрмитовы пространства

Стандартное скалярное произведение

Скалярное произведение векторов вводилось сперва в «обычном» трёхмерном пространстве:

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3;$$

затем в вещественном пространстве столбцов $\mathbb{R}^n$:

$$(\mathbf{x}, \mathbf{y}) = X^\top Y = \sum_{1 \leqslant k \leqslant n} x_k y_k;$$

и наконец, в комплексном пространстве столбцов $\mathbb{C}^n$:

$$(\mathbf{x}, \mathbf{y}) = X^\dagger Y = \sum_{1 \leqslant k \leqslant n} \bar{x}_k y_k.$$

Назовём эти скалярные произведения **стандартными**.

Необходимо не забывать о наличии в последней формуле комплексного сопряжения. Без него **норме** $\|\mathbf{x}\| = \sqrt{(\mathbf{x},\mathbf{x})}$ нельзя придать смысл длины вектора $\mathbf{x}$, ибо она не обязательно будет вещественным числом. Однако навесить комплексное сопряжение можно либо на компоненты первого вектора, либо на компоненты второго, и оба варианта распространены в математике. Мой выбор — всегда сопрягать первый вектор. Во-первых, запись $X^\dagger Y$ элегантнее, чем $X^\top \bar{Y}$. Во-вторых, в квантовой механике принято поступать именно так, хотя сами скалярные произведения обозначают иначе, следуя Дираку. Великий физик высоко ценил элегантность формул.

Свойства стандартного вырастают в аксиомы общего

Непосредственно из координатных формул видны основные свойства стандартных скалярных произведений. Это (анти)линейность по первому аргументу:

$$(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = (\mathbf{x}_1, \mathbf{y}) + (\mathbf{x}_2, \mathbf{y}), \quad (\alpha\mathbf{x}, \mathbf{y}) = \bar{\alpha}(\mathbf{x}, \mathbf{y});$$

линейность по второму аргументу:

$$(\mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2) = (\mathbf{x}, \mathbf{y}_1) + (\mathbf{x}, \mathbf{y}_2), \quad (\mathbf{x}, \beta\mathbf{y}) = \beta(\mathbf{x}, \mathbf{y});$$

эрмитова симметричность:

$$(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})};$$

положительная определённость:

$$(\mathbf{x}, \mathbf{x}) \geqslant 0, \text{ причём } (\mathbf{x}, \mathbf{x}) = 0 \implies \mathbf{x} = \mathbf{0}.$$

Эрмитовым (также унитарным) **пространством** называют линейное пространство $\mathcal{V}$ над $\mathbb{C}$ вместе с определённой на нём комплекснозначной функцией двух аргументов, обладающей перечисленными свойствами. Саму эту функцию называют скалярным произведением, или **эрмитовой структурой**. Аналогично, с заменой всего комплексного на всё вещественное, вводят **эвклидово пространство**.

Матрица Грама

Зафиксируем базис $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ эрмитова пространства $\mathcal{V}$. Выражая в этом базисе векторы и получаемые из них величины, будем применять обозначения как матричные, так и индексные (вместе с правилом Эйнштейна, то есть опуская знаки суммирования).

Пользуясь аксиомами, преобразуем скалярное произведение двух векторов $\mathbf{x} = x^i\mathbf{e}_i$ и $\mathbf{y} = y^j\mathbf{e}_j$:

$$(\mathbf{x},\mathbf{y}) = (x^i\mathbf{e}_i, y^j\mathbf{e}_j) = \bar{x}^i(\mathbf{e}_i, y^j\mathbf{e}_j) = \bar{x}^i y^j(\mathbf{e}_i,\mathbf{e}_j).$$

Значит, для вычисления любых скалярных произведений достаточно знать скалярные произведения $g_{ij} \leftrightharpoons (\mathbf{e}_i,\mathbf{e}_j)$. Матрицу

$$G = G(\mathbf{e}_1,\ldots,\mathbf{e}_n),$$

составленную из этих чисел, называют **матрицей Грама** выбранного базиса. По определению и аксиоме эрмитовой симметричности $G^\dagger = G$. Базис является ОНБ тогда и только тогда, когда $G = E$. В матричной записи получаем вычислительную формулу

$$(\mathbf{x},\mathbf{y}) = X^\dagger G Y.$$

Упражнение 11. *Выведите формулу $G' = S^\dagger G S$ изменения матрицы Грама при переходе к новому базису, где S есть матрица перехода.*

Упражнение 12. *Проследите пошагово поведение матрицы Грама в процессе ортогонализации как с нормированием ($\mathbf{v}_j \rightsquigarrow \mathbf{f}_j$), так и без него ($\mathbf{v}_j \rightsquigarrow \mathbf{h}_j$). Что происходит с её определителем?*

Лемма. *Матрица Грама всякого списка векторов неотрицательно полуопределена.*

Для $G = G(\mathbf{v}_1,\ldots,\mathbf{v}_k)$ это значит, что $X^\dagger G X \geqslant 0$ для всех столбцов $X \in \mathbb{C}^k$. В этом случае пишут $G \geqslant 0$.

Доказательство. Каждый столбец $X = (x_1,\ldots,x_k)^\top$ задаёт вектор $\mathbf{x} = x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k$, причём $(\mathbf{x},\mathbf{x}) = X^\dagger G X$ неотрицательно по последней из аксиом скалярного произведения. □

Упражнение 13. *Докажите, что $G(\mathcal{X})$ невырождена (значит, положительно определена) тогда и только тогда, когда список $\mathcal{X}$ линейно независим.*

Наоборот, начнём с любого линейного пространства $\mathcal{V}$ без эрмитовой структуры. При любом выборе в нём базиса и любом выборе положительно определённой эрмитовой матрицы G в качестве его матрицы Грама функция, вычисляемая по формуле $X^\dagger G Y$, задаёт эрмитову структуру на $\mathcal{V}$. А по сказанному выше, всякая эрмитова структура получается таким образом, других не бывает.

Неравенство Коши и углы

Теорема. *Для всех пар векторов эвклидова или эрмитова пространства выполнено неравенство*

$$(\mathbf{x},\mathbf{y})(\mathbf{y},\mathbf{x}) \leqslant (\mathbf{x},\mathbf{x})(\mathbf{y},\mathbf{y}).$$

Равенство достигается $\iff$ векторы линейно зависимы, то есть коллинеарны.

Доказательство. Неравенство утверждает, что неотрицателен определитель матрицы Грама

$$G(\mathbf{x},\mathbf{y}) = \begin{bmatrix} (\mathbf{x},\mathbf{x}) & (\mathbf{x},\mathbf{y}) \\ (\mathbf{y},\mathbf{x}) & (\mathbf{y},\mathbf{y}) \end{bmatrix},$$

а это установлено выше в лемме. Утверждение о равенстве следует из упражнения 13. □

Следствие (неравенство Коши)**.** $|(\mathbf{x},\mathbf{y})| \leqslant \|\mathbf{x}\|\|\mathbf{y}\|$.

Из этого неравенства ясно, что в эвклидовом пространстве углы находят по обычной формуле

$$\cos\varphi(\mathbf{x},\mathbf{y}) = \frac{(\mathbf{x},\mathbf{y})}{\|\mathbf{x}\|\|\mathbf{y}\|} \quad \Longrightarrow \quad -1 \leqslant \cos\varphi \leqslant 1.$$

В эрмитовом же пространстве приходится побороть невещественность скалярного произведения $(\mathbf{x},\mathbf{y})$ взятием его по модулю, что ведёт к отсутствию тупых углов: $\cos\varphi \geqslant 0$.

Следствие (неравенство треугольника)**.** $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$.

Доказательство. Прибавим $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ к неравенству

$$2\,\mathfrak{Re}(\mathbf{x},\mathbf{y}) \leqslant 2|(\mathbf{x},\mathbf{y})| \leqslant 2\|\mathbf{x}\|\|\mathbf{y}\|.$$

Получим слева $\|\mathbf{x}+\mathbf{y}\|^2$, а справа $(\|\mathbf{x}\| + \|\mathbf{y}\|)^2$. □

Объёмы и расстояния

Ортонормированность выбранного базиса выражается равенством $G = E$, а единичная матрица выпадает из формул, ведь в произведениях она просто не заметна. Так что одной из первых задач геометрии эвклидова пространства становится получение формул, применимых в любом базисе.

Для любого списка $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ векторов линейного пространства $\mathcal{V}$ назовём параллелепипедом на $\mathcal{X}$ множество

$$\mathcal{P}(\mathcal{X}) = \{\alpha_1\mathbf{x}_1 + \ldots + \alpha_k\mathbf{x}_k \mid 0 \leqslant \alpha_j \leqslant 1 \text{ для всех } 1 \leqslant j \leqslant k\}.$$

Не будем огорчаться, когда в случае линейно зависимого списка параллелепипед вырождается.

Теорема. *В эвклидовом или эрмитовом пространстве для любого списка $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ векторов квадрат k-мерного объёма параллелепипеда $\mathcal{P}(\mathcal{X})$ равен определителю матрицы Грама $G(\mathcal{X})$.*

Доказательство. Если список $\mathcal{X}$ линейно зависим, то и k-мерный объём, и определитель равны нулю. Если нет, то ограничим рассмотрение на линейную оболочку $\langle\mathcal{X}\rangle$. Обозначим через S матрицу перехода от какого-то ОНБ $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ этой оболочки к её базису $\mathcal{X}$. По формуле перехода получим $G(\mathcal{X}) = S^\dagger E S$, так что $\det G(\mathcal{X}) = |\det S|^2$. Определитель матрицы перехода равен по модулю коэффициенту искажения объёмов, то есть отношению объёма образа $\mathcal{P}(\mathcal{X})$ к объёму исходного кубика $\mathcal{P}(\mathbf{e}_1, \ldots, \mathbf{e}_k)$.

Можно рассуждать немного иначе, строя из $\mathcal{X}$ ортогональный базис методом Грама — Шмидта без нормирования. При этом сохраняются и объём, и $\det G$. В итоге параллелепипед прямоуголен, а G диагональна, так что обе части искомого равенства получаются перемножением компонент. □

Следствие. *В эвклидовом или эрмитовом пространстве квадрат расстояния от точки $\mathbf{x}_0$ до линейного подпространства с базисом $\mathcal{X}$ равен*

$$\operatorname{dist}(\mathbf{x}_0, \langle\mathcal{X}\rangle)^2 = \frac{\det G(\{\mathbf{x}_0\} \cup \mathcal{X})}{\det G(\mathcal{X})}.$$

Доказательство. Рассуждение аналогично выводу формулы для расстояния между скрещивающимися прямыми в $\mathbb{R}^3$. Искомое расстояние находится как высота параллелепипеда, равная его объёму, поделённому на «площадь» основания, которая также есть объём параллелепипеда, но имеющего размерность на единицу ниже. □

8.4. Нормальные операторы

Лекция 6
13.03.20

Сопряжённый оператор

Теорема. *На эвклидовом или эрмитовом пространстве $\mathcal{V}$ для каждого линейного оператора A найдётся единственный такой линейный оператор A^*, что*

$$(\mathsf{A}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathsf{A}^*\mathbf{y})$$

для всех векторов $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

Доказательство. Выберем в $\mathcal{V}$ ортонормированный базис, соответственно представим требуемое равенство в матричном виде $(AX)^\dagger Y = X^\dagger(A^*Y)$ и преобразуем левую часть в $X^\dagger A^\dagger Y$. Искомый **сопряжённый оператор** A^* задаётся эрмитово сопряжённой матрицей $A^\dagger$. □

Столь простое выражение матрицы сопряжённого оператора наводит на мысль, что этот объект весьма полезен. Найдём теперь формулу для матрицы сопряжённого оператора в произвольном базисе с матрицей Грама G:

$$\left.\begin{aligned}(\mathsf{A}\mathbf{x},\mathbf{y}) &= (AX)^\dagger GY = X^\dagger A^\dagger GY\\ (\mathbf{x},\mathsf{A}^*\mathbf{y}) &= X^\dagger G(A^*Y) = X^\dagger GA^*Y\end{aligned}\right\} \implies A^* = G^{-1}A^\dagger G.$$

Упражнение 14. *Проверьте, что всегда* $(\mathsf{A}^*)^* = \mathsf{A}$.

Упражнение 15. *Проверьте, что* $\operatorname{Spec}\mathsf{A}^* = \overline{\operatorname{Spec}\mathsf{A}}$ *для всякого линейного оператора* A. *Иными словами,* $\lambda \in \operatorname{Spec}\mathsf{A} \iff \bar\lambda \in \operatorname{Spec} A^*$.

В случае одномерного эрмитова пространства $\mathcal{V}$ матрица оператора A состоит из одного числа. Тогда в ОНБ из единственного единичного вектора оператор A^* задаётся 1×1 матрицей $\mathsf{A}^\dagger$, где пропадает транспонирование и остаётся лишь комплексное сопряжение. Видно, как уместна привычка физиков обозначать его через z^* и насколько естественнее при этом выглядят формулы предыдущего упражнения: $\operatorname{Spec}(\mathsf{A}^*) = (\operatorname{Spec}\mathsf{A})^*$ и $\lambda^* \in \operatorname{Spec}\mathsf{A}^*$.

Упражнение 16. *Проверьте, что* $\mathsf{A}^*(\mathcal{L}) \subseteq \mathcal{L} \iff \mathsf{A}(\mathcal{L}^\perp) \subseteq \mathcal{L}^\perp$ *для всякого линейного оператора* A *и всякого подпространства* $\mathcal{L}$.

Спектральная теорема

Эрмитово сопряжённая матрица $A^\dagger$ возникает в записях условий, определяющих принадлежность матрицы A к одному из трёх замечательных классов: эрмитовы ($A^\dagger = A$), косоэрмитовы ($A^\dagger = -A$), унитарные ($A^\dagger A = E$). Естественно выделить аналогичными условиями три класса операторов: эрмитовы ($\mathsf{A}^* = \mathsf{A}$), косоэрмитовы ($\mathsf{A}^* = -\mathsf{A}$), унитарные ($\mathsf{A}^*\mathsf{A} = \mathsf{E}$). Есть и альтернативные названия: самосопряжённые, кососамосопряжённые, **изометричные**, а на эвклидовом пространстве — симметричные, кососимметричные, ортогональные.

При знакомстве с этими классами можно приэкономить немало времени, заметив условие, обобщающее все три указанных. Оператор A на эвклидовом или эрмитовом пространстве называют **нормальным**, если $\mathsf{A}^*\mathsf{A} = \mathsf{A}\mathsf{A}^*$. Ключом к пониманию строения нормальных операторов служит спектральная теорема.

Теорема. *Следующие условия на оператор* A *на эрмитовом пространстве* $\mathcal{V}$ *равносильны:*

(1) $\mathsf{A}^*\mathsf{A} = \mathsf{A}\mathsf{A}^*$;
(2) *пространство* $\mathcal{V}$ *обладает собственным ортонормированным базисом для* A.

Напомним, что второе условие означает, что $\mathcal{V}$ обладает ортонормированным базисом, в котором матрица оператора A диагональна. В эвклидовом пространстве эта теорема верна только для операторов с полностью вещественным спектром.

Нам пригодится лемма из предыдущей главы, поэтому повторим её.

Лемма. *Если операторы* A *и* B *коммутируют, то каждое собственное подпространство для* A *инвариантно под действием* B.

Доказательство теоремы. ($2 \Rightarrow 1$) Когда $G = E$, матрица $A^* = A^\dagger$ диагональна одновременно с матрицей A, и тогда они коммутируют.

($1 \Rightarrow 2$) Выберем $\lambda \in \operatorname{Spec}\mathsf{A}$. Если это единственное собственное число, то собственное подпространство $\mathcal{V}^\lambda$ есть всё $\mathcal{V}$ и достаточно взять любой его ОНБ. Иначе заметим, что $\mathcal{V}^\lambda$ по определению инвариантно под действием A, а по лемме ещё и под действием A^*. Тогда его ортогональное дополнение $(\mathcal{V}^\lambda)^\perp$, которое является суммой остальных собственных подпространств для оператора A, инвариантно под действием как A, так и A^* по упражнениям 16 и 14.

Применим индукцию по размерности пространства $\mathcal{V}$: считаем, что теорема доказана для операторов на пространствах меньшей размерности. Значит, ограничение A на $(\mathcal{V}^\lambda)^\perp$ обладает собственным ОНБ. Дополняя его любым ортонормированным базисом $\mathcal{V}^\lambda$, получаем искомый собственный ОНБ для исходного оператора. □

Спектральные портреты

Первым шагом к пониманию устройства конкретного нормального оператора является знание его спектра. Изображение спектра на комплексной плоскости называют **спектральным портретом**.

Нам уже известно, что спектр эрмитова оператора вещественен. Это можно получить и напрямую, подставляя один и тот же собственный вектор вместо $\mathbf{x}$ и $\mathbf{y}$ в условие $(\mathsf{A}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathsf{A}\mathbf{y})$. Заметим, что если оператор A эрмитов, то оператор $i\mathsf{A}$ — косоэрмитов. Отсюда следует, что спектр косоэрмитова оператора чисто мнимый. Далее, унитарный оператор сохраняет нормы всех векторов: $(\mathsf{A}\mathbf{x}, \mathsf{A}\mathbf{x}) = (\mathbf{x}, \mathbf{x})$. Подставляя в

Класс	*Оператор*	*Спектр*
эрмитовы	$\mathsf{A}^* = \mathsf{A}$	$\bar\lambda = \lambda$
косоэрмитовы	$\mathsf{A}^* = -\mathsf{A}$	$\bar\lambda = -\lambda$
унитарные	$\mathsf{A}^*\mathsf{A} = \mathsf{E}$	$\bar\lambda\lambda = 1$
нормальные	$\mathsf{A}^*\mathsf{A} = \mathsf{A}\mathsf{A}^*$	$\bar\lambda\lambda = \lambda\bar\lambda$

это условие собственный вектор оператора A с собственным числом λ, мы увидим, что $\bar\lambda\lambda = 1$.

Возникшая табличка ясно указывает на то, что понятие сопряжённого оператора является операторным аналогом комплексного сопряжения. При этом эрмитов оператор является аналогом вещественного числа, а унитарный — числа с единичным модулем. Основное применение эрмитовых операторов в физике связано с их ролью наблюдаемых величин в квантовой механике, где унитарные также определяют эволюцию квантовой системы во времени.

Канонические виды

Казалось бы, спектральная теорема для нормальных операторов полностью закрывает вопрос о каноническом виде такого оператора: он диагонален; на диагонали стоят собственные числа; получены простые ограничения на их значения для трёх важнейших классов. Это так и есть, но только в комплексном случае. В вещественном случае утверждение верно для симметричных операторов, но для кососимметричных и ортогональных диагонализация требует перехода к базису комплексного пространства. Тем не менее, из последней можно извлечь блочный диагональный вещественный вид, называемый каноническим.

Теорема. *Для всякого ортогонального оператора на эвклидовом пространстве найдётся вещественный ОНБ, в котором его матрица блочная диагональная с блоками*

$$[\pm 1], \quad \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}.$$

Теорема. *Для всякого кососимметричного оператора на эвклидовом пространстве найдётся вещественный ОНБ, в котором его матрица блочная диагональная с блоками*

$$[0], \quad \begin{bmatrix} 0 & \rho \\ -\rho & 0 \end{bmatrix}.$$

Доказательство обеих. Одномерные блоки содержат немногие возможные вещественные корни.

Невещественные корни характеристического многочлена вещественного оператора всегда появляются комплексно-сопряжёнными парами $\{\lambda, \bar{\lambda}\}$; более того, собственные векторы для них также можно выбрать комплексно-сопряжёнными парами $\{\mathbf{v}, \bar{\mathbf{v}}\}$, ибо $A\mathbf{v} = \lambda\mathbf{v}$ тут же даёт $A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$. Двумерные подпространства $\langle \mathbf{v}, \bar{\mathbf{v}} \rangle$ инвариантны. В комплексном диагональном виде всегда можно расположить элементы пар рядом и затем сгруппировать их, заметив подобие матриц

$$\begin{bmatrix} \rho\cos\varphi & \rho\sin\varphi \\ -\rho\sin\varphi & \rho\cos\varphi \end{bmatrix} = U^{-1} \begin{bmatrix} \rho e^{i\varphi} & 0 \\ 0 & \rho e^{-i\varphi} \end{bmatrix} U, \quad \text{где} \quad U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}.$$

В двумерном подпространстве $\langle \mathbf{v}, \bar{\mathbf{v}} \rangle$ унитарная матрица U осуществляет «поворот» от малого ОНБ $\{\mathbf{v} = \mathbf{a} + i\mathbf{b}, \bar{\mathbf{v}} = \mathbf{a} - i\mathbf{b}\}$ к малому вещественному ОНБ

$$\{\tfrac{1}{\sqrt{2}}(\mathbf{v} + \bar{\mathbf{v}}), \tfrac{1}{i\sqrt{2}}(\mathbf{v} - \bar{\mathbf{v}})\} = \{\sqrt{2}\,\mathbf{a}, \sqrt{2}\,\mathbf{b}\}.$$

Композиция всех таких переходов внутри пар в итоге приводит от большого комплексного ОНБ, гарантированного спектральной теоремой, к большому вещественному ОНБ. Найдя его, мы можем «забыть» о том, что по пути использовалось комплексное пространство. □

Упражнение 17. *Почему в этом доказательстве* $\mathbf{a} \perp \mathbf{b}$*?*

Коммутирующие самосопряжённые операторы

Теорема. *Следующие условия на семейство самосопряжённых операторов на эвклидовом или эрмитовом пространстве* $\mathcal{V}$ *равносильны:*

(1) *все операторы семейства коммутируют друг с другом;*
(2) *пространство* $\mathcal{V}$ *обладает ортонормированным базисом, собственным одновременно для всех операторов семейства.*

Иными словами, коммутирующие самосопряжённые операторы диагонализуемы одновременно. Это свойство имеет фундаментальное значение для квантовой механики, где оно означает, что коммутирующие наблюдаемые можно сколь угодно точно измерить одновременно, а некоммутирующие — нельзя.

Доказательство. (2 ⇒ 1) Все диагональные матрицы коммутируют.

(1 ⇒ 2) Адаптируем к текущей ситуации доказательство спектральной теоремы для нормальных операторов. Возьмём из семейства лю-

бой оператор A и выберем $\lambda \in \operatorname{Spec}\mathsf{A}$. Собственное подпространство $\mathcal{V}^\lambda$ для оператора A по определению инвариантно под действием A, а по лемме — даже под действием всех операторов семейства. Ортогональное дополнение $(\mathcal{V}^\lambda)^\perp$, которое является суммой остальных собственных подпространств для оператора A, по упражнению 16 инвариантно под действием $\mathsf{A}^* = \mathsf{A}$ и также всех операторов семейства.

При $\mathcal{V}^\lambda \neq \mathcal{V}$ задача распадается как минимум на две, поскольку в соответствии с инвариантным разложением $\mathcal{V} = \mathcal{V}^\lambda \oplus (\mathcal{V}^\lambda)^\perp$ все операторы семейства обретают блочно диагональные виды. Применяя индукцию по размерности пространства $\mathcal{V}$, составим искомый общий для семейства собственный ОНБ из найденных таких ОНБ слагаемых.

При $\mathcal{V}^\lambda = \mathcal{V}$ любой ОНБ является собственным для A, поэтому достаточно диагонализовать все остальные операторы семейства, для чего можно применить индукцию по их количеству. Даже если семейство бесконечно (возможность чего и отличает семейство от набора), это можно сделать аккуратно: количество существенных шагов ограничено сверху размерностью пространства $\mathcal{V}$. □

Указанный последним шаг доказательства обращает наше внимание на то обстоятельство, что общие собственные векторы, принадлежащие одному собственному значению оператора A из такого коммутирующего семейства, могут принадлежать разным собственным значениям другого оператора семейства. Это типичная ситуация в квантовой механике; например, так, в общих чертах, возникает деление электронных оболочек в атоме на орбитали.

8.5. Разложения операторов и матриц

Углубим обнаруженную выше аналогию строения операторов с комплексными числами. При работе с последними применяются две формы записи: $z = x+iy$ и $z = \rho e^{i\varphi}$. По сути, это разложения комплексного числа в сумму и произведение более простых объектов. При этом

$$x = \mathfrak{Re}\, z = \tfrac{1}{2}(z+\bar z), \quad y = \mathfrak{Im}\, z = \tfrac{1}{2i}(z-\bar z), \quad \rho = \sqrt{z\bar z}, \quad e^{i\varphi} = \rho^{-1} z.$$

Эрмитово разложение оператора

Операторная аналогия разложения в сумму проходит без запинок: выписываются налагаемые требования, а из них решением элементарных уравнений находятся искомые компоненты.

Теорема. *Каждый линейный оператор* A *на эвклидовом или эрмитовом пространстве однозначно представляется в виде*

$$\mathsf{A} = \mathfrak{Re}\,\mathsf{A} + i\,\mathfrak{Im}\,\mathsf{A},$$

где операторы $\mathfrak{Re}\,\mathsf{A}$ *и* $\mathfrak{Im}\,\mathsf{A}$*, называемые вещественной и мнимой частями оператора* A*, эрмитовы.*

Упражнение 18. *Эти компоненты находятся по формулам*

$$\mathfrak{Re}\,\mathsf{A} = \tfrac{1}{2}(\mathsf{A} + \mathsf{A}^*), \quad \mathfrak{Im}\,\mathsf{A} = \tfrac{1}{2i}(\mathsf{A} - \mathsf{A}^*).$$

Упражнение 19. *Нормальность оператора* A *равносильна коммутированию его вещественной и мнимой частей.*

Алгебраические структуры на операторах

Множество всех эрмитовых операторов на фиксированном линейном пространстве $\mathcal{V}$ само является вещественным линейным пространством, ибо линейные комбинации эрмитовых операторов с вещественными коэффициентами также эрмитовы, что прямо следует из определения сопряжённого оператора и аксиом скалярного произведения. На операторах есть также операция умножения (композиции), но легко убедиться, что композиция эрмитовых операторов редко эрмитова.

Упражнение 20. *Эрмитовость произведения* AB *эрмитовых операторов равносильна их коммутированию.*

С помощью эрмитова разложения выделим из композиции вещественную часть. Операция

$$\mathsf{A} \circ \mathsf{B} = \mathfrak{Re}(\mathsf{AB}) = \tfrac{1}{2}(\mathsf{AB} + \mathsf{BA})$$

на эрмитовых операторах называется **йордановым произведением**. Её стандартное обозначение кружочком $\circ$ не слишком удачно: его легко перепутать с композицией. Половина нужна, чтобы получить $\mathsf{E}\circ\mathsf{E} = \mathsf{E}$. Эта операция дистрибутивна относительно линейных комбинаций и потому заслуживает называться произведением. Йорданово произведение оказывается неассоциативным:

$$(\mathsf{A} \circ \mathsf{B}) \circ \mathsf{C} \neq \mathsf{A} \circ (\mathsf{B} \circ \mathsf{C}),$$

хотя равенство выполнено (упражнение) при $\mathsf{A} = \mathsf{C}$. Итак, алгебраическая структура пространства наблюдаемых не попадает в изучавшиеся прежде классы. Теперь это класс **йордановых алгебр**.

Выделяя из композиции мнимую часть, получим $\frac{1}{2i}(\mathsf{AB} - \mathsf{BA})$. Присутствие мнимой единицы не украшает такую операцию. Но тут можно

схитрить, перейдя от эрмитовых операторов к косоэрмитовым. Множество косоэрмитовых операторов тоже является вещественным линейным пространством, но на нём определена операция

$$[\mathsf{A},\mathsf{B}] = \mathsf{AB} - \mathsf{BA},$$

называемая **коммутатором** или **скобкой Ли**. Она дистрибутивна относительно линейных комбинаций и неассоциативна, а также лишена единицы, но обладает свойствами

$$[\mathsf{A},\mathsf{B}] + [\mathsf{B},\mathsf{A}] = 0;$$
$$[[\mathsf{A},\mathsf{B}],\mathsf{C}] + [[\mathsf{B},\mathsf{C}],\mathsf{A}] + [[\mathsf{C},\mathsf{A}],\mathsf{B}] = 0.$$

Такая структура, очень важная и в математике, и в физике, называется **алгеброй Ли**. Вы уже видели пример алгебры Ли: векторное произведение в $\mathbb{R}^3$.

Что до множества унитарных операторов — оно не является линейным пространством, не содержа ни нуля, ни линейных комбинаций. На нём есть только операция композиции, которая ассоциативна и притом допускает единичный элемент (тождественный оператор) и обратный к каждому элементу. Это пример другого важнейшего типа структур, называемых **группами**.

Полярное разложение оператора

Операторная аналогия разложения в произведение может иметь вид $\mathsf{A} = \mathsf{PU}$ либо $\mathsf{A} = \mathsf{UP}$, но ввиду априорной некоммутативности умножения операторов компоненты первой записи могут отличаться от одноимённых компонент второй. Оператор U заменяет фазовый множитель $e^{i\varphi}$, а значит, он должен быть унитарен. Оператор P заменяет (неотрицательный вещественный) модуль комплексного числа, а значит, он должен быть эрмитов и, более того, неотрицателен, то есть, все его собственные числа должны быть вещественны и неотрицательны.

Теорема. *Каждый линейный оператор* A *на эвклидовом или эрмитовом пространстве имеет как* **левое** *полярное разложение* $\mathsf{A} = \mathsf{PU}$*, так и* **правое** *полярное разложение* $\mathsf{A} = \mathsf{U}'\mathsf{P}'$ *с неотрицательными эрмитовыми операторами* P, P' *и унитарными операторами* U, U'.

