ALGEBRA LINIOWA I GEOMETRIA

A.I.Kostrikin
J.I.Manin

ALGEBRA LINIOWA I GEOMETRIA

z rosyjskiego przełożył
Aleksander Strasburger

Warszawa 1993
Wydawnictwo Naukowe PWN

Dane oryginału

А. И. Кострикин, Ю. И. Манин
Линейная алгебра и геометрия
«Наука» Москва 1986

Okładkę i strony tytułowe projektował
Dariusz Litwiniec

Redaktor
Bożena Jagoszewska

Skład komputerowy wykonał w systemie TEX
Aleksander Strasburger

ISBN 83-01-10207-1

WYDAWNICTWO NAUKOWE PWN

Wydanie pierwsze
Arkuszy wydawniczych 22,5
Arkuszy drukarskich 21,75
Reprodukcja Dział Graficzny PWN
Druk ukończono w styczniu 1993 r.
Zam. 2488/92

WROCŁAWSKA DRUKARNIA NAUKOWA

Przedmowa

Każdą rzecz można oglądać z rozmaitych punktów widzenia, a temat naszej książki nie stanowi wyjątku od tej reguły. Dla studenta Wydziału Mechaniczno-Matematycznego Państwowego Uniwersytetu Moskiewskiego (MGU) algebra liniowa jest przedmiotem wykładanym na drugim semestrze pierwszego roku. Dla specjalisty — algebraika wykształconego na ideach Bourbakiego, algebra liniowa jest teorią dość szczególnych struktur algebraicznych, a mianowicie przestrzeni i odwzorowań liniowych lub, w nowomodnym języku, teorią kategorii liniowych.

Zadanie algebry liniowej można scharakteryzować w szerszej perspektywie jako opracowanie matematycznego języka, w którym dałaby się wyrazić jedna z najbardziej podstawowych idei nauk przyrodniczych — idea liniowości. Jej najważniejszym przypadkiem szczególnym jest, być może, zasada liniowości małych przyrostów — niemal każdy naturalny proces jest prawie zawsze lokalnie liniowy. Zasada ta legła u podstaw całej analizy matematycznej i jej zastosowań. Algebra wektorów w trójwymiarowej przestrzeni fizycznej, którą historia uczyniła kamieniem węgielnym w budowli algebry liniowej, wyrosła z tego samego źródła — dopiero od czasu Einsteina wiemy, że także przestrzeń fizyczna jest tylko w przybliżeniu liniowa w małym otoczeniu obserwatora. Na szczęście, to małe otoczenie jest dostatecznie duże.

Fizyka dwudziestego wieku radykalnie i dość nieoczekiwanie rozszerzyła zasięg zastosowania idei liniowości, dołączając do zasady liniowości małych przyrostów zasadę superpozycji wektorów stanu. Najogólniej mówiąc, przestrzeń stanów dowolnego układu kwantowo-mechanicznego jest przestrzenią liniową nad ciałem liczb zespolonych. W rezultacie prawie wszystkie konstrukcje zespolonej algebry liniowej stały się częścią aparatu matematycznego, stosowanego do formułowania fundamentalnych praw przyrody — od teorii dwoistości liniowej, wyjaśniającej kwantową zasadę komplementarności Bohra, do teorii reprezentacji grup, pozwalającej objaśnić tablicę Mendelejewa, „zoologię" cząstek elementarnych, a nawet samą strukturę czasoprzestrzeni.

Wybór materiału do tego wykładu jest określony poprzez dążenie autorów do przedstawienia nie tylko podstaw aparatu, który osiągnął prawie ukończoną postać już na początku tego wieku, ale także do wyrobienia poglądu na możliwości jego zastosowań, odnoszących się zazwyczaj do innych dyscyplin. Tradycja wykładania sprzyja rozczłonkowaniu żywego organizmu matematyki na izolowane organy, których możliwości życiowe muszą być sztucznie podtrzymywane. Odnosi się to szczególnie do tych „krytycznych okresów" w historii naszej nauki, które charakteryzują się zwróceniem szczególnej uwagi na logiczną zwartość i szczegółowe opracowanie podstaw teorii. Ostatnie pół wieku było okresem ukierunkowanej przez teorię mnogości przebudowy języka i podstawowych pojęć, a jedność matematyki upatrywano

przede wszystkim w jedności jej podstaw logicznych. Nie odcinając się od zauważalnych osiągnięć tego okresu, chcemy w tej książce wyrazić także rysujące się tendencje do syntezy matematyki jako środka poznania świata zewnętrznego. (Niestety, musieliśmy całkowicie pominąć teorię metod numerycznych w zagadnieniach algebry liniowej, która rozrosła się w samodzielną dyscyplinę.)

Z wymienionych względów do przedstawianej przez nas książki, podobnie jak i do *Wstępu do algebry* jednego z autorów, został włączony nie tylko materiał przeznaczony do wykładu, ale również przeznaczony do czytania w domu, a także dla zajęć seminaryjnych. Nie można tu narzucać ostrego podziału, ale w zgodzie ze standardowym programem wykładu (jeden semestr, lecz dwa wykłady tygodniowo), powinien się w nim znaleźć podstawowy materiał § § 1–9 części 1; § § 2–8 części 2; § § 1, 3, 5, 6 części 3 i § § 1, 3–6 części 4. Przy tym jako materiał podstawowy określamy tu nie tyle dowody trudnych twierdzeń, których nie ma za dużo w algebrze liniowej, co system pojęć, który należy sobie przyswoić. Dlatego wiele twierdzeń zawartych w tych rozdziałach można przedstawić w uproszczonych wariantach lub nawet całkowicie opuścić — z braku czasu taka praktyka jest w pełni dopuszczalna. Jak przy tym uniknąć zamienienia wykładu w nużące wyliczanie definicji stanowi dla wykładowcy poważny problem. Mamy nadzieję, że pozostałe fragmenty książki będą mu w tym pomocne.

Przy powtórnym wydaniu zostały poprawione błędy drukarskie i drobne stylistyczne usterki, na które zwrócili nam uwagę liczni czytelnicy. W dwóch miejscach byliśmy zmuszeni wnieść poprawki do dowodów. Wiele cennych spostrzeżeń przekazali nam współpracownicy Katedry algebry i teorii liczb Państwowego Uniwersytetu Leningradzkiego. Konstruktywna krytyka recenzentów, a także bardzo sumienna i wielce kompetentna praca redaktora wydania V. L. Popova w znacznym stopniu przyczyniły się do polepszenia jakości książki. Wszystkim tym osobom wyrażamy głęboką wdzięczność.

Odpowiedzialność za pozostałe w książce niedostatki autorzy przyjmują na swoje barki.

Spis literatury uzupełniającej

1. A. I. Kostrikin, *Wstęp do algebry*, PWN, Warszawa 1982.
2. S. Lang, *Algebra*, PWN, Warszawa 1984.
3. A. I. Malcev, *Osnovy linejnoj algebry*, Moskva, Nauka 1975.
4. I. M. Gelfand, *Wykłady z algebry liniowej*, PWN, Warszawa 1971.
5. P. Halmos, *Finite dimensional vector spaces*, Van Nostrand, Princeton 1958.
6. E. Artin, *Geometric algebra*, Interscience, New York 1957.
7. I. M. Glazman, J.I. Ljubič, *Konečnomernyj linejnyj analiz*, Moskva, Nauka 1969.
8. V. V. Voevodin, *Vyčislitelnye osnovy linejnoj algebry*, Moskva, Nauka 1977.

Część 1.

Przestrzenie liniowe i odwzorowania liniowe

§ 1. Przestrzenie liniowe

1. Wektory zaczepione w wybranym punkcie przestrzeni można mnożyć przez liczby i dodawać zgodnie z regułą równoległoboku. Stanowi to klasyczny model praw składania przemieszczeń, prędkości lub sił w mechanice. W określeniu ogólnej przestrzeni wektorowej (inaczej liniowej) liczby rzeczywiste zastępuje się elementami dowolnego ciała, a najprostsze własności dodawania i mnożenia wektorów wprowadza się jako aksjomaty. „Trójwymiarowość" przestrzeni fizycznej nie odciska żadnego śladu na definicji, gdyż pojęcie wymiaru wprowadza się i bada osobno.

. Wykłady poświęcone dwu- i trójwymiarowej geometrii analitycznej dostarczają wielu przykładów geometrycznej interpretacji algebraicznych związków między dwiema lub trzema zmiennymi. Jednakże w opinii N. Bourbakiego „ograniczenie się tylko do języka geometrycznego odpowiadającego przestrzeni trójwymiarowej byłoby dla współczesnego matematyka tak krępującym jarzmem, jak to, które przeszkadzało Grekom rozszerzyć pojęcie liczby na stosunki wielkości niewspółmiernych".

2. Definicja. *Przestrzenią liniową (lub wektorową) L nad ciałem $\boldsymbol{K}$* nazywamy zbiór L z dwoma wyróżnionymi działaniami: wewnętrznym $L \times L \to L$, zapisywanym zazwyczaj jako dodawanie: $(l_1, l_2) \mapsto l_1 + l_2$, i zewnętrznym $\boldsymbol{K} \times L \to L$, zapisywanym zazwyczaj jako mnożenie: $(a, l) \mapsto al$.

Zakładamy przy tym, że spełnione są następujące aksjomaty:

a) *Dodawanie elementów przestrzeni L, zwanych wektorami, wprowadza w L strukturę grupy abelowej (przemiennej). Element zerowy tej grupy jest zwykle oznaczany przez 0, a element odwrotny do l oznaczany jest przez $-l$.*

b) *Mnożenie wektorów przez elementy ciała $\boldsymbol{K}$, zwane skalarami, jest unitarne, tj. $1l = l$ dla każdego l, i łączne, tj. $a(bl) = (ab)l$ dla każdych $a, b \in \boldsymbol{K}, l \in L$.*

c) *Dodawanie i mnożenie są związane prawami rozdzielności, tj.*

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

dla każdych $a, a_1, a_2 \in \boldsymbol{K}$, $l, l_1, l_2 \in L$.

3. A oto niektóre najprostsze wnioski z definicji.

a) $0l = a0 = 0$ dla wszystkich $a \in \boldsymbol{K}, l \in L$. W istocie, $0l + 0l = (0+0)l = 0l$, skąd wynika $0l = 0$ na mocy prawa skracania w grupie abelowej. Analogicznie, $a0 + a0 = a(0+0) = a0 = 0$, tj. $a0 = 0$.

b) $(-1)l = -l$. W istocie, $l + (-1)l = 1l + (-1)l = (1 + (-1))l = 0l = 0$, a więc wektor $(-1)l$ jest przeciwny do wektora l.

c) Jeśli $al = 0$, to mamy $a = 0$ lub $l = 0$. W istocie, jeśli $a \neq 0$, to $0 = a^{-1}(al) = (a^{-1}a)l = 1l = l$.

d) Dla dowolnych $a_1, \ldots, a_n \in \boldsymbol{K}$; $l_1, \ldots, l_n \in L$ wyrażenie $a_1 l_1 + \cdots + a_n l_n = \sum_{i=1}^{n} a_i l_i$ jest jednoznacznie określone, gdyż dzięki łączności dodawania w grupie abelowej można nie wprowadzać nawiasów ustalających kolejność dodawania parami składników sumy. Podobnie jednoznacznie określony sens ma wyrażenie $a_1 a_2 \cdots a_n l$. Wyrażenie postaci $\sum_{i=1}^{n} a_i l_i$ nazywa się *kombinacją liniową* wektorów $l_1, \ldots, l_n$, a skalary $a_1, \ldots, a_n$ — *współczynnikami* tej kombinacji liniowej.

W dalszym ciągu często będą pojawiać się następujące przykłady przestrzeni liniowych.

4. Przestrzeń zerowymiarowa. Jest to grupa abelowa $L = \{0\}$ zawierająca jedynie element zerowy. Jedyną możliwą postacią mnożenia przez skalary jest $a0 = 0$ dla wszystkich $a \in \boldsymbol{K}$ (sprawdźcie spełnienie aksjomatów !).

Uwaga! Zerowymiarowe przestrzenie wektorowe określone nad *różnymi ciałami są różnymi przestrzeniami*, bo definicja przestrzeni liniowej obejmuje także zadanie ciała skalarów $\boldsymbol{K}$.

5. Ciało skalarów jako jednowymiarowa przestrzeń kartezjańska. Tu $L = \boldsymbol{K}$, dodawanie jest dodawaniem w ciele $\boldsymbol{K}$, a mnożenie przez skalary jest mnożeniem w $\boldsymbol{K}$. Spełnienie aksjomatów przestrzeni liniowej wynika z aksjomatów ciała dla $\boldsymbol{K}$.

Nieco ogólniej, jeśli dane jest ciało $\mathcal{K}$ i jego podciało $\boldsymbol{K}$, to $\mathcal{K}$ można traktować jako przestrzeń wektorową nad $\boldsymbol{K}$. Na przykład, ciało $\boldsymbol{C}$ liczb zespolonych jest przestrzenią wektorową nad ciałem liczb rzeczywistych $\boldsymbol{R}$, które z kolei jest przestrzenią wektorową nad ciałem liczb wymiernych $\boldsymbol{Q}$.

6. n-wymiarowa przestrzeń kartezjańska. Rozważmy $L = \boldsymbol{K}^n = \boldsymbol{K} \times \cdots \times \boldsymbol{K}$ (iloczyn kartezjański $n \geq 1$ czynników). Elementy L będziemy zapisywać bądź jako wiersze $(a_1, \ldots, a_n)$, $a_i \in \boldsymbol{K}$, bądź jako kolumny o wysokości n. Dodawanie i mnożenie przez skalary określamy za pomocą wzorów:

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

Dla $n = 1$ otrzymujemy poprzedni przykład. Jednowymiarowe przestrzenie nad $\boldsymbol{K}$ nazywają się *prostymi* albo *$\boldsymbol{K}$-prostymi*, a dwuwymiarowe — *$\boldsymbol{K}$-płaszczyznami*.

7. Przestrzenie funkcji. Niech S oznacza dowolny zbiór, a $F(S)$ — zbiór funkcji na S o wartościach w $\boldsymbol{K}$, tj. odwzorowań S w $\boldsymbol{K}$. Jak zazwyczaj, dla funkcji $f: S \to \boldsymbol{K}$ symbolem $f(s)$ oznaczona jest jej wartość w punkcie $s \in S$.

Dodawanie i mnożenie funkcji przez skalary jest określone punktowo, tj. dla dowolnych $f, g \in F(S)$ i każdego $a \in \boldsymbol{K}$ przyjmujemy

$$(f+g)(s) = f(s) + g(s) \quad \text{dla każdego } s \in S;$$
$$(af)(s) = a(f(s)) \quad \text{dla każdego } s \in S.$$

W przypadku gdy $S = \{1, \ldots, n\}$, możemy utożsamić $F(S)$ z $\boldsymbol{K}^n$, przyporządkowując funkcji f „wektor" $(f(1), \ldots, f(n))$ jej wartości we wszystkich punktach S. Określenia dodawania i mnożenia są zgodne z tym utożsamieniem.

Każdemu elementowi $s \in S$ można przyporządkować „funkcję delta δ_s, skupioną w $\{s\}$", którą określa się następująco: $\delta_s(s) = 1$, $\delta_s(t) = 0$, gdy $t \neq s$. Jeżeli $S = \{1, \ldots, n\}$, to zamiast $\delta_i(k)$ piszemy zazwyczaj δ_{ik} i funkcję delta nazywamy *symbolem (deltą) Kroneckera.*

Jeżeli zbiór S jest skończony, to każdą funkcję z $F(S)$ można przedstawić w postaci kombinacji liniowej funkcji delta: $f = \sum_{s \in S} f(s)\delta_s$. W istocie, wynika to z porównania dla dowolnego punktu $s \in S$ wartości funkcji występujących po obu stronach powyższej równości. I odwrotnie, jeśli $f = \sum_{s \in S} a_s \delta_s$, to obliczając wartości każdej ze stron w punkcie s, otrzymamy $f(s) = a_s$.

Jeśli zbiór S jest nieskończony, to powyższe stwierdzenie nie jest prawdziwe, a mówiąc dokładniej nie można go wyprowadzić z wyżej podanych definicji, bowiem sumy nieskończonej ilości wektorów nie zostały określone! Niekiedy można określić sumy nieskończone w przestrzeniach wektorowych wyposażonych dodatkowo w pojęcie granicy lub, co na jedno wychodzi, w topologię (por. § 10). Przestrzenie takie są zasadniczym obiektem studium analizy funkcjonalnej.

W przypadku gdy $S = \{1, \ldots, n\}$, funkcja δ_i reprezentowana jest przez wektor $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (jedynka na i-tym miejscu, a zera na wszystkich pozostałych), a równość $f = \sum_{s \in S} f(s)\delta_s$ można przepisać w postaci

$$(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i e_i.$$

8. Warunki liniowe i podprzestrzenie liniowe. W analizie najczęściej rozważa się funkcje o wartościach rzeczywistych, określone na osi $\boldsymbol{R}$ albo na odcinkach $(a, b) \subset \boldsymbol{R}$. Jednakże dla większości zastosowań przestrzeń wszystkich funkcji jest zbyt obszerna i wygodnie rozpatrywać, na przykład, tylko funkcje ciągłe albo tylko różniczkowalne. Zazwyczaj po wprowadzeniu odpowiednich określeń dowodzi się, że suma funkcji ciągłych jest ciągła i iloczyn funkcji ciągłej przez skalar jest funkcją ciągłą, podobnie jak w przypadku różniczkowalności.

Oznacza to, że wszystkie funkcje ciągłe, a także wszystkie różniczkowalne tworzą przestrzeń liniową.

Ogólniej, niech L będzie przestrzenią liniową nad $\boldsymbol{K}$, a $M \subset L$ — takim podzbiorem, który jest podgrupą i jednocześnie jest przeprowadzany w siebie poprzez mnożenie przez skalary. Wówczas M wyposażone w działania indukowane przez działania w L (inaczej mówiąc, z ograniczeniami do M działań określonych w L) nazywa się *podprzestrzenią liniową* L, a warunki określające należenie do M wektorów z L nazywają się *warunkami liniowymi.*

Oto przykład warunków liniowych w przestrzeni kartezjańskiej $\boldsymbol{K}^n$: ustalmy skalary $a_1, \ldots, a_n \in \boldsymbol{K}$ i określmy $M \subset L$ poprzez

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

Koniunkcja warunków liniowych jest także warunkiem liniowym. Innymi słowy, przecięcie dowolnej liczby podprzestrzeni liniowych jest podprzestrzenią liniową (sprawdzić to!). Później wykażemy, że dowolna podprzestrzeń w $\boldsymbol{K}^n$ jest zadana skończoną liczbą warunków postaci (1).

Opisana niżej konstrukcja jest ważnym przykładem warunków liniowych.

9. Przestrzeń sprzężona. Niech L będzie przestrzenią liniową nad $\boldsymbol{K}$. Rozważmy najpierw przestrzeń liniową $F(L)$ *wszystkich* funkcji określonych na L o wartościach w $\boldsymbol{K}$. Funkcję $f \in F(L)$ będziemy nazywać *liniową* (często będziemy ją także nazywać *funkcjonałem liniowym* bądź *formą liniową*), jeśli warunki

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

są spełnione dla każdych $l, l_1, l_2 \in L, a \in \boldsymbol{K}$. Przez indukcję względem liczby składników dowodzimy, że

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

Pokażmy, że *funkcje liniowe tworzą podprzestrzeń liniową w* $F(L)$, a mówiąc inaczej, że „*warunek liniowości funkcji jest warunkiem liniowym*". W istocie, jeśli f, f_1 oraz f_2 są liniowe, to

$$\begin{aligned}(f_1 + f_2)(l_1 + l_2) &= f_1(l_1 + l_2) + f_2(l_1 + l_2) = \\ &= f_1(l_1) + f_1(l_2) + f_2(l_1) + f_2(l_2) = \\ &= (f_1 + f_2)(l_1) + (f_1 + f_2)(l_2).\end{aligned}$$

(Posłużyliśmy się tutaj kolejno: określeniem dodawania funkcji, liniowością f_1, f_2, przemiennością i łącznością dodawania w ciele i powtórnie określeniem dodawania funkcji.) Analogicznie,

$$\begin{aligned}(af)(l_1 + l_2) &= a[f(l_1 + l_2)] + a[f(l_1) + f(l_2)] = a[f(l_1)] + a[f(l_2)] = \\ &= (af)(l_1) + (af)(l_2).\end{aligned}$$

W ten sposób wykazaliśmy, że $f_1 + f_2$ oraz af są funkcjami liniowymi.

Przestrzeń funkcji liniowych na przestrzeni liniowej L nazywa się *przestrzenią sprzężoną z (dualną)* lub *dwoistą do* L i oznacza przez L^*.

W dalszym ciągu poznamy wiele innych konstrukcji przestrzeni liniowych.

10. Uwagi na temat oznaczeń. Oznaczanie jednakowymi symbolami zera i dodawania w $\boldsymbol{K}$ i w L to niezupełnie poprawna, choć bardzo wygodna praktyka. Wszystkie wzory ze zwykłej algebry szkolnej, które mają sens w tej sytuacji pozostają prawdziwe (por. przykłady w p. 3).

A oto dwa przykłady, w których byłoby wskazane użycie innych oznaczeń.

a) Przyjmijmy $L = \{x \in \boldsymbol{R} \mid x > 0\}$. Zauważmy, że L jest grupą przemienną względem mnożenia i wprowadźmy w L mnożenie przez skalary z $\boldsymbol{R}$ wzorem $(a, x) \mapsto x^a$. Łatwo sprawdzić, że spełnione są wszystkie reguły działań wymienione w definicji z p. 2, choć w zwykłej notacji przyjmują niecodzienną postać: wektorem zerowym w L jest 1, zamiast $1l = l$ mamy $x^1 = x$, zamiast $a(bl) = (ab)l$ tożsamość $(x^a)^b = x^{ab}$, zamiast $(a + b)l = al + bl$ tożsamość $x^{a+b} = x^a x^b$, itd.

b) Niech L będzie przestrzenią wektorową nad ciałem liczb zespolonych $\boldsymbol{C}$. Określmy nową przestrzeń wektorową $\overline{L}$ nad $\boldsymbol{C}$ z taką samą grupą addytywną L, ale z mnożeniem zadanym za pomocą wzoru

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

gdzie $\overline{a}$ oznacza liczbę zespoloną sprzężoną z a. Ze wzorów $\overline{a + b} = \overline{a} + \overline{b}$ i $\overline{a \cdot b} = \overline{a} \cdot \overline{b}$ z łatwością wynika, że $\overline{L}$ jest przestrzenią wektorową. Gdybyśmy w jakiejś sytuacji musieli jednocześnie rozpatrywać przestrzeń L i $\overline{L}$, wówczas mogłoby się okazać, że zamiast $\overline{a}l$ wygodniej jest pisać np. $a * l$ albo $a \bullet l$.

11. Uwagi o rysunkach i intuicyjnych wyobrażeniach. Wiele z ogólnych pojęć i twierdzeń algebry liniowej wygodnie ilustrować za pomocą schematów rysunkowych czy szkiców. Chcemy jednak od razu uprzedzić czytelnika o niektórych niebezpieczeństwach związanych z takimi wyobrażeniami.

a) *Małe wymiary.* Żyjemy w przestrzeni trójwymiarowej i nasze rysunki zazwyczaj odwzorowują relacje dwu- lub trójwymiarowe. Natomiast w algebrze liniowej mamy do czynienia z przestrzeniami dowolnego, choć skończonego wymiaru, a w analizie funkcjonalnej także z przestrzeniami nieskończenie wymiarowymi. Naszą „małowymiarową" intuicję można znacznie rozwinąć, ale należy to robić z rozwagą.

Prosty przykład: jak można wyobrazić sobie wzajemne położenie dwóch płaszczyzn w przestrzeni czterowymiarowej? Wyobraźcie sobie dwie przecinające się wzdłuż prostej płaszczyzny w $\boldsymbol{R}^3$, które rozchodząc się w czwarty wymiar odrywają się od tej prostej wszędzie poza początkiem układu.

b) *Ciało liczb rzeczywistych.* Fizyczna przestrzeń $\boldsymbol{R}^3$ jest przestrzenią liniową nad ciałem liczb rzeczywistych. Niezwykłość geometrii przestrzeni liniowej nad ciałem $\boldsymbol{K}$ może być związana z własnościami ciała $\boldsymbol{K}$.

Jako kolejny przykład rozważmy najważniejszy dla mechaniki kwantowej przypadek $\boldsymbol{K} = C$. Prosta nad $\boldsymbol{C}$ to jednowymiarowa przestrzeń kartezjańska $\boldsymbol{C}^1$. Przyzwyczailiśmy się do wyobrażenia, że mnożenie punktów prostej $\boldsymbol{R}^1$ przez liczbę

rzeczywistą a jest albo rozciągnięciem a razy (gdy $a > 1$), albo ściśnięciem a^{-1} razy (gdy $0 < a < 1$), albo kombinacją jednego z tych dwóch z odbiciem prostej (gdy $a < 0$).

Tymczasem rozważane w $\boldsymbol{C}^1$ mnożenie przez liczbę zespoloną a ma naturalną interpretację przy użyciu geometrycznej realizacji $\boldsymbol{C}^1$ jako $\boldsymbol{R}^2$ („płaszczyzna Arganda", zwana inaczej „płaszczyzną zespoloną", której nie należy mylić z $\boldsymbol{C}^2$!). Przy tej interpretacji liczbie $z = x + iy \in \boldsymbol{C}$ odpowiada punkt $(x, y) \in \boldsymbol{R}^2$, a mnożeniu przez liczbę $a \neq 0$ odpowiada rozciągnięcie $|a|$ razy złożone z obrotem o kąt $\arg a$ przeciwnie do ruchu wskazówek zegara. W szczególności w ten sposób zauważamy, wybierając $a = -1$, że „odwrócenie" prostej rzeczywistej $\boldsymbol{R}^1$ jest śladem na $\boldsymbol{R}^1$ obrotu $\boldsymbol{C}^1$ o kąt 180°.

Ogólniej, często bywa pożytecznie wyobrazić sobie n-wymiarową zespoloną przestrzeń $\boldsymbol{C}^n$ jako $2n$-wymiarową rzeczywistą przestrzeń $\boldsymbol{R}^{2n}$ (por. § 12 o kompleksyfikacji i realifikacji).

Następnych przykładów dostarczają skończone ciała skalarów $\boldsymbol{K}$, a w szczególności ciało $\boldsymbol{F}_2 = \{0, 1\}$, ważne w teorii kodowania. W tych przypadkach przestrzenie kartezjańskie mają skończoną liczbę punktów, co czasem pozwala na wygodną interpretację geometrii liniowej nad $\boldsymbol{K}$ przy użyciu dyskretnych modeli. I tak, często bywa wygodnie utożsamić $\boldsymbol{F}_2^n$ z wierzchołkami jednostkowego wielościanu w $\boldsymbol{R}^n$, tj. ze zbiorem punktów postaci $(\varepsilon_1, \ldots, \varepsilon_n)$, gdzie $\varepsilon_i = 0$ lub 1. Dodawanie współrzędnych w $\boldsymbol{F}_2^n$ odbywa się zgodnie z regułami algebry Boole'a: $0 + 1 = 1 + 0 = 1$; $0+0 = 1+1 = 0$. Podprzestrzeń złożona z punktów spełniających $\varepsilon_1 + \cdots + \varepsilon_n = 0$ realizuje najprostszy kod z wykrywaniem błędów. Umawiając się, że punkty $(\varepsilon_1, \ldots, \varepsilon_n)$ przekazują informację tylko pod warunkiem spełnienia $\sum_{i=1}^{n} \varepsilon_i = 0$, w przypadku otrzymania sygnału $(\varepsilon'_1, \ldots, \varepsilon'_n)$, dla którego $\sum_{i=1}^{n} \varepsilon'_i \neq 0$, można być pewnym, że przy przekazywaniu tego sygnału wystąpiły zakłócenia.

c) *Euklidesowość przestrzeni fizycznej.* Mamy tu na myśli fakt, że w przestrzeni jest określone nie tylko dodawanie i mnożenie wektorów przez skalary, ale także ich długość, kąty między nimi, pola powierzchni czy objętości niektórych figur itp. Nasze rysunki z konieczności zawierają odwołania do tych własności „metrycznych", a my machinalnie je przyjmujemy, choć w aksjomatyce ogólnych przestrzeni wektorowych te własności wcale nie występują. Nie należy wyobrażać sobie, że jeden wektor jest od drugiego krótszy, czy że para wektorów tworzy kąt prosty, dopóki w przestrzeni nie została wprowadzona specjalna struktura, powiedzmy abstrakcyjny iloczyn skalarny. Omówieniu takich struktur poświęciliśmy drugą część tej książki.

Ćwiczenia

1. Czy są przestrzeniami wektorowymi nad $\boldsymbol{Q}$ następujące zbiory liczb rzeczywistych:

a) liczby rzeczywiste dodatnie;

b) liczby rzeczywiste nieujemne;

c) liczby całkowite;
d) liczby wymierne o mianowniku $\leq N$;
e) liczby postaci $a + b\pi$, gdzie a, b są dowolnymi liczbami wymiernymi?

2. Niech S oznacza dowolny zbiór, $F(S)$ — przestrzeń funkcji o wartościach w ciele $\boldsymbol{K}$. Które z podanych niżej warunków są warunkami liniowymi:
a) f zeruje się w zadanym punkcie;
b) f przyjmuje wartość jeden w zadanym punkcie;
c) f zeruje się we wszystkich punktach zadanego podzbioru $S_0 \subset S$;
d) f zeruje się co najmniej w jednym punkcie podzbioru $S_0 \subset S$;
e) $f(x) \to 0$ przy $|x| \to \infty$;
f) $f(x) \to 1$ przy $|x| \to \infty$;
g) f ma najwyżej skończoną ilość punktów nieciągłości;
(w punktach e) — g) zakładamy, że $S = \boldsymbol{R}$ i $\boldsymbol{K} = \boldsymbol{R}$)?

3. Niech L będzie przestrzenią liniową funkcji o wartościach rzeczywistych, określonych i ciągłych na odcinku $[-1, 1]$. Które z podanych niżej funkcjonałów na przestrzeni L są liniowe:

a) $f \mapsto \int_{-1}^{1} f(x)dx$;

b) $f \mapsto \int_{-1}^{1} f^2(x)dx$;

c) $f \mapsto f(0)$; (jest to tzw. „funkcja delta" Diraca);

d) $f \mapsto \int_{-1}^{1} f(x)g(x)\,dx$, gdzie g oznacza ustaloną funkcję ciągłą na $[-1, 1]$?

4. Niech $L = \boldsymbol{K}^n$. Które z poniższych warunków dla $(x_1, \ldots, x_n) \in L$ są warunkami liniowymi:
a) $\sum_{i=1}^{n} a_i x_i = 1$; $a_1, \ldots, a_n \in \boldsymbol{K}$;
b) $\sum_{i=1}^{n} x_i = 0$ (rozważyć osobno przypadki $\boldsymbol{K} = \boldsymbol{R}$, $\boldsymbol{K} = \boldsymbol{C}$, $\boldsymbol{K}$ — ciało dwuelementowe lub, ogólniej, ciało charakterystyki 2);
c) $x_3 = 2x_4$?

5. Niech $\boldsymbol{K}$ będzie ciałem o q elementach. Ile elementów zawiera przestrzeń $\boldsymbol{K}^n$? Ile rozwiązań ma równanie $\sum_{i=1}^{n} a_i x_i = 0$?

6. Niech $\boldsymbol{K}^\infty$ oznacza przestrzeń nieskończonych ciągów $(a_1, a_2, a_3, \ldots)$, $a_i \in \boldsymbol{K}$, z dodawaniem i mnożeniem „po współrzędnych". Które z poniższych warunków dla wektorów z $\boldsymbol{K}^\infty$ są warunkami liniowymi:
a) tylko skończona liczba współrzędnych a_i jest różna od zera;
b) tylko skończona liczba współrzędnych a_i jest równa zeru;
c) żadna ze współrzędnych a_i nie równa się 1;

d) warunek Cauchy'ego: dla każdego $\varepsilon > 0$ istnieje taka liczba $N > 0$, że $|a_n - a_m| < \varepsilon$ dla wszystkich $n, m > N$;

e) warunek Hilberta: szereg $\sum_{n=1}^{\infty} |a_n|^2$ jest zbieżny;

f) (a_i) jest ciągiem ograniczonym, tj. istnieja taka stała c zależna tylko od (a_i), że $|a_i| < c$ dla wszystkich i (w punktach d) - f) przyjmujemy $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$)?

7. Niech S oznacza zbiór skończony. Wykazać, że każdy funkcjonał liniowy na $F(S)$ jest jednoznacznie wyznaczony przez rodzinę $\{a_s \mid s \in S\}$ elementów ciała $\boldsymbol{K}$, taką że funkcji f jest przyporządkowana liczba $\sum_{s \in S} a_s f(s)$.

W przypadku gdy n jest liczbą elementów S oraz $a_s = \frac{1}{n}$ dla wszystkich s, funkcjonał $f \mapsto \frac{1}{n} \sum_{s \in S} f(s)$ jest nazywany średnią arytmetyczną funkcji.

Jeśli $\boldsymbol{K} = \boldsymbol{R}$, $a_s > 0$ i $\sum_{s \in S} a_s = 1$, to funkcjonał $\sum_{s \in S} a_s f(s)$ nazywa się średnią ważoną funkcji f (z wagami a_s).

§ 2. Baza i wymiar

1. Definicja. Rodzinę wektorów $\{e_1, \ldots, e_n\}$ przestrzeni liniowej L nazywamy *bazą (skończoną) przestrzeni* L, jeśli każdy wektor $l \in L$ można jednoznacznie przedstawić w postaci kombinacji liniowej $l = \sum_{i=1}^{n} a_i e_i$, $a_i \in \boldsymbol{K}$. Współczynniki a_i nazywamy *współrzędnymi wektora* l względem bazy $\{e_i\}$.

2. Przykłady. a) Wektory $e_i = (0, \ldots, 1, \ldots, 0)$, $1 \leq i \leq n$, stanowią bazę $\boldsymbol{K}^n$.

b) Jeśli zbiór S jest skończony, to funkcje $\delta_s \in F(S)$ stanowią bazę $F(S)$. Oba te stwierdzenia zostały wykazane w § 1.

Jeśli w L została wybrana baza złożona z n wektorów i wektory są zadawane przez współrzędne względem tej bazy, to dodawanie wektorów i mnożenie przez skalary wykonujemy na współrzędnych: $\sum_{i=1}^{n} a_i e_i + \sum_{i=1}^{n} b_i e_i = \sum_{i=1}^{n} (a_i + b_i) e_i$, $a\left(\sum_{i=1}^{n} a_i e_i\right) = \sum_{i=1}^{n} (a a_i) e_i$. Dlatego wybór bazy jest równoznaczny z wyborem utożsamienia L z n-wymiarową przestrzenią kartezjańską.

Zamiast równości $l = \sum_{i=1}^{n} a_i e_i$ będziemy często pisać $l = \vec{a}$, oznaczając przez $\vec{a}$ wektor kolumnowy

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

lub też wektor wierszowy

$$(a_1, \ldots, a_n) = [a_1, \ldots, a_n]^t$$

złożony ze współrzędnych $a_1, \ldots, a_n$; w tych oznaczeniach jawna zależność od bazy nie występuje.

3. Definicja. Przestrzeń L nazywamy *skończenie wymiarową*, jeśli jest albo zerowymiarowa (por. § 1, p. 4), albo ma skończoną bazę. Pozostałe przestrzenie nazywamy *nieskończenie wymiarowymi.*

Wygodnie traktować jako bazę przestrzeni zerowymiarowej zbiór pusty. Ponieważ dla takiej przestrzeni wszystkie rezultaty stają się trywialne, będziemy zazwyczaj ograniczać się do rozważania niepustych baz.

4. Twierdzenie. *W przestrzeni skończenie wymiarowej liczba elementów bazy nie zależy od bazy.*

Liczbę tę nazywamy wymiarem przestrzeni L i oznaczamy $\dim L$ lub $\dim_{\boldsymbol{K}} L$. Gdy $\dim L = n$, to przestrzeń L nazywamy n-wymiarową. W przypadku przestrzeni nieskończenie wymiarowej piszemy $\dim L = \infty$.

Dowód. Niech $\{e_1, \ldots, e_n\}$ będzie bazą L. Wykażemy, że żadna rodzina $\{e'_1, \ldots, e'_m\}$, taka że $m > n$ nie może być bazą przestrzeni L z następującego powodu: istniałoby wówczas przedstawienie wektora $0 = \sum_{i=1}^{m} x_i e'_i$, w którym nie wszystkie x_i byłyby równe zeru. Zatem 0 nie przedstawiałoby się jednoznacznie w postaci kombinacji liniowej wektorów $\{e'_i\}$, bo zawsze istnieje przedstawienie trywialne $0 = \sum_{i=1}^{m} 0e'_i$. Stąd już wynika w całości teza twierdzenia, ponieważ w ten sposób pokazujemy, że żadna baza nie może zawierać więcej elementów od jakiejkolwiek innej.

Przyjmijmy $e'_k = \sum_{i=1}^{n} a_{ik} e_i$ dla $k = 1, \ldots, m$. Dla dowolnych $x_k \in \boldsymbol{K}$ mamy

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

Ponieważ $\{e_i\}$ tworzą bazę przestrzeni L, wektor zerowy ma jednoznaczne przedstawienie $\sum_{k=1}^{n} 0e_k$ w postaci liniowej kombinacji $\{e_k\}$. Dlatego warunek $\sum_{k=1}^{m} x_k e'_k = 0$ jest równoważny z układem jednorodnych równań liniowych względem x_k:

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

Ponieważ liczba niewiadomych m w układzie jest większa od liczby równań n, ma on niezerowe rozwiązanie. To kończy dowód.

5. Uwagi. a) Można by było rozważać dowolne rodziny wektorów i nazywać taką rodzinę bazą, gdy każdy wektor przestrzeni ma jednoznaczne przedstawienie w postaci kombinacji liniowej wektorów tej rodziny. W takim znaczeniu każda przestrzeń liniowa ma bazę, a nieskończenie wymiarowa przestrzeń ma nieskończoną bazę. Jednakże takie pojęcie jest niezbyt użyteczne. Z reguły nieskończenie wymiarowe prze-

strzenie są wyposażone w jakąś topologię, a określenie bazy tak się formułuje, aby wziąć pod uwagę możliwość tworzenia nieskończonych kombinacji liniowych.

b) W tradycyjnych oznaczeniach wektory bazowe ogólnych przestrzeni liniowych indeksowane są liczbami naturalnymi od 1 do n (czasem także od 0 do n), ale nie jest to w istocie konieczne. Naturalny zbiór indeksujący dla bazy $\{\delta_s\}$ w $F(S)$ stanowi sam S.

Można także uważać bazę po prostu za podzbiór L, którego elementy nie są w żaden sposób indeksowane (por. p. 20). Natomiast indeksacja lub raczej porządek elementów bazy jest istotna przy posługiwaniu się formalizmem macierzowym (por. § 4). Jednakże w jeszcze innych zagadnieniach może być użyteczna inna struktura zbioru indeksów elmentów bazy. Tak jest np. gdy S jest grupą skończoną, bo wówczas jest istotne, jak zmieniają się przy mnożeniu indeksy $s \in S$ bazy $\{\delta_s\}$, a przypadkowa numeracja elementów S liczbami naturalnymi zazwyczaj zaciemnia tylko obraz tych relacji.

6. Przykłady. a) Wymiar $\boldsymbol{K}^n$ jest równy n.

b) Wymiar $F(S)$ jest równy liczbie n elementów zbioru S, gdy S jest zbiorem skończonym.

Później nauczymy się obliczać wymiary przestrzeni bez konstruowania ich baz. Jest to bardzo ważne, gdyż wiele niezmienników liczbowych w matematyce jest określonych właśnie jako wymiary („liczby Bettiego" w topologii, indeksy operatorów w teorii równań różniczkowych); tymczasem bazy odpowiednich przestrzeni mogą okazać się trudne do wyznaczenia lub też pozbawione szczególnego znaczenia. Na razie jednak będziemy posługiwać się bazami.

Sprawdzenie tego, że dany układ $\{e_1, \ldots, e_n\}$ jest bazą L rozpada się, zgodnie z defnicją, na dwa odzielne zagadnienia, których rozważanie prowadzi nas do następujących pojęć.

7. Definicja. *Powłoką liniową* rodziny wektorów w L nazywamy zbiór wszystkich ich kombinacji liniowych.

Łatwo sprawdzić, że powłoka liniowa jest podprzestrzenią liniową w L (por. § 1, p. 8). Powłokę liniową $\boldsymbol{K}e_1 + \boldsymbol{K}e_2 + \cdots$ będziemy nazywać także *podprzestrzenią rozpiętą* przez wektory $\{e_i\}$ albo *generowaną* przez rodzinę $\{e_i\}$. Można ją także określić jako *przecięcie wszystkich podprzestrzeni liniowych w L zawierających wszystkie wektory e_i* (wykazać to!). *Rzędem* rodziny wektorów nazywamy wymiar jej powłoki liniowej.

Pierwszą charakterystyczną własnością bazy jest: *jej powłoka liniowa jest równa całej przestrzeni L.*

8. Definicja. Rodzinę wektorów $\{e_i\}$ nazywamy *rodziną liniowo niezależną*, jeśli żadna nietrywialna jej kombinacja liniowa nie jest równa zeru, tj. jeśli z równości $0 = \sum_{i=1}^{n} a_i e_i$ wynika, że wszystkie $a_i = 0$. W przeciwnym przypadku nazywamy ją *liniowo zależną*.

Liniowa niezależność rodziny $\{e_i\}$ oznacza, że *wektor zerowy jednoznacznie wyraża się w postaci kombinacji liniowej elementów rodziny.* Wówczas także każdy inny wektor albo ma jednoznaczne przedstawienie w postaci kombinacji liniowej tych wektorów, albo nie ma żadnego. W istocie, porównując dwa takie wyrażenia $l = \sum_{i=1}^{n} a_i e_i = \sum_{i=1}^{n} a'_i e_i$, otrzymujemy $0 = \sum_{i=1}^{n} (a_i - a'_i) e_i$, skąd wynika $a_i = a'_i$.

Stąd także wynika druga charakterystyczna własność bazy: *jej elementy są liniowo niezależne.*

Koniunkcja tych własności jest równoważna z pierwotną definicją bazy.

Zauważmy jeszcze, że *rodzina wektorów jest liniowo niezależna wtedy i tylko wtedy, gdy jest bazą swojej powłoki liniowej.*

Rodzina wektorów $\{e_1, \ldots, e_n\}$ jest, oczywiście, liniowo zależna, jeśli wśród wektorów e_i jest wektor zerowy lub są przynajmniej dwa jednakowe (dlaczego?).

Ogólniej:

9. Lemat. a) *Rodzina wektorów $\{e_1, \ldots, e_n\}$ jest liniowo zależna wtedy i tylko wtedy, gdy choć jeden spośród wektorów e_i jest kombinacją liniową pozostałych.*

b) *Jeśli rodzina $\{e_1, \ldots, e_n\}$ jest liniowo niezależna, a rodzina $\{e_1, \ldots, e_n, e_{n+1}\}$ jest liniowo zależna, to e_{n+1} jest kombinacją liniową $e_1, \ldots, e_n$.*

Dowód. a) Jeśli $\sum_{i=1}^{n} a_i e_i = 0$ oraz $a_j \neq 0$, to $e_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} (-a_j^{-1} a_i) e_i$.

Odwrotnie, jeśli $e_j = \sum_{i \neq j} b_i e_i$, to $e_j - \sum_{i \neq j} b_i e_i = 0$.

b) Jeśli $\sum_{i=1}^{n+1} a_i e_i = 0$ i nie wszystkie współczynniki a_i są równe zeru, to z pewnością $a_{n+1} \neq 0$, gdyż w przeciwnym razie otrzymalibyśmy nietrywialną zależność liniową pomiędzy wektorami $e_1, \ldots, e_n$. Zatem $e_{n+1} = \sum_{i=1}^{n} (-a_{n+1}^{-1} a_i) e_i$, co kończy dowód lematu.

Niech $E = \{e_1, \ldots, e_n\}$ będzie dowolną skończoną rodziną wektorów przestrzeni L, a $F = \{e_{i_1}, \ldots, e_{i_m}\}$ — jej liniowo niezależną podrodziną. Nazwiemy F *maksymalną*, jeśli każdy element rodziny E wyraża się liniowo przez elementy z F.

10. Stwierdzenie. *Każda liniowo niezależna podrodzina $E' \subset E$ zawiera się w pewnej maksymalnej liniowo niezależnej podrodzinie $F \subset E$. Powłoki liniowe F i E są jednakowe.*

Dowód. Jeśli $E \setminus E'$ zawiera wektor, który nie wyraża się jako kombinacja liniowa elementów z E', to dołączamy go do E'. Na mocy części b) lematu p. 9 tak otrzymana rodzina E'' będzie nadal liniowo niezależna. Powtórzmy następnie tę konstrukcję dla E'', itd. Ponieważ rodzina E jest skończona, ten proces zakończy się otrzymaniem maksymalnej podrodziny F. Dowolny element powłoki liniowej E będzie, oczywiście, wyrażać się liniowo przez wektory należące do rodziny F.

W przypadku $E' = \emptyset$ jako E'' należy wybrać niezerowy element należący do E, jeśli taki istnieje, a w przeciwnym przypadku F jest rodziną pustą.

11. Uwaga. Wynik powyższy pozostaje prawdziwy i dla nieskończonych rodzin E. Dla dowodu należy zastosować indukcję pozaskończoną lub lemat Zorna; por. p. p. 18-20.

Maksymalna rodzina nie musi być jedyna: rozważmy $E = (1,0),(0,1),(1,1)\}$, $E' = \{(1,0)\}$ w $\boldsymbol{K}^2$. Wówczas E' zawiera się w dwóch maksymalnych niezależnych podrodzinach $\{(1,0),(0,1)\}$ i $\{(1,0),(1,1)\}$. Jednakże liczba elementów maksymalnej podrodziny jest jednoznacznie określona; jest ona równa wymiarowi powłoki liniowej E i nazywa się *rzędem* (*rangą*) rodziny E.

Często bywa użyteczne następujące twierdzenie.

12. Twierdzenie o uzupełnianiu do bazy. *Niech* $E' = \{e_1,\dots,e_m\}$ *będzie liniowo niezależną rodziną wektorów skończenie wymiarowej przestrzeni* L. *Wówczas istnieje baza* L *zawierająca* E'.

Dowód. Wybierzmy jakąkolwiek bazę $\{e_{m+1},\dots,e_{m+n}\}$ w L, oznaczmy $E = \{e_1,\dots,e_m,e_{m+1},\dots,e_{m+n}\}$ i niech F będzie maksymalną liniowo niezależną podrodziną E zawierającą E'. Jest ona poszukiwaną bazą dla L.

W istocie, wystarczy sprawdzić, że powłoka liniowa F jest równa L. Na mocy stwierdzenia p. 10 powłoki liniowe F i E są równe, a ta ostatnia jest równa L, gdyż zawiera bazę.

13. Wniosek (monotoniczność wymiaru). *Niech* M *będzie liniową podprzestrzenią* L. *Wówczas* $\dim M \leq \dim L$, *a jeśli* L *jest skończenie wymiarowa, to z równości* $\dim M = \dim L$ *wynika* $M = L$.

Dowód. Jeśli M ma wymiar nieskończony, to i L także. Rzeczywiście, pokażemy najpierw, że w M można znaleźć liniowo niezależną rodzinę wektorów o dowolnej liczbie elementów. Jeśli $\{e_1,\dots,e_n\}$ jest n-elementową liniowo niezależną rodziną, to jej powłoka liniowa $M' \subset M$ nie może być równa M, gdyż wtedy wymiar M równałby się n. Zatem M zawiera wektor e_{n+1} nie będący liniową kombinacją wektorów $\{e_1,\dots,e_n\}$, a teza b) lematu p. 9 pokazuje, że rodzina $\{e_1,\dots,e_n,e_{n+1}\}$ jest liniowo niezależna. Załóżmy teraz, że M ma wymiar nieskończony, a L ma wymiar n. Wówczas, powtarzając rozważania z końca dowodu twierdzenia p. 4, dowodzimy, że dowolne $n+1$ kombinacji liniowych elementów bazy L jest liniowo zależne, co przeczy nieskończoności wymiaru M.

Pozostaje do rozpatrzenia przypadek, gdy M i L są skończenie wymiarowe. Na mocy twierdzenia o uzupełnianiu do bazy możemy wtedy dowolną bazę M uzupełnić do bazy L, skąd wynika $\dim M \leq \dim L$. Wreszcie, jeśli $\dim M = \dim L$, to dowolna baza M musi być też bazą L, gdyż inaczej jej uzupełnienie do bazy w L zawierałoby $> \dim L$ elementów, co jest niemożliwe.

14. Bazy i flagi. Jeden ze standardowych sposobów badania zbiorów S wyposażonych w struktury algebraiczne polega na wyróżnieniu w nich ciągu podzbiorów $S_0 \subset S_1 \subset S_2 \subset \cdots$ lub też $S_0 \supset S_1 \supset S_2 \supset \cdots$ w taki sposób, że przejście od jednego podzbioru do następującego po nim jest szczególnie proste.

Takie ciągi nazwają się w ogólności *filtracjami* (rosnącymi lub malejącymi). W teorii przestrzeni liniowych silnie rosnące ciągi $L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n$ podprzestrzeni przestrzeni L nazywamy *flagami*. (Uzasadnieniem dla takiej nazwy jest ciąg $\{punkt\ 0\} \subset \{prosta\} \subset \{płaszczyzna\}$, co odpowiada fladze — „gwóźdź", „drzewce", „sztandar".)

Liczbę n nazywamy *długością flagi* $L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n$. Flagę $L_0 \subset L_1 \subset \ldots \subset L_n \subset \cdots$ nazywamy *maksymalną*, jeśli $L_0 = \{0\}$, $\cup L_i = L$ oraz dla każdego i między L_i a L_{i+1} nie uda się zawrzeć podprzestrzeni, tj. jeśli $L_i \subset M \subset L_{i+1}$, to albo $L_i = M$, albo $M = L_{i+1}$.

Wychodząc od dowolnej bazy $\{e_1, \ldots, e_n\}$ przestrzeni L, można zbudować flagę długości n, przyjmując $L_0 = \{0\}$ i definiując L_i dla $i \geq 1$ jako powłokę liniową $\{e_1, \ldots, e_i\}$. Z dowodu następnego twierdzenia możemy wywnioskować, że flaga ta jest maksymalna i że w ten sposób otrzymamy każdą flagę maksymalną.

15. Twierdzenie. *Wymiar przestrzeni L jest równy długości dowolnej maksymalnej flagi w L.*

Dowód. Niech $L_0 \subset L_1 \subset L_2 \subset \cdots$ będzie maksymalną flagą w L. Dla każdego $i \geq 1$ wybierzmy wektor $e_i \in L_i \setminus L_{i-1}$. Pokażemy, że $\{e_1, \ldots, e_i\}$ jest bazą dla przestrzeni L_i. Przede wszystkim zaobserwujmy, że powłoka liniowa rodziny $\{e_1, \ldots, e_i\}$ jest zawarta w L_{i-1}, a e_i nie należy do L_{i-1}, skąd za pomocą indukcji względem i (z uwzględnieniem $e_1 \neq 0$) wnioskujemy, że rodzina $\{e_1, \ldots, e_i\}$ jest liniowo niezależna dla każdego i.

Stosując jeszcze raz indukcję względem i pokażemy teraz, że $\{e_1, \ldots, e_i\}$ rozpina L_i. Załóżmy, że jest to prawdziwe dla $i-1$ i niech M oznacza powłokę liniową rodziny $\{e_1, \ldots, e_i\}$. Zatem $L_{i-1} \subset M$ na mocy założenia indukcyjnego oraz $L_{i-1} \neq M$, ponieważ $e_i \notin L_{i-1}$. Z określenia maksymalności flagi wynika zatem $M = L_i$.

Teraz już łatwo zakończyć dowód twierdzenia. Jeśli $L_0 \subset L_1 \subset L_2 \subset \cdots \subset L_n = L$ jest skończoną flagą maksymalną w L, to na mocy powyższego rodzina $\{e_1, \ldots, e_n\}$, $e_i \in L_i \setminus L_{i-1}$ jest bazą L, a więc $\dim L = n$. Jeśli natomiast w L zawarte są nieskończone flagi maksymalne, to powyższa konstrukcja pozwala otrzymać liniowo niezależne rodziny o dowolnej liczbie wektorów i w tym przypadku wymiar L jest nieskończony.

16. Uzupełnienie. W przestrzeni skończenie wymiarowej dowolną flagę można uzupełnić do maksymalnej i dlatego jej długość jest zawsze $\leq \dim L$.

W istocie, uzupełniajmy daną flagę pośrednimi przestrzeniami, tak długo jak to możliwe. Ponieważ tego procesu nie można kontynuować w nieskończoność, gdyż dla dowolnej flagi konstrukcja układów wektorów $\{e_1, \ldots, e_i\}$, $e_i \in L_i \setminus L_{i-1}$ użyta na początku dowodu twierdzenia p. 15 prowadzi do układów liniowo niezależnych, długość flagi nie może przekroczyć n.

17. Podstawowa zasada postępowania z nieskończenie wymiarowymi przestrzeniami: Lemat Zorna lub indukcja pozaskończona. Większość twierdzeń skończenie wymiarowej algebry liniowej wynika bez trudności z faktu istnienia

skończonych baz wraz z twierdzeniem p. 12 o uzupełnianiu do bazy — czytelnik napotka wiele przykładów tego podejścia w dalszym ciągu. Jednakże przyzwyczajenie do baz utrudnia przejście do analizy funkcjonalnej. Opiszemy teraz pewną zasadę teorii mnogości, która w wielu przypadkach zastępuje argumentację wykorzystującą pojęcie bazy.

Przypomnijmy (por. *Wstęp do algebry*, rozdz. 1, § 6), że *zbiorem częściowo uporządkowanym* nazywamy zbiór X wraz z dwuargumentową *relacją porządku* $\leq$ na X, która jest zwrotna ($x \leq x$), przechodnia (jeśli $x \leq y$ i $y \leq z$, to $x \leq z$) i antysymetryczna (jeśli $x \leq y$ i $y \leq x$, to $x = y$). Jednakże w pełni dopuszczalna jest sytuacja, że para elementów $x, y \in X$ nie spełnia ani warunku $x \leq y$, ani warunku $y \leq x$. Przeciwnie, jeśli dla dowolnej pary zachodzi $x \leq y$ lub $y \leq x$, to zbiór ten nazywamy *liniowo uporządkowanym* lub *łańcuchem*.

Ograniczeniem górnym podzbioru Y w częściowo uporządkowanym zbiorze X nazywamy taki element $x \in X$, że $y \leq x$ dla każdego $y \in Y$. Taki element może nie istnieć — w zbiorze $X = \boldsymbol{R}$ ze zwykłą relacją $\leq$ podzbiór $Y = \boldsymbol{Z}$ (liczby całkowite) nie ma ograniczenia górnego.

Elementem największym zbioru częściowo uporządkowanego X nazywamy taki element $n \in X$, że $x \leq n$ dla każdego $x \in X$, a *elementem maksymalnym* — taki element $m \in X$, że z warunku $m \leq x \in X$ wynika $x = m$. Element największy jest także maksymalny, ale niekoniecznie odwrotnie.

18. Przykład. Typowym przykładem zbioru częściowo uporządkowanego jest zbiór wszystkich podzbiorów $P(S)$ zbioru S lub jego część, z relacją $\subseteq$ jako częściowym porządkiem. Jeśli S ma więcej niż dwa elementy, to $P(S)$ jest częściowo uporządkowany, ale nie liniowo uporządkowany (dlaczego?). W tym przypadku elementem maksymalnym, a także największym jest S.

19. Lemat Zorna[1]. *Niech X będzie niepustym zbiorem częściowo uporządkowanym, w którym każdy łańcuch ma ograniczenie górne w X. Wówczas każdy łańcuch ma takie ograniczenie górne, które jest jednocześnie elementem maksymalnym w X.*

Lemat Zorna można wyprowadzić z innych, bardziej intuicyjnych aksjomatów teorii mnogości, jednakże faktycznie jest on równoważny, na gruncie pozostałych aksjomatów, z tak zwanym „pewnikiem wyboru". Dlatego, jak to się często robi, wygodnie zaliczyć go do podstawowych aksjomatów.

20. Przykład zastosowania lematu Zorna: istnienie baz w przestrzeniach liniowych nieskończonego wymiaru. Niech L będzie przestrzenią liniową nad $\boldsymbol{K}$. Oznaczymy przez $X \subseteq P(L)$ *zbiór liniowo niezależnych rodzin wektorów w L, częściowo uporządkowany przez relację $\subseteq$.*

Inaczej mówiąc, $Y \in X$, jeśli dowolna kombinacja liniowa wektorów z Y jest równa zeru tylko wtedy, gdy jej współczynniki są zerami. Sprawdźmy założenie le-

(1) W literaturze polskiej nazywa się go lematem Kuratowskiego–Zorna (*przyp. tłum.*).

matu Zorna: jeśli S jest łańcuchem w X, to w X istnieje ograniczenie górne dla S. Rzeczywiście, niech $Z = \bigcup_{Y \in S} Y$. Dla każdego $Y \in S$ mamy, oczywiście, $Y \subseteq Z$, a poza tym Z jest liniowo niezależną rodziną wektorów, gdyż dowolna skończona rodzina wektorów $\{y_1, \ldots, y_n\}$ należących do Z jest zawarta w jakiejś rodzinie $Y \in S$. W istocie, niech $y_i \in Y_i \in S$; ponieważ S jest łańcuchem, więc z każdych dwóch elementów $Y_i, Y_j \in S$ jeden jest zawarty w drugim; opuszczając kolejno we wszystkich takich parach mniejszą rodzinę, widzimy, że wśród Y_i jest rodzina największa, w której zawierają się wszystkie wektory $y_1, \ldots, y_n$, a więc są one liniowo niezależne. Zastosujmy teraz tezę lematu Zorna, a właściwie wystarczy nam jej część stwierdzająca istnienie w X elementu maksymalnego. Zgodnie z definicją, jest to taka liniowo niezależna rodzina wektorów $Y \in X$, która po dołączeniu do niej dowolnego wektora $l \in L$ przestaje być liniowo niezależna. Dokładnie te same argumenty, które były użyte przy dowodzie części b) lematu w p. 9 pokazują, że l jest wówczas skończoną kombinacją liniową elementów z Y, a więc Y jest bazą dla L.

Ćwiczenia

1. Niech L będzie przestrzenią wektorową wielomianów zmiennej x stopnia $\leq n-1$ o współczynnikach z ciała $\boldsymbol{K}$. Sprawdzić następujące stwierdzenia:

a) $1, x, \ldots, x^{n-1}$ tworzą bazę w L. Współrzędnymi wielomianu f względem tej bazy są współczynniki tego wielomianu.

b) $1, x-a, (x-a)^2, \ldots, (x-a)^{n-1}$ są bazą w L. Jeśli $\operatorname{char} \boldsymbol{K} = p \geq n$, to współrzędnymi wielomianu f względem tej bazy są $\left\{ f(a), f'(a), \frac{f''(a)}{2!}, \ldots, \frac{f^{(n-1)}(a)}{(n-1)!} \right\}$.

c) Dla danych parami różnych elementów $a_1, \ldots, a_n \in \boldsymbol{K}$ zdefiniujemy wielomiany $g_i(x)$, przyjmując $g_i(x) = \prod_{j \neq i} (x - a_j)(a_j - a_i)^{-1}$. Stanowią one bazę w L (zwaną bazą interpolacyjną). Współrzędnymi wielomianu f względem tej bazy są $\{f(a_1), \ldots, f(a_n)\}$.

2. Niech L będzie n-wymiarową przestrzenią i $f: L \to \boldsymbol{K}$ niezerowym funkcjonałem liniowym. Wykazać, że $M = \{l \in L \mid f(l) = 0\}$ jest $(n-1)$-wymiarową podprzestrzenią w L. Ponadto dowieść, że w ten sposób można otrzymać wszystkie $(n-1)$-wymiarowe podprzestrzenie L.

3. Niech L będzie n-wymiarową przestrzenią, $M \subset L$ — m-wymiarową podprzestrzenią. Wykazać istnienie takich funkcjonałów liniowych $f_1, \ldots, f_{n-m} \in L^*$, że $M = \{l \in L \mid f_1(l) = \cdots = f_{n-m}(l) = 0\}$.

4. Obliczyć wymiary następujących przestrzeni:

a) przestrzeni wielomianów n zmiennych stopnia $\leq p$;

b) przestrzeni jednorodnych wielomianów (form) n zmiennych stopnia $\leq p$;

c) przestrzeni funkcji należących do $F(S), gdzie |S| < \infty$, które zerują się we wszystkich punktach podzbioru $S_0 \subset S$.

5. Niech $\boldsymbol{K}$ będzie ciałem skończonym charakterystyki p. Wykazać, że liczba jego elementów jest równa p^n dla pewnego $n \geq 1$. (*Wskazówka.* Rozważyć $\boldsymbol{K}$ jako

przestrzeń liniową nad prostym podciałem, złożonym ze wszystkich "sum jedności" w $\boldsymbol{K}$, tj. $0, 1, 1+1, \dots$).

6. W przestrzeni nieskończenie wymiarowej zamiast pojęcia flagi używa się pojęcia łańcucha podprzestrzeni (uporządkowanych przez zawieranie). Posługując się lematem Zorna dowieść, że każdy łańcuch zawiera się w maksymalnym.

§ 3. Odwzorowania liniowe

1. Definicja. Niech L, M będą przestrzeniami liniowymi nad ciałem $\boldsymbol{K}$. Odwzorowanie $f: L \to M$ nazywamy *odwzorowaniem liniowym*, gdy dla każdych wektorów $l, l_1, l_2 \in L$ i dowolnego $a \in \boldsymbol{K}$ zachodzi

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

Odwzorowania liniowe są homomorfizmami grup addytywnych, mamy zatem $f(0) = 0f(0) = 0$ oraz $f(-l) = f((-1)l) = -f(l)$. Przez indukcję względem n dowodzimy, że dla dowolnych $a_i \in \boldsymbol{K}$, $l_i \in L$ spełniona jest równość $f\left(\sum_{i=1}^{n} a_i l_i\right) = \sum_{i=1}^{n} a_i f(l_i)$. Odwzorowania liniowe $f: L \to L$ nazywamy także *operatorami liniowymi* w przestrzeni L lub *endomorfizmami* tej przestrzeni.

2. Przykłady. a) Odwzorowanie liniowe $f: L \to M$ nazywamy *zerowym*, jeśli $f(l) = 0$ dla wszystkich $l \in L$, *tożsamościowym* — jeśli $f: L \to L$ oraz $f(l) = l$ dla każdego $l \in L$. To ostatnie oznaczamy id_L lub id (ang. *identity*). *Operator mnożenia przez skalar* $a \in \boldsymbol{K}$ lub *homotetia* to odwzorowanie $f: L \to L$, $f(l) = al$ dla każdego $l \in L$. Dla $a = 0$ jest to odwzorowanie zerowe, dla $a = 1$ — tożsamościowe.

b) Odwzorowania liniowe $f: L \to \boldsymbol{K}$ są to rozważane w § 1, p. 9 funkcje liniowe, inaczej funkcjonały liniowe. Niech L będzie przestrzenią o bazie $\{e_1, \dots, e_n\}$. Dla dowolnego $1 \leq i \leq n$ odwzorowanie $e^i: L \to \boldsymbol{K}$, gdzie przez $e^i(l)$ oznaczamy i-tą współrzędną l względem bazy $\{e_1, \dots, e_n\}$, jest funkcjonałem liniowym.

c) Niech $L = \{x \in \boldsymbol{R} \mid x > 0\}$ ze strukturą przestrzeni liniowej nad $\boldsymbol{R}$ opisaną w § 1, p. 10 a), $M = \boldsymbol{R}^1$. Odwzorowanie $\log: L \to M$, $x \mapsto \log x$, jest liniowe (nad $\boldsymbol{R}$).

d) Niech $S \subset T$ będą dwoma zbiorami. Odwzorowanie $\mathcal{F}(T) \to \mathcal{F}(S)$, które dowolnej funkcji określonej na zbiorze T przyporządkowuje jej obcięcie do S, jest liniowe. W szczególności, jeśli $S = \{s\}$, $s \in T$, $f \in \mathcal{F}(T)$, to odwzorowanie $f \mapsto$ (*wartość f w punkcie s*) jest liniowe.

Konstrukcje odwzorowań liniowych o zadanych własnościach często oparte są o następujący wynik.

3. Stwierdzenie. *Niech L, M będą dwiema przestrzeniami liniowymi nad ciałem $\boldsymbol{K}$, $\{l_1, \dots, l_n\} \subset L$ i $\{m_1, \dots, m_n\} \subset M$ — dwiema rodzinami wektorów o jednakowej liczbie elementów. Wtedy:*

a) *jeśli powłoka liniowa* $\{l_1, \ldots, l_n\}$ *jest równa* L, *to istnieje najwyżej jedno odwzorowanie liniowe* $f: L \to M$, *takie że* $f(l_i) = m_i$ *dla każdego* i;

b) *jeśli ponadto rodzina* $\{l_1, \ldots, l_n\}$ *jest liniowo niezależna, a więc tworzy bazę* L, *to takie odwzorowanie istnieje.*

Dowód. Niech f, f' będą takimi odwzorowaniami, że $f(l_i) = f'(l_i) = m_i$ dla każdego i. Rozważmy odwzorowanie $g = f - f'$, gdzie $(f - f')(l) = f(l) - f'(l)$, które, jak łatwo sprawdzić, jest liniowe. Ponadto odwzorowuje ono każdy z wektorów l_i, a więc także każdą ich kombinację liniową, na wektor zerowy. A więc f i f' przyjmują jednakowe wartości na każdym wektorze z L, tj. $f = f'$. Załóżmy teraz, że $\{l_1, \ldots, l_n\}$ tworzą bazę L. Ponieważ każdy wektor L zapisuje się jednoznacznie w postaci $\sum_{i=1}^{n} a_i l_i$, możemy zdefiniować odwzorowanie (zbiorów) $f: L \to M$ wzorem

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

Sprawdzenie, że to odwzorowanie jest liniowe nie przedstawia trudności.

W powyższym dowodzie posłużyliśmy się różnicą dwóch odwzorowań liniowych $L \to M$. Jest to szczególny przypadek następującej ogólniejszej konstrukcji.

4. Oznaczmy przez $\mathcal{L}(L, M)$ zbiór odwzorowań liniowych z L w M. Dla $f, g \in \mathcal{L}(L, M)$ i $a \in \boldsymbol{K}$ określimy af i $f + g$ wzorami

$$(f + g)(l) = f(l) + g(l), \quad (af)(l) = a(f(l))$$

dla dowolnego $l \in L$. Tak samo jak w § 1, p. 9 sprawdzamy, że af i $f + g$ są liniowe, co pokazuje, że $\mathcal{L}(L, M)$ jest przestrzenią liniową.

5. Niech $f \in \mathcal{L}(L, M)$ i $g \in \mathcal{L}(M, N)$. Teoriomnogościowe złożenie $g \circ f = gf: L \to N$ jest odwzorowaniem liniowym. Rzeczywiście,

$$(gf)(l_1 + l_2) = g[f(l_1 + l_2)] = g[f(l_1) + f(l_2)] = g[f(l_1)] + g[f(l_2)] = \\ = gf(l_1) + gf(l_2)$$

i analogicznie,

$$(gf)(al) = a(gf(l)).$$

Jest jasne, że $\mathrm{id}_M \circ f = f \circ \mathrm{id}_L = f$. Ponadto, $h(gf) = (hg)f$, jeśli obie strony równości są określone, tak że można opuścić nawiasy; to ogólna własność łączności teoriomnogościowego złożenia odwzorowań. Wreszcie, złożenie jest liniowe względem każdego z argumentów przy ustalonym drugim: np. $g \circ (af_1 + bf_2) = a(g \circ f_1) + b(g \circ f_2)$.

6. Niech $f \in \mathcal{L}(L, M)$ będzie odwzorowaniem bijektywnym, tak że odwzorowanie odwrotne $f^{-1}: M \to L$ jest dobrze określone (jako odwzorowanie zbiorów). Twierdzimy, że jest ono automatycznie liniowe. W tym celu należy sprawdzić, że

$$f^{-1}(m_1 + m_2) = f^{-1}(m_1) + f^{-1}(m_2), \quad f^{-1}(am_1) = af^{-1}(m_1)$$

dla każdych $m_1, m_2 \in M$ i $a \in \boldsymbol{K}$. Ponieważ f jest bijekcją, więc istnieją jednoznacznie wyznaczone wektory $l_1, l_2 \in L$, takie że $m_i = f(l_i)$. Napisawszy równości

$$f(l_1 + l_2) = f(l_1) + f(l_2), \quad f(al_1) = af(l_1),$$

zastosujmy do obu stron każdej z nich f^{-1}, skąd, zamieniając w otrzymanych wzorach l_i na m_i, otrzymamy żądany wynik.

Bijektywne odwzorowania liniowe $f: L \to M$ nazywamy *izomorfizmami*. Przestrzenie L i M nazywamy *izomorficznymi*, gdy istnieje między nimi izomorfizm.

Następujące twierdzenie pokazuje, że wymiar przestrzeni określa ją z dokładnością do izomorfizmu.

7. Twierdzenie. *Dwie skończenie wymiarowe przestrzenie liniowe nad ciałem $\boldsymbol{K}$ są izomorficzne wtedy i tylko wtedy, gdy mają jednakowy wymiar.*

D o w ó d. Izomorfizm $f: L \to M$ zachowuje wszystkie własności przestrzeni opisywane w terminach kombinacji liniowych. W szczególności przeprowadza on dowolną bazę L na bazę M, a więc wymiary L i M są równe. (Z tego argumentu także wynika, że przestrzeń skończenie wymiarowa nie może być izomorficzna z przestrzenią nieskończenie wymiarową.) Odwrotnie, niech L i M mają ten sam wymiar. Wybierzmy bazy $\{l_1, \ldots, l_n\}$ i $\{m_1, \ldots, m_n\}$ w L i odpowiednio w M. Wzór

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

określa, na mocy stwierdzenia p. 3, odwzorowanie liniowe. Jest ono także bijekcją, gdyż wzór

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

wyraża odwzorowanie odwrotne f^{-1}.

8. Ostrzeżenie. Nawet jeśli istnieje izomorfizm między przestrzeniami L i M, to jest on jednoznacznie wyznaczony tylko w dwóch przypadkach:

a) $L = M = \{0\}$,

b) L i M są jednowymiarowe i ciało $\boldsymbol{K}$ jest ciałem dwuelementowym (spróbujcie to wykazać!).

We wszystkich innych przypadkach istnieje wiele (a nawet, jeśli $\boldsymbol{K}$ jest nieskończone, to nieskończenie wiele) izomorfizmów. W szczególności istnieje wiele izomorfizmów przestrzeni L z nią samą. Na mocy wyników punktów 5 i 6 tworzą one grupę względem złożenia. Grupę tę nazywamy *pełną grupą liniową przestrzeni L*. Później będziemy mogli ją opisać w sposób bardziej konkretny jako grupę nieosobliwych macierzy kwadratowych.

Czasami zdarza się i tak, że między dwiema przestrzeniami liniowymi istnieje izomorfizm niezależny od jakichkolwiek szczególnych wyborów (takich jak wybór baz

w przestrzeniach L i M dokonany w dowodzie twierdzenia p. 7). Takie izomorfizmy będziemy nazywać *kanonicznymi* lub *naturalnymi* (jednak dokładne określenie tych terminów jest możliwe tylko przy użyciu języka teorii kategorii, zob. § 13). Należy starannie odróżniać izmorfizmy naturalne od „przypadkowych". Podamy teraz dwa charakterystyczne przykłady, bardzo ważne dla zrozumienia tej różnicy.

9. Izomorfizm „przypadkowy" przestrzeni z przestrzenią do niej sprzężoną. Niech L będzie przestrzenią skończenie wymiarową, a $\{e_1, \ldots, e_n\}$ jej bazą. Oznaczmy przez $e^i \in L^*$ funkcjonał liniowy $l \mapsto e^i(l)$, gdzie $e^i(l)$ oznacza i-tą współrzędną l względem bazy $\{e_i\}$ (nie należy mylić tego oznaczenia z i-tą potęgą, która w przestrzeni liniowej nie jest określona). Twierdzimy, że *funkcjonały liniowe* $\{e^1, \ldots, e^n\}$ *tworzą bazę przestrzeni* L^*, którą nazywać będziemy *bazą sprzężoną (dwoistą)* z $\{e_1, \ldots, e_n\}$. Funkcjonały $\{e^i\}$ można równoważnie opisać następująco: $e^i(e_k) = \delta_{ik}$ (jest to symbol Kroneckera: $\delta_{ik} = 1$ dla $i = k$ oraz $\delta_{ik} = 0$ dla $i \neq k$).

W istocie, każdy funkcjonał liniowy $f: L \to \boldsymbol{K}$ można zapisać jako kombinację liniową $\{e^i\}$:

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

Wynika to stąd, że wartości lewej i prawej strony są takie same dla każdej kombinacji liniowej $\sum_{k=1}^{n} a_k e_k$, bo $e^i\left(\sum_{k=1}^{n} a_k e_k\right) = a_i$ na mocy określenia e^i. Ponadto $\{e^i\}$ są liniowo niezależne: jeśli bowiem $\sum_{i=1}^{n} a_i e^i = 0$, to dla każdego k, $1 \leq k \leq n$, mamy $a_k = \left(\sum_{i=1}^{n} a_i e^i\right)(e_k) = 0$.

Zatem L i L^* mają ten sam wymiar n, i co więcej, istnieje izomorfizm $f: L \to L^*$, który przeprowadza e_i w e^i.

Jednakże *ten izomorfizm nie jest kanoniczny*: zamiana bazy $\{e_1, \ldots, e_n\}$ na inną zmienia, na ogół, także i ten izomorfizm. Na przykład, jeśli wymiar L jest równy 1, to dla dowolnego niezerowego wektora $e_1 \in L$ zbiór $\{e_1\}$ jest bazą. Niech $\{e_1\}$ oznacza bazę dwoistą do $\{e_1\}$, $e^1(e_1) = 1$. Wtedy bazą dwoistą do bazy $\{ae_1\}$, gdzie $a \in \boldsymbol{K} \setminus \{0\}$, jest baza $\{a^{-1}e^1\}$. Tymczasem odwzorowanie $f_1: e_1 \mapsto e^1$ jest różne od odwzorowania $f_2: ae_1 \mapsto a^{-1}e^1$, jeśli tylko $a^2 \neq 1$.