Упражнение 21. *Нормальность оператора* A *равносильна коммутированию его эрмитового и унитарного сомножителей.*

Геометрически полярное разложение можно понимать как реализацию любого линейного оператора композицией двух операторов: уни-

тарный сомножитель, сохраняющий нормы всех векторов, осуществляет набор поворотов (в вещественном пространстве он может проявиться с отражением); эрмитов сомножитель осуществляет растяжения и сжатия вдоль собственных направлений.

Упражнение 22. *Для любого линейного оператора* A *на эвклидовом или эрмитовом пространстве операторы* AA^* *и* $\mathsf{A}^*\mathsf{A}$ *эрмитовы и, более того, неотрицательны.*

Из таких операторов можно извлекать неотрицательные квадратные корни.

Упражнение 23. *Найдите формулы для вычисления в ОНБ матриц операторов* $\sqrt{\mathsf{AA}^*}$ *и* $\sqrt{\mathsf{A}^*\mathsf{A}}$ *по матрице положительного эрмитова оператора* A*. Однозначен ли полученный ответ в случае вырожденного неотрицательного оператора?*

Тогда если $\mathsf{A} = \mathsf{PU}$, то $\mathsf{AA}^* = \mathsf{PUU}^*\mathsf{P}^* = \mathsf{PEP} = \mathsf{P}^2$. Поэтому, можно однозначно найти $\mathsf{P} = \sqrt{\mathsf{AA}^*}$, а также и $\mathsf{U} = \mathsf{P}^{-1}\mathsf{A}$, но только в том случае, когда оператор P обратим. Для правого полярного разложения $\mathsf{A} = \mathsf{UP}$ аналогично находим $\mathsf{P} = \sqrt{\mathsf{A}^*\mathsf{A}}$ и $\mathsf{U} = \mathsf{AP}^{-1}$.

Для вырожденных операторов A полученные формулы для U непригодны, но само полярное разложение существует и может быть найдено через сингулярное разложение.

Спектральное разложение нормального оператора

Возьмём теперь произвольный нормальный оператор A на эрмитовом пространстве $\mathcal{V}$. Спектральная теорема даёт прямое разложение

$$\mathcal{V} = \bigoplus_{\lambda \in \operatorname{Spec}\mathsf{A}} \mathcal{V}^\lambda$$

всего пространства на собственные подпространства, причём векторы из различных собственных подпространств ортогональны друг другу. Для каждого $\lambda \in \operatorname{Spec}\mathsf{A}$ обозначим через P_λ оператор проектирования на $\mathcal{V}^\lambda$:

$$\mathsf{P}_\lambda(\mathbf{x}) = \begin{cases} \mathbf{x} \text{ для } \mathbf{x} \in \mathcal{V}^\lambda, \\ \mathbf{0} \text{ для } \mathbf{x} \in \mathcal{V}^\mu \text{ при } \lambda \neq \mu. \end{cases}$$

В собственном (диагонализующем) ОНБ пространства $\mathcal{V}$ для оператора A матрица каждого P_λ также диагональна, причём по диагонали стоят единицы «напротив» базиса подпространства $\mathcal{V}^\lambda$, а все остальные компоненты нулевые. Таким образом, $(\mathsf{P}_\lambda)^* = \mathsf{P}_\lambda$, $\mathsf{P}_\lambda\mathsf{P}_\lambda = \mathsf{P}_\lambda$, а $\mathsf{P}_\lambda\mathsf{P}_\mu = \mathsf{O}$ при $\lambda \neq \mu$. Кроме того, $\mathsf{P}_\lambda\mathsf{A} = \mathsf{AP}_\lambda = \lambda\mathsf{P}_\lambda$.

Сумма указанных проекторов равна тождественному оператору:

$$\mathsf{E} = \sum_{\lambda \in \operatorname{Spec} \mathsf{A}} \mathsf{P}_\lambda.$$

Умножая на A, получаем спектральное разложение оператора A:

$$\mathsf{A} = \sum_{\lambda \in \operatorname{Spec} \mathsf{A}} \lambda \mathsf{P}_\lambda.$$

Спектральное разложение чаще всего формулируют для эрмитовых операторов, но для вывода достаточно нормальности. Эти же рассуждения пригодны в вещественном случае для всякого симметричного оператора, поскольку для него имеется собственный ОНБ.

Функциональное исчисление

Возьмём оператор A, обладающий спектральным разложением, и произвольную (но обычно достаточно хорошую) комплекснозначную функцию, определённую на всех $\lambda \in \operatorname{Spec} \mathsf{A}$. Определим оператор $f(\mathsf{A})$ равенством

$$f(\mathsf{A}) = \sum_{\lambda \in \operatorname{Spec} \mathsf{A}} f(\lambda) \mathsf{P}_\lambda.$$

Например, выше при обсуждении полярного разложения нам пришлось вычислять квадратные корни из неотрицательных операторов. Для вырожденных операторов обнаружилось препятствие — неоднозначность корня, что объясняется плохим поведением комплексной функции $f(z) = \sqrt{z}$ в точке $z = 0$, называемой точкой ветвления.

Операторные функции удовлетворяют всем привычным функциональным соотношениям, не требующим коммутирования, например:

$$f(\mathsf{A})g(\mathsf{A}) = (fg)(\mathsf{A}), \quad \exp(i\mathsf{A}) = \cos \mathsf{A} + i \sin \mathsf{A}.$$

Упражнение 24. *Вещественные функции от эрмитова оператора также дают эрмитовы операторы.*

Глава 9. ОСНОВЫ ОБРАБОТКИ ДАННЫХ

9.1. Линейная регрессия

Центрирование данных

Выборкой данных будем называть набор из m точек $\{\mathbf{x}_k\}$ эвклидова пространства $\mathbb{R}^n$. Во многих методах обработки данных первым шагом, а лучше даже сказать нулевым, находят выборочное среднее

$$\mathbf{a}_0 = \tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} \mathbf{x}_k.$$

Эта точка доставляет минимум сумме квадратов расстояний до точек выборки. Задачу минимизации и её решение записывают в виде

$$f(\mathbf{a}) = \sum_{1 \leqslant k \leqslant m} \|\mathbf{x}_k - \mathbf{a}\|^2 \to \min_{\mathbf{a}}, \qquad \mathbf{a}_0 = \arg\min_{\mathbf{a}} f(\mathbf{a}).$$

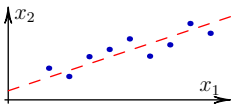
Новый символ в правой формуле означает аргумент минимума, то есть значение $\mathbf{a}$, при котором минимально значение целевой функции $f(\mathbf{a})$; он очень удобен и общепринят (в узких кругах). Показанное серым множество индексов суммирования далее вообще опускаем.

Упражнение 1. *Применив условия глобального минимума функции нескольких переменных, убедитесь, что получается именно среднее. По идее, достаточно изучить случаи двух или трёх точек.*

Часто перед дальнейшей обработкой данные смещаются переносом начала координат в найденное среднее, то есть заменой $\mathbf{x} \mapsto \mathbf{x} - \mathbf{a}_0$.

Линейная регрессия на плоскости

В простейшем варианте задачи линейной регрессии исходными данными является конечный набор $\{\mathbf{x}_k\}$ точек на плоскости, расположенных приблизительно на одной прямой, которую и требуется найти. А именно, среди всех прямых на плоскости нужно найти наиболее подходящую. Для сравнения кандидатов необходима числовая величина. Эту роль играет сумма квадратов расстояний до точек выборки, так что тут видна *прямая* аналогия с отысканием среднего.



Уравнение ищем в параметрическом виде

$$\mathbf{r}(t) = \mathbf{a}_0 + \mathbf{a}_1 t.$$

Логично взять в качестве начальной точки $\mathbf{a}_0$ выборочное среднее. Точнее тогда говорить не о поиске оптимальной прямой, а о поиске оптимальной пары из точки и содержащей её прямой. Определив центр, переносим туда начало координат, так что $\mathbf{a}_0$ исчезает из формул.

Направляющий вектор прямой будем искать среди направлений — векторов единичной длины. Каждый вектор $\mathbf{x}$ можно разложить на его проекцию $\mathbf{x}^{\parallel}$ на прямую и ортогональную составляющую

$$\mathbf{x}^{\perp} = \mathbf{x} - \mathbf{x}^{\parallel} = \mathbf{x} - (\mathbf{a}_1, \mathbf{x})\mathbf{a}_1,$$

дающую расстояние. Получаем задачу условной минимизации

$$\mathbf{a}_1 = \arg\min_{\|\mathbf{a}\|=1} \sum \|\mathbf{x}_k - (\mathbf{a}, \mathbf{x}_k)\mathbf{a}\|^2$$

с наложенным условием $\|\mathbf{a}\| = 1$. Каждое слагаемое есть квадратичная форма $(\mathbf{x}_k, \mathbf{x}_k) - (\mathbf{x}_k, \mathbf{a})^2$, а потому вся задача — это поиск минимума квадратичной формы на окружности. Его мы уже обсуждали как пример задачи, приводящей к собственным числам и векторам.

Линейная регрессия в любой размерности

Теперь исходные точки $\mathbf{x}_k$ лежат в эвклидовом пространстве $\mathbb{R}^3$. Найдём оптимальные точку $\mathcal{L}_0 = \{\mathbf{a}_0\}$ и прямую $\mathcal{L}_1 = \{\mathbf{a}_0 + \mathbf{a}_1 t_1\}$ как и выше, причём буквально по тем же формулам с тем лишь отличием, что уравнение $\|\mathbf{a}\| = 1$ теперь задаёт сферу и соответственно адаптируется задача на собственные числа и векторы.

Проектируя затем данные на плоскость, ортогональную $\mathcal{L}_1$, получим новый набор данных, состоящий из отклонений исходной выборки от линейного приближения. Его можно проанализировать таким же образом и найти новое направление $\mathbf{a}_2$, вероятно, второе по силе влияния на разброс исходных данных после $\mathbf{a}_1$. Плоскость

$$\mathcal{L}_2 = \{\mathbf{a}_0 + \mathbf{a}_1 t_1 + \mathbf{a}_2 t_2\}$$

будет наилучшим приближением к выборке в том же смысле, в каком выше была прямая: в эту плоскость вложена оптимальная прямая, содержащая оптимальную точку. При этом $\mathbf{a}_2 \perp \mathbf{a}_1$.

Подобную цепочку $\mathcal{L}_0 \subset \mathcal{L}_1 \subset \mathcal{L}_2$ вложенных линейных многообразий в геометрии называют **флагом** без требования оптимальности и вообще обычно вне связи с прикладной задачей линейной регрессии.

При размерности данных $n > 3$ построение флага оптимальных линейных многообразий можно продолжать, пока в проекциях остаётся ещё не учтённый разброс данных, либо можно остановиться, списав оставшееся на шум. Последнее весьма актуально, когда размерность велика и поставлена задача её снижения.

Зародыш метода наименьших квадратов

В случае линейной регрессии на плоскости можно получить явную формулу для искомого направления иным способом, в свою очередь представляющим из себя простейший случай другого важного метода. Вместо $\mathbf{x}_k$, здесь удобно обозначать данные на плоскости через (x_i, y_i): из двух замеряемых переменных, считаем x независимой, а y линейно зависящей от x. Зависимость будем искать в совсем школьном виде $y = kx + b$. Зная значения коэффициентов, мы могли бы выделить из выборки случайные ошибки $\varepsilon_i = y_i - (kx_i + b)$.

С ростом числа m точек в выборке (замеров) ошибки, если они действительно случайные, всегда уменьшаются усреднением вследствие одной из основных теорем теории вероятностей, называемой законом больших чисел. В таком случае уравнение

$$\sum(kx_i + b) = b \cdot m + k \cdot \sum x_i = \sum y_i$$

можно использовать для оценки коэффициентов.

Искомых коэффициентов, однако, два, и нужно второе уравнение. В качестве него можно взять

$$\sum x_i(kx_i + b) = b \cdot \sum x_i + k \cdot \sum x_i^2 = \sum x_i y_i.$$

Оно берётся из того же соотношения с ошибками, умноженного на x_i. Из этой пары уравнений можно выразить по правилу Крамера оценки для b и k. Их называют оценками по **методу наименьших квадратов**, что опять же связано с минимизацией суммы квадратов расстояний.

Особенно часто встречаются пропорциональные зависимости, где $b = 0$. Когда это известно, оценку k по методу наименьших квадратов делают из второго уравнения. Позже она нам пригодится, но перед тем будет переписана далее в этом разделе.

Матричное представление линейной регрессии

Усложним предыдущую ситуацию, сделав независимую переменную векторной: $X \in \mathbb{R}^n$. Заодно тут принято избавляться от константы b, отводя для неё фиктивную «ось», по которой замеры равны единице

всегда. Поскольку опять анонсировано изучение линейной регрессии, искомую зависимость предполагаем в виде скалярного произведения $y = (X, P) = X^\top P$, и требуется оценить вектор параметров P. Данные по-прежнему содержат случайные ошибки $\varepsilon_i = y_i - (X_i, P)$, усредняющиеся при большом количестве замеров.

Здесь напрашивается введение матричных обозначений. Векторы данных X_i сделаем строками $m \times n$ матрицы A. Это константы во всём процессе, хотя он и призван объяснить наблюдения, в которых они появились как значения переменных. То же самое относится к столбцу Y, составленному из значений y_i.

Нужная оценка по методу наименьших квадратов будет приближённым решением системы линейных уравнений $AP = Y$. В ней n неизвестных и m уравнений, причём вся деятельность ведётся в предположении наличия большой выборки. Поэтому $m \gg n$, система сильно **переопределена** и точного решения как правило не имеет. Приближённым решением естественно назвать значение P, доставляющее минимум **невязке** $\|Y - AP\|$. Итак,

$$\hat{P} = \arg\min_P \|AP - Y\|^2 = \arg\min_P (AP - Y)^\top (AP - Y).$$

В точке минимума равен нулю градиент целевой функции. Можно увидеть интересные вещи, наивно написав этот градиент по правилу дифференцирования произведения как

$$\nabla_P \big((AP - Y)^\top (AP - Y)\big) = A^\top (AP - Y) + (AP - Y)^\top A = S + S^\top.$$

Слагаемые здесь различаются лишь общим транспонированием, почему и введена буква S. Если подумать, что из условия $S + S^\top = 0$ как-то можно получить $S = 0$, то мы придём к **нормальным уравнениям**

$$A^\top AP = A^\top Y,$$

а в скалярном случае это превращается как раз в уравнения на b и k, выписанные выше. Матрица $A^\top A$ скорее всего невырождена, особенно при $m \gg n$, и точка минимума определяется по формуле

$$\hat{P} = (A^\top A)^{-1} A^\top Y.$$

Вычисления по ней эффективны благодаря положительной определённости и симметричности матрицы $A^\top A$.

Увы, такой ход мысли грубо некорректен. Нужно заметить, что S является столбцом, а $S^\top$ строкой, и складывать их никак нельзя! Тем не менее, желаемое заключение $S = 0$ можно спасти, развернув вы-

численение покомпонентно; тогда удаётся совершить перестановку индексов, фактически транспонирующую одно из слагаемых.

С другой стороны, столбец AP есть линейная комбинация столбцов матрицы A, поэтому геометрически минимизация невязки означает поиск ближайшей к Y точки в пространстве столбцов A, то есть проекции Y на это подпространство в $\mathbb{R}^m$. К полученной здесь формуле для этой проекции полезно вернуться после углубления теории в следующих разделах.

Упражнение 2. *Проведите обрисованное вычисление с индексами.*

9.2. Дисперсия, ковариация, корреляция

Разброс данных и дисперсия

Помимо выборочного среднего, в статистике применяются другие числовые характеристики выборки, описывающие разброс данных вокруг среднего. Простейшую и важнейшую из них называют **дисперсией**. Для скалярной (числовой) выборки $\{x_k\}$ её вычисляют по формуле

$$\tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} x_k^2 - \Big(\tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} x_k \Big)^2,$$

то есть это среднее от квадратов минус квадрат среднего. Больший разброс даёт бóльшую дисперсию.

Упражнение 3. *Проверьте, что дисперсия выборки не меняется при её смещении, и в частности центрировании.*

Попробуем воспользоваться символикой теории вероятностей и математической статистики, не вдаваясь в формальные детали, необходимые для её введения. Выборку обозначим через X, а через X^2 — набор из квадратов $\{x_k^2\}$. Выборочное среднее обозначают через $\mathrm{M}[X]$ в советской литературе (математическое ожидание) и через $\mathrm{E}[X]$ в иностранной (expectation), при этом ещё варьируя вид скобок и шрифты. Дисперсию (variance) обозначают соответственно:

$$\mathrm{D}[X] = \mathrm{M}[X^2] - \mathrm{M}[X]^2, \qquad \mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2.$$

По упражнению, формулу можно переписать в виде

$$\mathrm{D}[X] = \mathrm{M}\big[(X - \mathrm{M}[X])^2\big].$$

У центрированной выборки остаётся лишь среднее от квадратов.

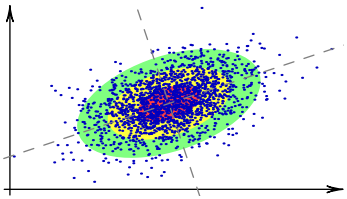
Для векторных выборок в $\mathbb{R}^n$ есть разные обобщения. К ним теперь и переходим.

Дисперсия вдоль направления

Зафиксировав направление $\mathbf{a} \in \mathbb{R}^2$, по двумерной выборке $X = \{\mathbf{x}_k\}$ вычислим скалярные произведения и напишем числовую выборку

$$(X, \mathbf{a}) = \{(\mathbf{x}_k, \mathbf{a})\}.$$

Дисперсия вдоль направления $\mathbf{a}$, определяемая как $\mathrm{D}[(X, \mathbf{a})]$, численно характеризует разброс данных вдоль направления $\mathbf{a}$.



Типичная двумерная выборка выглядит как пятно эллиптической формы. Направления, вдоль которых дисперсия экстремальна, являются главными осями приближающего эллипса. Задача поиска главных направлений математически совпадает с задачей линейной регрессии: минимизируемые величины отличаются постоянной, а зависимость от направления вся заключена в усреднении слагаемых $(\mathbf{x}_k, \mathbf{a})^2$. Дисперсия вдоль главных направлений $\mathbf{a}_1$ и $\mathbf{a}_2$ характеризует силу влияния скрытых факторов, отыскиваемых этим методом.

Разумеется, всё сказанное тут работает в любой размерности.

Ковариация и корреляция

Здесь возьмём центрированную векторную выборку $\{(x_k, y_k)\}$ в $\mathbb{R}^2$. Можно изучать дисперсию выборки вдоль каждой координаты:

$$\mathrm{Var}[X] = \mathrm{E}[X^2] = \tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} x_k^2, \qquad \mathrm{Var}[Y] = \mathrm{E}[Y^2] = \tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} y_k^2,$$

и это лишь частный случай дисперсии вдоль направления. Однако, глядя на эти выражения, нетрудно догадаться, что можно обобщить дисперсию иначе, рассмотрев

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] = \tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} x_k y_k.$$

Эта величина сопоставляет между собой распределение значений по двум координатам и называется **ковариацией**. Без условия центрированности работает общая формула

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y].$$

В отличие от дисперсии, которая всегда неотрицательна фактически по определению, ковариация бывает любого знака; если при увеличении значения x наблюдается тенденция к уменьшению значения y, то ковариация между ними отрицательна.

Теперь представим те же данные двумя векторами X, Y в пространстве столбцов $\mathbb{R}^m$. Тогда дисперсии и ковариация становятся скалярными произведениями с точностью до общего множителя:

$$\operatorname{Var}[X] = \tfrac{1}{m} X^\top X, \qquad \operatorname{Cov}[X,Y] = \tfrac{1}{m} X^\top Y, \qquad \operatorname{Var}[Y] = \tfrac{1}{m} Y^\top Y.$$

Следовательно, они удовлетворяют неравенству Коши:

$$\operatorname{Cov}[X,Y]^2 \leqslant \operatorname{Var}[X] \cdot \operatorname{Var}[Y].$$

Пирсон ввёл в обиход очень удобный **коэффициент корреляции**

$$R_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\operatorname{Var}[X]}\sqrt{\operatorname{Var}[Y]}} = \frac{X^\top Y}{\|X\| \cdot \|Y\|};$$

его значения безразмерны и, в силу предыдущего неравенства, всегда лежат в интервале $[-1, 1]$. В терминах эвклидова пространства $\mathbb{R}^m$, это косинус угла между векторами X и Y.

Положительная корреляция двух признаков часто свидетельствует в пользу наличия у них общей причины, а отрицательная — взаимоисключающих причин. Независимые признаки имеют нулевую корреляцию, но не коррелирующие признаки не обязательно независимы.

Другая безразмерная комбинация ковариации и дисперсии встретилась нам выше в методе наименьших квадратов: оценка углового коэффициента искомой прямой регрессии там даётся формулой

$$\hat{k} = \frac{\operatorname{Cov}[X,Y]}{\operatorname{Var}[X]} = \frac{X^\top Y}{X^\top X}.$$

Матрица ковариаций

Возьмём центрированную векторную выборку $\{X_k\} \subset \mathbb{R}^n$, помня, что выше собирали эти векторы как строки в матрицу A. Каждый её столбец тогда содержит данные всех замеров одного датчика. Соответственно, коллекцию ковариаций

$$C_{ij} = \tfrac{1}{m} \sum_{1 \leqslant k \leqslant m} a_{ki} a_{kj},$$

по всевозможным парам координат, вместе с дисперсиями вдоль координатных осей, можно собрать в матрицу C. Через матрицу A она выражается очень простым образом: $C = \frac{1}{m} A^\top A$. При этом главные

оси эллипсоида, приближающего данные, являются собственными векторами матрицы C. Переход в новую систему координат, где главные оси будут координатными, диагонализует матрицу ковариаций: все ковариации обнуляются, а остаются только дисперсии вдоль новых координатных осей. Тем самым выявляется набор скрытых факторов, влияющих на разброс данных независимо друг от друга.

Геометрический взгляд на всю эту деятельность получит дальнейшее развитие с изучением сингулярного разложения.

9.3. Сингулярное разложение

Предварительный эллипс

Лекция 7 (27.03.20)

Пример. Рассмотрим оператор A на обычной эвклидовой плоскости $\mathbb{R}^2$, заданный в стандартном ОНБ матрицей

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Его действие можно описать словом «перекашивание». Разумеется, этот оператор не нормален, ведь A есть жорданова клетка.

Изучим, как A искажает длины (нормы) векторов. Уже на первой паре рисунков разнесём векторы «до» и «после» в разные пространства, воспринимая A не как оператор, а как отображение из одной плоскости в другую. Ввиду линейности, достаточно понять, как выглядит образ единичной окружности. Это будет эллипс.

Однако удобнее оказывается не образ, а прообраз единичной окружности, заданной уравнением $\|\mathbf{y}\| = 1$. Решаем уравнение $\|\mathsf{A}\mathbf{x}\| = 1$, или в координатах $X^\top A^\top A X = 1$ с матрицей

$$A^\top A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Это уравнение, $x_1^2+2x_1x_2+2x_2^2=1$, задаёт эллипс. Полуоси направлены по собственным векторам $\mathbf{v}_1$, $\mathbf{v}_2$ оператора $\mathsf{A}^*\mathsf{A}$ и равны по длине соответствующим собственным числам в степени $-\frac{1}{2}$, ведь именно такой показатель возникает при приведении уравнения $\lambda_1\tilde{x}_1^2+\lambda_2\tilde{x}_2^2=1$ эллипса в главных осях к каноническому виду $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$.

Видно также, что отображение A равно композиции трёх более простых:

(1) поворот от собственного ОНБ $\{\mathbf{v}_1,\mathbf{v}_2\}$ для $A^\top A$ к стандартному базису;
(2) два растяжения вдоль координатных осей;
(3) поворот обратно к собственному ОНБ.

Повороты это в любом случае операторы, а смена пространства (отображение) совершается только на втором шаге. Завершающий поворот отличается величиной угла от начинающего, что тоже видно на рисунке, но они и не должны быть связаны, коли происходят в разных пространствах!

Сингулярные числа

Определение. Неотрицательный квадратный корень из собственного числа оператора $\mathsf{A}^*\mathsf{A}$ называют **сингулярным числом** оператора A, и аналогично для матриц: $\sigma_k(\mathsf{A})=\sqrt{\lambda_k(\mathsf{A}^*\mathsf{A})}$.

Обычно список сингулярных чисел конкретного оператора или матрицы упорядочивают по невозрастанию: $\sigma_1\geqslant\sigma_2\geqslant\ldots\geqslant\sigma_n\geqslant 0$. Их геометрический смысл ясен из обобщения рассмотренного примера на произвольную размерность: прообраз многомерной сферы есть не менее многомерный эллипсоид (либо «эллипсоидный цилиндр» когда оператор A вырожден). Сингулярные числа важны в разнообразных приложениях линейной алгебры. На практике, собственные чис-

ла во многом проигрывают сингулярным; достаточно упомянуть общность конструкции и зависимость от всяческих ошибок при расчётах (погрешности измерения, компьютерное округление).

Определение сингулярных чисел имеет смысл не только для оператора, но и для любого линейного отображения $\mathsf{A}\colon \mathcal{V} \to \mathcal{U}$ эвклидовых или эрмитовых пространств. Необходимое для того **сопряжённое отображение** $\mathsf{A}^*\colon \mathcal{U} \to \mathcal{V}$ вводится почти той же формулой $(\mathsf{A}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathsf{A}^*\mathbf{y})$, что и в частном случае оператора, хотя тут слева скалярное произведение в $\mathcal{U}$, а справа в $\mathcal{V}$. Множителями в матрицу сопряжённого отображения входят две матрицы Грама разных базисов разных пространств, никак между собой не связанные, но иных отличий от случая операторов нет.

Упражнение 4. *Для любой матрицы A, не обязательно квадратной, равны ранги матриц A, $A^\dagger A$ и $AA^\dagger$.*

Упражнение 5. *Для любого линейного отображения A проверьте равенства $\operatorname{Ker}\mathsf{A}^* = (\operatorname{Im}\mathsf{A})^\perp$ и $\operatorname{Im}\mathsf{A}^* = (\operatorname{Ker}\mathsf{A})^\perp$.*

Формулировка сингулярного разложения

Теорема. *Каждая комплексная матрица A представляется в виде $A = U\Sigma V^\dagger$, где матрицы U и V унитарны, а диагональная матрица Σ составлена из сингулярных чисел $\sigma_1(A), \ldots, \sigma_n(A)$.*

Для вещественной матрицы A вместо унитарных сомножителей такого разложения можно найти ортогональные.

Построение обобщает разложение оператора в разобранном выше двумерном примере. Подчеркнём, что сингулярным разложением обладают все прямоугольные матрицы, а не только квадратные. Отличие общего случая лишь в том, что блочная матрица

$$\Sigma = \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

не обязательно квадратна, а квадратен только её невырожденный блок

$$\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r),$$

составленный из положительных сингулярных чисел. В зависимости от соотношения размеров и ранга r исходной $m \times n$ матрицы A два или все три указанных нулевых блока могут отсутствовать.

Теперь можно найти обещанные формулы для полярных разложений произвольного оператора, включая вырожденные. Зная сингуляр-

ное разложение $A = U\Sigma V^\dagger$, положим $W = UV^\dagger$. Тогда

$$A = (U\Sigma U^\dagger)W = W(V\Sigma V^\dagger),$$

причём матрица W унитарна, а матрицы $U\Sigma U^\dagger$ и $V\Sigma V^\dagger$ неотрицательные эрмитовы.

Структура линейного отображения

Прежде чем заниматься доказательством сингулярного разложения, полезно вспомнить теорему о матрице линейного отображения, ранее показавшую возможность перейти к простому виду сменой базисов. Тогда мы не имели понятия длины, а потому смены базисов были свободнее, матрицы перехода были любые невырожденные, а на месте блока Σ_r в итоге оставалась единичная матрица.

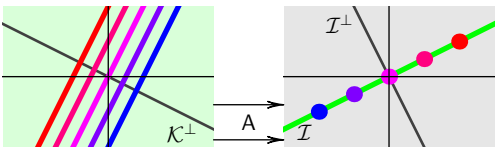
Следствие. *Для всякого отображения* $\mathsf{A}\colon \mathcal{V} \to \mathcal{U}$ *эвклидовых или эрмитовых пространств в них найдутся такие ОНБ, что матрица отображения принимает указанный выше вид* Σ.

Доказательство. Выбрав любые ОНБ, можно работать с произвольными пространствами как со стандартными. Затем нужно перейти к ОНБ $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ и $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, найденным как в показанном ниже построении сингулярного разложения матрицы. □

На пути к обещанному доказательству выделим одно промежуточное обстоятельство. Для этого опять вернёмся к теореме о структуре линейного отображения $\mathsf{A}\colon \mathcal{V} \to \mathcal{U}$. Там мы видели два разложения в прямые суммы. В нынешних буквах это $\mathcal{V} = \mathcal{K}' \oplus \mathcal{K}$ и $\mathcal{U} = \mathcal{I} \oplus \mathcal{I}'$, где $\mathcal{K} = \operatorname{Ker}\mathsf{A}$ и $\mathcal{I} = \operatorname{Im}\mathsf{A}$, причём A отображает $\mathcal{K}'$ на $\mathcal{I}$ взаимно однозначно. Однако теперь целесообразно брать не произвольные дополнения к подпространствам, а ортогональные: $\mathcal{K}' = \mathcal{K}^\perp$ и $\mathcal{I}' = \mathcal{I}^\perp$. Эти обозначения сохраним до конца главы и будем иногда использовать:

$$\mathcal{V} = \mathcal{K}^\perp \oplus \mathcal{K}, \quad \mathcal{U} = \mathcal{I} \oplus \mathcal{I}^\perp.$$

На рисунке вся зеленоватая плоскость переходит в зелёную прямую, а остальные точки сероватой плоскости лежат вне образа. Каждая из цветных прямых переходит в соответствующую жирную точку.



Матрица оператора A в любых базисах пространств $\mathcal{V}$ и $\mathcal{U}$, подчинённых обсуждаемым разложениям, имеет блочный вид

$$\Sigma = \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right],$$

причём блок Σ_r невырожден. Чтобы перейти отсюда к сингулярному разложению, остаётся лишь согласовать выбор базисов в $\mathcal{K}'$ и $\mathcal{I}$.

Доказательство сингулярного разложения

Лемма. *Для всякого линейного отображения $\mathsf{A}\colon \mathcal{V} \to \mathcal{U}$ эвклидовых или эрмитовых пространств найдётся такой ОНБ $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, что $(\mathsf{A}\mathbf{v}_i, \mathsf{A}\mathbf{v}_j) = 0$ при $i \neq j$.*

Доказательство. Искомым базисом служит собственный ОНБ неотрицательного эрмитова оператора $\mathsf{A}^*\mathsf{A}\colon \mathcal{V} \to \mathcal{V}$. Действительно, тогда

$$(\mathsf{A}\mathbf{v}_i, \mathsf{A}\mathbf{v}_j) = (\mathbf{v}_i, \mathsf{A}^*\mathsf{A}\mathbf{v}_j) = \sigma_j^2(\mathbf{v}_i, \mathbf{v}_j) = \begin{cases} \sigma_i^2 & \text{при } i = j, \\ 0 & \text{при } i \neq j, \end{cases}$$

где σ_i^2 есть собственное число оператора $\mathsf{A}^*\mathsf{A}$, причём оно вещественное и даже неотрицательное. Нет ничего страшного в том, что в общем случае некоторые образы $\mathsf{A}\mathbf{v}_i$ могут быть нулевыми. □

Доказательство теоремы. Применим лемму к линейному отображению $\mathsf{A}\colon \mathbb{C}^n \to \mathbb{C}^m$ стандартных эрмитовых пространств, представленному в стандартных ОНБ матрицей A. При этом упорядочим ОНБ $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ по невозрастанию длин образов $\mathsf{A}\mathbf{v}_i$, равных σ_i по доказательству леммы. Ненулевые образы линейно независимы ввиду их ортогональности. Нормируем их и дополним полученную ортонормированную систему до ОНБ $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ пространства $\mathbb{C}^m$. Для связи с предшествующим замечанием отметим, что

$$\langle \mathbf{v}_1, \ldots, \mathbf{v}_r \rangle = \mathcal{K}^\perp, \quad \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \mathcal{I}.$$

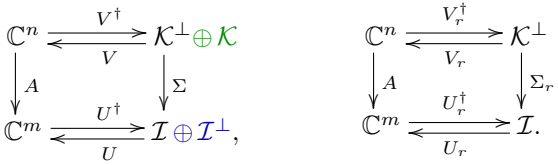
Унитарный оператор, переводящий стандартный ОНБ $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ в ОНБ $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, задаётся унитарной матрицей; это и будет V. Унитарный оператор, переводящий стандартный ОНБ $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$ в ОНБ $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, задаётся унитарной матрицей; это и будет U. Теперь в три шага получаем

$$\mathbf{e}_i \mapsto \mathsf{V}\mathbf{e}_i = \mathbf{v}_i \mapsto \mathsf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i \mapsto \mathsf{U}^{-1}(\sigma_i\mathbf{u}_i) = \sigma_i\mathbf{e}_i$$

для всех $1 \leqslant i \leqslant n$, причём $\sigma_i = 0$ при $i > r = \operatorname{rk} A$. Значит, $U^\dagger A\, V = \Sigma$ и $A = U\Sigma V^\dagger$. □

Сокращённое сингулярное разложение

Построение в доказательстве сингулярного разложения $A = U\Sigma V^\dagger$ перепишем в виде диаграммы отображений, а затем сократим её:



На левой диаграмме горизонтальные стрелки обратимы: это смены ОНБ, и они представлены умножениями на унитарные матрицы. На правой диаграмме стрелки вправо осуществляют проекции, а стрелки влево вложения. Базисы ядра $\mathcal{K}$ и дополнения $\mathcal{I}^\perp$ к образу в деле не участвуют; соответствующие им части матриц $V^\dagger$ и U гасятся нулевыми блоками матрицы Σ. Достаточно взять из U левый $m \times r$ блок U_r, а из $V^\dagger$ — верхний $r \times n$ блок $V_r^\dagger$, чтобы всё равно получить $A = U_r\Sigma_r V_r^\dagger$. Когда $r \ll n$ либо $r \ll m$, что типично в обработке данных, такое сокращение разложения даёт значительную экономию.

Упражнение 6. *Постройте полярное разложение прямоугольной матрицы в случаях $m < n$ и $m > n$.*

9.4. Ранговое приближение

Утверждения этого раздела теснее связывают сингулярное разложение с прикладными методами обработки данных из начала главы.

Ранговое разложение и разложение Шмидта

Про $m \times n$ матрицу A говорят, что она **полного ранга**, если её ранг максимальный из возможных, то есть $\operatorname{rk} A = \min\{m, n\}$. Например, в сокращённом сингулярном разложении $A = U_r\Sigma_r V_r^\dagger$ все три сомножителя полного ранга.

Следствие. *Всякую $m \times n$ матрицу ранга $r > 0$ можно представить как произведение $m \times r$ матрицы и $r \times n$ матрицы полных рангов.*

Доказательство. Матрицы $F = U_r\Sigma_r$ и $G = V_r^\dagger$ удовлетворяют требованиям: $A = FG$, причём F имеет полный ранг по столбцам, а G по строкам. С тем же успехом можно внести Σ_r в матрицу G вместо F. □

Можно найти **ранговое разложение** другим способом. Приведя матрицу A к ступенчатому виду A' элементарными преобразованиями строк, узнаём, какие столбцы в ней главные. Из главных столбцов A составим матрицу F, а из ненулевых строк A' составим матрицу G, сохраняя порядок в обоих случаях. Определение ступенчатого вида, данное в первом семестре, гарантирует его единственность; если пользоваться именно им, то $A = FG$ и оба сомножителя получились полного ранга.

Упражнение 7. *Докажите, что $A = FG$ в этом способе.*

Пример. У нас многократно возникало произведение строки на столбец, но не наоборот. Однако произведение столбца на строку осмысленно, причём без требований к размерностям:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}.$$

Ранг получившейся матрицы равен 1.

Упражнение 8. *Из наборов столбцов $X_k \in \mathbb{R}^m$ и $Y_k \in \mathbb{R}^n$ будем строить $m \times n$ матрицу $A = X_1 Y_1^\dagger + \ldots + X_r Y_r^\dagger$. Докажите, что:*

(1) *тогда* $\operatorname{rk} A \leqslant r$;
(2) *всякую матрицу ранга r можно представить в таком виде.*

Ранговое приближение Шмидта

Распишем сингулярное разложение $A = U\Sigma V^\dagger$ в духе предыдущего упражнения, разобрав унитарные матрицы U и V на столбцы:

$$A = \sigma_1 U_1 V_1^\dagger + \ldots + \sigma_r U_r V_r^\dagger.$$

Здесь $\|U_k\| = \|V_k\| = 1$ ввиду унитарности. Поскольку сингулярные числа всегда расположены в порядке убывания, может случиться, что последние слагаемые очень малы по сравнению с первыми, и тогда для грамотно подобранного $s < r$, либо даже $s \ll r$, матрица

$$A_s = \sigma_1 U_1 V_1^\dagger + \ldots + \sigma_s U_s V_s^\dagger$$

будет очень мало отличаться от исходной матрицы A.

Ниже теорема Шмидта утверждает, что это приближение является наилучшим. Тем самым она открывает путь к бесчисленным приложениям сингулярного разложения в самых разных областях науки и высоких технологий, включая далёкие друг от друга настолько, что общность применяемых в них математических методов непросто распознать за пеленой обозначений, терминологии и имён.

Определение. **Нормой Фробениуса** матрицы A называют неотрицательную величину $\|A\|$, определённую равносильными формулами

$$\|A\|^2 = \sum_{i,j} |a_{ij}|^2 = \operatorname{tr}(A^\dagger A) = \sum_k \sigma_k^2.$$

Упражнение 9. *Установите равносильность.*

Теорема (Schmidt)**.** *Для всякой $m \times n$ матрицы B_s ранга $s \leqslant r$ верно неравенство $\|A - B_s\| \geqslant \|A - A_s\|$.*

В двух следующих блоках обратимся к приложениям.

Метод главных компонент

Исходно имеется матрица данных A, каждая строка которой представляет собой запись результата эксперимента, например, состоящую из данных n датчиков. Данные предварительно обработаны так, что сумма в каждом столбце нулевая (столбцы центрированы). Полезную информацию о распределении данных извлекают из матрицы ковариаций $A^\dagger A$; хотя данные замеров и дальнейшие вычисления в этом методе обычно вещественные, мы сохраним комплексные обозначения.

При $n = 2$ можно изобразить данные набором точек на плоскости. Как уже говорилось, часто получается пятно эллиптической формы, что побудило Пирсона искать самый подходящий эллипс. Его главные направления и длины осей определяются диагонализацией эрмитовой матрицы $A^\dagger A$. Отсюда пошло название: метод главных компонент.

Сингулярное разложение $A = U\Sigma V^\dagger$ даёт $A^\dagger A = V\Sigma^\dagger \Sigma V^\dagger$, причём матрица $\Sigma^\dagger\Sigma$ диагональна. Главные направления, называемые **нагрузками** — собственные векторы матрицы $A^\dagger A$, то есть столбцы матрицы $V^\dagger$. Проекции столбцов данных на столбцы нагрузок называют **счетами**. Переход от исходного базиса (сенсоров) к базису из нагрузок выявляет независимые друг от друга факторы, в исходных данных как правило скрытые, а соответствующие сингулярные числа показывают силу влияния каждого фактора. Затем можно отбросить малозначимые факторы, применив приближение Шмидта.

Восстановление параметров модели

Одна из типичных задач прикладной математики требует определить неизвестные параметры модели по экспериментальным данным. В линейной модели связь параметров $X \in \mathbb{R}^n$ и данных $B \in \mathbb{R}^m$ записывается как уравнение $AX = B$ с известной матрицей моделирова-

ния A. Тогда параметры можно оценить по формуле $X = A^\natural B$. Псевдообратная матрица $A^\natural$ подробнее изучается в следующем разделе, но понять её устройство можно, и даже легче, из этого примера.

Воспользуемся сингулярным разложением матрицы A и перепишем уравнение $U\Sigma V^\dagger X = B$ как $\Sigma V^\dagger X = U^\dagger B$. Унитарные матрицы U и V дают переходы от стандартных ОНБ пространства данных и пространства параметров к тем ОНБ, в которых связь выражается проще всего: как бы диагональной матрицей Σ, а в координатах

$$b'_1 = \sigma_1 x'_1, \dots, b'_n = \sigma_n x'_n, \quad b'_{n+1} = 0, \dots, b'_m = 0.$$

Здесь комбинированный параметр x'_k влияет только на комбинированный элемент данных b'_k, причём с весом σ_k. Веса определяют чувствительность данных к модельным параметрам. При $m > n$ также выявляются лишние данные, от параметров модели не зависящие, а потому не помогающие их определить. Кроме того, на практике наличие разнообразных шумов приводит к очень малым сингулярным числам, а потому очень большим элементам σ_k^{-1} матрицы $\Sigma^\natural$. Соответствующие комбинированные параметры модели нельзя устойчиво восстановить. Для подавления шумов исходную матрицу A заменяют её ранговым приближением, округлив малые сингулярные числа до нулей.

Итерационный алгоритм

Слагаемые рангового приближения матрицы A естественно вычислять одно за другим до достижения требуемой точности, часто определяемой по отношению следующего слагаемого к уже накопленной сумме, либо по величине неучтённого остатка. Каждое слагаемое можно вычислять в цикле приближений, выполняемых по методу наименьших квадратов.

Покажем итерации, дающие ведущее слагаемое $\sigma_1 U_1 V_1{}^\top$, опуская индекс. Начнём поиск минимума отклонения

$$f(U,V) = \|A - UV^\top\|^2 = \sum_{i,j} (a_{ij} - u_i v_j)^2$$

с произвольного столбца U. В цикле сперва возьмём правый вектор

$$V = \arg\min_{Y\in\mathbb{R}^n} f(U,Y) \implies V = \frac{A^\top U}{U^\top U}$$

и затем обновим левый вектор, полагая

$$U = \arg\min_{X\in\mathbb{R}^n} f(X,V) \implies U = \frac{AV}{V^\top V}.$$

Процесс повторяем, пока не выполнится выбранный критерий остановки. На последнем шаге найдём $\sigma = \|U\| \cdot \|V\|$, а векторы нормируем.

Далее, разность $A - \sigma UV^\top$ примем за новую матрицу A и для неё таким же циклом найдём следующее ведущее слагаемое...

9.5. Псевдообратная матрица

Геометрическое построение псевдообратной

Отложим само определение псевдообратной и начнём с геометрического построения, как наиболее близкого к происходившему раньше. Затем зайдём с другой стороны, и даже двух сторон, подводящих к формальному определению, при дефиците сил и времени излишнему.

Отметим сперва псевдообратную к диагональной матрице, а именно, к Σ из сингулярного разложения. Блок Σ_r невырожден, поэтому

$$\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \implies \Sigma_r^\natural = \Sigma_r^{-1} = \operatorname{diag}(1/\sigma_1, \ldots, 1/\sigma_r),$$

а нулевые внедиагональные блоки меняются местами, как того требуют размерности. Помните восстановление параметров модели?

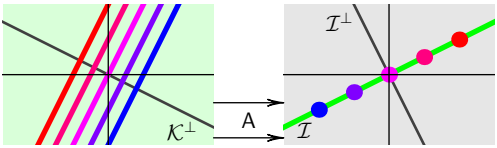
Теорема. *Сингулярные разложения A и $A^\natural$ всегда связаны:*

$$A = U\Sigma V^\dagger \Leftrightarrow A^\natural = V\Sigma^\natural U^\dagger.$$

Доказательство. Проверить четыре условия из Упражнения 11. □

Помимо довольно банального доказательства, здесь полезно внимательно посмотреть на ситуацию геометрически. Наша матрица A задаёт линейное отображение $\mathsf{A}\colon \mathbb{C}^n \to \mathbb{C}^m$ в стандартных базисах. Как и выше, ортогонально разложим оба пространства: $\mathbb{C}^n = \mathcal{K}^\perp \oplus \mathcal{K}$ и $\mathbb{C}^m = \mathcal{I} \oplus \mathcal{I}^\perp$, причём отображение $\mathsf{A}\colon \mathcal{K}^\perp \to \mathcal{I}$ обратимо и устроено весьма аналогично примеру с эллипсом. Обратное к нему отображение дополним нулём на $\mathcal{I}^\perp$.

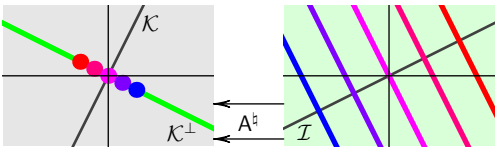
Матрица полученного в итоге отображения $\mathbb{C}^m \to \mathbb{C}^n$ в стандартных базисах совпадает с $A^\natural$. В самом деле, применим опять сингулярное разложение $A = U\Sigma V^\dagger$ вместе с его сокращённой версией $A = U_r\Sigma_r V_r^\dagger$. Матрица $U_r^\dagger$ проектирует на $\mathcal{I}$ а матрица $V_r^\dagger$ проектирует на $\mathcal{K}^\perp$. Поэтому первый шаг выражается умножением $B \mapsto U^\dagger B$. На втором шаге играет блок Σ_r^{-1}, расширенный нулями. На третьем шаге умножение на V возвращает к стандартному базису.



Пример. Первый рисунок в этой паре, ранее уже приводившийся, на самом деле вычислен для отображения $A\colon \mathbb{R}^2 \to \mathbb{R}^2$, в стандартных базисах заданного матрицей

$$A = \begin{bmatrix} -2 & 1 \\ -1 & 1/2 \end{bmatrix}.$$

Второй рисунок аналогичным образом иллюстрирует действие построенного псевдообратного отображения.



Действие псевдообратного отображения на точках $\mathbf{b} \in \mathbb{C}^m$ стоит описать подробнее, разбив его на три шага:

$n \mapsto m$
09.05.20

(1) Проектируем точку на $\mathcal{I} = \operatorname{Im} \mathsf{A}$, то есть $\mathbf{b} \mapsto \mathbf{p} = \mathsf{P}_{\mathcal{I}}(\mathbf{b})$.
(2) Берём полный прообраз $\mathcal{M} = \mathsf{A}^{-1}(\mathbf{p})$. Это даёт линейное многообразие, параллельное $\mathcal{K} = \operatorname{Ker} \mathsf{A}$, ибо разные прообразы отличаются на элементы ядра.
(3) Проектируем прообразы на $\mathcal{K}^\perp$ параллельно ядру $\mathcal{K} = \operatorname{Ker} \mathsf{A}$, то есть $\mathcal{M} \mapsto P_{\mathcal{K}^\perp}(\mathcal{M})$. При этом все точки полного прообраза проектируются в одну точку. Её можно охарактеризовать тем, что это ближайшая к началу координат точка в $\mathcal{M}$.

Упражнение 10. *Проследите эти шаги на приведённых рисунках.*

Первое условие псевдообращения

Вернёмся к системам линейных уравнений, $AX = B$, с произвольными $m \times n$ матрицами коэффициентов. Мы учились решать их методом исключения, но в частном случае квадратных матриц также получили готовую формулу $X = A^{-1}B$, пригодную только для невырожденных матриц. Теперь нас интересует возможность её обобщения.

Для начала отметим, что равенство $AA^{-1} = E$ влечёт $AA^{-1}A = A$, и попробуем использовать последнее. Может быть, матрица A° со свойством $AA^\circ A = A$ существует не только для невырожденных матриц A?

Например, если A полного ранга по строкам, то $AA^\dagger$ невырождена. Тогда $A^\circ = A^\dagger(AA^\dagger)^{-1}$ удовлетворяет первому условию, а $X = A^\circ B$ является решением системы.

Предположим, что исходная система имеет решение X_0 и существует требуемая матрица A°. Возьмём $X = A^\circ B$, тогда

$$AX = AA^\circ B = AA^\circ AX_0 = AX_0 = B,$$

и мы нашли решение системы, никак не связанное с X_0. Далее, когда система имеет множество решений, мы хотим описать его.

Теорема. *Для любой матрицы A° со свойством $AA^\circ A = A$:*

(1) *система $AX = B$ совместна $\Leftrightarrow$ $AA^\circ B = B$;*
(2) *тогда её общее решение имеет вид $X = A^\circ B + (E - A^\circ A)Y$ с произвольным вектором $Y \in \mathbb{R}^m$.*

Доказательство. Упражнение. □

Четыре условия Пенроуза

Одно лишь первое условие не ведёт к однозначности в поиске псевдообратных. Классическая обратная матрица имеет слишком много приложений, и совсем не удивительно, что различные подходы к обобщению выдвигают разные дополнительные условия. Пенроуз нашёл группу условий, ведущую к замечательному ответу, который и был разобран в начале раздела. Сперва дадим его же другими внезапно готовыми формулами, опираясь на теорему о ранговом разложении.

Определение (MacDuffee). Для произвольной комплексной матрицы A введём матрицу $A^\natural$ следующим образом:

(0) для $A = 0$ положим $A^\natural \leftrightharpoons A^\dagger = 0$;
(1) для матрицы G полного ранга по строкам $G^\natural \leftrightharpoons G^\dagger(GG^\dagger)^{-1}$;
(2) для матрицы F полного ранга по столбцам $F^\natural \leftrightharpoons (F^\dagger F)^{-1}F^\dagger$;
(3) для матрицы $A = FG$, рангово разложенной, $A^\natural \leftrightharpoons G^\natural F^\natural$.

Тогда матрицу $A^\natural$ называют **псевдообратной Мура — Пенроуза** для A.

Упражнение 11. *Проверьте для $Z = A^\natural$ четыре условия Пенроуза:*

$$AZA = A, \quad ZAZ = Z, \quad (AZ)^\dagger = AZ, \quad (ZA)^\dagger = ZA.$$

Комбинация этих условий делает псевдообратную Мура — Пенроуза чрезвычайно полезной; примеры того мы кратко рассмотрим ниже. Иных псевдообратных мы не коснёмся, а потому перестанем уточнять.

Лемма (Penrose)**.** *Для каждой заданной матрицы A одновременное решение всех четырёх уравнений на Z единственно; значит, $Z = A^\natural$.*

Упражнение 12. *Первое условие влечёт*

$$\operatorname{Im}(AZ) = \operatorname{Im} A, \qquad \operatorname{Im}(ZA)^\dagger = \operatorname{Im} A^\dagger$$
$$\operatorname{Ker}(ZA) = \operatorname{Ker} A, \qquad \operatorname{Ker}(AZ)^\dagger = \operatorname{Ker} A^\dagger.$$

Упражнение 13. *При выполнении первого и третьего условий AZ задаёт ортогональный проектор на образ A.*

Упражнение 14. *Для всех матриц Z_3 и Z_4, удовлетворяющих первому условию, установите два свойства матрицы $Z = Z_4AZ_3$:*

(1) *для Z выполнены первое и второе;*
(2) *если для Z_3 также выполнено третье условие, а для Z_4 четвёртое, то для Z выполнены все четыре, и тогда $Z = A^\natural$.*

Упражнение 15. *Докажите тождество $A^{\natural\natural} = A$.*

Пример. Для двух маленьких матриц A и B укажем псевдообратные:

$$A = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad A^\natural = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad B^\natural = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}.$$

Их можно проверить подстановкой в уравнения Пенроуза. При этом $(AB)^\natural \neq B^\natural A^\natural$, поскольку $(AB)^\natural = [1]$, но $B^\natural A^\natural = [1/2]$.

Метод наименьших квадратов

Повсюду в науке проводятся обработка и анализ наблюдательных данных, в которых часто возникают несовместные системы линейных уравнений. Следовательно, очень важна задача отыскания приближённого решения. Под этим понимают столбец X, придающий минимальное значение **невязке** $\|B - AX\|$, и называют его решением по методу наименьших квадратов.

Теорема. *Минимум невязки достигается при $X = A^\circ B$ для любой матрицы A° со свойствами $AA^\circ A = A$ и $(AA^\circ)^\dagger = AA^\circ$.*

Здесь перед нами предстали первое и третье условия Пенроуза. Поэтому в качестве решения годится то, в котором A° является именно псевдообратной Мура — Пенроуза, но гарантии единственности решения в общем случае нет, ибо наложено только два из четырёх условий.

Доказательство. Разложим вектор B как сумму $B = B_\parallel + B_\perp$ его проекций на $\mathcal{I} = \operatorname{Im} A$ и $\mathcal{I}^\perp$. Тогда $B_\parallel - AX \in \mathcal{I}$, поэтому

$$\|B - AX\|^2 = \|B_\parallel - AX\|^2 + \|B_\perp\|^2.$$

Минимум достигается, если и только если $B_{\|} = AX$. Подставляя сюда $X = A^\circ B$, получим $B_{\|} = AA^\circ B$, а это проверено в упражнении 13. □

Минимизация нормы приближённого решения

Проблему отсутствия единственности решения можно преодолеть, выбирая решение, ближайшее к началу координат, то есть имеющее минимальную норму.

Теорема. *Если система $AX = B$ совместна, то её решение минимальной нормы:*

(1) *единственно и лежит в $\operatorname{Im} A^\dagger = (\operatorname{Ker} A)^\perp$;*
(2) *равно $A^\circ B$ для любой матрицы A° со свойствами $AA^\circ A = A$ и $(A^\circ A)^\dagger = A^\circ A$.*

Здесь перед нами предстали первое и четвёртое условия Пенроуза.

Доказательство. Поскольку ограничение A на $\mathcal{K}^\perp = (\operatorname{Ker} A)^\perp$ есть биекция $\mathcal{K}^\perp \to \mathcal{I}$, существует единственное решение $X_0 \in \mathcal{K}^\perp$.

(1) Общее решение системы записывается как $X = X_0 + Y$ с произвольным $Y \in \mathcal{K}$. Тогда

$$\|X\|^2 = \|X_0\|^2 + \|Y\|^2,$$

поэтому среди всех решений именно X_0 минимизирует норму.

(2) Первая теорема раздела даёт $AA^\circ B = B$. Значит, $X = A^\circ B$ это решение, притом $X = A^\circ(AA^\circ B) = (A^\circ A)(A^\circ B)$ лежит в $\operatorname{Im} A^\circ A$. По четвёртому условию Пенроуза, а затем по его первому условию, да ещё посредством упражнения 12, совпадают образы:

$$\operatorname{Im} A^\circ A = \operatorname{Im}(A^\circ A)^\dagger = \operatorname{Im} A^\dagger = \mathcal{K}^\perp.$$

Теперь наблюдение о единственности влечёт $X = X_0$. □

Теорема (Penrose)**.** *Среди всех приближённых решений любой системы линейных уравнений $AX = B$ по методу наименьших квадратов, минимальной нормой обладает $X = A^\natural B$.*

Доказательство. По доказательству теоремы о минимуме невязки, нужно смотреть на точные решения системы $AX = AZ_3B$ со всеми матрицами Z_3, удовлетворяющими первому и третьему условиям. По предыдущей теореме, минимум нормы точного решения этой системы приносит $X = Z_4AZ_3B$ для любой матрицы Z_4, удовлетворяющей первому и четвёртому условиям. Упражнение 14 даёт $X = A^\natural B$. □

Глава 10. НАЧАЛА ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ

10.1. Линии в пространстве

Способы задания

Лекция 8+ (31.03.20)

Точки будем задавать координатами (x, y, z) или радиус-вектором $\mathbf{r} = (x, y, z)$. Параметрический способ задания линии наиболее актуален, но иногда линия возникает в неявном виде как пересечение двух поверхностей. Используемые функции должны быть разумными с точки зрения анализа, а именно, достаточное число раз непрерывно дифференцируемыми. Также в анализе обсуждаются техника переходов между различными способами задания и возникающие тонкости. Бывает и так, что линия задана по-разному на своих отдельных участках, называемых обычно *кусками*.

Способ	Смысл	Формула
Параметрический	Координаты выражены через одну независимую переменную	$\mathbf{r} = \mathbf{r}(u)$
Неявный	Координаты связаны уравнениями	$F_1(x, y, z) = 0$ $F_2(x, y, z) = 0$

Естественный параметр

Необходимо различать понятия линии как геометрического множества точек и параметризованной линии; последнее ближе к понятию траектории движущейся точки. Геометрическая линия может быть параметризована разными способами, что соответствует прохождению одной траектории с разными скоростями, а также в двух противоположных направлениях (выбор ориентации линии).

Если физически естественным параметром часто является время, то для математического исследования линии чаще удобна длина дуги, то есть длина пути, пройденного от начальной точки. Обозначим её через s. Этот параметр соответствует постоянной по модулю (единичной) скорости движения, что заметно упрощает формулы и теоретические вычисления. При произвольном параметре u длина дуги находится по формулам

$$ds = |\mathbf{r}'(u)|\,du, \quad s(b) - s(a) = \int_a^b |\mathbf{r}'(u)|\,du.$$

Производные по естественному параметру s будем обозначать точками: $\dot{\mathbf{r}}$, $\ddot{\mathbf{r}}$, $\dddot{\mathbf{r}}$, а по произвольному — штрихами: $\mathbf{r}'$, $\mathbf{r}''$, $\mathbf{r}'''$.

Сопровождающий трёхгранник

Исключим из рассмотрения неприятные и неинтересные случаи, в которых все построения этого раздела окажутся бессмысленны. Во-первых, при $\mathbf{r}' = \mathbf{0}$ у линии возможна особенность. Изучение особых (нерегулярных) точек линий весьма содержательно, но не является нашей целью; они могут быть разрешены лишь в качестве крайних. Во-вторых, если вектор ускорения $\mathbf{r}''$ коллинеарен вектору скорости $\mathbf{r}'$ на каком-то промежутке параметра, то заданный им участок линии прямолинеен. Запретить это можно, наложив условие $\mathbf{r}' \times \mathbf{r}'' \neq \mathbf{0}$.