10. Izomorfizm kanoniczny przestrzeni liniowej z jej dwusprzężoną. Niech L będzie przestrzenią liniową, L — przestrzenią funkcji liniowych na L, $L^{**} = (L^*)^*$ — przestrzenią funkcji liniowych na L^*, zwaną *dwusprzężoną* lub *drugą sprzężoną* z przestrzenią L. Skonstruujemy teraz odwzorowanie kanoniczne $\varepsilon_L: L \to L^{**}$ w sposób niezależny od jakichkolwiek dowolnych wyborów. *Przyporządkowuje ono każdemu wektorowi* $l \in L$ *funkcję na* L^*, *której wartością na funkcjonale* $f \in L^*$ *jest* $f(l)$, a w skrócie

$$\varepsilon_L : l \longmapsto [f \longmapsto f(l)].$$

Zauważmy, że ε_L ma następujące własności:

a) *Dla każdego* $l \in L$ *odwzorowanie* $\varepsilon_L(l)\colon L^* \to \boldsymbol{K}$ *jest liniowe.* Istotnie, oznacza to, że *wyrażenie* $f(l)$ *jest liniowe jako funkcja* f *przy ustalonym* l, a to wynika wprost z określenia dodawania funkcjonałów i mnożenia ich przez liczby (por. § 1, p. 7). W ten sposób wykazaliśmy, że ε_L rzeczywiście określa odwzorowanie L w L^{**}, jak twierdziliśmy wyżej.

b) *Odwzorowanie* $\varepsilon_L\colon L \to L^{**}$ *jest liniowe.* Rzeczywiście, oznacza to, że wyrażenie $f(l)$ jest liniowe jako funkcja l przy ustalonym f, a to zachodzi, gdyż $f \in L^*$.

c) *Jeśli* L *jest skończenie wymiarowa, to odwzorowanie* $\varepsilon_L\colon L \to L^{**}$ *jest izomorfizmem.* Rzeczywiście, niech $\{e_1, \ldots, e_n\}$ będzie bazą L, $\{e^1, \ldots, e^n\}$ — bazą L^* do niej dwoistą, $\{e'_1, \ldots, e'_n\}$ — bazą L^{**} dwoistą do bazy $\{e^1, \ldots, e^n\}$.

Pokażemy, że $\varepsilon_L(e_i) = e'_i$, skąd już wynika, że ε_L jest izomorfizmem (posługiwanie się w tym miejscu dowodu bazami jest bez znaczenia wobec tego, że w określeniu ε_L nie były one użyte!).

Istotnie, zgodnie z określeniem $\varepsilon_L(e_i)$ jest funkcjonałem na L^*, którego wartość na e^k jest równa $e^k(e_i) = \delta_{ik}$ (znowu „symbol Kroneckera"). Ale według określenia bazy dwoistej właśnie e'_i jest takim funkcjonałem na L^*.

Zauważmy, że jeśli L jest nieskończenie wymiarowa, to $\varepsilon_L\colon L \to L^{**}$ jest w dalszym ciągu injektywne, ale nie jest na ogół surjektywne (por. ćwiczenie 2). W analizie funkcjonalnej zamiast całej przestrzeni L^* zazwyczaj rozważa się tylko podprzestrzeń L' funkcjonałów liniowych ciągłych w odpowiednio wybranych topologiach na L i $\boldsymbol{K}$ i wtedy można określić odwzorowanie $L \to L''$, które niejednokrotnie także jest izomorfizmem. Tego rodzaju (topologiczne) przestrzenie liniowe nazywają się *refleksywnymi.* W ten sposób pokazaliśmy, że *przestrzenie skończenie wymiarowe (bez względu na topologię) są refleksywne.*

Rozważymy teraz związek między odwzorowaniami liniowymi a podprzestrzeniami liniowymi.

11. Definicja. Niech $f\colon L \to M$ będzie odwzorowaniem liniowym. Zbiór $\operatorname{Ker} f = \{l \in L \mid f(l) = 0\} \subset L$ nazywamy *jądrem* f, a zbiór $\operatorname{Im} f = \{m \in M \mid \exists l \in L, f(l) = m\} \subset M$ — *obrazem* f.

Nietrudno przekonać się, że jądro f jest podprzestrzenią liniową L, a obraz f — podprzestrzenią liniową M. Sprawdźmy, na przykład, drugi z wymienionych faktów. Niech $m_1, m_2 \in \operatorname{Im} f$, $a \in \boldsymbol{K}$. Wówczas istnieją takie wektory $l_1, l_2 \in L$, że $f(l_1) = m_1$, $f(l_2) = m_2$, a to znaczy, że $m_1 + m_2 = f(l_1 + l_2)$, $am_1 = f(al_1)$. A więc $m_1 + m_2 \in \operatorname{Im} f$ oraz $am_1 \in \operatorname{Im} f$.

Odwzorowanie f jest injektywne wtedy i tylko wtedy, gdy $\operatorname{Ker} f = \{0\}$. Istotnie, jeśli $f(l_1) = f(l_2)$, $l_1 \neq l_2$, to $0 \neq l_1 - l_2 \in \operatorname{Ker} f$. I na odwrót, jeśli $0 \neq l \in \operatorname{Ker} f$, to $f(l) = 0 = f(0)$.

12. Twierdzenie. *Niech* L *będzie przestrzenią skończenie wymiarową,* $f\colon L \to M$ *— odwzorowaniem liniowym. Wówczas* $\operatorname{Ker} f$ *i* $\operatorname{Im} f$ *są skończenie wymiarowe oraz*

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

Dowód. Skończoność wymiaru jądra f wynika z wniosku p. 13, § 2. Wybierzmy bazę $\{e_1, \ldots, e_m\}$ w $\operatorname{Ker} f$ i uzupełnijmy ją do bazy $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_{m+n}\}$ przestrzeni L zgodnie z twierdzeniem p. 12, § 2. Wykażemy, że wektory $f(e_{m+1})$, $\ldots, f(e_{m+n})$ tworzą bazę $\operatorname{Im} f$, skąd oczywiście wynika już twierdzenie.

Dowolny wektor należący do $\operatorname{Im} f$ ma postać

$$f\Big(\sum_{i=1}^{m+n} a_i e_i\Big) = \sum_{i=m+1}^{m+n} a_i f(e_i),$$

co pokazuje, że $f(e_{m+1}), \ldots, f(e_{m+n})$ generują $\operatorname{Im} f$.

Załóżmy, że $\sum_{i=m+1}^{m+n} a_i f(e_i) = 0$, a więc także $f\big(\sum_{i=m+1}^{m+n} a_i e_i\big) = 0$, a to oznacza, że $\sum_{i=m+1}^{m+n} a_i e_i \in \operatorname{Ker} f$, tj. $\sum_{i=m+1}^{m+n} a_i e_i = \sum_{j=1}^{m} a_j e_j$. Ta ostatnia równość jest możliwa tylko wtedy, gdy wszystkie współczynniki są zerami, gdyż $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_{m+n}\}$ jest bazą L. Wykazaliśmy zatem, że wektory $f(e_{m+1}), \ldots, f(e_{m+n})$ są liniowo niezależne, a wraz z tym i całe twierdzenie.

13. Wniosek. *Następujące własności f są równoważne (przy założeniu skończoności wymiaru L):*

a) *f jest injektywne;*

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

Dowód. Na mocy twierdzenia, $\dim L = \dim \operatorname{Im} f$ wtedy i tylko wtedy, gdy $\dim \operatorname{Ker} f = 0$, tj. $\operatorname{Ker} f = \{0\}$.

Ćwiczenia

1. Niech $f: \boldsymbol{R}^m \to \boldsymbol{R}^n$ będzie odwzorowaniem zadanym przez funkcje różniczkowalne, w ogólności nieliniowe, ale przeprowadzające zero na zero:

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

Przyporządkujmy mu odwzorowanie liniowe $df_0: \boldsymbol{R}^m \to \boldsymbol{R}^n$, nazywane *różniczką* f w punkcie 0 i zdefiniowane wzorem

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

gdzie przez $\{e_j\}$, $\{e_i'\}$ oznaczyliśmy, odpowiednio, standardowe bazy $\boldsymbol{R}^m$ i $\boldsymbol{R}^n$. Wykazać, że jeśli przeprowadzić zamianę baz w przestrzeniach $\boldsymbol{R}^m$ i $\boldsymbol{R}^n$ i obliczyć df_0 dla nowych baz według tego samego wzoru, to nowo otrzymane odwzorowanie liniowe df_0 będzie identyczne z poprzednim.

2. Wykazać, że przestrzeń wielomianów $\boldsymbol{Q}[x]$ nie jest izomorficzna ze swoją przestrzenią sprzężoną. (*Wskazówka.* Porównać moc tych przestrzeni.)

§ 4. Macierze

1. W tym paragrafie naszym celem będzie wprowadzenie opisu macierzowego i przedstawienie podstawowych zależności łączących go z formalizmem przestrzeni i odwzorowań liniowych. Po szczegóły i przykłady odsyłamy czytelnika do rozdziałów 2 i 3 *Wstępu do algebry*; w szczególności będziemy swobodnie posługiwać się rozwiniętą tam teorię wyznaczników nie powtarzając jej tutaj. Czytelnik powinien samodzielnie przekonać się, że wszystko co tam zostało powiedziane, można przenieść bez zmian z przypadku ciała liczb rzeczywistych na przypadek dowolnego ciała skalarów, z jedynym wyjątkiem tych stwierdzeń, które wykorzystywały takie specyficzne własności liczb rzeczywistych, jak uporządkowanie czy ciągłość.

2. Terminologia. *Macierzą A o wymiarach $m \times n$ i elementach ze zbioru S* nazywamy rodzinę (a_{ik}) elementów z S, indeksowaną parami uporządkowanymi liczb (i,k), gdzie $1 \leq i \leq m$, $1 \leq k \leq n$. Często będziemy pisać $A = (a_{ik})$, $1 \leq i \leq m$, $1 \leq k \leq n$, a jawne podanie wymiarów macierzy czasami będziemy opuszczać. Przy ustalonym i rodzinę $(a_{i1}, \ldots, a_{in})$ nazywamy *i-tym wierszem macierzy A*, a przy ustalonym k rodzinę $(a_{1k}, \ldots, a_{mk})$ — *k-tą kolumną macierzy A*. Macierze o wymiarach $1 \times n$ i $m \times 1$ nazywamy odpowiednio *wierszem* i *kolumną*.

Jeli $m = n$, to macierz A nazywamy *macierzą kwadratową*, a zamiast „macierz o wymiarach $n \times n$" będziemy częściej mówić „macierz kwadratowa stopnia n".

Jeśli A jest macierzą kwadratową stopnia n, $S = \boldsymbol{K}$ (ciało skalarów) i $a_{ik} = 0$ przy $i \neq k$, to A nazywamy *macierzą diagonalną*; czasem będziemy w tej sytuacji używać oznaczenia $\operatorname{diag}(a_{11}, \ldots, a_{nn})$. W ogólności elementy (a_{ii}) nazywamy *elementami głównej przekątnej* (*diagonali*). Elementy $a_{1,k+1}; a_{2,k+2}; \ldots$ dla $k > 0$ tworzą przekątną położoną powyżej głównej przekątnej, a elementy $a_{k+1,1}; a_{k+2,2}; \ldots$ dla $k > 0$ — przekątną położoną poniżej głównej przekątnej. Jeśli $S = \boldsymbol{K}$ i $a_{ik} = 0$ dla $k < i$, to macierz nazywamy *macierzą górnotrójkątną*, a jeśli $a_{ik} = 0$ dla $i < k$ — *dolnotrójkątną*. Diagonalną macierz kwadratową o elementach z $\boldsymbol{K}$, której wszystkie elementy leżące na diagonali są równe, nazywamy *macierzą skalarną*. Jeśli elementy te są równe jedności, to macierz nazywa się *jednostkowa*. Macierz jednostkowa stopnia n jest oznaczana E_n lub po prostu E, jeśli z kontekstu wynika, jaki jest jej stopień.

Wszystkie te nazwy są związane ze standardową formą zapisu macierzy w postaci tablicy

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

Macierz A^t transponowana do macierzy A o wymiarach $m \times n$ ma wymiary $n \times m$, a jej element położony na przecięciu i-tego wiersza i k-tej kolumny jest równy a_{ki}.

(Używany czasem sposób zapisu $A^t = (a_{ki})$ jest niejednoznaczny!)

3. Uwagi. Elementy macierzy występujących w teorii przestrzeni liniowych nad ciałem $\boldsymbol{K}$ są zazwyczaj elementami ciała $\boldsymbol{K}$. Jednakże zdarzają się też wyjątki — czasem, na przykład, będziemy traktować uporządkowaną bazę $\{e_1, \ldots, e_n\}$ przestrzeni L jako macierz o wymiarach $1 \times n$ i elementach z przestrzeni L. Inny przykład to tzw. *macierze blokowe*, których elementami są z kolei macierze — bloki wyjściowej macierzy. A mianowicie, każdy podział zbioru wskaźników wierszy $[1, \ldots, m] = I_1 \cup I_2 \cup \ldots \cup I_\mu$ oraz zbioru wskaźników kolumn $[1, \ldots, n] = J_1 \cup J_2 \cup \ldots \cup J_\nu$ na kolejno położone, rozłączne odcinki określa rozkład macierzy A na bloki

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

gdzie $A_{\alpha\beta}$ jest macierzą o elementach a_{ik}, $i \in I_\alpha$, $k \in J_\beta$. Jeśli $\mu = \nu$, to w oczywisty sposób można określić pojęcia macierzy blokowo-diagonalnych, blokowo górnotrójkątnych i blokowo dolnotrójkątnych. Ten sam przykład pokazuje także, że nie zawsze jest wygodnie indeksować wiersze i kolumny macierzy liczbami z zakresu 1 do m (odpowiednio n); często istotna jest tylko kolejność wierszy i kolumn.

4. Macierz odwzorowania liniowego. Niech N i M będą skończenie wymiarowymi przestrzeniami liniowymi nad $\boldsymbol{K}$ z wyróżnionymi bazami $\{e_1, \ldots, e_n\}$ i $\{e'_1, \ldots, e'_m\}$ odpowiednio. Dowolnemu odwzorowaniu liniowemu $f: N \to M$ przyporządkujemy w następujący sposób macierz A_f o wymiarach $m \times n$ i elementach z ciała skalarów $\boldsymbol{K}$ — zauważmy, że wymiarami macierzy są wymiary przestrzeni N, M wzięte w odwrotnej kolejności. Zapiszemy wektory $f(e_k)$ jako kombinacje liniowe wektorów $e_1, \ldots, e_m$, $f(e_k) = \sum_{i=1}^{m} a_{ik} e_i$ i przyjmiemy $A_f = (a_{ik})$. Inaczej mówiąc, współczynniki tych kombinacji liniowych rozmieszczamy w kolejnych kolumnach macierzy A_f.

Macierz A_f nazywamy *macierzą odwzorowania liniowego f względem baz* (albo *w bazach*) $\{e_k\}$, $\{e_i\}$.

Na mocy stwierdzenia § 3, p. 3 odwzorowanie liniowe f jest jednoznacznie wyznaczone przez wektory $f(e_k)$, przy czym jako te ostatnie można przyjąć dowolną rodzinę n wektorów przestrzeni M. Dlatego też wyżej opisane przyporządkowanie jest bijekcją między zbiorem $\mathcal{L}(N, M)$, a zbiorem macierzy o wymiarach $m \times n$ i elementach z $\boldsymbol{K}$ (lub, jak też mówimy, *macierzy nad $\boldsymbol{K}$*). Jednakże ta bijekcja zależy od wyboru baz (por. p. 8 poniżej).

Macierz A_f pozwala także opisać odwzorowanie liniowe f poprzez jego działanie na współrzędne wektorów. Jeśli przedstawić wektor $l \in L$ jako wektor-kolumnę

$\vec{x} = [x_1, \ldots, x_n]$ współrzędnych względem bazy $\{e_1, \ldots, e_n\}$, tj. $l = \sum_{i=1}^{n} x_i e_i$, to $f(l)$ można przedstawić w postaci wektora-kolumny $\vec{y} = [y_1, \ldots, y_m]$, przy czym

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

Innymi słowy $\vec{y} = A_f \cdot \vec{x}$ jest iloczynem macierzy A_f przez wektor-kolumnę $\vec{x}$.

Mówiąc o macierzy *operatora liniowego* $A = (a_{ik})$, będziemy milcząco zakładać, że w obu „egzemplarzach" przestrzeni M została wybrana jedna i ta sama baza. Macierz operatora liniowego jest macierzą kwadratową, a macierz operatora tożsamościowego jest macierzą jednostkową.

Zgodnie z § 3, p. 4 zbiór $\mathcal{L}(N, M)$ ma strukturę przestrzeni liniowej nad $\boldsymbol{K}$. Przy utożsamieniu elementów $\mathcal{L}(N, M)$ z macierzami tę strukturę można opisać następująco.

5. Dodawanie macierzy i mnożenie ich przez skalary. Niech $A = (a_{ik})$, $B = (b_{ik})$ będą dwiema macierzami o tych samych wymiarach nad ciałem $\boldsymbol{K}$ i niech $a \in \boldsymbol{K}$. Zdefiniujemy

$$A + B = (c_{ik}), \quad \text{gdzie } c_{ik} = a_{ik} + b_{ik},$$

oraz

$$aA = (aa_{ik}).$$

Działania te określają strukturę przestrzeni liniowej w zbiorze macierzy o ustalonych wymiarach. Ponadto łatwo sprawdzić, że jeśli $A = A_f$, $B = B_g$ (względem tej samej bazy), to

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

tak że opisana odpowiedniość (przypomnijmy, że jest ona bijekcją) jest izomorfizmem. W szczególności $\dim \mathcal{L}(N, M) = \dim N \cdot \dim M$, ponieważ przestrzeń macierzy (o wymiarach $m \times n$) jest izomorficzna z $\boldsymbol{K}^{mn}$.

Superpozycję odwzorowań liniowych opisujemy poprzez mnożenie macierzy.

6. Mnożenie macierzy. Iloczyn macierzy A o wymiarach $m \times n'$ i elementach z $\boldsymbol{K}$ przez macierz B o wymiarach $n \times p$ i elementach z $\boldsymbol{K}$ jest określony tylko wtedy, gdy $n'' = n' = n$, a iloczyn AB ma wówczas wymiary $m \times p$ i jest zdefiniowany wzorem

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

Jak nietrudno sprawdzić, $(AB)^t = B^t A^t$. Może się zdarzyć, że choć iloczyn AB jest określony, to iloczyn BA nie jest (gdy $m \neq p$) bądź też że oba iloczyny AB i BA są dobrze określone, ale mają inne wymiary (tak jest, gdy $m \neq n$), bądź nawet oba są dobrze określone i mają jednakowe wymiary, ale nie są równe. Innymi słowy,

mnożenie macierzy nie jest przemienne. Jednakże jest ono *łączne; pod warunkiem, że iloczyny AB i BC są określone, to także (AB)C i A(BC) są określone i jednakowe.* Istotnie, niech $A = (a_{ij})$, $B = (b_{jk})$, $C = (c_{kl})$. Sprawdzenie zgodności wymiarów A z BC i AB z C możemy pozostawić czytelnikowi, a zakładając, że z tym nie ma już wątpliwości, możemy obliczyć (il)-ty element $(AB)C$ ze wzoru

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

a (il)-ty element $A(BC)$ ze wzoru

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

Ponieważ mnożenie w $\boldsymbol{K}$ jest łączne, oba wyrażenia są równe. Korzystając z już udowodnionej łączności mnożenia macierzy, można sprawdzić, że „mnożenie blokami" macierzy blokowych jest także łączne (por. także ćwiczenie 1). Ponadto, iloczyn macierzy jest liniowy względem każdego z czynników z osobna:

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

Najistotniejszą cechą mnożenia macierzy jest to, że odpowiada ono składaniu odwzorowań liniowych. Także wiele innych problemów w algebrze liniowej wygodnie jest opisywać za pomocą mnożenia macierzy: to właśnie jest główną przyczyną unifikującej roli odgrywanej przez język algebry macierzowej i pewnej samodzielności algebry macierzowej wewnątrz algebry liniowej. Rozważmy niektóre z takich sytuacji.

7. Macierz złożenia odwzorowań liniowych. Niech P, N, M będą trzema skończenie wymiarowymi przestrzeniami liniowymi, $P \xrightarrow{g} N \xrightarrow{f} M$ — dwoma odwzorowaniami liniowymi. Wybierzmy bazy $\{e_i''\}$, $\{e_k'\}$, $\{e_m\}$ w P, N, M odpowiednio i oznaczmy przez A_g, A_f i A_{fg} macierze g, f i fg w tych bazach. Sprawdzimy, że $A_{fg} = A_f A_g$. W istocie, niech $A_f = (a_{ji})$, $A_g = (b_{ik})$. Mamy

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

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

Stąd wynika, że (j,k)-ty element macierzy A_{fg} jest równy $\sum_i a_{ji}b_{ik}$, tj. $A_{fg} = A_f A_g$.

Na mocy rezultatów p. p. 4–6 przez wybór bazy można utożsamić zbiór operatorów liniowych $\mathcal{L}(L,L)$ ze zbiorem macierzy kwadratowych $M_n(\boldsymbol{K})$ stopnia $n = \dim L$ nad ciałem $\boldsymbol{K}$. Zadane w obydwu tych zbiorach struktury przestrzeni liniowych i pierścieni są z tym utożsamieniem zgodne. Bijekcjom, tj. automorfizmom

liniowym $f: L \to L$, odpowiadają przy tym macierze odwracalne: jeśli $f \circ f^{-1} = \mathrm{id}_L$, to $A_f A_{f^{-1}} = E_n$, a więc $A_{f^{-1}} = A_f^{-1}$.

Przypomnijmy tu, że macierz jest odwracalna (inaczej nieosobliwa) wtedy i tylko wtedy, gdy $\det A \neq 0$.

8. a) *Wyrażenie odwzorowania liniowego we współrzędnych.* Używając oznaczeń p. 4, możemy zapisać wektory z przestrzeni M i N za pomocą wektorów kolumnowych

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

a wówczas działanie operatora f można wyrazić za pomocą mnożenia macierzowego wzorem

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

albo też skrótowo $\vec{y} = A_f \vec{x}$ (por. p. 4). Czasami bywa też wygodnie zapisywać analogiczny wzór za pomocą baz $\{e_i\}$, $\{e'_k\}$. Przyjmuje on wówczas postać

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

Przy stosowaniu tego ostatniego wzoru formalizm mnożenia macierzowego wymaga, aby w wyrażeniu po prawej stronie wzoru wektory z M mnożyć przez skalary z prawej strony, a nie z lewej; jest to dla nas bez znaczenia, gdyż będziemy po prostu przyjmować, że $e'a = ae'$ dla dowolnych $e' \in M$, $a \in \boldsymbol{K}$. Korzystając z tego rodzaju zapisu, będziemy musieli nieraz sprawdzać łączność lub liniowość takich „mieszanych" iloczynów macierzy względem czynników, których część stanowią wektory z L, a część skalary z $\boldsymbol{K}$. Na przykład,

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

lub

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

itp. Formalizm p. p. 4, 5 automatycznie przenosi się na te przypadki. Ta sama uwaga odnosi się także do macierzy blokowych.

b) *Współrzędne wektora w zmienionej bazie.* Załóżmy, że w przestrzeni L zostały wybrane dwie bazy $\{e_i\}$, $\{e'_i\}$. Dowolny wektor $l \in L$ można wyrazić, używając współrzędnych w każdej z tych baz: $l = \sum_{i=1}^{n} x_i e_i = \sum_{k=1}^{n} x'_k e'_k$. Wykażemy istnienie

takiej niezależnej od wektora l macierzy kwadratowej A stopnia n, że $\vec{x} = A\vec{x}'$. Istotnie, jeśli $e'_k = \sum_{i=1}^{n} a_{ik}e_i$, to $A = (a_{ik})$. Macierz A będziemy nazywać *macierzą przejścia* (od bazy $\{e_i\}$ („nieprimowanych" wektorów) do bazy $\{e'_k\}$ („primowanych" wektorów) lub od współrzędnych „nieprimowanych" do „primowanych"). Zauważmy, że jest to macierz odwracalna, a odwrotną do niej jest macierz przejścia od bazy „primowanych" do „nieprimowanych" wektorów.

Odnotujmy także i to, że wzór $\vec{x} = A\vec{x}'$ można interpretować jako wzór wyrażający współrzędne starego wektora kolumnowego $\vec{x}$ poprzez współrzędne wektora $f(\vec{x})$, gdzie f oznacza odwzorowanie liniowe $L \to L$ o macierzy A w bazie $\{e_i\}$.

W fizyce te dwa punkty widzenia nazywane są odpowiednio „biernym" i „czynnym". W tym pierwszym opisujemy *jeden i ten sam stan układu* (wektor l) z punktu widzenia *różnych obserwatorów* (o własnych układach odniesienia), a w tym drugim *obserwator jest ten sam*, a stan układu zostaje poddany przekształceniom polegającym, na przykład, na operacji symetrii przestrzeni stanów tego układu.

c) *Macierz odwzorowania liniowego względem zmienionych baz.* Wracając do sytuacji opisanej w p. 4, wyjaśnijmy, jak zmienia się macierz A_f odwzorowania liniowego przy przejściu od baz $\{e_k\}$, $\{e'_i\}$ do nowych baz $\{\overline{e}_k\}$, $\{\overline{e}'_i\}$ przestrzeni N i M. Niech B oznacza macierz przejścia od współrzędnych względem $\{e_k\}$ do współrzędnych względem $\{\overline{e}_k\}$, a C — macierz przejścia od współrzędnych względem $\{e'_i\}$ do współrzędnych względem $\{\overline{e}'_i\}$. Twierdzimy, że macierz $\overline{A}_f$ odwzorowania f względem baz $\{\overline{e}_k\}$, $\{\overline{e}'_k\}$ wyraża się wzorem

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

Istotnie, prowadząc rachunki dla baz, otrzymujemy

$$(\overline{e}_k)\overline{A}_f = f((\overline{e}_k)) = f((e_k)B) = (f(e_k))B =$$

$$= (e'_i)A_fB = (\overline{e}'_i)C^{-1}A_fB.$$

Czytelnikowi polecamy przeprowadzenie powyższego rachunku przy użyciu współrzędnych. Przypadek, gdy $N = M$ oraz $\{e_i\} = \{e'_i\}$, $\{\overline{e}_i\} = \{\overline{e}'_i\}$ i $B = C$, jest specjalnie ważny. Macierz odwzorowania liniowego f względem nowej bazy jest wyrażona wzorem

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

Odwzorowanie $M_n(\boldsymbol{K}) \to M_n(\boldsymbol{K}), A \mapsto B^{-1}AB$ nazywamy *sprzężeniem* (przez macierz nieosobliwą B). Sprzężenie jest *automorfizmem algebry macierzowej* $M_n(\boldsymbol{K})$:

$$B^{-1}\Big(\sum_{i=1}^{n} a_iA_i\Big)B = \sum_{i=1}^{n} a_iB^{-1}A_iB, \quad a_i \in \boldsymbol{K};$$

$$B^{-1}(A_1 \dots A_m)B = (B^{-1}A_1B)\dots(B^{-1}A_mB)$$

(w iloczynie po prawej wewnętrzne czynniki B i B^{-1} występują obok siebie i dzięki temu skracają się). Szczególną rolę odgrywają te spośród określonych na $M_n(\boldsymbol{K})$ funkcji, które nie zmieniają się przy zamianie macierzy na sprzężoną z nią, gdyż z nich można budować *niezmienniki* (*inwarianty*) *operatorów liniowych*: jeśli ϕ jest taką funkcją, to definiując $\phi(f) = \phi(A_f)$ otrzymujemy funkcję zależną tylko od f, a nie od bazy, w której wyznaczyliśmy A_f. A oto dwa ważne przykłady.

9. Wyznacznik i ślad operatora liniowego. Określimy ślad (ang. *trace*) operatora f wzorem

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

(ślad macierzy A to suma elementów jej głównej przekątnej), oraz wyznacznik

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

Niezmienniczość wyznacznika względem sprzężenia jest oczywista:

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

Dla wykazania niezmienniczości śladu wykażemy ogólniejszy fakt: *jeśli A i B są takimi macierzami, że AB i BA są określone, to* $\operatorname{Tr} AB = \operatorname{Tr} BA$. Rzeczywiście,

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

Jeśli B jest nieosobliwa, to, stosując tylko co wykazany fakt do macierzy $B^{-1}A$ i B, otrzymamy

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

Tworząc funkcje symetryczne wartości własnych macierzy i operatorów (zdefiniujemy je niżej w § 8), otrzymamy więcej przykładów funkcji niezmienniczych.

Na zakończenie tego paragrafu zbierzemy definicje, nazwy i standardowe oznaczenia dla pewnych klas macierzy o wyrazach rzeczywistych lub zespolonych, o podstawowym znaczeniu dla teorii grup i algebr Liego oraz jej wielorakich zastosowań, w szczególności w fizyce. Pierwszą klasę stanowią tak zwane *grupy klasyczne*: są to istotnie grupy względem mnożenia macierzy. Drugą klasę stanowią *algebry Liego*: są to przestrzenie liniowe zamknięte względem operacji brania *komutatora*: $[A, B] = AB - BA$. Równoległość wprowadzonych oznaczeń zostanie w pewnym zakresie wyjaśniona w § 11 oraz w ćwiczeniu 8.

10. Grupy klasyczne

a) *Pełna grupa liniowa* $\boldsymbol{GL}(n, \boldsymbol{K})$. Składa się ona ze wszystkich nieosobliwych macierzy kwadratowych stopnia n nad ciałem $\boldsymbol{K}$.

b) *Specjalna grupa liniowa* $\boldsymbol{SL}(n, \boldsymbol{K})$. Składa się ona z macierzy kwadratowych stopnia n nad ciałem $\boldsymbol{K}$ z wyznacznikiem 1.

W powyższych przykładach K mogło być dowolnym ciałem. Poniżej ograniczymy się do przypadku $K = R$ albo C, choć istnieją uogólnienia podanych definicji na przypadek innych ciał.

c) *Grupa ortogonalna* $O(n, K)$. Składa się ona z macierzy stopnia n spełniających warunek $AA^t = E_n$. Takie macierze istotnie tworzą grupę, gdyż

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

oraz

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

Dla $K = R$, C grupa ta nazywa się odpowiednio rzeczywistą lub zespoloną (grupą ortogonalną). Jej elementy noszą nazwę *macierzy ortogonalnych*. Zamiast $O(n, R)$ piszemy zazwyczaj $O(n)$.

d) *Specjalna grupa ortogonalna* $SO(n, K)$. Składa się z macierzy ortogonalnych o wyznaczniku równym jedności:

$$SO(n, K) = O(n, K) \cap SL(n, K).$$

Zamiast $SO(n, R)$ piszemy zazwyczaj $SO(n)$.

e) *Grupa unitarna* $U(n)$. Składa się z zespolonych macierzy stopnia n spełniających warunek $A\overline{A}^t = E_n$, gdzie $\overline{A}$ — macierz o elementach będących sprzężeniami odpowiadających elementów macierzy A: jeśli $A = (a_{ik})$, to $\overline{A} = (\overline{a}_{ik})$. Używając równości $\overline{A}\,\overline{B} = \overline{AB}$, nietrudno sprawdzić tym sposobem co w c), że także $U(n)$ jest grupą. Elementy $U(n)$ nazywamy *macierzami unitarnymi*. Macierz $\overline{A}^t$ często jest nazywana *macierzą hermitowsko sprzężoną* z macierzą A; matematycy oznaczają ją A^*, a fizycy $A^\dagger$. Zauważmy jeszcze, że operacja sprzężenia hermitowskiego jest określona dla macierzy zespolonych o dowolnych wymiarach.

f) *Specjalna grupa unitarna* $SU(n)$. Składa się z macierzy unitarnych o wyznaczniku równym jedności:

$$SU(n) = U(n) \cap SL(n, C).$$

Wprost z definicji wynika, że rzeczywiste macierze unitarne są macierzami ortogonalnymi:

$$O(n) = U(n) \cap GL(n, R), \qquad SO(n) = U(n) \cap SL(n, R).$$

11. Klasyczne algebry Liego. (Macierzową) algebrą Liego nazywamy dowolną addytywną podgrupę macierzy kwadratowych $M_n(K)$, zamkniętą względem operacji brania komutatora macierzy $[A, B] = AB - BA$. (Ogólną definicję znajdzie czytelnik w ćwiczeniu 14.) Następujące zbiory macierzy stanowią klasyczne algebry Liego; zazwyczaj są one także przestrzeniami wektorowymi nad K (czasem tylko nad R, chociaż $K = C$). Jednakże nie są one grupami względem mnożenia!

a) *Algebra* $\mathfrak{gl}(n, K)$ składa się ze wszystkich macierzy $M_n(K)$.

b) *Algebra* $\mathfrak{sl}(n, \boldsymbol{K})$ składa się ze wszystkich macierzy $M_n(\boldsymbol{K})$ o śladzie zero (czasem nazywamy je macierzami „bezśladowymi"). Zamkniętość względem brania komutatora wynika ze wzoru $\mathrm{Tr}[A, B] = 0$ wykazanego w p. 9. Biorąc pod uwagę, że Tr jest funkcją liniową na przestrzeni macierzy kwadratowych czy operatorów liniowych, wnioskujemy, że $\mathfrak{sl}(n, \boldsymbol{K})$ jest przestrzenią liniową nad $\boldsymbol{K}$.

c) *Algebra* $\mathfrak{o}(n, \boldsymbol{K})$ składa się ze wszystkich macierzy $M_n(\boldsymbol{K})$, spełniających warunek $A + A^t = 0$. Równoważnie: $A = (a_{ik})$, gdzie $a_{ii} = 0$ (jeśli charakterystyka $\boldsymbol{K}$ nie równa się 2), $a_{ik} = -a_{ki}$. Takie macierze będziemy nazywać antysymetrycznymi lub skośnie symetrycznymi. Odnotujmy jeszcze to, że $\mathrm{Tr}\, A = 0$ dla wszystkich $A \in \mathfrak{o}(n, \boldsymbol{K})$.

Jeśli $A^t = -A$, $B^t = -B$, to $[A, B]^t = [B^t, A^t] = [-B, -A] = -[A, B]$, więc macierz $[A, B]$ jest antysymetryczna. Te macierze tworzą przestrzeń liniową nad $\boldsymbol{K}$.

W tym kontekście odnotujmy jeszcze, że zbiór macierzy symetrycznych (A nazywamy symetryczną, gdy $A = A^t$) nie jest zamknięty względem operacji komutatora, ale jest zamknięty względem antykomutatora $AB + BA$ lub operacji Jordana $\frac{1}{2}(AB + BA)$.

d) *Algebra* $\mathfrak{u}(n)$ składa się z zespolonych macierzy kwadratowych stopnia n spełniających warunek $A + \overline{A}^t = 0$, lub równoważnie $a_{ik} = -\overline{a}_{ki}$. Takie macierze będziemy nazywać *hermitowsko antysymetrycznymi* lub *antyhermitowskimi* lub jeszcze *skośnie hermitowskimi*. Tworzą one przestrzeń liniową nad $\boldsymbol{R}$, ale nie nad $\boldsymbol{C}$. Jeśli $A^t = -\overline{A}$, $B^t = -\overline{B}$, to

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

a zatem $\mathfrak{u}(n)$ jest algebrą Liego.

Odnotujmy jeszcze, że macierz A nazywa się *hermitowsko symetryczną* lub po prostu *hermitowską*, jeśli $\overline{A}^t = A$, tj. $a_{ik} = \overline{a}_{ki}$. W szczególności, taka macierz ma na diagonali liczby czysto urojone. Jest jasne, że rzeczywiste macierze hermitowskie są symetryczne, a antyhermitowskie — antysymetryczne, skąd wynika

$$\mathfrak{o}(n) = \mathfrak{u}(n) \cap \mathfrak{sl}(n, \boldsymbol{R}).$$

Macierz A jest hermitowska wtedy i tylko wtedy, gdy iA jest antyhermitowska.

e) *Algebra* $\mathfrak{su}(n)$. Tak nazywamy algebrę bezśladowych i antyhermitowskich macierzy stopnia n, czyli $\mathfrak{u}(n) \cap \mathfrak{sl}(n, \boldsymbol{C})$. Jest to przestrzeń liniowa nad $\boldsymbol{R}$.

Badając w drugiej części książki przestrzenie liniowe wyposażone w metryki typu euklidesowego lub hermitowskiego, wyjaśnimy geometryczny sens operatorów reprezentowanych macierzami z opisanych tu klas, a przy okazji uzupełnimy powyższą listę.

Ćwiczenia

1. Sformułować precyzyjnie i udowodnić twierdzenie mówiące o tym, że macierze określone nad ciałem i rozbite na bloki można mnożyć blokami, jeśli wymiary i ilości

bloków są zgodne:

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

gdy liczba kolumn w bloku A_{ij} równa się liczbie wierszy w bloku B_{jk} i liczba wierszy blokowych w macierzy A jest równa liczbie kolumn blokowych macierzy B.

2. Wprowadzić pojęcie macierzy nieskończonej (z nieskończoną ilością wierszy i (lub) kolumn). Podać warunki, aby można było mnożyć takie macierze (przykłady: macierze skończone, tj. takie, dla których jedynie skończona liczba elementów macierzy nie jest zerem; macierze, których każdy wiersz i (lub) kolumna zawiera skończoną liczbę niezerowych elementów). Znaleźć warunki istnienia potrójnych iloczynów.

3. Wykazać, że równanie $XY - YX = E$ nie ma rozwiązań X, Y w zbiorze skończonych macierzy kwadratowych nad ciałem zerowej charakterystyki. (*Wskazówka.* Rozważyć ślady obydwóch stron.) Znaleźć rozwiązanie tego równania w zbiorze macierzy nieskończonych. (*Wskazówka.* Rozważyć operatory liniowe $\frac{d}{dx}$ i operator mnożenia przez x w przestrzeni wielomianów wszystkich stopni zmiennej x, posługując się tożsamością $\frac{d}{dx}(xf) - x\frac{d}{dx}f = f$.)

4. Podać jawny opis klasycznych grup i algebr Liego w przypadkach $n = 1$ i $n = 2$. Skonstruować izomorfizm grupy $\boldsymbol{U}(1)$ z $\boldsymbol{SO}(2, \boldsymbol{R})$.

5. Następujące macierze nad $\boldsymbol{C}$ są znane jako macierze Pauliego:

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

(Wprowadził je znany fizyk W. Pauli, jeden z twórców mechaniki kwantowej, dla potrzeb zaproponowanej przez siebie teorii spinu elektronowego.) Sprawdzić następujące ich własności:

a) $[\sigma_a, \sigma_b] = 2i\varepsilon_{abc}\sigma_c$, gdzie $\{a, b, c\} = \{1, 2, 3\}$, a przez ε_{abc} oznaczyliśmy znak permutacji $\begin{pmatrix} 1 & 2 & 3 \\ a & b & c \end{pmatrix}$;

b) $\sigma_a\sigma_b + \sigma_b\sigma_a = 2\delta_{ab}\sigma_0$ (δ_{ab} — symbol Kroneckera).

c) Macierze $i\sigma_1, i\sigma_2, i\sigma_3$ stanowią bazę nad $\boldsymbol{R}$ dla $\mathfrak{su}(2)$, a nad $\boldsymbol{C}$ — dla $\mathfrak{sl}(2)$; macierze $\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$ są bazą nad $\boldsymbol{R}$ dla $\mathfrak{u}(2)$, a nad $\boldsymbol{C}$ — bazą $\mathfrak{gl}(2)$.

6. Poniższe macierze stopnia 4 nad $\boldsymbol{C}$ znane są pod nazwą macierzy Diraca (σ_a są macierzami Pauliego):

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

(Wprowadził je znany fizyk P. A. M. Dirac, jeden z twórców mechaniki kwantowej, w swej teorii relatywistycznego elektronu ze spinem). Wykorzystując rezultaty ćwiczeń 5 i 1 sprawdzić następujące ich własności:

a) $\gamma_a\gamma_b + \gamma_b\gamma_a = 2g_{ab}E_4$, gdzie $g_{ab} = 0$ dla $a \neq b$, $g_{00} = 1$, $g_{11} = g_{22} = g_{33} = -1$;

b) Przyjmijmy $\gamma_5 = i\gamma_1\gamma_2\gamma_3\gamma_0$. Sprawdzić, że $\gamma_5 = -\begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}$;

c) $\gamma_a\gamma_5 = -\gamma_5\gamma_a$ dla $a = 0,1,2,3$; $(\gamma_5)^2 = E_4$.

7. Sprawdzić prawdziwość podanych w poniższej tabeli danych dotyczących wymiarów klasycznych algebr Liego (jako przestrzeni liniowych nad odpowiednimi ciałami).

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

8. Niech A oznacza macierz kwadratową stopnia n, ε — parametr rzeczywisty, $\varepsilon \to 0$. Wykazać, że macierz $U = E+\varepsilon A$ jest „unitarna z dokładnością do ε^2" wtedy i tylko wtedy, gdy A jest antyhermitowska:

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

Sformułować i wykazać analogiczne zależności dla pozostałych par klasycznych grup i algebr Liego.

9. Niech $U = E + \varepsilon A$, $V = E + \varepsilon B$, gdzie $\varepsilon \to 0$. Sprawdzić, że

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

(wyrażenie stojące po lewej stronie nazywa się komutatorem grupowym U i V).

10. Rząd $\mathrm{r}(A)$ macierzy A nad ciałem — to maksymalna liczba liniowo niezależnych kolumn tej macierzy. Wykazać, że $\mathrm{r}(A) = \dim \mathrm{Im}\, f$.

Wykazać, że macierz kwadratową rzędu 1 można przedstawić jako iloczyn kolumny i wiersza.

11. Niech A, B będą macierzami o wymiarach $m \times n$, $m_1 \times n_1$ nad dowolnym ciałem i niech będzie ustalony sposób indeksowania wszystkich mn elementów macierzy A i wszystkich m_1n_1 elementów macierzy B (np. kolejno wzdłuż wierszy). *Iloczynem tensorowym* albo *iloczynem Kroneckera* tych macierzy nazywamy macierz $A \otimes B$ o wymiarach $mn \times m_1n_1$ z elementami $a_{ik}b_{lm}$ umieszczonymi na miejscach $\alpha\beta$, gdzie α — indeks miejsca elementu a_{ik}, β — indeks b_{lm}. Sprawdzić następujące fakty:

a) $A \otimes B$ jest liniowy względem każdego ze swych argumentów przy ustalonym drugim;

b) Jeśli $m = n$, $m_1 = n_1$, to $\det(A \otimes B) = (\det A)^{m_1}(\det B)^m$.

12. Ile operacji potrzeba dla przemnożenia dwóch dużych macierzy? Poniższa seria zadań poświęcona jest przedstawieniu metody Strassena, pozwalającej znacznie zmniejszyć tę liczbę, jeśli macierze są rzeczywiście dużych rozmiarów.

a) Przemnożenie dwóch macierzy stopnia N zwykłym sposobem wymaga wykonania N^3 mnożeń i $N^2(N-1)$ dodawań.

b) Podana niżej formuła mnożenia dla $N = 2$ pokazuje, że można wykonać mniej mnożeń (7 zamiast 8) kosztem wykonania 18 zamiast 4 dodawań (nie jest wymagana przemienność mnożenia):

$$\begin{pmatrix} a & b \\ -c & -d \end{pmatrix}\begin{pmatrix} A & B \\ -C & -D \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

gdzie oznaczyliśmy

$$S_{11} = (a+d)(A+D) - (b+d)(C+D) - d(A-C) - (a-b)D,$$
$$S_{12} = (a-b)D - a(D-B), \quad S_{21} = (d-c)A - d(A-C),$$
$$S_{22} = (a+d)(A+D) - (a+c)(A+B) - a(D-B) - (d-c)A.$$

c) Stosując powyższą metodę do macierzy stopnia 2^n rozbitych na cztery bloki wymiarów $2^{n-1} \times 2^{n-1}$, wykazać, że obliczenie ich iloczynu wymaga wykonania 7^n mnożeń i $6(7^n - 4^n)$ dodawań.

d) Rozszerzając macierze stopnia N do najbliższego stopnia 2^n poprzez dopisanie zer, wykazać, że dla obliczenia ich iloczynu wystarcza wykonanie $O(N^{\log_2 7}) = O(N^{2,81})$ operacji.

A może uda Ci się, czytelniku, wymyślić coś lepszego?

13. Niech $L = M_n(\boldsymbol{K})$ będzie przestrzenią macierzy kwadratowych stopnia n. Dowieść, że dla dowolnego funkcjonału $f \in L^*$ istnieje jedyna macierz $A \in M_n(\boldsymbol{K})$, taka że

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

dla wszystkich $X \in M_n(\boldsymbol{K})$. Wyprowadzić stąd istnienie kanonicznego izomorfizmu

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

dla dowolnej skończenie wymiarowej przestrzeni L.

14. ***K*-algebrą *Liego*** nazywamy przestrzeń liniową L nad $\boldsymbol{K}$ wraz z operacją dwuargumentową (komutatorem): $L \times L \to L$ oznaczaną $[\ ,\ ]$ i spełniającą warunki:

a) komutator $[l, m]$ jest liniowy względem każdego z argumentów $l, m \in L$ przy ustalonym drugim argumencie;

b) $[l, m] = -[m, l]$ dla wszystkich l, m;

c) $[l_1, [l_2, l_3] + [l_3, [l_1, l_2] + [l_2, [l_3, l_1] = 0$ dla wszystkich l_1, l_2, $l_3 \in L$.

(Jest to tzw. tożsamość Jacobiego)

Sprawdzić, że opisane w punkcie 11 klasyczne algebry Liego są algebrami Liego w podanym tu znaczeniu. Ogólniej, sprawdzić, że komutator $[X, Y] = XY - YX$ w dowolnym pierścieniu łącznym spełnia tożsamość Jacobiego.

§ 5. Podprzestrzenie i sumy proste

1. W tym paragrafie zbadamy pewne geometryczne własności wzajemnego położenia podprzestrzeni skończenie wymiarowej przestrzeni liniowej L. Zilustrujmy w czym rzecz na najprostszym przykładzie. Niech $L_1, L_1' \subset L$ będą dwiema podprzestrzeniami.Jeśli istnieje taki automorfizm liniowy $f\colon L \to L$, który przeprowadza L_1 na L_1', to wydaje się naturalne przyjąć, że są one jednakowo położone w przestrzeni L. Potrzeba na to, aby $\dim L_1 = \dim L_1'$, gdyż f zachowuje wszystkie relacje liniowe,

a więc przeprowadza bazę L_1 na bazę L_1'. Ten warunek jest także wystarczający. Istotnie, wybierzmy bazy $\{e_1, \ldots, e_m\}$ w L_1 i $\{e_1', \ldots, e_m'\}$ w L_1'. Zgodnie z twierdzeniem § 2 p. 12 możemy je uzupełnić do baz przestrzeni L — $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ i $\{e_1', \ldots, e_m', e_{m+1}', \ldots, e_n'\}$. Na mocy stwierdzenia § 3 p. 3 istnieje takie odwzorowanie liniowe $f: L \to L$, że dla każdego i obrazem e_i jest e_i'. Jest ono odwracalne i przeprowadza L_1 na L_1'.

W ten sposób wnioskujemy, że *wszystkie podprzestrzenie liniowe jednego wymiaru są jednakowo położone w L.*

Innym naturalnym zagadnieniem jest rozważenie wzajemnego położenia (uporządkowanych) par podprzestrzeni $L_1, L_2 \subset L$. Tak jak powyżej będziemy uważali, że pary (L_1, L_2) i (L_1', L_2') są jednakowo położone, jeśli istnieje taki liniowy automorfizm $f: L \to L$, że $f(L_1) = L_1'$ i $f(L_2) = L_2'$ Jak poprzednio, równości $\dim L_1 = \dim L_1'$ i $\dim L_2 = \dim L_2'$ są konieczne, ale tym razem już na ogół niewystarczające. Rzeczywiście, jeśli (L_1, L_2) i (L_1, L_2) są jednakowo położone, to f przeprowadza podprzestrzeń $L_1 \cap L_2$ na $L_1' \cap L_2'$ i dlatego konieczny jest również warunek $\dim(L_1 \cap L_2) = \dim(L_1' \cap L_2')$. Jeśli na L_1 i L_2 nie nakładamy innych warunków poza ustaleniem $\dim L_1$ i $\dim L_2$, to $\dim(L_1 \cap L_2)$ może przyjmować, ogólnie mówiąc, różne wartości. Dla wyjaśnienia, jakie wartości są tu możliwe, wprowadzimy pojęcie sumy podprzestrzeni liniowych.

2. Definicja. Niech $L_1, \ldots, L_n \subset L$ będą podprzestrzeniami liniowymi L. Ich *sumą* nazywamy zbiór

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

Łatwo przekonać się, że suma podprzestrzeni jest także podprzestrzenią i że tak określona operacja sumowania jest łączna i przemienna, podobnie do operacji przekroju podprzestrzeni liniowych. Równoważnie można opisać sumę $L_1 + \cdots + L_n$ jako *najmniejszą podprzestrzeń w L zawierającą każdą z podprzestrzeni L_i*. Następujące twierdzenie podaje ważny związek wymiaru sumy dwóch podprzestrzeni z wymiarem ich przecięcia.

3. Twierdzenie. *Jeśli $L_1, L_2 \subset L$ są skończenie wymiarowe, to $L_1 \cap L_2$ i $L_1 + L_2$ są także skończenie wymiarowe i*

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

Dowód. $L_1 + L_2$ jest powłoką liniową sumy baz L_1 i L_2, więc jest skończenie wymiarowa; podobnie $L_1 \cap L_2$, gdyż zawiera się w skończenie wymiarowych przestrzeniach L_1 i L_2. Oznaczmy $m = \dim L_1 \cap L_2$, $n = \dim L_1$, $p = \dim L_2$ i obierzmy bazę $\{e_1, \ldots, e_m\}$ przestrzeni $L_1 \cap L_2$. Korzystając z twierdzenia § 2 p. 12, uzupełnimy ją do baz przestrzeni L_1 i L_2, które oznaczymy $\{e_1, \ldots, e_m, e_{m+1}', \ldots, e_n'\}$ i $\{e_1, \ldots, e_m, e_{m+1}'', \ldots, e_p''\}$. Taką parę baz w L_1 i L_2 będziemy nazywać *zgodną*.

Wykażemy teraz, że rodzina $\{e_1, \ldots, e_m, e'_{m+1}, \ldots, e'_n, e''_{m+1} \ldots, e''_p\}$ tworzy bazę przestrzeni $L_1 + L_2$, a stąd już wynika teza twierdzenia, wobec równości

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

Ponieważ każdy wektor z $L_1 + L_2$ jest sumą wektorów z L_1 i L_2, a więc jest sumą kombinacji liniowych $\{e_1, \ldots, e_m, e'_{m+1}, \ldots e'_n\}$ i $\{e_1, \ldots, e_m, e''_{m+1}, \ldots, e''_p\}$, suma tych rodzin generuje $L_1 + L_2$. Pozostaje więc tylko sprawdzić liniową niezależność. Załóżmy, że istnieje nietrywialna zależność liniowa

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

Wówczas z pewnością istnieją takie wskaźniki j oraz k, dla których $y_j \neq 0$ i $z_k \neq 0$, w przeciwnym bowiem razie otrzymalibyśmy nietrywialną zależność liniową dla elementów bazy L_1 lub L_2.

Wobec tego niezerowy wektor $\sum_{k=m+1}^{p} z_k e''_k \in L_2$ także należy do L_1, gdyż można go zapisać w postaci $-\left(\sum_{i=1}^{m} x_i e_i + \sum_{j=m+1}^{n} y_j e'_j\right)$. Należy on więc do $L_1 \cap L_2$ i jest kombinacją liniową wektorów $\{e_1, \ldots, e_m\}$ stanowiących bazę $L_1 \cap L_2$. Nasze przypuszczenie doprowadziło zatem do wniosku o nietrywialnej zależności liniowej między wektorami $\{e_1, \ldots, e_m, e''_{m+1}, \ldots, e''_p\}$, co byłoby sprzeczne z ich określeniem. W ten sposób twierdzenie zostało wykazane.

4. Wniosek. *Niech $n_1 \leq n_2 \leq n$ będą wymiarami przestrzeni L_1, L_2 i L odpowiednio. Wówczas liczby $i = \dim L_1 \cap L_2$ oraz $s = \dim(L_1 + L_2)$ mogą przyjmować dowolne wartości w zakresie ograniczonym warunkami $0 \leq i \leq n_1$, $n_2 \leq s \leq n$ oraz $i + s = n_1 + n_2$.*

Dowód. Konieczność tych warunków wynika z inkluzji $L_1 \cap L_2 \subset L_1$, $L_2 \subset L_1 + L_2 \subset L$ i z twierdzenia p. 3. Dla wykazania dostateczności wybierzemy $s = n_1 + n_2 - i$ niezależnych liniowo wektorów z L: $\{e_1, \ldots, e_i; e'_{i+1}, \ldots, e'_{n_1}; e''_{i+1}, \ldots, e''_{n_2}\}$ i oznaczymy przez L_1, L_2 powłoki liniowe $\{e_1 \ldots, e_i; e'_{i+1}, \ldots, e'_{n_1}\}$ i $\{e_1, \ldots, e_i; e''_{i+1}, \ldots, e''_{n_1}\}$ odpowiednio. Tak jak przy dowodzie twierdzenia nietrudno sprawdzić, że $L_1 \cap L_2$ jest powłoką liniową $\{e_1, \ldots, e_m\}$.

5. Teraz możemy wykazać, że niezmienniki $n_1 = \dim L_1$, $n_2 = \dim L_2$ oraz $i = \dim(L_1 \cap L_2)$ całkowicie charakteryzują wzajemne położenie pary podprzestrzeni (L_1, L_2) w L. W istocie, weźmy inną parę (L'_1, L'_2) z tymi samymi niezmiennikami i skonstruujmy zgodne pary baz dla L_1, L_2 i L'_1, L_2, a z nich, tak jak w dowodzie twierdzenia p.3, bazy $L_1 + L_2$ i $L'_1 + L'_2$, a w końcu te ostatnie uzupełnijmy do baz w L. Istnienie automorfizmu liniowego przeprowadzającego jedną z tak otrzymanych baz na drugą pokazuje, że te pary są jednakowo położone w L.

6. Położenie ogólne. W kontekście rozważań poprzedzającego punktu powiemy, że podprzestrzenie L_1, $L_2 \subset L$ znajdują się w *położeniu ogólnym*, jeśli ich

przecięcie ma najmniejszy z możliwych, a suma największy z możliwych wymiarów dopuszczanych nierównościami wniosku p. 4.

Na przykład, dwie płaszczyzny w trójwymiarowej przestrzeni znajdują się w położeniu ogólnym, jeśli ich przecięcie jest prostą, a w przestrzeni czterowymiarowej — jeśli jest punktem. Używając innego określenia dla wyrażenia tego samego, mówimy, że L_1 i L_2 *są transwersalne.*

Określenie „w ogólnym położeniu" pochodzi stąd, że w pewnym sensie większość par podprzestrzeni (L_1, L_2) znajduje się w takim właśnie położeniu, a inne położenie jest wyjątkowe (zdegenerowane). Stwierdzenie to można uściślić na wiele sposobów. Jeden z nich polega na tym, aby wprowadzić opis zbioru par podprzestrzeni przy użyciu pewnych parametrów i sprawdzić, że para nie znajduje się w położeniu ogólnym tylko wtedy, gdy te parametry spełniają dodatkowe warunki, nie spełnione przez ogólne wartości parametrów.

Inny sposób, przydatny dla przypadku $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$, wykorzystuje następujące rozumowanie: należy wybrać w L jakąś bazę, określić L_1 i L_2 dwoma układami równań liniowych i pokazać, że zmieniając dowolnie mało współczynniki tych równań („zaburzając" L_1 i L_2), można przeprowadzić wybraną parę w położenie ogólne.

Można byłoby także spróbować rozważać niezmienniki charakteryzujące położenie trójek, czwórek i większej liczby podprzestrzeni w L. Tutaj kombinatoryczne trudności szybko rosną i dla rozwiązania tego zagadnienie potrzeba innej techniki, a poza tym, poczynając od czwórek podprzestrzeni, ich wzajemne położenie nie daje się już scharakteryzować jedynie dyskretnymi niezmiennikami w rodzaju wymiarów różnych sum i przecięć.

Zwróćmy jeszcze uwagę na to, że jak podpowiada nasza „fizyczna" intuicja, położenie, powiedzmy prostej względem płaszczyzny, można scharakteryzować za pomocą kąta między nimi. Jednak podkreśliliśmy to już w § 1, że pojęcie kąta wykorzystuje dodatkową strukturę. W obrębie czysto liniowej teorii występuje tylko różnica między „zerowym" a „niezerowym" kątem.

Teraz zbadamy jeden szczególny, ale bardzo ważny przypadek wzajemnego położenia n-tek podprzestrzeni.

7. Definicja. Przestrzeń L nazywamy *sumą prostą podprzestrzeni* $L_1, \ldots, L_n$, jeśli każdy wektor $l \in L$ można jednoznacznie wyrazić w postaci $\sum_{i=1}^{n} l_i$, gdzie $l_i \in L_i$.

Jeśli warunek definicji jest spełniony, to piszemy $L = L_1 \oplus \cdots \oplus L_n$ lub $L = \bigoplus_{i=1}^{n} L_i$. W szczególności, jeśli $\{e_1, \ldots, e_n\}$ jest bazą L, a $L_i = \boldsymbol{K}e_i$ jest powłoką liniową wektora e_i, to $L = \bigoplus_{i=1}^{n} L_i$. Oczywiście, jeśli $L = \bigoplus_{i=1}^{n} L_i$, to także $L = \sum_{i=1}^{n} L_i$, ale ten drugi warunek jest słabszy.

8. Twierdzenie. *Jeśli* $L_1, \ldots, L_n \subset L$ *są podprzestrzeniami* L, *to* $L = \bigoplus_{i=1}^{n} L_i$ *wtedy i tylko wtedy, gdy spełniony jest którykolwiek z następujących (równoważnych) warunków:*

a) $\sum_{i=1}^{n} L_i = L$ *oraz* $L_j \cap \left(\sum_{i \neq j} L_i\right) = \{0\}$ *dla każdego* $1 \leq j \leq n$;

b) $\sum_{i=1}^{n} L_i = L$ *oraz* $\sum_{i=1}^{n} \dim L_i = \dim L$ (*tu zakładamy, że* L *jest skończenie wymiarowa*).

Dowód. a) Jednoznaczność zapisu dowolnego wektora $l \in L$ w postaci $\sum_{i=1}^{n} l_i$, $l_i \in L_i$ jest równoważna z jednoznacznością takiego zapisu dla wektora zerowego. Istotnie, jeśli $\sum_{i=1}^{n} l_i = \sum_{i=1}^{n} l'_i$, to także $0 = \sum_{i=1}^{n} (l_i - l'_i)$, i na odwrót. Przyjmując teraz, że istnieje nietrywialne przedstawienie $0 = \sum_{i=1}^{n} l_i$, w którym powiedzmy $l_j \neq 0$, otrzymujemy $l_j = -\sum_{i \neq j} l_i \in L_j \cap \left(\sum_{i \neq j} L_i\right)$, więc warunek a) nie jest spełniony. Odwracając to rozumowanie, widzimy, że z niespełnienia warunku a) wynika niejednoznaczność przedstawienia wektora zerowego.

b) Jeśli $\bigoplus_{i=1}^{n} L_i = L$, to z pewnością

$$\sum_{i=1}^{n} L_i = L \quad \text{oraz} \quad \sum_{i=1}^{n} \dim L_i \geq \dim L,$$

gdyż suma baz L_i generuje L, a więc zawiera bazę L. Z twierdzenia p. 3, zastosowanego do L_j i $\sum_{i \neq j} L_i$, otrzymujemy

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

Z już wykazanej części twierdzenia wynika, że przecięcie po lewej stronie jest zerowymiarowe. Ponadto, jeśli suma wszystkich przestrzeni L_i jest prosta, to także i suma wszystkich przestrzeni poza L_j jest prosta, a więc dzięki indukcji możemy przyjąć, że

$$\dim\left(\sum_{i \neq j} L_i\right) = \sum_{i \neq j} \dim L_i,$$

skąd wynika

$$\sum_{i} \dim L_i = \dim L.$$

Odwrotnie, jeśli $\sum_{i} \dim L_i = \dim L$, to suma baz wszystkich L_i składa się z $\dim L$ elementów, a że generuje L na mocy pierwszej części warunku, więc jest bazą L. Zatem z istnienia nietrywialnego przedstawienia zera $0 = \sum_{i=1}^{n} l_i$, $l_i \in L_i$, wynikałoby istnienie znikającej kombinacji liniowej elementów tej bazy z nietrywialnymi współczynnikami, co jest niemożliwe.

Rozpatrzymy teraz związek między rozkładami na sumą prostą a szczególnymi operatorami liniowymi — projektorami.

9. Definicja. Operator liniowy $p: L \to L$ nazywamy *projektorem* (*operatorem rzutowania*), jeśli $p^2 = p \circ p = p$.

Z rozkładem na sumę prostą $L = \bigoplus_{i=1}^{n} L_i$ związany jest w naturalny sposób układ n projektorów określonych następująco: dla dowolnych $l_i \in L_i$

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

Ponieważ dowolny element $l \in L$ można zapisać jednoznacznie w postaci sumy $\sum_{j=1}^{n} l_j$, $l_j \in L_j$, odwzorowania p_i są określone poprawnie. Ich liniowość i własność idempotentności $p^2 = p$ wynikają wprost z określenia. Oczywiście $L_i = \operatorname{Im} p_i$.

Co więcej, jeśli $i \neq j$, to $p_i p_j = 0$, gdyż wektor l_i można zapisać w postaci $l_i = \sum_{j=1}^{n} l'_j$, gdzie $l'_j = 0$ dla $i \neq j$, $l'_i = l_i$.

Wreszcie, $\sum_{i=1}^{n} p_i = \mathrm{id}$, ponieważ $\Big(\sum_{i=1}^{n} p_i\Big)\Big(\sum_{j=1}^{n} l_j\Big) = \sum_{j=1}^{n} l_j$, jeśli $l_j \in L_j$.

Także odwrotnie, taki układ projektorów wyznacza rozkład przestrzeni na sumę prostą.

10. Twierdzenie. *Niech $p_1, \ldots, p_n: L \to L$ będzie skończonym zbiorem projektorów spełniających warunki:*

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

Wtedy, oznaczając $L_i = \operatorname{Im} p_i$, mamy $L = \bigoplus_{i=1}^{n} L_i$.

Dowód. Stosując operator $\mathrm{id} = \sum_{i=1}^{n} p_i$ do dowolnego wektora $l \in L$, otrzymamy $l = \sum_{i=1}^{n} p_i(l)$, gdzie $p_i(l) \in L_i$. Dlatego $L = \sum_{i=1}^{n} L_i$. Dla wykazania, że ta suma jest prosta, zastosujemy kryterium a) twierdzenia z p. 8. Niech $l \in L_j \cap \Big(\sum_{i \neq j} L_i\Big)$. Z określenia podprzestrzeni L_i jako $\operatorname{Im} p_i$ wynika istnienie takich wektorów $l_1, \ldots, l_n$, że

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

Zastosujmy do tej równości operator p_j. Korzystając z tego, że $p_j^2 = p_j$, $p_i p_j = 0$ dla $j \neq i$, otrzymujemy

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

Zatem $l = 0$ i dowód zakończony.

11. Przestrzenie dopełniające. Jeśli L jest przestrzenią skończonego wymiaru, to dla dowolnej podprzestrzeni $L_1 \subset L$ można dobrać taką podprzestrzeń $L_2 \subset L$, że $L = L_1 \oplus L_2$; poza trywialnymi przypadkami $L_1 = \{0\}$ lub $L_1 = L$ ten wybór nie jest jednoznaczny. W istocie, wybierając bazę $\{e_1, \ldots, e_m\}$ w L_1 i dowolnie uzupełniając ją do bazy $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ w L, możemy przyjąć jako L_2 powłokę liniową zbioru $\{e_{m+1}, \ldots, e_n\}$.

12. Zewnętrzne sumy proste. Do tej pory zajmowaliśmy się przypadkiem danej rodziny podprzestrzeni $L_1, \ldots, L_n$ zawartych w jednej i tej samej przestrzeni L. Załóżmy teraz, że $L_1, \ldots, L_n$ są przestrzeniami liniowymi, nie zakładając tego, że są wszystkie zawarte we wspólnej przestrzeni. Określimy teraz ich *zewnętrzną sumę prostą* L w następujący sposób:

a) L jest zbiorem $L_1 \times \ldots \times L_n$, tzn. elementami L są ciągi $(l_1, \ldots, l_n)$, gdzie $l_i \in L_i$.

b) Dodawanie i mnożenie przez skalary wykonywane jest na składowych:

$$(l_1, \ldots, l_n) + (l'_1, \ldots, l'_n) = (l_1 + l'_1, \ldots, l_n + l'_n),$$
$$a(l_1, \ldots, l_n) = (al_1, \ldots, al_n).$$

Nietrudno sprawdzić, że L spełnia aksjomaty przestrzeni liniowej. Odwzorowanie $f_i: L_i \to L$, $f_i(l) = (0, \ldots, 0, l, 0, \ldots, 0)$ (l na i-tym miejscu) jest liniowym włożeniem L_i w L, więc z konstrukcji natychmiast wynika, że $L = \bigoplus_{i=1}^{n} f_i(L_i)$. Utożsamiając L_i z $f_i(L_i)$, możemy przyjąć, że L jest przestrzenią, która zawiera każdą z przestrzeni L_i i jest ich sumą prostą. To wyjaśnia pochodzenie nazwy zewnętrzna suma prosta. Często jest wygodnie oznaczać również zewnętrzną sumę prostą symbolem $\bigoplus_{i=1}^{n} L_i$.

13. Sumy proste odwzorowań liniowych. Niech $L = \bigoplus_{i=1}^{n} L_i$, $M = \bigoplus_{i=1}^{n} M_i$ i niech $f: L \to M$ będzie takim odwzorowaniem liniowym, dla którego $f(L_i) \subset M_i$. Oznaczymy przez f_i indukowane odwzorowanie $L_i \to M_i$. W takiej sytuacji będziemy pisać $f = \bigoplus_{i=1}^{n} f_i$. Podobnie można określić zewnętrzną sumę prostą odwzorowań liniowych. Wybierając w L i M bazy tak, aby były one sumami baz podprzestrzeni L_i i M_i odpowiednio, zauważamy, że macierz f składa się ze stojących na diagonali bloków — macierzy odwzorowań f_i, a pozostałe jej miejsca wypełnione są zerami.

14. Orientacja rzeczywistych przestrzeni liniowych. Niech L będzie skończenie wymiarową przestrzenią liniową nad ciałem liczb rzeczywistych. Dowolne dwie uporządkowane bazy $\{e_i\}$ i $\{e'_i\}$ przestrzeni L są w niej jednakowo położone w tym sensie, że istnieje jedyny izomorfizm liniowy $f: L \to L$ przeprowadzający e_i w e'_i. Rozważmy jednakże bardziej subtelne zagadnienie: kiedy można przeprowadzić bazę

$\{e_i\}$ na bazę $\{e'_i\}$ za pomocą *ciągłego przemieszczenia*, inaczej *deformacji*? Pod tym określeniem rozumiemy taką rodzinę izomorfizmów liniowych $f_t\colon L \to L$, zależną w sposób ciągły od parametru $t \in [0,1]$, że $f_0 = \mathrm{id}$, $f_1(e_i) = e'_i$ dla każdego i, a ciągła zależność rodziny od t oznacza po prostu ciągłą zależność elementów macierzy f_t w jakiejkolwiek ustalonej bazie. Oczywisty warunek konieczny wynika z następującej obserwacji; ponieważ przy zmianie t wyznacznik f_t zmienia się w sposób ciągły i nie przechodzi przez zero, znak $\det f_t$ powinien być równy znakowi $\det f_0 = 1$, tj. $\det f_t > 0$.

Zachodzi również stwierdzenie odwrotne: *jeśli macierz przejścia z bazy $\{e_i\}$ do bazy $\{e'_i\}$ ma dodatni wyznacznik, to za pomocą ciągłego przemieszczenia można przeprowadzić $\{e_i\}$ na $\{e'_i\}$*.

Ten fakt można oczywiście sformułować i w taki sposób: *dowolną macierz rzeczywistą o wyznaczniku dodatnim można połączyć z macierzą jednostkową krzywą ciągłą składającą się z macierzy nieosobliwych (zbiór nieosobliwych macierzy rzeczywistych o dodatnim wyznaczniku jest spójny)*. Aby przejść od sformułowania używającego baz do sformułowania używającego macierzy, wystarczy zamiast pary baz $\{e_i\}$, $\{f(e_i)\}$ rozważać macierz przejścia z pierwszej do drugiej bazy.

Dowód tego twierdzenia przedstawimy w kilku pośrednich etapach.

a) *Niech $A = B_1 \cdots B_n$, przy czym A, B_i są macierzami o dodatnich wyznacznikach. Jeśli każdą z macierzy B_i można połączyć z E krzywą ciągłą, to można to też zrobić z macierzą A.*

Rzeczywiście, niech $B_j(t)$ oznaczają takie krzywe ciągłe w przestrzeni macierzy nieosobliwych, że $B_j(0) = B_j$, $B_j(1) = E$. Wtedy krzywa $A(t) = B_1(t) \cdots B_n(t)$ łączy A z E.

b) *Jeśli A można połączyć krzywą ciągłą z B, a B z E, to można w ten sposób połączyć A z E.*

Rzeczywiście, jeśli $A(t)$ jest taką krzywą, że $A(0) = A$, $A(1) = B$, a $B(t)$ taką, że $B(0) = B$, $B(1) = E$, to krzywa

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

łączy A z E. Chwyt polegający na zmianie skali i początku parametryzacji wprowadzony został dlatego, że umówiliśmy się parametryzować krzywe macierzowe liczbami $t \in [0,1]$. Oczywiście moglibyśmy używać dowolnych pośrednich odcinków parametryzacji, przeprowadzać kolejno wszystkie potrzebne deformacje i dopiero na końcu zmienić skalę. Dlatego w dalszym ciągu nie będziemy zwracać uwagi na odcinki parametryzacji.

c) *Dowolną nieosobliwą macierz kwadratową A można przedstawić w postaci iloczynu skończonej liczby macierzy elementarnych, następujących, opisanych niżej typów: $F_{s,t}$, $F_{s,t}(\lambda)$, $F_s(\lambda)$, $\lambda \in \boldsymbol{R}$. Oznaczając przez E_{st} macierz, w której jedynka występuje na miejscu (s,t) i zera na pozostałych miejscach, wymienione powyżej*

macierze są określone wzorami

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

Fakt ten został wykazany we *Wstępie do algebry*, rozdz. 2, § 4, wniosek z twierdzenia p. 5.

Załóżmy teraz , że macierz A jest przedstawiona w postaci iloczynu macierzy elementarnych. Zakładając, że jej wyznacznik jest dodatni, pokażemy, posługując się rezultatami a) i b), jak połączyć ją z E za pomocą pewnej liczby następujących po sobie deformacji.

Przede wszystkim, $\det F_{s,t}(\lambda) = 1$ dla każdego λ i $F_{s,t}(0) = E$. Zatem, zmieniając w wyjściowych czynnikach parametr λ od wartości początkowej do zera, możemy zdeformować wszystkie te czynniki do E, a więc możemy przyjąć, że od samego początku te czynniki nie występowały.

Macierze $F_s(\lambda)$ są diagonalne — na miejscu (s, s) jest λ, a na pozostałych jedynka. Zmieniajmy λ od wartości początkowej do $+1$ albo do -1, odpowiednio do znaku początkowej wartości λ. Rezultatem tej deformacji będzie albo macierz jednostkowa, albo macierz odwzorowania liniowego, które zamienia jeden z wektorów bazy na przeciwny, nie zmieniając żadnego z pozostałych.

Ten etap deformacji macierzy A daje w wyniku macierz złożenia dwóch odwzorowań — jedno jest przestawieniem wektorów bazy ($F_{s,t}$ zamienia miejscami s-ty i t-ty wektor), drugie zamienia znaki części wektorów (złożenie $F_s(+1)$ i $F_t(-1)$).

Dowolną permutację można rozłożyć w iloczyn transpozycji. Macierz $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ transpozycji wektorów bazy płaszczyzny można połączyć z macierzą $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ krzywą $\begin{pmatrix} -\cos t & \sin t \\ \sin t & \cos t \end{pmatrix}$, gdzie $\pi/2 \geq t \geq 0$. Jest jasne, że rozmieszczając elementy tej ostatniej macierzy na miejscach $(s, s), (s, t), (t, s), (t, t)$, otrzymamy odpowiednią deformację znoszącą czynniki $F_{s,t}$.

Konstrukcje opisane dotychczas doprowadziły macierz A do postaci diagonalnej z elementami ± 1 na diagonali, przy tym liczba minus jedynek jest parzysta, gdyż wyznacznik jest dodatni. Macierz $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ można połączyć z $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ krzywą $\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$, gdzie $\pi \geq t \geq 0$. Zbierając wszystkie -1 w pary i przeprowadzając opisane deformacje dla wszystkich tych par, zakończymy dowód.

Wróćmy teraz do omawiania orientacji.

Powiemy, że bazy $\{e_i\}$, $\{e'_i\}$ są *jednakowo zorientowane*, jeśli wyznacznik macierzy przejścia od jednej do drugiej bazy jest dodatni. Jest jasne, że zbiór uporządkowanych baz przestrzeni L rozpada się dokładnie na dwie klasy zawierające

jednakowo zorientowane bazy, podczas gdy bazy należące do różnych klas są zorientowane przeciwnie.

Wybór jednej z tych klas nazywamy *orientacją przestrzeni* L.

Orientacja jednowymiarowej przestrzeni odpowiada wybraniu w niej „dodatniego kierunku" lub półprostej $\boldsymbol{R}_+e = \{ae \mid a > 0\}$, gdzie przez e oznaczamy dowolny wektor określający orientację.

Wybór orientacji w dwuwymiarowej przestrzeni, określony przez dokonanie wyboru bazy $\{e_1, e_2\}$, można wyobrażać sobie jako wyznaczenie „kierunku dodatniego obrotu" płaszczyzny — od e_1 do e_2. Jest to intuicyjnie zgodne z tym, że baza $\{e_2, e_1\}$ zadaje przeciwną orientację (wyznacznik macierzy przejścia $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ jest równy -1) i przeciwny kierunek obrotu.

W ogólnym przypadku przejście od bazy $\{e_i\}$ do bazy $\{e'_i\}$ składającej się z tych samych wektorów, ale inaczej uporządkowanych, zachowuje orientację, jeśli permutacja wektorów jest parzysta, i zmienia ją, jeśli permutacja ta jest nieparzysta. Zmiana znaku u jednego z wektorów bazy zamienia orientację na przeciwną.

W fizycznej trójwymiarowej przestrzeni wybór konkretnej orientacji można związać z pewną osobliwością naszej ludzkiej fizjologii — asymetrią lewej i prawej strony. Lewa strona to ta, gdzie u przeważającej większości ludzi znajduje się serce. Wielki, wskazujący i środkowy palec lewej ręki, zagięte w kierunku od dłoni i ustawione w liniowo niezależnym położeniu przedstawiają uporządkowaną bazę zadającą orientację („reguła lewej dłoni"). Pytanie o to, czy istnieją czysto fizyczne procesy pozwalające wybrać orientację przestrzeni, tj. procesy „nie inwariantne względem odbicia zwierciadlanego", znalazło pozytywną odpowiedź dopiero około dwudziestu lat temu dzięki eksperymentowi, który ku ogólnemu zaskoczeniu wykazał niezachowanie parzystości w słabych oddziaływaniach.

Ćwiczenia

1. Niech (L_1, L_2, L_3) będzie uporządkowaną trójką parami różnych płaszczyzn w $\boldsymbol{K}^3$. Wykazać,że istnieją dwa typy wzajemnego położenia takich trójek charakteryzujące się tym, że $\dim(L_1 \cap L_2 \cap L_3) = 0$ lub 1. Który z tych typów należy uważać za ogólny?