Для произвольного ненулевого вектора $\mathbf{a}$ обозначим через $\mathbf{e}(\mathbf{a})$ единичный вектор в направлении $\mathbf{a}$. Фиксируя на параметризованной линии точку $\mathbf{r}(u)$, определим взаимно ортогональные единичные векторы к данной линии в выбранной точке:

$$\begin{aligned} \mathbf{t} &= \mathbf{e}(\mathbf{r}'), && \text{— вектор касательной,} \\ \mathbf{b} &= \mathbf{e}(\mathbf{r}' \times \mathbf{r}'') && \text{— вектор бинормали,} \\ \mathbf{n} &= \mathbf{b} \times \mathbf{t}, && \text{— вектор главной нормали.} \end{aligned}$$

Здесь как раз необходимо иметь $\mathbf{r}'(u) \times \mathbf{r}''(u) \neq \mathbf{0}$. Векторы $\mathbf{t}$, $\mathbf{n}$ и $\mathbf{b}$ составляют базис особой (подвижной) декартовой системы координат с началом в выбранной точке линии, называемый **реперо́м Френе**.

artels 1831-
Frenet 1847

Упражнение 1. *Проверьте, что выражение* $|\mathbf{r}'|^{-3}(\mathbf{r}' \times \mathbf{r}'')$ *не зависит от параметризации линии.*

Упражнение 2. *Проверьте, что репер Френе зависит не от параметризации линии, а лишь от направления её прохождения.*

Термины **касательная**, **главная нормаль** и **бинормаль** применяют к прямым, проходящим через выбранную точку линии в направлениях векторов $\mathbf{t}$, $\mathbf{n}$ и $\mathbf{b}$. Полезны также три плоскости, проходящие через пару этих прямых ортогонально третьей:

Плоскость	Ортогональна	Содержит
Нормальная	$\mathbf{t}$	$\mathbf{n}$, $\mathbf{b}$
Спрямляющая	$\mathbf{n}$	$\mathbf{b}$, $\mathbf{t}$
Соприкасающаяся	$\mathbf{b}$	$\mathbf{t}$, $\mathbf{n}$

Движение сопровождающего трёхгранника

При движении точки $\mathbf{r}(u)$ вдоль линии переменные векторы репера Френе остаются единичными и взаимно ортогональными:

$$\mathbf{t}(u)\cdot\mathbf{t}(u)=1, \qquad \mathbf{n}(u)\cdot\mathbf{b}(u)=0,$$
$$\mathbf{n}(u)\cdot\mathbf{n}(u)=1, \qquad \mathbf{b}(u)\cdot\mathbf{t}(u)=0,$$
$$\mathbf{b}(u)\cdot\mathbf{b}(u)=1, \qquad \mathbf{t}(u)\cdot\mathbf{n}(u)=0.$$

Весь трёхгранник движется как твёрдое тело; его движение раскладывается на поступательное движение его вершины $\mathbf{r}(u)$ относительно неподвижной системы координат и вращение репера $\{\mathbf{t},\mathbf{n},\mathbf{b}\}$ вокруг вершины.

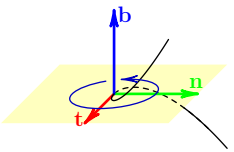
Скорости движения векторов репера выражаются через сами векторы $\mathbf{t}$, $\mathbf{n}$ и $\mathbf{b}$. Дифференцируя указанные постоянные скалярные произведения и опуская параметр u, получим выражения вида

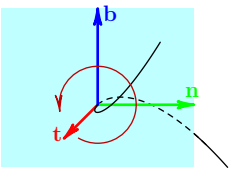
$$2\,\mathbf{t}\cdot\mathbf{t}'=0, \quad \mathbf{n}'\cdot\mathbf{b}+\mathbf{n}\cdot\mathbf{b}'=0.$$

Эти сведения можно собрать в (формальное) матричное равенство

$$\begin{bmatrix}\mathbf{t}'\\ \mathbf{n}'\\ \mathbf{b}'\end{bmatrix}=\begin{bmatrix}0 & k_1 & k_3\\ -k_1 & 0 & k_2\\ -k_3 & -k_2 & 0\end{bmatrix}\cdot\begin{bmatrix}\mathbf{t}\\ \mathbf{n}\\ \mathbf{b}\end{bmatrix}$$

с неизвестными пока функциями $k_i(u)$. Бинормаль по определению ортогональна как скорости $\mathbf{r}'$, так и ускорению $\mathbf{r}''$, а производная вектора $\mathbf{t}=\mathbf{r}'/|\mathbf{r}'|$ есть линейная комбинация $\mathbf{r}'$ и $\mathbf{r}''$. Значит, $k_3(u)\equiv 0$. Оставшиеся две функции могут быть какими угодно.





Теперь выберем длину дуги s в качестве параметра. В этом случае вместо k_1 и k_2 пишут $k=k(s)$ и $\varkappa=\varkappa(s)$. Эти функции называют соответственно **кривизной** и **кручением** данной линии в точке $\mathbf{r}(s)$. Они являются скоростями вращения соответственно касательной вокруг бинормали и бинормали вокруг касательной.

Наше матричное равенство эквивалентно **формулам Френе — Серре**:

Bartels 1831-
Frenet 1847
Serret 1851

$$\dot{\mathbf{t}}=k\,\mathbf{n}, \qquad \dot{\mathbf{n}}=\varkappa\,\mathbf{b}-k\,\mathbf{t}, \qquad \dot{\mathbf{b}}=-\varkappa\,\mathbf{n}.$$

Вектор $k\,\mathbf{n}$ называют **вектором кривизны** линии в точке.

Мгновенную угловую скорость $\boldsymbol{\omega}(s)$ вращения репера Френе называют **вектором Дарбу**. При этом каждый вектор $\mathbf{a}$, имеющий постоянные координаты относительно подвижного репера $\{\mathbf{t},\mathbf{n},\mathbf{b}\}$, изменяется относительно неподвижного репера $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ по формуле $\dot{\mathbf{a}} = \boldsymbol{\omega}\times\mathbf{a}$. Подставляя сюда по очереди векторы $\mathbf{a}=\mathbf{n}$ и $\mathbf{a}=\mathbf{t}$, находим выражение $\boldsymbol{\omega} = \varkappa\,\mathbf{t} + k\,\mathbf{b}$ для вектора Дарбу.

раздел 6.1

Упражнение 3. *Найдите координаты векторов* $\dot{\boldsymbol{\omega}}$, $\ddot{\boldsymbol{\omega}}$ *и* $\ddot{\boldsymbol{\omega}}+\dot{\boldsymbol{\omega}}\times\boldsymbol{\omega}$ *в базисе* $\{\mathbf{t},\mathbf{n},\mathbf{b}\}$.

В курсе дифференциальных уравнений доказывается теорема Пикара о существании и единственности решения системы линейных (обыкновенных дифференциальных) уравнений. Значит, благодаря ей формулы Френе — Серре гарантируют, что вместе с начальным положением репера Френе кривизна и кручение как функции длины дуги однозначно задают линию в пространстве.

Вычисление кривизны и кручения

Из первой формулы Френе — Серре видим, что $k = |\dot{\mathbf{t}}| = |\ddot{\mathbf{r}}|$. Поскольку $|\dot{\mathbf{r}}| = 1$, отсюда можно сперва перейти к $k = |\dot{\mathbf{r}}\times\ddot{\mathbf{r}}|$, а затем применить упражнение 1. Получим формулу

$$k = \frac{|\mathbf{r}'\times\mathbf{r}''|}{|\mathbf{r}'|^3}$$

для кривизны при произвольной параметризации $\mathbf{r}=\mathbf{r}(u)$. Она полезна в таком виде, хотя при $u=s$ упрощается до $k=|\ddot{\mathbf{r}}|$. Причина в том, что на практике переход от исходного параметра к естественному может быть громоздок и такими упрощениями не оправдывается.

Вычислим кручение при произвольной параметризации $\mathbf{r}=\mathbf{r}(u)$. Будем обозначать $|\mathbf{r}'(u)|$ через v. Используем несколько раз цепное правило, замалчивая явный вид скалярных функций $\alpha(u)$, $\beta(u)$ и $\gamma(u)$, всё равно сразу исчезающих ввиду свойств смешанного произведения:

$$(\mathbf{r}',\mathbf{r}'',\mathbf{r}''') = (v\,\dot{\mathbf{r}},\,\alpha\,\dot{\mathbf{r}}+v^2\,\ddot{\mathbf{r}},\,\beta\,\dot{\mathbf{r}}+\gamma\,\ddot{\mathbf{r}}+v^3\,\dddot{\mathbf{r}}) = 0 + v^6(\dot{\mathbf{r}},\ddot{\mathbf{r}},\dddot{\mathbf{r}}).$$

Далее по формулам Френе — Серре находим

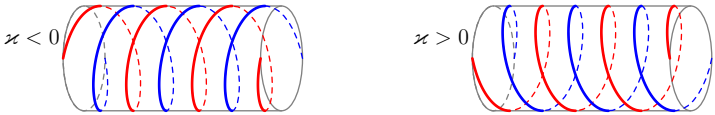
$$\begin{aligned}(\dot{\mathbf{r}},\ddot{\mathbf{r}},\dddot{\mathbf{r}}) = (\mathbf{t},\,k\,\mathbf{n},\,\dot{k}\,\mathbf{n}+k\,\dot{\mathbf{n}}) &= (\mathbf{t},\,k\,\mathbf{n},\,\dot{k}\,\mathbf{n}-k^2\,\mathbf{t}+k\varkappa\,\mathbf{b})\\ &= 0 + k^2\varkappa\,\{\mathbf{t},\mathbf{n},\mathbf{b}\} = k^2\varkappa.\end{aligned}$$

Подставляя уже известную формулу для k, получаем

$$\varkappa = \frac{(\mathbf{r}',\mathbf{r}'',\mathbf{r}''')}{(\mathbf{r}'\times\mathbf{r}'')^2}.$$

Кручение является рациональной функцией производных координат, в отличие от кривизны, содержащей квадратный корень при вычислении модулей.

Кручение бывает как положительным, так и отрицательным. Правый и левый винты различаются именно знаком кручения.



Особые случаи

Расширяя сказанное об исключениях, выделим четыре особых случая поведения линии в одной точке или на некотором куске.

Условие	В точке	На куске
$\dot{\boldsymbol{\omega}} = \mathbf{0}$	(Замалчивается?)	Винтовая линия
$\varkappa = 0$	Точка уплощения	Плоская линия
$k = 0$	Точка распрямления	Прямая
$\mathbf{r}' = \mathbf{0}$	Нерегулярная точка	—

Условие $\dot{\boldsymbol{\omega}} = \mathbf{0}$ равносильно постоянству кривизны и кручения, как отмечено в упражнении 3. Если $\varkappa = 0$ в одной точке, то вблизи неё линия более обычного похожа на плоскую. Если $k = 0$ в одной точке, что равносильно $\mathbf{r}' \times \mathbf{r}'' = \mathbf{0}$, то вблизи неё линия более обычного похожа на прямую; репер Френе там может изменяться скачком, а кручение не определено. В нерегулярной же точке всё может быть достаточно плохо помимо того, что кривизна не определена. В элементарных курсах обычно избегают рассмотрения нерегулярных точек.

Проекции линии на плоскости трёхгранника

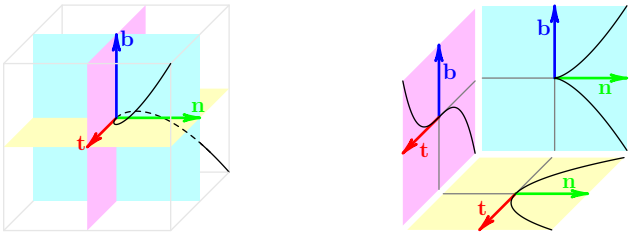
Выберем на линии точку $P = \mathbf{r}(s)$ и изучим поведение линии около неё относительно плоскостей сопровождающего трёхгранника. Разложим смещение $\boldsymbol{\delta} = \mathbf{r}(s+\varepsilon) - \mathbf{r}(s)$ по Тэйлору:

$$\boldsymbol{\delta} = \varepsilon\,\dot{\mathbf{r}} + \tfrac{1}{2}\varepsilon^2\,\ddot{\mathbf{r}} + \tfrac{1}{6}\varepsilon^3\,\dddot{\mathbf{r}} + \mathbf{o}(\varepsilon^3).$$

Выражения для $\dot{\mathbf{r}}$, $\ddot{\mathbf{r}}$ и $\dddot{\mathbf{r}}$ в точке P в базисе $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ мы нашли на пути к формуле для кручения. Подставив и собрав координаты, получим

$$\boldsymbol{\delta} = (\varepsilon + o(\varepsilon^2))\,\mathbf{t} + (\tfrac{1}{2}k\varepsilon^2 + o(\varepsilon^2))\,\mathbf{n} + (\tfrac{1}{6}k\varkappa\varepsilon^3 + o(\varepsilon^3))\,\mathbf{b}.$$

При $k \neq 0$ и $\varkappa \neq 0$ линия вблизи точки P похожа на **скрученную параболу** $\mathbf{r}(u) = u\,\mathbf{t} + \alpha u^2\mathbf{n} + \beta u^3\mathbf{b}$.

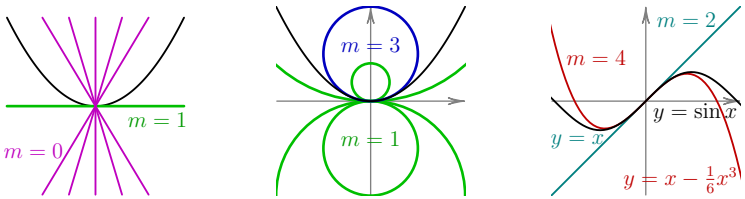


Проекциями на плоскости трёхгранника тоже будут линии, походящие на разные параболы. Проекция на соприкасающуюся плоскость имеет в точке P ту же кривизну, что и изучаемая пространственная линия. Проекция на спрямляющую плоскость имеет в P перегиб, и в частности, нулевую кривизну, чем и объясняется название этой плоскости. Направление перегиба соответствует знаку константы β, равному знаку кручения, ибо кривизна линии в пространстве не бывает отрицательной. У проекции на нормальную плоскость P является особой точкой, а именно, точкой возврата.

Плоскость проекции		Модельная линия	
$\langle \mathbf{t}, \mathbf{n} \rangle$	Соприкасающаяся	$x^2 = y$	Парабола обыкновенная
$\langle \mathbf{t}, \mathbf{b} \rangle$	Спрямляющая	$x^3 = z$	Парабола кубическая
$\langle \mathbf{n}, \mathbf{b} \rangle$	Нормальная	$y^3 = z^2$	Парабола полукубическая

Соприкасающаяся окружность

Для двух гладких линий $\mathbf{r}_1(s)$ и $\mathbf{r}_2(s)$ введём условие **касания m-го порядка** как совпадение в точке касания значений функций и их производных до порядка m включительно. Одна линия будет фиксирована, а вторую будем подбирать для изучения первой, то есть приближать её до нужного порядка.



Пример. На среднем рисунке синяя окружность $x^2 + (y - \frac{1}{2})^2 = (\frac{1}{2})^2$ имеет касание 3-го порядка с параболой $y = x^2$, потому что вблизи начала координат на этой окружности $y = x^2 + O(x^4)$.

Простейшее приближение первого порядка даёт обычную касательную — прямую линию. Для второго порядка касания прямая не годится, кроме точек распрямления. Поэтому простейшее приближение второго порядка ищут среди окружностей. Вы знакомились с ним в курсе механики: центр и радиус этой окружности называются **центром** и **радиусом кривизны**. Радиус есть $R = 1/k$, а центром в подвижной системе координат является $R\,\mathbf{n}$. Они имеют смысл и для плоских линий, когда ими и исчерпывается картинка. (Она развивается с изучением эволют и эвольвент, но мы не трогаем их.)

В пространстве же можно отметить, что, поскольку в точке касания касательная к этой окружности направлена по вектору $\mathbf{t}$, то и вся окружность лежит в соприкасающейся плоскости.

Соприкасающаяся сфера

Возьмём несложную линию или поверхность F, имеющую с линией $\mathbf{r}(s)$ общую точку P. Сравним расстояния от другой близкой точки линии до P и до фигуры F, обозначая их через ε и $\delta(\varepsilon)$ соответственно. Если второе есть бесконечно малое порядка (точно) $m{+}1$ относительно первого, $\delta(\varepsilon) = O(\varepsilon^{m+1})$, то исходная линия и новая фигура имеют в точке P **касание** (точно) m**-го порядка**. Если F тоже линия, получается условие, равносильное введённому выше.

Тип фигуры	m	Специфика фигуры	$\exists!$ при
Прямая	1	Касательная к линии	$\mathbf{r}' \neq 0$
Плоскость	1	Содержащая касательную	Не единственна
Плоскость	2	Соприкасающаяся	$k \neq 0$
Окружность	2	Соприкасающаяся	$k \neq 0$
Сфера	3	Соприкасающаяся	$\varkappa \neq 0$

Касательная прямая и соприкасающиеся плоскость, окружность и сфера выделяются среди фигур своего типа наивысшим порядком соприкосновения с данной линией в выбранной её точке. Для прямой и плоскости эти утверждения легко следуют из выписанного выше разложения смещения $\boldsymbol{\delta}(\varepsilon)$: достаточно посмотреть, куда входят ε и ε^2. Оттуда же можно определить параметры для окружности и сферы.

Пример. Найдём соприкасающуюся сферу для стандартной скрученной параболы $\mathbf{r}(u) = (u, u^2, u^3)$ в точке с $u = 0$. Репер $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ там совпадает со стандартным репером $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Поместим центр сферы в

точку $\mathbf{c} = (x, y, z)$ и раскроем квадрат расстояния до точки на линии:

$$|\mathbf{r}(u) - \mathbf{c}|^2 = (u - x)^2 + (u^2 - y)^2 + (u^3 - z)^2$$
$$= (x^2 + y^2 + z^2) - 2x \cdot u + (1 - 2y) \cdot u^2 - 2z \cdot u^3 + o(u^3).$$

Постоянное слагаемое тут равно квадрату радиуса искомой сферы. Зануляя остальные удержанные слагаемые, находим координаты её центра: $\mathbf{c} = (0, 1/2, 0)$, или $\mathbf{c} = \mathbf{n}/2$.

В рассмотренном примере центр сферы лежит на главной нормали, но в общем случае это не так, и центр сферы лежит в нормальной плоскости $\langle \mathbf{n}, \mathbf{b} \rangle$. Его проекция на соприкасающуюся плоскость $\langle \mathbf{t}, \mathbf{n} \rangle$ является центром соприкасающейся окружности.

Соприкасающаяся окружность является пересечением соприкасающейся сферы и соприкасающейся плоскости. Центр сферы и центр окружности соединяет параллельная $\mathbf{b}$ прямая, называемая **осью кривизны** пространственной линии в выбранной точке P.

Упражнение 4. *Докажите, что вектор из точки касания до центра соприкасающейся сферы равен*

$$\frac{\ddot{\mathbf{r}} \times \dot{\mathbf{r}}}{(\dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}})}.$$

Указание: смотрите формулы (295)–(297) *в учебнике Рашевского, либо распишите* $\boldsymbol{\delta}(\varepsilon)$ *точнее и приравняйте нулю слагаемые с* ε^2 *и* ε^3.

Как видно в этом упражнении, при $\varkappa \to 0$, что означает уплощение линии, радиус соприкасающейся сферы стремится к бесконечности, и в точке уплощения «сфера» совпадает с соприкасающейся плоскостью. Можно считать, что так происходит в каждой точке плоской линии. Другой интересный случай: линии, лежащие на сфере, притом не плоские, то есть не дуги окружностей. Эта сфера тогда является соприкасающейся в каждой точке линии.

Пример. Линия Вивиани, вероятно, наиболее широко известная такая линия, к тому же очень красивая. Обычно её представляют как пересечение сферы и кругового цилиндра, но любую из этих поверхностей можно заменить некоторыми другими, в том числе простыми квадратичными. Поэтому линия Вивиани является одной из простейших линий на сфере.

Если сфера равномерно вращается, а вокруг неё с тем же периодом вращается точка по круговой орбите, проходящей над полюсами, то проекция точки на сферу описывает линию Вивиани.

Упражнение 5. *Проверьте последнее утверждение.*

Пример. В геологии линии на сфере применяются для описания движения плит земной коры.

10.2. Геометрия на поверхности

Лекция 9 (03.04.20)

Способы задания

Как и для линий, параметрический способ задания остаётся основным, но доминирует слабее, потому что неявный способ тоже нередко приносит простые формулы. Явный способ задания поверхности — частный случай как параметрического, так и неявного.

Способ	Смысл	Формула
Параметрический	Координаты выражены через две независимые переменные	$\mathbf{r} = \mathbf{r}(u, v)$
Явный	Одна из координат выражена через остальные	$z = f(x, y)$
Неявный	Координаты связаны уравнением	$F(x, y, z) = 0$

Теоретической основой перехода от неявного задания к явному является одна из важнейших теорем классического анализа — теорема о неявной функции. В практических вычислениях лишь для небольшого списка типов поверхностей удаётся совершить переход от одного способа задания к другому. Типична ситуация, когда поверхность параметризуется не целиком, а по частям; даже сфера требует выделения не менее двух перекрывающихся кусков.

При изучении параметризованной поверхности постоянно используются частные производные радиус-вектора $\mathbf{r}$ по параметрам, так что для них нужны компактные обозначения. Примем один из стандартных вариантов: $\mathbf{r}_u$ и $\mathbf{r}_v$. Вторые частные производные обозначим через $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$ и $\mathbf{r}_{vv}$.

Нормаль и касательная плоскость

Касательным пространством $T_P S$ к поверхности S в точке P называется линейная оболочка касательных векторов к линиям, лежащим на поверхности и проходящим через P. Как показывает пример вершины конуса $x^2 + y^2 - z^2 = 0$, размерность касательного пространства может превышать число 2, которое интуиция выдвигает на роль размерно-

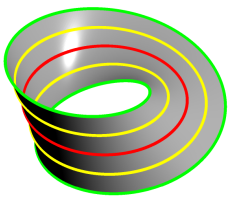
сти самой поверхности. Точки поверхности, в которых случается такое превышение, называют **особыми** или **сингулярными**, но таковых «мало», и мы очень стараемся их избегать. Остальные точки — **регулярные**.

Нормалью к поверхности S в точке P называют вектор, ортогональный ко всем касательным векторам к S в P. Таким вектором для поверхности $\mathbf{r} = \mathbf{r}(u, v)$ является векторное произведение $\mathbf{r}_u \times \mathbf{r}_v$, а для поверхности $F(x, y, z) = 0$ — градиент $\nabla F = (F_x, F_y, F_z)$.

Если указанные формулы в какой-то точке дают нулевой вектор, она может оказаться особой. При наличии ненулевой нормали в точке P касательное пространство обязательно имеет размерность 2 и называется **касательной плоскостью**, так что P гарантированно регулярна.

Ориентируемость

Вектор нормали часто удобно брать единичным, но в каждой точке возможно два направления. Выбор направления в близких друг другу точках желательно делать согласованно, что всегда возможно на отдельно взятом параметризованном куске поверхности. Однако при сшивании поверхности из кусков может обнаружиться невозможность согласовать направления нормалей всех кусков. Пример этого явления — лента Мёбиуса. В таком случае поверхность называется **неориентируемой**. Если поверхность ориентируема, то её **ориентацией** называют выбор одного из двух возможных согласований нормалей.



Практически все вопросы о поверхностях, которые затрагивает наш курс, относятся к куску поверхности, поэтому неориентируемые поверхности не рассматриваются.

Первая квадратичная форма

Вследствие выбора параметризации $\mathbf{r} = \mathbf{r}(u, v)$ (куска) поверхности, в каждой касательной плоскости появляется выделенный базис $(\mathbf{r}_u, \mathbf{r}_v)$. Приращение $d\mathbf{r}$ связано с приращениями du и dv параметров правилом дифференцирования $d\mathbf{r} = \mathbf{r}_u\, du + \mathbf{r}_v\, dv$, в чём нетрудно усмотреть запись касательного вектора $d\mathbf{r}$ в базисе $(\mathbf{r}_u, \mathbf{r}_v)$.

Первой квадратичной формой или **метрикой** параметризованной поверхности S называют функцию $\mathtt{I} = d\mathbf{r}\cdot d\mathbf{r}$. Как правило, она меняется от точки к точке, хотя нередко эту зависимость не указывают в обозначениях. Метрика возникает во всех геометрически важных формулах, связанных с поверхностью S, поэтому её коэффициентам присвоены устоявшиеся имена:

$$E = \mathbf{r}_u \cdot \mathbf{r}_u, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v,$$

или в матричном виде

$$\begin{bmatrix} E & F \\ F & G \end{bmatrix} = \begin{bmatrix} \mathbf{r}_u \\ \mathbf{r}_v \end{bmatrix} \begin{bmatrix} \mathbf{r}_u & \mathbf{r}_v \end{bmatrix}. \tag{4}$$

Значение функции $\mathtt{I}$ на касательном векторе $\boldsymbol{\xi} = \xi_1\,\mathbf{r}_u + \xi_2\,\mathbf{r}_v$ равно $\mathtt{I}(\boldsymbol{\xi}) = \boldsymbol{\xi}\cdot\boldsymbol{\xi}$, то есть

$$\mathtt{I}(\boldsymbol{\xi}) = E\,\xi_1^2 + 2F\,\xi_1\xi_2 + G\,\xi_2^2 = \begin{bmatrix} \xi_1 & \xi_2 \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}.$$

Ассоциированную симметричную билинейную форму обозначают той же буквой $\mathtt{I}$, так что для двух векторов

$$\boldsymbol{\xi} = \xi_1\,\mathbf{r}_u + \xi_2\,\mathbf{r}_v, \quad \boldsymbol{\eta} = \eta_1\,\mathbf{r}_u + \eta_2\,\mathbf{r}_v,$$

касательных в *одной и той же* точке поверхности, имеем значение $\mathtt{I}(\boldsymbol{\xi},\boldsymbol{\eta}) = \boldsymbol{\xi}\cdot\boldsymbol{\eta}$, то есть

$$\mathtt{I}(\boldsymbol{\xi},\boldsymbol{\eta}) = E\,\xi_1\eta_1 + F\,\xi_1\eta_2 + F\,\xi_2\eta_1 + G\,\xi_2\eta_2 = \begin{bmatrix} \xi_1 & \xi_2 \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}.$$

Смысл введения метрики в том, чтобы выделить часто повторяющийся в вычислениях фрагмент в статусный объект.

Пример (сфера). Параметризуем единичную сферу долготой φ и широтой θ (в ролях букв u и v):

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\theta\cos\varphi \\ \cos\theta\sin\varphi \\ \sin\theta \end{bmatrix}.$$

Находим частные производные

$$\mathbf{r}_\varphi = \begin{bmatrix} -\cos\theta\sin\varphi \\ \cos\theta\cos\varphi \\ 0 \end{bmatrix}, \quad \mathbf{r}_\theta = \begin{bmatrix} -\sin\theta\cos\varphi \\ -\sin\theta\sin\varphi \\ \cos\theta \end{bmatrix},$$

затем скалярные произведения:

$$\begin{bmatrix} E & F \\ F & G \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 \\ 0 & 1 \end{bmatrix}.$$

Длины, углы, площади

Конечно, длина касательного вектора $\boldsymbol{\xi}$ равна $\sqrt{\boldsymbol{\xi}\cdot\boldsymbol{\xi}} = \sqrt{\mathtt{I}(\boldsymbol{\xi})}$, а косинус угла между двумя касательными векторами $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$ в одной точке равен

$$\frac{\boldsymbol{\xi}\cdot\boldsymbol{\eta}}{|\boldsymbol{\xi}|\,|\boldsymbol{\eta}|} = \frac{\mathtt{I}(\boldsymbol{\xi},\boldsymbol{\eta})}{\sqrt{\mathtt{I}(\boldsymbol{\xi})\mathtt{I}(\boldsymbol{\eta})}}.$$

Лежащую на поверхности линию задают параметризацией $u = u(t)$, $v = v(t)$, а тогда элемент ds длины её дуги равен

$$|\mathbf{r}'(t)|\,dt = |u'(t)\,\mathbf{r}_u + v'(t)\,\mathbf{r}_v|\,dt,$$

где штрихи обозначают производные по t, то есть значению

$$\sqrt{E(t)u'(t)^2 + 2F(t)u'(t)v'(t) + G(t)v'(t)^2}\,dt.$$

Длину куска линии вычисляют интегрированием этого выражения.

Элемент площади определяется при помощи характеризации модуля векторного произведения как площади параллелограмма:

$$dS = |\mathbf{r}_u \times \mathbf{r}_v|\,du\,dv.$$

Тождество $\mathbf{a}^2\mathbf{b}^2 = (\mathbf{a}\cdot\mathbf{b})^2 + (\mathbf{a}\times\mathbf{b})^2$ позволяет выразить $|\mathbf{r}_u \times \mathbf{r}_v|$ через коэффициенты метрики. В итоге получается формула для вычисления площади куска D поверхности:

$$S(D) = \iint_D \sqrt{EG - F^2}\,du\,dv.$$

Как видно из приведённых формул, для основных геометрических вычислений на поверхности не обязательно знать её параметризацию! Достаточно знать метрику. Метрика поверхности определяет на ней неэвклидову геометрию (эвклидов, или плоский, случай имеет метрику $d\mathbf{r}^2 = du^2 + dv^2$ и соответствует единичной матрице). Это первый шаг к полезному представлению о поверхности как «самой по себе», а не как вложенной в привычное 3-мерное эвклидово пространство.

Смена параметризации

Один и тот же кусок поверхности можно параметризовать различными способами. Со временем пригодится чёткое понимание закона преобразования метрики при смене параметризации. Итак, при $\mathbf{r} = \mathbf{r}(u,v) = \mathbf{r}(\tilde{u},\tilde{v})$ правила замены переменных в дифференциалах

$$\left.\begin{aligned}\mathbf{r}_{\tilde{u}} &= u_{\tilde{u}}\,\mathbf{r}_u + v_{\tilde{u}}\,\mathbf{r}_v\\ \mathbf{r}_{\tilde{v}} &= u_{\tilde{v}}\,\mathbf{r}_u + v_{\tilde{v}}\,\mathbf{r}_v\end{aligned}\right\} \quad\Longleftrightarrow\quad \begin{bmatrix}\mathbf{r}_{\tilde{u}} & \mathbf{r}_{\tilde{v}}\end{bmatrix} = \begin{bmatrix}\mathbf{r}_u & \mathbf{r}_v\end{bmatrix}\begin{bmatrix}u_{\tilde{u}} & u_{\tilde{v}}\\ v_{\tilde{u}} & v_{\tilde{v}}\end{bmatrix}$$

дают матрицу перехода от базиса $(\mathbf{r}_u, \mathbf{r}_v)$ к базису $(\mathbf{r}_{\tilde{u}}, \mathbf{r}_{\tilde{v}})$. Это матрица Якоби $J = \frac{D(u,v)}{D(\tilde{u},\tilde{v})}$. Вычисляя коэффициенты метрики по определению (4), находим

$$\begin{bmatrix}\tilde{E} & \tilde{F}\\ \tilde{F} & \tilde{G}\end{bmatrix} = J^{\top}\begin{bmatrix}E & F\\ F & G\end{bmatrix}J.$$

10.3. Кривизна поверхности

Геодезическая и нормальная кривизны линии

Выберем на поверхности S точку P, затем проведём через неё лежащую на S линию. Не меняя точки, заинтересуемся зависимостью между кривизной линии и расположением её сопровождающего трёхгранника $\{\mathbf{t},\mathbf{n},\mathbf{b}\}$ относительно единичной нормали $\mathbf{m}$ к поверхности. Вектор кривизны $\mathbf{k} = k\,\mathbf{n}$ нашей линии естественно разложить на нормальную и касательную к поверхности компоненты:

$$\mathbf{k} = \mathbf{k}_{\mathrm{H}} + \mathbf{k}_{\Gamma}, \quad \mathbf{k}_{\mathrm{H}} = k_{\mathrm{H}}\,\mathbf{m}, \quad \mathbf{k}_{\Gamma} = \pm k_{\Gamma}\,\mathbf{m}\times\mathbf{t}.$$

Коэффициенты k_{H} и k_{Γ} называют соответственно **нормальной кривизной** и **геодезической кривизной** линии на поверхности, причём принято выбирать знак так, чтобы $k_{\Gamma} \geqslant 0$.

При естественной параметризации линии $\mathbf{r} = \mathbf{r}(s)$ на S компоненты вектора кривизны $\mathbf{k} = \ddot{\mathbf{r}}$ найдём по формулам проектирования вектора на направление:

$$k_{\mathrm{H}} = \ddot{\mathbf{r}}\cdot\mathbf{m}, \quad k_{\Gamma} = |\ddot{\mathbf{r}}\cdot(\dot{\mathbf{r}}\times\mathbf{m})| = |(\ddot{\mathbf{r}},\dot{\mathbf{r}},\mathbf{m})|.$$

При произвольной параметризации линии в знаменателях появляются поправки на скорость:

$$k_{\mathrm{H}} = \frac{\mathbf{r}''\cdot\mathbf{m}}{|\mathbf{r}'|^2}, \quad k_{\Gamma} = \frac{|(\mathbf{r}'',\mathbf{r}',\mathbf{m})|}{|\mathbf{r}'|^3}.$$

Упражнение 6. *Проекция линии* $\mathbf{r}(t)$ *на касательную плоскость* $T_P S$ *имеет в точке* P *кривизну, равную* k_Γ.

кручение?

Геодезические на поверхности

Геодезической на поверхности называют линию, геодезическая кривизна которой в каждой точке равна нулю. Есть несколько равносильных условий в терминах сопровождающего трёхгранника:

(а) главная нормаль $\mathbf{n}$ линии всегда коллинеарна нормали $\mathbf{m}$ к поверхности;
(б) соприкасающаяся плоскость линии всегда содержит $\mathbf{m}$;
(в) спрямляющая плоскость линии всегда совпадает с касательной плоскостью к поверхности.

Теорема. *Через каждую точку поверхности в каждом направлении проходит единственная геодезическая.*

Точнее будет сказать, что любые две достаточно короткие геодезические, проходящие через одну точку в одном направлении, являются дугами большей геодезической. Фактически геодезические продолжаются или неограниченно, или вплоть до края поверхности.

Доказательство. На параметризованной поверхности $\mathbf{r}(u,v)$ мы ищем уравнение геодезической в виде $v = g(u)$. Подстановка $\mathbf{r}(u, g(u))$ в числитель $(\mathbf{r}'', \mathbf{r}', \mathbf{m})$ левой части уравнения $k_\Gamma = 0$ даёт дифференциальное уравнение вида $g'' = F(g')$. Явный вид рациональной функции F здесь роли не играет; её коэффициенты зависят от $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$ и $\mathbf{r}_{vv}$. Существование в некоторой окрестности начальной точки единственного решения с заданными начальными условиями — точка и направление — обеспечивает теорема Пикара. □

Свойства геодезических на поверхности делают их аналогами прямых на плоскости.

Теорема. *Среди всех линий на поверхности, проходящих через выбранную точку в выбранном направлении, геодезическая имеет наименьшую по абсолютной величине кривизну.*

Доказательство. Следует из разложения $k^2 = k_\text{Н}^2 + k_\Gamma^2$ и определённости нормальной кривизны, установленной чуть ниже. □

Теорема. *Если две точки на поверхности достаточно близки, то кратчайшим путём между ними является дуга геодезической.*

Доказательство. Удобно выбрать в окрестности данных точек P_1 и P_2 особую (**полугеодезическую**) параметризацию, в которой искомая геодезическая служит координатной линией, а метрика имеет вид

$$ds^2 = du^2 + G(u,v)\,dv^2.$$

Приняв, что это всегда возможно, и помня, что $G > 0$ по определению, оценим длину произвольной линии, соединяющей P_1 и P_2:

$$L = \int_{P_1}^{P_2} \sqrt{du^2 + G\,dv^2} \geqslant \int_{P_1}^{P_2} \sqrt{du^2}.$$

В правой части стоит как раз длина геодезической, поскольку вдоль неё $dv \equiv 0$. □

Вторая квадратичная форма

Двигаясь вдоль линии $u = u(t)$, $v = v(t)$ на поверхности $\mathbf{r}(u,v)$, будем следить за изменением единичного вектора нормали $\mathbf{m}(u,v)$. Дифференцируя соотношения $\mathbf{m}\cdot\mathbf{m} = 1$ и $\mathbf{r}'\cdot\mathbf{m} = 0$, видим, что $\mathbf{m}\cdot\mathbf{m}' = 0$ и $\mathbf{r}''\cdot\mathbf{m} = -\mathbf{r}'\cdot\mathbf{m}'$. Поэтому нормальная кривизна линии выражается через пару $\mathbf{r}'$, $\mathbf{m}'$ касательных векторов к поверхности.

Отвлекаясь от линии, выделим второй статусный объект: **второй квадратичной формой** поверхности называют $\mathtt{II} = -d\mathbf{r}\cdot d\mathbf{m}$. Её коэффициентам также присвоены устоявшиеся имена:

$$\begin{aligned} L &= \mathbf{r}_{uu}\cdot\mathbf{m} = -\mathbf{r}_u\cdot\mathbf{m}_u,\\ M &= \mathbf{r}_{uv}\cdot\mathbf{m} = -\mathbf{r}_u\cdot\mathbf{m}_v = -\mathbf{r}_v\cdot\mathbf{m}_u,\\ N &= \mathbf{r}_{vv}\cdot\mathbf{m} = -\mathbf{r}_v\cdot\mathbf{m}_v, \end{aligned}$$

или в матричном виде

$$\begin{bmatrix} L & M \\ M & N \end{bmatrix} = -\begin{bmatrix} \mathbf{r}_u \\ \mathbf{r}_v \end{bmatrix} \begin{bmatrix} \mathbf{m}_u & \mathbf{m}_v \end{bmatrix}.$$

Оператор Родрига

Поскольку $\mathbf{r}_u$ и $\mathbf{r}_v$ составляют базис касательной плоскости T_PS, правило

$$\mathsf{A}\mathbf{r}_u = \mathbf{m}_u, \quad \mathsf{A}\mathbf{r}_v = \mathbf{m}_v$$

задаёт единственный линейный оператор на T_PS, переводящий приращение $d\mathbf{r} = \mathbf{r}_u\,du + \mathbf{r}_v\,dv$ радиус-вектора в соответствующее приращение $d\mathbf{m} = \mathbf{m}_u\,du + \mathbf{m}_v\,dv$ единичной нормали.

Теорема. *Линии на поверхности, имеющие в общей точке общую касательную, имеют в ней равные нормальные кривизны.*

Доказательство. В одну строчку: $k_{\text{н}} = \ddot{\mathbf{r}} \cdot \mathbf{m} = -\dot{\mathbf{r}} \cdot \dot{\mathbf{m}} = -\dot{\mathbf{r}} \cdot \mathsf{A}\dot{\mathbf{r}}$, но условием $\dot{\mathbf{r}}$ определён. □

Упражнение 7. *Проверьте, что в базисе* $(\mathbf{r}_u, \mathbf{r}_v)$ *оператор Родрига* $\mathsf{A}\colon d\mathbf{r} \mapsto d\mathbf{m}$ *задаётся матрицей*

$$-\begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} L & M \\ M & N \end{bmatrix}.$$

Упражнение 8. *Заметив, что*

$$\mathbf{m}_u \times \mathbf{m}_v = K\,(\mathbf{r}_u \times \mathbf{r}_v), \tag{5}$$

выразите скаляр $K = K(u, v)$ *через коэффициенты первой и второй квадратичных форм, либо через* A. *Указание: умножьте* $\mathbf{m}_u \times \mathbf{m}_v$ *и* $\mathbf{r}_u \times \mathbf{r}_v$ *скалярно на* $\mathbf{r}_u \times \mathbf{r}_v$ *и примените тождество*

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = \begin{vmatrix} \mathbf{a}\cdot\mathbf{c} & \mathbf{b}\cdot\mathbf{c} \\ \mathbf{a}\cdot\mathbf{d} & \mathbf{b}\cdot\mathbf{d} \end{vmatrix}.$$

Главные направления и главные кривизны поверхности

Ввиду только что доказанной теоремы осмысленно понятие нормальной кривизны *самой поверхности* в данном касательном направлении $\boldsymbol{\xi} = \xi_1\,\mathbf{r}_u + \xi_2\,\mathbf{r}_v$. Она находится по формуле

$$k_{\text{н}}(\boldsymbol{\xi}) = \frac{\mathtt{II}(\boldsymbol{\xi})}{\mathtt{I}(\boldsymbol{\xi})} = \frac{L\xi_1^2 + 2M\xi_1\xi_2 + N\xi_2^2}{E\xi_1^2 + 2F\xi_1\xi_2 + G\xi_2^2},$$

причём без потери общности можно ограничиться векторами единичной длины $\mathtt{I}(\boldsymbol{\xi})$.

Заменой базиса пространства — а здесь это касательная плоскость — квадратичная форма $\mathtt{II}$ приводится к каноническому (диагональному) виду:

$$\mathtt{II}(\boldsymbol{\xi}) = \tilde{L}\tilde{\xi}_1^2 + \tilde{N}\tilde{\xi}_2^2.$$

Если $\tilde{L} \neq \tilde{N}$, то векторы канонического базиса $\mathbf{t}_{\min}$ и $\mathbf{t}_{\max}$ есть единственные (с точностью до знака) направления, в которых нормальная кривизна достигает своих экстремумов $k_{\min}$ и $k_{\max}$. Эти направления и кривизны называют **главными**. Если в выбранной точке $\tilde{L} = \tilde{N}$, то нормальные кривизны во всех направлениях равны.

Поспешное заключение, что главные кривизны и направления находятся как собственные числа и векторы матрицы второй квадратичной формы, ошибочно: нельзя забывать о метрике.

Теорема (Rodrigues)**.** *Главные кривизны и главные направления есть собственные числа и собственные векторы оператора* $-\mathsf{A}$.

Доказательство. Симметричный оператор $-\mathsf{A}$ имеет ортонормированный базис из собственных векторов $\mathbf{v}_{\min}$ и $\mathbf{v}_{\max}$, соответствующих вещественным собственным числам $\lambda_{\min} \leqslant \lambda_{\max}$. Случай кратного корня окажется малоинтересен. Запишем вектор $-\mathsf{A}\boldsymbol{\xi}$ для произвольного единичного касательного вектора:

$$\boldsymbol{\xi} = \mathbf{v}_{\min}\cos\theta + \mathbf{v}_{\max}\sin\theta,$$
$$-\mathsf{A}\boldsymbol{\xi} = \mathbf{v}_{\min}\lambda_{\min}\cos\theta + \mathbf{v}_{\max}\lambda_{\max}\sin\theta.$$

Подставляя $\boldsymbol{\xi}$ на место $d\mathbf{r}$, имеем $\mathsf{A}\boldsymbol{\xi}$ на месте $d\mathbf{m}$. Нормальная кривизна в направлении $\boldsymbol{\xi}$ равна

$$k_{\text{н}}(\boldsymbol{\xi}) = -\frac{\boldsymbol{\xi}\cdot\mathsf{A}\boldsymbol{\xi}}{\boldsymbol{\xi}\cdot\boldsymbol{\xi}} = \lambda_{\min}\cos^2\theta + \lambda_{\max}\sin^2\theta.$$

Это выражение экстремально при углах, кратных $\frac{\pi}{2}$, то есть на направлениях $\pm\mathbf{v}_{\min}$ и $\pm\mathbf{v}_{\max}$. □

Теорема (Euler, 1760)**.** *Нормальная кривизна поверхности в направлении, составляющем угол θ с главным, равна $k_{\min}\cos^2\theta + k_{\max}\sin^2\theta$.*

Доказательство. Формула в конце предыдущего доказательства. □

Средняя и полная кривизна поверхности

Уравнение $-\mathsf{A}(d\mathbf{r}) = k\,d\mathbf{r}$ на собственные числа и векторы оператора $-\mathsf{A}$ скалярным умножением на $d\mathbf{r}$ преобразуется к $\mathtt{II}(d\mathbf{r}) = k\,\mathtt{I}(d\mathbf{r})$, поэтому главные кривизны проще найти по матрицам двух квадратичных форм, решая уравнение

$$\det(\mathtt{II} - k\,\mathtt{I}) = 0.$$

Определитель раскрывается в квадратное уравнение $k^2 - 2Hk + K = 0$ с коэффициентами

$$2H = \frac{EN - 2FM + GL}{EG - F^2} = \operatorname{tr}(-\mathsf{A}), \quad K = \frac{LN - M^2}{EG - F^2} = \det(-\mathsf{A}),$$

и корнями $k_{\min}$ и $k_{\max}$. Величины $H = \frac{1}{2}(k_{\min} + k_{\max})$ и $K = k_{\min}k_{\max}$ называются соответственно **средней** и **полной кривизной** поверхности в выбранной точке. Обратите внимание, что в формуле (5) скаляр K есть в точности полная кривизна.

Отыскание главных направлений

Найти главные направления $d\mathbf{r} = \mathbf{r}_u\,du + \mathbf{r}_v\,dv$ означает найти зависимость между du и dv, при которой $d\mathbf{m} = k\,d\mathbf{r}$. Умножая это уравнение отдельно на $\mathbf{r}_u$ и $\mathbf{r}_v$, получим систему

$$\begin{cases} L\,du + M\,dv = k(E\,du + F\,dv), \\ M\,du + N\,dv = k(F\,du + G\,dv). \end{cases}$$

Тогда

$$\begin{vmatrix} L\,du + M\,dv & E\,du + F\,dv \\ M\,du + N\,dv & F\,du + G\,dv \end{vmatrix} = 0,$$

что можно переписать в более компактном виде

$$\begin{vmatrix} -dv^2 & du\,dv & -du^2 \\ E & F & G \\ L & M & N \end{vmatrix} = 0.$$

Отсюда всегда можно найти два отношения $du : dv$, дающих искомые главные направления.

Типы точек и соприкасающийся параболоид

Значения главных кривизн определяют форму поверхности вблизи выбранной точки. В системе координат с началом в точке P поверхности и репером $(\mathbf{t}_{\min}, \mathbf{t}_{\max}, \mathbf{m})$ рассмотрим квадрику

$$2z = k_{\min}\,x^2 + k_{\max}\,y^2.$$

Это уравнение отличается от уравнения $z = f(x, y)$ исходной поверхности вблизи точки P величинами третьего порядка малости по x и y, поскольку является фактически разложением Тэйлора функции $f(x, y)$ до второго порядка. Поэтому среди всех квадрик, проходящих через точку P, эта — единственная, имеющая соприкосновение второго порядка с исходной поверхностью. Её называют **соприкасающимся параболоидом**. В соответствии с типом параболоида классифицированы

регулярные точки поверхности.

Кривизны	Квадрика	Тип точки
$k_{\min} = k_{\max} = 0$	Плоскость	Точка уплощения
$k_{\min} k_{\max} = 0$	Параболический цилиндр	Параболическая
$k_{\min} k_{\max} < 0$	Гиперболический параболоид	Гиперболическая
$k_{\min} k_{\max} > 0$	Эллиптический параболоид	Эллиптическая
$k_{\min} = k_{\max} \neq 0$	Параболоид вращения	Омбилическая

10.4. Внутренняя геометрия поверхности

Лекция 10 (10.04.20)

В этом разделе займёмся поиском ответов на два общих вопроса о поверхностях. Какие построения и величины определяются:

(1) заданием первой квадратичной формы;
(2) заданием первой и второй квадратичных форм?

Основным будет первый вопрос: занятие им составляет предмет *внутренней геометрии* поверхности. Ответ на второй вопрос структурно прост и технически попутен.

Изометричность и изгибание

Биективное отображение $\varphi\colon S \to S^*$ поверхностей называется **изометрией**, если оно сохраняет длины всех линий. Наглядное представление об изометриях даёт изгибание листа бумаги в куски цилиндра или конуса. Не менее интересен пример изометрии между кусками геликоида и катеноида.

Параметризации изометричных кусков $\mathbf{r}(u,v)$ и $\mathbf{r}^*(u,v)$ удобно согласовывать так, чтобы иметь тождество $\mathbf{r}^*(u,v) \equiv \varphi(\mathbf{r}(u,v))$. Тогда метрики $\mathtt{I}$ и $\mathtt{I}^*$ совпадают как функции u и v. Следовательно, все величины, выражающиеся через коэффициенты метрики и их производные любого порядка, неизменны (инвариантны) при изометриях. Перечислим некоторые из них:

- геодезическая кривизна линий;
- полная кривизна поверхности;
- параллельный перенос векторов вдоль линий.

Индексные обозначения

Если прежде мы могли позволить себе обозначать параметры и всякие коэффициенты различными буквами почти без индексов, то те-

перь характер возникающих формул требует перехода к более гибким и мощным индексным обозначениям.

Параметры поверхности	u^i	$u^1 = u$, $u^2 = v$
Операция дифференцирования	∂_i	$\partial_1 = \frac{\partial}{\partial u}$, $\partial_2 = \frac{\partial}{\partial v}$
Первые частные производные	$\mathbf{r}_i$	$\mathbf{r}_1 = \mathbf{r}_u$, $\mathbf{r}_2 = \mathbf{r}_v$
Вторые частные производные	$\mathbf{r}_{ij}$	$\mathbf{r}_{11} = \mathbf{r}_{uu}$, $\mathbf{r}_{12} = \mathbf{r}_{uv}$, $\mathbf{r}_{22} = \mathbf{r}_{vv}$
Компоненты метрики	g_{ij}	$g_{11} = E$, $g_{12} = F$, $g_{22} = G$
Компоненты «обратной» метрики	g^{ij}	(элементы обратной матрицы)
Компоненты формы II	b_{ij}	$b_{11} = L$, $b_{12} = M$, $b_{22} = N$
Компоненты оператора A	a^i_j	(выше явно не выписывались)

Какая будет выгода?

(1) Сокращение количества формул за счёт превращения нескольких похожих формул в одну формулу, имеющую «свободные» индекс(ы) в качестве параметров.
(2) Упрощение формул с использованием знака суммирования по «несвободным» индексам.
(3) Дальнейшее упрощение формул за счёт опускания (иногда многочисленных) знаков суммирования, достигаемое хитрыми соглашениями о расстановке индексов.

Для внимательного рассмотрения, вот примеры уже известных формул в новых обозначениях:

Смысл формулы	Формула	Индексы в формуле	
		спаренные	свободные
Определение метрики	$g_{ij} = \mathbf{r}_i \cdot \mathbf{r}_j$	нет	i, j
Определение формы II	$b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{m} = -\mathbf{r}_i \cdot \mathbf{m}_j$	нет	i, j
Касательный вектор	$\boldsymbol{\xi} = \xi^i \mathbf{r}_i$	i	нет
Координаты вектора $\boldsymbol{\xi}$	ξ^i	нет	i
Значения метрики	$\mathtt{I}(\boldsymbol{\xi}) = \boldsymbol{\xi} \cdot \boldsymbol{\xi} = g_{ij}\xi^i\xi^j$	i, j	нет
	$\mathtt{I}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \boldsymbol{\xi} \cdot \boldsymbol{\eta} = g_{ij}\xi^i\eta^j$	i, j	нет
Связь оператора A с формами I и II	$a^i_j = -g^{ik}b_{kj}$	k	i, j

Выводы о системе индексных обозначений:

- индексы делятся на нижние и верхние;
- наборы свободных индексов в левой и правой частях формулы должны совпадать;

- спаренный индекс — раз нижний и раз верхний — означает суммирование по всем значениям этого индекса.
- любой индекс можно молча заменять (всюду!) любой ещё не использованной буквой.

Принцип расположения каждого конкретного индекса проясняется при изучении тензорной алгебры и тензорного анализа.

Деривационные формулы

Исторически сложившаяся группа **деривационных формул** состоит из разложений по базису $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{m})$ в каждой точке поверхности частных производных самих этих векторов:

$$\mathbf{r}_{ij} = \Gamma^k_{ij}\,\mathbf{r}_k + b_{ij}\,\mathbf{m},$$
$$\mathbf{m}_j = a^i_j\,\mathbf{r}_i.$$

Последняя формула идентична определению оператора Родрига.

Неизвестные пока функции Γ^k_{ij} называют **коэффициентами связности**. Умножая это разложение скалярно на $\mathbf{r}_\ell$, получим $\mathbf{r}_{ij}\cdot\mathbf{r}_\ell = \Gamma^k_{ij}\,g_{k\ell}$. С другой стороны, $\partial_k(g_{ij}) = \mathbf{r}_{ik}\cdot\mathbf{r}_j + \mathbf{r}_{jk}\cdot\mathbf{r}_i$, поэтому в итоге Γ^k_{ij} выражаются через первые производные компонент метрики, а также компоненты «обратной» метрики, возникающие при решении системы линейных уравнений. При изометриях эти функции сохраняются и во внутренней геометрии поверхности они оказываются вездесущи.

Упражнение 9. *Получите явную формулу для Γ^k_{ij}, о которой идёт речь.*

Упражнение 10. *Покажите, что $\mathbf{r}_{11}\cdot\mathbf{r}_{22} - \mathbf{r}_{12}\cdot\mathbf{r}_{21}$ равно линейной комбинации вторых частных производных компонент метрики. Указание: рассмотрите подходящие производные $\partial_\ell(\mathbf{r}_{ij}\cdot\mathbf{r}_k)$.*

Упражнение 11. *Запишите $\partial_1(b_{2j}) - \partial_2(b_{1j})$ как линейную комбинацию произведений Γ^k_{ij} на b_{ij} с подходящими индексами. Указание: $b_{ij} = -\mathbf{r}_i\cdot\mathbf{m}_j$. Вот ответ в компактной, но необычной форме:*

$$\begin{vmatrix} \partial_1 & b_{1j} \\ \partial_2 & b_{2j} \end{vmatrix} = \begin{vmatrix} \Gamma^k_{1j} & b_{1k} \\ \Gamma^k_{2j} & b_{2k} \end{vmatrix}.$$

Здесь нужно формально раскрыть определители, чтобы получить корректную индексную запись.

Выражение геодезической кривизны

Вычислим кривизну линии $u^i(s)$ на поверхности $\mathbf{r}(u^1, u^2)$. Дифференцируем:

$$\dot{\mathbf{r}} = \dot{u}^i\,\mathbf{r}_i, \qquad \ddot{\mathbf{r}} = \ddot{u}^k\,\mathbf{r}_k + \dot{u}^i\dot{u}^j\,\mathbf{r}_{ij}.$$

Подставляя сюда $\mathbf{r}_{ij}$ из деривационных формул и разделяя компоненты, получим

$$\mathbf{k}_\mathrm{H} = b_{ij}\dot{u}^i\dot{u}^j\,\mathbf{m}, \qquad \mathbf{k}_\Gamma = w^k\,\mathbf{r}_k,$$

причём функции $w^k = \ddot{u}^k + \Gamma^k_{ij}\dot{u}^i\dot{u}^j$ сохраняются при изометриях. Наконец, из формул $k_\Gamma = |\mathbf{k}_\Gamma \times \dot{\mathbf{r}}|$ и $\mathbf{r}_1 \times \mathbf{r}_2 = \sqrt{\det \mathtt{I}}\;\mathbf{m}$ находим

$$k_\Gamma = \begin{vmatrix} \dot{u}^1 & w^1 \\ \dot{u}^2 & w^2 \end{vmatrix} \sqrt{\det \mathtt{I}}.$$

Теорема. *Геодезическая кривизна линии на поверхности сохраняется при изометриях.* □

Следствие. *Изометрии переводят геодезические в геодезические.* □

Theorema Egregium

Перемножая скалярно деривационные формулы для $\mathbf{r}_{1i}$ и $\mathbf{r}_{j2}$, видим

$$\mathbf{r}_{1i}\cdot\mathbf{r}_{j2} = \Gamma^k_{1i}\Gamma^\ell_{j2}\,\mathbf{r}_k\cdot\mathbf{r}_\ell + b_{1i}b_{j2}\,\mathbf{m}\cdot\mathbf{m} = \Gamma^k_{1i}\Gamma^\ell_{j2}\,g_{k\ell} + b_{1i}b_{j2}.$$

Разность двух формул, получаемых отсюда при $i \neq j$, даёт

$$\det \mathtt{II} = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = \begin{vmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{r}_{22} \end{vmatrix} - \begin{vmatrix} \Gamma^k_{11} & \Gamma^\ell_{12} \\ \Gamma^k_{21} & \Gamma^\ell_{22} \end{vmatrix} g_{k\ell}.$$

Выполнившие предпоследнее упражнение могут считать, что они доказали великую теорему, за которой закрепилось оригинальное латинское название, данное удивлённым первооткрывателем.

Теорема (Gauss; 1827)**.** *Полная кривизна поверхности сохраняется при изометриях.*

Доказательство. Числитель формулы

$$K = \frac{\det \mathtt{II}}{\det \mathtt{I}}$$

удалось выразить через компоненты $\mathtt{I}$ и их производные. □

Таким образом, полную кривизну поверхности можно измерить путём измерения расстояний и углов на ней. Мотивацией для этих исследований Гаусса была геодезия.

Оформление поверхности в пространстве

Рассмотрим две поверхности S и S^{**} в пространстве, обладающие такими параметризациями $\mathbf{r}(u,v)$ и $\mathbf{r}^{**}(u,v)$, что совпадают первые и вторые квадратичные формы.

Теорема (Петерсон, 1853)**.** *Если* $\mathtt{I} = \mathtt{I}^{**}$ *и* $\mathtt{II} = \mathtt{II}^{**}$ *как функции от* u, v *для поверхностей* S *и* S^{**}, *то найдётся движение* $\varphi\colon \mathbb{R}^3 \to \mathbb{R}^3$ *пространства как твёрдого тела, совмещающее эти поверхности:* $\mathbf{r}^{**}(u,v) \equiv \varphi(\mathbf{r}(u,v))$.

Доказательство. Выберем любую точку $P = \mathbf{r}(u,v)$ и соответствующую ей точку $P^{**} = \mathbf{r}^{**}(u,v)$. Сдвигом в пространстве совместим эти точки. Ввиду совпадения метрик найдётся поворот, совмещающий реперы $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{m})$ и $(\mathbf{r}_1^{**}, \mathbf{r}_2^{**}, \mathbf{m}^{**})$ в этих точках. Пусть φ будет композицией указанных сдвига и поворота.

Выберем линии $u = u(t)$, $v = v(t)$ на S и на S^{**}, проходящие при $t = 0$ через точку $P = P^{**}$. Вдоль них получим уравнения

$$\begin{aligned} \tfrac{d}{dt}(\mathbf{r}_i) &= \tfrac{d}{dt}(u^j)\,\mathbf{r}_{ij} \;= \tfrac{d}{dt}(u^j)\Gamma_{ij}^k\,\mathbf{r}_k + \tfrac{d}{dt}(u^j)b_{ij}\,\mathbf{m}, \\ \tfrac{d}{dt}(\mathbf{m}) &= \tfrac{d}{dt}(u^j)\,\mathbf{m}_j = \tfrac{d}{dt}(u^j)a_j^i\,\mathbf{r}_i \end{aligned}$$

и такую же точно систему линейных дифференциальных уравнений для $(\mathbf{r}_1^{**}, \mathbf{r}_2^{**}, \mathbf{m}^{**})$, поскольку условия $\mathtt{I} = \mathtt{I}^{**}$ и $\mathtt{II} = \mathtt{II}^{**}$ гарантируют идентичность деривационных формул для S и S^{**}. Благодаря единственности решения системы при заданных начальных условиях, наши реперы совпадают всюду вдоль выбранных линий. Тогда приращение радиус-вектора

$$\mathbf{r}(t) - \mathbf{r}(0) = \int_0^t \tfrac{d}{dt}(u^j)\,\mathbf{r}_j\,dt$$

совпадает с приращением $\mathbf{r}^{**}(t) - \mathbf{r}^{**}(0)$. Поэтому поверхности совмещены целиком, если употребить одинаковые области определения функций $\mathbf{r}(u,v)$ и $\mathbf{r}^{**}(u,v)$. □

Раз первая и вторая квадратичные формы полностью определяют вид поверхности в пространстве, естественно спросить, как сделанный выбор $\mathtt{I}$ ограничивает возможности для $\mathtt{II}$. Ограничения сводятся к полученной выше **формуле Гаусса**, выражающей $\det \mathtt{II}$ через метрику и её производные, и к **формулам Петерсона — Майнарди — Кодацци**, связывающим две комбинации производных от b_{ij} с самими b_{ij} посредством коэффициентов связности и приведённым в последнем из упражнений на деривационные формулы. Оставим этот факт без доказательства.