2. Wykazać, że trójki parami różnych prostych w $\boldsymbol{K}^3$ są wszystkie jednakowo położone, a dla czwórek stwierdzenie to nie ma miejsca.

3. Niech $L_1 \subset L_2 \subset \cdots \subset L_n$ będzie flagą w skończenie wymiarowej przestrzeni L, $m_i = \dim L_i$. Wykazać, że jeśli $L'_1 \subset L'_2 \subset \cdots \subset L'_n$ jest inną flagą, $m_i = \dim L_i$, to automorfizm L przeprowadzający jedną flagę na drugą istnieje wtedy i tylko wtedy, gdy $m_i = m'_i$ dla wszystkich i.

4. To samo zagadnienie dla rozkładu na sumę prostą.

5. Wykazać fakty podane w piątym akapicie p. 6.

6. Niech $p\colon L \to L$ będzie projektorem. Wykazać, że $L = \operatorname{Ker} p \oplus \operatorname{Im} p$.

Wyprowadzić stąd, że dowolny projektor w odpowiedniej bazie L ma przedstawienie macierzowe postaci

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

gdzie $r = \dim \operatorname{Im} p$.

7. Niech L będzie N-wymiarową przestrzenią nad ciałem skończonym o q elementach.

a) Obliczyć liczbę k-wymiarowych podprzestrzeni w L, $1 \leq k \leq n$.

b) Obliczyć liczbę par podprzestrzeni $L_1, L_2 \subset L$ o danych wymiarach L_1, L_2 i $L_1 \cap L_2$. Sprawdzić, że wśród wszystkich par o zadanych $\dim L_1$ i $\dim L_2$, względna liczba tych par, które znajdują się w położeniu ogólnym, dąży do 1 przy $q \to \infty$.

§ 6. Przestrzenie ilorazowe

1. Niech L będzie przestrzenią liniową, $M \subset L$ — jej podprzestrzenią liniową i $l \in L$. Przy rozmaitych okazjach będziemy mieć do czynienia ze zbiorami

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

„przesunięć" podprzestrzeni liniowej M o wektor l. Wkrótce przekonamy się, że takie przesunięcia nie są na ogół podprzestrzeniami liniowymi przestrzeni L; nazywają się one *podrozmaitościami liniowymi*. Zaczniemy od wykazania następującego lematu.

2. Lemat. *$l_1 + M_1 = l_2 + M_2$ wtedy i tylko wtedy, gdy $M_1 = M_2 = M$ oraz $l_1 - l_2 \in M$. Inaczej mówiąc, każda podrozmaitość liniowa wyznacza jednoznacznie tę podprzestrzeń liniową, której jest przesunięciem, natomiast wektor przesunięcia — tylko z dokładnością do elementu z tej podprzestrzeni.*

Dowód. Załóżmy najpierw, że $l_1 - l_2 \in M$ i oznaczmy $l_1 - l_2 = m_0$. Mamy

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

Kiedy m przebiega wszystkie wektory z przestrzeni M, wtedy także $m - m_0$ przebiega wszystkie wektory z M. Dlatego $l_1 + M = l_2 + M$. Odwrotnie, niech $l_1 + M = l_2 + M$. Oznaczmy $m_0 = l_1 - l_2$. Z definicji wynika, że $m_0 + M_1 = M_2$, a ponieważ $0 \in M_2$, więc musi być $m_0 \in M_1$. Zatem $m_0 + M_1 = M_1$ na podstawie argumentu z poprzedniego akapitu i w końcu $M_1 = M_2 = M$, co kończy dowód.

3. Definicja. *Przestrzenią ilorazową* L/M przestrzeni liniowej L przez podprzestrzeń M nazywamy zbiór wszystkich podrozmaitości liniowych w L, które są przesunięciami podprzestrzeni M, wyposażony w następujące operacje:

a) $(l_1 + M) + (l_2 + M) = (l_1 + l_2) + M$,

b) $a(l_1 + M) = al_1 + M$ dla dowolnych $l_1, l_2 \in L$, $a \in \boldsymbol{K}$.

Te operacje są określone poprawnie i zadają na L/M strukturę przestrzeni wektorowej nad ciałem K.

4. Sprawdzenie poprawności definicji. Składa się ono z następujących kroków:

a) Jeśli $l_1 + M = l_1' + M$ i $l_2 + M = l_2' + M$, to $l_1 + l_2 + M = l_1' + l_2' + M$.

W istocie, z lematu p. 2 wynika, że $l_1 - l_1' = m_1 \in M$ i $l_2 - l_2' = m_2 \in M$. Dlatego na mocy tegoż lematu

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

ponieważ $m_1 + m_2 \in M$.

b) Jeśli $l_1 + M = l_1' + M$, to $al_1 + M = al_1' + M$. Rzeczywiście, oznaczając ponownie $l_1 - l_1' = m \in M$, mamy $al_1 - al_1' = am \in M$, i teraz zastosowanie lematu p. 2 daje oczekiwany wynik.

W ten sposób pokazaliśmy, że dodawanie i mnożenie przez skalary z $\boldsymbol{K}$ są w istocie dobrze określone. Pozostają do sprawdzenia aksjomaty przestrzeni wektorowej, ale są one bezpośrednią konsekwencją spełnienia odpowiednich wzorów w L. Na przykład, jedną z formuł rozdzielności wykażemy w następujący sposób:

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

Wykorzystaliśmy tu kolejno: określenie dodawania w L/M, określenie mnożenia przez skalary w L/M, rozdzielność w L i znowu określenie dodawania i mnożenia przez skalary w L/M.

5. Uwagi. a) Z definicji wynika, że grupa addytywna L/M jest identyczna z grupą ilorazową grupy addytywnej przestrzeni L przez podgrupę M. W szczególności, podrozmaitość M jest zerem dla L/M.

b) Istnieje odwzorowanie kanoniczne $f: L \to L/M : f(l) = l + M$. Jest ono surjektywne, a jego warstwami, tj. przeciwobrazami elementów, są dokładnie podrozmaitości odpowiadające tym elementom. Istotnie, na mocy lematu p.2

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

Zauważmy, że przy pierwszym pojawieniu się w tym ciągu równości $l + M$ jest traktowany jako element zbioru L/M, natomiast w pozostałych przypadkach jako *podzbiór* L.

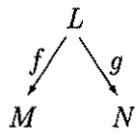
Z punktu 4 wynika, że f jest odwzorowaniem liniowym, a lemat p. 2 pokazuje, że $\operatorname{Ker} f = M$, bowiem $l_0 + M = M$ wtedy i tylko wtedy, gdy $l_0 \in M$.

6. Wniosek. *Jeśli L ma wymiar skończony, to* $\dim L/M = \dim L - \dim M$.

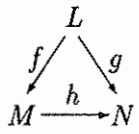
Dowód. Zastosować twierdzenie § 3, p. 12, do skonstruowanego wyżej odwzorowania $L \to L/M$.

W matematyce w wielu ważnych problemach mamy do czynienia z sytuacją, w której przestrzenie $M \subset L$ są nieskończenie wymiarowe, a przestrzeń ilorazowa L/M ma skończony wymiar. W takim przypadku nie można posłużyć się wnioskiem p. 6 i wtedy obliczenie $\dim L/M$ jest na ogół zadaniem nietrywialnym. W ogólności liczbę $\dim L/M$ nazywamy *kowymiarem* podprzestrzeni M w L i oznaczamy $\operatorname{codim} M$ lub $\operatorname{codim}_L M$.

7. Rozważmy zagadnienie, kiedy dla danych dwóch odwzorowań $f\colon L \to M$ i $g\colon L \to N$ istnieje takie odwzorowanie $h\colon M \to N$, że $g = hf$. W języku diagramów można je sformułować następująco: kiedy można włożyć diagram



w trójkąt przemienny



(por. § 13 o diagramach przemiennych). W przypadku odwzorowań liniowych odpowiedź daje następujący rezultat.

8. Stwierdzenie. *Na to, aby takie odwzorowanie h istniało, potrzeba i wystarcza, aby* $\operatorname{Ker} f \subset \operatorname{Ker} g$; *jeśli ten warunek jest spełniony i* $\operatorname{Im} F = M$, *to h jest jedyne.*

Dowód. Jeśli odwzorowanie h istnieje, to z $g = hf$ wynika, że $g(l) = hf(l) = 0$, jeśli tylko $f(l) = 0$. Dlatego $\operatorname{Ker} f \subset \operatorname{Ker} g$.

Odwrotnie, niech $\operatorname{Ker} f \subset \operatorname{Ker} g$. Najpierw określimy odwzorowanie h na podprzestrzeni $\operatorname{Im} f \subset M$, a tu jedyną możliwością jest zadanie h wzorem $h(m) = g(l)$, gdzie l jest tak wybrane, że $m = f(l)$. Teraz należy sprawdzić, że odwzorowanie h jest dobrze określone i jest liniowe na podprzestrzeni $\operatorname{Im} f$. To pierwsze wynika z faktu, że jeśli $m = f(l_1) = f(l)$, to $l_1 - l_2 \in \operatorname{Ker} f \subset \operatorname{Ker} g$, skąd $g(l_1) = g(l_2)$. To drugie wynika automatycznie z liniowości f i g.

Teraz wystarczy przedłużyć odwzorowanie h z podprzestrzeni $\operatorname{Im} f \subset M$ na całą przestrzeń, np. w taki sposób: wybierzmy bazę $\operatorname{Im} f$, uzupełnijmy ją do bazy w M i na dołączonych w ten sposób wektorach bazy przestrzeni M jako wartość h przyjmijmy zero.

9. Niech $g: L \to M$ będzie odwzorowaniem liniowym. Poprzednio zostały zdefiniowane jądro i obraz g, a teraz uzupełnimy tamte określenia, definiując

$$\text{koobraz } g : \operatorname{Coim} g = L/\operatorname{Ker} g,$$
$$\text{kojądro } g : \operatorname{Coker} g = M/\operatorname{Im} g.$$

Istnieje ciąg odwzorowań liniowych „rozkładających g na czynniki"

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

gdzie poza h wszystkie odwzorowania są kanonicznymi włożeniami lub rzutowaniami, a h jest jedynym takim odwzorowaniem, dla którego poniższy diagram jest przemienny

$$\begin{array}{ccc} & L & \\ {\scriptstyle c}\swarrow & & \searrow{\scriptstyle g} \\ \operatorname{Coim} g & \xrightarrow{\quad h \quad} & \operatorname{Im} g. \end{array}$$

Odwzorowanie h jest wyznaczone jednoznacznie, gdyż $\operatorname{Ker} c = \operatorname{Ker} g$, a ponadto jest ono izomorfizmem, bowiem odwzorowanie odwrotne istnieje i też jest jednoznacznie wyznaczone.

Sens łączenia tych przestrzeni w pary (z przedrostkiem „ko" i bez niego) wyjaśnia teoria dwoistości (por. następny paragraf i ćwiczenie 1 po nim).

10. Skończenie wymiarowa alternatywa Fredholma. Niech $g: L \to M$ będzie odwzorowaniem liniowym. Liczbę

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

nazywa się *indeksem* operatora g. Z poprzedniego punktu wynika, że *jeśli L i M są skończenie wymiarowe, to indeks g zależy tylko od L i M*:

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

W szczególności, jeśli $\dim M = \dim L$, np. gdy g jest liniowym operatorem na L, to $\operatorname{ind} g = 0$ dla dowolnego g. Stąd wynika tak zwana alternatywa Fredholma:

albo równanie $g(x) = y$ jest rozwiązalne dla wszystkich y i wówczas równanie $g(x) = 0$ ma tylko rozwiązanie zerowe,

albo równanie to jest rozwiązalne nie dla wszystkich y i wtedy jednorodne równanie $g(x) = 0$ ma niezerowe rozwiązania.

Dokładniej, *jeśli* $\operatorname{ind} g = 0$, *to wymiar przestrzeni rozwiązań równania jednorodnego jest równy kowymiarowi przestrzeni prawych stron, przy których równanie niejednorodne jest rozwiązalne.*

Ćwiczenia

1. Niech M, $N \subset L$. Wykazać, że następujące odwzorowanie jest izomorfizmem liniowym:

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

2. Niech $L = M \oplus N$. Wtedy odwzorowanie kanoniczne

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

jest izomorfizmem.

§ 7. Dwoistość

1. W paragrafie 1 określiliśmy dla dowolnej przestrzeni liniowej L przestrzeń $L^* = \mathcal{L}(L, \boldsymbol{K})$ sprzężoną z nią, a w § 3 przy założeniu $\dim L < \infty$ wykazaliśmy $\dim L^* = \dim L$ oraz podaliśmy konstrukcję kanonicznego izomorfizmu $\varepsilon_L : L \to L^{**}$. Tutaj rozszerzymy opis tej dwoistości, włączając do naszych rozważań odwzorowania liniowe, podprzestrzenie i przestrzenie ilorazowe.

Teoria dwoistości otrzymała swą nazwę dzięki temu, że poświęcona jest badaniu własności „dwustronnej symetrii" przestrzeni liniowych, niezbyt łatwych do przedstawienia w sposób obrazowy, choć zupełnie fundamentalnych. Wspomnimy tu tylko, że charakterystyczny dla mechaniki kwantowej dualizm „fala — cząstka" wyraża się adekwatnie właśnie w języku dwoistości liniowej nieskończenie wymiarowych przestrzeni liniowych, a dokładniej połączenia liniowej i grupowej dwoistości w metodach analizy fourierowskiej.

Wygodnie nam będzie śledzić tę symetrię, zmieniając nieco oznaczenia w porównaniu z tymi, które były wprowadzone paragrafach 1 i 3.

2. Symetria pomiędzy L i L^*. Niech $l \in L$, $f \in L^*$. Dla podkreślenia analogii do iloczynu skalarnego będziemy pisać (f, l) zamiast $f(l)$, choć wektory należą do różnych przestrzeni! W ten sposób określamy odwzorowanie $L^* \times L \to \boldsymbol{K}$, które jest liniowe względem każdego z argumentów przy ustalonym drugim:

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

W ogólności, odwzorowanie $L \times M \longrightarrow \boldsymbol{K}$ z takimi własnościami nazywa się *odwzorowaniem dwuliniowym*, a także *dwoistością przestrzeni L i M*. Wprowadzona właśnie dwoistość pomiędzy L a L^* nazywa się *kanoniczną* (zob. omówienie tego terminu w § 3, p. 8).

Rozważane w § 3, p. 10 odwzorowanie $\varepsilon_L: L \to L^{**}$ można, jak to wynika bezpośrednio z określenia, określić wzorem

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

gdzie po lewej stronie występuje symbol dwoistości pomiędzy L^{**} a L^*, a po prawej — między L^* a L. Jeśli $\dim L < \infty$, a więc gdy ε_L jest izomorfizmem i za jego pośrednictwem możemy utożsamić L^{**} z L, to wzór ten można zapisać w symetrycznej formie $(l, f) = (f, l)$. Innymi słowy, możemy również uważać, że L jest *przestrzenią dwoistą do* L^*.

3. Symetria między bazami dwoistymi. Niech $\{e_1, \ldots, e_n\}$ będzie bazą w L, $\{e^1, \ldots, e^n\}$ — bazą dwoistą w L^*. Zgodnie z definicją § 3, p. 9 jest ona określona wzorami

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

Wynikająca z konwencji przyjętej w poprzedzającym punkcie symetria $(e^i, e_k) = (e_k, e^i)$ oznacza, że baza $\{e_k\}$ jest bazą dwoistą do bazy $\{e^i\}$, gdy traktujemy L jako przestrzeń funkcjonałów liniowych na L^*. W ten sposób $\{e^i\}$ i $\{e_k\}$ tworzą *parę baz dwoistych*, a nadto relacja dwoistości pomiędzy bazami jest symetryczna.

Zapiszmy wektor $l^* \in L^*$ w postaci kombinacji liniowej $\sum_{i=1}^{n} b_i e^i$, a wektor $l \in L$ w postaci $\sum_{i=1}^{n} a_i e_i$. Wtedy

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

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

gdzie $\vec{a}$, $\vec{b}$ oznaczają wektory-kolumny odpowiednich współrzędnych. Ten wzór jest w istocie analogiczny do wzoru wyrażającego iloczyn skalarny dwóch wektorów w przestrzeni euklidesowej, jednakże tutaj wiąże ze sobą wektory z różnych przestrzeni.

4. Odwzorowanie sprzężone lub dwoiste. Niech $f: L \to M$ będzie odwzorowaniem liniowym przestrzeni liniowych. Pokażemy teraz, że istnieje jedoznacznie wyznaczone odwzorowanie liniowe $f^*: M^* \to L^*$, które spełnia warunek

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

dla wszystkich wektorów $m^* \in M^*$, $l \in L$.

a) *Jednoznaczność* f^*. Nich f_1^*, f_2^* będą dwoma takimi odwzorowaniami. Wtedy $(f_1^*(m^*), l) = (m^*, f(l)) = (f_2^*(m^*), l)$ dla wszystkich $m^* \in M^*$, $l \in L$, a więc

$\big((f_1^* - f_2^*)(m^*),\, l\big) = 0$. Ustalając w tej równości m^* i traktując l jako zmienną, widzimy, że $(f_1^* - f_2^*)(m^*) \in L^*$ jest funkcjonałem liniowym na L równym zeru w każdym punkcie, a więc jest zerem. Dlatego $f_1^* = f_2^*$.

b) *Istnienie* f^*. Ustalmy $m^* \in M$ i rozważmy wyrażenie $(m^*, f(l))$ jako funkcję na L. Na mocy liniowości f i dwuliniowości $(\cdot\,,\cdot)$ jest to funkcja liniowa, a więc należy do L^*. Oznaczając ją przez $f^*(m^*)$ i wykorzystując liniowość $(m^*, f(l))$ względem m^*, otrzymujemy

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

co dowodzi, że f^* jest odwzorowaniem liniowym.

Niech w L i M będą wybrane bazy, a w L^* i M^* bazy do nich dwoiste i niech macierz A wyraża w tych bazach odwzorowanie f. Pokażemy, że w bazach dwoistych macierzą f^* jest macierz transponowana A^t. Istotnie, niech B będzie macierzą f^*. Jeśli zgodnie z notacją p. 3 oznaczyć wektory-kolumny współrzędnych wektorów m^*, l przez $\vec{a}$, $\vec{b}$, to

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

Z łączności mnożenia macierzy i jednoznaczności f^* wynika, że $A = B^t$, tj. $B = A^t$.

Podstawowe własności konstrukcji odwzorowania sprzężonego z danym zostały zebrane w poniższym twierdzeniu.

5. Twierdzenie. a) $(f+g)^* = f^* + g^*$;

b) $(af)^* = af^*$; *tutaj* $f,\ g\colon L \to M$ *oraz* $a \in \boldsymbol{K}$;

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

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

e) *jeśli utożsamić* L^{**} *z* L *i* M^{**} *z* M *za pomocą kanonicznych izomorfizmów, to wówczas można także utożsamić* $f^{**}\colon L^{**} \to M^{**}$ *z* $f\colon L \to M$.

Dowód. Przyjmując, że L i M są skończenie wymiarowe, najprościej będzie sprawdzić te własności, używając przedstawienia f i g za pomocą macierzy w bazach dwoistych i korzystając po drodze z prostych własności operacji transponowania macierzy:

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

Niezmienniczy dowód pozostawiamy czytelnikowi jako ćwiczenie.

6. Dwoistość pomiędzy podprzestrzeniami w L i L^*. Niech $M \subset L$ będzie podprzestrzenią liniową. Będziemy oznaczać przez $M^\perp \subset L$ i nazywać *ortogonalnym dopełnieniem do* M zbiór funkcjonałów zerujących się na M. Innymi słowy,

$$m^* \in M^\perp \iff (m^*, m) = 0 \quad \text{dla każdego } m \in M.$$

Łatwo zobaczyć, że $M^\perp$ jest podprzestrzenią liniową. Poniżej zestawiamy najważniejsze własności tej konstrukcji (przy założeniu, że L ma wymiar skończony).

a) *Istnieje izomorfizm kanoniczny* $L^*/M^\perp \to M^*$. Konstrukcja przebiega następująco: przyporządkowujemy podrozmaitości $l^* + M^\perp$ obcięcie funkcjonału l^* do M, zauważając przy tym, że nie zależy ono od wyboru l^*, gdyż obcięcia do M funkcjonałów należących do $M^\perp$ znikają. Liniowość tego przyporządkowania jest oczywista, a surjektywność wynika stąd, że każdy funkcjonał liniowy na M można przedłużyć do funkcjonału na L.

W istocie, niech $\{e_1, \ldots, e_m\}$ będzie bazą w M, a $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ — jej uzupełnieniem do bazy L. Funkcjonał f określony na M i wyznaczony zadaniem wartości $f(e_1), \ldots, f(e_m)$ na wektorach bazy można przedłużyć na przestrzeń L, przyjmując, na przykład, $f(e_{m+1}) = \cdots = f(e_n) = 0$.

Na koniec, skonstruowane odwzorowanie jest injektywne. Rzeczywiście, ma ono zerowe jądro: jeśli obcięcie l^* do M jest zerem, to $l^* \in M^\perp$, więc $l^* + M^\perp = M^\perp$ jest zerem przestrzeni $L^*/M^\perp$.

b) $\dim M + \dim M^\perp = \dim L$. Ta równość wynika z poprzedniego stwierdzenia, wniosku § 6, p. 6 i równości $\dim L^* = \dim L$, $\dim M^* = \dim M$.

c) *Przy kanonicznym utożsamieniu* L^{**} *z* L *przestrzeń* $(M^\perp)^\perp$ *można utożsamić z przestrzenią* M.

Rzeczywiście, ponieważ $(m^*, m) = 0$ dla wszystkich m^* i ustalonego $m \in M$, więc oczywiście $M \subset (M^\perp)^\perp$. Ale ponadto, na mocy poprzedniej własności zastosowanej dwukrotnie

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

A to właśnie oznacza, że $M = (M^\perp)^\perp$.

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

Dowód pozostawiamy jako ćwiczenie.

Ćwiczenia

1. Niech z odwzorowaniem liniowym $g: L \to M$ skończenie wymiarowych przestrzeni będzie związany ciąg odwzorowań skonstruowany w § 6 p.8. Skonstruować izomorfizmy kanoniczne $\operatorname{Ker} g^* \to \operatorname{Coker} g$, $\operatorname{Coim} g^* \to \operatorname{Im} g$, $\operatorname{Im} g^* \to \operatorname{Coim} g$, $\operatorname{Coker} g^* \to \operatorname{Ker} g$.

2. Wyprowadzić stąd tzw. „trzecie twierdzenie Fredholma": na to aby równanie $g(x) = y$ miało rozwiązanie (względem x przy danym y), potrzeba i wystarcza, by y był ortogonalny do jądra odwzorowania sprzężonego $g^*: M^* \to L^*$.

3. Ciąg przestrzeni liniowych i odwzorowań $L \xrightarrow{f} M \xrightarrow{g} N$ nazywamy *dokładnym na miejscu* M, jeśli $\operatorname{Im} f = \operatorname{Ker} g$. Sprawdzić następujące fakty:

a) Ciąg $0 \longrightarrow L \xrightarrow{f} M$ jest dokładny na miejscu L wtedy i tylko wtedy, gdy f jest injekcją.

b) Ciąg $M \xrightarrow{f} N \longrightarrow 0$ jest dokładny na miejscu N wtedy i tylko wtedy, gdy g jest surjekcją.

c) Ciąg odwzorowań liniowych przestrzeni skończenie wymiarowych $0 \longrightarrow L \xrightarrow{f} M \xrightarrow{g} N \longrightarrow 0$ jest dokładny (na każdym miejscu) wtedy i tylko wtedy, gdy jest dokładny ciąg dwoisty $0 \longrightarrow N^* \xrightarrow{g^*} M^* \xrightarrow{f^*} L^* \longrightarrow 0$.

4. Przypomnijmy, że jeśli odwzorowanie $f: L \to M$ w pewnych bazach ma macierz A, to odwzorowanie f^* względem baz dwoistych ma macierz A^t. Wyprowadzić stąd, że rząd macierzy jest równy rzędowi macierzy transponowanej, tj. że maksymalne liczby liniowo niezależnych wierszy i kolumn są takie same.

§ 8. Struktura odwzorowania liniowego

1. W tym paragrafie rozpoczniemy rozważania, których celem jest podanie możliwie przejrzystego geometrycznego opisu odwzorowania liniowego $f: L \to M$. Jeśli L i M nie są ze sobą w żaden sposób związane, to rozwiązanie tego problemu jest bardzo proste i podajemy je w twierdzeniu p. 2 tego paragrafu. Dużo bardziej interesujący i różnorodny obraz powstaje, gdy $M = L$ (ten przypadek opisujemy w tym i następnym paragrafie), a także gdy $M = L^*$ (w części 2). Problem ten, sformułowany w języku macierzowym oznacza sprowadzenie macierzy f do możliwie najprostszej postaci poprzez odpowiedni wybór baz, specjalnie dobranych w zależności od struktury f. W pierwszym przypadku bazy w L i M można wybierać niezależnie od siebie, natomiast w drugim ograniczamy się do jednej bazy L lub też pary baz w L i L^* wzajemnie sprzężonych: mniejsza swoboda wyboru wyraża się w większej różnorodności odpowiedzi.

Używając języka § 5, nasze zadanie można sformułować nieco inaczej. Skonstruujmy zewnętrzną sumę prostą $L \oplus M$ i wraz z odwzorowaniem f rozważmy jego wykres Γ_f — zbiór wszystkich wektorów postaci $(l, f(l)) \in L \oplus M$. Łatwo sprawdzić, że Γ_f jest podprzestrzenią liniową $L \oplus M$. Interesować nas będą niezmienniki położenia Γ_f w $L \oplus M$. W przypadku gdy bazy L i M można zadawać niezależnie od siebie, sytuację opisuje następujące twierdzenie.

2. Twierdzenie. *Niech $f: L \to M$ będzie odwzorowaniem liniowym przestrzeni skończenie wymiarowych. Wówczas prawdziwe są następujące fakty:*

a) *Istnieją takie rozkłady na sumy proste $L = L_0 \oplus L_1$, $M = M_1 \oplus M_2$, że $\mathrm{Ker} f = L_0$, a f indukuje izomorfizm L_1 z M_1.*

b) *Istnieją takie bazy L i M, że macierz f względem tych baz ma postać (a_{ij}), gdzie $a_{ii} = 1$ dla $1 \leq i \leq r$, $a_{ij} = 0$ dla pozostałych i,j.*

c) *Niech A będzie macierzą o wymiarach $m \times n$. Wówczas istnieją takie nieosobliwe macierze B i C o wymiarach $m \times m$ i $n \times n$ i taka liczba $r \leq \min(m,n)$, że macierz BAC ma postać opisaną w poprzedzającym punkcie. Liczba r jest jednoznacznie wyznaczona i jest równa rzędowi A.*

Dowód. a) Przyjmijmy $L_0 = \mathrm{Ker}\, f$, a jako L_1 obierzmy podprzestrzeń dopełniającą do L_0, co jest zawsze możliwe na mocy § 5, p. 10. Następnie jako M_1 obierzemy $\mathrm{Im}\, f$, a jako M_2 jej podprzestrzeń dopełniającą. Pozostaje nam sprawdzić, że f wy-

znacza izomorfizm L_1 na M_1. Odwzorowanie $f: L_1 \to M_1$ jest injektywne, ponieważ przecięcie jądra f, którym jest L_0, z podprzestrzenią L_1 zawiera tylko wektor zerowy. Jest ono także surjektywne, gdyż jeśli $l = l_0 + l_1$, $l_0 \in L_0$, $l_1 \in L_1$, to $f(l) = f(l_1)$.

b) Przyjmijmy $r = \dim L_1 = \dim M_1$ i obierzmy bazę $\{e_1, \ldots, e_r, e_{r+1}, \ldots, e_n\}$ w L, tak aby pierwsze r wektorów tworzyło bazę L_1, a następne — bazę L_0. Wówczas także wektory $e'_i = f(e_i)$, $1 \leq i \leq r$, tworzą bazę $M = \operatorname{Im} f$. Tę ostatnią uzupełnimy do bazy M wektorami $\{e'_{r+1}, \ldots, e'_m\}$. Oczywiście,

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

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

a więc macierz f w wybranych bazach ma zadaną postać.

c) Dla zadanej macierzy A skonstruujmy odpowiadające jej odwzorowanie liniowe f przestrzeni kartezjańskich $\boldsymbol{K}^n \to \boldsymbol{K}^m$, do którego zastosujemy wykazaną już część b) twierdzenia. W nowych bazach macierz f będzie miała szukaną postać i będzie związana z A wzorem BAC, gdzie B, C oznaczają odpowiednie macierze przejścia, por. § 4, p. 8. I wreszcie, $\mathrm{r}(A) = \mathrm{r}(BAC) = \mathrm{r}(f) = \dim \operatorname{Im} f$, co kończy dowód.

Rozważanie operatorów liniowych zaczniemy od zdefiniowania najprostszej ich klasy, a mianowicie *operatorów diagonalizowalnych.*

Wprowadzimy najpierw jedno ogólne pojęcie. Powiemy, że podprzestrzeń $L_0 \subset L$ jest *podprzestrzenią niezmienniczą względem operatora* $f: L \to L$, jeśli $f(L_0) \subset L_0$.

3. Definicja. Operator liniowy $f: L \to L$ nazywamy *diagonalizowalnym*, jeśli spełniony jest którykolwiek z dwóch równoważnych warunków:

a) L ma rozkład na sumę prostą jednowymiarowych podprzestrzeni niezmienniczych dla f;

b) istnieje taka baza L, w której macierz operatora f jest diagonalna.

Równoważność tych warunków można sprawdzić bez trudu. Jeśli w bazie $\{e_i\}$ macierz operatora f jest diagonalna, to $f(e_i) = \lambda_i e_i$, co pokazuje, że jednowymiarowe podprzestrzenie rozpięte przez wektory e_i są niezmiennicze, a L jest ich sumą prostą. Odwrotnie, jeśli $L = \bigoplus_i L_i$ jest takim rozkładem, a e_i — dowolnym niezerowym wektorem z L_i, to $\{e_i\}$ jest bazą L.

Operatory diagonalizowalne stanowią najprostszą, ale z wielu względów najważniejszą klasę operatorów. Na przykład, zobaczymy wkrótce, że nad ciałem liczb zespolonych dowolny operator możemy uczynić diagonalizowalnym, ewentualnie nieznacznie zmieniając jego macierz, a więc operatory „w ogólnym położeniu" są diagonalizowalne. Dla wyjaśnienia, co może być przeszkodą dla diagonalizowalności operatora, wprowadzimy dwie definicje i udowodnimy jedno twierdzenie.

4. Definicja. a) *Podprzestrzenią własną operatora f nazywamy jednowymiarową podprzestrzeń $L_1 \subset L$, jeśli jest ona niezmiennicza dla f, tj. $f(L_1) \subset L_1$.*

Jeżeli tak jest, to wtedy f działa na niej jak mnożenie przez skalar $\lambda \in \boldsymbol{K}$. Ten skalar nazywamy *wartością własną operatora* f (na L_1).

b) Wektor $l \in L$ nazywamy *wektorem własnym* f, jeśli jego liniowa powłoka $\boldsymbol{K}l$ jest podprzestrzenią własną, lub równoważnie, jeśli $l \neq 0$ i $f(l) = \lambda l$ dla pewnego $\lambda \in \boldsymbol{K}$.

Zgodnie z określeniem p. 3 operator diagonalizowalny zadaje rozkład L na sumę prostą podprzestrzeni własnych. Zbadajmy, kiedy f ma choć jedną podprzestrzeń własną.

5. Definicja. Niech L będzie skończenie wymiarową przestrzenią liniową, $f: L \to L$ operatorem liniowym, A — jego macierzą w dowolnej bazie. *Wielomianem charakterystycznym operatora* f, również jego macierzy A, nazywamy wielomian zmiennej t o współczynnikach w ciele $\boldsymbol{K}$ równy $\det(tE - A)$ (gdzie det oznacza wyznacznik). Będzie on zazwyczaj oznaczany przez $P(t)$.

6. Twierdzenie. a) *Wielomian charakterystyczny operatora liniowego* f *nie zależy od wyboru bazy, w której jest wyznaczona jego macierz.*

b) *Każda wartość własna operatora jest pierwiastkiem jego wielomianu charakterystycznego* $P(t)$ *i dowolny pierwiastek* $P(t)$ *należący do ciała* $\boldsymbol{K}$ *jest wartością własną* f *odpowiadającą pewnej (niekoniecznie jedynej) podprzestrzeni własnej operatora* f *w* L.

Dowód. a) Zgodnie z §4 p.8 macierz operatora f w innej bazie wyraża się jako $B^{-1}AB$. Zatem korzystając z multyplikatywności wyznacznika, otrzymujemy

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

Odnotujmy, że $P(t) = t^n - \operatorname{Tr} f \cdot t^{n-1} + \cdots + (-1)^n \det f$ (w oznaczeniach z § 4 p. 9).

b) Jeśli $\lambda \in \boldsymbol{K}$ jest pierwiastkiem $P(t)$, to macierz odwzorowania $\lambda \cdot \mathrm{id} - f$ jest osobliwa, a zatem ma ono nietrywialne jądro. Jeśli $l \neq 0$ jest elementem tego jądra, to $f(l) = \lambda l$, a więc λ jest wartością własną f, a l — odpowiadającym jej wektorem własnym. Odwrotnie, jeśli $f(l) = \lambda l$, to wektor l należy do jądra $\lambda \cdot \mathrm{id} - f$, a więc $\det(\lambda \cdot \mathrm{id} - f) = P(\lambda) = 0$.

7. Teraz widzimy, że operator f może nie mieć wcale wartości własnych, a tym bardziej może nie być diagonalizowalny, jeśli jego wielomian charakterystyczny $P(t)$ nie ma pierwiastków w ciele $\boldsymbol{K}$. Taki przypadek może się zdarzyć nad algebraicznie niedomkniętym ciałem, jakim jest, powiedzmy, ciało $\boldsymbol{R}$ lub ciało skończone. Na przykład, rozważmy macierz $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ o wyrazach rzeczywistych. Wówczas

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

więc jeśli $(a + d)^2 - 4(ad - bc) = (a - d)^2 + 4bc < 0$, to A nie jest diagonalizowalna.

Tutaj spotykamy się po raz pierwszy z przypadkiem, w którym własności odwzorowań liniowych istotnie zależą do własności ciała.

Nie chcąc uzależniać naszych rozważań od tych ostatnich tak długo, jak to tylko będzie możliwe, przyjmiemy założenie, obowiązujące do p. 9 następnego paragrafu włącznie, że mamy do czynienia z algebraicznie domkniętym ciałem $\boldsymbol{K}$. Czytelnik, który nie zna innych niż $\boldsymbol{C}$ przykładów takich ciał, może wszędzie przyjmować $\boldsymbol{K} = \boldsymbol{C}$. Algebraiczną domkniętość $\boldsymbol{K}$ opisują następujące, wzajemnie równoważne warunki: a) dowolny wielomian $P(t)$ jednej zmiennej o współczynnikach w $\boldsymbol{K}$ posiada pierwiastek $\lambda \in \boldsymbol{K}$; b) dowolny wielomian $P(t)$ tego rodzaju można przedstawić w postaci $a \prod_{i=1}^{n} (t-\lambda_i)^{r_i}$, gdzie $a, \lambda_i \in \boldsymbol{K}$, $\lambda_i \neq \lambda_j$ dla $i \neq j$, przy czym to przedstawienie jest jednoznaczne jeśli $P(t) \not\equiv 0$. W takim przypadku liczbę r_i nazywa się *krotnością pierwiastka* λ_i wielomianu $P(t)$. Zbiór wszystkich pierwiastków wielomianu charakterystycznego nazywa się *widmem* (*spektrum*) *operatora* f. Jeśli wszystkie krotności są równe 1, to mówimy, że operator ma *widmo proste*.

Jeśli ciało $\boldsymbol{K}$ jest algebraicznie domknięte, to zgodnie z twierdzeniem p.6 dowolny operator liniowy $f: L \to L$ ma podprzestrzeń własną. Mimo to może się zdarzyć, że nie będzie on diagonalizowalny, gdyż suma wszystkich jego podprzestrzeni własnych będzie mniejsza od L, a dla operatora diagonalizowalnego jest ona zawsze równa L. Zanim jednak przejdziemy do ogólnej dyskusji, zapoznajmy się z przypadkiem macierzy 2×2 o wyrazach zespolonych.

8. Przykład. Niech L będzie dwuwymiarową przestrzenią z ustaloną bazą, w której operator $f: L \to L$ reprezentowany jest przez macierz $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Wielomian charakterystyczny f jest równy $t^2 - (a+d)t + (ad-bc)$, a jego pierwiastkami są $\lambda_{1,2} = -\frac{a+d}{2} \pm \sqrt{\frac{(a-d)^2}{4} + bc}$. Rozważymy kolejno następujące przypadki.

a) $\lambda_1 \neq \lambda_2$. Niech e_i dla $i = 1, 2$, będzie wektorem własnym o wartości własnej λ_i. Te wektory są liniowo niezależne, gdyż z $ae_1 + be_2 = 0$ wynika

$$f(ae_1 + be_2) = a\lambda_1 e_1 + b\lambda_2 e_2 = 0,$$

skąd otrzymujemy $\lambda_1(ae_1 + be_2) - (a\lambda_1 e_1 + b\lambda_2 e_2) = b(\lambda_1 - \lambda_2) = 0$, a więc $b = 0$ i analogicznie $a = 0$. Zatem w bazie $\{e_1, e_2\}$ macierz f jest diagonalna.

b) $\lambda_1 = \lambda_2 = \lambda$. Tym razem operator f jest diagonalizowalny tylko wtedy, gdy jego działanie polega na mnożeniu przez λ dowolnego wektora z L, co oznacza, że $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ i ostatecznie $a = d = \lambda$, $b = c = 0$. Jeśli te warunki nie są spełnione, a spełniony jest tylko słabszy warunek $(a-d)^2 + 4bc = 0$ zapewniający $\lambda_1 = \lambda_2$, to operator f może mieć tylko jeden, z dokładnością do proporcjonalności, wektor własny i wtedy f oczywiście nie jest diagonalizowalny.

Przykładem jest macierz $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, którą nazywa się *klatką Jordana o wymiarach* 2×2 (lub *stopnia* 2).

W następnym paragrafie wykażemy, że właśnie podobne macierze są podstawowymi składnikami, z których konstruuje się postać normalną dowolnego operatora liniowego nad ciałem algebraicznie domkniętym. Podajmy ogólne definicje.

9. Definicja. a) *Klatką Jordana* $J_r(\lambda)$ *stopnia* r *o wartości własnej* λ nazywamy macierz postaci

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

b) *Macierzą Jordana* nazywamy macierz składającą się z rozmieszczonych wzdłuż jej głównej przekątnej bloków $J_{r_i}(\lambda_i)$ i zer wszędzie poza tymi blokami:

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

c) *Bazą Jordana dla operatora* $f: L \to L$ nazywamy taką bazę przestrzeni L, w której macierz operatora f jest macierzą Jordana, lub, jak to się także określa, ma normalną postać Jordana.

d) *Sprowadzeniem macierzy kwadratowej* A *do normalnej postaci Jordana* nazywamy rozwiązanie równania macierzowego $X^{-1}AX = J$, gdzie X oznacza (nieznaną) macierz nieosobliwą, a J — także nieznaną macierz Jordana.

10. Przykład. Niech $L_n(\lambda)$ oznacza przestrzeń liniową funkcji zespolonych postaci $e^{\lambda x}f(x)$, gdzie $\lambda \in \boldsymbol{C}$, a f jest wielomianem stopnia $\leq n-1$. Ponieważ $\frac{d}{dx}(e^{\lambda x}f(x)) = e^{\lambda x}(\lambda f(x)+f'(x))$, różniczkowanie $\frac{d}{dx}$ jest operatorem liniowym w tej przestrzeni. Oznaczmy $e_{i+1} = \frac{x^i}{i!}e^{\lambda x}$ (przypominając, że $0! = 1$) dla $i = 0, \dots, n-1$. Oczywiście,

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

(dla $i = 0$ pierwszy składnik nie występuje), a więc

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

Tak więc funkcje $\left(\frac{x^i}{i!}e^{\lambda x}\right)$ tworzą bazę Jordana dla operatora $\frac{d}{dx}$ w rozważanej przestrzeni.

Ten przykład pokazuje szczególne znaczenie macierzy Jordana w teorii równań różniczkowych liniowych (por. ćwiczenia 1-3 do § 9).

11. Poza rozpatrzonymi już własnościami geometrycznymi będziemy potrzebować w następnym paragrafie pewnych algebraicznych własności funkcji wielomianowych argumentu operatorowego. Niech $f: L \to L$ będzie ustalonym operatorem.

a) Dla dowolnego wielomianu $Q(t) = \sum_{i=1}^{n} a_i t^i$ o współczynnikach z ciała $\boldsymbol{K}$ wyrażenie $\sum_{i=1}^{n} a_i f^i$ określa jednoznacznie element pierścienia $\mathcal{L}(L, L)$ endomorfizmów przestrzeni L. Ten endomorfizm będziemy oznaczać $Q(f)$.

b) Powiemy, że wielomian $Q(t)$ *anihiluje* (także *zeruje*) operator f, jeśli $Q(f) = 0$. Dla dowolnego operatora f w skończenie wymiarowej przestrzeni L niezerowy wielomian o tej własności zawsze istnieje. Istotnie, jeśli $\dim L = n$, to $\dim \mathcal{L}(L, L) = n^2$, a zatem operatory $\mathrm{id}, f, \ldots, f^{n^2}$ są liniowo zależne nad $\boldsymbol{K}$. To rozumowanie wskazuje, że istnieje wielomian stopnia $\leq n^2$ anihilujący f, choć w samej rzeczy z twierdzenia Hamiltona–Cayleya, którego dowiedziemy za chwilę, wynika istnienie wielomianu stopnia n anihilującego operator f.

c) Rozważmy wielomian $M(t)$ o współczynniku przy najwyższej potędze równym jedności, taki który anihiluje f i ma możliwie najmniejszy stopień. Taki wielomian nazywać będziemy *wielomianem minimalnym operatora* f. Oczywiście jest on jednoznacznie wyznaczony, gdyż jeśli $M_1(t)$, $M_2(t)$ byłyby dwoma takimi wielomianami, to $M_1(t) - M_2(t)$ zerowałby f i miałby stopień istotnie niższy, a więc $M_1(t) - M_2(t) = 0$.

d) Pokażemy teraz, że wielomian minimalny operatora f jest dzielnikiem dowolnego wielomianu anihilującego f. Rzeczywiście, niech $Q(f) = 0$. Przeprowadzając dzielenie z resztą Q przez M, zapisujemy $Q(t) = X(t)M(t) + R(t)$, $\deg R(t) < \deg M(t)$. Zatem $R(f) = Q(f) - X(f)M(f) = 0$, skąd $R = 0$.

12. Twierdzenie Hamiltona–Cayleya. *Wielomian charakterystyczny $P(t)$ operatora f anihiluje ten operator.*

Dowód. Będziemy posługiwać się tym twierdzeniem i dowiedziemy go tylko w przypadku ciała algebraicznie domkniętego, choć jest ono prawdziwe i bez tego dodatkowego założenia.

Przeprowadzimy indukcję względem $\dim L$. Jeśli L jest jednowymiarowa, to f jest mnożeniem przez skalar λ, $P(t) = t - \lambda$, a więc $P(f) = 0$.

Niech $\dim L = n \geq 2$ i załóżmy, że twierdzenie jest wykazane dla przestrzeni wymiaru $n - 1$. Wybierzmy wartość własną λ operatora f i jednowymiarową podprzestrzeń własną $L_1 \subset L$ odpowiadającą tej wartości własnej. Niech $\{e_1\}$ będzie bazą L_1, którą uzupełnimy do bazy $\{e_1, \ldots, e_n\}$ przestrzeni L. W tej bazie macierz operatora f ma postać

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

skąd wynika, że $P(t) = (t-\lambda)\det(tE-A)$. Operator f określa odwzorowanie liniowe $\overline{f}: L/L_1 \to L/L_1$, $\overline{f}(l+L_1) = f(l)+L_1$. Wektory $\overline{e}_i = e_i + L_1$ dla $i \geq 2$ tworzą bazę L/L_1, a macierzą operatora $\overline{f}$ w tej bazie jest A. Dlatego $\overline{P}(t) = \det(tE-A)$ jest wielomianem charakterystycznym operatora $\overline{f}$, a więc na mocy założenia indukcyjnego $\overline{P}(\overline{f}) = 0$. To oznacza, że dla dowolnego wektora $l \in L$ mamy $\overline{P}(\overline{f})l \in L_1$, skąd wynika

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

gdyż $f-\lambda$ przeprowadza dowolny wektor z L_1 na zero. To kończy dowód.

13. Przykłady. a) Niech $f = id_L$, $\dim L = n$. Wtedy wielomian charakterystyczny f jest równy $(t-1)^n$, a wielomian minimalny jest równy $t-1$, a więc są one różne dla $n > 1$.

b) Niech f będzie wyznaczony przez klatkę Jordana $J_r(\lambda)$. Wielomian charakterystyczny f jest równy $(t-\lambda)^r$. Dla obliczenia wielomianu minimalnego zauważmy, że $J_r(\lambda) = \lambda E_r + J_r(0)$. Co więcej,

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

gdzie jedynki stoją na k-tej przekątnej powyżej przekątnej głównej, więc $J_r(0)^k = 0$ dla $k \geq r$. Z drugiej strony

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

dla $0 \leq k \leq r-1$, a ponieważ wielomian minimalny jest dzielnikiem wielomianu charakterystycznego, więc muszą być równe.

Ćwiczenia

1. Niech $f: L \to L$ będzie operatorem diagonalnym z widmem prostym.

a) Wykazać, że dowolny operator $g: L \to L$, taki że $gf = fg$, można przedstawić w postaci wielomianu operatora f.

b) Dowieść, że wymiar przestrzeni takich operatorów wynosi $\dim L$.

Czy te stwierdzenia pozostają prawdziwe, jeśli widmo operatora nie jest proste?

2. Niech f, $g: L \to L$ będą operatorami liniowymi w przestrzeni wymiaru n nad ciałem charakterystyki 0. Załóżmy, że $f^n = 0$, $\dim \operatorname{Ker} f = 1$, $gf - fg = f$. Dowieść, że wartości własne g mają postać a, $a-1$, $a-2, \ldots, a-(n-1)$ dla pewnego $a \in \boldsymbol{K}$.

§ 9. Jordanowska postać normalna macierzy

Głównym celem tego paragrafu jest wykazanie następującego twierdzenia o istnieniu i jednoznaczności normalnej postaci jordanowskiej dla macierzy i operatorów liniowych.

1. Twierdzenie. *Niech $\boldsymbol{K}$ będzie ciałem algebraicznie domkniętym, L — skończenie wymiarową przestrzenią liniową nad $\boldsymbol{K}$, a $f: L \to L$ — operatorem liniowym. Wtedy:*

a) *istnieje baza Jordana dla operatora f, a dokładniej, macierz A operatora f w dowolnie wybranej bazie można sprowadzić do normalnej postaci Jordana J za pomocą odpowiednio dobranej macierzy przejścia X do innej bazy, tzn.*

$$X^{-1}AX = J;$$

b) *macierz J jest wyznaczona jednoznacznie z dokładnością do przestawienia wchodzących w jej skład klatek jordanowskich.*

2. Dowód twierdzenia podzielimy na szereg kroków. Zaczniemy od konstrukcji rozkładu na sumę prostą $L = \bigoplus_{i=1}^{n} L_i$, gdzie L_i są podprzestrzeniami niezmienniczymi dla f, z którymi później związemy klatki jordanowskie o jednej i tej samej liczbie λ na przekątnej. Dla wprowadzenia niezmienniczej charakteryzacji tych podprzestrzeni przypomnijmy, że $(J_r(\lambda) - \lambda E)^r = 0$. Operatory, których pewna potęga jest równa zeru, przyjęto nazywać *operatorami nilpotentnymi*. A więc obcięcie operatora $f - \lambda$ do podprzestrzeni odpowiadającej klatce $J_r(\lambda)$ jest operatorem nilpotentnym i to samo jest prawdziwe dla jego obcięcia do sumy podprzestrzeni z ustalonym λ. To uzasadnia poniższą definicję.

3. Definicja. Wektor $l \in L$ nazywamy *wektorem pierwiastkowym operatora f odpowiadającym liczbie $\lambda \in \boldsymbol{K}$*, jeśli istnieje liczba naturalna r, dla której $(f-\lambda)^r l = 0$ — tutaj $(f - \lambda)$ oznacza operator $(f - \lambda \mathrm{id})$.

Oczywiście każdy wektor własny jest wektorem pierwiastkowym.

4. Stwierdzenie. *Jeśli oznaczyć przez $L(\lambda)$ podzbiór przestrzeni L składający się z wektorów pierwiastkowych operatora f odpowiadających danemu λ, to $L(\lambda)$ jest podprzestrzenią liniową L, a $L(\lambda) \neq 0$ wtedy i tylko wtedy, gdy λ jest wartością własną f.*

Dowód. Przypuśćmy, że $(f - \lambda)^{r_1} l_1 = (f - \lambda)^{r_2} l_2 = 0$. Wówczas mamy $(f - \lambda)^r (l_1 + l_2) = 0$ dla $r = \max(r_1, r_2)$ oraz $(f - \lambda)^{r_1}(a l_1) = 0$, co dowodzi, że $L(\lambda)$ jest podprzestrzenią liniową.

Jeśli λ jest wartością własną f, to istnieje wektor własny odpowiadający λ, a więc $L(\lambda) \neq 0$. Odwrotnie, niech $l \in L(\lambda)$, $l \neq 0$. Wybierzmy najmniejsze r, dla którego $(f - \lambda)^r = 0$ — oczywiście $r \geq 1$. Wektor $l' = (f - \lambda)^{r-1} l$ jest wektorem własnym f odpowiadającym wartości własnej λ, bo $l' \neq 0$ na podstawie konstrukcji, a z $(f - \lambda) l' = 0$ wynika $f(l') = \lambda l'$.

5. Stwierdzenie. *Zachodzi rozkład* $L = \bigoplus L(\lambda_i)$, *gdzie* λ_i *przebiega wszystkie wartości własne operatora* f, *tj. wszystkie różne pierwiastki wielomianu charakterystycznego operatora* f.

Dowód. Niech $P(t) = \prod_{i=1}^{s}(t-\lambda_i)^{r_i}$ będzie wielomianem chrakterystycznym f, $\lambda_i \neq \lambda_j$ dla $j \neq i$. Oznaczmy $F_i(t) = P(t)(t-\lambda_i)^{-r_i}$, $f_i = F_i(f)$, $L_i = \operatorname{Im} f_i$. Sprawdzimy następujące fakty.

a) $(f-\lambda_i)^{r_i}L_i = 0$, tj. $L_i \subset L(\lambda_i)$. Rzeczywiście, $(f-\lambda_i)^{r_i}f_i = (f-\lambda_i)^{r_i}F_i(f) = P(f) = 0$ na podstawie twierdzenia Hamiltona–Cayleya.

b) $L = L_1 + \cdots + L_s$. Rzeczywiście, ponieważ wielomiany $F_i(t)$ są parami względnie pierwsze, więc istnieją takie wielomiany $X_i(t)$, że $\sum_{i=1}^{s} F_i(t)X_i(t) = 1$. Podstawiając w tej tożsamości operator f zamiast zmiennej t, otrzymujemy

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

Stosując tę równość do wektora $l \in L$, otrzymamy

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

c) $L = L_1 \oplus \cdots \oplus L_s$. Wystarczy sprawdzić, że dla dowolnego $1 \leq i \leq s$ mamy $L_i \cap \left(\sum_{i\neq j} L_j\right) = 0$. Rzeczywiście, jeśli wektor l należy do tego przecięcia, to

$$(f-\lambda_i)^{r_i}l = 0, \quad \text{gdyż } l \in L_i,$$
$$F_i(f)l = \prod_{j\neq i}^{s}(f-\lambda_i)^{r_i}l = 0, \quad \text{gdyż } l \in \sum_{j\neq i} L_j.$$

Ponieważ $(t-\lambda_i)^{r_i}$ i $F_i(t)$ są względnie pierwsze, więc istnieją takie wielomiany $X(t)$ i $Y(t)$, że $X(t)(t-\lambda_i)^{r_i} + Y(t)F_i(t) = 1$. Podstawiając f i stosując uzyskaną tożsamość operatorową do wektora l, otrzymamy $X(f)(0) + Y(f)(0) = l = 0$.

d) $L_i = L(\lambda_i)$. W istocie, już sprawdziliśmy, że $L_i \subset L(\lambda_i)$. Dla dowodu przeciwnej inkluzji wybierzmy wektor $l \in L(\lambda_i)$ i zapiszmy go w postaci $l = l' + l''$, gdzie $l' \in L_i$, $l'' \in \bigoplus_{i\neq j} L_j$. Istnieje taka liczba r', że $(f-\lambda_i)^{r'}l'' = 0$, ponieważ $l'' = l - l' \in L(\lambda_i)$. Ponadto $F_i(f)l'' = 0$. Zatem podstawiając w tożsamości $X(t)(t-\lambda_i)^{r'} + Y(t)F_i(t) = 1$ operator f zamiast zmiennej, a następnie stosując otrzymaną równość operatorową do wektora l'', otrzymamy $l'' = 0$, skąd $l = l' \in L_i$.

6. Wniosek. *Operator* f *z widmem prostym jest diagonalizowalny.*

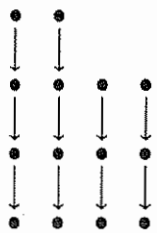
Dowód. Istotnie, liczba różnych wartości własnych f jest wtedy równa $n = \deg P(t) = \dim L$. Dlatego w rozkładzie $L = \bigoplus_{i=1}^{n} L(\lambda_i)$ wszystkie przestrzenie $L(\lambda_i)$

są jednowymiarowe, a że każda z nich zawiera wektor własny, więc w bazie złożonej z tych wektorów macierz f jest diagonalna.

Teraz ustalimy jedną wartość własną λ i pokażemy, że obcięcie f do $L(\lambda)$ ma bazę Jordana odpowiadającą λ. Aby nie wprowadzać nowych oznaczeń, będziemy przyjmować aż do końca p. 7, że *f ma tylko jedną wartość własną λ i $L = L(\lambda)$*. Co więcej, możemy nawet przyjąć, że $\lambda = 0$, gdyż dowolna baza Jordana dla operatora f jest jednocześnie bazą Jordana dla operatora $f+\mu$, gdzie μ jest dowolną stałą. Zatem operator f jest nilpotentny na mocy twierdzenia Hamiltona–Cayleya: $P(t) = t^n$, $f^n = 0$ i prawdziwy jest następujący fakt.

7. Stwierdzenie. *Dla operatora nilpotentnego f w skończenie wymiarowej przestrzeni L istnieje baza Jordana, a jego macierz względem tej bazy jest sumą klatek postaci $J_r(0)$.*

Dowód. Bazie Jordana w przestrzeni L wygodnie przyporządkować diagram podobny do przedstawionego poniżej:



W tym diagramie punkty reprezentują elementy bazy, a strzałki przedstawiają działanie f (a w ogólnym przypadku działanie $f - \lambda$). Elementy najniższego wiersza operator przeprowadza na zero, tj. w tym wierszu występują wektory własne f wchodzące w skład danej bazy. W ten sposób każda kolumna diagramu przedstawia bazę podprzestrzeni niezmienniczej odpowiadającej jednej klatce Jordana, której stopień jest równy wysokości tej kolumny (tj. liczbie występujących w niej punktów). Jeśli oznaczyć

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

to możemy zapisać

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

Ale i odwrotnie, jeśli znajdziemy bazę L, której elementy przechodzą na inne jej elementy albo na zero w taki sposób, że działanie f na elementy tej bazy wraz z nimi samymi można przedstawić za pomocą podobnego diagramu, to będzie ona bazą Jordana.

Dowód istnienia przeprowadzimy przez indukcję względem wymiaru L. Jeśli $\dim L = 1$, to operator nilpotentny f jest operatorem zerowym i dowolny niezerowy wektor L stanowi bazę Jordana. Rozważmy więc przypadek $\dim L = n > 1$ i załóżmy, że wykazaliśmy już istnienie baz Jordana dla przestrzeni o wymiarach $< n$. Oznaczmy przez L_0 podprzestrzeń w L złożoną z wektorów własnych f, tj. $\operatorname{Ker} f$. Ponieważ $\dim L_0 > 0$, więc $\dim L/L_0 < n$ i operator f indukuje operator $\overline{f}: L/L_0 \to L/L_0$, $\overline{f}(l + L_0) = f(l) + L_0$. (Poprawność określenia operatora $\overline{f}$ i jego liniowość łatwo wynika z niezmienniczości L_0 względem f.)

Na mocy założenia indukcyjnego $\overline{f}$ ma bazę Jordana i możemy przyjąć, że jest ona niepusta, gdyż w przeciwnym przypadku $L = L_0$ i dowolna baza L_0 będzie bazą Jordana dla f. Zbudujmy teraz diagram $\overline{D}$ dla elementów bazy Jordana operatora $\overline{f}$ i z każdej kolumny tego diagramu wybierzmy najwyżej położony wektor $\overline{e}_i$, $i = 1, \ldots, m$, dla którego obierzemy następnie wektor $e_i \in L$ tak aby $\overline{e}_i = e_i + L_0$. Teraz możemy z wektorów przestrzeni L zbudować diagram D w następujący sposób. Dla $i = 1, \ldots, m$ i-ta kolumna diagramu D będzie zawierać (w kierunku z góry na dół) wektory $e_i, f(e_i), \ldots, f^{h_i-1}(e_i), f^{h_i}(e_i)$, gdzie przez h_i oznaczyliśmy wysokość i-tej kolumny diagramu $\overline{D}$. Ponieważ $\overline{f}^{h_i}(\overline{e}_i) = 0$, więc $f^{h_i}(e_i) \in L_0$ i $f^{h_i+1}(e_i) = 0$. Następnie wybierzemy bazę powłoki liniowej wektorów $f^{h_1}(e_1), \ldots, f^{h_m}(e_m)$ i biorąc pod uwagę, że zawiera się ona w L_0, uzupełnimy ją do bazy L_0. Dołączone przy tym wektory umieścimy w dodatkowych kolumnach (o wysokości 1) w najniższym wierszu diagramu D — zauważmy, że f przeprowadza je na zero.

Skonstruowany w ten sposób diagram D wektorów przestrzeni L razem z działaniem operatora f na te wektory ma dokładnie taką postać, jaka określa bazę Jordana. Musimy jedynie jeszcze sprawdzić, że wektory diagramu D rzeczywiście stanowią bazę L.

Najpierw pokażemy, że powłoka liniowa wektorów diagramu D jest równa L. Niech $l \in L$, $\overline{l} = l + L_0$. Z założenia $\overline{l} = \sum_{i=1}^{m} \sum_{j=0}^{h_i-1} a_{ij} \overline{f}^j(\overline{e}_i)$. Z konstrukcji $\overline{f}$ wynika, że

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

Wszystkie wektory $f^j(e_i)$, $j \leq h_i - 1$, występują w wierszach diagramu D, poczynając od drugiego z dołu, a podprzestrzeń L_0 jest rozpięta przez wektory występujące w pierwszym od dołu wierszu na mocy konstrukcji. Dlatego też l można przedstawić jako kombinację elementów D.

Do sprawdzenia pozostała zatem tylko liniowa niezależność wektorów D. Przede wszystkim elementy dolnego wiersza są liniowo niezależne. Rzeczywiście, jeśli jakaś nietrywialna kombinacja liniowa tych wektorów jest zerem, to musi mieć postać $\sum_{i=1}^{m} a_i f^{h_i}(e_i) = 0$, gdyż pozostałe elementy tego wiersza uzupełniają bazę powłoki liniowej $\{f^{h_1}(e_1), \ldots, f^{h_m}(e_m)\}$ do bazy L_0. Ponieważ dla każdego i mamy $h_i \geq 1$,

zatem

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

skąd wynika

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

Z ostatniego z tych warunków wnioskujemy, że wszystkie $a_i = 0$, gdyż wektory $\overline{f}^{h_i-1}(\overline{e}_i)$ stanowią dolny wiersz diagramu D, a więc są częścią bazy $L\Lambda_0$.

Wreszcie wykażemy, że jeśli jakaś nietrywialna kombinacja wektorów D jest równa zeru, to można z niej otrzymać nietrywialną zależność liniową dla wektorów najniższego wiersza D. W istocie, odnajdźmy najwyższy wiersz D, w którym występują wektory wchodzące z niezerowymi współczynnikami w tę kombinację liniową i niech będzie to wiersz h, licząc od dołu diagramu. Zastosujmy do tej kombinacji liniowej operator f^{h-1}. Wtedy oczywiście ta część kombinacji liniowej, która jest złożona z wektorów zawartych w tym najwyższym wierszu przejdzie na nietrywialną kombinację liniową elementów najniższego wiersza, podczas gdy pozostałe składniki zostaną wyzerowane. To kończy dowód stwierdzenia.

Pozostała nam już tylko do sprawdzenia część twierdzenia z p. 1 odnosząca się do jednoznaczności.

8. Ustalmy dowolną bazę Jordana operatora f. Dowolny diagonalny element macierzy operatora f względem tej bazy jest oczywiście jedną z wartości własnych λ tego operatora. Wybierzmy tę część bazy, która odpowiada wszystkim blokom macierzy z tą wartością λ na diagonali i oznaczymy przez L_λ jej powłokę liniową. Ponieważ $(J_r(\lambda) - \lambda)^r = 0$, więc $L_\lambda \subset L(\lambda)$, gdzie $L(\lambda)$ oznacza podprzestrzeń pierwiastkową L. Ponadto $L = \bigoplus L_{\lambda_i}$ na mocy określenia bazy Jordana i $L = \bigoplus L(\lambda_i)$ na mocy stwierdzenia p. 5, przy czym w obu przypadkach sumy obejmują wszystkie wartości własne λ_i operatora f. A zatem $\dim L_\lambda = \dim L(\lambda_i)$ i $L_\lambda = L(\lambda_i)$. To znaczy, że suma wymiarów klatek Jordana odpowiadających każdej wartości λ_i nie zależy od wyboru bazy Jordana, a ponadto od wyboru tej bazy nie zależą też powłoki liniowe L_{λ_i}. Dlatego wystarczy sprawdzić twierdzenie w przypadku $L = L(\lambda)$ albo nawet $L = L(0)$.

Skonstruujmy diagram D odpowiadający danej bazie Jordana $L = L(0)$. Wymiary klatek Jordana są równe wysokościom jego kolumn; gdyby, jak na naszym rysunku, uporządkować kolumny malejąco, to wówczas ich wysokości można by jednoznacznie określić, posługując się znanymi długościami wierszy diagramu, uporządkowanymi malejąco, poczynając od najniższego wiersza. Pokażemy, że długość najniższego wiersza jest równa wymiarowi $L_0 = \operatorname{Ker} f$. W istocie, przedstawmy dowolny wektor własny l operatora f w postaci kombinacji liniowej elementów D. Do tej kombinacji liniowej ze współczynnikami zerowymi wejdą wszystkie te wektory, które znajdują się powyżej najniższego wiersza. Gdyby tak nie było, tj. najwyżej położone w diagramie wektory o niezerowych współczynnikach w tej kombinacji pochodziłyby

z wiersza o numerze $h \geq 2$, to wtedy wektor $f^{h-1}(l) = 0$ byłby nietrywialną kombinacją liniową elementów najniższego wiersza D (por. koniec dowodu stwierdzenia p. 7), a to jest sprzeczne z liniową niezależnością elementów D. Tak więc najniższy wiersz D tworzy bazę L_0 i dlatego jego długość jest jednakowa dla wszystkich baz Jordana. Dokładnie tak samo dowodzimy, że długość drugiego wiersza nie zależy od wyboru bazy, gdyż jest ona równa wymiarowi $\operatorname{Ker} \overline{f}$ w przestrzeni L/L_0, przy użyciu oznaczeń poprzedzającego punktu. To kończy dowód jednoznaczności, więc i całego twierdzenia p. 1.

9. Uwagi. a) Jeśli operator f jest reprezentowany przez macierz A w pewnej bazie, to problem sprowadzenia A do postaci Jordana wygodnie rozwiązywać, używając następującej procedury.

Najpierw wyznaczamy wielomian charakterystyczny A i jego pierwiastki.

Następnie obliczamy wymiary klatek Jordana odpowiadających pierwiastkom λ. W tym celu wystarczy obliczyć długości wierszy odpowiednich diagramów, to znaczy znaleźć $\dim \operatorname{Ker}(A-\lambda)$, $\dim \operatorname{Ker}(A-\lambda)^2 - \dim \operatorname{Ker}(A-\lambda)$, $\dim \operatorname{Ker}(A-\lambda)^3 - \dim \operatorname{Ker}(A-\lambda)^2$, itd.

Teraz możemy zapisać macierz A w postaci Jordana J i poszukać rozwiązań równania macierzowego $AX - XJ = 0$. Przestrzeń rozwiązań tego układu liniowego ma najczęściej wymiar większy niż jeden, a zbiór rozwiązań zawiera również macierze osobliwe. Jednakże istnienie wśród rozwiązań macierzy nieosobliwych jest zapewnione przez udowodnione wyżej twierdzenie, a dla naszego celu wystarczy dowolna z nich.

b) Jednym z ważniejszych zastosowań postaci Jordana macierzy jest obliczanie funkcji macierzy (dotychczas znane są nam tylko funkcje wielomianowe). Przyjrzyjmy się na przykład, jak obliczyć wysoką potęgę A^N macierzy A. Ponieważ potęgi macierzy Jordana oblicza się łatwo (zob. § 8, p. 13), zastosowanie wzoru $A^N = XJ^NX^{-1}$, gdzie $A = XAX^{-1}$, może okazać się ekonomiczne, gdyż macierz X oblicza się tylko raz i nie zależy ona od N. Ten sam wzór można zastosować do oszacowania szybkości wzrastania wyrazów macierzy A^N.

c) Wykorzystując znajomość formy Jordana, łatwo znaleźć wielomian minimalny macierzy A. Ograniczmy się dla prostoty do przypadku ciała charakterystyki zero. Wtedy wielomian minimalny $J_r(\lambda)$ jest równy $(t-\lambda)^r$ (por. § 8, p. 13), wielomian minimalny macierzy blokowej $(J_{r_i}(\lambda))$ wynosi $(t-\lambda)^{\max(r_i)}$, a wreszcie wielomian minimalny najogólniejszej macierzy Jordana z elementami diagonalnymi $\lambda_1, \ldots, \lambda_s$ ($\lambda_i \neq \lambda_j$ dla $i \neq j$) jest równy $\prod_{j=1}^{s} (t-\lambda_j)^{r_j}$, gdzie r_j jest największym spośród wymiarów klatek Jordana odpowiadających wartości λ_j.

10. Inne formy normalne. W tym punkcie krótko opiszemy inne normalne postacie macierzy, pożyteczne np. w przypadku ciał niedomkniętych algebraicznie.

a) *Przestrzenie cykliczne i klatki cykliczne.* Przestrzeń L nazywamy *cykliczną* względem operatora f, jeśli L zawiera wektor l, także nazywany *cyklicznym*, o tej własności, że wektory $l, f(l), \ldots, f^{n-1}(l)$ tworzą bazę L. Oznaczając $e_i = f^{n-i}(l)$,

$i = 1, \ldots, n = \dim L$, otrzymujemy

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

gdzie $a_i \in \boldsymbol{K}$ są jednoznacznie wyznaczone przez relację $f^n(l) = \sum_{i=0}^{n-1} a_i f^i(l)$. Macierz operatora f w takiej bazie nazywamy *klatką cykliczną*. Na odwrót, jeśli macierz operatora f w bazie $(e_1, \ldots, e_n)$ jest klatką cykliczną, to wektor $l = e_n$ jest cykliczny oraz $e_i = f^{n-i}(l)$ (indukcja względem i).

Wykażemy, że postać klatki cyklicznej odpowiadającej f nie zależy od początkowego wyboru wektora cyklicznego. W tym celu sprawdzimy, że w pierwszej kolumnie tej klatki stoją współczynniki wielomianu minimalnego operatora f — $M(t) = t^n - \sum_{i=0}^{n-1} a_i t^i$.

Istotnie, $M(f) = 0$, ponieważ $M(f)[f^i(l)] = f^i[M(f)l] = 0$, a wektory $f^i(l)$ generują L. Z drugiej strony, jeśli $N(t)$ jest wielomianem stopnia $< n$, to $N(f) \neq 0$, ponieważ w przeciwnym przypadku, działając operatorem $N(f) = 0$ na wektor cykliczny l, otrzymalibyśmy nietrywialną relację liniową dla wektorów bazy $l, f(l), \ldots, f^{n-1}(l)$.

b) *Kryterium cykliczności przestrzeni.* Zgodnie z poprzednimi rozważaniami, jeśli przestrzeń L jest cykliczna względem f, to jej wymiar jest równy stopniowi wielomianu minimalnego operatora f i w konsekwencji ten ostatni jest równy wielomianowi charakterystycznemu. Prawdziwe jest również odwrócenie tej obserwacji: jeśli operatory id, $f, \ldots, f^{n-1}$ są liniowo niezależne, to istnieje taki wektor l, że wektory l, $f(l), \ldots, f^{n-1}(l)$ są liniowo niezależne, a przestrzeń L jest cykliczna. Tego nie będziemy dowodzić.

c) *Macierz dowolnego operatora może być przez wybór odpowiedniej bazy sprowadzona do sumy klatek cyklicznych.* Dowód można przeprowadzić analogicznie do dowodu twierdzenia o postaci Jordana. Zamiast czynników $(t - \lambda_i)^{r_i}$ wielomianu charakterystycznego należy rozważać czynniki $p_i(t)^{r_i}$, gdzie $p_i(t)$ są nierozkładalnymi nad ciałem $\boldsymbol{K}$ dzielnikami wielomianu charakterystycznego. Twierdzenie o jednoznaczności także pozostaje prawdziwe, jeśli ograniczyć się do przypadku, gdy wielomiany minimalne wszystkich klatek cyklicznych są nierozkładalne. Bez tego ograniczenia twierdzenie nie jest prawdziwe: przestrzeń cykliczna może być sumą prostą dwóch podprzestrzeni cyklicznych, których wielomiany minimalne są względnie pierwsze.

Ćwiczenia

1. Niech L będzie skończenie wymiarową przestrzenią funkcji różniczkowalnych zmiennej zespolonej, taką że jeśli $f \in L$, to $\frac{df}{dx} \in L$. Pokazać, że istnieją takie liczby

zespolone $\lambda_1, \ldots, \lambda_s$ i liczby naturalne $r_1, \ldots, r_s \geq 1$, że $L = \bigoplus L_i$, gdzie L_i jest przestrzenią funkcji postaci $e^{\lambda_i x} P_i(x)$, a $P_i(x)$ oznacza dowolny wielomian stopnia $\leq r_i - 1$. (*Wskazówka.* Rozważyć bazę Jordana dla operatora $\frac{d}{dx}$ w L i kolejno wyznaczać postać wchodzących do L funkcji, zaczynając od najniższego wiersza diagramu.)

2. Niech $y(x)$ oznacza funkcję zmiennej zespolonej x spełniającą równanie różniczkowe postaci

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

Oznaczmy przez L przestrzeń liniową funkcji rozpiętą przez $\frac{d^i y}{dx^i}$ dla wszystkich $i \geq 0$. Dowieść, że L ma wymiar skończony i operator $\frac{d}{dx}$ przeprowadza ją w siebie.

3. Posługując się rezultatami ćwiczeń 1 i 2 wykazać, że $y(x)$ można przedstawić w postaci $\sum e^{\lambda_i x} P_i(x)$, gdzie P_i są wielomianami. Jak wiążą się liczby λ_i z postacią równania różniczkowego?

4. Niech $J_r(\lambda)$ będzie klatką Jordana nad $\boldsymbol{C}$. Wykazać, że dowolnie mało zmieniając jej elementy można doprowadzić do tego, aby otrzymana macierz była diagonalizowalna. (*Wskazówka.* Zmienić tak elementy diagonalne, aby były parami różne.)

5. Rozszerzyć wynik tego ćwiczenia na dowolne macierze nad $\boldsymbol{C}$, wykorzystując to, że współczynniki wielomianu charakterystycznego są funkcjami ciągłymi współczynników macierzy, a warunek niewystępowania wielokrotnych pierwiastków wielomianu sprowadza się do tego, aby jego wyróżnik był różny od 0.

6. Nadać precyzyjne znaczenie a następnie dowieść następujących stwierdzeń:

a) Generyczna macierz 2×2 nad $\boldsymbol{C}$ jest diagonalizowalna.

b) Generyczna macierz 2×2 z równymi wartościami własnymi nie jest diagonalizowalna.

§ 10. Przestrzenie liniowe unormowane

W tym paragrafie zbadamy te szczególne własności rzeczywistych lub zespolonych przestrzeni liniowych, które umożliwiają wprowadzenie pojęcia przejścia granicznego w tych przestrzeniach i uprawianie w nich analizy. Te zagadnienia odgrywają szczególną rolę w przypadku nieskończenie wymiarowym, tak że przedstawiony tu materiał jest w istocie elementarnym wstępem do analizy funkcjonalnej.

1. Definicja. Parę (E, d), gdzie E jest zbiorem, $d: E \times E \to \boldsymbol{R}$ — funkcją o wartościach rzeczywistych, nazywamy *przestrzenią metryczną*, jeśli dla dowolnych $x, y, z \in E$ spełnione są następujące warunki:

a) $d(x,y) = d(y,x)$ (*symetria*);

b) $d(x,x) = 0$; $d(x,y) > 0$, *jeśli* $x \neq y$ (*dodatniość*);

c) $d(x,z) \leq d(x,y) + d(y,z)$ (*nierówność trójkąta*).

Funkcję d spełniającą te warunki nazywamy *metryką*, a liczbę $d(x,y)$ — *odległością punktów* x, y.

2. Przykłady. a) $E = \boldsymbol{R}$ albo $\boldsymbol{C}$, $d(x,y) = |x-y|$.

b) $E = \boldsymbol{R}^n$ albo $\boldsymbol{C}^n$, $d(\vec{x},\vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2}$.

Jest to tak zwana metryka standardowa (naturalna). Będziemy ją rozważać systematycznie w drugiej części i tam też zbadamy jej uogólnienia na przypadek dowolnego ciała opisywane w teorii form kwadratowych. Innymi metrykami są

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

c) $E = C(a,b)$ — funkcje ciągłe na odcinku $[a,b]$. A oto trzy najważniejsze przykłady metryk w tym zbiorze:

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

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

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

(Sprawdzić aksjomaty. Dla d_2 w przykładzie b) i d_3 w przykładzie c) nierówność trójkąta będzie wykazana w następnej części książki.) d) E dowolny zbiór, $d(x,y) = 1$ dla $x \neq y$. Jest to jedna z tak zwanych dyskretnych metryk na E.

(Z każdą metryką jest związana pewna topologia na E, a opisana metryka indukuje topologię dyskretną.)