10.5. Абсолютная производная и параллельный перенос

Векторные поля на поверхности

Векторные поля в пространстве уже встречались вам в курсе механики и подробно изучаются в курсе основ математического анализа. Однако в дифференциальной геометрии, говоря о **векторном поле** на поверхности S, чаще всего подразумевают не только то, что оно имеет смысл лишь на S, но и то, что оно касательно к S в каждой точке.

На параметризованной поверхности $\mathbf{r} = \mathbf{r}(u^1, u^2)$ есть первородные векторные поля: частные производные $\mathbf{r}_i$. Выбор же на ней другого хорошего векторного поля $\mathbf{v} = v^i\,\mathbf{r}_i$ означает, что в каждой точке поверхности выбран касательный вектор, координаты которого в базисе $(\mathbf{r}_1, \mathbf{r}_2)$ есть гладкие функции параметров u^i.

Векторные поля можно складывать, просто складывая векторы в соответствующих точках, что алгебраически эквивалентно сложению координатных функций:

$$(v^i\,\mathbf{r}_i) + (w^i\,\mathbf{r}_i) = (v^i + w^i)\,\mathbf{r}_i.$$

Аналогично выполняется умножение на скаляры, поэтому множество $\mathrm{Vect}(S)$ векторных полей на данной поверхности S есть линейное пространство.

Поля *некасательных* векторов полезны при рассмотрении поверхности, вложенной в $\mathbb{R}^3$. Мы регулярно использовали единичное нормальное поле $\mathbf{m}$. Скалярное произведение двух векторных полей есть функция на поверхности, а в иных терминах — **скалярное поле**. Значения его вычисляются скалярным перемножением векторов данных полей в соответствующих точках. Эту операцию мы проводили, определяя первую и вторую квадратичные формы. Аналогично определено векторное произведение полей; если поля $\mathbf{p}$ и $\mathbf{q}$ касательные, то $\mathbf{p} \times \mathbf{q}$ есть нормальное поле к поверхности.

Выбирая на поверхности единичное касательное поле $\mathbf{p}$, то есть $|\mathbf{p}| \equiv 1$, и ортогональное ему единичное касательное поле $\mathbf{q} = \mathbf{m} \times \mathbf{p}$, получим подвижный ОНБ $(\mathbf{p}, \mathbf{q}, \mathbf{m})$.

Упражнение 12. *Выведите тождество*

$$(\mathbf{m}, \mathbf{m}_1, \mathbf{m}_2) = \mathbf{p}_1 \cdot \mathbf{q}_2 - \mathbf{p}_2 \cdot \mathbf{q}_1.$$

Индексы на $\mathbf{m}$, $\mathbf{p}$, *и* $\mathbf{q}$ *обозначают частные производные по* u^i.

Указание: введите мгновенные угловые скорости $\boldsymbol{\omega}_1$ *и* $\boldsymbol{\omega}_2$ *вращения репера* $(\mathbf{p}, \mathbf{q}, \mathbf{m})$ *при смещениях вдоль* $\mathbf{r}_1$ *и* $\mathbf{r}_2$, *тогда* $\mathbf{m}_i = \boldsymbol{\omega}_i \times \mathbf{m}$ *и*

т. п., так что все задействованные векторы выразятся через $\mathbf{p}$, $\mathbf{q}$, $\boldsymbol{\omega}_1$ *и* $\boldsymbol{\omega}_2$*; далее достаточно тождеств векторной алгебры.*

Абсолютная производная векторного поля

Посчитаем производную от касательного поля $\mathbf{v} = v^i\,\mathbf{r}_i$ по направлению координатной линии:

$$\partial_j(\mathbf{v}) = \partial_j(v^i\,\mathbf{r}_i) = \partial_j(v^i)\,\mathbf{r}_i + v^i\,\mathbf{r}_{ij} = \bigl(\partial_j(v^k) + \Gamma^k_{ij}v^i\bigr)\,\mathbf{r}_k + b_{ij}v^i\,\mathbf{m}.$$

Присутствие нормальной компоненты портит картину. В то время как вектор $\mathbf{v}$ принадлежит внутренней геометрии, поскольку функции v^i зависят лишь прямо от u^j и не замечают изгибаний, его обычная производная $\partial_j(\mathbf{v})$ чувствительна к положению поверхности в пространстве. Поэтому интересным объектом является

$$\nabla_j(\mathbf{v}) = \bigl(\partial_j(v^k) + \Gamma^k_{ij}v^i\bigr)\,\mathbf{r}_k,$$

то есть проекция $\partial_j(\mathbf{v})$ на касательную плоскость, называемая **абсолютной производной** векторного поля $\mathbf{v}$. Это тоже получается векторное поле. В выборе обозначений для его координат есть некая трудность; остановимся на одном из распространённых вариантов:

$$v^k_{;j} = \partial_j(v^k) + \Gamma^k_{ij}v^i.$$

При этом странный нижний индекс в $v^k_{;j}$ указывает на параметр, по которому взята производная, а точка с запятой указывает на её абсолютность.

Абсолютное дифференцирование ∇_j линейно:

$$\nabla_j(\alpha\,\mathbf{v} + \beta\,\mathbf{w}) = \alpha\nabla_j(\mathbf{v}) + \beta\nabla_j(\mathbf{w}),$$

так что оно является линейным оператором на пространстве $\mathrm{Vect}(S)$.

Дифференцирование одного векторного поля вдоль другого

Абсолютные дифференцирования можно скомпоновать в абсолютный дифференциал $\nabla = \nabla_j\,du^j$, но это объект непростой. Вместо него удобнее перейти к производным по любому векторному полю. Они берутся в каждой точке по тому же правилу, что и обычные производные по вектору в анализе. Итак, выбирая векторное поле $\mathbf{z} = z^j\,\mathbf{r}_j$, получаем соответствующий оператор дифференцирования

$$\nabla_{\mathbf{z}} = z^j\nabla_j\colon\ \mathrm{Vect}(S) \to \mathrm{Vect}(S).$$

Применение этого оператора к векторному полю $\mathbf{v} = v^i\,\mathbf{r}_i$ даёт новое векторное поле

$$\nabla_{\mathbf{z}}\mathbf{v} = z^j v^k_{;j}\,\mathbf{r}_k.$$

Параллельный перенос на поверхности

Понятие параллельного переноса векторов на поверхности имеет две стороны, геометрическую и аналитическую; конечно же, они удачно согласуются.

На поверхности возьмём линию $\mathbf{r}(t)$ и касательный вектор $\mathbf{q}(0)$ в одной её точке. Есть только одна возможность так определить векторы $\mathbf{q}(t)$, касательные к поверхности в соответствующих точках $\mathbf{r}(t)$, чтобы, двигаясь вдоль линии, всегда иметь $\mathbf{q}'(t) \parallel \mathbf{m}(t)$. Это и есть параллельный перенос вдоль линии. Результат переноса в общем случае зависит от пути. Название объясняется тем, что при этом вектор $\mathbf{q}+d\mathbf{q}$ получается параллельным переносом в пространстве вектора $\mathbf{q}$, приложенного в точке $\mathbf{r}(t)$, в точку $\mathbf{r}(t+dt)$. Если на самом деле векторное поле $\mathbf{q}$ определено не только вдоль линии, а на куске поверхности, то абсолютная производная от $\mathbf{q}(t)$ вдоль $\mathbf{r}'(t)$ по определению обязана равняться нулю.

Теорема. *Параллельный перенос на поверхности инвариантен относительно изометрий.*

Доказательство. Следует из инвариантности абсолютной производной. □

Неформально говоря, перенести затем изогнуть, или изогнуть затем перенести, — разницы нет.

Теорема. *При параллельном переносе вдоль линии на поверхности касательная плоскость к поверхности движется как твёрдое тело.*

Доказательство. Выберем пару касательных векторов $\mathbf{p}$, $\mathbf{q}$ и посмотрим, как они перенесутся:

$$(\mathbf{p} \cdot \mathbf{q})' = \mathbf{p}' \cdot \mathbf{q} + \mathbf{p} \cdot \mathbf{q}'.$$

В правой части обе производные коллинеарны нормали $\mathbf{m}$ к поверхности, а потому $(\mathbf{p} \cdot \mathbf{q})' \equiv 0$, что означает сохранение всех длин и углов при движении касательной плоскости. □

Упражнение 13. *Единичное касательное поле геодезической параллельно. (Это значит, что оно получается параллельным переносом себя вдоль геодезической из любой её начальной точки.)*

Обнесение по контуру и кривизна

Возьмём теперь *замкнутую* линию C на поверхности S — их часто называют **контурами** — и исследуем состояние касательной плоскости к S в выбранной точке C после обнесения её параллельно вокруг контура. Будучи твёрдым телом, касательная плоскость может лишь повернуться на какой-то угол $\Delta\varphi$ относительно начального положения; его и будем искать. Положительным считаем угол поворота от $\mathbf{r}_1$ к $\mathbf{r}_2$. Контур обходим в положительном направлении, как обычно в анализе. Наглядно: область остаётся слева, нормаль $\mathbf{m} = \mathbf{e}(\mathbf{r}_1 \times \mathbf{r}_2)$ направлена вверх.

Зафиксируем любое единичное векторное поле $\mathbf{p}$ как начало отсчёта углов в каждой касательной плоскости. Какой бы касательный вектор мы не несли параллельно вдоль нашего контура, дифференциал угла $d\varphi$ будет одним и тем же, зависящим только от $\mathbf{p}$ и контура. Можно проверить, что на самом деле $d\varphi = -\mathbf{q} \cdot d\mathbf{p}$, где $\mathbf{q} = \mathbf{m} \times \mathbf{p}$.

Теорема. *Угол поворота вектора, параллельно обнесённого по замкнутому контуру на поверхности, равен интегралу от полной кривизны по области D, охваченной этим контуром:*

$$\Delta\varphi = \iint_D K\, dS.$$

Доказательство. Подставим $d\varphi = -\mathbf{q} \cdot d\mathbf{p}$ и преобразуем полученный контурный интеграл в поверхностный по формуле Стокса:

$$\begin{aligned}\Delta\varphi = \oint_C d\varphi = -\oint_C \mathbf{q} \cdot d\mathbf{p} &= -\oint_C (\mathbf{q} \cdot \mathbf{p}_1\, du^1 + \mathbf{q} \cdot \mathbf{p}_2\, du^2) \\ &= \iint_D (\mathbf{p}_1 \cdot \mathbf{q}_2 - \mathbf{p}_2 \cdot \mathbf{q}_1)\, du^1\, du^2.\end{aligned}$$

По упражнениям 12 и 8 перепишем подынтегральную функцию:

$$\mathbf{p}_1 \cdot \mathbf{q}_2 - \mathbf{p}_2 \cdot \mathbf{q}_1 = \mathbf{m} \cdot (\mathbf{m}_1 \times \mathbf{m}_2) = K\, \mathbf{m} \cdot (\mathbf{r}_1 \times \mathbf{r}_2) = K\sqrt{\det \mathrm{I}}.$$

Наконец, $\sqrt{\det \mathrm{I}}\, du^1\, du^2$ есть в точности элемент площади поверхности dS. □

Локальная теорема Гаусса — Бонне

Рассмотрим теперь поведение касательного вектора $\mathbf{t} = \dot{\mathbf{r}}$ к контуру $\mathbf{r}(s)$ при обходе его с единичной скоростью. При однократном обходе угол между $\mathbf{t}$ и выбранным началом отсчёта $\mathbf{p}$ получает приращение 2π. Оно складывается из поворота касательной плоскости на угол $\Delta\varphi$

и приращения $\Delta\psi$ угла между $\mathbf{t}$ и каким-то параллельно переносимым вектором. Считаем контур кусочно гладким, то есть допускаем угловые точки со скачками $\Delta_i\psi$ направления $\mathbf{t}$, тогда

$$\Delta\psi = \oint_C d\psi + \sum \Delta_i\psi.$$

На каждом гладком куске контура имеем $d\psi = -(\mathbf{m}\times\mathbf{p})\cdot d\mathbf{p}$, как и выше, причём роль $\mathbf{p}$ может играть любое единичное касательное поле, например $\mathbf{t}$. Поэтому

$$d\psi = -(\mathbf{m}\times\mathbf{t})\cdot\dot{\mathbf{t}}\,ds = -(\mathbf{m},\mathbf{t},\dot{\mathbf{t}})\,ds = (\ddot{\mathbf{r}},\dot{\mathbf{r}},\mathbf{m})\,ds = k_\Gamma\,ds,$$

где геодезическая кривизна берётся вместе со знаком.

Теорема. *Для каждого контура C, ограничивающего элементарный кусок поверхности D, полная кривизна поверхности и геодезическая кривизна контура связаны формулой*

$$\iint_D K\,dS + \oint_C k_\Gamma\,ds + \sum \Delta_i\psi = 2\pi.$$

Доказательство. Предшествующие рассуждения и предыдущая теорема. □

Как интересный частный случай отсюда следует формула суммы углов геодезического многоугольника. Для геодезического треугольника T с внутренними углами α, β, γ получается

$$\alpha+\beta+\gamma = \pi + \iint_T K\,dS.$$

Минимизация площади поверхности при заданном крае

Изучим вопрос, имеющий непосредственные приложения в разных областях науки: можно ли уменьшить площадь конкретной поверхности, немного шевеля её вдали от края? Если нет, то поверхность называют **минимальной**. Такую форму принимают мыльные плёнки и разнообразные двумерные молекулярные и биологические структуры.

Теорема. *Средняя кривизна минимальной поверхности равна нулю во всех точках.*

Доказательство. Возьмём кусок D поверхности с параметризацией $\mathbf{r}(u^i)$. Зададим шевеление как небольшое смещение в направлении нормали, полагая $\mathbf{r}^\varepsilon = \mathbf{r} - \varepsilon\eta\mathbf{m}$. Гладкая функция $\eta = \eta(u^i)$ должна зануляться на краю куска, но в остальном произвольна. Малый параметр

деформации ε не зависит от точки поверхности, а причина выбора знака минус скоро прояснится.

Площадь S исходной поверхности минимальна, когда она меньше всех площадей S^ε близких к ней поверхностей, что возможно только если $\frac{1}{\varepsilon}(S^\varepsilon - S) \to 0$ при $\varepsilon \to 0$ для любой функции η. Подобные пределы называют вариационными производными.

Итак, мы ищем предел выражения

$$\frac{1}{\varepsilon}(S^\varepsilon - S) = \frac{1}{\varepsilon} \iint_D \bigl(\sqrt{\det g^\varepsilon} - \sqrt{\det g}\bigr)\, du^1\, du^2.$$

Для этого сперва найдём зависимость метрики от ε. Два первых слагаемых у частных производных

$$\mathbf{r}_i^\varepsilon = \mathbf{r}_i - \varepsilon\eta\mathbf{m}_i - \varepsilon\eta_i\mathbf{m}$$

касательны к поверхности, а третье направлено по нормали; поэтому в выражении компонент метрики $g_{ij}^\varepsilon = \mathbf{r}_i^\varepsilon \cdot \mathbf{r}_j^\varepsilon$ четыре слагаемых равны нулю. Из оставшихся, два содержат ε^2, и мы получаем

$$g_{ij}^\varepsilon = \mathbf{r}_i \cdot \mathbf{r}_j - \varepsilon\eta(\mathbf{r}_i \cdot \mathbf{m}_j + \mathbf{m}_i \cdot \mathbf{r}_j) + o(\varepsilon) = g_{ij} + 2\varepsilon\eta b_{ij} + o(\varepsilon).$$

Знак минус поглотило определение второй квадратичной формы.

Теперь вычислим определитель:

$$\begin{aligned}\det g^\varepsilon &= g_{11}^\varepsilon g_{22}^\varepsilon - (g_{12}^\varepsilon)^2 = (g_{11} + 2\varepsilon\eta b_{11})(g_{22} + 2\varepsilon\eta b_{22}) - (g_{12} + 2\varepsilon\eta b_{12})^2 \\ &= (g_{11}g_{22} - g_{12}^2) + 2\varepsilon\eta(g_{11}b_{22} - 2g_{12}b_{12} + g_{22}b_{11}) + o(\varepsilon).\end{aligned}$$

Распознаём во второй скобке числитель формулы, которой мы ввели среднюю кривизну, а в первой — её знаменатель. Значит,

$$\det g^\varepsilon = \det g \cdot \bigl(1 + 4H\varepsilon\eta + o(\varepsilon)\bigr).$$

Извлекая корень, воспользуемся разложением $\sqrt{1+2x} = 1 + x + o(x)$ для упрощения:

$$\sqrt{\det g^\varepsilon} = \sqrt{\det g} \cdot \bigl(1 + 2H\varepsilon\eta + o(\varepsilon)\bigr).$$

Поэтому искомый предел равен

$$\iint_D 2H\eta\, dS.$$

Равенство этого интеграла нулю для всех функций η означает, что $H = 0$ тождественно на всей поверхности. □

Глобальная теорема Гаусса — Бонне

Компактную гладкую поверхность без края всегда можно **триангулировать**, то есть разбить на треугольные кусочки. Число вершин, ребёр (сторон) и граней (треугольников) в таких случаях принято обозначать буквами V, E и F. Также поступают при разбиениях на многоугольники.

Упражнение 14. *Проверьте, что число $V - E + F$ зависит только от поверхности, а не от способа её разбиения на многоугольники.*

Это число называют **эйлеровой характеристикой** поверхности и обозначают через $\chi = \chi(S)$. Например, у сферы $\chi = 2$, а у тора $\chi = 0$.

Теорема. *Интеграл по всей поверхности от её полной кривизны пропорционален её эйлеровой характеристике, а именно,*

$$\iint_S K\, dS = 2\pi\chi(S)$$

для всякой компактной гладкой поверхности S.

Доказательство. Сделаем триангуляцию на геодезические треугольники. Интеграл по поверхности равен сумме интегралов по всем этим треугольникам. На каждом треугольнике

$$\iint_T K\, dS = \alpha + \beta + \gamma - \pi.$$

В сумме этих равенств по всем F треугольникам получим 2π вокруг каждой из V вершин, так что справа будет $\pi(2V - F)$. Затем заметим, что $3F = 2E$, ибо каждая сторона принадлежит ровно двум треугольникам, а у каждого треугольника ровно три стороны. Поэтому $2V - F$ превращается в $2(V - E + F) = 2\chi$. □

Произвольные деформации поверхности, в отличие от изгибаний, меняют её кривизну. Из этой теоремы следует, что в отсутствие разрывов и склеек интеграл от полной кривизны сохраняется, поскольку сохраняется эйлерова характеристика.

Глава 11. ГРУППЫ И АЛГЕБРЫ

11.1. Примеры разных типов групп

Лекция 11 (17.04.20)

Определение группы

Понятие группы — одно из основных в алгебре. На протяжении нашего курса мы неоднократно сталкивались с разнообразными группами, но по причинам практического характера сторонились как абстракции самого понятия, так и дальнейших примеров.

Главный источник групп в математике и её приложениях — преобразования различных объектов. Для преобразований характерно: наличие композиции, которая по самой своей природе ассоциативна; тождественного преобразования (ничего-не-делания); обратимости каждого преобразования. Эти три свойства и составляют аксиомы группы.

Определение (Cayley, 1858). **Группой** $\langle \mathcal{G}, \cdot \rangle$ называют множество $\mathcal{G}$ с одной бинарной операцией, обладающей свойствами:

- ассоциативность;
- есть нейтральный элемент $e \in \mathcal{G}$;
- каждый элемент обратим.

Подгруппой группы $\mathcal{G}$ называют её подмножество $\mathcal{H}$, являющееся группой с той же операцией. Это выражают записью $\mathcal{H} \leqslant \mathcal{G}$.

Вообще, операцию в группе называют **умножением**, а нейтральный элемент — **единицей**, и обозначают обратный к $g \in \mathcal{G}$ через g^{-1}. Если же операция в рассматриваемой группе коммутативна, то её обычно называют **сложением**, а нейтральный элемент — **нулём**, и обозначают обратный к $g \in \mathcal{G}$ через $-g$. Примеры именно этого класса групп мы уже видели, изучая кольца, поля, линейные пространства и алгебры. Пересмотрите списки аксиом — всё это коммутативные группы (их называют **абелевыми**) с различными дополнительными операциями «умножения».

Группы чрезвычайно разнообразны. Они бывают **конечные** и **бесконечные** просто как множества, **дискретные** и **непрерывные** по характеру преобразований, могут строиться из кусков совершенно разных типов. Мы рассмотрим несколько примеров из разных классов групп, попутно отмечая основные понятия этой области математики.

Группы поворотов плоскости

Число элементов конечной группы называют **порядком** этой группы. Следует отметить, что слово «порядок» в математике перегружено смыслами, что может приводить к путанице.

Пример. Группа порядка один обладает лишь нейтральным элементом, за что справедливо называется **тривиальной**.

Пример. Множество $\sqrt[n]{1}$ комплексных корней степени n из единицы есть группа (порядка n) относительно умножения комплексных чисел. В ней каждый элемент является степенью одного фиксированного корня $\varepsilon = \exp \frac{2\pi i}{n}$. Группу с таким свойством называют **циклической**. Всякая циклическая группа коммутативна, поскольку перемножение степеней сводится к сложению показателей: $\varepsilon^m \varepsilon^n = \varepsilon^{m+n}$. Это вообще простейшие группы; сюда входят и самые маленькие нетривиальные.

Порядком элемента группы любого типа — хоть конечной, хоть бесконечной — называют наименьший порядок его степени, равной нейтральному элементу.

Пример. В циклической группе комплексных корней из единицы степени 12 имеются: по одному элементу порядков 1 и 2; по два элемента порядков 3, 4 и 6; четыре элемента порядка 12.

рисунок

Пример. Комплексные числа модуля 1 образуют группу относительно умножения комплексных чисел. Она бесконечна и коммутативна.

Пример. Повороты плоскости вокруг одной фиксированной точки образуют группу. Это одна из простейших непрерывных групп.

Вспомним интерпретацию комплексных чисел как преобразований плоскости. Числу $e^{i\varphi}$ соответствует поворот на угол φ вокруг нуля. Группы из последних двух примеров, хотя и построены на разных множествах, по сути, лишь этим и отличаются, а по *структуре* совпадают. Такую пару групп называют **изоморфными**.

Группа симметрий прямоугольника

Пример (прямоугольник). Изучим движения плоскости, переводящие в себя заданный прямоугольник, притом не квадрат. Всего их четыре: нейтральный e, два отражения s_1, s_2, поворот r на 180 градусов.

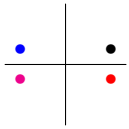
рисунок

Поскольку группы придуманы, чтобы описывать множества преобразований, для которых естественна операция композиции, запишем

все композиции двух симметрий прямоугольника в виде таблицы умножения (таблицы Кэли) его группы:

	e	r	s_1	s_2
e	e	r	s_1	s_2
r	r	e	s_2	s_1
s_1	s_1	s_2	e	r
s_2	s_2	s_1	r	e

e = тождественный
r = вращение на 180°
s_1 = отражение
s_2 = отражение



Видно, что тут нет элемента, степенями которого являлись бы три остальных: все квадраты равны e. Это самая маленькая нециклическая группа; её часто называют **четверная группа Клейна**.

Справа показаны множества точек плоскости, в которые элементы группы переводят чёрную точку. Такие множества называют **орбитами**. Перенумеровав вершины прямоугольника, можно записать действие каждого элемента группы. Нейтральный элемент оставляет на месте все вершины; закодируем его как $e = (1)(2)(3)(4)$. Отражение переставляет вершины парами: $s_1 = (14)(23)$ и $s_2 = (12)(34)$. Поворот тоже переставляет вершины парами: $r = (13)(24)$.

Группа симметрий равностороннего треугольника

рисунок

Пример (треугольник). Аналогично изучим движения плоскости, переводящие в себя равносторонний треугольник. Отражений тут три, а поворотов два. Выпишем таблицу умножения в этой группе:

	e	r_1	r_2	s_1	s_2	s_3
e	e	r_1	r_2	s_1	s_2	s_3
r_1	r_1	r_2	e	s_2	s_3	s_1
r_2	r_2	e	r_1	s_3	s_1	s_2
s_1	s_1	s_3	s_2	e	r_2	r_1
s_2	s_2	s_1	s_3	r_1	e	r_2
s_3	s_3	s_2	s_1	r_2	r_1	e

e = тождественный
r_1 = вращение на 120°
r_2 = вращение на 240°
s_1 = отражение
s_2 = отражение
s_3 = отражение

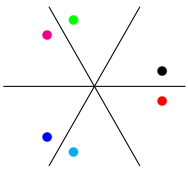


Таблица не симметрична относительно диагонали, потому что преобразования здесь не коммутируют! Это самая маленькая некоммутативная группа. Она изоморфна группе перестановок $\mathbb{S}_3$, обсуждаемой ниже; чтобы увидеть это, нужно записать, как её элементы переставляют вершины треугольника.

Упражнение. *Сколько симметрий у квадрата? Выпишите таблицу умножения его группы.*

Упражнение. *Сколько симметрий у правильного n-угольника? Опишите их.*

Группа перестановок

Пример. Строя теорию определителей, мы начали с множества $\mathbb{S}_n$ перестановок множества $\Omega = \{1, 2, \ldots, n\}$. Поскольку каждая перестановка есть биективное отображение $\Omega \to \Omega$, имеется естественная композиция перестановок, как отображений, и все аксиомы группы выполнены. Порядок группы перестановок $\mathbb{S}_n$ равен $n!$. При $n \geqslant 3$ она некоммутативна. Вычислим оба произведения для перестановок:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

Чтобы не запутаться, оговорим порядок записи композиции: перестановка $\sigma\tau$ определена правилом $\sigma\tau(k) = \sigma(\tau(k))$. Тогда

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \tau\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

На ранней стадии становления теории групп, развившейся из исследований разрешимости алгебраических уравнений в радикалах, под группой понимали подгруппу группы перестановок (Galois, 1828).

Теорема (Cayley, 1858)**.** *Каждая конечная группа из n элементов изоморфна подгруппе группы перестановок $\mathbb{S}_n$.*

Идея доказательства. Перенумеруем элементы группы в произвольном порядке: $g_1, \ldots, g_n$. Для каждого элемента a определим перестановку σ_a: если $ag_i = g_j$, то полагаем $\sigma_a(i) = j$. Из аксиом группы можно вывести, что действительно получится перестановка; примеры этого видны в приведённых выше таблицах умножения групп. Остаётся проверить, что $\sigma_{ab} = \sigma_a\sigma_b$. □

Конечные группы отражений и поворотов

Каждая прямая на плоскости Π задаёт операцию отражения. Единственное отражение s вместе с нейтральным элементом $e = s^2$ составляют (циклическую) группу порядка 2. Возьмём в Π набор прямых и рассмотрим всевозможные композиции отражений, ими заданных.

Упражнение. *Композиция двух отражений является поворотом, переносом, либо нейтральным в зависимости от взаимного расположения двух «зеркал».*

Композиции нескольких отражений ведут себя хаотично, если набросать прямые случайным образом. Однако при особых условиях возникают небольшие конечные группы. Их легко обнаружить как груп-

пы симметрий простых геометрических фигур. Все допускаемые здесь преобразования сохраняют расстояния между точками и называются **движениями**. При этом отражения меняют ориентацию плоскости, за что их относят к несобственным движениям.

Пример. Два отражения относительно перпендикулярных прямых порождают группу симметрий прямоугольника. Три отражения относительно прямых с общей точкой, делящих плоскость на шесть равных углов, порождают группу симметрий треугольника. С этими двумя группами мы познакомились выше.

Следующая группа этой серии получается из четырёх прямых, а также возникает как группа симметрий квадрата.

Аналогично для любого положительного целого числа $n \geqslant 3$ набор из n прямых на плоскости, проходящих через одну точку и делящих плоскость на $2n$ равных углов, порождает группу симметрии правильного n-угольника. Композиция двух различных отражений вращает плоскость вокруг центра P, причём различных поворотов получается $n-1$, а всего в группе $2n$ элементов. Её называют **группой диэдра** и обозначают через D_n. Повороты вместе с тождественным преобразованием образуют в ней циклическую подгруппу C_n.

Упражнение. *Получите группу диэдра D_n, начав всего с двух отражений.*

Пример. Аналогично предыдущему примеру рассмотрим группу, порождённую зеркальными отражениями трёхмерного пространства относительно плоскостей какого-то «хорошего» набора. Конечные группы получаются для наборов отражающих плоскостей, связанных с правильной n-угольной призмой или с правильным многогранником — тетраэдром, кубом или октаэдром, додекаэдром или икосаэдром.

Так получаются семейства конечных групп, порождённых отражениями, и в пространствах большего числа измерений (Coxeter, 1934).

Плоские решётки и орнаменты

Пример. Возьмём на плоскости ненулевой вектор $\mathbf{v}$ и обозначим через T параллельный перенос, то есть преобразование плоскости, заданное правилом $T(\mathbf{a}) = \mathbf{a} + \mathbf{v}$. Тогда перенос на $-\mathbf{v}$ является обратным преобразованием T^{-1}. Более того, для каждого целого числа n степень T^n является переносом на вектор $n\mathbf{v}$, причём $T^m T^n = T^{m+n}$. Таким образом, множество $\{T^n \mid n \in \mathbb{Z}\}$ образует бесконечную группу переносов, изоморфную группе целых чисел по сложению.

Это простейший пример бесконечной дискретной группы: если взять одну точку плоскости и отметить все точки, куда она переходит при таких переносах, то получится дискретное множество (каждая его точка изолирована).

Пример. Все параллельные переносы плоскости образуют бесконечную коммутативную группу $\mathsf{T}(\mathbb{R}^2)$. Выбрав базис $\{\mathbf{v}_1, \mathbf{v}_2\}$, мы можем задавать произвольный вектор его координатами в этом базисе. Значит, каждому переносу соответствует пара вещественных чисел.

Пример. Как и в предыдущем примере, выберем базис $\{\mathbf{v}_1, \mathbf{v}_2\}$ векторов на плоскости Π, но теперь ограничимся целыми коэффициентами. Положим

$$\Gamma = \{n_1\mathbf{v}_1 + n_2\mathbf{v}_2 \mid n_1, n_2 \in \mathbb{Z}\}.$$

Множество переносов Π на векторы из **решётки** Γ образует бесконечную подгруппу в $\mathsf{T}(\mathbb{R}^2)$. Это пример дискретной группы. Вся она получается композициями переносов на векторы $\mathbf{v}_1$, $\mathbf{v}_2$ и им обратных.

Пример. Симметрии любой решётки Γ не исчерпываются переносами. Всегда осуществим поворот на половину оборота вокруг любой точки Γ, но могут быть допустимы и другие преобразования, оставляющие на месте одну точку или прямую линию; это зависит от формы базисного параллелограмма.

Симметрии квадратной и правильной гексагональной решёток вместе с переносами включают отражения и повороты, как в группах D_4 и D_6. Подгруппы групп симметрии этих решёток реализуют различные комбинации (всего возможно 17 типов) переносов с отражениями и поворотами. Они дают различные типы орнаментов, известных с древности и широко представленных в культуре разных народов мира.

Трёхмерные аналоги этих групп называют **кристаллографическими группами**. Подробнее с ними знакомятся геологи, а также химики и физики, работающие с твёрдыми телами.

Классические группы матриц

Лекция 12 (24.04.20)

Самые важные примеры непрерывных групп называют **классическими группами**. Это некоторые подгруппы относительно умножения в кольцах матриц. В кольце ещё есть операция сложения, но здесь о ней напрочь забывают, только перемножают и берут обратные.

условием	внутри	выделена	обозначаемая
$\det A \neq 0$	$\mathbf{M}_n(\mathbb{F})$	**полная линейная группа**	$\mathbf{GL}_n(\mathbb{F})$
$\det A = 1$	$\mathbf{GL}_n(\mathbb{F})$	**специальная линейная группа**	$\mathbf{SL}_n(\mathbb{F})$
$A^\top A = E$	$\mathbf{GL}_n(\mathbb{F})$	**ортогональная группа**	$\mathbf{O}_n(\mathbb{F})$
$\det A = 1$	$\mathbf{O}_n(\mathbb{F})$	**специальная ортогональная группа**	$\mathbf{SO}_n(\mathbb{F})$
$A^\dagger A = E$	$\mathbf{GL}_n(\mathbb{C})$	**унитарная группа**	$\mathbf{U}_n$
$\det A = 1$	$\mathbf{U}_n$	**специальная унитарная группа**	$\mathbf{SU}_n$

Обобщение ортогональных групп

Модифицируя определение, ортогональную группу $\mathbf{O}_n(\mathbb{F})$ можно рассматривать как группу таких матриц A, что $A^\top E A = E$. Вспоминая обсуждавшиеся в декабре билинейные формы, можно понимать это условие как сохранение линейным преобразованием A на пространстве $\mathbb{F}^n$ билинейной формы, в стандартном базисе данной выражением

$$d^2(\mathbf{x},\mathbf{y}) = \sum_{1\leqslant k\leqslant n} x_k y_k.$$

Такой подход открывает путь к обобщениям ортогональной группы.

Для произвольной билинейной формы, заданной в стандартном базисе матрицей F, все матрицы $A \in \mathbf{GL}_n(\mathbb{R})$ с условием

$$A^\top F A = F$$

образуют группу. Наиболее интересными оказываются различные случаи симметричных и кососимметричных форм; оба класса содержат важные для физики примеры.

Пример. Симметричные формы работают в различных обобщениях эвклидовой геометрии в роли функций, измеряющих расстояния. Например, поэтому **группа Лоренца** всех линейных преобразований пространства $\mathbb{R}^4$, сохраняющих билинейную форму Минковского

$$d^2(\mathbf{x},\mathbf{y}) = x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3,$$

полезна в электродинамике и в специальной теории относительности.

Пример. **Симплектическая группа** всех линейных преобразований пространства $\mathbb{R}^{2n}$, сохраняющих кососимметричную билинейную форму

$$\omega(\mathbf{x},\mathbf{y}) = \sum_{1\leqslant k\leqslant n} x_k y_{k+n} - x_{k+n} y_k,$$

используется в гамильтоновой механике (одна половина переменных есть обобщённые координаты, вторая — обобщённые импульсы).

11.2. Морфизмы, действия, представления

Морфизмы групп

Определение. Такое отображение групп $\varphi\colon \langle \mathcal{G}, \cdot\rangle \to \langle \mathcal{H}, \times\rangle$, что

$$\varphi(g_1 \cdot g_2) = \varphi(g_1) \times \varphi(g_2)$$

для всех $g_i \in \mathcal{G}$, называют **морфизмом** (а раньше называли **гомоморфизмом**) групп. Биективный морфизм групп называют **изоморфизмом**.

Определение. **Ядром** морфизма групп $\varphi\colon \mathcal{G} \to \mathcal{H}$ называют подгруппу

$$\operatorname{Ker}\varphi = \{g \in \mathcal{G} \mid \varphi(g) = e \in \mathcal{H}\} \leqslant \mathcal{G}.$$

Примеры. В таблице собрано несколько примеров морфизмов групп. Первые три отображают группу вещественных чисел по сложению в три разных группы по умножению; они являются морфизмами, потому что $\exp(x+y) = \exp(x)\exp(y)$. Только второе отображение биективно, и это единственный изоморфизм в таблице. Ядро третьего морфизма состоит из всех таких вещественных чисел x, что $\exp(2\pi i x) = 1$.

Отображение $D_2 \to \mathbb{S}_4$ получается взятием одной орбиты из четырёх точек за множество, на котором действуют перестановки. Образы только двух отражений s_1 и s_2 указаны в таблице, но этого достаточно; чтобы быть морфизмом, отображение обязано переводить нейтральный элемент группы $\mathcal{G}$ в нейтральный элемент группы $\mathcal{H}$, а произведение $s_1 s_2$ — в произведение указанных перестановок. Однако эти образы нельзя выбирать произвольно: любое соотношение между **порождающими** s_1 и s_2 должно выполняться и для их образов при морфизме. В данном случае все такие соотношения следуют из

$$s_1^2 = e, \quad s_2^2 = e, \quad s_1 s_2 = s_2 s_1,$$

и указанные в таблице перестановки-образы удовлетворяют таким же соотношениям. Это общие рассуждения, применимые ко всем морфизмам $\varphi\colon \mathcal{G} \to \mathcal{H}$; от случая к случаю меняется только конкретный вид соотношений, задающих начальную группу $\mathcal{G}$.

$\mathcal{G}$	$\mathcal{H}$	отображение	ядро	ин-?	сюр-?	действие
$\langle\mathbb{R},+\rangle$	$\mathbb{R}^\times$	$x \mapsto \exp(x)$	$\{0\}$	да	нет	—
$\langle\mathbb{R},+\rangle$	$\langle\mathbb{R}_{>0},\times\rangle$	$x \mapsto \exp(x)$	$\{0\}$	да	да	—
$\langle\mathbb{R},+\rangle$	$\mathbb{C}^\times$	$x \mapsto \exp(2\pi i x)$	$\langle\mathbb{Z},+\rangle$	нет	нет	повороты плоскости
D_2	$\mathbb{S}_4$	$s_1 \mapsto \left(\begin{smallmatrix}1&2&3&4\\4&3&2&1\end{smallmatrix}\right)$ $s_2 \mapsto \left(\begin{smallmatrix}1&2&3&4\\2&1&4&3\end{smallmatrix}\right)$	$\{e\}$	да	нет	перестановки орбиты
D_2	$\mathbf{GL}_2(\mathbb{R})$	$s_1 \mapsto \left[\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right]$ $s_2 \mapsto \left[\begin{smallmatrix}-1&0\\0&1\end{smallmatrix}\right]$	$\{e\}$	да	нет	отражения плоскости
$\mathbb{S}_3$	$\mathbf{GL}_3(\mathbb{R})$	$\sigma \mapsto \left[\begin{smallmatrix}0&1&0\\1&0&0\\0&0&1\end{smallmatrix}\right]$ $\tau \mapsto \left[\begin{smallmatrix}1&0&0\\0&0&1\\0&1&0\end{smallmatrix}\right]$	$\{e\}$	да	нет	перестановки координат
$\mathbf{GL}_n(\mathbb{R})$	$\mathbb{R}^\times$	$A \mapsto \det A$	$\mathbf{SL}_n(\mathbb{R})$	нет	да	искажение объёма

Аналогично два следующих отображения однозначно определены указанными парами матриц. В качестве упражнения вычислите образы перестановок $\sigma\tau$ и $\tau\sigma$. Не перепутайте их! Затем сравните с картинками перестановок из главы про определители. Возможно, вы найдёте простой способ построить (инъективный) морфизм $\mathbb{S}_n \to \mathbf{GL}_n(\mathbb{R})$ для любого $n \in \mathbb{Z}_{>0}$.

Действия групп

Абстрактное понятие группы как отдельного объекта появилось в математике много позже изучения различных совокупностей преобразований, ныне называемых действиями групп.

Определение. **Действием** группы $\mathcal{G}$ на множестве $\mathcal{X}$ называют отображение, сопоставляющее каждому элементу $g \in \mathcal{G}$ какую-то биекцию $\hat{g}\colon \mathcal{X} \to \mathcal{X}$ таким образом, что композиции этих биекций между собой соответствуют закону умножения в группе.

Чтобы записать аксиомы действия группы формально, требуется удобное обозначение для элемента, в который $x \in \mathcal{X}$ переводится «под действием» $g \in \mathcal{G}$. Среди множества вариантов, которые доводилось

видеть или выдумывать, наиболее привлекательным сейчас кажется обозначение $g \rightharpoonup x$. Тогда аксиомы действия примут вид:

- $e \rightharpoonup x = x$;
- $(gh) \rightharpoonup x = g \rightharpoonup (h \rightharpoonup x)$.

На самом деле первое требование следует (упражнение) из второго и биективности всех отображений $\hat{g}\colon \mathcal{X} \to \mathcal{X}$.

Более точно следует говорить, что приведённые аксиомы определяют **действие слева**, а зеркально симметричные — **действие справа**: $x \leftharpoonup (hg) = (x \leftharpoonup h) \leftharpoonup g$. Различие между действиями слева и справа бывает существенно для некоммутативных групп, где возможно $gh \neq hg$. Выбранные здесь значки для действия позволяют отличать и левое от правого, и элементы группы от элементов $\mathcal{X}$.

Среди данных выше примеров групп многие на самом деле описывают не сами группы, а их конкретные действия. Это относится к группам перестановок, группам, порождённым отражениями, группе поворотов, группе переносов, кристаллографическим группам. Однако одна и та же группа может действовать на различных множествах, а также по-разному действовать на одном и том же множестве.

Пример. Аддитивная (значит, по сложению) группа $\langle \mathbb{R}, + \rangle$ может многими способами действовать на комплексной плоскости, например:

(1) действие переносами вдоль вещественной оси, $x \rightharpoonup z = z + x$;
(2) действие растяжениями, $x \rightharpoonup z = e^x z$;
(3) действие поворотами, $x \rightharpoonup z = e^{2\pi i x} z$.

Пример. Группа $\mathbf{GL}_n(\mathbb{R})$ невырожденных вещественных $n \times n$ матриц действует на пространстве $\mathbb{R}^n$ столбцов по правилу $A \rightharpoonup X = AX$, а на пространстве $\mathbb{R}^n$ строк — по правилу $X \leftharpoonup A = XA$.

Орбиты и стабилизаторы

Определение. **Орбитой** и **стабилизатором** элемента $x \in \mathcal{X}$ под действием группы $\mathcal{G}$ называют соответственно множества

$$\mathcal{G} \rightharpoonup x = \{g \rightharpoonup x \in \mathcal{X} \mid g \in \mathcal{G}\} \quad \text{и} \quad \mathcal{G}_x = \{g \in \mathcal{G} \mid g \rightharpoonup x = x\}.$$

В предпоследнем примере орбитами являются:

(1) прямые, параллельные вещественной оси;
(2) незамкнутые лучи, выходящие из начала координат, а также само начало как отдельная орбита;
(3) окружности с центром в начале координат, а также само начало как отдельная орбита.

Стабилизаторы тривиальны, то есть равны $\{0\}$, у всех точек $z \in \mathbb{C}$ при переносах, и у всех, кроме $z = 0$, при растяжениях. Стабилизаторы начала координат при растяжениях и поворотах равны всей действующей группе $\mathbb{R}$. Самые интересные стабилизаторы в этом примере имеют точки $z \neq 0$ при поворотах: это подгруппа целых чисел $\langle \mathbb{Z}, + \rangle$.

Лемма. *Стабилизатор каждой точки при любом действии $\mathcal{G}$ является подгруппой в $\mathcal{G}$.* □

Пример. Группа, порождённая отражениями куба относительно плоскостей его симметрии, действует на множествах его вершин, его рёбер, его граней, всех его точек. Это всё разные действия. Подсчитаем число элементов в нескольких орбитах и стабилизаторах:

x	$\mathcal{G} \rightharpoonup x$	$\mathcal{G}_x$
вершина	8	6
ребро	12	4
грань	6	8

Постаравшись, здесь можно увидеть и структуру стабилизаторов. Они изоморфны группам диэдра D_3, D_2 и D_4 соответственно.

Теорема. *Если конечная группа $\mathcal{G}$ действует на конечном множестве $\mathcal{X}$, то для каждого $x \in \mathcal{X}$ порядки орбиты, стабилизатора и всей группы связаны соотношением*

$$^{\#}(\mathcal{G} \rightharpoonup x) \cdot {}^{\#}(\mathcal{G}_x) = {}^{\#}(\mathcal{G}).$$

Пример. Группа $\mathbf{SO}_3(\mathbb{R})$ действует поворотами на пространстве $\mathbb{R}^3$. Орбита начала координат состоит из одной точки. Орбита любой другой точки есть сфера с центром в начале координат.

Можно изучать действие $\mathbf{SO}_3$ на одной сфере. Стабилизатор каждой её точки есть подгруппа поворотов, ось которых проходит через эту точку, поэтому он изоморфен группе $\mathbf{SO}_2$ поворотов плоскости.

Действия группы на себе

Определение действия группы $\mathcal{G}$ на множестве $\mathcal{X}$ можно формально переписать как отображение

$$\mathcal{G} \times \mathcal{X} \to \mathcal{X}, \quad (g, x) \mapsto g \rightharpoonup x.$$

А что в случае $\mathcal{X} = \mathcal{G}$? Оказывается, каждая группа действует на себе, да не одним способом. Саму групповую операцию, умножение

$$\mathcal{G} \times \mathcal{G} \to \mathcal{G}, \quad (g, h) \mapsto gh,$$

можно двояко интерпретировать как действие **сдвигами**:

$$h \mapsto gh = (g \rightharpoonup h) \text{ есть сдвиг всей группы } \mathcal{G} \text{ слева на } g;$$

$$g \mapsto gh = (g \leftharpoondown h) \text{ есть сдвиг всей группы } \mathcal{G} \text{ справа на } h.$$

Аксиома действия в этом случае есть ассоциативность умножения в $\mathcal{G}$.

Другой важный пример (левого) действия даёт отображение

$$\mathcal{G} \times \mathcal{G} \to \mathcal{G}, \quad (g, h) \mapsto ghg^{-1} = g \rightharpoonup h;$$

оно называется действием **сопряжениями**. Сопряжения исключительно важны при изучении групп. Мы пользовались подобием матриц в теории операторов и конгруэнтностью матриц в теории билинейных форм. В случае невырожденных матриц $\mathcal{G} = \mathcal{X} = \mathbf{GL}_n(\mathbb{F})$ и подобие есть в точности сопряжённость. Конгруэнтность связана с действием $\mathbf{O}_n$ на $M_n(\mathbb{F})$.

Матричные представления группы

Если множество $\mathcal{X}$ наделено какой-то структурой, то среди различных возможных действий произвольной группы $\mathcal{G}$ на нём наиболее интересны те, которые уважают эту структуру (действия посредством **автоморфизмов**). Важнейший случай — действие на линейном пространстве $\mathcal{V}$, когда каждый элемент группы $\mathcal{G}$ представляется линейным оператором на $\mathcal{V}$. Часто при этом базис пространства заранее выбран и операторы представляются матрицами; поэтому задание такого действия становится эквивалентно заданию морфизма групп $\mathcal{G} \to \mathbf{GL}_n(\mathbb{F})$, где $\mathbb{F}$ и n есть основное поле и размерность пространства $\mathcal{V}$ соответственно.

Определение. Произвольный морфизм групп $\mathcal{G} \to \mathbf{GL}_n(\mathbb{F})$ называют n-мерным **представлением** группы $\mathcal{G}$ над полем $\mathbb{F}$.

Пример. Среди примеров морфизмов групп в таблице почти все являются линейными представлениями. В двух случаях в таблице явно написано, что $\mathcal{H}$ есть полная линейная группа. Кроме того, первые три примера и последний тоже попадают в этот класс: это одномерные представления, ибо $\mathbb{R}^\times = \mathbf{GL}_1(\mathbb{R})$ и $\mathbb{C}^\times = \mathbf{GL}_1(\mathbb{C})$.

Группы кос?

11.3. Алгебра кватернионов

Лекция 13 (08.05.20)

Сфера в четырёх измерениях

Начнём с описания общего вида элементов группы $\mathbf{SU}_2$. Из условия унитарности $A^\dagger A = E$ и равенства $\det A = 1$ выводим (упражнение)

$$\mathbf{SU}_2 = \left\{ \begin{bmatrix} \bar u & \bar v \\ -v & u \end{bmatrix} \in \mathbf{GL}_2(\mathbb{C}) \,\middle|\, \bar u u + \bar v v = 1 \right\}.$$

Записав комплексные числа u и v в алгебраической форме, получим

$$\mathbf{SU}_2 = \left\{ \begin{bmatrix} x_0 - ix_1 & x_2 - ix_3 \\ -x_2 - ix_3 & x_0 + ix_1 \end{bmatrix} \,\middle|\, x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1,\ x_s \in \mathbb{R} \right\}.$$

Поэтому можно рассматривать элементы группы $\mathbf{SU}_2$ как точки на трёхмерной сфере радиуса 1 в четырёхмерном пространстве.

Правила Гамильтона

Отбрасывая теперь условие с радиусом, рассмотрим само четырёхмерное вещественное пространство матриц:

$$(\diamondsuit) \qquad \mathbb{H} = \left\{ \begin{bmatrix} x_0 - ix_1 & x_2 - ix_3 \\ -x_2 - ix_3 & x_0 + ix_1 \end{bmatrix} \,\middle|\, x_s \in \mathbb{R} \right\}.$$

Все такие матрицы имеют вид $x_0 E + x_1 I + x_2 J + x_3 K$ для $x_s \in \mathbb{R}$ и

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}.$$

При знакомстве с комплексными числами осенью мы отметили их представление вещественными 2×2 матрицами специального вида, что можно считать одним из подходов к определению поля комплексных чисел. Аналогично проверим аксиомы поля для множества $\mathbb{H}$. Выполненными оказываются все аксиомы, кроме коммутативности умножения. Впрочем, матричное представление их все сразу обеспечивает, кроме возможности деления, то есть наличия обратного по умножению для каждого ненулевого элемента.

Правила умножения в $\mathbb{H}$ обычно записывают для базисных элементов E, I, J и K:

$$(\heartsuit) \qquad I^2 = J^2 = K^2 = IJK = -E.$$

Отсюда, в частности, следует, что

$$IJ = K = -JI, \qquad JK = I = -KJ, \qquad KI = J = -IK.$$

Определение (Hamilton, 1843). **Алгеброй кватернионов** $\mathbb{H}$ называют четырёхмерное вещественное пространство с базисом $\{E, I, J, K\}$ и заданной законами (♡) структурой алгебры над $\mathbb{R}$ с единицей E.

Сопряжение и норма

Определение. Для кватернионов имеются понятия, аналогичные комплексному сопряжению (**сопряжение**) и модулю (**норма**):

$$q = x_0E + x_1I + x_2J + x_3K \implies \begin{cases} \bar{q} = x_0E - x_1I - x_2J - x_3K \in \mathbb{H}, \\ |q|^2 = x_0^2 + x_1^2 + x_2^2 + x_3^2 \in \mathbb{R}_{\geqslant 0}. \end{cases}$$

Квадрат нормы равен определителю представляющей 2×2 матрицы в (♢), поэтому $|q_1q_2| = |q_1| \cdot |q_2|$ для любых кватернионов q_1 и q_2.

Вычисление обратных элементов по умножению как обратных матриц даёт формулу, идентичную своему комплексному аналогу:

$$q^{-1} = \bar{q}/|q|^2.$$

В частности, все ненулевые элементы $\mathbb{H}$ обратимы.

Группа $\mathbf{SU}_2$ есть мультипликативная подгруппа в $\mathbb{H}$, состоящая из всех кватернионов нормы 1, называемых **унитарными**. По аналогии с экспоненциальной записью комплексных чисел каждый ненулевой кватернион представим в виде $q = \rho u$, где положительное вещественное число ρ есть норма $|q|$, а унитарный кватернион $u = q/|q|$ есть «направление» кватерниона q.

Чисто мнимые кватернионы

Обозначим через $\mathbb{H}_0 = \langle I, J, K \rangle$ трёхмерное вещественное пространство **чисто мнимых** кватернионов; у них $x_0 = 0$. В матричном представлении (♢) им соответствуют матрицы с нулевым следом, поскольку $\operatorname{tr} I = \operatorname{tr} J = \operatorname{tr} K = 0$, но $\operatorname{tr} E = 2$. Иногда удобно записывать произвольный кватернион как сумму его вещественной и мнимой частей, $q = x_0 + p$, опуская для краткости единичный элемент E при x_0. Кроме того, легко заметить равенства

$$E^\dagger = E, \quad I^\dagger = -I, \quad J^\dagger = -J, \quad K^\dagger = -K.$$

Сопряжение кватерниона соответствует эрмитову сопряжению представляющей его комплексной матрицы, а разложение $q = x_0 + p$ совпадает с разложением матрицы в сумму эрмитовой и косоэрмитовой.

Стоит отметить связь умножения чисто мнимых кватернионов со скалярным и векторным произведениями в $\mathbb{H}_0 \simeq \mathbb{R}^3$:

$$p\,p' = p \times p' - p \cdot p'.$$

Выходит, что привычные скалярное и векторное произведения есть вещественная и мнимая части произведения кватернионов. Больше того! Скалярное и векторное произведения именно так и появились в науке. Сперва в физике: Максвелл написал свои уравнения электродинамики в кватернионах, а затем их принялись упрощать.

Трёхмерное представление

Теперь проиллюстрируем полезность алгебры кватернионов на примере удобного описания вращений трёхмерного пространства.

Лемма. *Отображение*

$$\varphi\colon \mathbf{SU}_2 \times \mathbb{H} \to \mathbb{H}, \quad (u, q) \mapsto uqu^{-1} = u \rightharpoonup q$$

есть левое действие группы $\mathbf{SU}_2$ *линейными операторами на пространстве* $\mathbb{H}$*, при котором инвариантно подпространство* $\mathbb{H}_0 \simeq \mathbb{R}^3$*.*

Доказательство. Выполнение аксиомы левого действия

$$(u_2u_1) \rightharpoonup q = u_2 \rightharpoonup (u_1 \rightharpoonup q)$$

обеспечивается ассоциативностью алгебры $\mathbb{H}$. Оно линейно, поскольку

$$u \rightharpoonup (\alpha_1 q_1 + \alpha_2 q_2) = \alpha_1 (u \rightharpoonup q_1) + \alpha_2 (u \rightharpoonup q_2)$$

для всех $\alpha_i \in \mathbb{R}$ и $q_i \in \mathbb{H}$. Вместо $u \in \mathbf{SU}_2$ здесь годится любой ненулевой кватернион ρu, но получающийся линейный оператор на $\mathbb{H}$ зависит только от направления u, поэтому можно ограничиться действием унитарных кватернионов.

Инвариантность $\mathbb{H}_0$ означает, что наше отображение сохраняет чистую мнимость. Проверить это можно как минимум двумя путями, причём оба просты. Косоэрмитовость матрицы $P = x_1 I + x_2 J + x_3 K$ наследуется матрицей $UPU^\dagger$:

$$(UPU^\dagger)^\dagger = UP^\dagger U^\dagger = -UPU^\dagger.$$

Ну а след обладает свойством $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ для любых квадратных матриц A и B одинакового размера. Поэтому $\operatorname{tr}(upu^{-1}) = \operatorname{tr} p$. □

Итак, имеется трёхмерное линейное представление

$$\mathbf{SU}_2 \to \mathbf{GL}_3(\mathbb{R}).$$

Вращения трёхмерного пространства

Чтобы вычислить матрицу, представляющую действие унитарного кватерниона $u = x_0E + x_1I + x_2J + x_3K$ в базисе $\{I, J, K\}$, разобьём отображение $q \mapsto uqu^{-1} = uq\bar{u}$ на два шага, вводя операторы

$$\mathsf{L}_u \colon q \mapsto uq, \quad \mathsf{R}_u \colon q \mapsto q\bar{u}$$

на всём пространстве $\mathbb{H}$ и используя (♡) для вычисления их матриц

$$L_u = \begin{bmatrix} x_0 & -x_1 & -x_2 & -x_3 \\ x_1 & x_0 & -x_3 & x_2 \\ x_2 & x_3 & x_0 & -x_1 \\ x_3 & -x_2 & x_1 & x_0 \end{bmatrix}, \qquad R_u = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ -x_1 & x_0 & -x_3 & x_2 \\ -x_2 & x_3 & x_0 & -x_1 \\ -x_3 & -x_2 & x_1 & x_0 \end{bmatrix}.$$

Эти замечательные матрицы ортогональны и коммутируют:

$$L_u^{-1} = L_u^\top, \quad R_u^{-1} = R_u^\top, \quad L_uR_u = R_uL_u.$$

Кроме того, $\det L_u = \det R_u = |u|^4 = 1$. Поэтому L_u и R_u принадлежат группе $\mathbf{SO}_4(\mathbb{R})$. Однако в итоге нас интересует произведение L_uR_u. Перемножая, находим блочно диагональную матрицу с блоками [1] и

$$A_u = \begin{bmatrix} x_0^2 + x_1^2 - x_2^2 - x_3^2 & 2x_1x_2 - 2x_0x_3 & 2x_1x_3 + 2x_0x_2 \\ 2x_1x_2 + 2x_0x_3 & x_0^2 - x_1^2 + x_2^2 - x_3^2 & 2x_2x_3 - 2x_0x_1 \\ 2x_1x_3 - 2x_0x_2 & 2x_2x_3 + 2x_0x_1 & x_0^2 - x_1^2 - x_2^2 + x_3^2 \end{bmatrix}.$$

Блок A_u есть искомая матрица, представляющая действие $p \mapsto up\bar{u}$ в базисе $\{I, J, K\}$. То, что она ортогональна и имеет определитель 1, не видно сразу по матрице, но прямо следует из способа её получения. Ортогональность также следует из того, что оператор A_u сохраняет длины: $|up\bar{u}| = |u| \cdot |p| \cdot |\bar{u}| = |p|$. Итак, имеется морфизм групп

$$\pi \colon \mathbf{SU}_2 \to \mathbf{SO}_3, \quad u \mapsto A_u.$$

Упражнение. *Проверьте, что действие $p \mapsto -upu$ на чисто мнимых кватернионах, где кватернион $u = x_1I + x_2J + x_3K$ одновременно унитарный и чисто мнимый, является отражением относительно плоскости, и найдите вектор нормали к ней.*

Через кватернионы легко явно найти поворот, равный композиции двух данных: нужно просто перемножить соответствующие унитарные кватернионы. Без них решение будет громоздким. Аналогично решается задача поиска поворота из одного заданного состояния в другое: ответ находится делением кватернионов. Представление поворотов кватернионами экономичнее и эффективнее, чем ортогональными матрицами, а также свободно от недостатков представления углами

Эйлера, поэтому оно активно используется в аэродинамике и космонавтике, компьютерной графике и робототехнике.