3. Kule, ograniczoność i zupełność. W przestrzeni metrycznej E z metryką d wyróżniamy podzbiory

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

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

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

które będziemy nazywać odpowiednio *kulą otwartą*, *kulą domkniętą* i *sferą* o środku w punkcie x_0 i promieniu r. Nie należy jednak z tymi nazwami łączyć zbyt ściśle

intuicyjnych wyobrażeń zaczerpniętych z przypadku trójwymiarowego, gdyż np. w przestrzeni opisanej w punkcie d) p. 2 wszystkie sfery o promieniu $r \neq 1$ są zbiorami pustymi.

Podzbiór $F \subset E$ nazywamy *ograniczonym*, jeśli zawiera się w pewnej kuli (o skończonym promieniu).

Ciąg punktów $x_1, x_2, \ldots, x_n, \ldots$ przestrzeni E jest *zbieżny* do punktu $a \in E$, jeśli $\lim_{n\to\infty} d(x_n, a) = 0$. Ciąg nazywamy *ciągiem Cauchy'ego*, jeśli dla każdego $\varepsilon > 0$ istnieje $N = N(\varepsilon)$, takie że $d(x_n, x_m) < \varepsilon$ dla $m, n > N(\varepsilon)$.

Przestrzeń metryczną nazywamy *zupełną*, jeśli dowolny ciąg Cauchy'ego w tej przestrzeni jest zbieżny. Z zupełności $\boldsymbol{R}$ i $\boldsymbol{C}$ dowodzonej w analizie wynika, że przestrzenie $\boldsymbol{R}^n$ i $\boldsymbol{C}^n$ z dowolną z metryk d, d_1, d_2 przykładu b) p. 2 są zupełne.

4. Unormowane przestrzenie liniowe. Niech L oznacza przestrzeń liniową nad $\boldsymbol{R}$ lub $\boldsymbol{C}$. Szczególnie ważną rolę grają metryki na L spełniające następujące dwa warunki:

a) $d(l_1, l_2) = d(l_1 + l, l_2 + l)$ dla dowolnych $l, l_1, l_2 \in L$ (niezmienniczość względem przesunięć);

b) $d(al_1, al_2) = |a| d(l_1, l_2)$ (mnożenie przez skalar a zwiększa odległości $|a|$ razy).

Niech d będzie taką metryką. Nazwiemy normą wektora l (względem metryki d) liczbę $d(l, 0)$, którą będziemy oznaczać przez $\|l\|$. Z aksjomatów metryki (p. 2) i warunków a), b) wynikają następujące własności normy:

$$\|0\| = 0, \quad \|l\| > 0, \qquad \text{jeśli } l \neq 0;$$

$$\|al\| = |a| \cdot \|l\| \qquad \text{dla każdego } a \in \boldsymbol{K},\ l \in L,$$

$$\|l_1 + l_2\| \leq \|l_1\| + \|l_2\|, \qquad \text{dla każdych } l_1,\ l_2 \in L.$$

Pierwsze dwie własności są oczywiste, a trzecią sprawdza się następująco: $\|l_1 + l_2\| = d(l_1 + l_2, 0) = d(l_1, -l_2) \leq d(l_1, 0) + d(0, -l_2) = \|l_1\| + \|l_2\|$.

Przestrzeń liniową wyposażoną w funkcję $\| \ \|\colon L \to \boldsymbol{R}$, spełniającą podane powyżej trzy warunki, nazywamy *przestrzenią unormowaną*.

Na odwrót, norma wyznacza metrykę: wprowadzając odległość wzorem $d(l_1, l_2) = \|l_1 - l_2\|$, łatwo sprawdzić, że spełnione są aksjomaty metryki i zależność $d(l, 0) = \|l\|$.

Przestrzeń unormowaną zupełną nazywamy *przestrzenią Banacha*. Przestrzenie $\boldsymbol{R}^n$ i $\boldsymbol{C}^n$ z każdą z norm wyznaczonych przez metryki określone w p. 2 są przestrzeniami Banacha.

Ogólne pojęcie zbieżności w przestrzeni metrycznej wprowadzone w p. 3, w zastosowaniu do przypadku unormowanych przestrzeni liniowych nazywa się *zbieżnością względem normy*. Struktura liniowa pozwala wprowadzić silniejszy warunek zbieżności szeregu niż zbieżność ciągu jego sum częściowych względem normy. A mianowicie powiemy, że szereg $\sum_{i=1}^{\infty} l_i$ jest bezwzględnie zbieżny, gdy zbieżny jest szereg $\sum_{i=1}^{\infty} \|l_i\|$.

5. Norma i wypukłość. Nietrudno opisać wszystkie normy na jednowymiarowej przestrzeni L, bowiem *dowolne dwie różnią się od siebie o czynnik dodatni.*

W istocie, niech $l \in L$ będzie niezerowym wektorem i $\| \ \|_1$, $\| \ \|_2$ dwiema normami na L. Jeśli $\|l\|_1 = c\|l\|_2$, to $\|al\|_1 = |a| \cdot \|l\|_1 = c|a| \cdot \|l\|_2 = \|al\|_2$ dla każdego $a \in K$.

W jednowymiarowej przestrzeni L z dowolną normą *kołami* (odpowiednio *okręgami*) będziemy nazywać koła (odpowiednio sfery) o niezerowym promieniu i środku w zerze. Jak wynika z poprzedzającego rozumowania, *zbiory wszystkich kół i okręgów są jednakowe dla każdej z norm*. Zamiast zadawać normę można wyróżnić odpowiadające jej koło jednostkowe B albo okrąg jednostkowy S: znając B, otrzymujemy okrąg S jako granicę B, a znając S, wyznaczamy B jako zbiór $\{al \mid l \in S, |a| \leq 1\}$. Zaobserwujmy, że dla $\boldsymbol{K} = \boldsymbol{R}$ koła są odcinkami o środku w zerze, a okręgi to pary punktów symetrycznych względem zera.

Dla przeniesienia tego opisu na przestrzenie dowolnego wymiaru posłużymy się pojęciem wypukłości. Podzbiór $E \subset L$ nazywamy *wypukłym*, jeśli dla dowolnych dwóch wektorów $l_1, l_2 \in E$ i dla dowolnej liczby $0 \leq a \leq 1$ wektor $al_1 + (1-a)l_2$ należy do E. To określenie jest zgodne ze zwyczajnym pojęciem wypukłości w $\boldsymbol{R}^2$ i $\boldsymbol{R}^3$: wraz z dowolnymi dwoma punktami („końcami wektorów l_1 i l_2") zbiór E ma zawierać cały łączący je odcinek („końce wektorów $al_1 + (1-a)l_2$").

Niech $\| \ \|$ oznacza dowolną normę na L. Oznaczmy $B = \{l \in L \mid \|l\| \leq 1\}$, $S = \{l \in L \mid \|l\| = 1\}$. Obcięcie $\| \ \|$ na dowolną podprzestrzeń liniową $L_0 \subset L$ indukuje normę na L_0. Wnioskujemy stąd, że dla dowolnej *jednowymiarowej* podprzestrzeni $L_0 \subset L$ zbiór $L_0 \cap B$ jest kołem w L_0, a zbiór $L_0 \cap S$ jest okręgiem — w znaczeniu poprzedzającej definicji. Poza tym z nierówności trójkąta wynika, że jeśli $l_1, l_2 \in B$, $0 \leq a \leq 1$, to

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

tj. $al_1 + (1-a)l_2 \in B$, a więc B jest zbiorem wypukłym.

Spełnione jest także twierdzenie odwrotne.

6. Twierdzenie. *Niech $S \subset L$ będzie zbiorem spełniającym warunki:*

a) *przecięcie $S \cap L_0$ z dowolną jednowymiarową podprzestrzenią L_0 jest okręgiem,*

b) *zbiór $B = \{al \mid |a| \leq 1, l \in S\}$ jest wypukły.*

Wtedy istnieje jedyna norma $\| \ \|$ na L, dla której B jest kulą jednostkową, a S — sferą jednostkową.

Dowód. Oznaczmy przez $\| \ \|: L \to \boldsymbol{R}$ funkcję, która na każdej jednowymiarowej przestrzeni L_0 jest normą o jednostkowym okręgu równym $S \cap L_0$. Jest oczywiste, że istnieje jedyna taka funkcja, potrzebujemy więc tylko sprawdzić, czy spełnia ona nierówność trójkąta. Niech $l_1, l_2 \in L$, $\|l_1\| = N_1$, $\|l_2\| = N$, $N_i \neq 0$. Stosując warunek wypukłości B do wektorów $N^{-1}l_1$, $N_2^{-1}l_2 \in S$, otrzymamy

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

skąd wynika

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

7. Twierdzenie. *Dowolne dwie normy* $\| \|_1$ *i* $\| \|_2$ *na skończenie wymiarowej przestrzeni liniowej* L *są równoważne w tym sensie, że dla pewnych stałych dodatnich* $0 < c \leq c'$ *nierówność*

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

jest spełniona dla każdego $l \in L$. *W szczególności topologie, tj. pojęcia zbieżności, odpowiadające obu normom są jednakowe, a nadto wszystkie skończenie wymiarowe przestrzenie unormowane są przestrzeniami Banacha.*

Dowód. Ustalmy bazę w L i rozważmy naturalną normę $\|\vec{x}\| = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$, gdzie przez x_i oznaczyliśmy współrzędne w tej bazie. Wystarczy sprawdzić, że dowolna norma $\| \|_1$ jest równoważna z naturalną. Ograniczenie tej normy do sfery jednostkowej względem $\| \|$ jest ciągłą funkcją współrzędnych $\vec{x}$, przyjmującą tylko dodatnie wartości (ciągłość wynika z nierówności trójkąta). Zatem jest ona oddzielona od 0 stałą $c > 0$ i ograniczona z góry stałą $c' > 0$ na podstawie twierdzenia Bolzano–Weierstrassa (sfera jednostkowa jest zbiorem ograniczonym i domkniętym względem $\| \|$). Z nierówności $c \leq \|l\|_1 \leq c'$ spełnionej dla wszystkich $l \in S$ wynika nierówność $c\|l\| \leq \|l\|_1 \leq c'\|l\|$ dla wszystkich $l \in L$. Ponieważ L jest zupełna w topologii zadanej przez normę $\| \|$, a pojęcia zbieżności względem równoważnych norm są identyczne, L jest zupełna względem każdej z tych norm.

8. Norma odwzorowania liniowego. Niech L, M będą unormowanymi przestrzeniami liniowymi nad tym samym ciałem $\boldsymbol{R}$ albo $\boldsymbol{C}$. O odwzorowaniu liniowym $f: L \to M$ powiemy, że jest *ograniczone*, jeśli istnieje taka liczba rzeczywista $N \geq 0$, że dla każdego $l \in L$ spełniona jest nierówność $\|f(l)\| \leq N\|l\|$ (po lewej norma w M, po prawej w L). Oznaczymy przez $\mathcal{L}^1(L, M)$ zbiór wszystkich odwzorowań ograniczonych $L \longrightarrow M$ i dla każdego $f \in \mathcal{L}^1(L, M)$ przez $\|f\|$ będziemy oznaczać kres dolny wszystkich tych N, dla których zachodzi nierówność $\|f(l)\| \leq N\|l\|$, $l \in L$.

9. Twierdzenie. a) $\mathcal{L}^1(L, M)$ *jest przestrzenią unormowaną względem funkcji* $\|f\|$, *nazywanej normą indukowaną.*

b) *Jeśli* L *jest skończenie wymiarowa, to* $\mathcal{L}^1(L, M) = \mathcal{L}(L, M)$, *tj. każde odwzorowanie liniowe jest ograniczone.*

Dowód. a) Niech $f, g \in \mathcal{L}^1(L, M)$. Jeśli $\|f(l)\| \leq N_1\|l\|$ i $\|g(l)\| \leq N_2\|l\|$ dla każdego l, to

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

Zatem $f + g$ i af są ograniczone, a przechodząc do kresów dolnych, otrzymujemy

$$\|f+g\| \leq \|f+g\|, \quad \|af\| = |a|\|f\|.$$

Jeśli $\|f\| = 0$, to dla dowolnego $\varepsilon > 0$ $\|f(l)\| \leq \varepsilon\|l\|$, a to oznacza, że $\|f(l)\| = 0$, więc $f = 0$.

b) Na sferze jednostkowej w L odwzorowanie $l \mapsto \|f(l)\|$ jest funkcją ciągłą. Ponieważ sfera jest zbiorem domkniętym i ograniczonym, więc ta funkcja jest ograniczona i przyjmuje wartość maksymalną. Dlatego na sferze $\|f(l)\| \leq N$, skąd wynika $\|f(l)\| \leq N\|l\|$ dla wszystkich $l \in L$.

W trakcie dowodu wykazaliśmy też, że

$$\|f\| = \max\{\|f(l)\| \mid l \in \text{sfera jednostkowa w } L\}.$$

10. Przykłady. a) W skończenie wymiarowej przestrzeni L ciąg wektorów $l_1, \ldots, l_n, \ldots$ zbiega do wektora l wtedy i tylko wtedy, gdy w pewnej (a więc także w każdej) bazie ciąg i-tych współrzędnych wektorów l_j zbiega do i-tej współrzędnej wektora l, a inaczej mówiąc, gdy $f(l_1), \ldots, f(l_n), \ldots$ jest zbieżny dla każdego funkcjonału $f \in L^*$. Tego ostatniego warunku można używać także i w odniesieniu do nieskończenie wymiarowych przestrzeni, postulując zbieżność ciągu $f(l_i)$ tylko dla funkcjonałów ograniczonych f. Jednakże na ogół prowadzi to do innej topologii na L, nazywanej *topologią słabą*.

b) Niech L oznacza przestrzeń rzeczywistych funkcji różniczkowalnych na $[0,1]$ wyposażoną w normę $\|f\| = \left(\int_0^1 f(t)^2\,dt\right)^{1/2}$. Wtedy operator mnożenia przez t jest ograniczony, gdyż $\int_0^1 t^2 f(t)^2\,dt \leq \int_0^1 f(t)^2\,dt$, a operator $\frac{d}{dt}$ jest nieograniczony. Istotnie, dla dowolnego całkowitego $n \geq 0$ funkcja $\sqrt{2n+1}\,t^n$ należy do sfery jednostkowej, a norma jej pochodnej $n\sqrt{\frac{2n+1}{2n-1}} \to \infty$ dla $n \to \infty$.

11. Twierdzenie. *Niech $L \xrightarrow{f} M \xrightarrow{g} N$ będą ograniczonymi odwzorowaniami liniowymi przestrzeni unormowanych. Wtedy ich złożenie jest odwzorowaniem ograniczonym oraz*

$$\|g \circ f\| \leq \|g\| \cdot \|f\|.$$

Dowód. Jeśli $\|f(l)\| \leq N_1\|l\|$ i $\|g(m)\| \leq N_2\|m\|$ dla wszystkich $l \in L$, $m \in M$, to

$$\|g \circ f(l)\| \leq N_2\|f(l)\| \leq N_2 N_1\|l\|,$$

skąd po przejściu do kresów dolnych otrzymujemy tezę twierdzenia.

Ćwiczenia

1. Wyznaczyć normy na $\boldsymbol{R}^2$, dla których sferami jednostkowymi są odpowiednio następujące zbiory:

a) $x^2 + y^2 \leq 1$;

b) $x^2 + y^2 \leq r^2$;

c) kwadrat o wierzchołkach $(\pm 1, \pm 1)$;

d) kwadrat o wierzchołkach $(0, \pm 1)$, $(\pm 1, 0)$.

2. Niech $f(x) \geq 0$ będzie dwukrotnie różniczkowalną funkcją rzeczywistą na $[a,b] \subset \boldsymbol{R}$ i $f''(x) \leq 0$. Dowieść, że zbiór $\{(x,y) \mid a \leq x \leq b, 0 \leq y \leq f(x)\} \subset \boldsymbol{R}^2$ jest wypukły.

3. Korzystając z rezultatu ćwiczenia 2, wykazać, że przy $p > 1$ zbiór $\{(x,y) \mid |x|^p + |y|^p \leq 1\} \subset \boldsymbol{R}^2$ jest kulą jednostkową dla pewnej normy. Znaleźć tę normę i wykazać nierówność Minkowskiego

$$(|x_1 + y_1|^p + |x_2 + y_2|^p)^{1/p} \leq (|x_1|^p + |y_1|^p)^{1/p} + (|x_2|^p + |y_2|^p)^{1/p}.$$

4. Uogólnić rezultat ćwiczenia 3 na przypadek $\boldsymbol{R}^n$.

5. Niech B oznacza sferę jednostkową względem danej normy w L, B^* — sferę jednostkową dla indukowanej normy w $L^* = \mathcal{L}(L, \boldsymbol{K})$, $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$. Podać jawny opis B^* oraz obliczyć B^* dla norm z ćwiczeń 1 i 3.

§ 11. Funkcje operatorów liniowych

1. W paragrafach 8 i 9 zdefiniowaliśmy operatory postaci $Q(f)$, gdzie $f: L \to L$ jest operatorem liniowym, a Q — wielomianem o współczynnikach z ciała podstawowego $\boldsymbol{K}$. Jeśli $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$, przestrzeń jest unormowana i operator f jest ograniczony, to $Q(f)$ można określić dla szerszej klasy funkcji Q za pomocą przejścia granicznego.

Ograniczymy się tutaj do rozpatrzenia funkcji holomorficznych Q, zadanych przez szereg potęgowy o niezerowym promieniu zbieżności: $Q(t) = \sum_{i=1}^{\infty} a_i t^i$. Określmy $Q(f) = \sum_{i=1}^{\infty} a_i f^i$, jeśli ten szereg operatorowy jest zbieżny bezwzględnie, tzn. jeśli jest zbieżny szereg $\sum_{i=1}^{\infty} a\|f\|^i$. (W najbardziej interesującym dla nas przypadku skończenie wymiarowej przestrzeni L, $\mathcal{L}^1(L,L) = \mathcal{L}(L,L)$ i przestrzeń wszystkich operatorów liniowych jest skończenie wymiarową przestrzenią Banacha; por. § 10, część b) twierdzenia p. 9.).

2. Przykłady. a) Niech f będzie operatorem nilpotentnym. Wtedy $\|f^i\| = 0$ dla dostatecznie dużych i, więc szereg $Q(f)$ jest bezwzględnie zbieżny, a w istocie jest on równy jednej ze swoich sum częściowych.

b) Niech $\|f\| < 1$. Szereg $\sum_{i=0}^{\infty} f^i$ jest bezwzględnie zbieżny oraz

$$(\mathrm{id} - f) \sum_{i=0}^{\infty} f^i = \Big(\sum_{i=0}^{\infty} f^i\Big)(\mathrm{id} - f) = \mathrm{id}.$$

Rzeczywiście,

$$(\mathrm{id} - f) \sum_{i=0}^{N} f^i = \mathrm{id} - f^{N+1} = \Big(\sum_{i=0}^{N} f^i\Big)(\mathrm{id} - f)$$

i przejście do granicy przy $N \to \infty$ kończy dowód. W szczególności, jeśli $\|f\| < 1$, to operator $\mathrm{id} - f$ jest odwracalny.

c) Nazwiemy *operatorową funkcją wykładniczą* funkcję, przyporządkowującą operatorowi ograniczonemu f sumę szeregu

$$e^f = \exp(f) = \sum_{n=0}^{\infty} \frac{1}{n!} f^n.$$

Ponieważ $\|f^n\| \leq \|f\|^n$ (por. twierdzenie § 10, p. 11), a szereg liczbowy określający funkcję wykładniczą jest zbieżny jednostajnie na każdym zbiorze ograniczonym, więc operator $\exp(f)$ jest określony dla każdego operatora ograniczonego f i jest funkcją ciągłą swojego argumentu.

Na przykład, szereg Taylora $\sum_{i=0}^{\infty} \frac{(\Delta t)^i}{i!} \varphi^{(i)}(t)$ przedstawiający $\varphi(t + \Delta t)$ można formalnie zapisać w postaci $\exp\left(\Delta t \frac{d}{dt}\right)\varphi$. Oczywiście, aby nadać sens tej równości, należy wyposażyć przestrzeń funkcji nieskończenie różniczkowalnych φ w normę i wykazać zbieżność względem tej normy.

Jako szczególny przypadek mamy $\exp(a\,\mathrm{id}) = e^a\,\mathrm{id}$ (tutaj a jest skalarem), a także $\exp(\mathrm{diag}\,(a_1, \ldots, a_n)) = \mathrm{diag}\,(\exp a_1, \ldots, \exp a_n)$. Fundamantelna własność liczbowej funkcji wykładniczej, tj. $e^a e^b = e^{a+b}$, w ogólności nie jest prawdziwa dla operatorowej funkcji wykładniczej. Jednakże w pewnej ważnej, choć szczególnej sytuacji jest ona spełniona.

3. Twierdzenie. *Jeśli operatory f, $g: L \to L$ są przemienne, tj. $fg = gf$, to*

$$\exp(f) \circ \exp(g) = \exp(f + g).$$

Dowód. Stosując wzór dwumianowy Newtona i korzystając z możliwości przestawiania wyrazów szeregu bezwzględnie zbieżnego otrzymujemy

$$\begin{aligned}
\exp(f) \circ \exp(g) &= \left(\sum_{i\geq 0} \frac{1}{i!} f^i\right)\left(\sum_{k\geq 0} \frac{1}{k!} g^k\right) = \sum_{i,k\geq 0} \frac{1}{i!k!} f^i g^k = \\
&= \sum_{m\geq 0} \sum_{i=0}^{m} \frac{1}{i!(m-i)!} f^i g^{m-i} = \sum_{m\geq 0} \frac{1}{m!} \sum_{i=0}^{m} \frac{m!}{i!(m-i)!} f^i g^{m-i} = \\
&= \sum_{m\geq 0} \frac{1}{m!} (f+g)^m = \exp(f+g).
\end{aligned}$$

Przemienność g i f została wykorzystana poprzez zastosowanie wzoru dwumianowego do obliczenia $(f + g)^m$.

4. Wniosek. *Niech $f: L \to L$ będzie operatorem ograniczonym. Wtedy odwzorowanie $\boldsymbol{R} \to \mathcal{L}^1(L, L) : t \mapsto \exp(tf)$ jest homomorfizmem grupy $\boldsymbol{R}$ na pewną podgrupę multyplikatywnej grupy operatorów odwracalnych w $\mathcal{L}^1(L, L)$.*

Zbiór operatorów $\{\exp tf \mid t \in \boldsymbol{R}\}$ nazywamy *jednoparametrową podgrupą operatorów.*

5. Widmo. Niech f będzie jakimś operatorem w skończenie wymiarowej przestrzeni, $Q(t)$ — szeregiem potęgowym, dla którego szereg $Q(f)$ jest zbieżny bezwzględnie. Łatwo spostrzec, że jeśli $Q(t)$ jest wielomianem, to w bazie jordanowskiej dla f macierz $Q(f)$ jest macierzą górnotrójkątną z liczbami $Q(\lambda_i)$ na przekątnej, gdzie λ_i oznaczają wartości własne f. Stosując tę obserwację do sum częściowych szeregu Q i przechodząc do granicy, wnioskujemy, że to samo jest prawdziwe i dla dowolnego szeregu $Q(t)$. W szczególności, jeśli $S(f)$ jest widmem f, to $(Q(f)) = Q(S(f)) = Q(\lambda) \mid \lambda \in S(f)$. Co więcej, jeśli uwzględnić krotności pierwiastków wielomianu charakterystycznego λ_i, to $Q(S(f))$ będzie widmem $Q(f)$ z uwzględnieniem prawidłowych krotności. W szczególności,

$$\det(\exp f) = \prod_{i=1}^{n} \exp \lambda_i = \exp\left(\sum_{i=1}^{n} \lambda_i\right) = \exp \operatorname{Tr} f.$$

Przechodząc do formalizmu macierzowego, odnotujemy jeszcze dwie proste własności, których dowodzi się w podobny sposób:

a) $Q(A^t) = Q(A)^t$;

b) $Q(\overline{A}) = \overline{Q(A)}$, gdzie kreska nad symbolem oznacza sprzężenie zespolone i gdzie założyliśmy, że współczynniki szeregu Q są rzeczywiste.

Używając notacji wprowadzonej w § 4 i korzystając z powyższych własności odwzorowania exp, udowodnimy następujące twierdzenie odnoszące się do teorii klasycznych grup Liego (tutaj $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$).

6. Twierdzenie. *Odwzorowanie* exp *przeprowadza* $\mathfrak{gl}(n, \boldsymbol{K}), \mathfrak{sl}(n, \boldsymbol{K}), \mathfrak{o}(n, \boldsymbol{K}), \mathfrak{u}(n), \mathfrak{su}(n)$ *odpowiednio w* $\boldsymbol{GL}(n, \boldsymbol{K})$, $\boldsymbol{SL}(n, \boldsymbol{K})$, $\boldsymbol{SO}(n, \boldsymbol{K})$, $\boldsymbol{U}(n)$, $\boldsymbol{SU}(n)$.

Dowód. Przestrzeń $\mathfrak{gl}(n, \boldsymbol{K})$ jest odwzorowywana w $\boldsymbol{GL}(n, \boldsymbol{K})$, gdyż zgodnie z wnioskiem p. 4 macierze postaci $\exp A$ są odwracalne. Jeśli $\operatorname{Tr} A = 0$, to $\det \exp A = 1$, jak to pokazaliśmy w poprzedzającym punkcie. Z warunku $A + A^t = 0$ wynika, że $(\exp A)(\exp A)^t = 1$, a z warunku $A + \overline{A}^t = 0$ wynika, że $(\exp -A)(\overline{\exp A})^t = 1$. To kończy dowód.

7. Uwaga. We wszystkich wymienionych przypadkach obraz exp zawiera pewne otoczenie jedności odpowiedniej grupy. Dla dowodu można wprowadzić funkcję logarytm dla operatorów f spełniających warunek $\|f - \mathrm{id}\| < 1$, używając znanego wzoru $\log f = \sum_{n \geq 1} (-1)^{n-1} \frac{(f - \mathrm{id})^n}{n}$ i dowieść, że $f = \exp(\log f)$.

Jednakże obrazem odwzorowania exp nie jest na ogół cała grupa, tzn. nie jest ono surjektywne. Na przykład, nie istnieje macierz $A \in \mathfrak{sl}(2, \boldsymbol{C})$, taka że $\exp A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \in \boldsymbol{SL}(2, \boldsymbol{C})$. Istotnie, A nie może być diagonalizowalna, gdyż wówczas także i $\exp A$ byłaby diagonalizowalna. A zatem wartości własne A muszą być równe, a że ślad A jest równy zeru, więc muszą być równe zeru. Ale wówczas wartości własne $\exp A$ byłyby równe 1, podczas gdy wartości własne $\begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$ są równe -1.

§ 12. Rozszerzenie zespolone i zwężenie rzeczywiste

1. W paragrafach 8 i 9 przekonaliśmy się, że w przypadku ciała algebraicznie domkniętego struktura geometryczna operatorów liniowych jest odzwierciedlona w przejrzystej postaci kanonicznej ich macierzy. Dlatego nawet wtedy, gdy rozważamy przypadek rzeczywistych skalarów, bywa dogodnie wprowadzić liczby zespolone. W tym paragrafie rozpatrzymy dwie podstawowe operacje, rozszerzenie i zwężenie ciała skalarów, w zastosowaniu do przestrzeni i operatorów liniowych. Najwięcej uwagi poświęcimy na przejście od $\boldsymbol{R}$ do $\boldsymbol{C}$ (*kompleksyfikacja* inaczej *rozszerzenie zespolone*) oraz przejście od $\boldsymbol{C}$ do $\boldsymbol{R}$ (*realifikacja*, czyli *zwężenie rzeczywiste*) i tylko krótko wspomnimy o ogólniejszej sytuacji.

2. Zwężenie rzeczywiste. Niech L będzie przestrzenią liniową nad $\boldsymbol{C}$. Pomińmy możliwość mnożenia wektorów z L przez dowolne liczby zespolone i ograniczmy się tylko do mnożenia przez elementy $\boldsymbol{R}$. Oczywiście otrzymamy w ten sposób przestrzeń liniową nad $\boldsymbol{R}$, którą będziemy oznaczać przez $L_{\boldsymbol{R}}$ i nazywać *realifikacją* lub *zwężeniem rzeczywistym* L.

Niech L, M będą dwiema przestrzeniami liniowymi nad $\boldsymbol{C}$, $f\colon L \to M$ — odwzorowaniem liniowym. Oczywiście, f rozpatrywane jako odwzorowanie $L_{\boldsymbol{R}} \to M_{\boldsymbol{R}}$ jest w dalszym ciągu liniowe. Będziemy je oznaczać $f_{\boldsymbol{R}}$ i nazywać *zwężeniem rzeczywistym* f. Jest jasne, że mamy $\mathrm{id}_{\boldsymbol{R}} = \mathrm{id}$, $(f \circ g)_{\boldsymbol{R}} = f_{\boldsymbol{R}} \circ g_{\boldsymbol{R}}$; $(af + bg)_{\boldsymbol{R}} = af_{\boldsymbol{R}} + bg_{\boldsymbol{R}}$, jeśli $a,\ b \in \boldsymbol{R}$.

3. Twierdzenie. a) *Niech* $\{e_1, \ldots, e_m\}$ *będzie bazą przestrzeni* L *nad* $\boldsymbol{C}$. *Wtedy* $\{e_1, \ldots, e_m, ie_1, \ldots, ie_m\}$ *jest bazą przestrzeni* $L_{\boldsymbol{R}}$ *nad* $\boldsymbol{R}$ *i w szczególności,*

$$\dim_{\boldsymbol{R}} L_{\boldsymbol{R}} = 2 \dim_{\boldsymbol{C}} L.$$

b) *Niech* $A = B + iC$ *będzie macierzą odwzorowania liniowego* $f\colon L \to M$ *w bazach* $\{e_1, \ldots, e_m\}$, $\{e'_1, \ldots, e'_n\}$ *nad* $\boldsymbol{C}$, *gdzie* B *oraz* C *są macierzami rzeczywistymi. Wtedy macierzą odwzorowania liniowego* $f_{\boldsymbol{R}}\colon L_{\boldsymbol{R}} \to M_{\boldsymbol{R}}$ *w bazach* $\{e_1, \ldots, e_m, ie_1, \ldots, ie_m\}$, $\{e'_1, \ldots, e'_n, ie'_1, \ldots, ie'_n\}$ *jest*

$$\begin{pmatrix} B & -C \\ C & B \end{pmatrix}.$$

Dowód. a) Biorąc dowolny wektor $l \in L$, możemy zapisać

$$l = \sum_{k=1}^{m} a_k e_k = \sum_{k=1}^{m} (b_k + ic_k) e_k = \sum_{k=1}^{m} b_k e_k + \sum_{k=1}^{m} c_k (ie_k),$$

gdzie b_k, c_k są odpowiednio rzeczywistą i urojoną częścią a_k. Dlatego zbiór $\{e_k, ie_k\}$ rozpina $L_{\boldsymbol{R}}$. Jeśli $\sum_{k=1}^{m} b_k e_k + \sum_{k=1}^{m} c_k (ie_k) = 0$, gdzie b_k, $c_k \in \boldsymbol{R}$, to $b_k + ic_k = 0$ na mocy liniowej niezależności $\{e_1, \ldots, e_m\}$ nad $\boldsymbol{C}$, skąd wynika $b_k = c_k = 0$ dla każdego k.

b) Zgodnie z definicją A możemy napisać

$$f(e_1,\dots,e_m) = (e'_1,\dots,e'_n)(B+iC),$$

skąd, na mocy liniowości nad $\boldsymbol{C}$,

$$f(ie_1,\dots,ie_m) = (ie'_1,\dots,ie'_n)(-C+iB).$$

Dlatego mamy

$$(f(e_1),\dots,f(e_m),f(ie_1),\dots,f(ie_m)) = (e'_1,\dots,e'_n,ie'_1,\dots,ie'_n)\begin{pmatrix} B & -C \\ C & B \end{pmatrix},$$

co kończy dowód.

4. Wniosek. *Niech $f\colon L \to L$ będzie operatorem liniowym w skończenie wymiarowej przestrzeni L. Wtedy* $\det(f_{\boldsymbol{R}}) = |\det f|^2$.

Dowód. Niech f będzie przedstawione w bazie $\{e_1,\dots,e_m\}$ przez macierz $B+iC$ (gdzie B i C są rzeczywiste). Wtedy, stosując elementarne operacje (od razu w odniesieniu do bloków macierzy) najpierw na wierszach, a potem na kolumnach otrzymamy

$$\begin{aligned}\det(f_{\boldsymbol{R}}) &= \det\begin{pmatrix} B & -C \\ C & B \end{pmatrix} = \det\begin{pmatrix} B+iC & -C+iB \\ C & B \end{pmatrix} = \\ &= \det\begin{pmatrix} B+iC & 0 \\ C & B-iC \end{pmatrix} = \det(B+iC)\cdot\det(B-iC) = \\ &= \det f\cdot\overline{\det f} = |\det f|^2.\end{aligned}$$

5. Zwężenie ciała skalarów: przypadek ogólny. Jest z pewnością jasne, jak możemy uogólnić rozważania p. 2. Niech $\boldsymbol{K}$ będzie dowolnym ciałem, $\mathcal{K}$ — jego podciałem, a L — przestrzenią liniową nad $\boldsymbol{K}$. Pomijając możliwość mnożenia przez dowolne elementy ciała $\boldsymbol{K}$ i ograniczając je tylko do elementów z $\mathcal{K}$, otrzymamy przestrzeń liniową $L_{\mathcal{K}}$ nad ciałem $\mathcal{K}$. Analogicznie, odwzorowanie $f\colon L \to M$ liniowe nad $\boldsymbol{K}$ przechodzi w odwzorowanie liniowe $f\colon L_{\mathcal{K}} \to M_{\mathcal{K}}$. Jednym z używanych określeń tej operacji jest *zwężenie ciała skalarów* (z $\boldsymbol{K}$ do $\mathcal{K}$). Jest jasne, że mamy $\mathrm{id}_{\mathcal{K}} = \mathrm{id}$, $(f\circ g)_{\mathcal{K}} = f_{\mathcal{K}}\circ g_{\mathcal{K}}$; $(af+bg)_{\mathcal{K}} = af_{\mathcal{K}} + bg_{\mathcal{K}}$, jeśli $a, b \in \mathcal{K}$.

Również i ciało $\boldsymbol{K}$ można uważać za przestrzeń liniową nad $\mathcal{K}$. Jeśli ta przestrzeń ma wymiar skończony, to wymiary $\dim_{\boldsymbol{K}} L$ i $\dim_{\mathcal{K}} L_{\mathcal{K}}$ są związane wzorem

$$\dim_{\mathcal{K}} L_{\mathcal{K}} = \dim_{\mathcal{K}} \boldsymbol{K}\cdot\dim_{\boldsymbol{K}} L.$$

Dla dowodu wystarczy sprawdzić, że jeśli $\{e_1,\dots,e_n\}$ jest bazą L nad $\boldsymbol{K}$, a $\{b_1,\dots,b_m\}$ bazą $\boldsymbol{K}$ nad $\mathcal{K}$, to $\{b_1e_1,\dots,b_1e_n;\dots;b_me_1,\dots,b_me_n\}$ stanowi bazę L nad $\mathcal{K}$.

6. Struktura zespolona na rzeczywistej przestrzeni liniowej. Niech L będzie zespoloną przestrzenią liniową, a $L_{\boldsymbol{R}}$ — jej zwężeniem rzeczywistym. Dla odtworzenia w pełni mnożenia przez liczby zespolone w $L_{\boldsymbol{R}}$ wystarczy znać operator $J: L_{\boldsymbol{R}} \to L_{\boldsymbol{R}}$ mnożenia przez i: $J(l) = il$. Ten operator jest oczywiście liniowy nad $\boldsymbol{R}$ i spełnia warunek $J^2 = -\mathrm{id}$; jeśli jest on wyznaczony, to dla dowolnej liczby zespolonej $a + ib$, $a, b \in \boldsymbol{R}$ możemy napisać

$$(a + ib)l = al + bJ(l).$$

Ta obserwacja prowadzi nas do następującego ważnego pojęcia.

7. Definicja. Niech L będzie rzeczywistą przestrzenią liniową. Wybór operatora liniowego $J: L \to L$ spełniającego warunek $J^2 = -\mathrm{id}$ nazywamy *zadaniem w L struktury zespolonej.*

Opisana wyżej struktura zespolona w $L_{\boldsymbol{R}}$ jest nazywana *kanoniczną.* Motywacji dla wprowadzenia takiej definicji dostarcza następujące twierdzenie:

8. Twierdzenie. *Niech (L, J) będzie rzeczywistą przestrzenią liniową ze strukturą zespoloną. Jeśli zadać na L operację mnożenia przez liczby zespolone z $\boldsymbol{C}$ za pomocą wzoru*

$$(a + bi)l = al + bJ(l),$$

to w ten sposób zostaje określona w (zbiorze) L struktura zespolonej przestrzeni liniowej $\tilde{L}$, dla której $\tilde{L}_{\boldsymbol{R}} = L$.

Dowód. Oba aksjomaty rozdzielności wynikają łatwo z liniowości J i wzoru na dodawanie liczb zespolonych. Sprawdzimy aksjomat rozdzielności mnożenia:

$$\begin{aligned}(a + bi)[(c + di)l] &= (a + bi)[cl + dJ(l)] = a[cl + dJ(l)] + bJ[cl + dJ(l)] = \\ &= acl + adJ(l) + bcJ(l) - bdl = (ac - bd)l + (ad + bc)J(l) = \\ &= [ac - bd + (ad + bc)i]l = [(a + bi)(c + di)]l.\end{aligned}$$

Pozostałe aksjomaty są automatycznie spełnione, ponieważ L i $\tilde{L}$ są identyczne jako grupy względem dodawania.

9. Wniosek. *Jeśli (L, J) jest skończenie wymiarową przestrzenią ze strukturą zespoloną, to $\dim_{\boldsymbol{R}} L = 2n$ jest liczbą parzystą, a macierzą J w odpowiednio dobranej bazie jest macierz*

$$\begin{pmatrix} 0 & -E_n \\ E_n & 0 \end{pmatrix}.$$

Dowód. Istotnie, $\dim_{\boldsymbol{R}} L = 2\dim_{\boldsymbol{C}} \tilde{L}$ na mocy twierdzenia p. 7 i punktu a) twierdzenia p. 3 (skończoność wymiaru $\tilde{L}$ wynika stąd, że dowolna baza nad $\boldsymbol{R}$ przestrzeni L rozpina $\tilde{L}$ nad $\boldsymbol{C}$). Wybierzmy teraz bazę $\{e_1, \ldots, e_n\}$ przestrzeni $\tilde{L}$

nad C. Macierzą mnożenia przez i jest w tej bazie macierz iE_n, skąd na mocy punktu b) twierdzenia p. 3 wynika, że macierz operatora J w bazie $\{e_1, \ldots, e_n; ie_1, \ldots, ie_n\}$ przestrzeni L ma podaną postać.

10. Uwagi. a) Niech L będzie przestrzenią zespoloną, $g: L \to L_{\boldsymbol{R}}$ — odwzorowaniem $\boldsymbol{R}$-liniowym. Rozważmy problem, kiedy istnieje takie odwzorowanie liniowe $f: L \to L$, że $g = f_{\boldsymbol{R}}$. Oczywiście jest konieczne, aby g było przemienne z operatorem naturalnej struktury zespolonej J w $L_{\boldsymbol{R}}$, gdyż $g(il) = g(Jl) = ig(l) = Jg(l)$ dla wszystkich $l \in L$. Ale jest to także warunek dostateczny, gdyż z niego automatycznie wynika liniowość g nad $\boldsymbol{C}$:

$$g((a+bi)l) = ag(l) + bg(il) = ag(l) + bgJ(l) = ag(l) + bJg(l) =$$
$$= (a + bJ)g(l) = (a + bi)g(l).$$

b) Niech teraz L będzie rzeczywistą przestrzenią parzystego wymiaru, a $f: L \to L$ — $\boldsymbol{R}$-liniowym operatorem. Rozważmy zagadnienie, kiedy można na L wprowadzić taką strukturę zespoloną J, że f będzie zwężeniem rzeczywistym $\boldsymbol{C}$-liniowego odwzorowania $g: \tilde{L} \to \tilde{L}$, gdzie tak jak wyżej $\tilde{L}$ oznacza przestrzeń zespoloną zbudowaną z L za pomocą J. Częściowa odpowiedź odnosząca się do przypadku $\dim_{\boldsymbol{R}} L = 2$ jest następująca: poszukiwana struktura zespolona istnieje, jeśli L nie zawiera wektorów własnych f. Istotnie, wówczas f ma dwie wzajemnie sprzężone wartości własne $\lambda \pm i\mu$, $\lambda, \mu \in \boldsymbol{R}$, $\mu \neq 0$. Zdefiniujmy $J = \mu^{-1}(f - \lambda \mathrm{id})$. Na mocy twierdzenia Hamiltona–Cayleya, $f^2 - 2\lambda f + (\lambda^2 + \mu^2)\,\mathrm{id} = 0$, skąd

$$J^2 = \mu^{-2}(f^2 - 2\lambda f + \lambda^2\,\mathrm{id}) = -\,\mathrm{id}\,.$$

Ponadto J jest przemienne z f, co kończy dowód.

11. Kompleksyfikacja. Rozważmy teraz ustaloną rzeczywistą przestrzeń liniową L i wprowadźmy strukturę zespoloną J na zewnętrznej sumie prostej $L \oplus L$ za pomocą wzoru

$$J(l_1, l_2) = (-l_2, l_1).$$

Jest oczywiste, że $J^2 = -1$. *Kompleksyfikacją przestrzeni* L będziemy nazywać zespoloną przestrzeń $L \widetilde{\oplus} L$ wyznaczoną przez tę strukturę. Tę przestrzeń będziemy oznaczać przez $L^{\boldsymbol{C}}$, a innymi standardowymi oznaczeniami są $\boldsymbol{C} \otimes_{\boldsymbol{R}} L$ lub $\boldsymbol{C} \oplus L$; ich pochodzenie wyjaśni się po wprowadzeniu iloczynów tensorowych przestrzeni liniowych. Utożsamiając L z podprzestrzenią wektorów postaci $(l, 0)$ w $L \oplus L$ i korzystając z tego, że $i(l, 0) = J(l, 0) = (0, l)$, możemy przedstawić dowolny wektor $L^{\boldsymbol{C}}$ w postaci

$$(l_1, l_2) = (l_1, 0) + (0, l_2) = (l_1, 0) + i(l_2, 0) = l_1 + il_2.$$

Innymi słowy, $L^{\boldsymbol{C}} = L \oplus iL$, ale ta suma jest sumą prostą nad $\boldsymbol{R}$, a nie nad $\boldsymbol{C}$!

Dowolna baza L nad $\boldsymbol{R}$ jest bazą $L^{\boldsymbol{C}}$ nad $\boldsymbol{C}$, skąd wynika $\dim_{\boldsymbol{R}} L = \dim_{\boldsymbol{C}} L^{\boldsymbol{C}}$.

Rozważmy teraz odwzorowanie liniowe $f: L \to M$ przestrzeni liniowych nad ciałem $\boldsymbol{R}$. Wówczas odwzorowanie $f^{\boldsymbol{C}}$ (lub $f \otimes \boldsymbol{C}$): $L^{\boldsymbol{C}} \to M^{\boldsymbol{C}}$, określone wzorem

$$f^{\boldsymbol{C}}(l_1, l_2) = (f(l_1), f(l_2)),$$

jest liniowe nad $\boldsymbol{R}$ i komutuje z J, gdyż

$$fJ(l_1, l_2) = f(-l_2, l_1) = (-f(l_2), f(l_1)) = Jf(l_1, l_2).$$

Zatem jest ono $\boldsymbol{C}$-liniowe — nazywamy je *kompleksyfikacją* odwzorowania f. Oczywiście, $\mathrm{id}^{\boldsymbol{C}} = \mathrm{id}$, $(af + bg)^{\boldsymbol{C}} = af^{\boldsymbol{C}} + bg^{\boldsymbol{C}}$ dla $a,b \in \boldsymbol{R}$ i $(fg)^{\boldsymbol{C}} = f^{\boldsymbol{C}} g^{\boldsymbol{C}}$. Traktując bazy L i M jako bazy odpowiednio w $L^{\boldsymbol{C}}$ i $M^{\boldsymbol{C}}$, wnioskujemy, że macierz odwzorowania f względem wyjściowej pary baz jest identyczna z macierzą $f^{\boldsymbol{C}}$ względem tej "nowej" pary baz. W szczególności, (zespolone) wartości własne odwzorowań f i $f^{\boldsymbol{C}}$ i ich formy jordanowskie są identyczne.

Przyjrzyjmy się teraz temu, co dzieje się przy składaniu operacji zwężenia rzeczywistego i rozszerzenia zespolonego w jednej i drugiej z możliwych kolejności.

12. Najpierw kompleksyfikacja, potem realifikacja. Niech L będzie przestrzenią rzeczywistą. Twierdzimy, że istnieje naturalny izomorfizm

$$(L^{\boldsymbol{C}})_{\boldsymbol{R}} \longrightarrow L \oplus L.$$

Rzeczywiście, na mocy konstrukcji, $L^{\boldsymbol{C}}$ jako przestrzeń rzeczywista jest identyczna z $L \oplus L$. Analogicznie, $(f^{\boldsymbol{C}})_{\boldsymbol{R}} \to f \oplus f$ (w znaczeniu powyższego utożsamienia) dla dowolnego $\boldsymbol{R}$-liniowego odwzorowania $f: L \to M$.

Złożenie w odwrotnej kolejności prowadzi do nieco mniej oczywistej odpowiedzi. Wprowadźmy następujące pojęcie.

13. Definicja. Niech L będzie przestrzenią zespoloną. *Sprzężoną przestrzenią zespoloną* $\overline{L}$ nazywamy zbiór L z tą samą strukturą grupy addytywnej, lecz z innym mnożeniem przez skalary z $\boldsymbol{C}$, które chwilowo oznaczymy przez $a * l$:

$$a * l = \overline{a}l \quad \text{dla dowolnych } a \in \boldsymbol{C},\ l \in L.$$

Aksjomaty można sprawdzić bez trudu, jeśli wykorzystać znane własności sprzężenia $\overline{ab} = \overline{a} \cdot \overline{b}$ i $\overline{a+b} = \overline{a} + \overline{b}$.

Analogicznie, jeśli (L, J) jest rzeczywistą przestrzenią ze strukturą zespoloną J, to operator $-J$ także określa strukturę zespoloną, którą nazywa się sprzężoną z wyjściową. W oznaczeniach twierdzenia p. 7, jeśli $\tilde{L}$ jest przestrzenią zespoloną odpowiadającą (L, J), to $\overline{\tilde{L}}$ odpowiada $(L, -J)$.

14. Najpierw realifikacja, potem kompleksyfikacja. Teraz możemy dla dowolnej zespolonej przestrzeni liniowej L zbudować kanoniczny izomorfizm $\boldsymbol{C}$-liniowy

$$f: (L_{\boldsymbol{R}})^{\boldsymbol{C}} \to L \oplus \overline{L}.$$

W tym celu zauważmy, że na $(L_{\boldsymbol{R}})^{\boldsymbol{C}}$ istnieją dwa $\boldsymbol{R}$-liniowe operatory: operator kanonicznej struktury zespolonej $J(l_1, l_2) = (-l_2, l_1)$ i operator mnożenia przez i, odpowiadający wyjściowej strukturze przestrzeni zespolonej w L: $i(l_1, l_2) = (il_1, il_2)$. Ponieważ J komutuje z i, jest on $\boldsymbol{C}$-liniowy względem tej struktury. Z $J^2 = -\mathrm{id}$ wnioskujemy, że ma on wartości własne $\pm i$. Wprowadźmy standardowe oznaczenia dla podprzestrzeni odpowiadających tym wartościom własnym:

$$L^{1,0} = \{(l_1, l_2) \in (L_{\boldsymbol{R}})^{\boldsymbol{C}} \mid J(l_1, l_2) = i(l_1, l_2)\},$$
$$L^{0,1} = \{(l_1, l_2) \in (L_{\boldsymbol{R}})^{\boldsymbol{C}} \mid J(l_1, l_2) = -i(l_1, l_2)\}.$$

Oba zbiory $L^{1,0}$ i $L^{0,1}$ są zespolonymi podprzestrzeniami w $(L_{\boldsymbol{R}})^{\boldsymbol{C}}$ i jest jasne, że są one domknięte względem dodawania i mnożenia przez liczby rzeczywiste, a domknięte względem mnożenia przez liczby zespolone wynika stąd, że J jest przemienne z mnożeniem przez i. Pokażemy teraz, że $(L_{\boldsymbol{R}})^{\boldsymbol{C}} = L^{1,0} \oplus L^{0,1}$, a ponadto, że $L^{1,0}$ jest w naturalny sposób izomorficzne z L, a $L^{0,1}$ jest w naturalny sposób izomorficzne z $\overline{L}$.

Z określenia od razu wynika, że $L^{1,0}$ składa się z wektorów postaci $(l, -il)$, a $L^{0,1}$ z wektorów postaci (m, im). Dla danych $l_1, l_2 \in L$ równanie $(l_1, l_2) = (l, -il) + (m, im)$ o niewiadomych l i m ma jednoznaczne rozwiązanie $l = \frac{1}{2}(l_1 + il_2), m = \frac{1}{2}(l_1 - il_2)$. To dowodzi, że $(L_{\boldsymbol{R}})^{\boldsymbol{C}} = L^{1,0} \oplus L^{0,1}$. Odwzorowania $L \to L^{1,0} : l \mapsto (l, -il)$ oraz $\overline{L} \to L^{0,1} : l \mapsto (l, il)$ są $\boldsymbol{R}$-liniowymi izomorfizmami, a poza tym komutują z działaniem i na L, $\overline{L}$ oraz z działaniem J na $L^{1,0}$, $L^{0,1}$ na mocy definicji. W ten sposób nasza konstrukcja jest zakończona.

15. Półliniowe odwzorowania przestrzeni zespolonych. Niech L, M będą zespolonymi przestrzeniami liniowymi. Odwzorowanie $f: L \to M$ nazywa się *odwzorowaniem półliniowym* (lub *antyliniowym*), jeśli jest ono odwzorowaniem liniowym $f: L \to \overline{M}$. Innymi słowy f jest homomorfizmem grup addytywnych i zachodzi

$$f(al) = \overline{a} f(l)$$

dla każdych $a \in \boldsymbol{C}$, $l \in L$. Szczególne znaczenie półliniowych odwzorowań wyjaśnimy w drugiej części tej książki, w trakcie rozważania zespolonych przestrzeni hermitowskich.

16. Rozszerzenie ciała skalarów: przypadek ogólny. Niech tak jak w p. 4 $\boldsymbol{K}$ będzie dowolnym ciałem, a $\mathcal{K}$ — jego podciałem. Wtedy dla dowolnej przestrzeni liniowej L nad $\mathcal{K}$ można określić przestrzeń liniową $\boldsymbol{K} \otimes_{\mathcal{K}} L$, inaczej $L^{\boldsymbol{K}}$, z ciałem skalarów $\boldsymbol{K}$ i o niezmienionym wymiarze. Przed wprowadzeniem języka iloczynów tensorowych podanie ogólnej definicji $L^{\boldsymbol{K}}$ jest niewygodne, ale dla praktycznych celów może wystarczyć następujący substytut: jeśli $\{e_1, \ldots, e_n\}$ jest bazą L nad $\mathcal{K}$, to $L^{\boldsymbol{K}}$ składa się ze wszystkich formalnych kombinacji liniowych $\left\{\sum_{i=1}^{n} a_i e_i \mid a_i \in \boldsymbol{K}\right\}$, tj. posiada tę samą bazę nad $\boldsymbol{K}$. W szczególności $(\mathcal{K}^n)^{\boldsymbol{K}} = \boldsymbol{K}^n$. Każdemu

K-liniowemu odwzorowaniu $f: L \to M$ można przyporządkować $\boldsymbol{K}$-liniowe odwzorowanie $f^{\boldsymbol{K}}: L^{\boldsymbol{K}} \to M^{\boldsymbol{K}}$: jeśli f jest zadane przez macierz względem pewnych baz L i M, to $f^{\boldsymbol{K}}$ jest zadane przez tę samą macierz.

Na zakończenie podamy jeszcze jedną własność kompleksyfikacji.

17. Stwierdzenie. *Niech $f: L \to L$ będzie operatorem liniowym w rzeczywistej przestrzeni wymiaru ≥ 1. Wówczas f ma podprzestrzeń niezmienniczą wymiaru 1 lub 2.*

Dowód. Jeśli f ma rzeczywistą wartość własną, to podprzestrzeń rozpięta przez odpowiadający jej wektor własny jest niezmiennicza. W przeciwnym przypadku wszystkie wartości własne są zespolone. Wybierzmy jedną z nich $\lambda + i\mu$. Jest ona także wartością własną $f^{\boldsymbol{C}}$ w $L^{\boldsymbol{C}}$. Niech $l_1 + il_2 \in L^{\boldsymbol{C}}$, $l_1, l_2 \in L$, będzie wektorem własnym odpowiadającym tej wartości własnej. Zgodnie z określeniem

$$f^{\boldsymbol{C}}(l_1 + il_2) = f(l_1) + if(l_2) = (\lambda + i\mu)(l_1 + il_2) =$$
$$= (\lambda l_1 - \mu l_2) + i(\mu l_1 + i\lambda l_2).$$

Stąd wnioskujemy, że $f(l_1) = (\lambda l_1 - \mu l_2)$, $f(l_2) = \mu l_1 + \lambda l_2$, tj. powłoka liniowa $\{l_1, l_2\}$ w L jest f-niezmiennicza.

§ 13. Język teorii kategorii

1. Określenie kategorii. Kategorię C stanowią następujące dane:

a) Klasa (albo zbiór) $\operatorname{Ob} C$, którego elementy nazywają się *obiektami* kategorii.

b) Klasa (albo zbiór) $\operatorname{Mor} C$, którego elementy nazywają się *morfizmami kategorii* lub *strzałkami*.

c) Każdej uporządkowanej parze obiektów $X, Y \in \operatorname{Ob} C$ jest przyporządkowany zbiór $\operatorname{Hom}_C(X, Y) \subset \operatorname{Mor} C$, którego elementy nazywają się *morfizmami z X w Y* i które są oznaczane $X \to Y$ lub $f: X \to Y$, lub także $X \xrightarrow{f} Y$.

d) Dla każdej uporządkowanej trójki obiektów $X, Y, Z \in \operatorname{Ob} C$ zadane jest odwzorowanie

$$\operatorname{Hom}_C(X, Y) \times \operatorname{Hom}_C(Y, Z) \longrightarrow \operatorname{Hom}_C(X, Z),$$

przyporządkowujące każdej parze morfizmów (f, g) morfizm gf albo $g \circ f$, nazywany ich *superpozycją* lub *złożeniem*.

Te dane związane są następującymi warunkami:

e) $\operatorname{Mor} C$ jest sumą rozłączną $\bigcup \operatorname{Hom}_C(X, Y)$, przy czym sumuje się po zbiorze wszystkich par uporządkowanych $X, Y \in \operatorname{Ob} C$. Innymi słowy, dla każdego morfizmu f istnieją jednoznacznie wyznaczone obiekty X, Y, takie że $f \in \operatorname{Hom}_C(X, Y)$: *początek X i koniec Y* strzałki f.

f) Składanie morfizmów jest łączne.

g) Dla każdego obiektu X istnieje morfizm identycznościowy $\mathrm{id}_X \in \mathrm{Hom}_C(X, X)$, to znaczy taki, że $\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$, jeśli te złożenia są określone. Nietrudno dostrzec, że taki morfizm jest tylko jeden: jeśli id'_X jest innym morfizmem o tych własnościach, to $\mathrm{id}'_X = \mathrm{id}'_X \circ \mathrm{id}_X = \mathrm{id}_X$.

Morfizm $f: X \to Y$ nazywa sie izomorfizmem, jeśli istnieje taki morfizm $g: Y \to X$, że $g \circ f = \mathrm{id}_X$, $f \circ g = \mathrm{id}_Y$.

2. Przykłady. a) *Kategoria zbiorów Set.* Jej obiektami są zbiory, a morfizmami są odwzorowania zbiorów.

b) *Kategoria $\mathcal{L}in_K$ przestrzeni liniowych nad ciałem K.* Jej obiektami są przestrzenie liniowe, a morfizmami odwzorowania liniowe.

c) *Kategoria grup.*

d) *Kategoria grup abelowych.*

Rozróżnienie pomiędzy klasą a zbiorem jest dyskutowane w aksjomatycznej teorii mnogości i jest konieczne dla ominięcia znanego paradoksu Russella. Nie każda kolekcja elementów tworzy zbiór, gdyż pojęcie „zbiór wszystkich zbiorów, które nie zawierają samego siebie jako elementu" jest samo w sobie sprzeczne. W aksjomatyce Gödela–Bernaysa takie kolekcje zbiorów nazywa się *klasami.* Technika teorii kategorii wymaga grupowania obiektów w kolekcje, leżące w niebezpiecznej bliskości do takich paradoksalnych sytuacji. Jednakże my nie będziemy poświęcać uwagi takim subtelnościom.

3. Diagramy. Ponieważ aksjomaty kategorii nie mówią nic o teoriomnogościowej strukturze obiektów, na ogół nie można odwoływać się do pojęcia „elementu" obiektu. Wszystkie podstawowe konstrukcje ogólnej teorii kategorii i ich zastosowania do konkretnych kategorii przeważnie formułuje się w terminach morfizmów i ich superpozycji. Wygodnym językiem dla takich sformułowań jest język diagramów. Na przykład, zamiast mówić, że mamy cztery obiekty X, Y, U, V i cztery morfizmy $f \in \mathrm{Hom}_C(X, Y)$, $g \in \mathrm{Hom}_C(Y, V)$, $h \in \mathrm{Hom}_C(X, U)$ i $d \in \mathrm{Hom}_C(U, V)$, takie że $gf = dh$, powiemy, że dany jest *kwadrat przemienny*

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle g} \\ U & \xrightarrow{d} & V \end{array}$$

Tutaj „przemienność" oznacza tyle, co równość $gf = dh$, a to z kolei znaczy, że „dwie drogi wzdłuż strzałek" od X do V prowadzą do jednego i tego samego rezultatu. Nieco ogólniej, *diagram* to zorientowany graf, którego wierzchołki są obiektami C, a krawędzie są morfizmami, np.

$$\begin{array}{ccccc} X & \longrightarrow & Y & \longrightarrow & Z \\ & & \downarrow & & \downarrow \\ U & \longrightarrow & V & \longrightarrow & W \end{array}$$

Diagram nazywamy *komutatywnym* (*przemiennym*), jeśli dowolne dwie drogi prowadzące wzdłuż jego strzałek z jednakowym początkiem i końcem odpowiadają jednakowym morfizmom.

W kategorii przestrzeni liniowych, a także w kategorii grup abelowych szczególnie ważna jest klasa diagramów nazywanych *kompleksami*. *Kompleks* ma postać ciągu, skończonego lub nie, obiektów i strzałek

$$\ldots X \longrightarrow Y \longrightarrow U \longrightarrow V \longrightarrow \ldots,$$

spełniającego następujący warunek: *superpozycja dowolnych dwóch kolejnych strzałek jest morfizmem zerowym*. Zauważmy, że pojęcie zerowego morfizmu nie należy do ogólnych kategorii, lecz jest specyficzne dla przestrzeni liniowych i grup abelowych oraz dla specjalnej kasy kategorii — tzw. kategorii addytywnych. Często obiekty i morfizmy występujące w kompleksie indeksowane są kolejnymi liczbami całkowitymi wziętymi z pewnego przedziału:

$$\ldots \longrightarrow X_{-1} \xrightarrow{f_{-1}} X_0 \xrightarrow{f_0} X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} \ldots$$

Taki kompleks przestrzeni liniowych (albo grup abelowych) nazywa się *dokładnym w wyrazie* X_i, jeśli $\operatorname{Im} f_{i-1} = \operatorname{Ker} f_i$ (zauważmy przy tym, że występujący w określeniu kompleksu warunek $f_i \circ f_{i-1} = 0$ oznacza tylko, że $\operatorname{Im} f_{i-1} \subset \operatorname{Ker} f_i$). Kompleks, który jest dokładny we wszystkich wyrazach nazywa się *dokładnym* lub *acyklicznym*, albo *ciągiem dokładnym*.

Oto trzy najprostsze przykłady:

a) Ciąg $0 \longrightarrow L \xrightarrow{i} M$ jest zawsze kompleksem, a jest dokładny w wyrazie L wtedy i tylko wtedy, gdy $\operatorname{Ker} i$ jest obrazem przestrzeni zerowej. Innymi słowami, dokładność jest tutaj tylko innym określeniem injektywności i.

b) Ciąg $M \xrightarrow{j} N \longrightarrow 0$ jest zawsze kompleksem; dokładność w wyrazie N oznacza, że $\operatorname{Im} j = N$, tj. że j jest surjekcją.

c) Kompleks $0 \longrightarrow L \xrightarrow{i} M \xrightarrow{j} N \longrightarrow 0$ jest dokładny, jeśli i jest injekcją, j jest surjekcją oraz $\operatorname{Im} i = \operatorname{Ker} j$. Utożsamiając L z obrazem i — podprzestrzenią w M, możemy następnie utożsamić N z przestrzenią ilorazową M/L i w taki sposób podobne *„trójki dokładne"* są odpowiednikami trójek $(L \subset M, M/L)$.

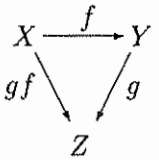
4. Naturalne konstrukcje i funktory. Takie konstrukcje, które po zastosowaniu do obiektów pewnej kategorii dają znowu obiekty kategorii (tej samej lub innej) mają w matematyce niezwykle ważne znaczenie. Jeśli te konstrukcje są jednoznaczne (nie są zależne od dokonywania nieokreślonych wyborów) i są uniwersalnie stosowalne, to często okazuje się, że przenoszą się one także i na morfizmy. Aksjomatyzacja szeregu przykładów doprowadziła do ważnego pojęcia funktora, które zresztą jest także naturalne z punktu widzenia samej teorii kategorii.

5. Definicja funktora. Niech C, D będą dwiema kategoriami. *Funktorem* F z kategorii C do kategorii D nazywamy parę odwzorowań (zazwyczaj oznaczanych także przez F) $\operatorname{Ob} C \to \operatorname{Ob} D$, $\operatorname{Mor} C \to \operatorname{Mor} D$, spełniających następujące warunki:

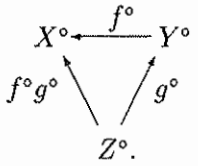
a) jeśli $f \in \mathrm{Hom}_C(X,Y)$, to $F(f) \in \mathrm{Hom}_C(F(X),F(Y))$;
b) $F(gf) = F(g)F(f)$ zawsze gdy złożenie gf jest określone oraz $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ dla wszystkich $X \in \mathrm{Ob}\,C$.

Tak określone funktory zazwyczaj nazywane są *funktorami kowariantnymi*. Określa się także *funktory kontrawariantne*, „odwracające strzałki", dla których warunki a) i b) zamienione zostają na:
a') jeśli $f \in \mathrm{Hom}_C(X,Y)$, to $F(f) \in \mathrm{Hom}_C(F(Y),F(X))$;
b') $F(gf) = F(f)F(g)$ i $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$.

Można uniknąć tego rozróżnienia przez wprowadzenie konstrukcji, przyporządkowującej każdej kategorii C *kategorię* C° *do niej dualną* zgodnie z następującym określeniem: $\mathrm{Ob}\,C = \mathrm{Ob}\,C^\circ$, $\mathrm{Mor}\,C = \mathrm{Mor}\,C^\circ$ i $\mathrm{Hom}_C(X,Y) = \mathrm{Hom}_{C^\circ}(Y,X)$, a przy tym superpozycji gf morfizmów w C odpowiada superpozycja fg tych samych morfizmów w C°, ale składanych w odwrotnym porządku. Wygodnie oznaczać przez X^0, f^0 obiekty i morfizmy w C° odpowiadające obiektom i morfizmom X, f w C. Wtedy diagram przemienny w C



odpowiada następującemu diagramowi przemiennemu w C°:



(Kowariantny) funktor $F: C \to D^\circ$ można utożsamić z kontrawariantnym funktorem $F: C \to D$, określonym przez podaną wyżej definicję.

6. Przykłady. a) Niech $\boldsymbol{K}$ będzie ciałem, $\mathcal{L}in_{\boldsymbol{K}}$ — kategorią przestrzeni liniowych nad $\boldsymbol{K}$, $\mathcal{S}et$ — kategorią zbiorów. W paragrafie 1 opisaliśmy, jak z każdym zbiorem $S \in \mathrm{Ob}\,\mathcal{S}et$ jest związana przestrzeń liniowa $F(S) \in \mathrm{Ob}\,\mathcal{L}in_{\boldsymbol{K}}$ funkcji na S o wartościach w $\boldsymbol{K}$. Ponieważ jest to naturalna konstrukcja, można oczekiwać, że istnieje jej rozszerzenie do funktora. Tak też jest rzeczywiście. Funktor ten jest *kontrawariantny*: morfizmowi $f: S \to T$ przypisuje odwzorowanie liniowe $F(f): F(T) \to F(S)$, nazywane *podniesieniem* albo *obrazem odwrotnym* funkcji, a które najczęściej bywa oznaczane przez f^*:

$$f^*(\varphi) = \varphi \circ f, \quad \text{gdzie } f: S \to T,\ \varphi: T \to \boldsymbol{K}.$$

Innymi słowy, $f^*(\varphi)$ to taka funkcja na S, której wartości są stałe na „włóknach" $f^{-1}(t)$ odwzorowania f i na takim włóknie wynoszą $\varphi(t)$. Dobrym zadaniem dla

czytelnika będzie sprawdzenie, że w ten sposób rzeczywiście został skonstruowany funktor.

b) Odwzorowanie dwoistości $\mathcal{L}in_{\boldsymbol{K}} \to \mathcal{L}in_{\boldsymbol{K}}$, zadawane na obiektach wzorem $L \mapsto L^* = \mathcal{L}(L, \boldsymbol{K})$, a na morfizmach wzorem $f \mapsto f^*$, jest *kontrawariantnym funktorem* z kategorii $\mathcal{L}in_{\boldsymbol{K}}$ w siebie. W istocie zostało to udowodnione w § 7.

c) Operacje kompleksyfikacji i realifikacji rozważane w § 12 określają funktory $\mathcal{L}in_{\boldsymbol{R}} \to \mathcal{L}in_{\boldsymbol{C}}$ i $\mathcal{L}in_{\boldsymbol{C}} \to \mathcal{L}in_{\boldsymbol{R}}$ odpowiednio. To samo dotyczy ogólnych konstrukcji rozszerzenia i zwężenia ciała skalarów, opisanych pokrótce w tymże § 12.

d) Dla dowolnej kategorii C i dowolnego obiektu $X \in \operatorname{Ob} C$ są określone dwa funktory z C do kategorii zbiorów: kowariantny $h_X : C \to \mathcal{S}et$ i kontrawariantny $h^X : C \to \mathcal{S}et^\circ$.

Definiuje się je następująco: $h_X(Y) = \operatorname{Hom}_C(X, Y)$, $h_X(f : Y \to Z)$ jest odwzorowaniem $h_X(Y) = \operatorname{Hom}_C(X, Y) \to h_X(Z) = \operatorname{Hom}_C(X, Z)$, które przypisuje morfizmowi $X \to Y$ jego złożenie z morfizmem $f : Y \to Z$.

Analogicznie $h^X(Y) = \operatorname{Hom}_C(Y, X)$, natomiast $h^X(f : Y \to Z)$ jest odwzorowaniem $h^X(Z) = \operatorname{Hom}_C(Z, X) \to h^X(Y) = \operatorname{Hom}_C(Y, X)$, które przypisuje morfizmowi $Z \to X$ jego złożenie z morfizmem $f : Y \to Z$.

Czytelnik sprawdzi, że h_X i h^X rzeczywiście są funktorami. Nazywają się one *funktorami reprezentującymi* (*przedstawiającymi*) obiekt X kategorii.

Zauważmy, że dla kategorii $C = \mathcal{L}in_{\boldsymbol{K}}$ można traktować h_X i h^X jako funktory o wartościach w $\mathcal{L}in_{\boldsymbol{K}}$, a nie jedynie w $\mathcal{S}et$.

7. Składanie funktorów. Dla danych $C_1 \xrightarrow{F} C_2 \xrightarrow{G} C_3$ trzech kategorii i dwóch funktorów między nimi określimy złożenie $GF : C_1 \to C_3$ jako teoriomnogościowe złożenie odwzorowań na obiektach i morfizmach. Trywialne sprawdzenie pokazuje, że jest ono także funktorem.

Można wprowadzić „kategorię kategorii", której obiektami są kategorie, a morfizmami są funktory!

Jednakże jeszcze ważniejszy jest następny stopień tej wysokiej drabiny abstrakcji, a mianowicie *kategoria funktorów*. Poniżej opiszemy tylko pokrótce, czym są morfizmy funktorów.

8. Naturalne przekształcenia naturalnych konstrukcji albo morfizmy funktorów. Niech $F, G : C \to D$ będą dwoma funktorami o wspólnym początku i końcu. *Morfizmem funktorów* $\varphi : F \to G$ nazywamy klasę morfizmów obiektów $\varphi(X)$: $F(X) \to G(X)$ z kategorii D, która zawiera dla każdego obiektu X kategorii C tylko jeden morfizm, spełniających następujący warunek: dla każdego morfizmu $f : X \to Y$ w kategorii C kwadrat

$$\begin{array}{ccc} F(X) & \xrightarrow{\varphi(X)} & G(X) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(Y) & \xrightarrow{\varphi(Y)} & G(Y) \end{array}$$

jest przemienny. Morfizm funktorów nazywamy *izomorfizmem*, jeśli wszystkie $\varphi(X)$ są izomorfizmami.

9. Przykład. Niech $**: \mathcal{L}in_K \to \mathcal{L}in_K$ oznacza funktor „dwukrotnego sprzężenia": $L \mapsto L^{**}$, $f \mapsto f^{**}$. W § 7 dla każdej przestrzeni liniowej L skonstruowaliśmy kanoniczne odwzorowanie liniowe $\varepsilon_L: L \to L^{**}$. Określa ono morfizm funktorów $\varepsilon_M: \mathrm{Id} \to **$, gdzie Id oznacza funktor tożsamościowy na $\mathcal{L}in_K$, który przyporządkowuje każdej przestrzeni liniowej tę samą przestrzeń i każdemu odwzorowaniu liniowemu to samo odwzorowanie. Rzeczywiście, w zgodzie z określeniem powinniśmy sprawdzić przemienność wszystkich możliwych kwadratów postaci

$$\begin{array}{ccc} L & \xrightarrow{\varepsilon_L} & L^{**} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f^{**}} \\ M & \xrightarrow{\varepsilon_M} & M^{**} \end{array}$$

Dla przestrzeni skończonego wymiaru wniosek taki otrzymujemy na mocy części d) twierdzenia § 7, p. 5, a uzasadnienie w ogólnym przypadku pozostawiamy czytelnikowi.

Ćwiczenia

1. Oznaczmy przez $\mathcal{S}et_0$ kategorię, w której obiektami są zbiory, a morfizmami takie odwzorowania zbiorów $f: S \to T$, że dla dowolnego punktu $t \in T$ włókno $f^{-1}(t)$ jest skończone. Dowieść, że w następujący sposób zostaje określony *kowariantny* funktor $F_0: \mathcal{S}et_0 \to \mathcal{L}in_K$.

a) $F_0(S) = F(S)$ — zbiór funkcji na S o wartościach w K;

b) dla dowolnego morfizmu $f: S \to T$ i funkcji $\varphi: S \to K$ funkcja $F_0(f)(\varphi) = f_*(\varphi) \in F(T)$ jest określona wzorem

$$f_*(\varphi)(t) = \sum_{s \in f^{-1}(t)} \varphi(s)$$

(„całkowanie po włóknach").

2. Wykazać, że zwężenie ciała skalarów z K do K (por. § 12) definiuje funktor $\mathcal{L}in_K \to \mathcal{L}in_K$.

§ 14. Kategoryjne własności przestrzeni liniowych

1. W tym paragrafie zbierzemy pewne fakty dotyczące kategorii wszystkich przestrzeni liniowych $\mathcal{L}in_K$ lub kategorii skończenie wymiarowych przestrzeni liniowych $\mathcal{L}in f_K$ nad ustalonym ciałem K. W większej części są one tylko przetłumaczeniem na język kategorii stwierdzeń wykazanych już wcześniej. Ich wybór był uwarunkowany następującym kryterium: są to przede wszystkim te własności, które *nie*

zachowują się przy przejściu do najbliższych kategorii takich jak kategoria moduł nad ogólnymi pierścieniami (np. nad $\boldsymbol{Z}$, tj. kategoria grup abelowych), czy naw kategoria nieskończenie wymiarowych topologicznych przestrzeni liniowych. Szcz gółowe badanie tych odstępstw dla kategorii modułów stanowi podstawowe zadan *algebry homologicznej*, a w analizie funkcjonalnej często prowadzi do poszukiwan nowych definicji, które pozwalają odtworzyć „dobre" własności $\mathcal{L}inf_{\boldsymbol{K}}$ (takim je np. pojęcie przestrzeni nuklearnych).

2. Twierdzenie o przedłużaniu odwzorowań. a) *Niech P, M, N będą prz strzeniami liniowymi, przy czym P jest skończenie wymiarowa, a $j\colon M \to N$ nie będzie liniową surjekcją. Wtedy dowolne odwzorowanie liniowe $g\colon P \to N$ można po nieść do takiego odwzorowania liniowego $h\colon P \to M$, dla którego $g = jh$. Innym słowy, poniższy diagram o dokładnym wierszu*

$$\begin{array}{ccccc} & & P & & \\ & & \downarrow{\scriptstyle g} & & \\ M & \xrightarrow{j} & N & \longrightarrow & 0 \end{array}$$

można uzupełnić do diagramu przemiennego

$$\begin{array}{ccccc} & & P & & \\ & \swarrow{\scriptstyle h} & \downarrow{\scriptstyle g} & & \\ M & \xrightarrow{j} & N & \longrightarrow & 0. \end{array}$$

b) *Niech P, L, M będą przestrzeniami liniowymi, przy czym M jest skończen wymiarowa, a $i\colon N \to M$ niech będzie liniową injekcją. Wtedy dowolne odwzorowan liniowe $g\colon N \to P$ można przedłużyć do takiego odwzorowania liniowego $h\colon M \to$ dla którego $g = hi$. Innymi słowy, poniższy diagram o dokładnym wierszu*

$$\begin{array}{ccccc} & & P & & \\ & & \uparrow{\scriptstyle g} & & \\ M & \xleftarrow{i} & L & \longleftarrow & 0 \end{array}$$

można uzupełnić do diagramu przemiennego

$$\begin{array}{ccccc} & & P & & \\ & \nearrow{\scriptstyle h} & \uparrow{\scriptstyle g} & & \\ M & \xleftarrow{i} & L & \longleftarrow & 0. \end{array}$$

Dowód. a) Wybierzmy bazę $\{e_1, \ldots, e_n\}$ P i oznaczmy $e'_i = g(e_i) \in N$. Na mocy surjektywności j istnieją takie wektory $e''_i \in M$, dla których $(e''_i) = e'_i$, $i = 1, \ldots, n$. Ze stwierdzenia § 3, p. 3. wynika, że istnieje jedyne odwzorowanie liniowe $h\colon P \to M$, dla którego $h(e_i) = e''_i$, $i = 1, \ldots, n$. Z konstrukcji $jh(e_i) = j(e''_i) = e'_i = g(e_i)$. Ponieważ $\{e_i\}$ tworzą bazę P, wnioskujemy, że $jh = g$.

b) Wybierzmy bazę $\{e'_1, \ldots, e'_m\}$ przestrzeni L i uzupełnijmy $e_k = i(e'_k)$, $1 \leq k \leq m$, do bazy $\{e_1, \ldots, e_m; e_{m+1}, \ldots, e_n\}$ przestrzeni M. Przyjmijmy $h(e_i) = g(e'_i)$ dla $1 \leq i \leq m$, $h(e_i) = 0$ dla $m+1 \leq i \leq n$. Istnienie takiego odwzorowania wynika znowu z tego samego stwierdzenia § 3, p. 3, choć można by również zastosować stwierdzenie § 6, p. 8. W ten sposób twierdzenie zostało wykazane.

W kategorii modułów obiekty P mające własność a) twierdzenia (dla wszystkich M, N) nazywane są *projektywnymi*, a obiekty o własności b) — *injektywnymi*. Wykazaliśmy więc, że w kategorii skończenie wymiarowych przestrzeni liniowych wszystkie obiekty są zarówno projektywne, jak i injektywne.

3. Twierdzenie o dokładności funktora L. *Niech* $0 \longrightarrow L \xrightarrow{i} M \xrightarrow{j} N \longrightarrow 0$ *oznacza dokładną trójkę skończenie wymiarowych przestrzeni liniowych, a* P *— dowolną skończenie wymiarową przestrzeń liniową nad tym samym ciałem* K. *Wtedy* $\mathcal{L}$, *traktowany jako funktor ze względu na pierwszy lub drugi argument w oddzielności, indukuje dokładne trójki przestrzeni liniowych:*

a) $$0 \longrightarrow \mathcal{L}(P, L) \xrightarrow{i_1} \mathcal{L}(P, M) \xrightarrow{j_1} \mathcal{L}(P, N) \longrightarrow 0,$$

b) $$0 \longleftarrow \mathcal{L}(L, P) \xleftarrow{i_2} \mathcal{L}(M, P) \xleftarrow{j_2} \mathcal{L}(N, P) \longleftarrow 0.$$

Dowód. a) Przypomnijmy (por. przykład d) § 13 p. 6), że i_1 przyporządkowuje morfizmowi $P \to L$ jego złożenie z $i : L \to M$, a j_1 przyporządkowuje morfizmowi $P \to M$ jego złożenie z $j\colon M \to N$. i_1 jest injekcją, gdyż i nią jest, bowiem jeśli superpozycja $P \to L \to M$ jest morfizmem zerowym, to $P \to L$ też nim jest. Odwzorowanie j_1 jest surjekcją na mocy części a) twierdzenia p. 2: dowolny morfizm $g\colon P \to N$ można podnieść do morfizmu $P \to M$, który po złożeniu z j daje wyjściowy morfizm. Złożenie $j_1 i_1$ jest zerem, bo przeprowadza strzałkę $P \to L$ w strzałkę $P \to N$, będącą złożeniem $P \longrightarrow L \xrightarrow{i} M \xrightarrow{j} N$, gdzie $ji = 0$.

W ten sposób sprawdziliśmy, że ciąg a) jest kompleksem i pozostaje do wykazania jego dokładność w środkowym wyrazie, tj. równość $\operatorname{Ker} j_1 = \operatorname{Im} i_1$. Wiemy już, że $\operatorname{Ker} j_1 \supset \operatorname{Im} i_1$. Dla wykazania odwrotnej inkluzji zauważmy, że jeśli strzałka $P \to M$ należy do jądra j_1, to złożenie tej strzałki z $j\colon M \to N$ jest zerem i dlatego obraz P w M jest zawarty w jądrze j. Jednakże jądro j jest jednocześnie obrazem $i(L) \subset M$ na mocy dokładności wyjściowej trójki. To oznacza, że P jest odwzorowywane w podprzestrzeń $i(L)$, a więc strzałkę $P \to M$ można podnieść do strzałki $P \to L$, która po złożeniu z i daje wyjściową strzałkę. To zaś oznacza, że ta ostatnia należy do obrazu i_1.

b) Tutaj rozważania są analogiczne lub, lepiej, dwoiste. Odwzorowanie i_2 jest surjekcją na mocy części b) twierdzenia p. 2. Odwzorowanie j_2 jest injektywne, gdyż

jeśli złożenie $M \xrightarrow{j} N \longrightarrow P$ jest zerowe, to także strzałka $N \to P$ jest zerowa, bo j jest surjekcją. Złożenie $i_2 j_2$ jest zerem, bowiem złożenie $L \longrightarrow M \xrightarrow{j} N \longrightarrow P$ jest zerem dla dowolnej strzałki na ostatnim miejscu. Zatem pozostaje dowieść, że $\operatorname{Ker} i_2 \subset \operatorname{Im} j_2$ (przeciwną inkluzję właśnie udowodniliśmy). Jeśli złożenie $L \longrightarrow M \xrightarrow{f} P$ jest zerem, to strzałka $M \xrightarrow{f} P$ jest zawarta w jądrze i_2. To oznacza, że $L = \operatorname{Ker} j$ zawiera się w jądrze f. Zdefiniujmy odwzorowanie $\overline{f}: N \to P$ za pomocą wzoru $\overline{f}(n) = f(j^{-1}(n))$, gdzie $j^{-1}(n)$ oznacza dowolny punkt przeciwobrazu n. Od jego wyboru nic nie zależy, bo $\operatorname{Ker} j \subset \operatorname{Ker} f$. Łatwo się sprawdza, że $\overline{f}$ jest liniowy oraz $j_2(\bar{f}) = f$; w istocie, $j_2(\bar{f})$ jest złożeniem $M \xrightarrow{j} N \xrightarrow{\overline{f}} P$, które przeprowadza $m \in M$ na $\overline{f} j(m) = f(j^{-1}(j(m))) = f(m)$. To dowodzi twierdzenia.

4. Kategoryjna charakterystyka wymiaru. Niech G będzie jakąś grupą abelową zapisywaną addytywnie, $\chi: \operatorname{Ob} \mathcal{L}in f_{\boldsymbol{K}} \to G$ — funkcją określoną na skończenie wymiarowych przestrzeniach liniowych i spełniającą dwa warunki:

a) jeśli L i M sa izomorficzne, to $\chi(L) = \chi(M)$;

b) dla dowolnej trójki przestrzeni $0 \to L \to M \to N \to 0$ zachodzi równość $\chi(M) = \chi(L) + \chi(N)$ (takie funkcje nazywają się *addytywnymi*).

Spełnione jest następujące

5. Twierdzenie. *Dla dowolnej funkcji addytywnej χ mamy*

$$\chi(L) = dim_{\boldsymbol{K}} L \cdot \chi(\boldsymbol{K}^1),$$

jeśli L jest przestrzenią skończonego wymiaru.

Dowód. Przeprowadzimy indukcję względem wymiaru L. Jeśli L jest jednowymiarowa, to jest izomorficzna z $\boldsymbol{K}^1$, tak że $\chi(L) = \chi(\boldsymbol{K}^1) = dim_{\boldsymbol{K}} L \cdot \chi(\boldsymbol{K}^1)$. Załóżmy, że twierdzenie jest wykazane dla wszystkich przestrzeni wymiaru n. Jeśli wymiar L jest równy $n+1$, to weźmy jednowymiarową podprzestrzeń $L_0 \subset L$ i rozważmy trójkę

$$0 \longrightarrow L_0 \xrightarrow{i} L \xrightarrow{j} L/L_0 \longrightarrow 0,$$

gdzie i jest włożeniem L_0, a $j(l) = l + L_0 \in L/L_0$. Na mocy addytywności χ i założenia indukcyjnego

$$\chi(L) = \chi(L_0) + \chi(L/L_0) = \chi(\boldsymbol{K}^1) + dim_{\boldsymbol{K}}(L/L_0)\chi(\boldsymbol{K}^1) =$$

$$= \chi(\boldsymbol{K}^1) + n\chi(\boldsymbol{K}^1) = (n+1)\chi(\boldsymbol{K}^1) = dim_{\boldsymbol{K}} L \cdot \chi(\boldsymbol{K}^1).$$

Twierdzenie zostało wykazane.

Ten wynik jest początkowym rezultatem szybko rozwijającej się obecnie obszernej algebraicznej teorii, zwanej **K**-teorią, która leży na styku topologii i algebry.

Ćwiczenia

1. Jeśli $K\colon 0 \xrightarrow{d_0} L_1 \xrightarrow{d_1} L_2 \xrightarrow{d_2} \cdots \xrightarrow{d_{n-1}} L_n \xrightarrow{d_n} 0$ jest kompleksem skończenie wymiarowych przestrzeni liniowych, to określoną dla $i = 1, \ldots, n$ przestrzeń ilorazową $H^i(K) = \operatorname{Ker} d_i / \operatorname{Im} d_{i-1}$ nazywamy *i-tą przestrzenią kohomologii kompleksu* K, a liczba $\chi(K) = \sum_{i=1}^{n} (-1)^i \dim L_i$ — *charakterystyką Eulera* tego kompleksu. Dowieść, że

$$\chi(K) = \sum_{i=1}^{n} (-1)^i \dim H^i(K).$$

2. „Lemat o wężu". Niech będzie dany diagram przemienny przestrzeni liniowych

$$\begin{array}{ccccccccc} & & L & \xrightarrow{d_1} & M & \xrightarrow{d_2} & N & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} & & \\ 0 & \longrightarrow & L' & \xrightarrow{d_1'} & M' & \xrightarrow{d_2'} & N' & & \end{array}$$

o dokładnych wierszach. Wykazać istnienie ciągu dokładnego przestrzeni

$$\operatorname{Ker} f \longrightarrow \operatorname{Ker} g \longrightarrow \operatorname{Ker} h \xrightarrow{\delta} \operatorname{Coker} f \longrightarrow \operatorname{Coker} g \longrightarrow \operatorname{Coker} h,$$

w którym wszystkie strzałki poza δ są indukowane odpowiednio przez d_1, d_2, d_1', d_2', a homomorfizm związujący δ (nazywany także operatorem kogranicy) jest określony w następujący sposób. Dla zadania wartości $\delta(n)$ w punkcie $n \in \operatorname{Ker} h$ należy znaleźć $m \in M$, dla którego $n = d_2(m)$, wyznaczyć $g(m) \in M'$, znaleźć takie $l' \in L'$, że $d_1'(l') = g(m)$ i przyjąć $\delta(n) = l' + \operatorname{Im} f \in \operatorname{Coker} f$. W szczególności należy sprawdzić istnienie $\delta(n)$ i jego niezależność od dowolności wyborów uczynionych w konstrukcji.

3. Niech $K\colon \ldots \longrightarrow L_i \xrightarrow{d_i} L_{i+1} \longrightarrow \ldots$ i $K''\colon \ldots \longrightarrow L_i' \xrightarrow{d_i'} L_{i+1}' \longrightarrow \ldots$ będą dwoma kompleksami. Morfizmem $f\colon K \to K'$ nazywa się taki układ odwzorowań liniowych $f_i\colon L_i \to L_i'$, że wszystkie kwadraty

$$\begin{array}{ccc} L_i & \xrightarrow{d_i} & L_{i+1} \\ \downarrow{\scriptstyle f_i} & & \downarrow{\scriptstyle f_{i+1}} \\ L_i' & \xrightarrow{d_i'} & L_{i+1}' \end{array}$$

są przemienne. Pokazać, że kompleksy i ich morfizmy tworzą kategorię.

4. Wykazać, że odwzorowanie $K \longrightarrow H^i(K)$ można przedłużyć do funktora z kategorii kompleksów do kategorii przestrzeni liniowych.

5. Niech $0 \longrightarrow K \xrightarrow{f} K' \xrightarrow{g} K'' \longrightarrow 0$ będzie dokładną trójką kompleksów i ich morfizmów. Z definicji oznacza to, że dla każdego i trójki przestrzeni liniowych

$$0 \longrightarrow L_i \xrightarrow{f_i} L_i' \xrightarrow{g_i} L_i'' \longrightarrow 0$$

są dokładne. Niech H^i oznaczają odpowiednie grupy kohomologii. Używając lematu o wężu, zbudować ciąg przestrzeni kohomologii

$$\ldots \longrightarrow H^i(K) \longrightarrow H^i(K') \longrightarrow H^i(K'') \xrightarrow{\delta} H^{i+1}(K) \longrightarrow \ldots$$

i wykazać, że jest to ciąg dokładny.

Część 2.

Geometria przestrzeni z iloczynem skalarnym

§ 1. O geometrii

1. Ta i następna część naszego wykładu są poświęcone tematowi, który można by nazwać „geometrią liniową". Jednak zanim do niego przejdziemy warto omówić pokrótce współczesne znaczenie słów „geometria" i „geometryczny". W ciągu wielu stuleci słowo „geometria" oznaczało wyłącznie geometrię Euklidesa na płaszczyźnie i w przestrzeni. Takie znaczenie ma ono ciągle jeszcze w zwykłym programie szkolnym i dlatego wygodniej nam będzie śledzić ewolucję pojęć geometrycznych, posługując się przykładami charakterystycznych własności tego, obecnie już bardzo specjalnego, działu geometrii.

2. „Figury". Szkolną geometrię zaczyna się poznawać, studiując własności takich figur płaszczyzny jak proste, kąty, trójkąty, koła, okręgi itp. Naturalnym uogólnieniem tego podejścia jest wybór pewnej przestrzeni M, „przestrzeni obejmującej" dla naszej geometrii, i pewnego zbioru podzbiorów M składającego się z figur badanych w tej przestrzeni.

3. „Ruchy". Drugim istotnym składnikiem szkolnej geometrii jest mierzenie długości i kątów oraz objaśnianie związków pomiędzy liniowymi a kątowymi elementami rozmaitych figur. Dopiero długi proces historyczny pozwolił zrozumieć, że u podstawy tych pomiarów leży istnienie specyficznego obiektu matematycznego, a mianowicie grupy ruchów płaszczyzny euklidesowej czy przestrzeni euklidesowej jako całości, i że wszystkie pojęcia metryczne mogą być określone w terminach tej grupy. Na przykład, odległość punktów jest jedyną funkcją par punktów, która jest niezmiennicza względem grupy ruchów euklidesowych (jeśli żądać ciągłości tej funkcji i ustalić „jednostkę długości" — odległość między wybraną parą punktów). *Program z Erlangen* F. Kleina (1872) utrwalił znaczenie tej fundamentalnej zasady i od tego momentu już na dobre „geometrią" stało się badanie przestrzeni M z dostatecznie dużą grupą symetrii oraz tych własności figur, które są niezmiennicze względem działania tej grupy, włączając w to kąty, odległości i objętości.

4. „Liczby". Równie fundamentalnym odkryciem, a przy tym o wiele wcześniejszym, była kartezjańska „metoda współrzędnych" i oparta na niej geometria analityczna płaszczyzny i przestrzeni. Ze współczesnego punktu widzenia współrzędne są pewnymi funkcjami określonymi na przestrzeni M (lub na jej podzbiorach) o wartościach rzeczywistych, zespolonych lub jeszcze ogólniejszych. Zadanie konkretnych wartości tych funkcji określa punkt przestrzeni, a zadanie zależności między tymi wartościami określa zbiór punktów. Opis klasy figur w M rozważanych w danej geometrii można zastąpić wyróżnieniem klasy tych zależności między współrzędnymi, które zadają interesujące nas figury. Zadziwiająca elastyczność i moc metody Kartezjusza wypływa stąd, że funkcje można dodawać i mnożyć, całkować, różniczkować, a także stosować do nich inne sposoby przejścia granicznego, co sprawia, że przy ich badaniu możemy rozporządzać pełną mocą analizy matematycznej. Wszystkie współczesne dyscypliny geometryczne — topologia, geometria różniczkowa, zespolona geometria analityczna czy geometria algebraiczna — za punkt wyjścia przyjmują określenie obiektu geometrycznego jako układu przestrzeni M i zadanej na niej rodziny F (lokalnych) funkcji.

5. „Odwzorowania". Jeśli (M_1, F_1) i (M_2, F_2) są dwoma obiektami geometrycznymi opisanego wyżej rodzaju, to można rozważać odwzorowania $M_1 \to M_2$, które mają tę własność, że odwzorowanie „cofania funkcji" przeprowadza elementy z F_2 na elementy z F_1. W najbardziej zadowalających logicznie systemach wśród takich odwzorowań znajdują się zarówno grupy symetrii F. Kleina, jak i same funkcje współrzędnościowe (jako odwzorowania M w $\boldsymbol{R}$ lub $\boldsymbol{C}$). Obiekty geometryczne stanowią kategorię, a morfizmy tej kategorii są wystarczająco dokładnym zamiennikiem symetrii nawet wtedy, gdy tych symetrii nie jest tak wiele (jak w ogólnych przestrzeniach riemannowskich, gdzie można mierzyć długości, kąty i objętości, ale mówiąc najogólniej, nie ma dostatecznej ilości ruchów).

6. Geometrie liniowe. Teraz można już scharakteryzować miejsce geometrii liniowych w tym ogólnym pejzażu. Geometrie liniowe można w pewnym sensie zaliczyć do bezpośredniego potomstwa geometrii Euklidesa. Rozważane w nich przestrzenie M są bądź to *przestrzeniami liniowymi* (tym razem już nad dowolnymi ciałami, ale jak poprzednio $\boldsymbol{R}$ i $\boldsymbol{C}$ pozostają w centrum uwagi, choćby ze względu na wielorakie zastosowania), bądź przestrzeniami wywodzącymi się od przestrzeni liniowych: *afinicznymi* („przestrzenie liniowe bez wyróżnionego początku układu współrzędnych") i *rzutowymi* („przestrzenie afiniczne uzupełnione nieskończenie oddalonymi punktami"). Grupami symetrii są podgrupy grupy liniowej, zachowujące ustalony „iloczyn skalarny", a także ich rozszerzenia o przesunięcia (grupy afiniczne) oraz grupy ilorazowe przez homotetie (grupy rzutowe). Rozważane funkcje są liniowe lub bliskie liniowym, czasem także kwadratowe. Figurami są podprzestrzenie liniowe i rozmaitości (uogólnienie prostych na płaszczyźnie euklidesowej) i kwadryki (uogólnienie okręgów). Można byłoby przypuszczać, że te uogólnienia geometrii euklidesowej to wynik czysto logicznej analizy, gdyż ukształtowany formalizm geometrii liniowych ma zadziwiający urok i zwartość. Jednakże siły życiowe tej gałęzi matematyki

są w znacznej mierze związane z istnieniem wielorakich jej zastosowań w naukach ścisłych. Pojęcie iloczynu skalarnego, leżące u podstaw całej części drugiej tego wykładu, można stosować do mierzenia kątów w abstrakcyjnych przestrzeniach euklidesowych. Jednak matematyk, który nie wie, że mierzy on także prawdopodobieństwa zdarzeń (w modelach mechaniki kwantowej), prędkości (w przestrzeni Minkowskiego szczególnej teorii względności) i współczynniki korelacji zmiennych losowych (w teorii prawdopodobieństwa), nie tylko zawęża swój horyzont poznawczy, ale pozbawia elastyczności swą czysto matematyczną intuicję. Dlatego uznaliśmy za konieczne dołączenie wiadomości także o tych interpretacjach.

§ 2. Iloczyny skalarne

1. Odwzorowania wieloliniowe. Niech $L_1, \ldots, L_n, M$ będą przestrzeniami liniowymi nad wspólnym ciałem $\boldsymbol{K}$. *Odwzorowaniem wieloliniowym* (przy $n = 2$ *dwuliniowym* lub *biliniowym*) nazywamy odwzorowanie

$$f\colon L_1 \times \cdots \times L_n \to M, \quad (l_1, \ldots, l_n) \mapsto f(l_1, \ldots, l_n) \in M,$$

jeśli jest ono liniowe jako funkcja każdego z argumentów $l_i \in L_i$ przy ustalonych wszystkich pozostałych $l_j \in L_j$, $j = 1, \ldots, n$, $j \neq i$. Inaczej mówiąc, spełnione są równości

$$f(l_1, \ldots, l_i + l_i', l_{i+1}, \ldots, l_n) = f(l_1, \ldots, l_i, \ldots, l_n) + f(l_1, \ldots, l_i', \ldots, l_n),$$
$$f(l_1, \ldots, al_i, l_{i+1}, \ldots, l_n) = af(l_1, \ldots, l_i, l_{i+1}, \ldots, l_n)$$

dla $i = 1, \ldots, n$; $a \in \boldsymbol{K}$. W przypadku $M = \boldsymbol{K}$ wieloliniowe odwzorowania nazywają się *funkcjami* (*formami*) *wieloliniowymi*.

W pierwszej części spotkaliśmy się już z odwzorowaniami dwuliniowymi

$$L \times L \longrightarrow \boldsymbol{K} : (f, l) \mapsto f(l), \quad f \in L^*, \, l \in L,$$
$$\mathcal{L}(L, M) \times L \longrightarrow M : (f, l) \mapsto f(l), \quad f \in \mathcal{L}(L, M), \, l \in L.$$

Wyznacznik macierzy kwadratowej jest funkcją wieloliniową jej wierszy lub kolumn. Jeszcze innego przykładu dostarcza odwzorowanie

$$\boldsymbol{K}^n \times \boldsymbol{K}^m \longrightarrow \boldsymbol{K} : (\vec{x}, \vec{y}) \mapsto \sum_{i,j} g_{ij} x_i y_j = \vec{x}^{\,t} G \vec{y},$$

gdzie G jest dowolną macierzą o wymiarach $n \times m$ nad ciałem $\boldsymbol{K}$, a wektory z $\boldsymbol{K}^n$ i $\boldsymbol{K}^m$ są reprezentowane przez wektory-kolumny współrzędnych.

Ogólne odwzorowania wieloliniowe będziemy badać później, w części poświęconej algebrze tensorowej. Tu natomiast zajmiemy się najważniejszą dla zastosowań klasą

funkcji dwuliniowych $L \times L \to \boldsymbol{K}$, lub w przypadku $\boldsymbol{K} = \boldsymbol{C}$, funkcji $L \times \overline{L} \to \boldsymbol{C}$, gdzie przez $\overline{L}$ oznaczona jest przestrzeń sprzężona z L (por. część 1, § 12). Taką funkcję nazywamy *iloczynem skalarnym* albo *metryką*[1] w przestrzeni L, a para (L, iloczyn skalarny) jest traktowana jako jeden obiekt geometryczny. Metryka w tym sensie jedynie w szczególnych przypadkach jest metryką w znaczeniu definicji wprowadzonej w części 1, § 10, p. 1 i czytelnik nie powinien dać się wprowadzić w błąd tym homonimom.

Iloczyn skalarny $L \times \overline{L} \to \boldsymbol{C}$ najczęściej rozpatrujemy jako *półtoraliniowe* odwzorowanie $g\colon L \times L \to \boldsymbol{C}$, liniowe ze względu na pierwszy argument, a półliniowe ze względu na drugi: $g(al_1, bl_2) = a\overline{b}g(l_1, l_2)$.

2. Sposoby wprowadzania iloczynu skalarnego. a) Niech $g\colon L \times L \to \boldsymbol{K}$ (lub $L \times \overline{L} \to \boldsymbol{C}$) będzie iloczynem skalarnym w skończenie wymiarowej przestrzeni L. Dla dowolnej bazy $\{e_1, \ldots, e_n\}$ przestrzeni L zdefiniujemy macierz

$$G = (g(e_i, e_j)), \quad i, j = 1, \ldots, n,$$

którą będziemy nazywać *macierzą Grama* bazy $\{e_1, \ldots, e_n\}$ względem g lub, inaczej, *macierzą g w bazie* $\{e_1, \ldots, e_n\}$. Wybór $\{e_i\}$ i G całkowicie wyznacza g, gdyż na mocy warunku dwuliniowości

$$g(\vec{x}, \vec{y}) = g\Big(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\Big) = \sum_{i,j=1}^{n} x_i y_j g(e_i, e_j) = \vec{x}^{\,t} G \vec{y}.$$

W przypadku formy półtoraliniowej analogiczny wzór ma postać

$$g(\vec{x}, \vec{y}) = g\Big(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\Big) = \sum_{i,j=1}^{n} x_i \overline{y_j} g(e_i, e_j) = \vec{x}^{\,t} G \overline{\vec{y}}.$$

Odwrotnie, jeśli baza $\{e_1, \ldots, e_n\}$ jest ustalona, a G jest dowolną macierzą kwadratową stopnia n nad $\boldsymbol{K}$, to odwzorowanie $(\vec{x}, \vec{y}) \to \vec{x}^{\,t} G \vec{y}$ (albo $\vec{x}^{\,t} G \overline{\vec{y}}$ w przypadku półtoraliniowym) zadaje iloczyn skalarny w L, którego macierzą względem tej bazy jest, jak łatwo sprawdzić, G. W ten sposób nasza konstrukcja ustala bijekcję między iloczynami skalarnymi (dwuliniowymi czy półtoraliniowymi) w n-wymiarowej przestrzeni z wybraną bazą a macierzami kwadratowymi stopnia n.

Wyjaśnijmy, jak zmienia się G przy zamianie bazy. Niech A będzie macierzą przejścia do bazy primowanej. Współrzędne są związane wzorem $\vec{x} = A\vec{x}'$, gdzie $\vec{x}$ oznacza kolumnę współrzędnych w starej bazie, a $\vec{x}'$ — kolumnę współrzędnych tego samego wektora w nowej bazie. Wtedy w przypadku formy dwuliniowej

$$(\vec{x}, \vec{y}) = \vec{x}^{\,t} G \vec{y} = (A\vec{x}')^t G (A\vec{y}') = (\vec{x}')^t A^t G A \vec{y}',$$

a więc macierz Grama bazy primowanej jest równa $A^t G A$. Analogicznie, w przypadku półtoraliniowej formy jest ona równa $A^t G \overline{A}$.

(1) W polskiej terminologii raczej formą metryczną (*przyp. tłum.*).

W pierwszej części wykładu macierze zasadniczo były wykorzystywane do zapisywania odwzorowań liniowych, nasuwa się więc naturalne pytanie, czy nie ma tu jakiegoś kanonicznego odwzorowania liniowego związanego z g i odpowiadającego macierzy Grama G. Jest tak rzeczywiście, a jego konstrukcja daje równoważną metodę wprowadzenia iloczynu skalarnego.

b) Niech $g: L \times L \to \boldsymbol{K}$ będzie iloczynem skalarnym. Przyporządkujmy każdemu wektorowi $l \in L$ funkcję $g_l: L \to \boldsymbol{K}$, dla której

$$g_l(m) = g(l, m), \quad m \in L.$$

Jest to funkcja liniowa względem m w przypadku dwuliniowym i półliniowa w półtoraliniowym, tj. $g_l \in L^*$ lub $g_l \in \overline{L}^*$ dla każdego l. Ponadto odwzorowanie

$$\tilde{g}: L \to L^* \text{ lub } L \longrightarrow \overline{L}^* : l \mapsto g_l = \tilde{g}(l)$$

jest liniowe, skonstruowane w kanoniczny sposób z g i jednoznacznie wyznacza g poprzez

$$g(l, m) = (\tilde{g}(l), m),$$

gdzie zewnętrzny nawias po prawej stronie oznacza kanoniczne dwuliniowe odwzorowanie $L^* \times L \to \boldsymbol{K}$ lub $\overline{L}^* \times L \to \boldsymbol{C}$.

Odwrotnie, dowolne odwzorowanie liniowe $\tilde{g}: L \to L^*$ (lub $L \to \overline{L}^*$) jednoznacznie wyznacza odpowiadające mu odwzorowanie dwuliniowe $g: L \times L \to \boldsymbol{K}$ (lub $L \times \overline{L} \to \boldsymbol{C}$) za pomocą tego samego wzoru

$$g(l, m) = (\tilde{g}(l), m).$$

Związek z poprzedzającą konstrukcją jest widoczny. *Jeśli w L została wybrana baza $\{e_1, \ldots, e_n\}$ i g wyraża się w tej bazie macierzą G, to macierzą $\tilde{g}$ względem pary baz dwoistych $\{e_1, \ldots, e_n\}$, $\{e^1, \ldots, e^n\}$ jest G^t.*

Rzeczywiście, jeśli $\tilde{g}$ jest wyznaczona przez macierz G^t względem tej pary baz dwoistych, to odpowiadający $\tilde{g}$ iloczyn skalarny g wyraża się poprzez wektory współrzędnych w bazie $\{e_1, \ldots, e_n\}$ wzorem

$$\begin{aligned} g(\vec{x}, \vec{y}) &= (\tilde{g}(\vec{x}), \vec{y}) = (\tilde{g}(\vec{x}))^t \vec{y} \ \left(\text{lub } (\tilde{g}(\vec{x}))^t \overline{\vec{y}} \right) = \\ &= (G^t \vec{x})^t \vec{y} \ \left(\text{lub } (G^t \vec{x})^t \overline{\vec{y}} \right) = \vec{x}^t G \vec{y} \ \left(\text{lub } \vec{x}^t G \overline{\vec{y}} \right), \end{aligned}$$

co dowodzi wypowiedzianego stwierdzenia. Po drodze posłużyliśmy się obserwacją uczynioną w części 1, § 7, że kanoniczne odwzorowanie $L^* \times L \to \boldsymbol{K}$ wyraża się poprzez współrzędne względem pary baz dwoistych wzorem $(\vec{x}, \vec{y}) = \vec{x}^t \vec{y}$.

3. Typy symetrii iloczynów skalarnych. Przestawienie argumentów dwuliniowego iloczynu skalarnego g prowadzi do nowego iloczynu skalarnego, który oznaczymy przez g^t:

$$g^t(l, m) = g(m, l).$$

W półtoraliniowym przypadku ta operacja zamienia także miejsca „liniowego" i „półliniowego" argumentu i jeśli chcemy tego uniknąć, to wygodniej rozważać $\overline{g}^t$:

$$\overline{g}(l, m) = \overline{g(m, l)}.$$

Funkcja $\overline{g}^t$ zależy liniowo od argumentu stojącego na pierwszym miejscu, a półliniowo — na drugim, jeśli taka sama zależność obowiązuje dla g. Odpowiedniość $g \mapsto g^t$, a także $g \mapsto \overline{g}^t$, łatwo opisać za pomocą macierzy Grama: odpowiada ona operacji $G \mapsto G^t$, w drugim przypadku $G \mapsto \overline{G}^t$ (oczywiście zakładamy, że g, g^t, $\overline{g}^t$ są wyrażone poprzez macierze w jednej i tej samej bazie przestrzeni L). Istotnie,

$$g^t(\vec{x}, \vec{y}) = g(\vec{y}, \vec{x}) = \vec{y}^t G\vec{x} = (\vec{y}^t G\vec{x})^t = \vec{x}^t G\vec{y},$$
$$\overline{g}^t(\vec{x}, \vec{y}) = \overline{g(\vec{y}, \vec{x})} = \overline{\vec{y}^t G\vec{x}} = (\overline{\vec{y}^t G\vec{x}})^t = \vec{x}^t \overline{G}^t \vec{y}.$$

Będziemy zajmować się prawie wyłącznie iloczynami skalarnymi, które spełniają jeden z następujących warunków symetrii względem opisanej wyżej operacji:

a) $g = g^t$. Takie iloczyny skalarne nazywamy *symetrycznymi*, a geometrię przestrzeni wyposażonych w symetryczny iloczyn skalarny nazywamy *geometrią ortogonalną*. Symetryczne iloczyny skalarne można zadawać za pomocą symetrycznych macierzy Grama.

b) $g = -g^t$. Takie iloczyny skalarne nazywamy *antysymetrycznymi* lub *symplektycznymi*, a odpowiadającą im geometrię nazywamy *geometrią symplektyczną*. Odpowiadają im antysymetryczne macierze Grama.

Przypadek półtoraliniowy:

c) $\overline{g}^t = g$. Takie iloczyny skalarne nazywamy *hermitowsko symetrycznymi* albo *hermitowskimi*, a odpowiadającą im geometrię — *geometrią hermitowską*. Odpowiadają im hermitowskie macierze Grama. Z warunku $\overline{g}^t = g$ wynika, że $\overline{g}(l, l) = g(l, l)$ dla każdego $l \in L$, tj. wszystkie wartości $g(l, l)$ są rzeczywiste.

Hermitowsko antysymetrycznych iloczynów skalarnych nie potrzeba rozważać osobno, gdyż odwzorowanie $g \mapsto ig$ zadaje bijekcję między nimi a iloczynami skalarnymi hermitowsko symetrycznymi:

$$\overline{g}^t = g \iff (\overline{ig})^t = -ig.$$

Geometryczne własności iloczynów skalarnych, różniących się od siebie tylko czynnikiem skalarnym, są praktycznie te same. Natomiast geometria ortogonalna różni się pod wieloma względami w istotny sposób od geometrii symplektycznej — warunków $g^t = g$ i $g^t = -g$ nie można takim prostym sposobem sprowadzić jeden do drugiego.

4. Ortogonalność. Niech (L, g) będzie przestrzenią wektorową z iloczynem skalarnym g. Wektory l_1, $l_2 \in L$ będziemy nazywać *wektorami ortogonalnymi* (względem g), jeśli $g(l_1, l_2) = 0$. Podprzestrzenie L_1, $L_2 \subset L$ nazwiemy *ortogonalnymi*, jeśli $g(l_1, l_2) = 0$ dla każdych $l_1 \in L_1$, $l_2 \in L_2$. Zasadniczym powodem, dla którego

ważne są przede wszystkim iloczyny skalarne mające jedną z wymienionych własności symetrii jest to, że dla nich *własność ortogonalności wektorów lub podprzestrzeni jest symetryczna względem tych wektorów lub podprzestrzeni.* Istotnie, jeśli $g = \pm g^t$ albo $g = \overline{g}^t$, to

$$g(l,m) = 0 \iff \pm g^t(l,m) = 0 \iff g(m,l) = 0,$$

i analogicznie w hermitowskim przypadku (w związku z odwrotną implikacją por. ćwiczenie 3).

Jeśli nie będzie wyraźnie zaznaczone przeciwnie, to w dalszym ciągu omawiać będziemy wyłącznie symetryczne, hermitowskie lub symplektyczne iloczyny skalarne. Pierwszym zastosowaniem pojęcia ortogonalności jest następująca definicja.

5. Definicja. a) *Jądrem iloczynu skalarnego* g w przestrzeni L nazywamy zbiór wszystkich wektorów $l \in L$ ortogonalnych do każdego wektora przestrzeni L.

b) Iloczyn skalarny g nazywamy *niezdegenerowanym*, jeśli jądro formy g jest trywialne, tj. zawiera tylko wektor zerowy.

Oczywiście jądro formy g jest identyczne z jądrem odwzorowania liniowego

$$\tilde{g} : L \longrightarrow L^* \quad (\text{lub} \quad L \longrightarrow \overline{L}^*),$$

skąd wnioskujemy, że jest podprzestrzenią liniową L. Dlatego też zamiast zadawać *niezdegenerowaną formę* g można zadawać *izomorfizm* $L \to L^*$ (lub $\overline{L}^*$). Ponieważ macierzą $\tilde{g}$ jest macierz G^t transponowana do macierzy Grama bazy L, *niezdegenerowanie* g *jest równoważne z niezdegenerowniem macierzy Grama* dowolnej bazy. W algebrze tensorowej i jej zastosowaniach do geometrii różniczkowej i fizyki bardzo często wykorzystuje się to, że niezdegenerowana symetryczna forma g określa izomorfizm $L \to L^*$ — jest to podstawą techniki „podnoszenia i opuszczania wskaźników".

Rząd g określamy jako wymiar obrazu $\tilde{g}$ lub jako rząd macierzy Grama G.

6. Problem klasyfikacji. Niech (L_1, g_1), (L_2, g_2) będą dwiema przestrzeniami liniowymi z iloczynem skalarnym nad tym samym ciałem $\boldsymbol{K}$. Izomorfizm liniowy $f: L_1 \to L_2$, który zachowuje wartości iloczynów skalarnych dowolnych dwóch wektorów, tzn.

$$g_1(l, l') = g_2(f(l), f(l')) \quad \text{dla każdych } l, l' \in L_1,$$

nazywamy *izometrią* tych przestrzeni, a przestrzenie, między którymi istnieje izometria, nazywamy *przestrzeniami izometrycznymi.* Oczywiście, odwzorowanie identycznościowe jest izometrią, superpozycja izometrii jest izometrią i odwzorowanie liniowe odwrotne do izometrii jest izometrią. W następnym paragrafie opiszemy rozwiązanie problemu klasyfikacji przestrzeni z dokładnością do izometrii, a następnie zbadamy grupy izometrii przestrzeni z nią samą i pokażemy, że obejmują one te grupy klasyczne, które zostały opisane w części 1, § 4.

Klasyczne rozwiązanie problemu klasyfikacji polega na tym, że dowolna przestrzeń z iloczynem skalarnym może być zapisana jako suma prosta parami ortogonalnych podprzestrzeni małych wymiarów (jednowymiarowych w ortogonalnym i hermitowskim przypadku, jedno lub dwuwymiarowych w przypadku symplektycznym). Z tego względu zakończymy ten paragraf jawnym opisem przestrzeni małych wymiarów z metryką.

7. Jednowymiarowe przestrzenie ortogonalne. Niech $\dim L = 1$ i niech g będzie symetrycznym iloczynem skalarnym w L. Wybierzmy dowolny niezerowy wektor $l \in L$ — jeśli $g(l, l) = 0$, to $g \equiv 0$, a więc g jest zdegenerowany, a nawet zerowy. Jeśli $g(l, l) = a \neq 0$, to dla dowolnego $x \in \boldsymbol{K}$ mamy $g(xl, xl) = ax^2$, a więc wartości $g(l, l)$ dla wszystkich niezerowych wektorów w L stanowią warstwę multyplikatywnej grupy $\boldsymbol{K}^* = \boldsymbol{K} \setminus \{0\}$ ciała $\boldsymbol{K}$ względem podgrupy składającej się z kwadratów elementów $\boldsymbol{K}^*$: $\{ax^2 \mid x \in \boldsymbol{K}^*\} \in \boldsymbol{K}^*/(\boldsymbol{K}^*)^2$. *Ta warstwa całkowicie charakteryzuje niezdegenerowany symetryczny iloczyn skalarny w jednowymiarowej przestrzeni L:* dla (L_1, g_1) i (L_2, g_2) dwie takie warstwy są identyczne wtedy i tylko wtedy, gdy te przestrzenie są izometryczne. Istotnie, jeśli $g_1(l_1, l_1) = ax^2$ i $g_2(l_2, l_2) = ay^2$, gdzie $l_i \in L_i$, to odwzorowanie $f\colon l_1 \mapsto y^{-1}xl_2$ jest izometrią przestrzeni L_1 z L_2, co dowodzi dostateczności warunku. Konieczność jest oczywista.

Wobec $\boldsymbol{R}^*/(\boldsymbol{R}^*)^2 = \{\pm 1\}$ i $\boldsymbol{C}^* = (\boldsymbol{C}^*)^2$ otrzymujemy następujące ważne szczególne przypadki powyższej klasyfikacji.

Dowolna jednowymiarowa przestrzeń ortogonalna nad $\boldsymbol{R}$ jest izometryczna z jednowymiarową przestrzenią kartezjańską wyposażoną w jeden z trzech iloczynów skalarnych: xy, $-xy$, 0.

Dowolna jednowymiarowa przestrzeń ortogonalna nad $\boldsymbol{C}$ jest izometryczna z jednowymiarową przestrzenią kartezjańską wyposażoną w jeden z dwóch iloczynów skalarnych: xy, 0.

8. Jednowymiarowe przestrzenie hermitowskie. W tym przypadku rozważania są analogiczne. Ciałem podstawowym jest $\boldsymbol{C}$; zdegenerowanie formy jest równoważne z jej zerowaniem się. Jeśli forma jest niezdegenerowana, to zbiór wartości $g(l, l)$ dla niezerowych wektorów $l \in L$ jest warstwą podgrupy $\boldsymbol{R}_+^* = \{x \in \boldsymbol{R}^* \mid x > 0\}$ w grupie $\boldsymbol{C}^*$, bo $g(al, al) = a\bar{a}g(l, l) = |a|^2 g(l, l)$, przy czym $|a|^2$ przyjmuje wszystkie wartości z $\boldsymbol{R}_+^*$, gdy a przebiega $\boldsymbol{C}^*$. Każdą różną od zera liczbę zespoloną można jednoznacznie zapisać w postaci $re^{i\varphi}$, gdzie $r \in \boldsymbol{R}^*$, a $e^{i\varphi}$ należy do zespolonego okręgu jednostkowego, który oznaczymy przez

$$\boldsymbol{C}_1^* = \{z \in \boldsymbol{C} \mid |z| = 1\}.$$

W języku teorii grup wyraża się to mówiąc o rozkładzie na iloczyn prosty $\boldsymbol{C}^* = \boldsymbol{R}_+^* \times \boldsymbol{C}_1^*$ i izomorfizmie $\boldsymbol{C}^*/\boldsymbol{R}_+^*$ z $\boldsymbol{C}_1^*$. W ten sposób pokazaliśmy, że niezdegenerowane półtoraliniowe formy można klasyfikować liczbami zespolonymi o module 1. Jednakże nie została jeszcze uwzględniona własność hermitowskiej symetrii, z której wynika, że $g(l, l) = \overline{g(l, l)}$, tj. że każda wartość $g(l, l)$ jest rzeczywista. Dlatego formom

hermitowskim odpowiadają tylko liczby $\{\pm 1\}$ w C_1^*, tak samo jak w przypadku ortogonalnym nad $\boldsymbol{R}$. Końcowy rezultat brzmi:

Dowolna jednowymiarowa przestrzeń hermitowska nad $\boldsymbol{C}$ jest izometryczna z jednowymiarową przestrzenią kartezjańską wyposażoną w jeden z trzech następujących iloczynów skalarnych: $x\overline{y}$, $-x\overline{y}$, 0.

Jednowymiarowe przestrzenie ortogonalne nad $\boldsymbol{R}$ (lub hermitowskie nad $\boldsymbol{C}$) z iloczynem skalarnym przyjmującym w odpowiednio dobranej bazie postać xy, $-xy$, 0 (albo $x\overline{y}$, $-x\overline{y}$, 0) będziemy nazywać odpowiednio *dodatnimi*, *ujemnymi* i *zerowymi*. Iloczyny skalarne dowolnego niezerowego wektora z nim samym są w takiej przestrzeni odpowiednio zawsze liczbami dodatnimi, ujemnymi albo zerem.

9. Jednowymiarowe przestrzenie symplektyczne. Tu napotykamy na nowe zjawisko — każda antysymetryczna forma na przestrzeni wymiaru 1 nad ciałem charakterystyki $\neq 2$ jest tożsamościowo równa zeru, a w szczególności zdegenerowana! Rzeczywiście,

$$g(l, l) = -g(l, l) \Rightarrow 2g(l, l) = 0,$$

$$g(al, bl) = abg(l, l) = 0.$$

W przypadku charakterystyki 2 warunek antysymetrii $g(l, m) = -g(m, l)$ jest równoważny z warunkiem symetrii $g(l, m) = g(m, l)$, a więc nad takimi ciałami geometria symplektyczna nie różni się od geometrii ortogonalnej. Jednakże geometria ortogonalna ma w tym przypadku pewne osobliwości i zazwyczaj będziemy ją wykluczać z naszych rozważań.

Jest zatem zrozumiałe, że jednowymiarowe przestrzenie symplektyczne nie mogą być tymi składnikami, z których są zbudowane ogólne przestrzenie symplektyczne, a zatem należy pójść przynajmniej o krok dalej.

10. Dwuwymiarowe przestrzenie symplektyczne. Niech (L, g) będzie przestrzenią dwuwymiarową z antysymetryczną formą g nad ciałem $\boldsymbol{K}$ charakterystyki $\neq 2$. Jeśli forma jest zdegenerowana, to jest automatycznie zerowa. I rzeczywiście, niech $l \neq 0$ będzie takim wektorem, że $g(l, m) = 0$ dla każdego $m \in L$. Niech $\{l, l'\}$ będzie bazą L zawierającą l. Biorąc pod uwagę, że $g(l, l) = g(l', l') = 0$ na mocy poprzedniego punktu, mamy dla każdych a, a', b, $b' \in \boldsymbol{K}$

$$g(al + a'l', bl + b'l') = abg(l, l) + ab'g(l, l') - a'bg(l, l') + bb'g(l', l') = 0.$$

Załóżmy teraz, że g jest niezerowa, a więc niezdegenerowana. Zatem istnieje para wektorów e_1, e_2, dla których $g(e_1, e_2) = a \neq 0$, a nawet takich, że $a = 1$; istotnie, $g(a^{-1}e_1, e_2) = a^{-1}a = 1$.

Niech teraz $g(e_1, e_2) = 1$. Wówczas wektory e_1, e_2 są liniowo niezależne, a więc tworzą bazę L — jeśli bowiem np. $e_1 = ae_2$, to $g(ae_2, e_2) = ag(e_2, e_2) = 0$. Zatem iloczyn skalarny g wyraża się poprzez współrzędne względem tej bazy wzorem

$$g(x_1e_1 + x_2e_2, y_1e_1 + y_2e_2) = x_1y_2 - x_2y_1,$$

a jego macierzą Grama jest

$$G = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Ostatecznie otrzymujemy:

Dowolna dwuwymiarowa przestrzeń symplektyczna nad ciałem $\boldsymbol{K}$ charakterystyki $\neq 2$ jest izometryczna z przestrzenią kartezjańską $\boldsymbol{K}^2$ wyposażoną w jeden z dwóch iloczynów skalarnych $x_1y_2 - x_2y_1$ albo 0.

Ćwiczenia

1. Niech L, M będą skończenie wymiarowymi przestrzeniami nad ciałem $\boldsymbol{K}$ i $g\colon L \times M \to \boldsymbol{K}$ — odwzorowaniem dwuliniowym. Nazwijmy *lewostronnym jądrem* g zbiór $L_0 = \{l \in L \mid g(l, m) = 0 \text{ dla każdego } m \in M\}$, a *prawostronnym jądrem* — zbiór $M_0 = \{m \in M \mid g(l, m) = 0 \text{ dla każdego } l \in L\}$. Wykazać następujące stwierdzenia:

a) $\dim L/L_0 = \dim M/M_0$.

b) g indukuje dwuliniowe odwzorowanie $g'\colon L/L_0 \times M/M_0 \to \boldsymbol{K}$, którego lewostronne i prawostronne jądro jest zerowe.

2. Wykazać, że każdy dwuliniowy iloczyn skalarny $g\colon L \times L \to \boldsymbol{K}$ (nad ciałem charakterystyki $\neq 2$) jednoznacznie rozkłada się na sumę symetrycznego i antysymetrycznego iloczynu skalarnego.

3. Niech $g\colon L \times L \to \boldsymbol{K}$ będzie takim dwuliniowym iloczynem skalarnym, że własność ortogonalności wektorów jest symetryczna, tj. z $g(l_1, l_2) = 0$ wynika $g(l_2, l_1) = 0$. Wykazać, że wtedy g jest albo symetryczny, albo antysymetryczny. (*Wskazówki.* a) Niech $l, m, n \in L$. Wykazać, że $g(l, g(l, n)m - g(l, m)n) = 0$. Wykorzystując symetrię relacji ortogonalności, wykazać dalej, że $g(l, n)g(m, l) = g(n, l)g(l, m)$. b) Przyjmując $l = n$, wyprowadzić stąd, że jeśli $g(m, l) \neq g(l, m)$, to $g(l, l) = 0$. c) Wykazać, że $g(n, n) = 0$ dla każdego $n \in L$, jeśli g jest niesymetryczny. W tym celu należy wybrać takie l, m, że $g(m, l) \neq g(l, m)$ i rozważyć osobno przypadki $g(l, n) \neq g(n, l)$, $g(l, n) = g(n, l)$. d) Wykazać, że jeśli $g(n, n) = 0$ dla wszystkich $n \in L$, to g jest antysymetryczny.)

4. Podać klasyfikację jednowymiarowych przestrzeni ortogonalnych nad ciałem skończonym $\boldsymbol{K}$ charakterystyki $\neq 2$, wykazując, że $\boldsymbol{K}^*/(\boldsymbol{K}^*)^2$ jest grupą cykliczną rzędu 2. (*Wskazówka.* Pokazać, że jądro homomorfizmu $\boldsymbol{K}^* \to \boldsymbol{K}^* : x \mapsto x^2$ jest podgrupą o indeksie 2, wykorzystując to, że dowolny wielomian ma w ciele nie więcej pierwiastków, niż wynosi jego stopień.)

5. Niech (L, g) będzie n-wymiarową przestrzenią liniową z niezdegenerowanym iloczynem skalarnym. Wykazać, że układ wektorów $\{e_1, \ldots, e_n\}$ przestrzeni L jest liniowo niezależny wtedy i tylko wtedy, gdy macierz $(g(e_i, e_j))$ jest niezdegenerowana.

§ 3. Twierdzenia klasyfikacyjne

1. Zasadniczym celem tego paragrafu jest przedstawienie klasyfikacji — z dokładnością do izomorfizmu — skończenie wymiarowych przestrzeni ortogonalnych, hermitowskich lub symplektycznych. Niech (L, g) będzie przestrzenią tego typu i $L_0 \subset L$ podprzestrzenią liniową. Obcięcie g do L_0 jest iloczynem skalarnym w L_0 i jeśli jest on niezdegenerowany, to L_0 nazywamy *podprzestrzenią niezdegenerowaną*, a jeśli obcięcie g do L_0 jest formą zerową, to nazywamy ją *podprzestrzenią osobliwą*. Jest przy tym istotne, że nawet jeśli L jest przestrzenią niezdegenerowaną, to obcięcia g do nietrywialnych podprzestrzeni mogą być zdegenerowane, a nawet zerowe. Na przykład, w przypadku symplektycznym wszystkie podprzestrzenie jednowymiarowe są zdegenerowane, a w przypadku ortogonalnej przestrzeni $\boldsymbol{R}^2$ z iloczynem $x_1y_1 - x_2y_2$ podprzestrzeń rozpięta przez wektor $(1, 1)$ jest zdegenerowana(1).

Dopełnieniem ortogonalnym podprzestrzeni $L_0 \subset L$ nazywamy zbiór

$$L_0^{\perp} = \{l \in L \mid g(l_0, l) = 0 \quad \text{dla każdego } l_0 \in L_0\}$$

(nie należy mylić z wprowadzonym w pierwszej części dopełnieniem ortogonalnym L_0, które jest podprzestrzenią L^* i którego nie będziemy tu używać). Łatwo zauważyć, że $L_0^{\perp}$ jest podprzestrzenią liniową L_0.

2. Stwierdzenie. *Załóżmy, że* (L, g) *ma wymiar skończony.*
a) *Jeśli podprzestrzeń* $L_0 \subset L$ *jest niezdegenerowana, to* $L = L_0 + L_0^{\perp}$.
b) *Jeśli obie podprzestrzenie* L_0 *i* $L_0^{\perp}$ *są niezdegenerowane, to* $(L_0^{\perp})^{\perp} = L_0$.

Dowód. a) Niech $\tilde{g}: L \to L^*$ (lub $(\overline{L^*})$ będzie odwzorowaniem wyznaczonym przez g, tak jak w poprzedzającym paragrafie. Oznaczmy przez $\tilde{g}_0$ jego ograniczenie do L_0, $\tilde{g}_0: L_0 \to L_0^*$ (lub $\overline{L^*}$). Jeśli L_0 jest niezdegenerowana, to $\operatorname{Ker} \tilde{g}_0 = 0$, gdyż inaczej w L_0 znalazłby się wektor ortogonalny do całej przestrzeni L, więc tym bardziej do L_0. Zatem $\dim \operatorname{Im} \tilde{g}_0 = \dim L_0$. To oznacza, że dla l_0 przebiegającego L_0 formy liniowe $g(l_0, \cdot)$ (argumentem formy jest wektor z L lub $\overline{L}$) stanowią przestrzeń form liniowych na L lub $\overline{L}$ wymiaru $\dim L_0$. Ponieważ $L_0^{\perp}$ jest częścią wspólną jąder tych form, więc $\dim L_0^{\perp} = \dim L - \dim L_0$, tj.

$$\dim L_0 + \dim L^{\perp} = \dim L.$$

Z drugiej strony, z niezdegenerowania L_0 wynika, że $L_0 \cap L_0^{\perp} = \{0\}$, bo $L_0 \cap L_0^{\perp}$ jest jądrem obcięcia g do L_0. Dlatego $L_0 + L_0^{\perp}$ jest sumą prostą, a ponieważ jej wymiar jest równy $\dim L$, więc $L_0 \oplus L_0^{\perp} = L$.

b) Z określenia jest jasne, że $L_0 \subset (L_0^{\perp})^{\perp}$, a z drugiej strony, jeśli L_0 i $L_0^{\perp}$ są niezdegenerowane, to z poprzedzającego rozumowania otrzymujemy

$$\dim(L_0^{\perp})^{\perp} = \dim L - \dim L_0^{\perp} = \dim L_0,$$

co kończy dowód.

(1) Nawet osobliwa (*przyp. tłum.*).

3. Twierdzenie. *Niech (L, g) będzie skończenie wymiarową przestrzenią ortogonalną (nad ciałem charakterystyki $\neq 2$), hermitowską lub symplektyczną. Wtedy istnieje rozkład L na sumę prostą parami ortogonalnych podprzestrzeni*

$$L = L_1 \oplus \cdots \oplus L_m,$$

jednowymiarowych w przypadku ortogonalnym i hermitowskim, a jednowymiarowych zdegenerowanych lub dwuwymiarowych niezdegenerowanych w przypadku symplektycznym.

Dowód. Przeprowadzimy indukcję względem $\dim L$. Przypadek $\dim L = 1$ jest trywialny, niech więc $\dim L \geq 2$. Jeśli g jest formą zerową, to nie ma nic do dowodzenia. Jeśli g jest niezerowa, to w przypadku symplektycznym istnieje para wektorów $l_1, l_2 \in L$, dla których $g(l_1, l_2) \neq 0$. Zgodnie z p. 10 poprzedniego paragrafu podprzestrzeń L_0 rozpięta na tych wektorach jest niezdegenerowana. Na mocy stwierdzenia p. 2 $L = L_0 \oplus L_0^{\perp}$, a na mocy założenia indukcyjnego można dalej rozłożyć $L_0^{\perp}$ jak w tezie twierdzenia, co prowadzi do żądanego rozkładu L.

Dla ortogonalnego i hermitowskiego przypadku wykażemy, że z niezdegenerowania g wynika istnienie niezdegenerowanej jednowymiarowej podprzestrzeni L_0, a sprawdziwszy to, będziemy mogli przyjąć $L = L_0 \oplus L_0^{\perp}$ i zastosować poprzednie rozumowanie, tj. indukcję względem wymiaru L.

W samej rzeczy, przypuśćmy, że $g(l, l) = 0$ dla wszystkich $l \in L$. Wówczas $g \equiv 0$, bo dla każdych $l_1, l_2 \in L$ mamy

$$0 = g(l_1 + l_2, l_1 + l_2) = g(l_1, l_1) + 2g(l_1, l_2) + g(l_2, l_2) = 2g(l_1, l_2)$$

lub

$$0 = g(l_1 + l_2, l_1 + l_2) = g(l_1, l_1) + 2\operatorname{Re} g(l_1, l_2) + g(l_2, l_2) = 2\operatorname{Re} g(l_1, l_2).$$

W przypadku ortogonalnym wynika stąd od razu $g(l_1, l_2) = 0$, natomiast w przypadku hermitowskim otrzymujemy tylko $\operatorname{Re} g(l_1, l_2) = 0$, tj. $g(l_1, l_2) = ia$, $a \in \boldsymbol{R}$. Jednak jeśli $a \neq 0$, to także

$$0 = \operatorname{Re} g((ia)^{-1} l_1, l_2) = \operatorname{Re}(ia)^{-1} g(l_1, l_2) = 1.$$

Otrzymana sprzeczność dowodzi twierdzenia.

Przejdźmy obecnie do zagadnienia jednoznaczności. Otrzymany w powyższym twierdzeniu rozkład na ortogonalną sumę prostą jest w dużym stopniu niejednoznaczny, poza trywialnymi przypadkami wymiaru 1 (i 2 w przypadku symplektycznym). W przypadku geometrii ortogonalnej nad ogólnymi ciałami niejednoznacznie wyznaczony jest nawet zbiór niezmienników $a_i \in \boldsymbol{K}^*/(\boldsymbol{K}^*)^2$, które charakteryzują ograniczenia g na jednowymiarowe przestrzenie L_i. Dokładne rozwiązanie problemu klasyfikacji przestrzeni ortogonalnych w istotny sposób zależy od własności ciała

podstawowego; np. dla $K = Q$ jest związane z takimi subtelnymi zagadnieniami teorioliczbowej natury jak prawo wzajemności kwadratowej. Dlatego w przypadku ortogonalnym ograniczymy się tylko do opisania wyniku dla $K = R$ i C (pewne dalsze szczegóły można znaleźć w § 14).

4. Niezmienniki przestrzeni z metryką. Niech (L, g) będzie przestrzenią z iloczynem skalarnym. Oznaczmy $n = \dim L$, $r_0 = \dim L_0$, gdzie przez L_0 oznaczyliśmy jądro g. Poza tym wprowadzimy jeszcze dwa dodatkowe niezmienniki, odnoszące się tylko do przypadków geometrii ortogonalnej nad ciałem R i geometrii hermitowskiej, a mianowicie r_+ oraz r_-, oznaczające liczbę dodatnich, odpowiednio ujemnych, jednowymiarowych podprzestrzeni L_i w rozkładzie L na ortogonalną sumę prostą danym w twierdzeniu p. 3.

Mamy oczywiście $r_0 \leq n$ i $n = r_0 + r_+ + r_-$ dla geometrii hermitowskiej i ortogonalnej nad R. Układ (r_0, r_+, r_-) będziemy nazywać *sygnaturą przestrzeni.* Przy $r_0 = 0$ sygnaturą nazywamy nieraz także (r_+, r_-) albo nawet $r_+ - r_-$ (traktując $n = r_+ + r_-$ jako znane).

Możemy teraz sformułować twierdzenie o jednoznaczności.

5. Twierdzenie. a) *Przestrzenie symplektyczne nad dowolnym ciałem, a także przestrzenie ortogonalne nad C są określone z dokładnością do izometrii przez dwie liczby całkowite n, r_0, tj. wymiary przestrzeni i jądra iloczynu skalarnego.*

b) *Przestrzenie ortogonalne nad ciałem R i przestrzenie hermitowskie nad C są określone z dokładnością do izometrii przez sygnaturę (r_0, r_+, r_-), która nie zależy od wyboru rozkładu ortogonalnego (ta ostatnia część twierdzenia nazywa się twierdzeniem*[1]) *o bezwładności formy).*

Dowód. a) Niech (L, g) będzie przestrzenią symplektyczną lub ortogonalną nad ciałem C. Rozważmy rozkład na sumę prostą $L = \bigoplus_{i=1}^{n} L_i$, opisany w twierdzeniu p. 3. Wykażemy, że r_0 jest równa liczbie jednowymiarowych zdegenerowanych względem g podprzestrzeni występujących w tym rozkładzie, a nawet, że suma L_0 tych przestrzeni jest jądrem g. Rzeczywiście, jest jasne, że ta suma jest zawarta w jądrze g, bo elementy L_0 są ortogonalne zarówno do L_0, jak i do pozostałych składników tej sumy. Z drugiej strony , jeśli $L_0 = \bigoplus_{i=1}^{r_0} L_i$ oraz

$$l = \sum_{j=1}^{n} l_j, \quad l_j \in L_j, \quad \exists\, j > r_0,\ l_j \neq 0,$$

to wówczas w przypadku ortogonalnym

$$g(l, l_j) = g(l_j, l_j) \neq 0,$$

(1) Sylvestera (*przyp. tłum.*).

a w przypadku symplektycznym istnieje taki wektor $l'_j \in L_j$, że

$$g(l, l'_j) = g(l_j, l'_j) \neq 0.$$

W przeciwnym przypadku jądro ograniczenia g do L_j byłoby nietrywialne, więc g byłaby formą zerową na L_j na mocy § 2, p. 10, co byłoby sprzeczne z tym, że $j > r_0$. Zatem $l \notin$ (jądro g), więc $L_0 =$ (jądro g).

Niech teraz (L, g) i (L', g') będą dwiema takimi, jak założyliśmy, przestrzeniami o jednakowych n i r_0. Konstruując dla nich rozkłady na ortogonalne sumy proste $L = \bigoplus_{i=1}^{n} L_i$ i $L' = \bigoplus_{i=1}^{n} L'_i$, dla których (jądro g) $= \bigoplus_{i=1}^{r_0} L_i$ oraz (jądro g') $= \bigoplus_{i=1}^{r_0} L'_i$, możemy zdefiniować izometrię (L, g) z (L', g') jako sumę prostą $\bigoplus f_i$ izometrii $f_i: L_i \to L'_i$, których istnienie jest zagwarantowane rezultatami § 2, p. p. 7, 10.

b) Niech teraz (L, g) i (L', g') będą ortogonalnymi przestrzeniami nad $\boldsymbol{R}$ lub hermitowskimi przestrzeniami nad $\boldsymbol{C}$ o sygnaturach (r_0, r_+, r_-) i (r'_0, r'_+, r'_-) określonych za pomocą rozkładów ortogonalnych $L = \bigoplus L_i$, $L' = \bigoplus L'_i$, o których była mowa w twierdzeniu p. 3. Załóżmy istnienie izometrii między tymi przestrzeniami. Wówczas po pierwsze $\dim L = \dim L'$, a więc $r_0 + r_+ + r_- = r'_0 + r'_+ + r'_-$. Dalej, dokładnie tak samo jak w poprzednim punkcie, można sprawdzić, że r_0 jest wymiarem jądra g, a r'_0 — wymiarem jądra g', a same te przestrzenie są sumami przestrzeni zerowych L_i i L'_i w odpowiednich rozkładach. Z założenia o istnieniu izometrii wnioskujemy, że jądra g i g' są liniowo izomorficzne, skąd dalej wynika, że $r_0 = r'_0$ i $r_+ + r_- = r'_+ + r'_-$. Pozostaje sprawdzić, że $r_+ = r'_+$ oraz $r_- = r'_-$. Niech $L = L_0 \oplus L_+ \oplus L_-$, $L' = L'_0 \oplus L'_+ \oplus L'_-$, gdzie przez L_0, L_+, L_- oznaczyliśmy sumy zerowych, dodatnich i ujemnych podprzestrzeni w wyjściowym rozkładzie L i podobnie dla L'. Wykażemy teraz, że założenie $r_+ = \dim L_+ > r'_+ = \dim L'_+$ prowadzi do sprzeczności — w podobny sposób można wykluczyć możliwość $r_+ < r'_+$. Ograniczmy izometrię $f: L \to L'$ do $L_+ \subset L$. Każdy wektor $f(l)$ można jednoznacznie zapisać jako sumę

$$f(l) = f(l)_0 + f(l)_+ + f(l)_-,$$

gdzie $f(l)_+ \in L_+$, itd. Odwzorowanie $L_+ \to L_+ : l \mapsto f(l)_+$ jest liniowe. Z założonej nierówności $\dim L_+ > \dim L'_+$ wynika, że istnieje niezerowy wektor $l \in L_+$, dla którego $f(l)_+ = 0$, a zatem

$$f(l) = f(l)_0 + f(l)_-.$$

Jednakże $g(l, l) > 0$, ponieważ $l \in L_+$, a L_+ jest sumą prostą ortogonalną dodatnich jednowymiarowych podprzestrzeni. Ponieważ f jest izometrią, więc powinniśmy mieć także $g'(f(l), f(l)) > 0$. Z drugiej strony,

$$g'(f(l), f(l)) = g'(f(l)_0 + f(l), f(l)_0 + f(l)_-) = g'(f(l)_-, f(l)_-) \leq 0.$$

Ta sprzeczność kończy dowód tego, że sygnatury izometrycznych przestrzeni, wyznaczone w dowolnych bazach ortogonalnych, są jednakowe. Odwrotnie, jeśli (L, g),

(L', g') są dwiema przestrzeniami z jednakowymi sygnaturami i ortogonalnymi rozkładami $L = \bigoplus L_i$, $L' = \bigoplus L'_i$, to między podprzestrzeniami występującymi w ich ortogonalnych rozkładach można określić wzajemnie jednoznaczną odpowiedniość $L_i \longleftrightarrow L'_i$, która zachowuje znaki obcięcia g do L_i i g' do L'_i odpowiednio. Na mocy rezultatów §2 p. p. 7, 8 istnieją izometrie $f_i: L_i \to L'_i$, a ich suma prosta $\bigoplus f_i$ jest izometrią L na L'.

Teraz wyprowadzimy pewne wnioski i podamy inne sformułowania twierdzeń z p. p. 3, 5, aby wydobyć różnorodne aspekty rozważanej sytuacji.

6. Bazy. Niech (L, g) będzie przestrzenią z iloczynem skalarnym. Bazę $\{e_1, \ldots, e_n\}$ przestrzeni L nazwiemy *bazą ortogonalną* (*prostopadłą*), jeśli $g(e_i, e_j) = 0$ dla każdej pary $i \neq j$. Z twierdzenia p. 5 wynika, że *każda przestrzeń ortogonalna lub hermitowska posiada bazę ortogonalną.* Istotnie, wystarczy skonstruować rozkład $L = \bigoplus L_i$ na jednowymiarowe ortogonalne podprzestrzenie i wybrać $e_i \in L_i$, $e_i \neq 0$.

Bazę ortogonalną nazywamy *ortonormalną*, jeśli $g(e_i, e_i) = 0, \pm 1$ dla każdego i. Rozważania z końca § 2 pokazują, że *każda przestrzeń ortogonalna nad R lub C i każda przestrzeń hermitowska posiada bazę ortonormalną.* Twierdzenie p. 5 pokazuje, że liczby elementów bazy ortonormalnej, dla których $g(e_i, e_i) = 0, 1$ lub -1 nie zależą od bazy dla $\boldsymbol{K} = \boldsymbol{R}$ (przypadek ortogonalny) i $\boldsymbol{K} = \boldsymbol{C}$ (przypadek hermitowski). W przypadku ortogonalnym nad ciałem $\boldsymbol{C}$ zawsze można doprowadzić do tego, aby $g(e_i, e_i) = 0$ lub 1, a liczby odpowiednich wektorów dla danej bazy nie zależały od jej wyboru. Macierz Grama bazy ortonormalnej ma postać

$$\left(\begin{array}{c|c|c} E_{r_+} & 0 & 0 \\ \hline 0 & -E_{r_-} & 0 \\ \hline 0 & 0 & 0 \end{array}\right),$$

(przy właściwym uporządkowaniu). Najczęściej pojęcia bazy ortonormalnej używa się w niezdegenerowanym przypadku, gdy nie występują wektory e_i, dla których $g(e_i, e_i) = 0$. Następująca ważna, pomimo swojej prostoty, formuła pozwala napisać jawną postać współczynników rozkładu dowolnego wektora $e \in L$ w bazie ortonormalnej (w przypadku niezdegenerowanym):

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

Rzeczywiście, iloczyny skalarne wyrażeń z lewej i prawej strony z każdym wektorem e_i są jednakowe, a z niezdegenerowania formy g wynika, że jeśli $g(e, e_i) = g(e', e_i)$ dla każdego i, to $e = e'$, bo $e - e'$ należy do jądra formy g.

W przestrzeni symplektycznej baza ortogonalna może istnieć tylko wtedy, gdy $g = 0$. Jednak twierdzenie p. 3 zapewnia istnienie *bazy symplektycznej* $\{e_1, e_2, \ldots,$

$e_r; e_{r+1}, \ldots, e_{2r}; e_{2r+1}, \ldots, e_n\}$, scharakteryzowanej spełnieniem równości

$$g(e_i, e_{r+i}) = -g(e_{r+i}, e_i) = 1, \quad i = 1, \ldots, r,$$

i równością zeru iloczynów skalarnych wszystkich pozostałych par wektorów bazy. Rzeczywiście, wystarczy rozłożyć L na sumę prostą ortogonalną dwuwymiarowych niezdegenerowanych podprzestrzeni L_i, $1 \leq i \leq r$, jednowymiarowych zdegenerowanych L_j, $2r \leq j \leq n$, jako $\{e_i, e_{i+r}\}$, $1 \leq i \leq r$, przyjąć bazę przestrzeni L_i, skonstruowaną w § 2, p. 10, a jako e_j dla $2r \leq j \leq n$ wziąć dowolny niezerowy wektor przestrzeni L_j. Macierz Grama bazy symplektycznej ma postać

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

Na podstawie twierdzenia p. 5 rząd formy symplektycznej jest równy $2r$. W szczególności, *niezdegenerowana przestrzeń symplektyczna ma wymiar parzysty.*

Niech L będzie niezdegenerowaną przestrzenią symplektyczną, a $\{e_1, e_2, \ldots, e_r; e_{r+1}, \ldots, e_{2r}\}$ — jej bazą symplektyczną. Niech L_1 będzie powłoką liniową $\{e_1, e_2, \ldots, e_r\}$, a L_2 — powłoką liniową $\{e_{r+1}, \ldots, e_{2r}\}$. Oczywiście przestrzenie L_1 i L_2 są osobliwe, wymiar każdej z nich jest połową wymiaru L oraz $L = L_1 \oplus L_2$. Odwzorowanie kanoniczne $\tilde{g}: L \to L^*$ wyznacza odwzorowanie

$$\widetilde{g_1}: L_2 \to L_1^*, \quad \widetilde{g_1}(l_2)(l_1) = g(l_2, l_1).$$

Jest ono izomorfizmem, bo $\dim L_2 = \dim L_1 = \dim L_1^*$ i $\operatorname{Ker} \widetilde{g_1} = \{0\}$: wektor należący do $\operatorname{Ker} \widetilde{g_1}$ jest ortogonalny do L_2, bo L_2 jest osobliwe, i do L_1 z konstrukcji, a L jest niezdegenerowana.

Stąd wynika, że *każda niezdegenerowana przestrzeń symplektyczna jest izometryczna z przestrzenią postaci* $L = L_1^* \oplus L_1$ *z formą symplektyczną*

$$g((f, l), (f', l')) = f(l') - f'(l), \quad f, f' \in L_1^*, \; l, l' \in L_1.$$

Dalsze szczegóły podajemy w § 12.

7. Macierze. Opisując iloczyny skalarne za pomocą macierzy Grama i przechodząc od dowolnie wybranej bazy do bazy ortogonalnej lub symplektycznej, otrzymujemy, opierając się na rezultatach § 2, p. p. 2, 6 tego paragrafu, następujące fakty:

a) Każdą macierz kwadratową symetryczną G nad ciałem $\boldsymbol{K}$ można sprowadzić do postaci diagonalnej za pomocą przekształcenia $G \mapsto A^t G A$, gdzie A jest nieosobliwa. Dla $\boldsymbol{K} = \boldsymbol{R}$ można osiągnąć i to, aby na diagonali występowały wyłącznie 0 i ± 1, a dla $\boldsymbol{K} = \boldsymbol{C}$ tylko 0 i 1; liczby elementów równych 0 i ± 1 (0 i 1 odpowiednio) zależą tylko od G, a nie zależą od A.

b) Każdą antysymetryczną macierz kwadratową G nad ciałem $\boldsymbol{K}$ charakterystyki $\neq 2$ można sprowadzić za pomocą przekształcenia $G \mapsto A^tGA$, gdzie A jest nieosobliwa, do postaci

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

Liczba $2r$ jest równa rzędowi G.

c) Każdą macierz hermitowską G nad $\boldsymbol{C}$ można sprowadzić do postaci diagonalnej z diagonalnymi wyrazami równymi 0 i ± 1 za pomocą przekształcenia $G \mapsto A^tG\overline{A}$, gdzie A nieosobliwa. Liczby elementów na diagonali równych odpowiednio 0 i ± 1 zależą tylko od G.

8. Formy dwuliniowe. Jeśli wektory przestrzeni (L, g) z ustaloną bazą zadawać współrzędnymi w tej bazie, to g wyraża się przez te współrzędne jako forma dwuliniowa $2n$ zmiennych, $n = \dim L$, wzorem

$$g(x_1, \ldots, x_n; y_1, \ldots, y_n) = \sum_{i,j=1}^{n} g_{ij}x_iy_j = \vec{x}^{\,t}G\vec{y},$$

gdzie G jest macierzą Grama tej bazy. Zmiana bazy prowadzi do liniowej zamiany zmiennych $x_1, \ldots, x_n$ i $y_1, \ldots, y_n$ za pomocą tej samej macierzy nieosobliwej A w przypadku dwuliniowym (a macierzy A dla $\vec{x}$ i macierzy $\overline{A}$ dla $\overline{\vec{y}}$ w przypadku półtoraliniowym). Dotychczas wykazane rezultaty oznaczają, że w zależności od symetrii macierzy G formę można doprowadzić za pomocą takiego przekształcenia zmiennych do jednej z następujących postaci, zwanych kanonicznymi.

Przypadek symetryczny nad dowolnym ciałem:

$$g(\vec{x}, \vec{y}) = \sum_{i=1}^{n} a_ix_iy_i,$$

co więcej, nad ciałem $\boldsymbol{R}$ można doprowadzić, aby $a_i = 0$ albo ± 1; nad ciałem $\boldsymbol{C}$ — aby $a_i = 0$ albo 1.

Przypadek hermitowski (forma półtoraliniowa):

$$g(\vec{x}, \vec{y}) = \sum_{i=1}^{n} a_ix_i\overline{y_i},$$

gdzie $a_i = 0$ albo 1.

Przypadek symplektyczny: $n = 2r + r_0$ i forma ma postać

$$g(\vec{x}, \vec{y}) = \sum_{i=1}^{r}(x_iy_{r+i} - y_ix_{r+i}).$$

9. Formy kwadratowe. Odwzorowanie $q: L \to \boldsymbol{K}$ nazywamy *formą kwadratową* na przestrzeni L, jeśli istnieje taka forma dwuliniowa $h: L \times L \to \boldsymbol{K}$, że

$$g(l) = h(l, l) \quad \text{dla każdego } l \in L.$$

Wykażemy, że jeśli charakterystyka ciała $\boldsymbol{K}$ jest $\neq 2$, to dla każdej formy kwadratowej q istnieje *jedyna symetryczna* forma dwuliniowa g, nazywana *formą biegunową* q, dla której $q(l) = g(l, l)$.

Dla wykazania istnienia takiej formy, wybierzmy dowolną formę dwuliniową h, dla której $q(l) = h(l, l)$ i zdefiniujmy formę dwuliniową g wzorem

$$g(l, m) = \frac{1}{2}\big[h(l, m) + h(m, l)\big].$$

Oczywiście, $g(l, m) = g(m, l)$, tj. forma g jest symetryczna. Ponadto

$$g(l, l) = \frac{1}{2}\big[h(l, l) + h(l, l)\big] = q(l),$$

a dwuliniowość natychmiast wynika z dwuliniowości h.