Спинорное накрытие

Посмотрим, как легко найти по кватерниону ось и угол доставляемого им поворота. Представление ортогональными матрицами такой простоты не даёт, ведь там требуется искать собственный вектор.

Теорема. *Для каждого унитарного кватерниона $u = x_0 + p$ с ненулевой мнимой частью оператор A_u является поворотом пространства $\mathbb{H}_0 \simeq \mathbb{R}^3$ на угол $\alpha = 2\arccos x_0$ вокруг оси, проходящей через p.*

Доказательство. Все нетривиальные элементы группы $\mathbf{SO}_3$ представляют в стандартном ОНБ повороты. След матрицы $A \in \mathbf{SO}_3$, задающей поворот на угол α, равен $1 + 2\cos\alpha$, а след матрицы A_u есть

$$\operatorname{tr} A_u = 3x_0^2 - x_1^2 - x_2^2 - x_3^2 = 4x_0^2 - 1;$$

отсюда следует заявленная связь угла с x_0. Вычисление

$$\begin{aligned}(x_0 + p) \rightharpoonup p &= (x_0 + p)\, p\, \overline{(x_0 + p)} \\ &= p\, (x_0 + p)\, \overline{(x_0 + p)} \\ &= p|x_0 + p|^2\end{aligned}$$

показывает, что p есть собственный вектор — ось поворота A_{x_0+p}. □

Следствие. *Образом морфизма π является вся группа $\mathbf{SO}_3$.* □

Так как $A_u = A_{-u}$ для каждого $u \in \mathbf{SU}_2$, наличие двух направлений по каждой оси и противоположность знаков соответствующих обоим выборам углов поворота аккуратно учтены морфизмом π.

Следствие. *Ядром морфизма π является подгруппа $\{\pm E\}$.*

Доказательство. Существует лишь два унитарных кватерниона с нулевой мнимой частью. □

Тополог назовёт полученное отображение $\pi\colon \mathbf{SU}_2 \to \mathbf{SO}_3$ «двулистным накрытием», ибо классические группы есть гладкие многообразия, а (сюръективное) отображение π непрерывно. Физический смысл действия $u \rightharpoonup p = uru^{-1}$, по сути, совпадающего с законом изменения матрицы оператора при смене базиса, вскрывается с введением двумерного комплексного «пространства спиноров». Тогда $\mathbb{H}_0 \simeq \mathbb{R}^3$ становится пространством косоэрмитовых операторов на спинорах.

Трюк с бокалом или ремнями, показанный на лекции, обычно связывают с Дираком, который его популяризовал, рассказывая о спинорах. Объяснение трюка чисто топологическое. Положение повернувшейся кисти руки представляют точкой в группе поворотов $\mathbf{SO}_3$, а всё её движение из естественного положения в вывернутое — петлёй в этой группе, которую невозможно стянуть в точку. Концы петли, совпадающие в $\mathbf{SO}_3$, оказываются разными в $\mathbf{SU}_2$. Второй оборот кисти замыкает петлю уже в $\mathbf{SU}_2$, а там всякая петля оказывается стягиваемой, и кисть приходит в естественное положение.

11.4. Примеры алгебр математической физики

Поля и кольца

Лекция 14 (15.05.20)

Числовые системы, лежащие в основах линейной алгебры, являются примерами колец и полей, и эти слова проникали в наш курс, особенно последнее. Такие системы включают две операции: сложение и умножение. Как уже говорилось в примерах групп, оставляя в числовых системах лишь одну операцию, мы получаем группы по сложению и по умножению. Аксиомы поля и аксиомы кольца налагают дополнительные требования. Дистрибутивность требуется всегда. Помимо неё, от умножения в кольце не требуется ничего, а потому кольца бывают очень разные по своей алгебраической структуре.

Целые числа образуют только кольцо, ибо нет обратной операции к делению. Рациональные, вещественные и комплексные числа образуют поля: после удаления нуля они остаются коммутативной группой по умножению. Существуют и другие поля, но им не находится места в нашем курсе, как, видимо, и вообще на физическом факультете.

При этом другие кольца появляются повсеместно, чаще всего — кольцо многочленов и кольцо матриц. Однако и то, и другое множество уже упоминались как примеры линейных пространств. Выходит, в них есть не две, а три операции: сложение элементов; перемножение элементов; умножение элементов на скаляры (живущие в поле, а может быть даже кольце). Алгебра — структура с дополнительным требованием к согласованию перемножения и умножения на скаляры. Алгебры оказываются особенно полезны, и не только алгебры многочленов и матриц. В этом разделе кратко представлены несколько более сложных классов алгебр, применяемых в физике; где-то это происходит явно, где-то абстракции остаются совершенно закулисными.

Алгебры и подалгебры

Определение. **Алгеброй** (над полем $\mathbb{F}$) называют множество, являющееся одновременно кольцом и линейным пространством над $\mathbb{F}$ так, что умножение на скаляры из $\mathbb{F}$ и умножение в кольце согласованы:

- $\lambda(\mathsf{AB}) = (\lambda\mathsf{A})\mathsf{B} = \mathsf{A}(\lambda\mathsf{B})$.

Размерностью алгебры над $\mathbb{F}$ называют её размерность как линейного пространства над $\mathbb{F}$.

Пример. Алгебра $M_n(\mathbb{R})$ матриц: размерности n^2, ассоциативная, некоммутативная, с единицей E.

Пример. Алгебра $\mathbb{R}[x]$ многочленов: размерности ∞, ассоциативная, коммутативная, с единицей 1.

Пример. Алгебра $\mathcal{L}(\mathcal{V})$, или $\operatorname{End}\mathcal{V}$, линейных операторов на линейном пространстве $\mathcal{V}$: размерности $(\dim\mathcal{V})^2$, ассоциативная, некоммутативная, с единицей E.

Когда в $\mathcal{V}$ выбран базис, все линейные операторы на $\mathcal{V}$ записываются матрицами. Линейной комбинации операторов соответствует та же линейная комбинация матриц, а композиции — произведение:

$$\left.\begin{matrix}\mathsf{A}\leftrightarrow A,\\ \mathsf{B}\leftrightarrow B\end{matrix}\right\} \Longrightarrow \begin{cases}\alpha\mathsf{A}+\beta\mathsf{B}\leftrightarrow \alpha A+\beta B,\\ \mathsf{AB}\leftrightarrow AB.\end{cases}$$

Это выражают фразой: выбор базиса в $\mathcal{V}$ устанавливает *изоморфизм алгебр* между $\mathcal{L}(\mathcal{V})$ и алгеброй матриц.

Определение. **Подалгеброй** алгебры $\mathcal{A}$ называют подпространство $\mathcal{A}'$ в $\mathcal{A}$, одновременно являющееся подкольцом (то есть произведение элементов $\mathcal{A}'$ также лежит в $\mathcal{A}'$).

Примеры. Много полезных подалгебр включает алгебра матриц. Первые шесть подпространств в $M_n(\mathbb{R})$ в их списке в конце первого семестра есть подалгебры. Аналогичны им подалгебры блочных матриц. Скажем, блочные верхнетреугольные и блочные диагональные матрицы имеют вид

$$\left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13}\\ \hline 0 & A_{22} & A_{23}\\ \hline 0 & 0 & A_{33}\end{array}\right] \quad\text{и}\quad \left[\begin{array}{c|c|c} A_{11} & 0 & 0\\ \hline 0 & A_{22} & 0\\ \hline 0 & 0 & A_{33}\end{array}\right],$$

где A_{ij} — любые матрицы, но диагональные блоки A_{ii} квадратные. Фиксируя число блоков и размеры каждого, получаем подалгебру.

Классические матричные алгебры Ли

На алгебре матриц $M_n(\mathbb{F})$ рассмотрим операцию коммутирования

$$[A, B] = AB - BA.$$

Она задаёт на линейном пространстве $M_n(\mathbb{F})$ структуру *неассоциативной* алгебры $\mathfrak{gl}_n(\mathbb{F})$, обладая свойствами:

- $[A, B] + [B, A] = 0$ (**антикоммутативность**),
- $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$ (**тождество Якоби**),

что и определяет принадлежность $\mathfrak{gl}_n(\mathbb{F})$ к особому классу **алгебр Ли**. Один пример алгебры Ли мы видели давно: линейное пространство $\mathbb{R}^3$ с операцией векторного умножения. В этом семестре такая структура обнаружилась на косоэрмитовых операторах.

Матричной алгеброй Ли называют подалгебру в $\mathfrak{gl}_n(\mathbb{F})$, то есть, линейное пространство матриц, замкнутое относительно коммутирования. Существует глубокая связь между классическими матричными алгебрами Ли и классическими матричными группами. Её усвоение требует владения важными аналитическими понятиями *гладкого многообразия* и *касательного пространства* к нему и хотя бы поэтому выходит довольно далеко за пределы нашего курса. Здесь мы ограничимся списком, как и для групп, и некоторыми намёками.

условием	внутри	выделена алгебра	над полем
—	$M_n(\mathbb{F})$	$\mathfrak{gl}_n(\mathbb{F})$	$\mathbb{F}$
$\operatorname{tr} A = 0$	$\mathfrak{gl}_n(\mathbb{F})$	$\mathfrak{sl}_n(\mathbb{F})$	$\mathbb{F}$
$A^\top + A = 0$	$\mathfrak{gl}_n(\mathbb{F})$	$\mathfrak{o}_n(\mathbb{F})$	$\mathbb{F}$
$A^\dagger + A = 0$	$\mathfrak{gl}_n(\mathbb{C})$	$\mathfrak{u}_n$	$\mathbb{R}$
$\operatorname{tr} A = 0$	$\mathfrak{u}_n$	$\mathfrak{su}_n$	$\mathbb{R}$

Условие $A^\dagger + A = 0$ **косоэрмитовости** не выдерживает умножения A на i, поэтому $\mathfrak{u}_n$ является алгеброй не над $\mathbb{C}$, а только над $\mathbb{R}$.

Сравнивая эту таблицу с таблицей классических групп, замечаем отсутствие алгебры $\mathfrak{so}_n(\mathbb{F})$. Дело в том, что $A^\top + A = 0$ влечёт $2 \operatorname{tr} A = 0$, поэтому $\mathfrak{so}_n(\mathbb{F}) = \mathfrak{o}_n(\mathbb{F})$, если $2 \neq 0$ в поле $\mathbb{F}$. Ограничиваясь полями $\mathbb{R}$ и $\mathbb{C}$, мы и не заметим последнего условия, даже не подумав о том, что существуют поля, где оно нарушается. Такие поля чаще всего встречаются в дискретной математике; в качестве примера упомянем простейшее — состоящее из двух элементов 0 и 1, например, понимаемых как остатки от деления целых чисел на 2.

Простая трёхмерная алгебра Ли

Теорема. *Матричные алгебры Ли* $\mathfrak{su}_2$ *и* $\mathfrak{so}_3(\mathbb{R})$ *изоморфны.*

Доказательство. Обе алгебры трёхмерны над $\mathbb{R}$, а $\mathfrak{su}_2$ ещё и совпадает с пространством $\mathbb{H}_0 = \langle I, J, K\rangle$ чисто мнимых кватернионов. Определим линейное отображение $\psi\colon \mathfrak{su}_2 \to \mathfrak{so}_3$, задав его значения

$$(\spadesuit)\qquad \psi(I) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 2 & 0 \end{bmatrix}, \quad \psi(J) = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix}, \quad \psi(K) = \begin{bmatrix} 0 & -2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

на базисных элементах

$$I = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}.$$

Требуемое свойство

$$\psi([A, B]) = [\psi(A), \psi(B)]$$

морфизма алгебр Ли достаточно проверить непосредственным вычислением для $A, B \in \{I, J, K\}$. Тогда линейность ψ обеспечит выполнение этого равенства для всех $A, B \in \langle I, J, K\rangle = \mathfrak{su}_2$.

Поскольку линейная оболочка $\langle \psi(I), \psi(J), \psi(K)\rangle$ совпадает с $\mathfrak{so}_3$, отображение ψ биективно. Значит, это изоморфизм алгебр Ли. □

Связь групп Ли и алгебр Ли

Если A есть одна из 3×3 матриц в (♠), то $\exp(At) \in \mathbf{SO}_3$ для всех $t \in \mathbb{R}$. В задании 7 вы должны увидеть в большей общности, как экспонента превращает косоэрмитовы матрицы в унитарные. Для кососимметричных и ортогональных матриц ситуация совершенно аналогичная. Таким образом, имеются отображения

$$\exp\colon \mathfrak{su}_n \to \mathbf{SU}_n, \qquad \exp\colon \mathfrak{so}_n \to \mathbf{SO}_n.$$

В отличие от этих двух алгебр, эти две группы не изоморфны, но мы не будем в это вникать (спинорного накрытия тут не достаточно).

Вообще, можно проверить, что матричная экспонента

$$\exp\colon \mathfrak{g} \to \mathbf{G}$$

связывает *соответствующие* классические матричные алгебру Ли $\mathfrak{g}$ и группу (тоже Ли) $\mathbf{G}$, чем и объясняются схожие обозначения при различающихся определяющих условиях.

Обратный переход — от группы Ли к алгебре Ли — делается средствами математического анализа. Дело в том, что классические мат-

ричные группы, как и другие группы Ли, являются также гладкими многообразиями. Попробуем увидеть на двух примерах из таблицы, как из условия на принадлежность к группе получается условие на принадлежность к соответствующей алгебре. Для этого рассмотрим матрицы, мало отличающиеся от единичной в «направлении» A, поэтому ниже $\varepsilon \ll 1$, а лучше даже $\varepsilon \to 0$.

Пример. Раскрывая определитель, отделим слагаемые при нулевой, первой и высших степенях ε:

$$\det(E + \varepsilon A) = \det E + \varepsilon \operatorname{tr} A + O(\varepsilon^2).$$

Касательные направления к группе $\mathbf{SL}_n(\mathbb{F})$ даются матрицами с условием $\operatorname{tr} A = 0$, появившимся при первой степени ε. Это алгебра $\mathfrak{sl}_n(\mathbb{F})$.

Пример. Аналогично поступим со свойством ортогональности:

$$(E + \varepsilon A)^\top (E + \varepsilon A) = E + \varepsilon(A^\top + A) + O(\varepsilon^2).$$

Условие при первой степени ε получается $A^\top + A = 0$.

Для трёхмерного пространства такая связь ортогональности и кососимметричности уже встретилась нам при изучении движения сопровождающего трёхгранника линии. Конечно, тут не в линии дело: эта связь проявляется везде, где есть вращения.

В приложениях группы и алгебры возникают не сами по себе: обычно возникают их действия и особенно линейные представления. Матричная экспонента напрямую связывает представления группы Ли и её алгебры Ли, поэтому те и другие часто встречаются бок о бок, а иногда даже не так уж чётко различаются. Объектом более фундаментальным, но и более сложным в изучении тут являются группы.

Группа Гейзенберга

Соответствие между группами Ли и алгебрами Ли покрывает не только классические группы; есть ключевые для физики примеры и иных сортов. Предварительно рассмотрим скучные множества матриц

$$\mathfrak{g} = \left\{ \begin{bmatrix} 0 & t \\ 0 & 0 \end{bmatrix} \;\middle|\; t \in \mathbb{R} \right\}, \qquad \mathbf{G} = \left\{ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \;\middle|\; t \in \mathbb{R} \right\}.$$

Поскольку $A^2 = 0$ для каждой матрицы $A \in \mathfrak{g}$, матричная экспонента сразу вычисляется, $\exp(A) = E + A$, и оказывается лежащей в $\mathbf{G}$. Кроме того, $[A, B] = 0$ для всех $A, B \in \mathfrak{g}$, а группа $\mathbf{G}$ изоморфна группе вещественных чисел по сложению.

Гораздо интереснее аналоги в следующей размерности. Рассмотрим теперь вещественные матрицы видов

$$\begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}.$$

Все матрицы правого типа образуют группу $\mathbf{H}$ по умножению, а все матрицы левого типа образуют соответствующую матричную алгебру Ли $\mathfrak{h}$. Возьмём в $\mathfrak{h}$ естественный базис

$$X = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

и вычислим **коммутаторы**:

$$[X, Y] = Z, \quad [X, Z] = 0, \quad [Y, Z] = 0.$$

Элемент Z коммутирует со всеми элементами алгебры, и за это его называют **центральным**.

Эти соотношения примечательны своей ролью в квантовой механике. Вместо X и Y нужно взять оператор координаты $\hat{x}$ и оператор импульса $\hat{p}$ частицы в одномерной задаче; тогда $[\hat{x}, \hat{p}] = i\hbar$ согласно принципу неопределённости Гейзенберга, причём оператор $i\hbar$ централен, ибо пропорционален тождественному.

Упражнение. *Коммутативна ли группа Гейзенберга* $\mathbf{H}$*?*

Ещё на заре квантовой теории появились различные её интерпретации и возникла проблема установления связей между ними. Ответ оказался математическим и выражается теоремой об эквивалентности представлений группы Гейзенберга. Которые бесконечномерны...

Алгебры Вейля

Алгебры Вейля тоже являются частью математического аппарата, примыкающего к основаниям квантовой механики (принципу неопределённости), где их называют **алгебрами коммутаторов**. Однако сами эти алгебры ассоциативные, а коммутаторы строятся по тому же правилу $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$, что и в матричных алгебрах Ли. Из первой алгебры Вейля таким образом получается алгебра Гейзенберга $\mathfrak{h}$.

В основе построений этого раздела лежит алгебра многочленов $\mathbb{C}[x]$. Как и любая алгебра, она является линейным пространством (над $\mathbb{C}$); обозначим это пространство через $\mathcal{V}$. Начнём с двух простых линейных операторов на $\mathcal{V}$:

- $\mathsf{X}\colon f(x) \mapsto xf(x)$ — умножение многочлена на x;
- $\mathsf{D}\colon f(x) \mapsto \frac{d}{dx}f(x)$ — дифференцирование многочлена по x.

Построим из них более сложные линейные операторы на $\mathcal{V}$, используя только композицию и линейные комбинации операторов. Иными словами, рассмотрим алгебру $\mathcal{A}_1$ операторов на $\mathcal{V}$, порождённую операторами X и D. Она и называется **первой алгеброй Вейля**.

Примеры. Операторы $\mathsf{X}^3 + \mathsf{X}$, $\mathsf{D}^2 + 4\mathsf{E}$, $\mathsf{DX} - \mathsf{XD}$ принадлежат этой алгебре. Из них первый — оператор умножения многочлена на $x^3 + x$, второй действует по правилу

$$f(x) \mapsto (\mathsf{D}^2 + 4\mathsf{E})f(x) = f''(x) + 4f(x),$$

а третий (важное упражнение) равен тождественному оператору.

Каждому многочлену $a(x) = \sum a_k x^k$ соответствуют оператор умножения на него, действующий по правилу

$$a(\mathsf{X})\colon f(x) \mapsto a(x)f(x),$$

и **дифференциальный оператор** $a(\mathsf{D})$, действующий по правилу

$$a(\mathsf{D})\colon f(x) \mapsto \sum a_k f^{(k)}(x),$$

где $f^{(k)}(x) = \mathsf{D}^k f(x)$ обозначает k-ю производную по x.

Для любых многочленов $a(x)$, $b(x) \in \mathbb{C}[x]$ выполнены соотношения

$$[a(\mathsf{X}), b(\mathsf{X})] = \mathsf{O},$$
$$[a(\mathsf{D}), b(\mathsf{D})] = \mathsf{O};$$

в этой записи используется коммутатор операторов: $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$. Однако неперестановочность умножений и дифференцирований видна уже на примере $[\mathsf{D}, \mathsf{X}] = \mathsf{E}$.

Упражнение. *Докажите, что соотношение* $[\mathsf{A}, \mathsf{B}] = \mathsf{E}$ *невозможно реализовать в алгебре матриц. Иными словами, алгебры Вейля и Гейзенберга не допускают конечномерных линейных представлений.*

Теорема. *Всякий оператор в алгебре Вейля* $\mathcal{A}_1$ *представим в виде* $g(\mathsf{X}; \mathsf{D})$ *для какого-то многочлена* $g(x, y) \in \mathbb{C}[x, y]$ *от двух букв, причём в каждом мономе все вхождения* X *левее всех вхождений* D.

При этом многочлен, которым запишется композиция операторов, не равен произведению многочленов, соответствующих множителям, кроме тривиального случая композиции $a(\mathsf{X})b(\mathsf{D})$.

Доказательство. Достаточно переписать в таком виде оператор $\mathsf{D}^k\mathsf{X}^l$ для любых натуральных чисел k и l, используя правило дифференцирования произведения, чтобы перенести все умножения левее всех дифференцирований. □

Исходя из алгебры многочленов $\mathbb{C}[x_1, \ldots, x_n]$ от n переменных, аналогично строят n-ую алгебру Вейля, порождённую операторами умножения $\mathsf{X}_1, \ldots, \mathsf{X}_n$ и дифференцирования $\mathsf{D}_1, \ldots, \mathsf{D}_n$. При этом

$$[\mathsf{X}_i, \mathsf{X}_j] = \mathsf{O}, \quad [\mathsf{D}_i, \mathsf{D}_j] = \mathsf{O}, \quad [\mathsf{D}_i, \mathsf{X}_j] = \delta_{ij}\mathsf{E}.$$

11.5. Алгебры Грассмана

Внешние формы

Обозначим через $\mathcal{V}$ пространство столбцов $\mathbb{R}^n$. Соберём столбцы $X_1, \ldots, X_n$ в матрицу A. Как вы можете помнить из первого семестра, функция $A \mapsto \det A$ от n переменных $X_i \in \mathcal{V}$ линейна по каждому аргументу и меняет знак при перестановке любой их пары, то есть полилинейна и кососимметрична, причём все такие функции пропорциональны определителю. Но есть ли на $\mathcal{V}$ какие-нибудь функции $k \neq n$ аргументов, имеющие эти же свойства?

Определение. Полилинейная и кососимметричная функция k аргументов на пространстве $\mathcal{V}$ называется **внешней k-формой** на $\mathcal{V}$.

Пример. Для всякого $U = [u_1, \ldots, u_n]^\top \in \mathbb{R}^n$ функция

$$X = [x_1, \ldots, x_n]^\top \mapsto U^\top X = u_1x_1 + \ldots + u_nx_n$$

есть 1-форма на $\mathbb{R}^n$. «Внешность» 1-форм редко подчёркивается, ибо кососимметричность тут тривиальна: не хватает аргументов, которые хочется переставлять.

Пример. При $1 \leqslant k \leqslant n$ простейшие внешние k-формы на $\mathbb{R}^n$ — миноры порядка k матрицы $[X_1, \ldots, X_k]$, где $X_i \in \mathcal{V}$.

Поскольку полилинейность и кососимметричность наследуются линейными комбинациями, внешние k-формы на $\mathbb{R}^n$ образуют линейное пространство Λ^k_n. При $k > n$ пространство Λ^k_n нулевое: оно состоит из единственной функции, равной нулю тождественно.

Структура **алгебры Грассмана** задаётся на линейном пространстве

$$\Lambda_n = \bigoplus_{0 \leqslant k \leqslant n} \Lambda^k_n.$$

В курсе основ математического анализа рассказывают, как на основе алгебры Грассмана строится техника дифференциальных форм, повсеместно применяемая в электродинамике, гидродинамике и других дисциплинах. Фундаментальный алгебраический формализм в приложениях часто не пытаются вскрыть; на мой взгляд, это может быть непродуктивно. С другой стороны, бессмысленно перегружать формальностями младшие курсы.

Внешняя алгебра трёхмерного пространства

Пример. Алгебра Грассмана пространства $\mathbb{R}^3$ восьмимерна. Элементы её стандартного базиса перечислены в таблице.

степень	0	1	2	3
базисные формы	$\omega_\varnothing$	ω_1 ω_2 ω_3	ω_{12} ω_{13} ω_{23}	ω_{123}
размерность	1	3	3	1

Нуль-формами считаются константы (функции 0 аргументов), так что $\omega_\varnothing \equiv 1$. Базисные 1-формы заданы правилом

$$\omega_i([x_1, x_2, x_3]^\top) = x_i.$$

Все 3-формы пропорциональны определителю, так что пусть определитель и будет ω_{123}.

Базисные 2-формы ω_{12}, ω_{13} и ω_{23} заданы 2×2 минорами. На векторах $X = [x_1, x_2, x_3]^\top$ и $Y = [y_1, y_2, y_3]^\top$ они принимают значения

$$\omega_{12}(X,Y) = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}, \quad \omega_{13}(X,Y) = \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix}, \quad \omega_{23}(X,Y) = \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}$$

соответственно. Подобно определителю, эти 2-формы имеют простой геометрический смысл: значение $w_{12}(X,Y)$ равно площади ориентированного параллелограмма, полученного проекцией векторов X и Y на плоскость $\{x_3 = 0\}$. Для любых α_1, α_2, $\alpha_3 \in \mathbb{R}$ функция

$$\alpha_1\omega_{23} + \alpha_2\omega_{13} + \alpha_3\omega_{12}$$

кососимметрична и полилинейна. Эту 2-форму тоже можно представить ориентированной площадью. (В случае $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$ это площадь проекции параллелограмма на паре аргументов на плоскость, имеющую углы $\arccos \alpha_i$ с тремя координатными плоскостями.)

Таблицей умножения базисных элементов зададим на пространстве $\Lambda_3 = \Lambda_3^0 \oplus \Lambda_3^1 \oplus \Lambda_3^2 \oplus \Lambda_3^3$ структуру ассоциативной алгебры с единицей:

	1	ω_1	ω_2	ω_3	ω_{12}	ω_{13}	ω_{23}	ω_{123}
1	1	ω_1	ω_2	ω_3	ω_{12}	ω_{13}	ω_{23}	ω_{123}
ω_1	ω_1	0	ω_{12}	ω_{13}	0	0	ω_{123}	0
ω_2	ω_2	$-\omega_{12}$	0	ω_{23}	0	$-\omega_{123}$	0	0
ω_3	ω_3	$-\omega_{13}$	$-\omega_{23}$	0	ω_{123}	0	0	0
ω_{12}	ω_{12}	0	0	ω_{123}	0	0	0	0
ω_{13}	ω_{13}	0	$-\omega_{123}$	0	0	0	0	0
ω_{23}	ω_{23}	ω_{123}	0	0	0	0	0	0
ω_{123}	ω_{123}	0	0	0	0	0	0	0

Внешнее умножение

Определим теперь формально операцию **внешнего умножения** внешних форм, с помощью которой, в частности, вычисляется приведённая таблица умножения алгебры Λ_3. В простейшем случае по правилу

$$(\omega_1 \wedge \omega_2)(X_1, X_2) = \begin{vmatrix} \omega_1(X_1)\, \omega_1(X_2) \\ \omega_2(X_1)\, \omega_2(X_2) \end{vmatrix}$$

из двух 1-форм ω_1 и ω_2 делают одну 2-форму $\omega_1 \wedge \omega_2$. Аналогичным определителем вычисляются значения

$$(\omega_1 \wedge \ldots \wedge \omega_k)(X_1, \ldots, X_k) = \det[\omega_i(X_j)]$$

базисной k-формы $\omega_1 \wedge \ldots \wedge \omega_k$ на любом списке векторов $X_1, \ldots, X_k$, так что произведениями базисных 1-форм являются как раз миноры. Используя разложение произвольной формы в линейную комбинацию базисных, можно определить умножение произвольных форм

$$\Lambda_n^k \times \Lambda_n^l \to \Lambda_n^{k+l}.$$

Внешнее умножение ассоциативно. Коммутативности нет, но

$$\omega \in \Lambda_n^k \quad \text{и} \quad \omega' \in \Lambda_n^l \implies \omega \wedge \omega' = (-1)^{kl} \omega' \wedge \omega.$$

Перестановка вниз первых k строк матрицы, всего имеющей $k+l$ строк, требует kl обменов местами соседних строк; отсюда и возникает множитель $(-1)^{kl}$. Это свойство форм называют **суперкоммутативностью**.

При внешнем умножении складываются степени форм, поэтому оно не делает алгеброй отдельно взятое пространство Λ_n^k, но превращает их прямую сумму Λ_n в алгебру над $\mathbb{R}$ размерности 2^n. Это и есть алгебра Грассмана, названная первооткрывателем **внешней алгеброй**.

Больше миноров

В случае произвольной размерности удобно перечислить все k-элементные множества, индексирующие упомянутые миноры:

$$\mathcal{C}(n,k) = \bigl\{ I \subseteq \{1,\dots,n\} \mid {}^{\#}(I) = k \bigr\}.$$

Для каждого $I \in \mathcal{C}(n,k)$ обозначим через M_I минор, составленный из строк матрицы $[X_1,\dots,X_k]$, номера которых входят в I.

Теорема. *Множество $\{M_I \mid I \in \mathcal{C}(n,k)\}$ есть базис пространства Λ_n^k внешних форм степени k на n-мерном пространстве.*

Доказательство. Обозначим через $E^{(J)}$ список взятых для всех $j \in J$, в порядке возрастания индекса, векторов $E^{(j)}$ стандартного базиса $\mathbb{R}^n$. Тогда

$$M_I(E^{(J)}) = \begin{cases} 1 & \text{при } I = J, \\ 0 & \text{при } I \neq J, \end{cases}$$

и линейная независимость указанного множества миноров следует.

Значение произвольной полилинейной функции на наборе векторов из $\mathcal{V}$ определяется однозначно по её значениям на наборах базисных векторов. Поэтому из свойств определителей следует, что всякая форма $\omega \in \Lambda_n^k$ однозначно представима в виде

$$\omega = \sum_{I \in \mathcal{C}(n,k)} \omega(E^{(I)}) M_I. \qquad \square$$