Dla wykazania jedyności, zauważmy, że jeśli $q(l) = g_1(l, l) = g_2(l, l)$, gdzie g_1, g_2 są dwuliniowe i symetryczne, to forma dwuliniowa $g = g_1 - g_2$ jest też symetryczna i $g(l, l) = 0$ dla każdego $l \in L$. Zgodnie z rozważaniami w dowodzie twierdzenia p. 3 wynika stąd, że $g(l, m) = 0$ dla każdych $l, m \in L$, co dowodzi tezy. Zauważmy jeszcze, że jeśli $q(l) = g(l, l)$ dla symetrycznej g, to

$$g(l, m) = \frac{1}{2}\big[q(l + m) - q(l) - q(m)\big].$$

W ten sposób dowiedliśmy, że geometrie ortogonalne (nad ciałami charakterystyki $\neq 2$) można traktować jako geometrie par (L, q), gdzie $q: L \to \boldsymbol{K}$ jest formą kwadratową. Formę kwadratową można wyrazić poprzez współrzędne wzorem

$$q(\vec{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

a macierz (a_{ij}) jest wyznaczona jednoznacznie pod warunkiem symetrii $a_{ij} = a_{ji}$, np.

$$q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2.$$

Z twierdzeń klasyfikacyjnych wynika, że formę kwadratową można sprowadzić do postaci sumy kwadratów zmiennych z pewnymi współczynnikami

$$q(\vec{x}) = \sum_{i=1}^{n} a_i x_i^2,$$

za pomocą liniowej zamiany zmiennych o nieosobliwej macierzy. Jeśli $\boldsymbol{K} = \boldsymbol{R}$, to

można przyjąć, że $a_i = 0, \pm 1$; liczby r_0, r_+, r_- współczynników równych zeru, plus i minus jedynce są określone jednoznacznie i składają się na sygnaturę wyjściowej formy kwadratowej, a $r_+ + r_-$ jest jej rzędem. Jeśli $\boldsymbol{K} = \boldsymbol{C}$, to można przyjąć, że $a_i = 0, 1$, a liczba jedynek to rząd formy, który także jest określony jednoznacznie.

§ 4. Algorytm ortogonalizacji i wielomiany ortogonalne

W tym paragrafie opiszemy klasyczne algorytmy stosowane dla otrzymywania baz ortogonalnych oraz ważne przykłady takich baz w przestrzeniach funkcji.

1. Sprowadzenie formy kwadratowej do sumy kwadratów. Niech

$$q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j, \quad a_{ij} = a_{ji},$$

będzie formą kwadratową nad ciałem $\boldsymbol{K}$ charakterystyki $\neq 2$. Następująca procedura daje wygodny sposób wyznaczenia takiej liniowej zamiany zmiennych x_i, która sprowadza q do sumy kwadratów (ze współczynnikami).

Przypadek 1. *Istnieje niezerowy współczynnik diagonalny.* Ewentualnie zmieniając numerację zmiennych, możemy przyjąć, że $a_{11} \neq 0$. Wtedy

$$q(x_1, \ldots, x_n) = a_{11} x_1^2 + x_1(2a_{12}x_2 + \cdots + 2a_{1n}x_n) + q'(x_2, \ldots, x_n),$$

gdzie q' jest formą kwadratową nie więcej niż $n-1$ zmiennych. Uzupełniając do pełnego kwadratu, mamy

$$q(x_1, \ldots, x_n) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + q''(x_2, \ldots, x_n),$$

gdzie q'' oznacza nową formę kwadratową $\leq n-1$ zmiennych. Przyjmując

$$y_1 = x_1 + a_{11}^{-1}(a_{12}x_2 + \cdots + a_{1n}x_n), \quad y_2 = x_2, \ldots, y_n = x_n,$$

otrzymujemy w nowych zmiennych formę

$$a_{11}y_1^2 + q''(y_2, \ldots, y_n),$$

i następny krok polega na zastosowaniu algorytmu do q''.

Przypadek 2. *Wszystkie elementy diagonalne są równe zeru.* Jeśli w ogóle $q = 0$, to nie ma nic do zrobienia, bo $q = \sum_{i=1}^{n} 0 \cdot x_i^2$. W przeciwnym przypadku, ewentualnie zmieniając kolejność zmiennych, można założyć, że $a_{12} \neq 0$. Wówczas

$$q(x_1, \ldots, x_n) = 2a_{12}x_1x_2 + x_1 l_1(x_3, \ldots, x_n) + x_2 l_2(x_3, \ldots, x_n) + q'(x_3, \ldots, x_n),$$

gdzie l_1, l_2 są formami liniowymi, a q' — formą kwadratową. Przyjmijmy

$$x_1 = y_1 + y_2, \quad x_2 = y_1 - y_2, \quad x_i = y_i, \quad i \geq 3.$$

W tak określonych nowych zmiennych forma q przyjmuje postać

$$2a_{12}(y_1^2 - y_2^2) + q''(y_1, y_2, \ldots, y_n),$$

przy czym w q'' nie występują wyrazy zawierające y_1^2 ani y_2^2. Można więc do tej ostatniej formy zastosować sposób polegający na wyłączeniu pełnego kwadratu, który był opisany wyżej i znowu zmniejszyć liczbę zmiennych. Kolejne zastosowanie tych kroków doprowadzi formę do postaci $\sum_{i=1}^{n} a_i z_i^2$. Ostateczna zamiana zmiennych jest nieosobliwa, ponieważ wszystkie pośrednie zamiany były takie.

Ostatnia zamiana zmiennych postaci $u_i = \sqrt{|a_i|}z_i$ dla $a_i \neq 0$ w przypadku $\boldsymbol{K} = \boldsymbol{R}$ i $u_i = \sqrt{a_i}z_i$ dla $a_i \neq 0$ w przypadku $\boldsymbol{K} = \boldsymbol{C}$ sprowadzi formę do sumy kwadratów ze współczynnikami 0, ± 1 lub 0, 1.

2. Algorytm ortogonalizacji Grama–Schmidta. Jest on bardzo bliski do opisanego w poprzednim punkcie, ale sformułowany zostanie w sposób bardziej geometryczny. Będziemy rozpatrywać łącznie przypadek ortogonalny i hermitowski.

Wyjściowymi danymi do tej konstrukcji są przestrzeń (L, g) z metryką symetryczną lub hermitowską i wyróżniona w niej baza $\{e'_1, \ldots, e'_n\}$. Niech L_i oznacza podprzestrzeń rozpiętą przez wektory $e'_1, \ldots, e'_i$, $i = 1, \ldots, n$. Proces ortogonalizacji można traktować jako konstruktywny dowód następującego faktu:

3. Stwierdzenie. *Używając wprowadzonych wyżej oznaczeń, załóżmy, że każda podprzestrzeń L_i jest niezdegenerowana. Wówczas istnieje taka baza ortogonalna $\{e_1, \ldots, e_n\}$ przestrzeni L, że powłoka liniowa $\{e_1, \ldots, e_i\}$ jest równa L_i dla każdego $i = 1, \ldots, n$. Mówimy, że ta baza została otrzymana w wyniku ortogonalizacji bazy wyjściowej $\{e'_1, \ldots, e'_n\}$. Każdy wektor e_i jest wyznaczony jednoznacznie z dokładnością do czynnika będącego niezerowym skalarem.*

Dowód. Wektory e_i skonstruujemy przez indukcję względem i. Jako e_1 możemy wziąć e'_1. Jeśli $e_1, \ldots, e_{i-1}$ są już wyznaczone, to szukamy e_i w postaci

$$e_i = e'_i - \sum_{j=1}^{i-1} x_j e'_j, \quad x_j \in \boldsymbol{K}.$$

Ponieważ $\{e'_1, \ldots, e'_i\}$ rozpinają L_i, a $\{e'_1, \ldots, e'_{i-1}\}$ i $\{e_1, \ldots, e_{i-1}\}$ rozpinają L_{i-1}, więc wektory $e_1, \ldots, e_{i-1}$ wraz z dołączonym do nich dowolnym wektorem e_i tej postaci będą rozpinać L_i. Dlatego wystarczy tak dobrać współczynniki x_i, żeby był on ortogonalny do $e_1, \ldots, e_{i-1}$ lub, równoważnie, do $e'_1, \ldots, e'_{i-1}$. Te warunki oznaczają, że $g(e_i, e'_k) = 0$, $k = 1, \ldots, i-1$, czyli

$$\sum_{j=1}^{i-1} x_j g(e'_j, e'_k) = g(e'_i, e'_k), \quad k = 1, \ldots, i-1.$$

Jest to układ $i-1$ równań liniowych z $i-1$ niewiadomymi x_i i macierzą współczynników układu równą macierzy Grama bazy $\{e'_1, \ldots, e'_{i-1}\}$ przestrzeni L_{i-1}. Z założenia

jest ona nieosobliwa, więc rozwiązania x_i istnieją i są jednoznacznie wyznaczone. Dowolny niezerowy wektor $\tilde{e}_i$ ortogonalny do L_{i-1} musi być proporcjonalny do e_i.

Prostszy i bezpośrednio rozwiązywalny układ równań otrzymujemy, postulując następującą postać e_i

$$e_i = e'_i - \sum_{j=1}^{i-1} y_j e_j, \quad y_j \in \boldsymbol{K},$$

przy czym traktujemy wektory $e_1, \ldots, e_{i-1}$ jako wyznaczone poprzednio. Ponieważ $e_1, \ldots, e_{i-1}$ są parami ortogonalne, z warunków $g(e_i, e_j) = 0$, $1 \leq j \leq i-1$, otrzymujemy

$$y_i = \frac{g(e'_i, e_j)}{g(e_j, e_j)}, \quad j = 1, \ldots, i-1.$$

Całym pomysłem, na którym opiera się ten dowód, jest jawne napisanie układów równań liniowych, których rozwiązania określają e_i. Zauważmy, że macierzami współczynników pierwszego układu są kolejne minory diagonalne macierzy Grama bazy wyjściowej:

$$G_i = (g(e'_j, e'_k)), \quad 1 \leq j, k \leq i.$$

Gdyby nie nasze dążenie do algorytmizacji konstrukcji, jeszcze prościej byłoby rozumować w sposób następujący: na mocy stwierdzenia § 3, p. 2 i niezdegenerowania L_{i-1} mamy

$$L_i = L_{i-1} \oplus L_{i-1}^{\perp}, \quad \dim L_{i-1}^{\perp} = \dim L_i - \dim L_{i-1}.$$

Teraz można wziąć jako e_i dowolny niezerowy wektor z $L_{i-1}^{\perp}$.

4. Uwagi i wnioski. a) Proces ortogonalizacji Grama–Schmidta najczęściej stosuje się w przypadku, gdy $g(l, l) > 0$ dla każdego $l \in L$, $l \neq 0$, tj. w przypadku przestrzeni euklidesowych i unitarnych, które będziemy szczegółowo badać później. W tym przypadku *każda podprzestrzeń przestrzeni L jest* automatycznie *niezdegenerowana*, a ortogonalizować można dowolną bazę. Formę g o tej własności nazywamy *dodatnio określoną*, a jej macierze Grama nazywają się *macierzami dodatnio określonymi*.

b) W przypadku $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$ można od początku konstruować bazę ortonormalną. W tym celu należy zastąpić wektor e_i wyznaczony metodą podaną w dowodzie twierdzenia wektorem $|g(e_i, e_i)|^{-1/2} e_i$ lub, dla przestrzeni ortogonalnej nad $\boldsymbol{C}$, wektorem $g(e_i, e_i)^{-1/2} e_i$.

c) *Każdą bazę ortogonalną niezdegenerowanej podprzestrzeni $L_0 \subset L$ można uzupełnić do bazy ortogonalnej przestrzeni L.*

Rzeczywiście, $L = L_0 \oplus L_0^{\perp}$, więc jako szukane uzupełnienie można przyjąć bazę ortogonalną $L_0^{\perp}$. Można ją wyznaczyć metodą Grama–Schmidta, jeśli najpierw uzupełnić bazę L_0 do bazy L (dowolnie, ale tak by pośrednie podprzestrzenie były niezdegenerowane).

d) Niech $\{e'_1, \ldots, e'_n\}$ będzie bazą (L, g), a $\{e_1, \ldots, e_n\}$ — jej ortogonalizacją. Jedyne niezerowe elementy macierzy Grama tej bazy, którymi są $g(e_i, e_i)$, oznaczymy

przez a_i. Jeśli przyjąć, że g jest hermitowska lub ortogonalna nad $\boldsymbol{R}$, to współczynniki a_i są rzeczywiste, a sygnatura g jest wyznaczona liczbą dodatnich i ujemnych współczynników a_i. Pokażemy jak ją wyrazić za pomocą minorów wyjściowej macierzy Grama $G = (g(e'_i, e'_k))$. Przez G_i oznaczymy i-tą diagonalną podmacierz G, tj. macierz Grama $\{e'_1, \ldots, e'_i\}$. Jeśli przez A_i oznaczyć macierz przejścia do bazy $\{e_1, \ldots, e_i\}$, to

$$\det(g(e_k, e_j))_{1 \le k,j \le i} = a_1 \cdots a_i = \det(A_i^t G_i A_i) = \det G_i (\det A_i)^2$$

w przypadku ortogonalnym oraz

$$a_1 \cdots a_i = \det(A_i^t G_i \overline{A_i}) = \det G_i |\det A_i|^2$$

w przypadku hermitowskim. Dlatego w każdym przypadku mamy

$$\operatorname{sgn}(a_1 \cdots a_i) = \operatorname{sgn}(\det G_i)$$

(tutaj $\operatorname{sgn}(t)$ oznacza znak liczby $t \in \boldsymbol{R}$). A więc *sygnatura formy g jest wyznaczona przez liczbę dodatnich i ujemnych elementów ciągu*

$$\det G_1, \quad \frac{detG_2}{detG_1}, \ \ldots, \ \frac{detG_n}{detG_{n-1}}.$$

W szczególności, *forma g* (i *jej macierz G*) *jest dodatnio określona wtedy i tylko wtedy, gdy każdy minor główny* $\det G_i$ *jest dodatni* (przypomnijmy, że tutaj G jest albo rzeczywista symetryczna, albo zespolona i hermitowsko symetryczna). Ten wynik nazywa się *kryterium Sylvestera określoności* (dodatniej albo ujemnej) *macierzy*.

Ogólniej, dla nieosobliwej formy kwadratowej nad dowolnym ciałem tożsamość

$$a_1 \cdots a_i = \det G_i (\det A_i)^2$$

pokazuje, że *dowolną formę o symetrycznej macierzy G i niezerowych minorach głównych* $\det G_i$ *można za pomocą liniowej zamiany zmiennych sprowadzić do postaci*

$$\sum_{i=1}^{n} \frac{\det G_i}{\det G_{i-1}} (y_i)^2, \quad \det G_0 = 1,$$

bowiem kwadraty $(\det A_i)^2$, uniemożliwiające bezpośrednie wyrażenie wyrazów a_i przez $\det G_j$, można wprowadzić jako czynniki do definicji nowych zmiennych. Ten wynik nazywa się *twierdzeniem Jacobiego*.

5. Formy dwuliniowe na przestrzeniach funkcyjnych. Niech f_1, f_2 będą funkcjami określonymi na odcinku (a, b) osi rzeczywistej (dopuszczamy możliwość $a = -\infty$, $b = \infty$) o wartościach rzeczywistych lub zespolonych. Ponadto niech

$G(x)$ będzie ustaloną funkcją na odcinku (a, b). W analizie często występują formy biliniowe na przestrzeniach funkcji zadane wyrażeniami typu

$$g(f_1, f_2) = \int_a^b G(x) f_1(x) f_2(x)\, dx$$

lub (przypadek półtoraliniowy)

$$g(f_1, f_2) = \int_a^b G(x) f_1(x) \overline{f_2(x)}\, dx.$$

Naturalnie, G, f_1 i f_2 muszą spełniać pewne warunki całkowalności, ale w poniższych przykładach będą one spełnione automatycznie.

Funkcja G nazywa się *wagą* formy g. Wyrażenie

$$g(f, f) = \int_a^b G(x) f(x)^2\, dx \quad \text{lub} \quad \int_a^b G(x) |f(x)|^2\, dx$$

nazywamy *średnią kwadratową ważoną* funkcji f (z wagą G); jeśli $G \geq 0$, to tę wielkość możemy traktować jako pewną całkową miarę odchylenia f od zera (funkcji zerowej). Typowy problem *aproksymacji* funkcji f kombinacjami liniowymi zadanego układu funkcji $f_1, \ldots, f_n, \ldots$ polega na poszukiwaniu takich współczynników $a_1, \ldots, a_n, \ldots$, które przy zadanym n minimalizują średnią kwadratową ważoną funkcji

$$f - \sum a_i f_i.$$

Zobaczymy później, że współczynniki a_i wyrażają się szczególnie prosto, jeśli $\{f_i\}$ stanowią układ ortogonalny lub ortonormalny względem iloczynu skalarnego g. W tym paragrafie ograniczymy się do jawnego opisania niektórych ważnych układów ortogonalnych.

6. Wielomiany trygonometryczne. Tutaj $G = 1$, $(a, b) = (0, 2\pi)$. *Wielomianami trygonometrycznymi* (lub *wielomianami Fouriera*) będziemy nazywać skończone kombinacje liniowe funkcji $\sin nx$, $\cos nx$ lub skończone kombinacje liniowe funkcji e^{inx}, $n \in \boldsymbol{Z}$. Zazwyczaj te pierwsze stosuje się w zagadnieniach dotyczących funkcji rzeczywistych, a drugie — funkcji zespolonych. Ponieważ $e^{inx} = \cos nx + i \sin nx$, nad $\boldsymbol{C}$ obie przestrzenie wielomianów Fouriera są identyczne. Nad $\boldsymbol{R}$ używamy metryki dwuliniowej, nad $\boldsymbol{C}$ — półtoraliniowej. Funkcje $\{1, \cos nx, \sin nx \mid n \geq 1\}$ i $\{e^{inx} \mid n \in \boldsymbol{Z}\}$ są liniowo niezależne, zarówno nad $\boldsymbol{R}$, jak i nad $\boldsymbol{C}$ i tworzą

układ ortogonalny, jak to wynika z łatwych do sprawdzenia wzorów:

$$\int_0^{2\pi} \cos nx \cos mx\, dx = \int_0^{2\pi} \sin nx \sin mx\, dx = \begin{cases} \pi, & \text{gdy } m = n > 0, \\ 0, & \text{gdy } m \neq n, \end{cases}$$

$$\int_0^{2\pi} \cos nx \sin mx\, dx = 0,$$

$$\int_0^{2\pi} e^{inx}\, \overline{e^{imx}}\, dx = \int_0^{2\pi} e^{i(n-m)x}\, dx = \begin{cases} 2\pi, & \text{gdy } m = n \geq 0, \\ 0, & \text{gdy } m \neq n \end{cases}$$

Zatem układy funkcji

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos nx, \frac{1}{\sqrt{\pi}} \sin nx, \,\middle|\, n \geq 1 \right\} \quad \text{oraz} \quad \left\{ \frac{1}{\sqrt{2\pi}} e^{inx} \,\middle|\, n \in \boldsymbol{Z} \right\}$$

są ortonormalne. Iloczyny skalarne funkcji f zadanej na $(0, 2\pi)$ z elementami tych ortonormalnych układów nazywa się *współczynnikami Fouriera* tej funkcji i oznacza się je przez

$$a_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)dx,$$

$$a_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \cos nx\, dx, \quad n \geq 1,$$

$$b_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \sin nx\, dx, \quad n \geq 1,$$

dla funkcji rzeczywistych f oraz

$$a_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-inx}\, dx, \quad n \in \boldsymbol{Z},$$

dla funkcji zespolonych. Jeśli sama funkcja jest wielomianem Fouriera, to na mocy wzoru z § 3, p. 6 mamy

$$f(x) = \frac{1}{\sqrt{2\pi}} a_0 + \frac{1}{\sqrt{\pi}} \sum_{n \geq 1} (a_n \cos nx + b_n \sin nx)$$

dla funkcji rzeczywistej f, a

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{n=\infty} a_n e^{inx}$$

dla funkcji zespolonej f. Sumy występujące po prawej są, oczywiście, skończone w rozważanym przypadku funkcji wielomianowych. Nieskończone szeregi takiej postaci nazywają się *szeregami Fouriera*. Problem zbieżności takich szeregów w ogóle, a w szczególności zbieżności do tej funkcji f, której współczynnikami Fouriera są a_n i b_n, stanowi jedno z głównych zagadnień w jednym z ważniejszych działów analizy.

7. Wielomiany Legendre'a. Tu $G = 1$, $(a, b) = (-1, 1)$. W rezultacie ortogonalizacji bazy $\{1, x, x^2, \ldots\}$ przestrzeni wielomianów o współczynnikach rzeczywistych otrzymujemy wielomiany $P_0(x)$, $P_1(x)$, $P_2(x), \ldots$ zwane wielomianami Legendre'a. Zazwyczaj normalizuje się je, nakładając warunek $P_n(1) = 1$. Tak znormalizowane wielomiany Legendre'a są opisane w jawny sposób przez następujące

8. Stwierdzenie. $P_0(x) = 1$, $P_n(x) = \dfrac{1}{2^n n!}\dfrac{d^n}{dx^n}(x^2-1)^n$, $n \geq 1$.

Dowód. Ponieważ stopień wielomianu $(x^2-1)^n$ jest równy $2n$, a stopień $\dfrac{d^n}{dx^n}(x^2-1)^n$ wynosi n, więc $P_0, \ldots, P_i$ rozpinają tę samą przestrzeń nad $\boldsymbol{R}$ co układ $1, x, \ldots, x^i$. Dlatego do sprawdzenia ortogonalności P_i, P_j dla $i \neq j$, wystarczy sprawdzić, że

$$\int_{-1}^{1} x^k P_n(x)\,dx = 0 \quad \text{dla } k < n.$$

Całkując przez części, otrzymamy

$$\int_{-1}^{1} x^k \frac{d^n}{dx^n}(x^2-1)^n\,dx = x^k \frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\Big|_{-1}^{1} - k\int_{-1}^{1} x^{k-1}\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n\,dx.$$

Pierwszy składnik znika, gdyż $(x^2-1)^n$ ma w punktach ± 1 pierwiastek n-krotny, a każde różniczkowanie zmniejsza krotność pierwiastka o jeden. Do drugiego składnika można zastosować analogiczną procedurę i po k krokach otrzymamy całkę proporcjonalną do

$$\int_{-1}^{1} \frac{d^{n-k}}{dx^{n-k}}(x^2-1)^n\,dx = \frac{d^{n-k-1}}{dx^{n-k-1}}(x^2-1)^n\Big|_{-1}^{1} = 0.$$

Ze wzoru Leibniza otrzymujemy następnie

$$\frac{d^n}{dx^n}(x-1)^n(x+1)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{d^k}{dx^k}(x-1)^n(x+1)^n.$$

W punkcie $x = 1$ nie znika tylko składnik ze wskaźnikiem $k = n$, a więc

$$P_n(1) = \frac{1}{2^n n!}\binom{n}{n}\left[\frac{d^n}{dx^n}(x-1)^n\right](x+1)^n\Big|_{x=1} = \frac{1}{2^n n!}\cdot 1 \cdot n! \cdot 2^n = 1,$$

co kończy dowód.

9. Wielomiany Czebyszewa. $G = \dfrac{1}{\sqrt{1-x^2}}$, $(a,b) = (-1,1)$. Wielomiany $T_n(x)$, $n \geq 0$, otrzymane przez ortogonalizację bazy $\{1, x, x^2, \ldots\}$ są dane jawnym wzorem

$$T_n(x) = \frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\frac{d^n}{dx^n}(x^2-1)^{n-1/2} = \cos(n \arccos x),$$

a ich normalizację wyrażają związki

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} 0, & \text{gdy } m \neq n, \\ \pi/2, & \text{gdy } m = n \neq 0, \\ \pi, & \text{gdy } m = n = 0. \end{cases}$$

10. Wielomiany Hermite'a. $G = e^{-x^2}$, $(a,b) = (-\infty, \infty)$. Wielomiany Hermite'a otrzymuje się w wyniku ortogonalizacji bazy $\{1, x, x^2, \ldots\}$. W jawnej postaci są one dane wzorami

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}(e^{x^2}),$$

a ich normalizację wyrażają zależności

$$\int_{-\infty}^{\infty} e^{x^2} H_n(x)H_m(x)\,dx = \begin{cases} 0, & \text{gdy } m \neq n, \\ 2^n n!\sqrt{\pi}, & \text{gdy } m = n. \end{cases}$$

Pozostawiamy czytelnikowi jako ćwiczenie podanie dowodów tych faktów.

Ćwiczenia

1. Dowieść, że hermitowska lub symetryczna forma g jest nieujemnie określona, tj. $g(l, l) \geq 0$ dla każdego $l \in L$, wtedy i tylko wtedy, gdy wszystkie minory diagonalne jej macierzy Grama są nieujemne.

2. Wykazać stwierdzenia sformułowane w p. p. 9, 10 tego paragrafu.

§ 5. Przestrzenie euklidesowe

1. Definicja. *Przestrzenią euklidesową* nazywamy skończenie wymiarową rzeczywistą przestrzeń liniową L wraz z symetrycznym i dodatnio określonym iloczynem skalarnym.

Będziemy pisać (l, m) zamiast $g(l, m)$ i $|l|$ zamiast $(l,l)^{1/2}$; liczbę $|l|$ będziemy nazywać *długością wektora* l.

Z rezultatów wyprowadzonych w paragrafach 3, 4 wynika, że

a) w każdej przestrzeni euklidesowej istnieje baza ortonormalna, której wektory mają długość 1; a zatem

b) każda taka przestrzeń jest izometryczna z kartezjańską przestrzenią euklidesową $\boldsymbol{R}^n$, dla $n = \dim L$, w której

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i y_i, \quad |\vec{x}| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}.$$

Kluczem do wielu własności przestrzeni euklidesowej jest wielokrotnie na nowo odkrywana nierówność Cauchy'ego–Buniakowskiego–Schwarza.

2. Stwierdzenie. *Dla dowolnych $l_1, l_2 \in L$ zachodzi nierówność*

$$(l_1, l_2) \leq |l_1| \cdot |l_2|.$$

Równość zachodzi wtedy i tylko wtedy, gdy wektory l_1, l_2 są liniowo zależne.

Dowód. W przypadku $l_1 = 0$, lewa i prawa strona są równe, a wektory są rzeczywiście liniowo zależne. Będziemy rozważać przypadek $l_1 \neq 0$. Dla każdej liczby rzeczywistej t mamy

$$|tl_1 + l_2|^2 = (tl_1 + l_2, tl_1 + l_2) = t^2|l_1|^2 + 2t(l_1, l_2) + |l_2|^2 \geq 0$$

na mocy dodatniej określoności iloczynu skalarnego. Wyróżnik trójmianu kwadratowego po prawej stronie ostatniej równości jest więc niedodatni, tj.

$$(l_1, l_2)^2 - |l_1|^2|l_2|^2 \leq 0.$$

Równość zachodzi wtedy i tylko wtedy, gdy trójmian ten ma pierwiastek rzeczywisty t_0. W takim przypadku

$$|t_0 l_1 + l_2|^2 = 0 \iff l_2 = -t_0 l_1,$$

co kończy dowód.

3. Wniosek (nierówność trójkąta). *Dla każdych $l_1, l_2 \in L$*

$$|l_1 + l_2| \leq |l_1| + |l_2|, \quad |l_1 - l_3| \leq |l_1 - l_2| + |l_2 - l_3|.$$

Dowód. Mamy

$$|l_1 + l_2|^2 = |l_1|^2 + 2(l_1, l_2) + |l_2|^2 \leq |l_1|^2 + 2|l_1||l_2| + |l_2|^2 = (|l_1| + |l_2|)^2.$$

Zamieniając w powyższym l_1 na $l_1 - l_2$ i l_2 na $l_2 - l_3$ otrzymamy drugą nierówność.

4. Wniosek. *Długość euklidesowa wektora $|l|$ jest normą w sensie definicji podanej w części 1, § 10, p. 4, a funkcja $d(l, m) = |l - m|$ jest metryką w znaczeniu definicji p. 1, loc. cit.*

D o w ó d. Nie została jeszcze wykazana tylko własność jednorodności metryki, tj. $|al| = |a||l|$ dla każdego $a \in \boldsymbol{R}$, która wynika z

$$|al| = (al, al)^{1/2} = (a^2|l|)^{1/2} = |a| \cdot |l|.$$

5. Kąty i odległości. Niech $l_1, l_2 \in L$ będą niezerowymi wektorami. Na mocy stwierdzenia p. 2

$$-1 \leq \frac{(l, l)}{|l||l|} \leq 1.$$

Istnieje zatem jednoznacznie wyznaczony kąt φ, $0 \leq \varphi \leq \pi$, taki że

$$\cos\varphi = \frac{(l, l)}{|l||l|}.$$

Nazywamy go kątem pomiędzy wektorami l_1, l_2. Ze względu na to, że iloczyn skalarny jest symetryczny, jest to „kąt niezorientowany", co wyjaśnia zakres jego zmienności w przedziale $(0, \pi)$. W zgodzie ze szkolną geometrią kąt między ortogonalnymi wektorami wynosi $\pi/2$. Można systematycznie konstruować geometrię euklidesową, wychodząc od podanych tu definicji kąta i odległości, i przekonać się, że w dwóch i trzech wymiarach jest ona identyczna z klasyczną.

Na przykład, wielowymiarowe twierdzenie Pitagorasa jest trywialnym wnioskiem z definicji: jeśli wektory $l_1, \ldots, l_n$ są parami ortogonalne, to

$$\left|\sum_{i=1}^{n} l_i\right|^2 = \sum_{i=1}^{n} |l_i|^2.$$

Zastosowanie znanego z geometrii płaszczyzny wzoru cosinusów do trójkąta o bokach l_1, l_2, l_3 daje

$$|l_3|^2 = |l_1|^2 + |l_2|^2 - 2|l_1||l_2|cos\varphi,$$

gdzie φ jest kątem między l_1 i l_2. W sformułowaniu wektorowym $l_3 = l_1 - l_2$ i wzór powyższy staje się tożsamością

$$|l_1 - l_2|^2 = |l_1|^2 + |l_2|^2 - 2(l_1, l_2)$$

na mocy naszego określenia kąta.

Niech U, $V \subset L$ będą dwoma podzbiorami w przestrzeni euklidesowej. *Odległością* między nimi nazywamy liczbę nieujemną

$$d(U, V) = \inf\{|l_1 - l_2| \mid l_1 \in U, l_2 \in V\}.$$

Rozważmy szczególny przypadek, gdy $U = \{l\}$ (jeden wektor), a $V = L_0 \subset L$ jest podprzestrzenią liniową. Na mocy stwierdzenia § 3, p. 2 mamy $L = L_0 \oplus L_0^{\perp}$ i $l = l_0 + l_0'$, gdzie $l_0 \in L_0$, $l_0' \in L_0'$. Wektory l_0, l_0' nazywamy *rzutami ortogonalnymi* wektora l na L_0 i $L_0^{\perp}$, odpowiednio.

6. Stwierdzenie. *Odległość wektora l od L_0 jest równa długości rzutu ortogonalnego l na $L_0^\perp$.*

Dowód. Dla dowolnego wektora $m \in L_0$ równość

$$|l - m|^2 = |l_0 + l_0' - m|^2 = |l_0 - m|^2 + |l_0'|^2$$

zachodzi na mocy twierdzenia Pitagorasa, bo wektory $l_0 - m \in L_0$ i $l_0' \in L_0^\perp$ są prostopadłe. Zatem

$$|l_0 - m|^2 \geq |l'|^2,$$

a równość zachodzi tylko wtedy, gdy $m = l_0$, co dowodzi tezy.

Jeśli w L_0 jest wybrana baza ortonormalna $\{e_1, \ldots, e_m\}$, to rzuty na L_0 wyznacza się według wzoru

$$l_0 = \sum_{i=1}^{m} (l, e_i) e_i.$$

W istocie, lewa i prawa strona ma jednakowe iloczyny skalarne ze wszystkimi wektorami e_i, więc ich różnica należy do $L_0^\perp$. Wreszcie,

$$d(l, L_0) = \left| l - \sum_{i=1}^{m} (l, e_i) e_i \right|$$

jest najmniejszą wartością $|l - m|$, gdy m przebiega wektory podprzestrzeni L_0. Ponieważ znowu na mocy twierdzenia Pitagorasa $|l_0|^2 \leq |l|^2$, więc

$$\sum_{i=1}^{m} (l, e_i)^2 \leq |l|^2.$$

7. Zastosowania do przestrzeni funkcyjnych. Jako przykład rozważymy przestrzeń funkcji ciągłych o wartościach rzeczywistych określonych na $(a, b) \subset \boldsymbol{R}$ z iloczynem skalarnym

$$(f, g) = \int_a^b fg \, dx.$$

Jest ona wprawdzie nieskończenie wymiarowa, ale wszystkie nasze nierówności będą odnosić się do skończonych układów funkcji z tej przestrzeni, możemy więc za każdym razem przyjmować, że rozważamy skończenie wymiarową przestrzeń euklidesową. Mamy $\int_a^b f^2(x)\,dx \geq 0$ i jeśli $\int_a^b f^2(x)\,dx = 0$, to $f(x) \equiv 0$. Nierówność Cauchy'ego–Buniakowskiego–Schwarza przybiera postać

$$\left(\int_a^b f(x) g(x) dx \right)^2 \leq \int_a^b f^2(x)\,dx \int_a^b g^2(x)\,dx,$$

a nierówność trójkąta — postać

$$\left(\int_a^b (f(x)+g(x))\,dx\right)^{1/2} \leq \left(\int_a^b f^2(x)\,dx\right)^{1/2} + \left(\int_a^b g^2(x)\,dx\right)^{1/2}$$

Jeśli $(a,b) = (0, 2\pi)$ i a_i, b_i są współczynnikami Fouriera funkcji $f(x)$, tak jak w § 4, p. 6, to wielomian Fouriera

$$f_N(x) = \frac{1}{\sqrt{2\pi}}a_0 + \frac{1}{\sqrt{\pi}}\sum_{n=1}^{N}(a_n \cos nx + b_n \sin nx)$$

jest rzutem ortogonalnym $f(x)$ na powłokę liniową zbioru $\{1, \cos nx, \sin nx \mid 1 \leq n \leq N\}$. Zatem współczynniki Fouriera $f(x)$ przy każdym N minimalizują średnią wartość odchylenia kwadratowego $f(x)$ od wielomianów Fouriera „stopnia" $\leq N$. Nierówność $|f_N|^2 \leq |f|^2$ przyjmuje postać

$$a_0^2 + \sum_{n=1}^{N}(a_n^2 + b_n^2) \leq \int_0^{2\pi} f^2(x)\,dx.$$

Ponieważ prawa strona jest niezależna od N i a_n^2, $b_n^2 \geq 0$, więc szereg

$$a_0^2 + \sum_{n=1}^{\infty}(a_n^2 + b_n^2)$$

jest zbieżny dla każdej funkcji ciągłej $f(x)$ na $[0, 2\pi]$. Można wykazać, że jego granicą jest dokładnie $\int_0^{2\pi} f^2(x)\,dx$.

W pełni analogiczne rozważania można zastosować do przypadku wielomianów Legendre'a, Czebyszewa i Hermite'a. Pozostawimy to czytelnikowi jako ćwiczenie.

8. Metoda najmniejszych kwadratów. Rozważmy układ m równań liniowych z n niewiadomymi o współczynnikach rzeczywistych

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, m.$$

Załóżymy, że układ ten jest „nadokreślony", tj. $m > n$, a rząd macierzy współczynników jest równy n. Taki układ na ogół nie ma rozwiązań. Jednakże można próbować znaleźć takie wartości niewiadomych $x_1^0, \ldots, x_n^0$, żeby sumaryczne średnie kwadratowe odchylenie lewych stron od prawych

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}x_j^0 - b_i\right)^2$$

przyjmowało możliwie najmniejszą wartość. Ten problem ma ważne zastosowania w praktyce. Na przykład, przy pomiarach geodezyjnych dzieli się daną okolicę na sieć trójkątów i mierzy pewne elementy tych trójkątów, a inne oblicza za pomocą wzorów trygonometrycznych. Ponieważ wszystkie pomiary są obarczone pewnym błędem, zaleca się wykonanie pomiarów większej liczby elementów, niż jest to teoretycznie konieczne dla wyznaczenia pozostałych, ale wtedy równania dla pozostałych elementów prawie na pewno okażą się niezgodne. Metoda najmniejszych kwadratów pozwala w takim przypadku otrzymać „przybliżone rozwiązanie", bardziej wiarygodne z powodu większej ilości informacji wprowadzonych do zagadnienia.

Pokażemy teraz, że to zadanie można rozwiązać przy użyciu rezultatów p. 7. Potraktujemy kolumny macierzy współczynników $e_i = (a_{1i}, \ldots, a_{mi})$ i kolumnę wyrazów wolnych $f = (b_1, \ldots, b_m)$ jako wektory kartezjańskiej przestrzeni euklidesowej $\boldsymbol{R}^m$ ze standardowym iloczynem skalarnym. Przyjmując

$$e = \sum_{i=1}^{n} x_i e_i,$$

otrzymamy

$$\sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} x_j - b_i \Big)^2 = \Big| \sum_{i=1}^{n} x_i e_i - f \Big|^2.$$

Stąd wynika, że minimum średniokwadratowego odchylenia jest osiągane wtedy, gdy $\sum_{i=1}^{n} x_i^0 e_i$ jest rzutem ortogonalnym f na podprzestrzeń rozpiętą przez wektory e_i. To oznacza, że współrzędne x_i^0 należy wyznaczyć jako rozwiązania układu n równań z n niewiadomymi

$$\Big(\sum_{i=1}^{n} x_i^0 e_i, e_j \Big) = (f, e_j), \quad j = 1, \ldots, n,$$

tak nazywanego „*układu normalnego*". Wyznacznikiem tego układu jest wyznacznik macierzy Grama $((e_i, e_j))$, gdzie

$$(e_i, e_j) = \sum_{k=1}^{m} a_{ki} a_{kj}.$$

Jest on różny od zera, bo założyliśmy, że rząd wyjściowego układu, tj. rząd układu wektorów (e_i) jest równy n (por. ćwiczenie 5 do § 2). Dlatego rozwiązanie istnieje i jest jedyne.

Zajmiemy się teraz problemem „mierzenia w przestrzeni euklidesowej".

9. n-wymiarowa objętość. Najprostszym figurom zawartym w jednowymiarowej przestrzeni euklidesowej, takim jak odcinki i ich skończone sumy, można przypisać długości i, odpowiednio, sumy ich długości. Szkolna geometria uczy jak mierzyć na płaszczyźnie pola takich figur jak prostokąty, trójkąty i, z pewnym trudem, także koła. Uogólnienie tych pojęć przynosi głęboka *teoria miary*, na którą tutaj nie ma

miejsca. Ograniczymy się tylko do zestawienia podstawowych własności i do elementarnych obliczeń związanych ze specjalną miarą figur w n-wymiarowej przestrzeni euklidesowej — ich n-wymiarową objętością.

Będziemy nazywać n-wymiarową objętością funkcję vol^n, określoną na *pewnych* podzbiorach n-wymiarowej przestrzeni euklidesowej L, nazywanych *mierzalnymi*, i przyjmująca wartości będące liczbami rzeczywistymi nieujemnymi lub ∞ (na zbiorach mierzalnych ograniczonych tylko wartości skończone). Rodzina zbiorów mierzalnych powinna być dostatecznie bogata. Tutaj po prostu postulujemy następujący spis własności funkcji vol^n i mierzalność wchodzących w ten spis zbiorów, nie dowodząc istnienia funkcji o tych własnościach ani nie wskazując naturalnej dziedziny jej określoności.

a) Funkcja vol^n jest przeliczalnie addytywna, tj.

$$\mathrm{vol}^n\left(\bigcup_{i=1}^{\infty} U_i\right) = \sum_{i=1}^{\infty} \mathrm{vol}^n U_i, \quad \text{jeśli } U_i \cap U_j = \emptyset \text{ dla } i \neq j;$$

vol^1 (punkt) $= 0$; vol^1 (odcinek) = długość odcinka. Odcinkiem w jednowymiarowej przestrzeni euklidesowej jest zbiór punktów postaci $tl_1 + (1-t)l_2$, $0 \leq t \leq 1$, a jego długością nazywamy liczbę $|l_1 - l_2|$.

b) Jeśli $U \subseteq V$, to $\mathrm{vol}^n U \leq \mathrm{vol}^n V$.

c) Jeśli $L = L_1 \oplus L_2$ (ortogonalna suma prosta), $\dim L_1 = m$, $\dim L_2 = n$, $U \subset L_1$, $V \subset L_2$, to dla $U \times V = \{(l_1, l_2) | l_1 \in U, l_2 \in V\} \subset L_1 \oplus L_2$ zachodzi

$$\mathrm{vol}^{m+n}(U \times V) = \mathrm{vol}^m U \cdot \mathrm{vol}^n V.$$

d) Jeśli $f: L \to L$ jest dowolnym operatorem liniowym, to

$$\mathrm{vol}^n(f(U)) = |\det f| \, \mathrm{vol}^n U, \quad n = \dim L.$$

Własności a), b) nie wymagają komentarzy. Własność c) jest daleko idącym uogólnieniem wzoru wyrażającego pole prostokąta jako iloczyn długości jego boków albo objętość prostopadłościanu jako iloczyn pola podstawy przez wysokość. Zauważmy, że z własności c) wynika, iż $(m+n)$-wymiarowa objętość zbioru ograniczonego W w L, zawartego w podprzestrzeni L_1 wymiaru $m < m+n$, jest równa zeru. Rzeczywiście, wtedy $L = L_1 \oplus L_1^{\perp}$ i $W = V \times \{0\}$ oraz $\mathrm{vol}^n\{0\} = 0$ dla $n > 0$ na mocy a) i b).

Sens własności d) jest mniej oczywisty. Jest to zasadniczy wkład algebry liniowej do teorii objętości euklidesowej, dostarczający objaśnienia dla występowania jakobianów w formalizmie całek wielowymiarowych. Być może najbardziej intuicyjnym wyjaśnieniem tej własności jest obserwacja, że operator rozciągnięcia a razy ($a \in \boldsymbol{R}$) w kierunku któregoś z wektorów bazy ortonormalnej powoduje zwiększenie objętości $|a|$ razy na mocy własności a) i b). Ponieważ dowolny niezerowy wektor można uzupełnić do bazy ortogonalnej, więc operator diagonalizowalny f z wartościami własnymi $a_1, \ldots, a_n \in \boldsymbol{R}$ powinien zwiększać objętość $|a_1| \cdots |a_n| = |\det f|$ razy.

Z drugiej strony izometrie powinny zachowywać objętości, a jak przekonamy się później, każdy operator w przestrzeni euklidesowej jest złożeniem izometrii i operatora diagonalizowalnego (por. § 8, ćwiczenie 11).

A teraz, wykorzystując te aksjomaty, podamy spis objętości szeregu najprostszych i najważniejszych n-wymiarowych figur.

10. Kostka jednostkowa. Tak nazywamy zbiór $\{t_1e_1 + \cdots + t_ne_n \mid 0 \leq t_i \leq 1\}$, gdzie $\{e_1, \ldots, e_n\}$ jest jakąś bazą ortonormalną L. Z własności a) i c) p. 10 od razu wynika, że jej objętość jest równa jedności.

Przyjmując, że t_i zmienia się w przedziale $0 \leq t_i \leq a$, otrzymamy kostkę o boku a. Ponieważ jest ona obrazem kostki jednostkowej względem homotetii — mnożenia przez a — jej objętość wynosi a^n.

11. n-wymiarowy prostopadłościan o krawędziach $\{l_1, \ldots, l_n\}$ jest to zbiór $\{t_1l_1 + \cdots + t_nl_n \mid t0 \leq t_i \leq 1\}$. Pokażemy, że jego objętość jest równa $\sqrt{|\det G|}$, gdzie $G = ((l_i, l_j))$ oznacza macierz Grama układu krawędzi. W istocie, jeśli układ $\{l_1, \ldots, l_n\}$ jest liniowo zależny, to wyznaczony przez niego prostopadłościan jest zawarty w podprzestrzeni wymiaru $< \dim L$ i jego n-wymiarowa objętość równa jest zeru na mocy uwagi z p.9, a jednocześnie macierz Grama układu jest zdegenerowana. Pozostaje do rozważenia przypadek liniowo niezależnego układu $\{l_1, \ldots, l_n\}$. Niech $\{e_1, \ldots, e_n\}$ oznacza bazę ortonormalną przestrzeni L, a f — odwzorowanie liniowe $L \to L$, przeprowadzające e_i w l_i, dla $i = 1, \ldots, n$. Jeśli A oznacza macierz tego odwzorowania w bazie $\{e_i\}$:

$$(l_1, \ldots, l_n) = (e_1, \ldots, e_n)A,$$

to macierz Grama $\{l_i\}$ jest równa A^tA, bo macierz Grama $\{e_i\}$ jest jednostkowa. Zatem,

$$\sqrt{|\det G|} = \sqrt{|\det(A^tA)|} = |\det A|.$$

Z drugiej strony, $|\det A| = |\det f|$ i f odwzorowuje kostkę jednostkową w nasz prostopadłościan. Na mocy własności d) z p. 9 objętość prostopadłościanu wynosi $|\det f|$, co chcieliśmy wykazać.

12. n-wymiarowa kula o promieniu r jest to zbiór wektorów

$$B^n(r) = \{l \mid |l| \leq r\}$$

lub, we współrzędnych ortogonalnych,

$$B^n(r) = \left\{(x_1, \ldots, x_n) \;\middle|\; \sum_{i=1}^{n} x_i^2 \leq r^2\right\}.$$

Ponieważ $B^n(r)$ można otrzymać z $B^n(1)$ przez rozciągnięcie r razy, więc

$$\operatorname{vol}^n B^n(r) = \operatorname{vol}^n B^n(1) r^n.$$

Czynnik $\mathrm{vol}^n B^n(1) = b_n$ można obliczyć tylko przy wykorzystaniu środków analitycznych. Przecinając $(n+1)$-wymiarową kulę n-wymiarowymi podrozmaitościami liniowymi, ortogonalnymi do ustalonego kierunku, otrzymujemy wzór indukcyjny

$$b_{n+1} = \left[2\int_0^1 \left(\sqrt{1-x^2}\right)^n dx_{n+1}\right] b_n, \quad n \geq 1.$$

Oczywiście, $b_1 = 2$, $b_2 = \pi$, $b_3 = \frac{4}{3}\pi$.

13. n-wymiarowa elipsoida z półosiami $r_1, \ldots, r_n$. We współrzędnych ortogonalnych jest ona zdefiniowana przez nierówność

$$\sum_{i=1}^n \left(\frac{x_i}{r_i}\right)^2 \leq 1.$$

Ponieważ można ją otrzymać z kuli $B^n(1)$, rozciągając tę ostatnią w stosunku r_i w kierunku każdej z półosi, jej objętość jest równa $b_n r_1 \cdots r_n$.

14. Pewna własność n-wymiarowej objętości. Polega ona na tym, że przy bardzo dużych n *„objętość n-wymiarowej figury jest skupiona w pobliżu jej powierzchni"*. Na przykład, objętość pierścienia kulistego zawartego pomiędzy sferami o promieniu 1 i $1-\varepsilon$ wynosi $b_n[1-(1-\varepsilon)^n]$, co przy ustalonym dowolnie małym ε i przy rosnącym n dąży do b_n. Dwudziestowymiarowy „arbuz" o promieniu 20 cm ze skórką grubości 1 cm prawie w dwóch trzecich składa się z samej skórki:

$$1 - \left(1 - \frac{1}{20}\right)^{20} \approx 1 - e^{-1}.$$

Ta własność odgrywa znaczącą rolę w mechanice statystycznej. Rozważmy na przykład najprostszy model gazu zamkniętego w rezerwuarze i składającego się z n atomów, które będziemy traktować jako punkty materialne o masie 2 (w odpowiednim układzie jednostek). Chwilowe położenie gazu można reprezentować poprzez n trójwymiarowych wektorów prędkości $(v_1, \ldots, v_n)$, tj. poprzez wektor $3n$-wymiarowej przestrzeni kartezjańskiej $\boldsymbol{R}^{3n}$. Kwadrat długości wektora z $\boldsymbol{R}^{3n}$ ma bezpośrednią interpretację *energii* układu (sumy energii kinetycznej atomów):

$$E = \sum_{i=1}^n |v_i|^2.$$

Dla makroskopowej objętości gazu w normalnych warunkach n jest rzędu 10^{23} (liczba Avogadro), tak że stan układu opisywany jest punktem sfery ogromnego wymiaru, o promieniu równym pierwiastkowi kwadratowemu z energii.

Rozważmy dwa (makroskopowe) rezerwuary połączone tak, że suma ich energii $E_1 + E_2 = E$ jest stała, ale możliwa jest wymiana energii bez wymiany atomów. Wówczas energie E_1 i E_2 większą część czasu są bliskie takim wartościom, które

maksymalizują „objętość przestrzeni stanów" dostępną dla połączonych układów, tj. iloczyn

$$\operatorname{vol}^{n_1} B(E^{1/2}) \operatorname{vol}^{n_2} B(E_2^{1/2})$$

(tu zamieniliśmy pola sfer objętościami kul, co nie jest poprawne, ale prawie nie wpływa na rezultat). Ze względu na to, że ze wzrostem E_1 i maleniem E_2 ($E_1 + E_2 = \text{const}$) pierwsza objętość nieprawdopodobnie szybko rośnie, gdy druga maleje, dla pewnych wartości E_1, E_2 pojawia się ostre maksimum tego iloczynu odpowiadające „najbardziej prawdopodobnemu" stanowi połączonego układu. Oczywiście zachodzi to wtedy, gdy

$$\frac{d}{dE_1} \log \operatorname{vol}^{n_1} B(E_1^{1/2}) = \frac{d}{dE_2} \log \operatorname{vol}^{n_2} B(E_2^{1/2}).$$

Odwrotnymi do tych wartości są (z dokładnością do proporcjonalności) *temperatury* rezerwuarów, a najbardziej prawdopodobne położenie odpowiada równości temperatur.

Ćwiczenia

1. Wykazać, że kąt φ nachylenia prostej, wychodzącej z początku układu współrzędnych na płaszczyźnie $\boldsymbol{R}^2$ i mającej możliwie najmniejsze średniokwadratowe odchylenie od m zadanych punktów (a_i, b_i), $i = 1, \ldots, m$, jest dany wzorem

$$\operatorname{tg} \varphi = \Big(\sum_{i=1}^{m} a_i b_i\Big) \Big/ \Big(\sum_{i=1}^{m} a_i\Big).$$

(*Wskazówka*. Znaleźć najlepsze „przybliżone rozwiązanie" układu równań $a_i x = b_i$.)

2. Niech $P_n(x)$ będzie n-tym wielomianem Legendre'a. Wykazać, że najwyższy współczynnik wielomianu $u_n(x) = \dfrac{2^n (n!)^2}{(2n)!} P_n(x)$ jest równy jedności i że minimum całki $I(u) = \displaystyle\int_{-1}^{1} u(x)^2 \, dx$ w zbiorze wielomianów $u(x)$ stopnia n z najwyższym współczynnikiem 1 jest osiągane dla wielomianu $u = u_n$.

(*Wskazówka*. Rozłożyć u na sumę wielomianów Legendre'a stopni $\leq n$.)

3. Niech (S, μ) oznacza parę składającą się ze skończonego zbioru S i rzeczywistej funkcji $\mu: S \to \boldsymbol{R}$, spełniającej dwa warunki: $\mu(s) \geq 0$ dla każdego $s \in S$ i $\sum_{s \in S} \mu(s) = 1$. Rozważmy funkcjonał liniowy $E: F(S) \to \boldsymbol{R}$ określony (na przestrzeni funkcji rzeczywistych $F(S)$ na S i przybierający wartości w $\boldsymbol{R}$) za pomocą wzoru

$$E(f) = \sum_{s \in S} \mu(s) f(s)$$

i oznaczmy przez $F_0(S)$ jego jądro.

(S, μ) nazywa się *skończoną przestrzenią probabilistyczną*, elementy $F(S)$ — *zmiennymi losowymi*, elementy $F_0(S)$ — *unormowanymi zmiennymi losowymi*, a

liczbę $E(f)$ — *oczekiwaniem matematycznym* zmiennej f. Zmienne losowe tworzą pierścień względem punktowego mnożenia funkcji.

Wykazać następujące fakty.

a) $F_0(S)$ i $F(S)$ mają strukturę przestrzeni ortogonalnej, w której kwadrat długości wektora f jest dany wzorem $E(f^2)$. Przestrzeń $F(S)$ jest euklidesowa wtedy i tylko wtedy, gdy $\mu(s) > 0$ dla wszystkich $s \in S$.

b) Dla dowolnych zmiennych losowych $f, g \in F(S)$ i $a, b \in \boldsymbol{R}$ przyjmiemy

$$P(f = a) = \sum_{f(s)=a} \mu(s), \quad P(f = a; g = b) = \sum_{\substack{f(s)=a \\ f(s)=b}} \mu(s);$$

są to, odpowiednio, prawdopodobieństwa tego, że f przyjmuje wartość a lub $f = a$ i „jednocześnie" $g = b$. Zmienne losowe f, g nazwiemy *niezależnymi*, jeśli

$$P(f = a; g = b) = P(f = a)P(g = b)$$

dla wszystkich $a, b \in \boldsymbol{R}$. Wykazać, że jeśli unormowane zmienne losowe f, $g \in F_0(S)$ są niezależne, to są ortogonalne.

Skonstruować przykład pokazujący, że odwrotna implikacja nie zachodzi.

Iloczyn skalarny zmiennych losowych f, $g \in F_0(S)$ nazywa się ich *kowariancją*, a cosinus kąta między nimi — ich *współczynnikiem korelacji*.

§ 6. Przestrzenie unitarne

1. Definicja. *Przestrzenią unitarną* nazywamy zespoloną przestrzeń liniową L z dodatnio określonym hermitowskim iloczynem skalarnym.

Podobnie jak w § 5 będziemy pisać (l, m) zamiast $g(l, m)$ i $|l|$ zamiast $(l, l)^{1/2}$. Później przekonamy się, że $|l|$ jest normą w L w znaczeniu definicji wprowadzonej w § 10 części 1. Przestrzenie unitarne zupełne względem tej normy nazywają się *przestrzeniami Hilberta* (*hilbertowskimi*). W szczególności, skończenie wymiarowe przestrzenie unitarne są przestrzeniami hilbertowskimi.

Z rezultatów wyprowadzonych w § 3 i § 4 wynika, że:

a) w każdej skończenie wymiarowej przestrzeni unitarnej istnieje baza ortonormalna, której wektory mają długość 1;

a zatem

b) każda taka przestrzeń jest izometryczna z kartezjańską przestrzenią unitarną $\boldsymbol{C}^n$ $(n = \dim L)$, w której

$$(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i \overline{y}_i, \quad |\vec{x}| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

Przestrzenie unitarne mają wiele wspólnych własności z przestrzeniami euklidesowymi z następującego powodu. Jeśli L jest skończenie wymiarową przestrzenią unitarną, to na jej realifikacji $L_{\boldsymbol{R}}$ istnieje (jedyna) struktura przestrzeni euklidesowej,

dla której norma $|l|$ jest identyczna jak norma w L. Istnienie takiej normy wynika z obserwacji zanotowanej w punktach a) i b) powyżej: jeśli $\{e_1, \ldots, e_n\}$ jest ortonormalną bazą L, a $\{e_1, ie_1, \ldots, e_n, ie_n\}$ — odpowiadającą jej bazą $L_{\boldsymbol{R}}$, to

$$\left|\sum_{j=1}^{n} x_j e_j\right|^2 = \sum_{j=1}^{n} |x_j|^2 = \sum_{j=1}^{n}((\operatorname{Re} x_j)^2 + (\operatorname{Im} x_j)^2),$$

a suma po prawej przedstawia kwadrat euklidesowej normy wektora $\sum_{j=1}^{n} \operatorname{Re} x_j e_j + \sum_{j=1}^{n} \operatorname{Im} x_j(ie_j)$ wyrażonego w bazie ortonormalnej $\{e_i, ie_i\}$. Jednoznaczność struktury euklidesowej odpowiadającej tej normie wynika z rezultatu § 3, p. 9.

Jednakże *iloczyny skalarne* w przestrzeni unitarnej L i przestrzeni euklidesowej $L_{\boldsymbol{R}}$ *nie są tożsame*: drugi z nich przyjmuje tylko wartości rzeczywiste, podczas gdy pierwszy — także zespolone. W rzeczywistości hermitowski iloczyn skalarny na L wyznacza nie tylko ortogonalną strukturę na $L_{\boldsymbol{R}}$, ale również symplektyczną, co wynika z następującej konstrukcji.

Wracając chwilowo do oznaczenia $g(l, m)$ dla hermitowskiego iloczynu skalarnego w L wprowadzimy oznaczenia

$$a(l, m) = \operatorname{Re} g(l, m), \quad b(l, m) = \operatorname{Im} g(l, m).$$

Mamy wówczas

2. Stwierdzenie. a) *$a(l, m)$ jest symetrycznym, a $b(l, m)$ — antysymetrycznym iloczynem skalarnym na $L_{\boldsymbol{R}}$; oba te iloczyny są niezmiennicze względem mnożenia przez i, tj. względem kanonicznej struktury zespolonej na $L_{\boldsymbol{R}}$:*

$$a(il, im) = a(l, m), \quad b(il, im) = b(l, m);$$

b) *formy a i b spełniają następujące zależności:*

$$a(l, m) = b(il, m), \quad b(l, m) = -a(il, m);$$

c) *dowolna para i-niezmienniczych form dwuliniowych a, b na $L_{\boldsymbol{R}}$, spełniających zależności sformułowane w* b) *i takich, że a jest symetryczna, b — antysymetryczna, definiuje iloczyn skalarny na L za pomocą wzoru*

$$g(l, m) = a(l, m) + ib(l, m);$$

d) *forma g jest dodatnio określona wtedy i tylko wtedy, gdy forma a jest dodatnio określona.*

Dowód. Warunek hermitowskiej symetrii $g(l, m) = \overline{g(m, l)}$ jest równoważny z warunkiem

$$a(l, m) + ib(l, m) = a(m, l) - ib(m, l)$$

symetrii a i antysymetrii b. Warunek $g(il, im) = i\bar{i}g(l, m) = g(l, m)$ jest równoważny z i-niezmienniczością a i b. Warunek $\boldsymbol{C}$-liniowości g względem pierwszego argumentu pociąga za sobą $\boldsymbol{R}$-liniowość oraz liniowość względem mnożenia przez i, tj.

$$a(il, m) + ib(il, m) = g(il, m) = ig(l, m) = -b(l, m) + ia(l, m),$$

skąd wynikają zależności b) i punkt c). Wreszcie d) wynika stąd, że $g(l, l) = a(l, l)$ na mocy antysymetrii b.

3. Wniosek. *W oznaczeniach punktu poprzedzającego, jeśli g jest dodatnio określona, a $\{e_1, \ldots, e_n\}$ jest bazą ortonormalną dla g, to $\{e_1, \ldots, e_n, ie_1, \ldots, ie_n\}$ jest bazą ortonormalną dla a i symplektyczną dla b.*

Odwrotnie, niech L będzie $2n$-wymiarową przestrzenią rzeczywistą z wyróżnioną formą euklidesową a i formą symplektyczną b oraz bazą $\{e_1, \ldots, e_n, e_{n+1}, \ldots, e_{2n}\}$, która jest bazą ortonormalną dla a i bazą symplektyczną dla b. Wówczas, wprowadzając na L strukturę zespoloną za pomocą operatora

$$J(e_j) = e_{n+j}, \quad 1 \leq j \leq n; \quad J(e_j) = -e_{j-n}, \quad n+1 \leq j \leq 2n,$$

oraz iloczyn skalarny za pomocą $g(l, m) = a(l, m) + ib(l, m)$, otrzymamy przestrzeń zespoloną z dodatnio określoną formą hermitowską g, w której $\{e_1, \ldots, e_n\}$ będzie bazą ortonormalną nad $\boldsymbol{C}$.

Dowód, który wymaga jedynie prostego sprawdzenia opartego na poprzedzającym stwierdzeniu, pozostawiamy czytelnikowi.

Wrócimy teraz do rozważania przestrzeni unitarnej L. Nierówność Cauchy'ego-Buniakowskiego–Schwarza w przypadku zespolonym przybiera następującą postać:

4. Stwierdzenie. *Dla każdych $l_1, l_2 \in L$ zachodzi nierówność*

$$|(l_1, l_2)|^2 \leq |l_1|^2 \cdot |l_2|^2,$$

a przy tym równość zachodzi wtedy i tylko wtedy, gdy wektory l_1, l_2 są proporcjonalne.

Dowód. Tak jak w § 5, p. 2, dla każdego rzeczywistego t zachodzi

$$|tl_1 + l_2|^2 = t^2|l_1|^2 + 2t\,\mathrm{Re}(l_1, l_2) + |l_2|^2 \geq 0.$$

Przypadek $l_1 = 0$ jest trywialny. Natomiast jeśli $l_1 \neq 0$, to widzimy, że

$$(\mathrm{Re}(l_1, l_2))^2 \leq |l_1|^2\,|l_2|^2.$$

Jeśli $(l_1, l_2) = |(l_1, l_2)|e^{i\varphi}$, $\varphi \in \boldsymbol{R}$, to $\mathrm{Re}(e^{-i\varphi}l_1, l_2) = |(l_1, l_2)|$. Zatem

$$|(l_1, l_2)|^2 \leq |e^{-i\varphi}l_1|^2|l_2|^2 \leq |l_1|^2|l_2|^2.$$

Równość jest osiągana wtedy i tylko wtedy, gdy $|t_0e^{-i\varphi}l_1 + l_2| = 0$ dla pewnego $t_0 \in \boldsymbol{R}$, co kończy dowód.

Dokładnie tak samo, jak w przypadku euklidesowym można wyprowadzić wnioski:

5. Wniosek (nierówność trójkąta). *Dla każdych* $l_1, l_2 \in L$

$$|l_1 + l_2| \leq |l_1| + |l_2|, \quad |l_1 - l_3| \leq |l_1 - l_2| + |l_2 - l_3|.$$

6. Wniosek. *Długość wektora* $|l|$ *określona za pomocą struktury unitarnej w* L *jest normą w* L *w znaczeniu określenia § 10, p. 4, części 1.*
(Tu niewielkiej zmiany wymaga dowód własności $|al| = |a| \cdot |l|$:

$$|al| = (al, al)^{1/2} = (a\bar{a}(l, l))^{1/2} = |a| \cdot |l|.)$$

7. Kąty. Niech $l_1, l_2 \in L$ będą wektorami niezerowymi. Na mocy p. 4 mamy

$$0 \leq \frac{|(l_1, l_2)|}{|l_1||l_2|} \leq 1.$$

Zatem istnieje jednoznacznie wyznaczony kąt φ, $0 \leq \varphi \leq \pi/2$, dla którego

$$\cos\varphi = \frac{|(l_1, l_2)|}{|l_1||l_2|}.$$

Jednakże w tych najważniejszych modelach przyrodniczych, gdzie stosuje się formalizm przestrzeni unitarnych, liczbę $\frac{|(l_1, l_2)|}{|l_1||l_2|}$ interpretuje się nie jako cosinus kąta, ale jako *prawdopodobieństwo*. Opiszemy pokrótce postulaty mechaniki kwantowej, które wykorzystują taką właśnie interpretację.

8. Przestrzeń stanów układu kwantowego. W mechanice kwantowej przyjmuje się, że z takimi układami fizycznymi jak elektron, atom wodoru i in. można związać (choć niejednoznacznie!) model matematyczny składający się z następujących danych:

a) przestrzeni unitarnej $\mathcal{H}$ nazywanej *przestrzenią stanów układu*. Te przestrzenie, które można znaleźć w standardowych podręcznikach, w przeważającej części są nieskończenie wymiarowymi przestrzeniami Hilberta, zrealizowanymi jako przestrzenie funkcji określonych na modelu przestrzeni „fizycznej" lub czasoprzestrzeni. Skończenie wymiarowe przestrzenie $\mathcal{H}$ pojawiają się jako przestrzenie *wewnętrznych stopni swobody układu*, jeśli rozważać go jako układ zlokalizowany lub też jeśli z takich czy innych względów można zaniedbać znaczenie jego ruchu w przestrzeni fizycznej. Taką przestrzenią jest np. dwuwymiarowa przestrzeń unitarna „stanów spinowych" elektronu, do której jeszcze wrócimy.

b) promieni, tj. jednowymiarowych podprzestrzeni zespolonych przestrzeni $\mathcal{H}$, które są nazywane (czystymi) *stanami układu*.

Całą dostępną informację o stanie układu w ustalonej chwili można przekazać podając promień $L \subset \mathcal{H}$ lub niezerowy wektor $\psi \in L$, który bywa także nazywany *ψ-funkcją* opisującą dany stan lub jeszcze inaczej — *wektorem stanu.*

Podstawowy postulat, głoszący, że wektory stanu stanowią zespoloną przestrzeń liniową, jest nazywany *zasadą superpozycji*, a kombinacja liniowa $\sum_{j=1}^{n} a_j\psi_j$, $a_j \in \boldsymbol{C}$, opisuje *superpozycję* stanów $\psi_1, \ldots, \psi_n$. Ze względu na to, że fizyczne znaczenie mają nie same wektory ψ_j, a tylko promienie $\boldsymbol{C}\psi_j$, współczynniki a_j nie są jednoznacznie określone. Jednakże jeśli przyjąć, że wektory ψ_j są unormowane, tj. $|\psi_j| = 1$, i liniowo niezależne oraz jeśli unormować $\sum_{j=1}^{n} a_j\psi_j$, to dowolność przy wyborze wektora ψ_j z zadanego promienia redukuje się do pomnożenia wektora przez liczbę $e^{i\varphi_j}$, nazywaną *czynnikiem fazowym*; taka jest też dowolność w wyborze współczynników a_j, a więc możemy przyjąć, że te współczynniki są rzeczywiste i nieujemne, a to razem z warunkiem unormowania $|\sum_{j=1}^{n} a_j\psi_j| = 1$ określa je jednoznacznie.

Bardzo wyidealizowane założenie o związku tego schematu z rzeczywistością polega na przyjęciu, że dysponujemy przyborami laboratoryjnymi A_ψ („piecami"), pozwalającymi przygotować wiele egzemplarzy naszego układu w chwilowym stanie ψ (dokładniej $\boldsymbol{C}\psi$) dla różnych $\psi \in \mathcal{H}$ .Ponadto dysponujemy przyrządami laboratoryjmymi B_χ („filtrami"), do których wprowadzamy układ w stanie ψ, a po wyjściu z niego otrzymujemy układ w, być może różnym od wejściowego, stanie χ lub też nie otrzymujemy niczego (układ nie został przepuszczony przez filtr B_χ).

Następny po zasadzie superpozycji zasadniczy postulat mechaniki kwantowej głosi, że *układ, który został przygotowany w stanie $\psi \in \mathcal{H}$ może zostać bezpośrednio potem zaobserwowany w stanie $\chi \in \mathcal{H}$ z prawdopodobieństwem*

$$\frac{|(\psi, \chi)|^2}{|\psi|^2|\chi|^2} = \cos^2\theta, \quad \textit{gdzie } \theta \textit{ jest kątem między } \psi \textit{ a } \chi.$$

W dalszym ciągu, w miarę wprowadzania nowych pojęć geometrycznych, będziemy w stanie sprecyzować matematyczne znaczenie „pieców" i „filtrów". Ponadto będziemy w stanie objaśnić, co się dzieje, gdy układ przygotowany w stanie ψ wprowadzić do filtru nie bezpośrednio po przygotowaniu, ale po upływie czasu t: okaże się, że w tym czasie stan ψ, a wraz z nim i iloczyn skalarny (ψ, χ) będą się zmieniać i że tę zmianę można również w piękny sposób opisać w terminach algebry liniowej.

Jeśli ψ, χ są unormowane, to wspomniane wyżej prawdopodobieństwo równe jest $|(\psi, \chi)|^2$, a sam iloczyn skalarny (ψ, χ), będący liczbą zespoloną, nazywa się *amplitudą prawdopodobieństwa* (przejścia od ψ do χ). Zauważmy, że w ślad za Diakiem fizycy zazwyczaj rozważają iloczyn skalarny jako funkcję antyliniową względem *pierwszego* argumentu i zapisują nie w postaci (ψ, χ), a w postaci $\langle\chi \mid \psi\rangle$, tak że początkowy i końcowy stan układu są zapisane w kolejności od prawej strony do lewej. Nawiasy $\langle\ \rangle$ nazywają się po angielsku „bracket". W związku z tym Dirac nazwał symbol $|\psi\rangle$ — „ket-wektorem", a symbol $\langle\chi|$ — „bra-wektorem". Z mate-

matycznego punktu widzenia $|\psi\rangle$ przedstawia element $\mathcal{H}$, a $\langle\chi|$ odpowiadający mu element przestrzeni antyliniowych funkcjonałów $\overline{\mathcal{H}}^*$, a $\langle\chi\,|\rangle$ jest wartością χ na ψ.

Jeśli ψ, χ są ortogonalne, tj. $(\psi, \chi) = 0$, to układu przygotowanego w stanie ψ nie uda się (bezpośrednio po przygotowaniu) zaobserwować w stanie χ, tj. filtr B_χ nie przepuści tego układu (ale przejdzie on z prawdopodobieństwem 1 przez filtr B_ψ). We wszystkich pozostałych przypadkach istnieje niezerowe prawdopodobieństwo przejścia od ψ do χ.

Elementy dowolnej ortonormalnej bazy $\{\psi_1, \ldots, \psi_n\}$ stanowią *system stanów bazowych* układu. Załóżmy, że rozporządzamy filtrami $B_{\psi_1}, \ldots, B_{\psi_n}$. Przy wielokrotnym przepuszczaniu przez te filtry układów przygotowanych w stanie $\psi = \sum_{i=1}^{n} a_i\psi_i$, $0 \leq a_i \leq 1$ (przy założeniu, że wektor jest unormowany), zaobserwujemy stan ψ_i z prawdopodobieństwem a_i^2. W ten sposób współczynniki tej kombinacji liniowej można wyznaczyć eksperymentalnie, ale w eksperymencie o zasadniczo statystycznym charakterze. Jest to jeden z powodów, dla których pomiary w mechanice kwantowej wymagają opracowania wielkiej liczby danych statystycznych. Jednakże bywa i tak, że układy w stanie ψ są przepuszczane przez filtr jednym strumieniem i otrzymane przy wyjściu prawdopodobieństwa a_i^2 są mierzone jako natężenia czegoś w rodzaju „linii spektralnych" — te natężenia są już same rezultatem uśredniania statystycznego. W dalszym ciągu uściślimy związek tego schematu z teorią spektralną operatorów liniowych.

9. Reguły Feynmana. Załóżmy, że $\{\psi_1, \ldots, \psi_n\}$ jest ustaloną bazą ortonormalną przestrzeni stanów $\mathcal{H}$. Dowolny wektor $\psi \in \mathcal{H}$ możemy zapisać

$$\psi = \sum_{i=1}^{n} (\psi, \psi_i)\psi_i,$$

skąd otrzymujemy

$$(\psi, \chi) = \sum_{i=1}^{n} (\psi, \psi_i)(\psi_i, \chi).$$

Analogicznie mamy $(\psi, \psi_i) = \sum_{j=1}^{n} (\psi, \psi_j)(\psi_j, \psi_i)$ i podstawiając ten wzór do poprzedniego, otrzymujemy

$$(\psi, \chi) = \sum_{i_1, i_2=1}^{n} (\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2})(\psi_{i_2}, \chi),$$

a ogólniej dla dowolnego $m \geq 1$

$$(\psi, \chi) = \sum_{i_1, \ldots, i_m=1}^{n} (\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2}) \cdots (\psi_{i_1}, \chi).$$

Te proste wzory algebry liniowej można interpretować, zgodnie z sugestią Feynmana, jako prawa „zespolonej teorii prawdopodobieństwa", odnoszącej się do amplitud

zamiast do wartości prawdopodobieństw. Rzeczywiście, potraktujmy ciągi postaci $(\psi, \psi_{i_1}, \psi_{i_2}, \ldots, \psi_{i_m}, \chi)$ jako „klasyczne trajektorie" układu przechodzącego kolejno przez stany zawarte w nawiasie, a liczbę $(\psi, \psi_{i_1})(\psi_{i_1}, \psi_{i_2}) \cdots (\psi_{i_m}, \chi)$ — jako amplitudę prawdopodobieństwa przejścia od ψ do χ wzdłuż odpowiedniej trajektorii klasycznej, zadaną jako iloczyn amplitud prawdopodobieństwa przejść wzdłuż kolejnych odcinków trajektorii.

Przy takiej interpretacji podany wyżej wzór dla (ψ, χ) oznacza, że *ta amplituda prawdopodobieństwa przejścia jest sumą amplitud prawdopodobieństw przejścia od ψ do χ po wszystkich dostępnych trajektoriach klasycznych* („o jednakowej długości").

Nieskończenie wymiarowy i dużo bardziej wyrafinowany wariant tej obserwacji, w którym główną rolę odgrywają czasoprzestrzenne (lub pędowo-energetyczne) wielkości obserwowalne, został przyjęty przez R. Feynmana jako podstawa jego na wpół heurystycznej techniki wyrażania amplitud przez „całki funkcjonalne po trajektoriach klasycznych". Przestrzeń trajektorii jest nieskończenie wymiarową przestrzenią funkcyjną, a matematykom nie udało się do tej pory zbudować takiej teorii, która nadawałaby ścisły sens tym wszystkim zadziwiającym rachunkom fizyków.

10. Odległości. Odległość między podzbiorami w przestrzeni unitarnej L można zdefiniować dokładnie tak samo, jak w przestrzeni euklidesowej:

$$d(U, V) = \inf\{|l_1 - l_2| \mid l_1 \in U, l_2 \in V\}.$$

Odległość wektora l od podprzestrzeni L_0 także jest równa długości rzutu ortogonalnego l na $L_0^{\perp}$. Dowód tego niczym nie różni się od dowodu w przypadku euklidesowym. W szczególności, jeśli $\{e_1, \ldots, e_m\}$ oznacza bazę ortonormalną L_0, to

$$d(l, L_0) = \left| l - \sum_{i=1}^{m} (l, e_i) e_i \right|,$$

tak jak w przypadku euklidesowym, a także

$$\left| \sum_{i=1}^{m} (l, e_i) e_i \right|^2 = \sum_{i=1}^{m} |(l, e_i)|^2 \leq |l|^2$$

na mocy twierdzenia Pitagorasa.

11. Zastosowanie do przestrzeni funkcyjnych. Tak jak w § § 4, 5 można wykazać następujące nierówności spełniane przez funkcje o wartościach zespolonych:

$$\left| \int_a^b f(x) g(x)\, dx \right|^2 \leq \int_a^b |f(x)|^2\, dx \int_a^b |g(x)|^2\, dx,$$

$$\left(\int_a^b |f(x) + g(x)|^2\, dx \right)^{1/2} \leq \left(\int_a^b |f(x)|^2\, dx \right)^{1/2} + \left(\int_a^b |g(x)|^2\, dx \right)^{1/2}$$

a także przez ich współczynniki Fouriera. Dla funkcji na odcinku $[0, 2\pi]$ określimy

$$a_n = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)e^{-inx}\,dx,$$

a wobec tego, jeśli rozważać przestrzeń tych funkcji z iloczynem skalarnym zadanym wzorem $\int_0^{2\pi} f(x)\overline{g(x)}\,dx$, suma

$$f_N(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} a_n e^{inx}$$

wyraża rzut ortogonalny f na przestrzeń wielomianów Fouriera „stopnia $\leq N$" i minimalizuje średnie kwadratowe odchylenie f od tej przestrzeni. W szczególności,

$$\sum_{n=-N}^{N} |a_n|^2 \leq \int_0^{2\pi} |f(x)|^2\,dx,$$

co pokazuje, że szereg $\sum_{-\infty}^{\infty} |a_n|^2$ jest zbieżny.

§ 7. Operatory ortogonalne i unitarne

1. Niech L będzie przestrzenią liniową z iloczynem skalarnym g. Zbiór wszystkich izometrii $f\colon L \to L$, tj. odwzorowań liniowych odwracalnych, które spełniają warunek

$$g(f(l_1), f(l_2)) = g(l_1, l_2)$$

dla wszystkich $l_1, l_2 \in L$, oczywiście stanowi grupę. Jeśli L jest przestrzenią euklidesową, to takie operatory nazywamy *ortogonalnymi*, a jeśli L jest unitarne, to *unitarnymi*. Symplektyczne izometrie będą rozważane później.

2. Stwierdzenie. *Niech L będzie skończenie wymiarową przestrzenią liniową z niezdegenerowanym iloczynem skalarnym $(\ ,\ )$, symetrycznym lub hermitowskim. Na to, by operator $f\colon L \to L$ był izometrią, potrzeba i wystarcza, by spełniał jeden z następujących warunków:*

a) $(f(l), f(l)) = (l, l)$ *dla każdego $l \in L$ (tu zakładamy, że charakterystyka ciała skalarów jest różna od 2);*

b) *Niech $\{e_1, \ldots, e_n\}$ oznacza bazę L z macierzą Grama G, a A — macierz operatora f w tej bazie. Wtedy*

$$A^t G A = G \quad \textit{lub} \quad A^t G \overline{A} = G;$$

c) *f odwzorowuje jakąś bazę ortonormalną na bazę ortonormalną;*

d) *jeśli sygnatura iloczynu skalarnego wynosi (p, q), to macierz operatora f w dowolnej bazie ortonormalnej $\{e_1, \ldots, e_p, e_{p+1}, \ldots, e_{p+q}\}$, gdzie $(e_i, e_i) = +1$, gdy $i \leq p$, oraz $(e_i, e_i) = -1$, gdy $p+1 \leq i \leq p+q$, spełnia warunek*

$$A^t \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix} A = \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix}$$

lub

$$A^t \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix} \overline{A} = \begin{pmatrix} E_p & 0 \\ 0 & -E_q \end{pmatrix}$$

odpowiednio w symetrycznym lub hermitowskim przypadku.

D o w ó d. a) W przypadku symetrycznym teza wynika z § 3, p. 9: jeśli f zachowuje formę kwadratową $(l, l) = q(l)$, to zachowuje jej formę spolaryzowaną

$$(l, m) = \frac{1}{2}[q(l+m) - q(l) - q(m)].$$

W przypadku hermitowskim mamy analogicznie

$$\mathrm{Re}(l, m) = \frac{1}{2}[q(l+m) - q(l) - q(m)].$$

a stwierdzenie § 6, p. 2 pokazuje, że (l, m) jest jednoznacznie wyznaczone przez $\mathrm{Re}(l, m)$ za pomocą wzoru

$$(l, m) = \mathrm{Re}(l, m) - i\,\mathrm{Re}(il, m),$$

a więc f zachowuje także (l, m).

b) Jeśli f jest izometrią, to macierze Grama baz $\{e_1, \ldots, e_n\}$ i $\{f(e_1), \ldots, f(e_n)\}$ są jednakowe. Ostatnia z tych macierzy jest równa A^tGA w przypadku symetrycznym i $A^tG\overline{A}$ w przypadku hermitowskim. Odwrotnie, jeśli f odwzorowuje bazę $\{e_1, \ldots, e_n\}$ na bazę $\{e'_1, \ldots, e'_n\}$ i macierze Grama obu tych baz są jednakowe, to f jest izometrią na mocy wzorów z § 2, p. 2 wyrażających iloczyn skalarny we współrzędnych.

Punkty c) i d) są szczególnymi przypadkami poprzednich.

Ze stwierdzenia p. 2 wynika, że operatory ortogonalne (odpowiednio unitarne) to takie operatory, które w jakiejś (a więc także w każdej) bazie *ortonormalnej* wyrażają się poprzez macierze *ortogonalne* (odpowiednio *unitarne*), tj. takie macierze U, które spełniają warunki

$$UU^t = E_n \quad \text{lub} \quad U\overline{U}^t = E_n.$$

Zbiory macierzy tego typu zostały wprowadzone już w § 4 części 1, gdzie oznaczane były przez $\boldsymbol{O}(n)$ i $\boldsymbol{U}(n)$ odpowiednio. Macierze izometrii w ortonormalnych bazach z sygnaturą (p, q), spełniające warunek d) stwierdzenia 2 oznaczamy analogicznie

przez $\boldsymbol{O}(p,q)$ i $\boldsymbol{U}(p,q)$; dla $p,q \neq 0$ nazywane są też czasem *pseudoortogonalnymi* lub *pseudounitarnymi* macierzami odpowiednio. W tym paragrafie będą nas interesować tylko grupy $\boldsymbol{O}(n)$ i $\boldsymbol{U}(n)$. Podstawową w fizyce grupę $\boldsymbol{O}(1,3)$ omówimy w § 12.

3. Grupy $\boldsymbol{U}(1)$, $\boldsymbol{O}(1)$ i $\boldsymbol{O}(2)$. Wprost z określenia wynika, że

$$\boldsymbol{U}(1) = \{a \in \boldsymbol{C} \mid |a| = 1\} = \{e^{i\varphi} \mid \varphi \in \boldsymbol{R}\},$$

$$\boldsymbol{O}(1) = \{\pm 1\} = \boldsymbol{U}(1) \cap \boldsymbol{R}.$$

Dalej dla $U \in \boldsymbol{O}(n)$ z $UU^t = E_n$ wynika, że $(\det U)^2 = 1$, więc $\det U = \pm 1$. Jeśli $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ jest macierzą ortogonalną o wyznaczniku -1, to $\begin{pmatrix} a & b \\ -c & -d \end{pmatrix}$ jest macierzą ortogonalną o wyznaczniku 1 należącą do $\boldsymbol{SO}(2)$. Macierze z $\boldsymbol{SO}(2)$ można zapisać w postaci

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; ad - bc = a^2 + b^2 = c^2 + d^2 = 1, ac + bd = 0 \right\}.$$

Oczywiście, dowolną taką macierz można także zapisać w postaci

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \quad \varphi \in [0, 2\pi),$$

tj. w postaci macierzy odpowiadającej euklidesowemu obrotowi o kąt φ. Odwzorowanie

$$\boldsymbol{U}(1) \longrightarrow \boldsymbol{SO}(2) : \; e^{i\varphi} \mapsto \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

jest izomorfizmem. Jego sens geometryczny można objaśnić, odwołując się do następującej obserwacji: realifikacja jednowymiarowej przestrzeni unitarnej $(\boldsymbol{C}^1, |z|^2)$ jest dwuwymiarową przestrzenią euklidesową $(\boldsymbol{R}^2, x_1^2 + x_2^2)$, a realifikacja odwzorowania unitarnego $z \mapsto e^{i\varphi} z$ wyraża się poprzez macierz obrotu o kąt φ.

W paragrafie 11 skonstruujemy znacznie mniej trywialną surjekcję (epimorfizm) $\boldsymbol{SU}(2) \to \boldsymbol{SO}(3)$ z jądrem $\{\pm 1\}$.

Obroty $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ o kąt $\varphi \neq 0, \pi$ nie mają wektorów własnych w $\boldsymbol{R}^2$, a zatem nie są diagonalizowalne. Przeciwnie, każda macierz $U \in \boldsymbol{O}(2)$ z $\det U = -1$ jest diagonalizowalna. Dokładniej mówiąc, wielomian charakterystyczny macierzy

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ -\sin\varphi & -\cos\varphi \end{pmatrix}$$

jest równy $t^2 - 1$ i ma pierwiastki ± 1. Łatwo sprawdzić bezpośrednim rachunkiem, że odpowiadające tym pierwiastkom podprzestrzenie własne są ortogonalne – później

wykażemy to w dużo większej ogólności. Dlatego każdy operator z $\boldsymbol{O}(2)$ o wyznaczniku równym -1 jest *odbiciem* względem pewnej prostej: działa jako identyczność na wektory tej prostej i zmienia znak wektorów do niej ortogonalnych.

Posługując się tą informacją możemy teraz opisać w pełnej ogólności strukturę operatorów ortogonalnych i unitarnych.

4. Twierdzenie. a) *Na to, by operator f w przestrzeni unitarnej był unitarny, potrzeba i wystarcza, by można było go sprowadzić do postaci diagonalnej przez wybór bazy ortonormalnej, a jego widmo było zawarte w okręgu jednostkowym $\boldsymbol{C}$.*

b) *Na to, by operator f w przestrzeni euklidesowej był ortogonalny, potrzeba i wystarcza, by w pewnej bazie ortonormalnej jego macierz miała postać blokowo diagonalną*

$$\begin{pmatrix} A(\varphi_1) & & & & & & & \\ & \ddots & & & & \mathbf{0} & & \\ & & A(\varphi_m) & & & & & \\ & & & 1 & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & & \\ & & & & & & -1 & \\ & \mathbf{0} & & & & & & \ddots \\ & & & & & & & & -1 \end{pmatrix},$$

gdzie $A(\varphi)$ oznacza macierz postaci

$$A(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

a pozostałe elementy są równe zeru.

c) *Wektory własne ortogonalnego lub unitarnego operatora, odpowiadające różnym wartościom własnym są do siebie ortogonalne.*

Dowód. a) Dostateczność warunku jest oczywista: jeśli $U = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, $|\lambda_i|^2 = 1$, to $U\overline{U}^t = E_n$, a więc U jest macierzą operatora unitarnego. Odwrotnie, niech f będzie operatorem unitarnym, λ — jego wartością własną odpowiadającą podprzestrzeni własnej L_λ. Ze stwierdzenia § 3, p. 2 wynika, że $L = L_\lambda \oplus L_\lambda^\perp$. Podprzestrzeń L_λ jest jednowymiarowa, f-niezmiennicza, a obcięcie f do L_λ jest operatorem unitarnym w przestrzeni jednowymiarowej, więc $\lambda \in \boldsymbol{U}(1)$, tj. $|\lambda|^2 = 1$. Jeśli wykażemy, że podprzestrzeń $L_\lambda^\perp$ jest także f-niezmiennicza, to dalej przez indukcję względem $\dim L$ wynika, że L rozkłada się na sumę prostą f-niezmienniczych, parami ortogonalnych, jednowymiarowych podprzestrzeni, co dowiedzie tezy a).

Załóżmy więc, że $l_0 \in L_\lambda$, $l_0 \neq 0$ i $(l_0, l) = 0$. Wtedy

$$(l_0, f(l)) = (f(\lambda^{-1} l_0), f(l)) = (\lambda^{-1} l_0, l) = \lambda^{-1}(l_0, l) = 0$$

pokazuje, że $f(l) \in L_\lambda^\perp$.

b) W przypadku ortogonalnym rozumowanie jest analogiczne; dostateczność warunku sprawdzamy wprost, a dalej dowodzimy przez indukcję względem wymiaru $\dim L$. Przypadki $\dim L = 1$ i 2 zostały już rozpatrzone w poprzednim punkcie. Jeśli $\dim L \geq 3$ i f ma rzeczywistą wartość własną λ, to należy znowu przyjąć $L = L_\lambda \oplus L_\lambda^\perp$ i argumentować jak wyżej (zauważmy przy tym, że koniecznie $\lambda = \pm 1$). Jeśli natomiast f nie ma rzeczywistych wartości własnych, to należy wybrać dwuwymiarową f-niezmienniczą podprzestrzeń własną $L_0 \subset L$, której istnienie otrzymujemy ze stwierdzenia p. 16, § 12 części 1. Macierz obcięcia f do tej podprzestrzeni w dowolnej bazie ortonormalnej ma postać $A(\varphi)$ na mocy poprzedzającego punktu. Pozostaje zatem do wykazania tylko f-niezmienniczość $L_0^\perp$. I rzeczywiście, jeśli $(l_0, l) = 0$ dla wszystkich $l_0 \in L_0$, to

$$(l_0, f(l)) = (f(f^{-1}(l_0)), f(l)) = (f^{-1}(l_0), l) = 0,$$

bowiem $f^{-1}(l_0) \in L_0$ dla każdego $l_0 \in L_0$. To kończy dowód części b).

c) Niech $f(l_i) = \lambda_i l_i$, $i = 1, 2$. Wtedy

$$(l_1, l_2) = (f(l_1), f(l_2)) = \lambda_1 \bar{\lambda}_2 (l_1, l_2).$$

Ponieważ $|\lambda_i|^2 = 1$, więc dla $\lambda_1 \neq \lambda_2$ mamy $\lambda_1 \bar{\lambda}_2 \neq 1$, a zatem $(l_1, l_2) = 0$. Ponieważ ten argument ma zastosowanie zarówno w unitarnym, jak i ortogonalnym przypadku, dowód został zakończony.

5. Wniosek („twierdzenie Eulera"). *W trójwymiarowej przestrzeni euklidesowej każde nie zmieniające orientacji odwzorowanie ortogonalne (tj. element grupy* $\boldsymbol{SO}(3)$*), jest obrotem wokół pewnej osi.*

Dowód. Ponieważ stopień wielomianu charakterystycznego f wynosi 3, więc ma on przynajmniej jeden pierwiastek rzeczywisty. Jeśli jest tylko jeden taki pierwiastek, to musi być równy 1, bo $\det f = 1$. Jeśli jest ich więcej, to wszystkie pierwiastki wielomianu charakterystycznego są rzeczywiste i możliwe są następujące kombinacje ich wartości: $(1, 1, 1)$ lub $(1, -1, -1)$. W każdym przypadku 1 jest wartością własną f, a odpowiadająca jej podprzestrzeń własna jest osią obrotu, natomiast w płaszczyźnie do niej ortogonalnej indukowane jest odwzorowanie o macierzy z $\boldsymbol{SO}(2)$, tj. obrót płaszczyzny.

§ 8. Operatory samosprzężone

1. W pierwszej części zaobserwowaliśmy, że najprostszą i jednocześnie najważniejszą klasę operatorów liniowych stanowią operatory diagonalizowalne. Okazuje się, że w przestrzeniach euklidesowych i unitarnych wyróżnioną rolę odgrywają operatory o widmie rzeczywistym, diagonalizowalne w pewnej bazie ortonormalnej. Są to te operatory, które opisują rozciągnięcia przestrzeni ze współczynnikami rzeczywistymi w kierunkach osi układu ortonormalnego.

Niech $\{e_1, \ldots, e_n\}$ będzie bazą ortonormalną L i $f: L \to L$ operatorem spełniającym $f(e_i) = \lambda_i e_i$, $\lambda_i \in \boldsymbol{R}$, $i = 1, \ldots, n$. Łatwo przekonać się, że taki operator ma następującą prostą własność:

$$(f(l_1), l_2) = (l_1, f(l_2)) \quad \text{dla każdej pary } l_1,\ l_2 \in L. \tag{1}$$

Rzeczywiście,

$$\Big(f(\sum_{i=1}^{n} x_i e_i), \sum_{i=1}^{n} y_i e_i\Big) = \sum_{i=1}^{n} \lambda_i x_i y_i \quad \text{lub} \quad \sum_{i=1}^{n} \lambda_i x_i \overline{y}_i,$$

$$\Big(\sum_{i=1}^{n} x_i e_i,\ f(\sum_{i=1}^{n} y_i e_i)\Big) = \sum_{i=1}^{n} \lambda_i x_i y_i \quad \text{lub} \quad \sum_{i=1}^{n} \lambda_i x_i \overline{y}_i$$

(ten fakt, że w przypadku unitarnym $\lambda_i \in \boldsymbol{R}$ został wykorzystany w drugim wzorze). Operatory o własności 1 nazywane są *samosprzężonymi*, a więc pokazaliśmy, że *operatory o widmie rzeczywistym, diagonalizowalne w ortonormalnej bazie, są samosprzężone*. Wkrótce wykażemy stwierdzenie odwrotne, ale najpierw zbadamy własność samosprzężoności bardziej systematycznie.

2. Operatory samosprzężone w przestrzeni z formą dwuliniową. W pierwszej części wykładu wykazaliśmy, że dla każdego odwzorowania liniowego $f: L \to M$ istnieje jedyne odwzorowanie liniowe $f^*: M^* \to L^*$, takie że

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

gdzie $m^* \in M^*$, $l \in L$ i gdzie nawiasy oznaczają kanoniczne odwzorowania dwuliniowe $L^* \times L \to \boldsymbol{K}$, $M^* \times M \to \boldsymbol{K}$.

W szczególności, w przypadku $M = L$ operatorowi $f: L \to L$ odpowiada operator $f^*: L^* \to L^*$. Załóżmy teraz, że na L jest zadana niezdegenerowana forma dwuliniowa $g: L \times L \to \boldsymbol{K}$, która określa izomorfizm $\tilde{g}: L \to L^*$. Wówczas, utożsamiając L^* z L za pomocą $\tilde{g}^{-1}$, możemy traktować operator f^*, a w rzeczywistości $\tilde{g}^{-1} \circ f^* \circ \tilde{g}$, jako operator w przestrzeni L. Będziemy go w dalszym ciągu oznaczać f^*, mimo że byłoby właściwiej oznaczać go np. f_g^*, ale operator f^* w poprzednim znaczeniu nie będzie więcej występował w tym paragrafie. Oczywiście, nowy operator f^* jest jednoznacznie wyznaczony wzorem

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

Tak jak poprzednio, będziemy go nazywać sprzężonym z f (względem iloczynu skalarnego g).

W przypadku półtoraliniowym $\tilde{g}$ jest izomorfizmem L z $\overline{L}^*$, a nie z L^*. Dlatego używając $\tilde{g}$ trzeba przenosić do L operator $\overline{f}^*: \overline{L}^* \to \overline{L}^*$ zdefiniowany jako $\overline{f}^*(m) = f^*(m)$. Przeniesiony operator $\tilde{g}^{-1} \circ \overline{f}^* \circ \tilde{g}: L \to L$ jest liniowy. Należałoby oznaczyć go f^+, ale zachowamy tradycyjne oznaczenie f^*. Wtedy także w półtoraliniowym przypadku będzie spełniona równość

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

Operacja $f \mapsto f^*$ jest liniowa, jeśli g jest biliniowa, a antyliniowa, jeśli g jest półtoraliniowa.

Operatory $f: L \to L$ o własności $f^* = f$ w euklidesowych i skończenie wymiarowych przestrzeniach unitarnych nazywane są *samosprzężonymi*, w przypadku euklidesowym nazywa się je też *symetrycznymi*, a w unitarnym — *hermitowskimi*. Tę terminologię wyjaśnia następująca prosta obserwacja.

3. Stwierdzenie. *Jeśli operator $F: L \to L$ ma w bazie ortonormalnej macierz A, to operator A^* ma w tej samej bazie macierz A^t w przypadku euklidesowym i $\overline{A}^t$ w przypadku unitarnym.*

W szczególności, operator jest samosprzężony wtedy i tylko wtedy, gdy jego macierz w pewnej bazie jest symetryczna lub hermitowska.

Dowód. Oznaczając iloczyn skalarny w L nawiasem i reprezentując wektory za pomocą wektorów-kolumn ich współrzędnych względem bazy ortonormalnej, mamy

$$(f(\vec{x}), \vec{y}) = (A\vec{x})^t\vec{y} = (\vec{x}^t A^t)\vec{y} = \vec{x}(A^t\vec{y}) = (\vec{x}, f^*(\vec{y}))$$

(w przypadku euklidesowym). Stąd już wynika, że macierz f^* jest równa A^t — w przypadku unitarnym argument jest analogiczny.

4. Operatory samosprzężone i iloczyny skalarne. Niech L będzie przestrzenią z symetrycznym lub hermitowskim iloczynem skalarnym $(\ ,\ )$. Dla dowolnego operatora $f: L \to L$ możemy określić nowy iloczyn skalarny $(\ ,\ )_f$ na L, przyjmując

$$(l_1, l_2)_f = (f(l_1), l_2).$$

Załóżmy, że L jest niezdegenerowana, tak że można posługiwać się pojęciem operatora sprzężonego. Wtedy

$$(l_2, l_1)_f = (f(l_2), l_1) = (l_2, f^*(l_1)) = (f^*(l_1), l_2) = (l_1, l_2)_f.$$

w przypadku euklidesowym i analogicznie

$$\overline{(l_2, l_1)_f} = \overline{(f(l_2), l_1)} = \overline{(l_2, f^*(l_1))} = (f^*(l_1), l_2) = (l_1, l_2)_f.$$

w przypadku unitarnym. Wnioskujemy stąd, że *jeśli operator f jest samosprzężony, to zbudowana z niego nowa metryka $(l_1, l_2)_f$ będzie jak poprzednio symetryczna lub hermitowska.* Również i odwrócenie tego stwierdzenia jest prawdziwe, o czym nietrudno przekonać się wprost lub przy użyciu stwierdzenia p.3.

W ten sposób skonstruowaliśmy bijekcję zbioru operatorów samosprzężonych na zbiór symetrycznych iloczynów skalarnych w przestrzeni z zadanym niezdegenerowanym iloczynem skalarnym. W przypadku euklidesowym lub unitarnym po dokonaniu wyboru bazy ortonormalnej łatwo opisać tę bijekcję w języku macierzowym: macierz Grama $(\ ,\ )_f$ jest macierzą transponowaną macierzy odwzorowania f.

Teraz wykażemy zasadnicze twierdzenie o operatorach samosprzężonych, analogiczne do twierdzenia o ortogonalnych i unitarnych operatorach z p. 7, § 7 i blisko z nim związane.

5. Twierdzenie. a) *Na to, by operator f w euklidesowej lub skończenie wymiarowej przestrzeni unitarnej był samosprzężony, potrzeba i wystarcza, by można było go zdiagonalizować poprzez wybór bazy ortonormalnej i jego widmo było rzeczywiste.*

b) *Wektory własne samosprzężonego operatora odpowiadające różnym wartościom własnym są ortogonalne.*

Dowód. a) Dostateczność tego warunku sprawdziliśmy na początku tego paragrafu. Łatwo wykazać rzeczywistość widma w przypadku unitarnym: niech λ będzie wartością własną operatora f odpowiadającą wektorowi własnemu $l \in L$. Wtedy

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

skąd $\lambda = \bar{\lambda}$, bo $(l, l) \neq 0$. Przypadek ortogonalny sprowadza się do unitarnego za pomocą następującego triku: rozważmy skompleksyfikowaną przestrzeń L^C i wprowadźmy w niej półtoraliniowy iloczyn skalarny za pomocą wzoru

$$(l_1 + il_2, l_3 + il_4) = (l_1, l_3) + (l_2, l_4) + i(l_2, l_3) - i(l_1, l_4).$$

Łatwo sprawdzić, że w ten sposób L^C otrzymuje strukturę przestrzeni unitarnej, a operator f^C jest hermitowski względem tego iloczynu skalarnego. Widma operatorów f i f^C są identyczne, bo w każdej bazie L nad $\boldsymbol{R}$, będącej jednocześnie bazą L^C nad $\boldsymbol{C}$, macierze f i f^C są jednakowe. Dlatego widmo f jest rzeczywiste.

Dalej obydwa przypadki można rozpatrywać równolegle i zastosować indukcję względem $\dim L$. Przypadek $\dim L = 1$ jest trywialny, załóżmy zatem $\dim L > 1$ i wybierzmy wartość własną λ oraz odpowiadającą jej podprzestrzeń własną L_0 i przyjmijmy $L_1 = L_0^{\perp}$. Na mocy stwierdzenia § 3, p. 2 mamy $L = L_0 \oplus L_1$. Podprzestrzeń L_1 jest niezmiennicza względem f, gdyż jeśli $l_0 \in L_0$, $l_0 \neq 0$ i $l \in L_1$, tj. $(l_0, l) = 0$, to

$$(l_0, f(l)) = (f(l_0), l) = \lambda(l_0, l) = 0,$$

skąd wynika $f(l) \in L_1$. Na mocy założenia indukcyjnego obcięcie f do L_1 można zdiagonalizować, wybierając bazę ortonormalną w L_1. Dołączając do niej wektor $l_0 \in L_0$, $|l_0| = 1$, otrzymujemy szukaną bazę w L.

b) Niech $f(l_1) = \lambda_1 l_1$, $f(l_2) = \lambda_2 l_2$. Wówczas

$$\lambda_1(l_1, l_2) = (f(l_1), l_2) = (l_1, f(l_2)) = \lambda_2(l_1, l_2),$$

skąd wynika, że jeśli $\lambda_1 \neq \lambda_2$, to $(l_1, l_2) = 0$.

6. Wniosek. *Każda rzeczywista symetryczna lub zespolona hermitowska macierz ma widmo rzeczywiste i jest diagonalizowalna.*

Dowód. Dla danej macierzy A skonstruujemy wyznaczony przez nią operator samosprzężony w przestrzeni kartezjańskiej $\boldsymbol{R}^n$ lub $\boldsymbol{C}^n$ wyposażonej w kanoniczną metrykę euklidesową lub hermitowską i zastosujemy twierdzenie p. 5. Wynika z niego nawet więcej, a mianowicie to, że macierz X, taką że $X^{-1}AX$ jest diagonalna, można znaleźć w $\boldsymbol{O}(n)$ lub $\boldsymbol{U}(n)$ odpowiednio.

7. Wniosek. *Odwzorowanie* $\exp\colon \mathfrak{u}(n) \to \boldsymbol{U}(n)$ *jest surjektywne.*

Dowód. Algebra Liego $\mathfrak{u}(n)$ składa się z macierzy antyhermitowskich (por. § 4, część 1), a każdą taką macierz można zapisać w postaci iA dla hermitowskiej macierzy A. Dla znalezienia rozwiązania A równania $\exp(iA) = U$, gdzie $U \in \boldsymbol{U}(n)$, rozważymy U jako operator unitarny f w kartezjańskiej przestrzeni unitarnej $\boldsymbol{C}^n$. Wykorzystując twierdzenie § 7, p. 4, wybierzemy w $\boldsymbol{C}^n$ nową bazę ortonormalną $\{e_1, \ldots, e_n\}$, w której macierz operatora f ma postać $\operatorname{diag}(e^{i\varphi_1}, \ldots, e^{i\varphi_n})$ i zdefiniujemy operator g, przyjmując jako jego macierz w tej bazie macierz $\operatorname{diag}(i\varphi_1, \ldots, i\varphi_n)$. Oznaczając przez A macierz operatora g w wyjściowej (kanonicznej) bazie $\boldsymbol{C}^n$, widzimy, że $\exp(ig) = f$, a zatem $\exp(iA) = U$.

8. Wniosek. a) *Niech* g_1, g_2 *będą dwiema symetrycznymi lub hermitowskimi formami w skończenie wymiarowej przestrzeni* L, *przy czym jedna z nich, powiedzmy* g_1, *jest dodatnio określona. Wówczas w* L *istnieje taka baza, której macierz Grama względem* g_1 *jest macierzą jednostkową, a względem* g_2 *jest diagonalna i rzeczywista.*

b) *Niech* g_1, g_2 *będą dwiema rzeczywistymi symetrycznymi lub zespolonymi hermitowskimi formami względem zmiennych* $x_1, \ldots, x_n; y_1, \ldots, y_n$ *i niech* g_1 *będzie dodatnio określona. Wówczas za pomocą liniowej nieosobliwej zamiany zmiennych (wspólnej dla* $\vec{x}$ *i* $\vec{y}$*) obie te formy można sprowadzić do postaci*

$$g_1(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i y_i, \quad g_2(\vec{x}, \vec{y}) = \sum_{i=1}^{n} \lambda_i x_i y_i, \quad \lambda_i \in \boldsymbol{R},$$

lub

$$g_1(\vec{x}, \vec{y}) = \sum_{i=1}^{n} x_i \overline{y}_i, \quad g_2(\vec{x}, \vec{y}) = \sum_{i=1}^{n} \lambda_i x_i \overline{y}_i, \quad \lambda_i \in \boldsymbol{R},$$

Dowód. Oczywiście oba sformułowanie są równoważne. Dla dowodu rozważymy (L, g_1) jako przestrzeń ortogonalną lub unitarną, przy czym zamiast $g_1(l_1, l_2)$ będziemy pisać (l_1, l_2), a następnie zapiszemy $g_2(l_1, l_2)$ w postaci $(l_1, l_2)_f$, gdzie $f\colon L \to L$ jest operatorem samosprzężonym, podobnie jak w p. 4. Wybierając teraz bazę ortonormalną L, w której f ma postać diagonalną, widzimy, odwołując się do uwagi zrobionej na końcu p. 4, że ta baza spełnia warunki sformułowane we wniosku.

9. Projektory ortogonalne. Niech L będzie przestrzenią liniową nad $\boldsymbol{K}$ i niech będzie zadany jej rozkład $L_1 \oplus L_2$ na sumę prostą. Jak to pokazaliśmy w części 1, taki rozkład wyznacza dwa projektory $p_i\colon L \to L$, takie że $\operatorname{Im} p_i = L_i$, $\mathrm{id}_L = p_1 + p_2$, $p_1 p_2 = p_2 p_1 = 0$, $p_i^2 = p_i$. Wartości własne projektorów są równe 0 lub 1. Jeśli L jest

przestrzenią euklidesową lub unitarną i $L_2 = L_1^{\perp}$, to odpowiadające temu rozkładowi projektory ortogonalne są diagonalne w bazie ortonormalnej L, będącej sumą takich baz dla L_1, L_2, więc są samosprzężone. Także odwrotnie, *każdy samosprzężony projektor p jest operatorem rzutu ortogonalnego na podprzestrzeń.* Rzeczywiście, $\operatorname{Ker} p$ i $\operatorname{Im} p$ są rozpięte przez wektory własne p, odpowiadające wartościom własnym 0, 1 odpowiednio, więc są ortogonalne na mocy twierdzenia p. 5 oraz $L = \operatorname{Ker} p \oplus \operatorname{Im} p$.

Co więcej, jeśli operator samosprzężony f diagonalizuje się w bazie ortonormalnej $\{e_i\}$, tj. $f(e_i) = \lambda_i e_i$, i jeśli oznaczyć przez p_i ortogonalny projektor na podprzestrzeń rozpiętą przez e_i, to zachodzi równość

$$f = \sum_{i=1}^{n} \lambda_i p_i, \tag{2}$$

którą nazywamy *rozkładem spektralnym operatora f.*

Można też przyjąć, że λ_i przebiega tylko parami różne wartości własne, a p_i jest operatorem rzutu ortogonalnego na przestrzeń pierwiastkową $L(\lambda_i)$; wzór (2) pozostaje także wówczas prawdziwy.

Twierdzenie p. 5 daje się uogólnić także na operatory samosprzężone ograniczone względem normy w nieskończenie wymiarowej przestrzeni Hilberta (a z pewnymi komplikacjami także i na nieograniczone). Jednakże takie uogólnienie wymaga bardzo nietrywialnych modyfikacji niektórych podstawowych pojęć. Główne problemy pojawiają się w związku ze strukturą widma: w przypadku skończenie wymiarowym λ jest wartością własną f wtedy i tylko wtedy, gdy operator $\lambda \operatorname{id} - f$ jest nieodwracalny, a tymczasem w przypadku nieskończenie wymiarowym zbiór wartości $\lambda \in \boldsymbol{C}$, dla których operator $\lambda \operatorname{id} - f$ jest osobliwy (tj. nie jest odwracalny), może być większy od zbioru wartości własnych f. Ogólnie rzecz biorąc, nieizolowanym punktom widma nie odpowiadają żadne wektory własne. Jednakże właśnie zbiór punktów osobliwych operatora $\lambda \operatorname{id} - f$ stanowi właściwe uogólnienie pojęcia widma w przypadku nieskończonego wymiaru. Ten niedostatek wektorów własnych zmusza do zmiany wielu sformułowań, jednakże zasadniczy rezultat jest uogólnieniem wzoru (2), w którym sumowanie jest zastąpione całkowaniem.

Ograniczymy się tutaj do zilustrowania kilku ważnych koncepcji poprzez przykłady, gdzie takie trudności nie występują.

10. Formalnie sprzężone operatory różniczkowe. Rozważmy pewną przestrzeń funkcji rzeczywistych na odcinku $[a, b]$ z iloczynem skalarnym

$$(f, g) = \int_a^b f(x) g(x)\, dx,$$

o której założymy, że operator $\dfrac{d}{dx}$ przeprowadza ją w siebie. Ze wzoru na całkowanie przez części otrzymujemy

$$\left(\frac{df}{dx}, g\right) + \left(f, \frac{dg}{dx}\right) = fg\Big|_a^b = f(b)g(b) - f(a)g(a).$$

Jeśli więc nasza przestrzeń zawiera wyłącznie funkcje przyjmujące jednakowe wartości na końcach odcinka, to

$$\left(\frac{df}{dx}, g\right) = \left(f, -\frac{dg}{dx}\right),$$

tj. na tej przestrzeni operator $-\frac{d}{dx}$ jest sprzężony z operatorem $\frac{d}{dx}$.

Wykorzystując wielokrotnie wzór całkowania przez części lub też korzystając z formalnej identyczności operatorowej $(f_1 \circ \cdots \circ f_n)^* = f_n^* \circ \cdots \circ f_1^*$ otrzymamy, dla takich przestrzeni wzór

$$\left[\sum_{i=0}^{n} a_i(x)\frac{d^i}{dx^i}\right]^* = \sum_{i=0}^{n}(-1)^i \frac{d^i}{dx^i} \circ a_i(x), \tag{3}$$

gdzie zapis $\frac{d^i}{dx^i} \circ a_i(x)$ dla operatora oznacza, że stosując go do funkcji $f(x)$, najpierw mnożymy ją przez $a_i(x)$, a następnie różniczkujemy iloczyn i-krotnie względem x. Wzór (3) określa operację (*formalnego*) *sprzężenia* operatorów różniczkowych: $D \mapsto D^*$. Operator D nazywa się (*formalnie*) *samosprzężonym*, jeśli $D = D^*$. Określenie „formalnie" ma przypominać, że w definicji nie jest jawnie zadana przestrzeń, w której D działa jako operator liniowy.

Jeśli iloczyn skalarny jest wprowadzony za pomocą wagi $G(x)$:

$$(f, g)_G = \int_a^b G(x)f(x)g(x)\,dx,$$

to oczywiste rachunki prowadzą do wniosku, że zamiast D^* należy rozpatrywać operator $G^{-1} \circ D^* \circ G$ (przyjmując, że G nie zeruje się w rozważanym przedziale) i że on właśnie jest kandydatem do roli operatora sprzężonego z D względem $(f, g)_G$.

Pokażemy, że ortogonalne układy funkcji rozważane w § 4 składają się z funkcji własnych samosprzężonych operatorów różniczkowych.

a) *Rzeczywiste wielomiany Fouriera stopnia* $\leq N$. Formalnie samosprzężony operator $\frac{d^2}{dx^2}$ przeprowadza tę przestrzeń w siebie i jest w niej samosprzężony. Jego wartości własne są równe 0 (z krotnością 1) i $-1^2, -2^2, \ldots, -N^2$ (z krotnością 2). Odpowiadające im wektory własne to 1 i $\{\cos nx, \sin nx\}$, $1 \leq n \leq N$.

b) *Wielomiany Legendre'a*. Operator

$$(x^2-1)\frac{d^2}{dx^2} + 2x\frac{d}{dx} = \frac{d}{dx} \circ \left[(x^2-1)\frac{d}{dx}\right]$$

jest formalnie samosprzężony i przeprowadza przestrzeń wielomianów stopnia $\leq N$ w siebie. Po zastosowaniu wzoru Leibniza do obu stron oczywistej równości

$$(x^2-1)\frac{d}{dx}(x^2-1)^n = 2nx(x^2-1)^n,$$

otrzymamy

$$\frac{d^{n+1}}{dx^{n+1}}\left[(x^2-1)\frac{d}{dx}(x^2-1)^n\right] =$$
$$= (x^2-1)\frac{d^{n+2}}{dx^{n+2}}(x^2-1)^n + 2(n+1)x\frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n +$$
$$+ n(n+1)\frac{d^n}{dx^n}(x^2-1)^n = 2nx\frac{d^{n+1}}{dx^{n+1}}(x^2-1)^n + 2n(n+1)\frac{d^n}{dx^n}(x^2-1)^n.$$

Dzieląc obie strony równości przez $2^n n!$ i odwołując się do definicji wielomianów Legendre'a, otrzymujemy

$$\left[(x^2-1)\frac{d^2}{dx^2} + 2x\frac{d}{dx}\right]P_n(x) = n(n+1)P_n(x).$$

Wykazaliśmy więc, że *operator* $(x^2-1)\frac{d^2}{dx^2} + 2x\frac{d}{dx}$ *w przestrzeni wielomianów stopnia* $\leq N$ *jest diagonalny w bazie ortogonalnej złożonej z wielomianów Legendre'a i ma widmo rzeczywiste proste. W szczególności jest on samosprzężony.*

Oczywiście, można by sprawdzić samosprzężoność w tej przestrzeni bezpośrednim całkowaniem przez części: człon typu $fg\Big|_{-1}^{1}$ znika ze względu na obecność czynnika (x^2-1) we współczynnikach operatora. W ten sposób z rezultatów tego punktu i twierdzenia p. 4 otrzymujemy inny dowód ortogonalności wielomianów Legendre'a.

Zostawimy czytelnikowi sprawdzenie części sformułowanych własności i ich interpretację w terminach algebry liniowej dla wielomianów Hermite'a i Czebyszewa (nie należy zapomnieć o czynniku wagowym $G(x)$!).

c) *Wielomian Hermite'a* $H_n(x) = (-1)^n e^{-x^2}\frac{d^n}{dx^n}(e^{-x^2})$ *jest wektorem własnym odpowiadającym wartości własnej* $-2n$ *operatora*

$$K = \frac{d^2}{dx^2} - 2x\frac{d}{dx}.$$

Funkcja $e^{-x^2/2}H_n(x) = (-1)^n e^{x^2/2}\frac{d^n}{dx^n}(e^{-x^2})$ *jest funkcją własną operatora*

$$H = \frac{d^2}{dx^2} - x^2$$

odpowiadającą wartości własnej $-(2n+1)$.

Pierwsze stwierdzenie otrzymujemy za pomocą prostej indukcji względem n i ten fragment opuszczamy. Dla dowodu drugiego stwierdzenia rozważymy pomocniczy operator

$$M = \frac{d}{dx} - x.$$

Łatwo można sprawdzić, że

$$[H, M] = HM - MH = -2\left(\frac{d}{dx} - x\right) = -2M.$$

Stąd wynika, że jeśli f jest funkcją własną operatora H odpowiadającą wartości własnej λ, to Mf jest funkcją własną operatora H odpowiadającą wartości własnej $\lambda - 2$:

$$HMf = [H, M]f + MHf = -2Mf + \lambda Mf = (\lambda - 2)Mf.$$

Ponieważ $H(e^{-x^2/2}) = -e^{-x^2/2}$, więc funkcja $M^n(e^{-x^2/2})$ jest funkcją własną H odpowiadającą wartości własnej $-(2n+1)$ dla każdego $n \geq 0$. Z drugiej strony proste sprawdzenie pokazuje, że

$$e^{x^2/2}M(e^{-x^2/2}f(x)) = e^{x^2}\frac{d}{dx}\left(e^{-x^2}f(x)\right),$$

skąd wynika, że

$$e^{-x^2/2}H_n(x) = (-1)^n M^n(e^{-x^2/2}),$$

co kończy dowód drugiego ze sformułowanych wyżej punktów.

d) *Wielomian Czebyszewa*

$$T_n(x) = \frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\frac{d^n}{dx^n}(1-x^2)^{n-1/2}$$

jest wektorem własnym o wartości własnej $-n^2$ *operatora*

$$(1-x^2)\frac{d^2}{dx^2} - x\frac{d}{dx}.$$

11. Operatory normalne. W przestrzeni unitarnej zarówno operatory unitarne, jak i operatory samosprzężone są szczególnym przypadkiem *operatorów normalnych*, które można opisać jednym z dwóch równoważnych warunków:

a) *Są to operatory diagonalizowalne poprzez wybór bazy ortonormalnej.*

b) *Są to operatory komutujące ze swoim operatorem sprzężonym.*

Sprawdzimy równoważność obu warunków.

Jeśli $\{e_i\}$ jest bazą ortonormalną i $f(e_i) = \lambda_i e_i$, to $f^*(e_i) = \bar{\lambda}_i e_i$, skąd wynika $[f, f^*] = 0$, tj. a) pociąga za sobą b).

Dla dowodu implikacji odwrotnej wybierzemy wartość własną λ operatora f i przyjmiemy

$$L_\lambda = \{l \in L \mid f(l) = \lambda l\}.$$

Sprawdźmy, że $f^*(L_\lambda) \subset L_\lambda$. Istotnie, jeśli $l \in L_\lambda$, to

$$f(f^*(l)) = f^*(f(l)) = f^*(\lambda l) = \lambda f^*(l),$$

bo $f^*f = ff^*$. Stąd wynika, że $L_\lambda^\perp$ jest f-niezmiennicza: jeśli $(l, l_0) = 0$ dla każdego $l_0 \in L_\lambda$, to

$$(f(l), l_0) = (l, f^*(l_0)) = 0.$$

Dokładnie tak samo pokazać można, że $L_\lambda^\perp$ jest f^*-niezmiennicza. Obcięcia f i f^* do $L_\lambda^\perp$ są oczywiście także przemienne, zatem stosując indukcję względem wymiaru L, można założyć, że w $L_\lambda^\perp$ f jest diagonalne w odpowiedniej bazie ortonormalnej. Ponieważ jest to także prawdą dla L_λ, więc dowód został zakończony.

Ćwiczenia

1. Niech $f: L \to L$ będzie operatorem w przestrzeni liniowej. Dowieść, że jeśli $|(f(l), l)| \leq c|l|^2$ dla każdego $l \in L$ i pewnego $c > 0$, to

$$|(f(l), m)| + |(l, f(m))| \leq 2c|l| \cdot |m|$$

dla każdych l, $m \in L$.

2. Niech $f: L \to L$ będzie operatorem samosprzężonym. Wykazać, że

$$|(f(l), l)| \leq |f| \cdot |l|^2$$

dla każdego $l \in L$, gdzie $|f|$ jest indukowaną normą f, a jeśli $c < |f|$, to istnieje wektor $l \in L$, taki że

$$|(f(l), l)| > c|l|^2.$$

3. Operator samosprzężony nazywamy *nieujemnym*, $f \geq 0$, jeśli $(f(l), l) \geq 0$ dla każdego l. Dowieść, że ten warunek jest równoważny z nieujemnością wszystkich punktów widma f.

4. Wykazać, że relacja $f \geq g$, zadana jako $f - g \geq 0$, jest relacją częściowego porządku w zbiorze operatorów samosprzężonych.

5. Wykazać, że iloczyn dwóch przemiennych ze sobą operatorów samosprzężonych nieujemnych jest operatorem nieujemnym.

6. Wykazać, że z każdego operatora samosprzężonego nieujemnego można wyciągnąć jedyny nieujemny pierwiastek kwadratowy.

7. Wyrazić jawnym wzorem poprawkę drugiego rzędu do wektora własnego i wartości własnej operatora $H_0 + \varepsilon H_1$.

8. Niech f oznacza operator samosprzężony, $\omega \in C$, $\operatorname{Im}\omega \neq 0$. Dowieść, że operator

$$g = (f - \overline{\omega}\,\mathrm{id})(f - \omega\,\mathrm{id})^{-1}$$

jest unitarny, jego widmo nie zawiera jedności oraz

$$f = (\omega g - \overline{\omega})(g - \mathrm{id})^{-1}.$$

9. Odwrotnie, niech g będzie operatorem unitarnym, którego widmo nie zawiera jedności. Dowieść, że operator

$$f = (\omega g - \overline{\omega}\,\mathrm{id})(g - \mathrm{id})^{-1}$$

jest samosprzężony oraz

$$g = (f - \overline{\omega}\,\mathrm{id})(f - \omega\,\mathrm{id})^{-1}.$$

(Wprowadzone powyżej odwzorowania, wiążące operatory unitarne z samosprzężonymi, nazywają się *transformacją Cayleya*. W jednowymiarowym przypadku ich odpowiednikiem jest odwzorowanie $a \mapsto \dfrac{a-\overline{\omega}}{a-\omega}$ przeprowadzające oś rzeczywistą w okrąg jednostkowy.)

10. Niech $f: L \to L$ będzie dowolnym operatorem w przestrzeni unitarnej. Dowieść, że f^*f jest operatorem samosprzężonym nieujemnym i że jest on dodatni wtedy i tylko wtedy, gdy f jest odwracalny.

11. Niech f będzie odwracalny oraz $r_1^2 = ff^*$, $r_2^2 = f^*f$, gdzie przez r_1, r_2 oznaczyliśmy operatory samosprzężone dodatnie. Dowieść, że

$$f = r_1 u_1 = u_2 r_2,$$

gdzie u_1, u_2 są unitarne. (Te równości wyrażają tzw. *rozkład biegunowy operatora liniowego* f, a u_1, u_2 są odpowiednio prawym i lewym *czynnikiem fazowym* f. W przypadku jednowymiarowym otrzymujemy przedstawienie niezerowej liczby zespolonej w postaci $re^{i\varphi}$.)

12. Dowieść, że każdy z rozkładów biegunowych $f = r_1u_1 = u_2r_2$ są jednoznaczne.

13. Wykazać, że również operatory nieodwracalne f mają rozkład biegunowy, ale w tym przypadku jednoznacznie określone są tylko r_1 i r_2, a nie u_1 i u_2.

§ 9. Operatory samosprzężone w mechanice kwantowej

1. Kontynuujemy tu omawianie podstawowych postulatów mechaniki kwantowej rozpoczęte w § 6, p. 8.

Niech $\mathcal{H}$ oznacza przestrzeń unitarną stanów pewnego układu kwantowego. W fizyce dla scharakteryzowania poszczególnych stanów posługujemy się możliwością określenia (zmierzenia) dla danego stanu układu wartości pewnych *wielkości fizycznych* takich jak energia, spin, współrzędna, pęd, itp. Po wybraniu jednostek i punktu odniesienia („zera") dla każdej z tych wielkości, wynikami pomiarów są *liczby rzeczywiste* (jest to w istocie definicja wielkości skalarnych), co odtąd będziemy zakładać.

Trzeci — po zasadzie superpozycji i interpretacji iloczynów skalarnych jako amplitud prawdopodobieństwa — postulat mechaniki kwantowej formułuje się w następujący sposób.

Każdej fizycznej wielkości skalarnej, której wartości można mierzyć w stanach układu o przestrzeni stanów $\mathcal{H}$, można przypisać operator samosprzężony $f: \mathcal{H} \to \mathcal{H}$ o następujących własnościach:

a) *Widmem operatora f jest zbiór wszystkich wartości możliwych do otrzymania w wyniku pomiarów odpowiadającej mu wielkości fizycznej przeprowadzonych w różnych stanach tego układu.*

b) *Jeśli $\psi \in \mathcal{H}$ jest wektorem własnym operatora f odpowiadającym wartości własnej λ, to przy pomiarze tej wielkości fizycznej w stanie ψ układu otrzymamy wartość λ z prawdopodobieństwem jeden.*

c) *Ogólniej, dokonując pomiaru wielkości f w stanie układu opisanym przez wektor ψ, $|\psi| = 1$, można otrzymać wartość λ należącą do widma operatora f z prawdopodobieństwem równym kwadratowi normy rzutu ortogonalnego wektora ψ na podprzestrzeń własną $\mathcal{H}(\lambda)$ odpowiadającą λ.*

Ponieważ na mocy twierdzenia § 8, p. 4 przestrzeń $\mathcal{H}$ rozkłada się na sumę prostą $\bigoplus_{i=1}^{m} \mathcal{H}(\lambda_i)$, $\lambda_i \neq \lambda_j$ dla $i \neq j$, więc wektor ψ wyraża się jako suma rzutów $\psi_i \in \mathcal{H}(\lambda_i)$, $i = 1, \ldots, m$, na podprzestrzenie własne. Twierdzenie Pitagorasa

$$1 = |\psi|^2 = \sum_{i=1}^{m} |\psi_i|^2$$

można zatem zinterpretować jako wypowiedź o tym, że przeprowadzając pomiary wielkości f w dowolnym stanie układu, otrzymamy z prawdopodobieństwem 1 którąś z możliwych wartości f.

Wielkości fizyczne, o których była mowa wyżej, a także odpowiadające im operatory samosprzężone nazywane są *wielkościami obserwowalnymi* lub *obserwablami*. Postulat o obserwablach nieraz interpretuje się szerzej, uważając, że każdemu operatorowi samosprzężonemu odpowiada pewna obserwowalna wielkość fizyczna, obserwabla.

W przypadku nieskończenie wymiarowych przestrzeni $\mathcal{H}$ podane wyżej postulaty wymagają pewnych zmian. Na przykład, zamiast sformułowań podanych w warunkach b) i c) trzeba rozważać prawdopodobieństwo tego, że przy pomiarze f w stanie ψ otrzymamy wartości z pewnego przedziału $(a, b) \subset \boldsymbol{R}$. Także i temu przedziałowi można przyporządkować podprzestrzeń $\mathcal{H}_{(a,b)} \subset \mathcal{H}$ — w skończenie wymiarowym przypadku będącą obrazem rzutu ortogonalnego $p_{(a,b)}$ na $\bigoplus_{\lambda_i \in (a,b)} \mathcal{H}(\lambda_i)$, — a szukane prawdopodobieństwo wynosi

$$|p_{(a,b)}\psi|^2 = (\psi, p_{(a,b)}\psi).$$

Ponadto może okazać się, że w nieskończenie wymiarowym przypadku operatory wielkości obserwowalnych są określone jedynie na pewnej podprzestrzeni $\mathcal{H}_0 \subset \mathcal{H}$.

Związek tej terminologii z pojęciami wprowadzonymi w § 6, p. 8 jest następujący. *Filtr* B_χ jest to przyrząd do pomiaru wielkości obserwowalnej odpowiadającej ortogonalnemu projektorowi na podprzestrzeń rozpiętą na wektorze χ. Przypisuje

się jej wartość 1, jeśli układ przeszedł przez filtr, a 0 w przeciwnym przypadku. *Piec* A_ψ — to połączenie przyrządu przygotowującego układy, w ogólności w rozmaitych stanach, z filtrem B_ψ, który przepuszcza układy tylko wtedy, gdy znajdują się one w stanie ψ. Sposób obliczenia prawdopodobieństw podany w § 6, p. 8 jest w oczywisty sposób tożsamy ze sposobem podanym w warunkach b) i c) powyżej.

Z tego przykładu wnioskujemy, że przyrząd, powiedzmy B_χ, do pomiaru wielkości obserwowalnej w stanie ψ, na ogół *zmienia* stan układu: z prawdopodobieństwem $|(\psi,\chi)|^2$ stan ψ przechodzi w stan χ, a z prawdopodobieństwem $1-|(\psi,\chi)|^2$ układ zostaje „zniszczony". Dlatego użycie nazwy „pomiar" w odniesieniu do takiego procesu oddziaływania układu z przyrządem może doprowadzić do zasadniczo nieprawidłowych wyobrażeń. Fizyka klasyczna jest oparta na założeniu, że sam pomiar można w zasadzie tak przeprowadzić, aby dowolnie mało zaburzyć stan układu poddawanego pomiarowi. Niemniej jednak nazwa „pomiar" jest ogólnie przyjęta w tekstach fizycznych i uznaliśmy za właściwe posługiwanie się nią także tutaj, zaczynając jednakże od nie tak powszechnie stosowanych, lecz intuicyjnie bardziej zrozumiałych „pieców" i „filtrów".

2. Wartości średnie i zasada nieoznaczoności. Niech f oznacza pewną obserwablę o widmie $\{\lambda_i\}$ i odpowiadającym jej rozkładzie ortogonalnym $\mathcal{H} = \bigoplus_i \mathcal{H}(\lambda_i)$. Jak już powiedziano, f przyjmuje w stanie ψ, $|\psi| = 1$, wartość λ_i z prawdopodobieństwem $(\psi, p_i\psi)$, gdzie przez p_i został oznaczony projektor ortogonalny na $\mathcal{H}(\lambda_i)$. Dlatego *wartość średnią* $\widehat{f}_\psi$ wielkości f w stanie ψ, obliczoną dla wielu pomiarów, można wyrazić następująco:

$$\widehat{f}_\psi = \sum_i \lambda_i(\psi, p_i\psi) = \sum_i(\psi, \lambda_i p_i\psi) = (\psi, f(\psi))$$

(przypominamy, że $|\psi| = 1$).

Wartość oznaczoną wyżej przez $(\psi, f(\chi))$ w oznaczeniach Diraca zapisuje się $\langle\chi|f|\psi\rangle$. Część tego symbolu — $f|\psi\rangle$ oznacza wynik działania operatora f na ketwektor $|\psi\rangle$, a $\langle\chi|f$ — wynik działania operatora f na bra-wektor $\langle\chi|$.

Wróćmy do wartości średnich. Jeśli operatory f i g są samosprzężone, to na ogół operator fg nie jest samosprzężony:

$$(fg)^* = g^*f^* = gf \neq fg,$$

jeśli f nie komutuje z g. Jednakże f^2, $f-\lambda$ ($\lambda \in \boldsymbol{R}$) i komutator $\frac{1}{i}[f,g] = \frac{1}{i}(fg-gf)$ są już samosprzężone.

Średnia wartość $[(f-\widehat{f}_\psi)^2]\widehat{}_\psi$ wielkości mierzalnej $(f-\widehat{f}_\psi)^2$ w stanie ψ jest *średnim kwadratowym odchyleniem wartości f od jej wartości średniej*, inaczej dyspersją (rozrzutem) wartości f. Oznaczymy

$$\widehat{\Delta f}_\psi = \sqrt{[(f-\widehat{f}_\psi)^2]\widehat{}_\psi}.$$

3. Stwierdzenie (zasada nieoznaczoności Heisenberga). *Dla dowolnych operatorów samosprzężonych f, g w przestrzeni unitarnej*

$$\widehat{\Delta f}_\psi \cdot \widehat{\Delta g}_\psi \geq \frac{1}{2}\left|([f,g]\psi, \psi)\right|.$$

Dowód. Wyrażenie przedstawiające średnią wartość komutatora

$$[f - \widehat{f}_\psi, g - \widehat{g}_\psi] = [f, g]$$

w stanie ψ przepiszemy, używając oznaczeń $f_1 = f - \widehat{f}_\psi$, $g_1 = g - \widehat{g}_\psi$, a następnie, po uwzględnieniu samosprzężoności operatorów f, g, zastosujemy nierówność Cauchy'ego–Buniakowskiego–Schwarza

$$|([f, g]\psi, \psi)| = |((f_1 g_1 - g_1 f_1)\psi, \psi)| = |(g_1\psi, f_1\psi) - (f_1\psi, g_1\psi)| =$$

$$= 2|\mathrm{Im}(g_1\psi, f_1\psi)| \leq 2|(g_1\psi, f_1\psi)| \leq$$

$$\leq 2\sqrt{(f_1\psi, f_1\psi)}\sqrt{(g_1\psi, g_1\psi)} = 2\widehat{\Delta f}_\psi \widehat{\Delta g}_\psi .$$

Ta nierówność pokazuje, że średnie rozrzuty wartości nieprzemiennych wielkości mierzalnych f, g, nie mogą być jednocześnie dowolnie zmniejszane. Mówi się jeszcze, że nieprzemienne wielkości nie są jednocześnie mierzalne; do tego sformułowania należy jednak odnosić się z taką samą ostrożnością, jak do terminu „pomiar".

Szczególną rolę odgrywają zastosowania nierówności Heisenberga dla takich par zmiennych, które spełniają relację $\frac{1}{i}[f, g] = \mathrm{id}$ — są one nazywane *parami obserwabli kanonicznie sprzężonych.* Spełniają one nierówność

$$2\widehat{\Delta f}_\psi \widehat{\Delta g}_\psi \geq 1/2$$

dla dowolnie wybranego stanu ψ. Zauważmy, że w przestrzeniach skończenie wymiarowych takich par nie ma, bo $\mathrm{Tr}\,[f, g] = 0$, $\mathrm{Tr}\,\mathrm{id} = \dim \mathcal{H}$. Jednakże w przestrzeniach nieskończenie wymiarowych takie pary obserwabli istnieją, a klasycznym przykładem jest relacja komutacji

$$\frac{1}{i}\left[x, \frac{1}{i}\frac{d}{dx}\right] = \mathrm{id}.$$

Występujące w niej operatory pojawiają się w modelach kwantowych tych układów fizycznych, które w języku fizyki klasycznej noszą nazwę „cząstki poruszającej się w jednowymiarowym polu potencjalnym".

Te i pewne inne obserwable opiszemy teraz bardziej szczegółowo.

4. a) *Operator położenia.* Tak nazywa się operator mnożenia przez x w przestrzeni funkcji zespolonych określonych na $\boldsymbol{R}$ (lub jakimś podzbiorze $\boldsymbol{R}$) z iloczynem skalarnym $\int f(x)\overline{g}(x)\,dx$. Przyjmuje się, że rozważany jest układ odpowiadający „cząstce poruszającej się po linii prostej w polu zewnętrznym".

b) *Operator pędu.* Tak nazywa się operator $\frac{1}{i}\frac{d}{dx}$ działający w tych samych przestrzeniach funkcji. (Zazwyczaj uzupełnia się go o czynnik $\hbar$ nazywany stałą Plancka — to zakłada właściwy wybór jednostek, o czym teraz nie będziemy wspominać.)

c) *Operator energii oscylatora kwantowego.* Jest on dany wyrażeniem (tak jak poprzednio przy założeniu odpowiedniego wyboru jednostek) $\frac{1}{2}\left[-\frac{d^2}{dx^2}+x^2\right]$.

d) *Operator rzutu spinu* dla układu „cząstka o spinie 1/2". Jest nim dowolny samosprzężony operator o wartościach własnych $\pm 1/2$ w dwuwymiarowej przestrzeni unitarnej. Później opiszemy ten operator bardziej szczegółowo.

W przykładach a) — c) celowo nie podaliśmy precyzyjnego opisu przestrzeni unitarnych, w których działają nasze operatory. W istocie są one nieskończenie wymiarowe i należy je badać metodami analizy funkcjonalnej. O przykładzie d) powiemy coś więcej niżej.

5. Operator energii i ewolucja układu w czasie. Zasadnicze znaczenie dla opisania dowolnego układu kwantowego ma, oprócz wprowadzenia jego przestrzeni stanów $\mathcal{H}$, także wyróżnienie fundamentalnej obserwabli $H\colon \mathcal{H} \to \mathcal{H}$, nazywanej *operatorem energii, operatorem Hamiltona* lub *hamiltonianem.*

Właśnie ten operator używany jest do sformułowania ostatniego z podstawowych postulatów mechaniki kwantowej.

Jeśli w chwili 0 *układ znajdował się w stanie* ψ *i jego rozwój w czasie do chwili* t *odbywał się w izolacji, a w szczególności nie były nad nim przeprowadzane pomiary, to w chwili* t *układ znajdzie się w stanie* $\exp(-itH)\psi$, *gdzie*

$$\exp(-itH) = \sum_{n=0}^{\infty} \frac{(-itH)^n}{n!} \colon \mathcal{H} \longrightarrow \mathcal{H}$$

(por. § 11, część 1). $U(t) = \exp(-itH)$ jest operatorem unitarnym. Jednoparametrowa grupa operatorów unitarnych $\{U(t) \mid t \in \boldsymbol{R}\}$ całkowicie określa ewolucję czasową układu izolowanego.

Wielkość fizyczna o wymiarze (energia)×(czas) nazywana jest działaniem. Wiele eksperymentów pozwala wyznaczyć uniwersalną jednostkę działania — sławną stałą Plancka $\hbar = 1{,}055 \cdot 10^{-34}$ J·s. Pisząc nasz wzór, przyjęliśmy, że Ht jest wyrażone w jednostkach $\hbar$, choć częściej zapisuje się go w postaci $\exp\left(-\frac{Ht}{i\hbar}\right)\psi$. Jednakże my będziemy opuszczać $\hbar$ dla skrócenia zapisu.

Zauważymy jeszcze, że na mocy liniowości operator e^{-itH} przeprowadza promienie w $\mathcal{H}$ na promienie, więc w istocie zadaje działanie na stanach układu, a nie tylko na wektorach ψ.

Wzór opisujący ewolucję układu można napisać w postaci różniczkowej

$$\frac{d}{dt}(e^{-itH}) = -iH(e^{-itH})$$

lub, określając $\psi(t) = e^{-itH}\psi$,

$$\frac{d\psi}{dt} = -iH\psi \quad \left(\frac{1}{i\hbar}H\psi, \text{ jeśli uwzględnić jednostki}\right).$$

To ostatnie równanie nazywa się *równaniem Schrödingera*. Po raz pierwszy zostało napisane w przypadku, gdy wektory stanu ψ były opisywane przez funkcje na przestrzeni fizycznej, a H był wyrażony jako operator różniczkowy w tych współrzędnych.

Jak zazwyczaj, poniższe komentarze będą odnosić się zasadniczo do przypadku skończenie wymiarowej przestrzeni stanów $\mathcal{H}$.

6. Widmo energetyczne i stany stacjonarne układu. *Widmem energetycznym* układu nazywamy widmo związanego z nim hamiltonianu H. *Stanami stacjonarnymi* układu nazywamy te stany, które nie zmieniają się w trakcie ewolucji czasowej układu. Odpowiadające im promienie są niezmiennicze względem operatora e^{itH}, tj. są jednowymiarowymi podprzestrzeniami własnymi tego operatora lub, co jest tym samym, podprzestrzeniami własnymi operatora H. Wartości własnej E_j hamiltonianu, zwanej także *poziomem energetycznym* układu, odpowiada wartość własna $e_{itE_j} = \cos tE_j + i \sin tE_j$ operatora ewolucji, zmieniająca się w czasie.

Jeśli H ma widmo proste, to przestrzeń $\mathcal{H}$ ma kanoniczną bazę ortonormalną składającą się z wektorów odpowiadających stanom stacjonarnym (są one wyznaczone z dokładnością do czynnika fazowego $e^{i\varphi}$). Jeśli krotność poziomu energetycznego E jest większa od jedności, to ten poziom i odpowiadające mu stany nazywają się *zdegenerowanymi*, a krotność E — stopniem degeneracji.

Wszystkie stany odpowiadające najniższemu poziomowi energetycznemu, tj. najmniejszej wartości własnej H, nazywają się *stanami podstawowymi* układu — stan podstawowy jest jedyny, gdy poziom najniższy jest niezdegenerowany. Ta nazwa jest odbiciem przekonania o tym, że układ kwantowy nigdy nie jest izolowany od świata zewnętrznego — z określonym prawdopodobieństwem może on wypromieniować lub pochłonąć pewną dawkę energii. W pewnych warunkach prawdopodobieństwo utraty energii jest znacznie większe niż prawdopodobieństwo jej pochłonięcia, i wtedy układ będzie wykazywał tendencję do „spadania" do swego najniższego stanu i pozostawania w nim w dalszym ciągu. Z tego też powodu stany wyższe od najniższego nazywają się *stanami wzbudzonymi*.

Wróćmy do omawianego w przykładzie d) p. 4. operatora Hamiltona oscylatora kwantowego: $\frac{1}{2}\left[-\frac{d^2}{dx^2} + x^2\right]$. W § 8, p. 10 c) pokazaliśmy, że funkcje $e^{-x^2/2}H_n(x)$ stanowią układ stanów stacjonarnych oscylatora o poziomach energii $E_n = n + \frac{1}{2}$, $n = 0, 1, 2, \ldots$ (Dokładniejsza analiza pokazuje, że energia powinna być tu mierzona w jednostkach $\hbar\omega$, gdzie stała ω jest częstością drgań oscylatora klasycznego.)

Określiwszy we właściwy sposób przestrzeń unitarną, do której odnoszone są rozważania, należy następnie wykazać, że te funkcje stanowią zupełny układ stanów stacjonarnych. Przy $n > 0$ oscylator, przechodząc ze stanu ψ_n w stan ψ_m, może wypromieniować porcję energii $E_n - E_m = (n-m)\hbar\omega$. W zastosowaniu do kwantowej teorii pola elektromagnetycznego proces ten opisuje się, mówiąc o „*wypromieniowaniu $n-m$ fotonów o częstotliwości ω*". Proces odwrotny jest pochłonięciem $n-m$ fotonów, przy czym oscylator przechodzi do wyższego (wzbudzonego) stanu energetycznego. Ten fakt, że energia może być pochłonięta lub stracona tylko w ilościach będących całkowitoliczbowymi wielokrotnościami $\hbar\omega$, ma przy tym zupełnie podstawowe znaczenie.

W stanie podstawowym oscylator ma niezerową energię $\frac{1}{2}\hbar\omega$, której jednakże nie może przekazać, bo nie ma niższych stanów energetycznych. W modelach kwantowych pole elektromagnetyczne traktuje się jako superpozycję nieskończenie wielu oscylatorów (w szczególności odpowiadających różnym częstotliwościom). W stanie podstawowym — próżni — pole ma zatem nieskończoną energię, choć z klasycznego punktu widzenie jest ono zerowe, gdyż skoro jest niemożliwe zabranie od niego energii, pole nie może na nic oddziaływać. Tak przedstawia się w najprostszym ujęciu jedna z najgłębszych trudności współczesnej kwantowej teorii pola. Ani aparat matematyczny, ani interpretacja fizyczna kwantowej teorii pola nie osiągnęły jeszcze etapu kompletności. Jest to otwarta i pociągająca teoria.

7. Formalizm teorii zaburzeń. W mechanice kwantowej znaczące miejsce zajmują te modele, dla których hamiltonian H można traktować jako sumę $H_0 + \varepsilon H_1$, gdzie przez H_0 oznaczyliśmy „niezaburzony" hamiltonian, a εH oznacza małą poprawkę — „zaburzenie". Z fizycznego punktu widzenia zaburzenie często opisuje oddziaływanie układu ze „światem zewnętrznym" (np. zewnętrznym polem magnetycznym) lub oddziaływanie części układu między sobą (a wtedy H_0 odpowiada wyidealizowanemu przypadkowi układu, którego składowe nie oddziałują ze sobą). Natomiast z matematycznego punktu widzenia takie założenie jest dopuszczalne wtedy, gdy analiza widma niezaburzonego hamiltonianu jest prostsza niż analiza H, a charakterystyki widmowe H można wyrazić w wygodny sposób poprzez szeregi względem parametru ε, których pierwsze wyrazy są określone przez H_0. Tutaj ograniczymy się do najczęściej używanych formuł i pewnych jakościowych spostrzeżeń.

a) *Poprawki pierwszego rzędu.* Niech $H_0 e_0 = \lambda_0 e_0$, $|e_0| = 1$. Spróbujemy znaleźć wektor własny i wartość własną dla $H_0 + \varepsilon H_1$, bliskie do e_0 i λ_0, odpowiednio, z dokładnością do wyrazów drugiego rzędu wielkości względem ε, tj. rozwiązać równanie

$$(H_0 + \varepsilon H_1)(e_0 + \varepsilon e_1) = (\lambda_0 + \varepsilon\lambda_1)(e_0 + \varepsilon e_1) + o(\varepsilon^2).$$

Porównując współczynniki przy ε, otrzymamy

$$(H_0 - \lambda_0)e_1 = (\lambda_1 - H_1)e_0.$$

Tutaj niewiadomymi są liczba λ_1 i wektor e_1. Można je wyznaczyć, używając następującego sposobu. Po pomnożeniu skalarnie obu stron ostatniego równania przez e_0,

na mocy samosprzężoności operatora $H_0 - \lambda_0$, otrzymamy po lewej stronie tego równania zero:

$$\big((H_0 - \lambda_0)e_1,\, e_0\big) = \big(e_1,\, (H_0 - \lambda_0)e_0\big) = 0.$$

Zatem $\big((\lambda_1 - H_1)e_0,\, e_0\big) = 0$, skąd na mocy unormowania e_0 wynika

$$\lambda_1 = (H_1 e_0,\, e_0).$$

Otrzymaliśmy poprawkę pierwszego rzędu do wartości własnej λ_0: „przesunięcie poziomu energetycznego" $\varepsilon\lambda_1$ wynosi $(\varepsilon H_1 e_0,\, e_0)$, tj. zgodnie z rezultatami p. 12 *jest równe średniej wartości „energii zaburzającej" εH_1 w stanie e_0.*

W celu wyznaczenia e_1 należałoby teraz odwrócić operator $H_0 - \lambda_0$. Jednakże jest on oczywiście nieodwracalny, bo λ_0 jest wartością własną H_0. Ponieważ jednak *prawa strona równania, tj.* $(\lambda_1 - H_1)e_0$, *jest ortogonalna do* e_0, więc wystarczy, aby $(H_0 - \lambda_0)$ był odwracalny na ortogonalnym dopełnieniu do e_0, które oznaczymy $e_0^{\perp}$. A to jest oczywiście równoważne (w skończenie wymiarowym przypadku) z tym, że *wartość własna λ_0 operatora H_0 ma krotność jeden*, tj. by *poziom energetyczny λ_0 jest niezdegenerowany.*

Jeśli ten warunek jest spełniony, to

$$e_1 = \Big((H_0 - \lambda_0)\Big|_{e_0^{\perp}}\Big)^{-1}(\lambda_1 - H_1)e_0,$$

co przedstawia poprawkę pierwszego rzędu do wektora własnego.

Wybierzmy bazę ortonormalną $\{e_0 = e^{(0)}, e^{(1)}, \ldots, e^{(n)}\}$, w której H_0 ma postać diagonalną z wartościami własnymi $\lambda_0 = \lambda^{(0)}, \lambda^{(1)}, \ldots, \lambda^{(n)}$. W bazie $\{e^{(1)}, \ldots, e^{(n)}\}$ przestrzeni $e_0^{\perp}$ spełniona jest równość

$$(\lambda_1 - H_1)e_0 = \sum_{i=1}^{n}\big((\lambda_1 - H_1)e_0,\, e^{(i)}\big)\,e^{(i)} = -\sum_{i=1}^{n}\big(H_1 e_0,\, e^{(i)}\big)\,e^{(i)},$$

a zatem po uwzględnieniu poprzedniej równości otrzymujemy

$$e_1 = \sum_{i=1}^{n}\frac{(H_1 e_0,\, e^{(i)})}{\lambda_0 - \lambda^{(i)}}e^{(i)}.$$

Jest intuicyjnie jasne, że ta poprawka pierwszego rzędu będzie dobrym przybliżeniem, jeśli energia zaburzenia jest mała w porównaniu z odległością od poziomu λ_0 do najbliższego do niego — ε powinno zostać skompensowane przez licznik $\lambda_0 - \lambda^{(i)}$ i takie założenie zazwyczaj robią fizycy.

b) *Poprawki wyższych rzędów.* Analogicznie do sytuacji rozważanej powyżej pokażemy, że jeśli wartość własna λ_0 jest niezdegenerowana, to można indukcyjnie wyznaczyć poprawkę $(i+1)$-go rzędu do (e_0, λ_0), opierając się na już wyznaczonych poprawkach rzędów $\leq i$. Niech zatem $i \geq 1$. Rozwiążemy równanie

$$(H_0 + \varepsilon H_1)\Big(\sum_{k=1}^{i+1} \varepsilon^k e_k\Big) = \Big(\sum_{l=1}^{i+1} \varepsilon^l \lambda_l\Big)\Big(\sum_{j=1}^{i+1} \varepsilon^j e_j\Big) + o(\varepsilon^{i+2})$$

względem e_{i+1}, λ_{i+1}. Porównując współczynniki przy ε^{i+1}, otrzymamy

$$(H_0 - \lambda_0)e_{i+1} = (\lambda_1 - H_1)e_i + \sum_{l=2}^{i} \lambda_l e_{i+1-l} + \lambda_{i+1} e_0.$$

Jak poprzednio, lewa strona jest ortogonalna do prawej, więc

$$\lambda_{i+1} = ((H_1 - \lambda_1)e_i,\, e_0) - \sum_{l=2}^{i} \lambda_l (e_{i+1-l},\, e_0),$$

$$e_{i+1} = \Big((H_0 - \lambda_0)\Big|_{e_0^\perp}\Big)^{-1} \Big[(\lambda_1 - H_1)e_i + \sum_{l=2}^{i+1} \lambda_l e_{i+1-l}\Big].$$

Wykazaliśmy więc, że poprawki dowolnego rzędu istnieją i są jednoznacznie wyznaczone.

c) *Szeregi teorii zaburzeń.* Formalne szeregi potęgowe względem ε

$$\sum_{i=0}^{\infty} \lambda_i \varepsilon^i, \quad \sum_{i=0}^{\infty} e_i \varepsilon^i,$$

w których λ_i i e_i są wyznaczone z podanych wyżej wzorów rekurencyjnych, nazywa się szeregami teorii zaburzeń. Można wykazać, że w skończenie wymiarowej przestrzeni te szeregi są zbieżne dla dostatecznie małych ε. Natomiast w nieskończenie wymiarowej przestrzeni mogą okazać się rozbieżne, ale suma pierwszych kilku wyrazów szeregu często daje przybliżenie dobrze zgodne z eksperymentem. Fizyczne znaczenie szeregów rachunku zaburzeń w kwantowej teorii pola jest duże, a badanie ich matematyczngo znaczenie prowadzi do wielu interesujących i ważkich problemów.

d) *Wielokrotne wartości własne i geometria.* Wyłączenie z poprzedzających rozważań przypadku wielokrotnej wartości własnej λ_0 wyniknęło z potrzeby odwrócenia operatora $H_0 - \lambda_0$ na przestrzeni $e_0^\perp$ i wyrażało się formalnie pojawieniem się w mianowniku różnicy $\lambda_0 - \lambda^{(i)}$. Można by otrzymać wzory także i w ogólnym przypadku, modyfikując w odpowiedni sposób rozważania, jednakże ograniczymy się tutaj do dyskusji geometrycznych efektów wielokrotności wartości własnych. Dostrzec je można już na typowym przykładzie $H_0 = \mathrm{id}$, gdzie wszystkie wartości własne są równe jedności. Mała zmiana H_0 prowadzi do następujących efektów.

O ile ta zmiana była dostatecznie ogólna, to wartości własne stają się parami różne — ten efekt jest nazywany w fizyce „rozszczepieniem poziomów" lub „usunięciem degeneracji". Na przykład, jedna linia spektralna może się rozszczepić na dwie lub więcej bądź to wskutek zwiększenia zdolności rozdzielczej przyrządu, bądź przez

umieszczenie układu w zewnętrznym polu. Matematyczny model w obu przypadkach jest ten sam i polega na dodaniu małej, poprzednio nie uwzględnianej, poprawki do H_0 (a być może także na modyfikacji przestrzeni stanów).

Teraz zastanówmy się, co może się zdarzyć z wektorami własnymi. W przypadku całkowitej degeneracji, H_0 jest diagonalne w dowolnej bazie ortonormalnej. Mała zmiana H_0 usuwająca degenerację oznacza wybór bazy ortonormalnej, której osie określają kierunki rozciągnięć, oraz współczynników tych rozciągnięć. Współczynniki powinny mało różnić się od wyjściowej wartości λ_0, ale same osie mogą być usytuowane w dowolny sposób. W ten sposób w pobliżu zdegenerowanego stanu własnego kierunki własne silnie zależą od zaburzenia. Dwa dowolnie małe zaburzenia operatora jednostkowego, oba z prostym widmem, mogą przyjmować postać diagonalną w dwóch ustalonych i dowolnie położonych względem siebie bazach ortonormalnych. To wskazuje na wewnętrzną przyczynę wystąpienia w mianownikach naszych wzorów różnicy $\lambda_0 - \lambda^{(i)}$.

§ 10. Geometria form kwadratowych i wartości własne operatorów samosprzężonych

1. Ten paragraf jest poświęcony badaniu geometrii wykresów form kwadratowych q w rzeczywistej przestrzeni liniowej, tj. zbiorów postaci $x_{n+1} = q(x_1, \ldots, x_n)$ w $\boldsymbol{R}^{n+1}$ i omówieniu pewnych zastosowań. Jedną z głównych przyczyn, dla których te klasyczne rezultaty wciąż jeszcze wywołują zainteresowanie, jest to, że formy kwadratowe są kolejnym, po liniowym, przybliżeniem dowolnej dwukrotnie różniczkowalnej funkcji i przez to stanowią klucz do wyjaśnienia geometrii „zakrzywienia" dowolnej gładkiej powierzchni wielowymiarowej. W paragrafach 3–6 wykazaliśmy już ogólne twierdzenia o klasyfikacji form kwadratowych za pomocą ogólnych liniowych lub tylko ortogonalnych odwzorowań, dlatego tutaj zaczniemy od wyjaśnienia ich geometrycznych implikacji. Przyjmiemy, że rozważamy przestrzeń euklidesową ze standardową metryką $\sum_{i=1}^{n} x_i^2$. Pominięcie struktury euklidesowej oznaczałoby tylko użycie mniej subtelnej relacji równoważności między wykresami. Strategia naszej dyskusji polega na tym, aby najpierw rozpatrzyć przypadek małych wymiarów, gdzie kształty wykresów można uchwycić obrazowo, a następnie rozważyć nisko wymiarowe przekroje w różnych kierunkach wielowymiarowych wykresów. Czytelnikowi zalecamy szkicowanie rysunków ilustrujących tekst. Przyjmujemy, że oś x_{n+1} jest skierowana do góry, a przestrzeń $\boldsymbol{R}^n$ położona poziomo.

2. Przypadek jednowymiarowy. Wykres krzywej $x_2 = \lambda x_1^2$ w $\boldsymbol{R}^2$ ma trzy podstawowe formy: „czarę" — wypukłość do dołu — przy $\lambda > 0$, „kopułę" — wypukłość do góry — przy $\lambda < 0$ i prostej poziomej przy $\lambda = 0$. Przy użyciu klasyfikacji liniowej, która dopuszcza dowolną zmianę skali wzdłuż osi x_1, te trzy przypadki wyczerpują wszystkie możliwości i można przyjąć, że $\lambda = \pm 1$ lub 0. Przy

klasyfikacji ortogonalnej λ jest niezmiennikiem: $|\lambda|$ określa stromiznę ścian czary albo kopuły, przy czym jest ona tym większa, im większa jest wartość $|\lambda|$. Inna charakteryzacja $|\lambda|$ wynika z obserwacji, że $\frac{1}{2|\lambda|}$ jest promieniem krzywizny wykresu na dnie albo w wierzchołku $(0,0)$. Rzeczywiście, równanie okręgu o promieniu R stycznego do osi x_1 w początku układu ma postać $x_1^2+(x_2-R)^2 = R^2$, a w otoczeniu zera mamy $x_2 \approx \frac{x_1^2}{2R}$.

3. Przypadek dwuwymiarowy. Dla zorientowania się w możliwościach, wybierzmy bazę ortonormalną w $\boldsymbol{R}^2$, w której q wyraża się przez sumę kwadratów z pewnymi współczynnikami: $q(y_1,y_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2$. Proste rozpięte przez wektory tej bazy nazywają się *osiami głównymi* formy q i są w ogólności obrócone względem wyjściowego położenia osi układu. Liczby λ_1, λ_2 są wyznaczone jednoznacznie jako wartości własne operatora samosprzężonego A, dla którego $q(\vec{x}) = \vec{x}^t A\vec{x}$ (w starych współrzędnych). W przypadku $\lambda_1 \neq \lambda_2$ same osie też są wyznaczone jednoznacznie, lecz przy $\lambda_1 = \lambda_2$ można je wybrać dowolnie, byleby tylko były ortogonalne. Dla każdego ze współczynników mamy zasadniczo trzy możliwości ($\lambda_i > 0, \lambda_i < 0, \lambda_i = 0$), ale ze względu na symetrię można ograniczyć się tylko do rozważenia czterech przypadków, spośród których tylko dwa są niezdegenerowane.

a) $x_3 = \lambda_1 x_1^2+\lambda_2 x_2^2$, $\lambda_1, \lambda_2 > 0$. Wykres jest nazywany *paraboloidą eliptyczną* i ma kształt czary. Przydawka „eliptyczna" pochodzi stąd, że rzuty przekrojów poziomych $\lambda_1 y_1^2 + \lambda_2 y_2^2 = c$ przy $c > 0$ są elipsami o półosiach $\sqrt{c\lambda_i^{-1}}$, położonych wzdłuż głównych osi formy q (a przy $\lambda_1 = \lambda_2$ okręgami). (Te rzuty są poziomicami funkcji q.) Nazwa „paraboloida" pochodzi stąd, że przekroje wykresu płaszczyznami pionowymi $ay_1 + by_2 = 0$ są parabolami (przy $\lambda_1 = \lambda_2$ wykres jest paraboloidą obrotową).

W przypadku $\lambda_1, \lambda_2 < 0$ jest to ta sama czara, tylko odwrócona.

b) $x_3 = \lambda_1 y_1^2 - \lambda_2 y_2^2$, $\lambda_1, \lambda_2 > 0$. Wykres jest nazywany *paraboloidą hiperboliczną*. Poziomicami q są hiperbole, niepuste dla każdej wartości x_3, tak że wykres przechodzi zarówno powyżej, jak i poniżej płaszczyzny $x_3 = 0$. Przekroje płaszczyznami pionowymi są jak poprzednio parabolami. Poziomica $x_3 = 0$ jest „zwyrodniałą hiperbolą" redukującą się do swych asymptot — dwóch prostych $\sqrt{\lambda_1}y_1 \pm \sqrt{\lambda_2}y_2 = 0$. Te proste w $\boldsymbol{R}^2$ nazywają się „kierunkami asymptotycznymi" formy q. Jeśli traktować q jako (nieokreśloną) metrykę na $\boldsymbol{R}^2$, to proste asymptotyczne składają się ze wszystkich wektorów o długości zerowej. Proste asymptotyczne dzielą $\boldsymbol{R}^2$ na cztery sektory. Poziomice $q = x_3$ przy $x_3 > 0$ położone są wewnątrz pary przeciwległych sektorów; gdy $x_3 \to +0$ (z góry), poziomice przybliżają się do asymptot, aby „przylgnąć" do nich przy $x_3 = 0$, a następnie „przeniknąwszy" przez nie wyłonić się w drugiej parze przeciwległych sektorów. Przekroje wykresu pionowymi płaszczyznami przechodzącymi przez proste asymptotyczne są tymi samymi prostymi — „wyprostowanymi parabolami".

Przypadek $x_3 = -\lambda_1 y_1^2 + \lambda_2 y_2^2$, $\lambda_1, \lambda_2 > 0$, można otrzymać z poprzedniego przez zamianę znaku x_3.

c) $x_3 = \lambda y_1^2$, $\lambda > 0$. Ponieważ funkcja nie zależy od zmiennej y_2, przekroje wykresu płaszczyznami pionowymi $y_2 = \text{const}$ mają jedną i tę samą postać: cały wykres, który jest zakreślony przy przesuwaniu wzdłuż osi y_2 przez położoną w płaszczyźnie (y_1, x_3) parabolę daną równaniem $x_3 = \lambda y_1^2$ nazywa się *walcem parabolicznym*. Poziomice są parami prostych $y_1 = \pm\sqrt{x_3\lambda^{-1}}$; przy $x_3 = 0$ sklejają się w jedną prostą; cały wykres położony jest nad płaszczyzną $x_3 = 0$.

Przypadek $x_3 = \lambda y_1^2$, $\lambda < 0$ otrzymuje się przez odwrócenie.

d) $x_3 = 0$ — jest to płaszczyzna.

4. Przypadek ogólny. Teraz jesteśmy w stanie objaśnić geometrię wykresu $x_{n+1} = q(x_1, \ldots, x_n)$ dla dowolnych n.

Przejdziemy do układu osi głównych w $\boldsymbol{R}^n$, tj. do bazy ortonormalnej, w której $q(y_1, \ldots, y_n) = \sum_{i=1}^{m} \lambda_i y_i^2$, $\lambda_1, \ldots, \lambda_m \neq 0$. Jak widzieliśmy, osie te są wyznaczone jednoznacznie, jeśli $m = n$ i $\lambda_i \neq \lambda_j$ przy $i \neq j$ lub jeśli $m = n-1$ i $\lambda_i \neq \lambda_j$ przy $i \neq j$. Wartości formy nie zależą od współrzędnych $y_{m+1}, \ldots, y_n$, więc jej wykres powstaje przez przesuwanie wykresu $\sum_{i=1}^{m} \lambda_i y_i^2$ w $\boldsymbol{R}^{m+1}$ we wszystkich możliwych kierunkach w podprzestrzeni rozpiętej przez wektory $\{e_{m+1}, \ldots, e_n\}$. Innymi słowy wykres jest „walcowy" wzdłuż tej podprzestrzeni. Nietrudno sprawdzić, że ta podprzestrzeń jest jądrem formy biegunowej dla q i jest trywialna wtedy i tylko wtedy, gdy q jest niezdegenerowana.

Niech q będzie niezdegenerowana, tj. $m = n$. Możemy założyć, że $\lambda_1, \ldots, \lambda_r > 0$, $\lambda_{r+1}, \ldots, \lambda_{r+s} < 0$, tj. (r, s) jest sygnaturą formy q. Jeśli forma jest dodatnio określona, tj. $r = n$, $s = 0$, to wykres ma postać n-wymiarowej czary; każdy jego przekrój pionową płaszczyzną jest parabolą, a poziomice $q = c > 0$ są elipsoidami o półosiach $\sqrt{c\lambda_i^{-1}}$, skierowanych wzdłuż osi głównych. Równanie takiej elipsoidy ma postać

$$\sum_{i=1}^{n} \left(\frac{x_i}{\sqrt{c\lambda_i^{-1}}} \right)^2 = 1,$$

tj. można ją otrzymać z kuli jednostkowej rozciągnięciami wzdłuż prostopadłych kierunków. W szczególności jest ona zbiorem ograniczonym i zawiera się w prostokątnym równoległościanie $|x_i| \leq \sqrt{c\lambda_i^{-1}}$, $i = 1, \ldots, n$. Jak zobaczymy niżej, badanie zmian długości półosi różnych przekrojów elipsoidy (też będących elipsoidami) przynosi użyteczne informacje o wartościach własnych operatorów samosprzężonych.

Dla $r = 0$, $s = n$ otrzymujemy kopułę. W obydwu przypadkach wykres nosi nazwę (n-wymiarowej) paraboloidy eliptycznej.

Pośrednie przypadki $rs \neq 0$ prowadzą do wielowymiarowych paraboloid hiperbolicznych różnych sygnatur. Kluczem do ich geometrii jest znowu struktura *stożka kierunków asymptotycznych* C w $\boldsymbol{R}^n$, tj. poziomicy zerowej $q(y_1, \ldots, y_n) = 0$ formy.

Nazywa się ją stożkiem, gdyż jest zakreślona przez swoje *tworzące*: prosta, zawierająca jeden wektor z C jest w niej zawarta w całości. Dla uzyskania wyobrażenia

o podstawie tego stożka rozważmy jego przecięcie, powiedzmy, z podrozmaitością liniową $y_n = 1$:

$$-\lambda_n^{-1} \sum_{i=1}^{n-1} \lambda_i y_i^2 = 1.$$

Jak widać, podstawa jest poziomicą formy kwadratowej $n-1$ zmiennych. Z najprostszym przypadkiem mamy do czynienia wtedy, gdy ta forma jest dodatnio określona, bo wówczas ta poziomica jest elipsoidą, a w szczególności jest ograniczona i nasz stożek przypomina „szkolne" trójwymiarowe stożki. Ten przypadek odpowiada sygnaturze $(n-1,1)$ lub $(1,n-1)$; dla $n=4$ przestrzeń $(\boldsymbol{R}^4, q)$ jest sławną *przestrzenią Minkowskiego*, która będzie szczegółowo omawiana później. Przy innej sygnaturze struktura stożka C jest znacznie bardziej złożona, bowiem jego podstawa „rozciąga się do nieskończoności". Przekroje wykresu q pionowymi płaszczyznami przechodzącymi przez tworzące C pokrywają się z tymi tworzącymi. Dla przekrojów innymi płaszczyznami otrzymujemy albo „czary", albo „kopuły" — kierunki asymptotyczne rozdzielają te dwa przypadki. Dlatego stożek C dzieli przestrzeń $\boldsymbol{R}^n \setminus C$ na dwie części, w całości zakreślone przez linie proste, na których forma q jest odpowiednio ujemna lub dodatnia. Jeden z tych obszarów nazywa się *wnętrzem stożka* C,a drugi jego *zewnętrzem. Geometryczny sens sygnatury* (r,s) *można w przybliżeniu, lecz obrazowo oddać następującym zdaniem: wykres formy* q *w kierunkach* r *oddala się w górę, a w kierunkach* s *— w dół.*

Chociaż nasze rozważania odnosiły się do rzeczywistych form kwadratowych, to wyniki można zastosować także do zespolonych form hermitowskich. Rzeczywiście, realifikacją $\boldsymbol{C}^n$ jest $\boldsymbol{R}^{2n}$, a realifikacją formy hermitowskiej $\sum_{i=1}^{n} a_{ij} x_i \overline{x}_j$, $a_{ij} = \overline{a}_{ji}$, jest rzeczywista forma kwadratowa. Przy realifikacji wszystkie wymiary ulegają podwojeniu, a w szczególności sygnatura zespolona (r,s) przechodzi w sygnaturę rzeczywistą $(2r,2s)$.

Opiszemy teraz pokrótce i bez dowodów dwa zastosowania tej teorii do mechaniki i topologii.

5. Ruch drgający. Dla ilustracji rozważmy kulkę, która może ślizgać się pod wpływem siły ciężkości w rynnie o kształcie $x_2 = \lambda x_1^2$ położonej w płaszczyźnie $\boldsymbol{R}^2$. Punkt $(0,0)$ jest jednym z możliwych torów ruchu kulki — *położeniem równowagi*. W przypadku $\lambda > 0$, ta równowaga jest trwała (stabilna) — przy niewielkim początkowym odchyleniu kulki lub nadaniu niewielkiej prędkości dochodzi do drgania kulki wokół dna rynny. Dla $\lambda > 0$ to położenie jest nietrwałe — kulka spada po jednej z dwóch gałęzi paraboli. Przy $\lambda = 0$ jest ono obojętne względem odchylenia przestrzennego, ale nietrwałe względem zmiany prędkości: kulka może pozostawać w spoczynku w dowolnym punkcie płaszczyzny $x_2 = 0$ albo poruszać się ruchem jednostajnym w dowolnym kierunku.

Okazuje się, że matematyczny opis dużej klasy układów mechanicznych w pobliżu położenia równowagi jest jakościowo dobrze modelowany wielowymiarowym

uogólnieniem powyższego obrazka: ruchem kulki pod działaniem siły ciężkości po powierzchni $x_{n+1} = q(x_1, \ldots, x_n)$ w pobliżu początku układu współrzędnych. Jeśli q jest dodatnio określona, to każdy „nieznaczny" ruch będzie bliski superpozycji małych drgań wzdłuż osi głównych formy q. Wzdłuż przestrzeni zerowej formy jest możliwa ucieczka do nieskończoności ze stałą prędkością. Wzdłuż kierunków, gdzie q jest ujemna, jest możliwe spadanie w dół. Zarówno obecność przestrzeni zerowej, jak i ujemnej składowej sygnatury formy wskazuje na niestabilność położenia równowagi lub nieprzydatność przybliżenia „małych drgań". Jednakże jest ważne, że małe zmiany formy rynny, w której ślizga się kulka (lub, bardziej technicznie, potencjału naszego układu), nie naruszają tej stabilności.

Żeby to wyjaśnić bliżej, powrócimy do wspomnianego na początku paragrafu przybliżenia dowolnej (powiedzmy, trzykrotnie różniczkowalnej) funkcji rzeczywistej $f(x_1, \ldots, x_n)$. W pobliżu zera ma ona postać

$$f(x_1, \ldots, x_n) = f(0, \ldots, 0) + \sum_{i=1}^{n} a_i x_i + \sum_{j,i=1}^{n} b_{ij} x_i x_j + o\Big(\sum_{i=1}^{n} |x_i|^2\Big),$$

gdzie

$$a_i = \frac{\partial f}{\partial x_i}(0, \ldots, 0), \quad b_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(0, \ldots, 0).$$

Odejmując od f jej wartość w zerze i część liniową, widzimy, że reszta jest z dokładnością do wyrazów wyższego rzędu kwadratowa. To odejmowanie oznacza, że rozpatrujemy odchylenie wykresu f od hiperpłaszczyzny stycznej do wykresu w zerze. Oznaczając tę hiperpłaszczyznę przez $\boldsymbol{R}^n$, zauważamy, że *zachowanie f w pobliżu zera jest zadane przez formę kwadratową o macierzy* $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(0, \ldots, 0)\right)$, *przynajmniej wtedy, gdy ta forma nie jest zdegenerowana* — w przeciwnym przypadku należy wziąć pod uwagę także wyrazy wyższego rzędu. (Na przykład, wykres $x_2 = x_1^3$ na lewo od zera spada w dół, a na prawo podnosi się do góry, a wykresy funkcji kwadratowych tak się nie zachowują. Dwuwymiarowy wykres $x_3 = x_1^2 + x_2^3$ to tzw. „małpie siodło", opadający w dół w tym krzywoliniowym sektorze, gdzie $x_1^2 + x_2^3 < 0$ — „miejsce na ogon".)

Punkt gdzie zeruje się różniczka $df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$ (tj. gdzie $\frac{\partial f}{\partial x_i} = 0$ dla wszystkich $i = 1, \ldots, n$), jest nazywany *punktem krytycznym* funkcji f (w naszych przykładach był to początek układu współrzędnych). Jest on *niezdegenerowany*, jeśli w tym punkcie forma kwadratowa $\sum_{j,i=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} \Delta x_i \Delta x_j$ jest niezdegenerowana. Poprzedzające rozważania dają się podsumować jednym zdaniem: *w pobliżu niezdegenerowanego punktu krytycznego wykres funkcji jest tak samo położony względem hiperpłaszczyzny stycznej, jak wykres jej części kwadratowej.* Następnie można wykazać, że niewielka modyfikacja funkcji (wraz z jej pierwszymi i drugimi pochodnymi) może tylko nieznacznie przesunąć położenie niezdegenerowanego punktu krytycz-

nego, ale nie zmienia sygnatury odpowiadającej mu formy kwadratowej i dlatego także ogólnego zachowania wykresu (lokalnego).

Można także dowieść, że w otoczeniu niezdegenerowanego punktu krytycznego można przeprowadzić taką gładką i mającą gładką odwrotną (choć na ogół nieliniową) zamianę współrzędnych $y_i = y_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$, że w nowych współrzędnych funkcja wyrazi się dokładnie poprzez funkcję kwadratową:

$$f(y_1, \ldots, y_n) = f(0, \ldots, 0) + \sum_{i,j=1}^{n} b_{ij} y_i y_j.$$

Ścisłe wyprowadzenie teorii małych drgań czytelnik może znaleźć w podręczniku W. I. Arnolda, *Metody matematyczne mechaniki klasycznej*, PWN, Warszawa 1981.

6. Teoria Morse'a. Wyobraźmy sobie n-wymiarową gładką, ograniczoną powierzchnię V zawartą w $(n+1)$-wymiarowej przestrzeni euklidesowej, np. coś w rodzaju skorupki jajka lub obwarzanka (torusa) w $\boldsymbol{R}^3$. Rozważmy przekroje V hiperpłaszczyznami $x_{n+1} = \text{const}$. Założymy, że istnieje tylko skończony zbiór takich wartości $c_1, \ldots, c_m$, że hiperpłaszczyzny $x_{n+1} = c_i$ są styczne do V i to w jedynym punkcie $v_i \in V$. W pobliżu punktu styczności można przybliżać V wykresem formy kwadratowej $x_{n+1} = c_i + q_i(x_1 - x_1(v_i), \ldots, x_n - x_n(v_i))$, jeśli tylko V jest w dostatecznie ogólnym położeniu (np. obwarzanek nie powinien leżeć poziomo). Okazuje się, że *niektóre z ważniejszych własności topologicznych* V, a w szczególności tzw. typ homotopijny V — *są całkowicie wyznaczone przez układ sygnatur form* q_i, tj. przez wskazanie tych kierunków w otoczeniu v_i, wzdłuż których V opada w dół i tych, wzdłuż których wznosi się w górę. Najbardziej zadziwiające jest to, że chociaż informacja o sygnaturach ma czysto lokalny charakter, to uzyskuje się z niej globalną charakterystykę V, jaką jest jej typ homotopijny. Na przykład, jeśli występują tylko dwa punkty krytyczne c_1, c_2 o sygnaturach $(n, 0)$ i $(0, n)$, to V jest topologicznie zbudowana jak n-wymiarowa sfera.

Szczegóły czytelnik może znaleźć w książce J. Milnora *Morse theory* (w jęz. ang.).

7. Operatory samosprzężone i wielowymiarowe kwadryki. Niech L oznacza przestrzeń euklidesową lub skończenie wymiarową unitarną, $f\colon L \to L$ — operator samosprzężony. Zbadamy własności widma tego operatora. Uporządkujmy wartości własne f malejąco z uwzględnieniem ich krotności: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ i obierzmy odpowiednio bazę ortonormalną złożoną z wektorów własnych $\{e_1, \ldots, e_n\}$. Powrócimy do punktu widzenia z § 8, p. 4, w którym zadanie operatora f traktowaliśmy jako równoważne z zadaniem formy symetrycznej lub hermitowskiej $(f(l_1), l_2)$ albo formy kwadratowej $q_f(l) = (f(l), l)$ (dla przypadku unitarnego jest ona kwadratowa na realifikacji przestrzeni). W bazie $\{e_1, \ldots, e_n\}$ ta forma przyjmuje postać

$$q_f(x, \ldots, x) = \sum_{i=1}^{n} \lambda_i x_i^2 \quad \text{lub} \quad \sum_{i=1}^{n} \lambda_i |x_i|^2,$$

skąd wnioskujemy, że kierunki $\boldsymbol{R}e_i$ (albo $\boldsymbol{C}e_i$) są *osiami głównymi* q_f.

Jedną z prostszych własności wariacyjnych zbioru wartości własnych λ_i wyraża następujące stwierdzenie.

8. Stwierdzenie. *Niech* $S = \{l \in L \mid |l| = 1\}$ *oznacza sferę jednostkową przestrzeni* L. *Wtedy*

$$\lambda_1 = \max_{l \in S} q_f(l), \quad \lambda_n = \min_{l \in S} q_f(l).$$

Dowód. Ponieważ $|x_i|^2 \geq 0$ i $\lambda_1 \geq \ldots \geq \lambda_n$, więc oczywiście

$$\lambda_n\Big(\sum_{i=1}^{n} |x_i|^2\Big) \leq \sum_{i=1}^{n} \lambda_i |x_i|^2 \leq \lambda_1\Big(\sum_{i=1}^{n} |x_i|^2\Big).$$

Na sferze jednostkowej lewa strona jest równa λ_n, a prawa λ_1 i te wartości są osiągane dla wektorów $(0, \ldots, 0, 1)$, $(1, 0, \ldots, 0)$ odpowiednio (współrzędne odnoszą się do bazy $\{e_1, \ldots, e_n\}$ diagonalizującej f).

9. Wniosek. *Oznaczmy przez* L_k^- *powłokę liniową rodziny* $\{e_1, \ldots, e_k\}$, *a przez* L_k^+ *— powłokę liniową rodziny* $\{e_k, \ldots, e_n\}$. *Wtedy*

$$\lambda_k = \max\{q_f(l) \mid l \in S \cap L_k^+\} = \min\{q_f(l) \mid l \in S \cap L_k^-\}.$$

Dowód. Rzeczywiście, przy oczywistym wyborze współrzędnych obcięcie q_f do L_k^+ ma postać $\sum_{i=1}^{n} \lambda_i |x_i|^2$, a obcięcie do L_k^- — postać $\sum_{i=1}^{k} \lambda_i |x_i|^2$.

Podane niżej ważne wzmocnienie tego wyniku, gdzie zamiast L_k^+ rozważana jest dowolna podprzestrzeń liniowa w L kowymiaru $k-1$, nosi nazwę *twierdzenia Fischera–Couranta*. Daje ono charakterystykę wartości własnych operatorów różniczkowych w postaci warunku typu „minimax".

10. Twierdzenie. *Dla dowolnej podprzestrzeni* $L' \subset L$ *kowymiaru* $k-1$ *są spełnione nierówności*

$$\lambda_k \leq \max\{q_f(l) \mid l \in S \cap L'\}, \quad \lambda_{n-k+1} \geq \min\{q_f(l) \mid l \in S \cap L'\}.$$

Te oszacowania są dokładne dla pewnych L' (*np. odpowiednio dla* L_k^+ *i* L_k^-), *tak więc*

$$\lambda_k = \min_{L'} \max\{q_f(l) | l \in S \cap L'\}, \quad \lambda_{n-k+1} = \max_{L'} \min\{q_f(l) | l \in S \cap L'\}.$$

Dowód. Ponieważ

$$\dim L' + \dim L_k^- = (n-k+1) + k = n+1,$$

a $\dim(L' + L_k^-) \leq \dim L = n$, więc z twierdzenia części 1, § 5, p. 3 wynika, że $\dim(L' \cap L_k^-) \geq 1$. Wybierzmy wektor $l_0 \in L' \cap L_k^- \cap S$. Zgodnie z wnioskiem

p. 9 mamy $\lambda_k = \min\{q_f(l) \mid l \in S \cap L_k^-\}$, a więc $\lambda_k \leq q_f(l_0)$, a tym bardziej $\lambda_k \leq \max\{q_f(l) \mid l \in S \cap L'\}$. Drugą nierówność twierdzenia najprościej wyprowadzić stosując już wykazaną nierówność do operatora $-f$ i zauważając, że znaki i kolejność wartości własnych ulegają przy tym odwróceniu.

11. Wniosek. *Niech* $\dim L/L_0 = 1$ *i niech* p *oznacza operator rzutowania* $L \to L_0$. *Oznaczmy przez* $\lambda'_1 \geq \lambda'_2 \geq \ldots \geq \lambda'_{n-1}$ *wartości własne operatora samosprzężonego* $pf: L_0 \to L_0$. *Wtedy*

$$\lambda_1 \geq \lambda'_1 \geq \lambda_2 \geq \lambda'_2 \geq \ldots \geq \lambda'_{n-1} \geq \lambda_n,$$

tj. wartości własne operatorów f *i* pf *przeplatają się.*

Dowód. Obcięcie formy q_f do L_0 jest równe formie q_{pf}: $(f(l), l) = (pf(l), l)$, jeżeli $l \in L_0$. Dlatego

$$\lambda'_k = \max\{q_{pf}(l) | l \in S \cap L'\} = \max\{q_f(l) | l \in S \cap L'\}$$

dla odpowiednio wybranej podprzestrzeni $L' \subset L_0$ o kowymiarze $k-1$ względem L_0. To oznacza, że jej kowymiar względem L jest równy k, skąd wynika $\lambda_{k+1} \leq \lambda'_k$. Stosując tę samą nierówność do $-f$ zamiast f, otrzymamy $-\lambda_k \leq -\lambda'_k$, tj. $\lambda'_k \leq \lambda_k$, co kończy dowód.

Pozostawiamy czytelnikowi możliwość sprawdzenia tego, że wniosek p. 11 ma następującą prostą interpretację geometryczną. Załóżmy, że $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$ i zamiast funkcji $q_f(l)$ na S rozważmy elipsoidę ϵ: $q_f(l) = 1$. Jej przecięcie ϵ_0 z podprzestrzenią L_0 jest także elipsoidą, której półosie mają długości przeplatające się z długościami półosi elipsoidy ϵ. Można wyobrażać sobie np. elipsoidę w $\boldsymbol{R}^3$ i elipsę ϵ_0 powstającą z przecięcia jej płaszczyzną. Wielka półoś ϵ_0 nie jest dłuższa od największej półosi ϵ („oczywiste"), ale i nie krótsza od średniej półosi ϵ. Natomiast mała półoś ϵ_0 jest nie krótsza od małej półosi ϵ („oczywiste"), ale i nie większa od średniej półosi ϵ. Pytanie kontrolne: jak otrzymać w przekroju okrąg?

§ 11. Trójwymiarowa przestrzeń euklidesowa

1. Trójwymiarowa przestrzeń euklidesowa $\mathcal{E}$ jest podstawowym modelem fizycznej przestrzeni Newtona i Galileusza. Czterowymiarowa przestrzeń Minkowskiego $\mathcal{M}$ wyposażona w symetryczną metrykę o sygnaturze $(r_+, r_-) = (1, 3)$, jest modelem czasoprzestrzeni fizyki relatywistycznej. Przynajmniej z tego względu zasługują one na uważniejsze zbadanie. Z matematycznego punktu widzenia mają także szczególne własności, istotne dla zrozumienia budowy świata, w którym żyjemy: związek obrotów $\mathcal{E}$ z kwaternionami i istnienie iloczynu wektorowego; geometria wektorów zerowej długości w $\mathcal{M}$.

Te szczególne własności wygodnie badać w świetle związku geometrii $\mathcal{E}$ i $\mathcal{M}$ z geometrią pomocniczej *dwuwymiarowej przestrzeni unitarnej* $\mathcal{H}$, *nazywanej przestrzenią spinorów*. Ten związek ma także głębokie znaczenie fizyczne, które odkryto

dopiero po narodzinach mechaniki kwantowej. Tu wybraliśmy właśnie tę drogę wykładu.

2. Ustalimy na początku dwuwymiarową przestrzeń unitarną $\mathcal{H}$. Dalej oznaczymy przez $\mathcal{E}$ *rzeczywistą przestrzeń liniową operatorów samosprzężonych w* $\mathcal{H}$ *ze śladem* 0. Każdy operator $f \in \mathcal{H}$ ma dwie rzeczywiste wartości własne, które różnią się tylko znakiem, bo ich suma, równa śladowi operatora, jest zerem. Zdefiniujemy

$$|f| = \sqrt{|\det f|}$$

(jest to dodatnia wartość własna f).

3. Stwierdzenie. *$\mathcal{E}$ wraz z normą $|\cdot|$ jest trójwymiarową przestrzenią euklidesową.*

Dowód. Względem bazy ortonormalnej $\mathcal{H}$ operatory f są reprezentowane macierzami hermitowskimi postaci

$$\begin{pmatrix} a & \overline{b} \\ b & -a \end{pmatrix}, \quad a \in \boldsymbol{R},\ b \in \boldsymbol{C},$$

tj. kombinacjami liniowymi

$$\operatorname{Re} b \cdot \sigma_1 + \operatorname{Im} b \cdot \sigma_2 + a \cdot \sigma_3,$$

gdzie $\sigma_1, \sigma_2, \sigma_3$ oznaczają macierze Pauliego (por. ćwiczenie 5 do § 4, części 1.);

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

Ponieważ $\sigma_1, \sigma_2, \sigma_3$ są liniowo niezależne nad $\boldsymbol{R}$, więc $\dim_{\boldsymbol{R}} \mathcal{E} = 3$. Określimy dalej

$$(f, g) = \operatorname{Tr}(fg).$$

Jest to dwuliniowy symetryczny iloczyn skalarny, więc jeśli wartości własne f oznaczymy przez $\pm\lambda$, to

$$|f|^2 = \frac{1}{2}\operatorname{Tr}(f^2) = \frac{1}{2}(\lambda^2 + \lambda^2) = |\det f|.$$

Oczywiście $\lambda^2 = 0$ wtedy i tylko wtedy, gdy $f = 0$, co kończy dowód.

Nazwiemy kierunkiem w $\mathcal{E}$ zbiór wektorów postaci

$$\boldsymbol{R}_+ f = \{af \mid a > 0\},$$

gdzie f jest niezerowym wektorem z $\mathcal{E}$. Innymi słowy, kierunek to półprosta w $\mathcal{E}$. Kierunkiem przeciwnym do $\boldsymbol{R}_+ f$ jest $\boldsymbol{R}_+(-f)$.

4. Stwierdzenie. *Istnieje wzajemnie jednoznaczna odpowiedniość między kierunkami w $\mathcal{E}$ a rozkładami $\mathcal{H}$ na sumę prostą dwóch ortogonalnych jednowymiarowych podprzestrzeni $\mathcal{H}_+ \oplus \mathcal{H}_-$. A mianowicie, kierunkowi $\boldsymbol{R}_+ f$ przyporządkowujemy $\mathcal{H}_+$ — podprzestrzeń własną $\mathcal{H}$ odpowiadającą dodatniej wartości własnej f i podprzestrzeń $\mathcal{H}_-$ odpowiadającą ujemnej wartości własnej f.*

Dowód. $\mathcal{H}_+$ i $\mathcal{H}_-$ są ortogonalne na mocy twierdzenia § 7, p. 4, a zamiana f na af, gdzie $a > 0$, nie zmienia $\mathcal{H}_+$ ani $\mathcal{H}_-$. Odwrotnie, jeśli jest zadany rozkład ortogonalny $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$, to zbiór tych operatorów $f \in \mathcal{E}$, które rozciągają $\mathcal{H}$ w kierunku $\mathcal{H}_+$ $\lambda > 0$ razy oraz $-\lambda < 0$ razy w kierunku $\mathcal{H}_-$, jest kierunkiem w $\mathcal{E}$.

5. Interpretacja fizyczna. Utożsamimy $\mathcal{E}$ z przestrzenią fizyczną, np. wybierając w nich ortogonalne współrzędne. Dalej utożsamiamy $\mathcal{H}$ z przestrzenią stanów wewnętrznych układu kwantowego nazywanego „cząstka o spinie 1/2, zlokalizowana w pobliżu początku układu współrzędnych" (np. elektron). Przyjmiemy także, że w przestrzeni fizycznej jest włączone pole magnetyczne o wybranym, ustalonym z góry kierunku $\boldsymbol{R}_+ f \subset \mathcal{E}$, . W tym polu układ ma dwa stany stacjonarne, którymi są właśnie $\mathcal{H}_+$ i $\mathcal{H}_-$.

Powiedzmy, że wybrany kierunek $\boldsymbol{R}_+ f$ odpowiada górnej pionowej półosi wybranego układu współrzędnych w przestrzeni fizycznej („oś z"). Wówczas stan $\mathcal{H}_+$ będziemy nazywać stanem „o rzucie spinu $+1/2$ na kierunek osi z" (albo „spin do góry") i odpowiednio stanem „o rzucie spinu $-1/2$" (albo „spin do dołu") — $\mathcal{H}_-$. Taka tradycyjna terminologia jest reliktem przedkwantowych wyobrażeń o tym, że obserwabla spinu odpowiada klasycznej wielkości obserwowalnej zwanej „momentem pędu", będącej charakterystyką wewnętrznego ruchu obrotowego układu, i dlatego można ją samą reprezentować poprzez wektor z $\mathcal{E}$ i mówić o jego rzutach na osie współrzędnych w $\mathcal{E}$. Jest to istotne nadużycie, bo stany układu są promieniami w $\mathcal{H}$, a nie wektorami w $\mathcal{E}$. Rozdźwięk z teorią klasyczną staje się jeszcze bardziej widoczny przy rozważaniu układów ze spinem $s/2$, dla których $\dim \mathcal{H} = s + 1$. Dokładnym sformułowaniem tego faktu jest właśnie stwierdzenie p. 4.

Wyżej podaliśmy wyidealizowany opis klasycznego ekperymentu Sterna i Gerlacha (1922). Zamiast elektronów użyto w nim jonów srebra, przepuszczanych między okładkami elektromagnesu. Ze względu na niejednorodność pola magnetycznego jony wychodzące w stanach bliskich stanom $\mathcal{H}_+$ i odpowiednio, $\mathcal{H}_-$ rozdzielały się na dwa pęki, co właśnie pozwoliło rozróżnić te stany. Srebro było wyparowywane w elektrycznym piecu, a pole magnetyczne pełniło rolę dwóch filtrów rozdzielających stany $\mathcal{H}_+$ i $\mathcal{H}_-$.

Powrócimy do badania przestrzeni euklidesowej $\mathcal{E}$.

6. Stwierdzenie. $(f, g) = 0$ *wtedy i tylko wtedy, gdy* $fg + gf = 0$.

Dowód. Mamy

$$(f, g) = \frac{1}{2}\operatorname{Tr}(fg + gf) = \frac{1}{4}\operatorname{Tr}[(f + g)^2 - f^2 - g^2].$$

Ponieważ f^2 ma jedyną wartość własną $|f|^2$, więc każdy kwadrat operatora z $\mathcal{E}$ jest operatorem skalarnym, a więc także $fg + gf$ jest operatorem skalarnym. Jest on równy zeru wtedy i tylko wtedy, gdy jego ślad jest równy zeru.

7. Bazy ortonormalne w $\mathcal{E}$. Z dowodu stwierdzenia p. 5 wynika natychmiast, że operatory $\{e_1, e_2, e_3\}$ tworzą bazę ortonormalną wtedy i tylko wtedy, gdy

$$e_1^2 = e_2^2 = e_3^2 = \mathrm{id}, \quad e_i e_j + e_j e_i = 0, \quad i \neq j.$$

W szczególności, jeśli w $\mathcal{H}$ została wybrana baza ortonormalna, to operatory w $\mathcal{H}$ wyznaczone przez macierze Pauliego $\sigma_1, \sigma_2, \sigma_3$ tworzą bazę ortonormalną w $\mathcal{E}$:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_i \sigma_j + \sigma_j \sigma_i = 0, \quad i \neq j.$$

Możemy teraz wyjaśnić matematyczne znaczenie macierzy Pauliego, dowodząc twierdzenia odwrotnego.

8. Stwierdzenie. *Dla każdej ortonormalnej bazy $\{e_1, e_2, e_3\}$ przestrzeni $\mathcal{E}$ istnieje baza ortonormalna $\{h_1, h_2\}$ przestrzeni $\mathcal{H}$, taka że*

$$A_{e_1} = \sigma_1, \quad A_{e_2} = \sigma_2 \text{ lub } -\sigma_2, \quad A_{e_3} = \sigma_3,$$

gdzie A_e oznacza macierz operatora e względem bazy $\{h_1, h_2\}$. Ta baza jest wyznaczona jednoznacznie z dokładnością do czynnika, będącego liczbą zespoloną o module jeden.

Dowód. Wartościami własnymi e_i są ± 1. Zatem $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$, gdzie e_3 jest identycznością na $\mathcal{H}_+$, a minus identycznością na $\mathcal{H}_-$. Na początek wybierzemy wektory $h_1' \in \mathcal{H}_+$, $h_2' \in \mathcal{H}_-$, spełniające $|h_1'| = |h_2'| = 1$. Są one wyznaczone z dokładnością do czynników $e^{i\varphi_1}$ oraz $e^{i\varphi_2}$, a macierzą e_3 w tej bazie jest σ_3. Dalej,

$$e_1(h_1') = e_1 e_3(h_1') = -e_3 e_1(h_1'),$$

co pokazuje, że $e_1(h_1')$ jest niezerowym wektorem własnym dla e_3 odpowiadającym wartości własnej -1. Zatem $e_1(h_1') = \alpha h_2'$ i analogicznie $e_1(h_2') = \beta h_1'$. Macierz e_1 względem bazy $\{h_1', h_2'\}$ jest hermitowska, więc $\alpha = \overline{\beta}$. Ponieważ $e_1^2 = \mathrm{id}$, zatem $\alpha\beta = 1 = |\alpha| = |\beta|$. Zamieniając $\{h_1', h_2'\}$ na $\{h_1, h_2\} = \{xh_1', yh_2'\}$, gdzie $|x| = |y| = 1$, po to aby jako macierz e_1 w nowej bazie otrzymać σ_1, dostajemy

$$e_1(h_1) = xe_1(h_1') = x\alpha h_2' = \alpha x y^{-1} h_2,$$
$$e_1(h_2) = ye_1(h_2') = y\beta h_1' = \beta y x^{-1} h_1.$$

Jeżeli x, y będą jeszcze spełniać warunek $xy^{-1} = \alpha^{-1}$, to otrzymamy wówczas automatycznie $\alpha x y^{-1} = \beta y x^{-1} = 1$. Możemy zatem przyjąć, np. $x = 1$, $y = \alpha$.

W ten sposób w bazie $\{h_1, h_2\}$ mamy $A_{e_3} = \sigma_3$, $A_{e_1} = \sigma_1$, a przy tym ta baza jest wyznaczona z dokładnością do czynnika skalarnego o module jeden. Te same argumenty, co w przypadku e_1, pokazują, że w tej bazie A_{e_2} jest macierzą $\begin{pmatrix} 0 & \gamma \\ \overline{\gamma} & 0 \end{pmatrix}$, gdzie $|\gamma|^2 = 1$. W takim razie z warunku ortogonalności $e_1e_2 + e_2e_1 = 0$ otrzymujemy

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & \gamma \\ \overline{\gamma} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \gamma \\ \overline{\gamma} & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

tj. $\gamma + \overline{\gamma} = 0$, skąd $\gamma = i$ lub $\gamma = -i$. Zatem $A_{e_2} = \sigma_2$ lub $A_{e_2} = -\sigma_2$.

9. Wniosek. *Przestrzeń $\mathcal{E}$ ma wyróżnioną orientację: ortonormalna baza $\{e_1, e_2, e_3\}$ należy do klasy zadającej tę orientację wtedy i tylko wtedy, gdy istnieje ortonormalna baza $\{h_1, h_2\}$ przestrzeni $\mathcal{H}$, względem której $A_{e_a} = \sigma_a$, $a = 1, 2, 3$.*

Dowód. Musimy sprawdzić, że jeśli $\{e_a\}$ w bazie $\{h_b\}$ i $\{e'_a\}$ w bazie $\{h'_b\}$ wyrażają się dokładnie macierzami Pauliego, to wyznacznik macierzy przejścia od $\{e_a\}$ do $\{e'_a\}$ jest dodatni lub też, że istnieje ciągła deformacja przeprowadzająca $\{e_a\}$ w $\{e'_a\}$. Skonstruujemy taką deformację pokazując, że $\{h_b\}$ można przeprowadzić w $\{h'_b\}$ ciągłą deformacją unitarną, tj. że istnieje taka rodzina operatorów unitarnych $f_t: \mathcal{H} \to \mathcal{H}$, indeksowana parametrem $t \in [0, 1]$ i spełniająca warunki: $f_0 = \mathrm{id}$, $f_1(h_b) = h'_b$, $\{f_t(h_1), f_t(h_2)\}$ jest bazą ortonormalną w $\mathcal{H}$ dla każdego t. Wtedy, oznaczając przez $\{g_t(e_1), g_t(e_2), g_t(e_3)\}$ bazę ortonormalną $\mathcal{E}$ wyrażoną przez macierze Pauliego w bazie $\{f_t(h_1), f_t(h_2)\}$, otrzymamy szukaną deformację przestrzeni $\mathcal{E}$.

Niech $\{h'_1, h'_2\} = \{h_1, h_2\}U$. Ze względu na to, że obie bazy są ortonormalne, macierz przejścia U jest unitarna. Na mocy wniosku § 8, p. 7 można ją zapisać w postaci $\exp(iA)$ z macierzą hermitowską A. Wtedy macierz tA jest hermitowska, a operator $\exp(itA)$ jest unitarny dla każdego t, co pozwala przyjąć

$$f_t\{h_1, h_2\} = \{h_1, h_2\}\exp(itA), \quad 0 \leq t \leq 1.$$

To kończy dowód.

Operatory $\sigma_1/2$, $\sigma_2/2$, $\sigma_3/2$ w $\mathcal{H}$ nazywają się *operatorami rzutu spinu* na odpowiednie osie w $\mathcal{E}$ — nazwa ta pochodzi od kwantowo-mechanicznej interpretacji tych operatorów, która była przedstawiona w p. 5. Czynnik $1/2$ jest wprowadzony po to, by wartościami własnymi tych operatorów były liczby $\pm 1/2$.

10. Iloczyn wektorowy. Niech $\{e_1, e_2, e_3\}$ będzie bazą ortonormalną w $\mathcal{E}$, zgodną z wyróżnioną orientacją. Iloczyn wektorowy w $\mathcal{E}$ jest zdefiniowany klasycznym wzorem

$$(x_1e_1 + x_2e_2 + x_3e_3) \times (y_1e_1 + y_2e_2 + y_3e_3) =$$
$$= (x_2y_3 - x_3y_2)e_1 + (x_3y_1 - x_1y_3)e_2 + (x_1y_2 - x_2y_1)e_3.$$

Zamiana bazy na inną, ale jednakowo zorientowaną, nie zmienia iloczynu wektorowego, ale jeśli nowa baza jest przeciwnie zorientowana, to iloczyn zmienia znak.

Nietrudno podać niezmienniczą konstrukcję iloczynu wektorowego opartą na omawianym formalizmie. Przypomnijmy, że *antyhermitowskie* operatory w $\mathcal{H}$ ze śladem 0 tworzą algebrę Liego $\mathfrak{su}(2)$ (por. § 4, część 1). Przestrzeń $\mathcal{E}$ możemy utożsamić z tą algebrą Liego dzieląc każdy operator należący do $\mathcal{E}$ przez i. Zatem w $\frac{1}{i}\mathcal{E}$ jest określona struktura algebry Liego dla której

$$\left[\frac{1}{i}\sigma_a, \frac{1}{i}\sigma_b\right] = 2\varepsilon_{abc}\frac{1}{i}\sigma_c,$$

gdzie $\varepsilon_{123} = 1$ oraz ε_{abc} jest skośnie symetryczny względem wszystkich wskaźników. Inaczej

$$[\sigma_a, \sigma_b] = 2i\varepsilon_{abc}\sigma_c.$$

Dlatego

$$\left[\sum_{a=1}^{3} x_a\sigma_a, \sum_{a=1}^{3} y_a\sigma_a\right] = 2i\left(\sum_{a=1}^{3} x_a\sigma_a\right) \times \left(\sum_{a=1}^{3} y_a\sigma_a\right),$$

a więc *iloczyn wektorowy z dokładością do trywialnego czynnika jest zwykłym komutatorem operatorów.* To pozwala na napisanie bez żadnych obliczeń klasycznych tożsamości

$$\vec{x} \times \vec{y} = -\vec{y} \times \vec{x},$$

$$\vec{x} \times (\vec{y} \times \vec{z}) + \vec{z} \times (\vec{x} \times \vec{y}) + \vec{y} \times (\vec{z} \times \vec{x}) = 0.$$

Inny sposób wprowadzenia iloczynu wektorowego polega na powiązaniu go z iloczynem skalarnym i kwaternionami.

11. Kwaterniony. Podobnie jak komutowanie, tak i mnożenie operatorów należących do $\mathcal{E}$ na ogół wyprowadza poza $\mathcal{E}$ — nie są zachowane jednocześnie ani hermitowskość operatorów ani warunek znikania ich śladu. W rzeczywistości iloczyn operatorów należących do $\mathcal{E}$ jest elementem przestrzeni $\boldsymbol{R}\mathrm{id} + i\mathcal{E}$, przy czym jego „część rzeczywista" jest iloczynem skalarnym, a „część urojona" — iloczynem wektorowym. W istocie,

$$\sigma_a\sigma_b = i\varepsilon_{abc}\sigma_c \quad \text{dla } a \neq b, \quad \{a,b,c\} = \{1,2,3\},$$

$$\sigma_a^2 = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad a = 1,2,3,$$

skąd otrzymujemy

$$\left(\sum_{a=1}^{3} x_a\sigma_a\right) \cdot \left(\sum_{a=1}^{3} y_a\sigma_a\right) = \left(\sum_{a=1}^{3} x_a y_a\right)\sigma_0 + i\left(\sum_{a=1}^{3} x_a\sigma_a\right) \times \left(\sum_{a=1}^{3} y_a\sigma_a\right)$$

albo tak jak to piszą fizycy,

$$(\vec{x} \cdot \vec{\sigma})(\vec{y} \cdot \vec{\sigma}) = (\vec{x} \cdot \vec{y})\sigma_0 + i(\vec{x} \times \vec{y}) \cdot \vec{\sigma}, \quad \text{gdzie } \vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3).$$

Stąd wynika, że rzeczywista przestrzeń operatorów $\boldsymbol{R}\mathrm{id}+i\mathcal{E}$ jest zamknięta względem mnożenia. Jej bazę tworzą elementy w klasycznej notacji zapisywane jako

$$\mathbf{1}=\sigma_0, \quad \boldsymbol{i}=-i\sigma_1, \quad \boldsymbol{j}=-i\sigma_2, \quad \boldsymbol{k}=-i\sigma_3$$

z tabliczką mnożenia

$$\boldsymbol{i}^2=\boldsymbol{j}^2=\boldsymbol{k}^2=-\mathbf{1}, \quad \boldsymbol{ij}=-\boldsymbol{ji}=\boldsymbol{k}, \quad \boldsymbol{ki}=-\boldsymbol{ik}=\boldsymbol{j}, \quad \boldsymbol{jk}=-\boldsymbol{kj}=\boldsymbol{i}.$$

Innymi słowy, otrzymaliśmy *algebrę kwaternionów* w jednej z tradycyjnych reprezentacji macierzowych (zob. *Wstęp do algebry*, rozdz. 9, § 4).

12. Homomorfizm $\boldsymbol{SU}(2) \to \boldsymbol{SO}(3)$. Ustalmy bazę ortonormalną $\{h_1, h_2\}$ w $\mathcal{H}$ i odpowiadającą jej bazę ortonormalną $\{e_1, e_2, e_3\}$ w $\mathcal{E}$, dla której $A_{e_i}=\sigma_i$. Każdy operator unitarny $U\colon \mathcal{H} \to \mathcal{H}$ przeprowadza bazę $\{h_1, h_2\}$ na inną bazę $\{h_1', h_2'\}$, której odpowiada z kolei baza $\{e_1', e_2', e_3'\}$ i operator $s(U)\colon \mathcal{E} \to \mathcal{E}$ przeprowadzający $\{e_i\}$ na $\{e_i'\}$. Na mocy wniosku p. 8 $s(U) \in \boldsymbol{SO}(3)$, bo wyznacznik $s(U)$ jest dodatni.

Przedstawiając elementy $\mathcal{E}$ jako macierze względem bazy $\{h_1, h_2\}$ możemy działanie $s(U)$ na $\mathcal{E}$ zapisać za pomocą prostego wzoru

$$s(U)(A)=UAU^{-1}$$

dla dowolnych $A \in \mathcal{E}$. Jest to w istocie szczególny przypadek ogólnego wzoru na transformację macierzy operatora przy zmianie bazy. Teraz możemy wykazać następujący ważny rezultat.

13. Twierdzenie. *Obcięcie odwzorowania s do $\boldsymbol{SU}(2)$ jest surjektywnym homomorfizmem grupy $\boldsymbol{SU}(2) \longrightarrow \boldsymbol{SO}(3)$ z jądrem $\{\pm E_2\}$.*

Dowód. Ze wzoru $s(U)(A)=UAU^{-1}$ od razu widać, że $s(E)=\mathrm{id}$ i $s(UV)=s(U)s(V)$, skąd wynika, że s jest homomorfizmem grup. Surjektywność można wykazać następująco.

Wybierzmy element $g \in \boldsymbol{SO}(3)$ i niech g przeprowadza bazę $(\sigma_1, \sigma_2, \sigma_3)$ przestrzeni $\mathcal{E}$ na bazę $(\sigma_1', \sigma_2', \sigma_3')$. Zaczynając od tej drugiej bazy, skonstruujmy bazę $\{h_1', h_2'\}$ przestrzeni $\mathcal{H}$, w której operatory σ_i' są reprezentowane macierzami σ_i. Na mocy stwierdzenia p. 7 można taką bazę $\{h_1', h_2'\}$ znaleźć, przy czym być może, macierzą operatora σ_2' będzie $-\sigma_2$, a nie σ_2. Jednakże ta druga możliwość jest wykluczona przez wniosek p. 8, bo $g \in \boldsymbol{SO}(3)$ zachowuje orientację $\mathcal{E}$. Operatór U, przeprowadzający $\{h_1, h_2\}$ na $\{h_1', h_2'\}$ spełnia $s(U)=g$, ale, być może, nie należy do $\boldsymbol{SU}(2)$, a tylko do $\boldsymbol{U}(2)$. Jeśli tak jest, to $\det U=e^{i\varphi}$, więc $e^{-i\varphi/2}U \in \boldsymbol{SU}(2)$. Macierz $e^{-i\varphi/2}U$ przeprowadza bazę $\{h_1, h_2\}$ na bazę $\{e^{-i\varphi/2}h_1', e^{-i\varphi/2}h_2'\}$, a tej bazie jak poprzednio odpowiada baza $\{\sigma_1', \sigma_2', \sigma_3'\}$ przestrzeni $\mathcal{E}$. Stąd wynika, że $s(e^{-i\varphi/2}U)=g$, a więc $s\colon \boldsymbol{SU}(2) \to \boldsymbol{SO}(3)$ jest, jak twierdziliśmy, surjektywny.

Jądro homomorfizmu $s\colon \boldsymbol{SU}(2) \to \boldsymbol{SO}(3)$ zawiera tylko operatory skalarne $e^{i\varphi}\,\mathrm{id}$, co wynika ze stwierdzenia p. 7, zgodnie z którym baza $\{h_1', h_2'\}$ jest wyznaczona przez

$\{e'_1, e'_2, e'_3\}$ właśnie z dokładnością do czynnika $e^{i\varphi}$. Ponieważ częścią wspólną grupy $\{e^{i\varphi}\}$ z grupą $\boldsymbol{SU}(2)$ jest $\{\pm \mathrm{id}\}$, dowód został zakończony.

Znaczenie opisanego homomorfizmu staje się jaśniejsze przez odwołanie do topologii: grupa $\boldsymbol{SU}(2)$ jest *jednospójna*, tj. każdą zamkniętą krzywą w tej grupie można w sposób ciągły ściągnąć do punktu, co dla grupy $\boldsymbol{SO}(3)$ jest niemożliwe. A więc $\boldsymbol{SU}(2)$ jest uniwersalną grupą nakrywającą $\boldsymbol{SO}(3)$.

Skorzystamy z udowodnionego twierdzenia w tym celu, by rozjaśnić obraz wewnętrznej struktury grupy $\boldsymbol{SO}(3)$, ponieważ grupa $\boldsymbol{SU}(2)$ jest zbudowana prościej. Tu warto zacytować R. Feynmana:

„To raczej dziwne, bo przecież żyjemy w trzech wymiarach, a mimo to trudno nam ocenić, co się dzieje, jeżeli obrócimy się w taki sposób, a później w taki sposób. Być może, gdybyśmy byli rybami lub ptakami i mieli prawdziwe wyczucie tego, co się dzieje, gdy wywijamy w przestrzeni koziołki, łatwiej moglibyśmy zdawać sobie sprawę z takich rzeczy." R. Feynman, R. Leighton, M. Sands, *Feynmana wykłady z fizyki*, t. III, str. 90, PWN, Warszawa 1972.

14. Struktura $\boldsymbol{SU}(2)$. Po pierwsze, elementami $\boldsymbol{SU}(2)$ są macierze stopnia 2 o wyrazach zespolonych, spełniające $U^t = \overline{U}$ i $\det U = 1$. Stąd od razu wynika, że

$$\boldsymbol{SU}(2) = \left\{ \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \;\middle|\; |a|^2 + |b|^2 = 1 \right\}.$$

Zbiór par $\{(a,b) \mid |a|^2 + |b|^2 = 1\} \subset \boldsymbol{C}^2$ można uważać za sferę jednostkową w przestrzeni $\boldsymbol{R}^4$ traktowanej jako realifikacja $\boldsymbol{C}^2$:

$$(\mathrm{Re}\, a)^2 + (\mathrm{Im}\, a)^2 + (\mathrm{Re}\, b)^2 + (\mathrm{Im}\, b)^2 = 1.$$

Stąd wnioskujemy, że grupa $\boldsymbol{SU}(2)$ *jest zbudowana topologicznie tak jak trójwymiarowa sfera w czterowymiarowej przestrzeni euklidesowej.*

Wniosek § 8, p. 7, zgodnie z którym odwzorowanie $\exp\colon \mathfrak{u}(2) \to \boldsymbol{U}(2)$ jest surjektywne, sugeruje pewien wybór układu generatorów grupy $\boldsymbol{SU}(2)$. Bezpośrednie obliczenie funkcji wykładniczej dla trzech tworzących przestrzeni $\mathfrak{su}(2)$ daje

$$\exp\left(\tfrac{1}{2} i t \sigma_1\right) = \begin{pmatrix} \cos\frac{t}{2} & i\sin\frac{t}{2} \\ i\sin\frac{t}{2} & \cos\frac{t}{2} \end{pmatrix},$$

$$\exp\left(\tfrac{1}{2} i t \sigma_2\right) = \begin{pmatrix} \cos\frac{t}{2} & \sin\frac{t}{2} \\ -\sin\frac{t}{2} & \cos\frac{t}{2} \end{pmatrix},$$

$$\exp\left(\tfrac{1}{2} i t \sigma_3\right) = \begin{pmatrix} e^{it/2} & 0 \\ 0 & e^{-it/2} \end{pmatrix}.$$

Każdy element $\begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \in \boldsymbol{SU}(2)$, taki że $ab \neq 0$ można zapisać w postaci

$$\exp\left(\frac{1}{2}i\varphi\sigma_3\right)\exp\left(\frac{1}{2}i\theta\sigma_1\right)\exp\left(\frac{1}{2}i\psi\sigma_3\right) = \begin{pmatrix} \cos\frac{\theta}{2}e^{i\frac{\varphi+\psi}{2}} & i\sin\frac{\theta}{2}e^{i\frac{\varphi-\psi}{2}} \\ i\sin\frac{\theta}{2}e^{i\frac{\psi-\varphi}{2}} & i\cos\frac{\theta}{2}e^{-i\frac{\psi-\varphi}{2}} \end{pmatrix},$$

gdzie $0 \leq \varphi < 2\pi$, $0 < \theta < \pi$, $-2\pi \leq \psi < 2\pi$. W tym celu wystarczy przyjąć $|a| = \cos\frac{\theta}{2}$, $\arg a = \frac{\varphi+\psi}{2}$, $\arg b = \frac{\varphi-\psi+\pi}{2}$. (Oczywiście, te elementy $\boldsymbol{SU}(2)$, dla których $b = 0$ mają postać $\exp\left(\frac{1}{2}i\varphi\sigma_3\right)$, a czytelnikowi pozostawiamy możliwość samodzielnego rozważenia przypadku $a = 0$.)

Kąty φ, θ, ψ nazywają się *kątami Eulera* dla grupy $\boldsymbol{SU}(2)$.

15. Struktura $\boldsymbol{SO}(3)$. Z topologicznego punktu widzenia utożsamiliśmy $\boldsymbol{SU}(2)$ z trójwymiarową sferą. Homomorfizm $s\colon \boldsymbol{SU}(2) \to \boldsymbol{SO}(3)$ przeprowadza parę punktów $\pm U \in \boldsymbol{SU}(2)$ na jeden punkt $\boldsymbol{SO}(3)$. Te punkty są końcami jednej średnicy sfery, można więc powiedzieć, że z topologicznego punktu widzenia $\boldsymbol{SO}(3)$ *powstaje przez sklejenia par przeciwległych punktów trójwymiarowej sfery.* Z drugiej strony, pary przeciwległych punktów sfery są we wzajemnie jednoznacznej odpowiedniości z prostymi łączącymi punkty tej pary w rzeczywistej przestrzeni czterowymiarowej. Zbiór takich prostych nazywa się *trójwymiarową rzeczywistą przestrzenią rzutową* i jest oznaczany $\boldsymbol{RP}^3$ — dokładniej zbadamy przestrzenie rzutowe później. W ten sposób $\boldsymbol{SO}(3)$ jest topologicznie równoważna z $\boldsymbol{RP}^3$.

Przyjrzyjmy się teraz, na co przeprowadza homomorfizm s opisane wyżej tworzące $\boldsymbol{SU}(2)$. W standardowej bazie $\{\sigma_1, \sigma_2, \sigma_3\}$ przestrzeni $\mathcal{E}$ spełnione są związki

$$\exp\left(\tfrac{1}{2}it\sigma_1\right)\sigma_1\exp\left(-\tfrac{1}{2}it\sigma_1\right) = \sigma_1,$$

$$\exp\left(\tfrac{1}{2}it\sigma_1\right)\sigma_2\exp\left(-\tfrac{1}{2}it\sigma_1\right) = (\cos t)\sigma_2 - (\sin t)\sigma_3,$$

$$\exp\left(\tfrac{1}{2}it\sigma_1\right)\sigma_3\exp\left(-\tfrac{1}{2}it\sigma_1\right) = (\sin t)\sigma_2 + (\cos t)\sigma_3.$$

A zatem

$$s\left(\exp\left(\tfrac{1}{2}it\sigma_1\right)\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

jest obrotem $\mathcal{E}$ o kąt t wokół osi $\boldsymbol{R}\sigma_1$. Całkowicie analogiczne sprawdzenie pokazuje, że także dla $k = 2, 3$ $s\left(\exp\left(\frac{1}{2}it\sigma_k\right)\right)$ jest obrotem $\mathcal{E}$ o kąt t wokół osi σ_k. W szczególności, dowolny obrót z $\boldsymbol{SO}(3)$ można przedstawić jako iloczyn trzech obrotów wokół $\sigma_1, \sigma_2, \sigma_3$ o kąty Eulera ψ, θ, φ, przy czym ψ można obrać w przedziale 0 do 2π.

Ćwiczenia

1. Dowieść tożsamości

$$(\vec{x} \times \vec{y}, \vec{z}) = (\vec{x}, \vec{y} \times \vec{z}),$$
$$\vec{x} \times (\vec{y} \times \vec{z}) - \vec{x} \times (\vec{y} \times \vec{z}) = (\vec{x}, \vec{y})\vec{z} - \vec{x}(\vec{y}, \vec{z}).$$

(*Wskazówka*. Skorzystać z łączności mnożenia w algebrze kwaternionów.)

2. W trójwymiarowej przestrzeni euklidesowej wyróżniono dwie osie z i z' tworzące między sobą kąt φ. Wiązka elektronów z rzutem spinu $+1/2$ wzdłuż osi z pada na filtr, przepuszczający jedynie elektrony o rzucie spinu $+1/2$ wzdłuż osi z'. Wykazać, że stosunek liczby elektronów przepuszczonych przez filtr do liczby elektronów zawartych w wiązce wynosi $\cos^2 \frac{\varphi}{2}$.

§ 12. Przestrzeń Minkowskiego

1. *Przestrzenią Minkowskiego* $\mathcal{M}$ nazywamy czterowymiarową rzeczywistą przestrzeń liniową z niezdegenerowaną symetryczną metryką o sygnaturze $(1,3)$ (czasami używa się sygnatury $(3,1)$). Zanim przystąpimy do omawiania własności tej przestrzeni z matematycznego punktu widzenia, wskażemy na te podstawowe zasady jej interpretacji fizycznej, które leżą u podstaw *szczególnej teorii względności* Einsteina.

a) *Punkty.* Punkt (albo wektor) przestrzeni $\mathcal{M}$ jest idealizacją fizycznego zdarzenia zlokalizowanego w czasie i przestrzeni, takiego jak „rozbłysk", „wypromieniowanie fotonu przez atom", „zderzenie dwóch cząstek elementarnych" itp. Początek układu współrzędnych w $\mathcal{M}$ można traktować jako zdarzenie zachodzące „tu i teraz" dla pewnego obserwatora — ustala ono jednocześnie początek odliczania czasu i początek odmierzania współrzędnych przestrzennych.

b) *Jednostki wymiarowe.* W fizyce klasycznej czas i długość są mierzone w innych jednostkach. Ze względu na to, że $\mathcal{M}$ jest modelem czasoprzestrzeni, w szczególnej teorii względności należy ustalić sposób przeliczania jednostek przestrzennych na czasowe i odwrotnie. Przyjęty sposób jest równoważny z postulatem „stałej prędkości światła c" i polega na tym, że wybranej jednostce czasu t_0 przyporządkowuje się jednostkę długości $l_0 = ct_0$ równą odległości przebywanej przez światło w czasie t_0 (np. „sekunda świetlna"). Jedną z jednostek t_0 lub l_0 uznajemy za wybraną raz na zawsze, a wskutek tego druga jest zadana warunkiem $l_0 = ct_0$ i prędkość światła w tych jednostkach staje się równa jedności.

c) *Interwał czasoprzestrzenny.* Dla dowolnych punktów $l_1, l_2 \in \mathcal{M}$ przestrzeni Minkowskiego iloczyn skalarny różnicy $(l_1 - l_2, l_1 - l_2)$ nazywa się kwadratem interwału czasoprzestrzennego pomiędzy nimi. Ten kwadrat może być liczbą dodatnią, ujemną lub zerem — w fizycznej terminologii odpowiednio *czasowym, przestrzennym*

lub zerowym. (Pewne uzasadnienie dla używania takich terminów będzie podane niżej.) Jeśli $l_2 = 0$, to te nazwy stosuje się także do samego wektora l_1 w zależności od znaku (l_1, l_1).

d) *Linie świata obserwatorów inercjalnych.* Jeśli choć jeden wektor prostej $L \subset \mathcal{M}$ jest czasowy, to i każdy inny jest też czasowy. Taki proste nazywamy *liniami świata obserwatorów inercjalnych.* Za dobre przybliżenie odcinka takiej linii można uważać zbiór zdarzeń zachodzących w statku kosmicznym, poruszającym się swobodnie (z wyłączonymi silnikami) w dużej odległości od ciał niebieskich (uwzględnienie ich przyciągania wymagałoby zmiany schematu matematycznego opisu czasoprzestrzeni i, w szczególności, przejścia do „zakrzywionych" modeli ogólnej teorii względności). Zauważmy, że na razie wprowadziliśmy do naszych rozważań jedynie linie świata wychodzące z początku układu współrzędnych. Inercjalny obserwator, który nie przeszedł „tu i teraz", porusza się po $l+L$ — przesuniętej o pewien wektor prostej czasowej L. Niech l_1, l_2 oznaczają dwa punkty należące do linii świata obserwatora inercjalnego. Wówczas $(l_1 - l_2, l_1 - l_2) > 0$ i interwał $|l_1 - l_2| = (l_1 - l_2, l_1 - l_2)^{1/2}$ ma interpretację *czasu, który upłynął pomiędzy zdarzeniami* l_1, l_2, *mierzonego według wskazań zegara poruszającego się z tym obserwatorem,* zwanego *czasem własnym obserwatora.* Linia świata obserwatora inercjalnego jest jego własnym „strumieniem czasu".

Konstatacja fizycznego faktu występowania „kierunku" czasu (z przeszłości do przyszłości) wyraża się matematycznie wyróżnieniem orientacji każdej prostej czasowej, tak że długość $|l|$ wektora czasowego może być wyposażona w znak dla odróżnienia wektorów skierowanych w przyszłość od tych skierowanych w przeszłość. Poniżej przekonamy się, że jest sensowne mówienie o uzgodnieniu tych orientacji, tj. o istnieniu wspólnego kierunku czasu — ale nie samego czasu! — dla różnych obserwatorów inercjalnych.

e) *Przestrzeń fizyczna obserwatora inercjalnego.* Podrozmaitość liniowa

$$\mathcal{E}_l = l + L^\perp \subset \mathcal{M}$$

jest interpretowana jako zbiór punktów „chwilowej przestrzeni fizycznej" obserwatora inercjalnego, znajdującego się w punkcie l swojej linii świata L. Ortogonalne dopełnienie jest, naturalnie, wzięte względem metryki Minkowskiego w $\mathcal{M}$. Nietrudno przekonać się, że $\mathcal{M} = L \oplus L^\perp$ oraz że na $L^\perp$ jest indukowana struktura trójwymiarowej przestrzeni euklidesowej (jednakże z ujemnie określoną metryką zamiast zwyczajnej dodatnio określonej). Wszystkie zdarzenia odpowiadające punktom $L^\perp$ są interpretowane przez obserwatora jako zachodzące „teraz", ale dla innego obserwatora nie będą one jednoczesne, bo $L_1^\perp \neq L_2^\perp$, gdy $L_1 \neq L_2$.

f) *Inercjalne układy współrzędnych.* Niech L oznacza prostą czasową z wybraną orientacją, e_0 — dodatnio zorientowany wektor czasowy na tej prostej o długości jeden, $\{e_1, e, e_3\}$ — ortonormalną bazę w $L^\perp$; $(e_i, e_i) = -1$ dla $i = 1, 2, 3$. Układ współrzędnych w $\mathcal{M}$ odpowiadający bazie $\{e_0, \ldots, e_3\}$ nazywa się *układem inercjal-*

nym. We współrzędnych w tym układzie

$$\left(\sum_{i=0}^{3} x_i e_i, \sum_{i=0}^{3} y_i e_i\right) = x_0 y_0 - \sum_{i=1}^{3} x_i y_i.$$

Ponieważ $x_0 = ct_0$, gdzie t_0 jest czasem własnym, interwał czasoprzestrzenny od początku układu do punktu $\sum_{i=0}^{3} x_i e_i$ jest równy $\left(c^2 t_0^2 - \sum_{i=1}^{3} x_i^2\right)^{1/2}$. Każdy inercjalny układ współrzędnych w $\mathcal{M}$ zadaje utożsamienie $\mathcal{M}$ z kartezjańską przestrzenią Minkowskiego $\left(\boldsymbol{R}^4, x^2 - \sum_{i=1}^{3} x_i^2\right)$. *Izometrie* $\mathcal{M}$ (lub jej kartezjańskiego odpowiednika) tworzą *grupę Lorentza*, a izometrie zachowujące orientacją czasu — *jej podgrupę ortochroniczną*.

g) *Stożek świetlny*. Zbiór punktów $l \in \mathcal{M}$, takich że $(l, l) = 0$, nazywamy *stożkiem świetlnym* C (początku układu). W dowolnym układzie inercjalnym C jest dany równaniem

$$x_0^2 = \sum_{i=1}^{3} x_i^2.$$

Przy $x_0 > 0$ punkt (x_0, x_1, x_2, x_3) należący do stożka jest oddzielony od położenia obserwatora $(x_0, 0, 0, 0)$ interwałem przestrzennym, którego kwadrat wyraża się wzorem $\sum_{i=1}^{3} x_i^2 = -x_0^2$, tj. punkt znajduje się w takiej odległości, jaką przejdzie w czasie x_0 kwant światła wypuszczony z początku układu w początkowej chwili czasu. (Przy $x_0 < 0$ zbiór takich punktów odpowiada rozbłyskom, które zaszły w momencie czasu własnego x_0 i które mogły zostać zaobserwowane w początku układu współrzędnych: „promieniowanie dochodzące".) Odpowiednio „proste zerowe", w całości leżące na stożku C, to linie świata cząstek wypuszczonych z początku układu współrzędnych i poruszających się z prędkością światła, np. fotonów. Czytelnik może zobaczyć podstawę „dochodzącej powłoki" stożka świetlnego wyglądając za okno — jest nią sfera niebieska.

Proste w $\mathcal{M}$ składające się z wektorów z ujemnym kwadratem interwału nie mają interpretacji fizycznej. Mogłyby one odpowiadać liniom świata cząstek poruszających się z prędkością większą od prędkości światła — hipotetycznych „tachionów", których jednak nie udało się dotychczas zaobserwować.

Przejdźmy teraz do zbadania $\mathcal{M}$ od strony matematycznej.

2. Realizacja $\mathcal{M}$ jako przestrzeni metryk. Tak jak w § 11, ustalimy dwuwymiarową przestrzeń zespoloną $\mathcal{H}$ i rozważymy zbiór $\mathcal{M}$ *hermitowskich iloczynów skalarnych* zadanych na $\mathcal{H}$. Jest to rzeczywista przestrzeń liniowa. Macierze Grama tych metryk w ustalonej bazie $\{h_1, h_2\}$ w $\mathcal{H}$ tworzą zbiór wszystkich macierzy hermitowskich 2×2. Metryce $l \in \mathcal{M}$ przyporządkujemy wyznacznik jej macierzy Grama G, który oznaczymy przez $\det l$. Przejście do bazy $\{h_1', h_2'\} = \{h_1, h_2\}V$ prowadzi do zmiany G na $G' = V^t G \overline{V}$, a $\det G' = |\det V|^2 \det G$. W szczególności, jeśli $V \in \boldsymbol{SL}(2, \boldsymbol{C})$, to $\det G = \det G'$. Zatem obliczenie $\det l$ w każdej z baz $\mathcal{H}$ należących do jednej klasy względem działania $\boldsymbol{SL}(2, \boldsymbol{C})$ daje jeden i ten sam wynik.

W dalszym ciągu *ustalamy taką klasę baz* $\mathcal{H}$ i wszystkie wyznaczniki będą obliczane względem niej. Zmiana klasy prowadzi tylko do pomnożenia otrzymanych wartości wyznaczników przez dodatni skalar.

3. Stwierdzenie. a) $\mathcal{M}$ *jest czterowymiarową przestrzenią rzeczywistą.*

b) *Na* $\mathcal{M}$ *istnieje jedyna symetryczna metryka* (l, m), *dla której* $(l, l) = \det l$. *Jej sygnatura jest równa* $(1,3)$, *a więc* $\mathcal{M}$ *jest przestrzenią Minkowskiego.*

Dowód. a) W przestrzeni macierzy hermitowskich 2×2 bazę stanowią macierze $\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E_2$, $\sigma_1, \sigma_2, \sigma_3$, gdzie σ_i dla $i \geq 1$ oznaczają macierze Pauliego. Dlatego $\dim \mathcal{M} = 4$.

b) Pokażemy, że w macierzowej realizacji $\mathcal{M}$ funkcja $\det l$ jest formą kwadratową, której forma biegunowa wyraża się wzorem

$$(l, m) = \frac{1}{2}(\operatorname{Tr} l \operatorname{Tr} m - \operatorname{Tr} lm),$$

w oczywisty sposób przedstawiającym formę dwuliniową symetryczną. W istocie, jeśli λ, μ są wartościami własnymi l, to mamy $\det l = \lambda\mu$, $\operatorname{Tr} l = \lambda+\mu$, $\operatorname{Tr} l^2 = \lambda^2+\mu^2$, tak że

$$\lambda\mu = \det l = \frac{1}{2}((\lambda+\mu)^2 - \lambda^2 - \mu^2) = \frac{1}{2}((\operatorname{Tr} l)^2 - \operatorname{Tr} l^2) = (l, l).$$

Teraz jest więc oczywiste, że $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ jest ortonormalną bazą $\mathcal{M}$ z macierzą Grama $\operatorname{diag}(1,-1,-1,-1)$, skąd wynika, że sygnatura naszej metryki jest równa $(1,3)$. To kończy dowód.

4. Wniosek. *Niech* $L \subset \mathcal{M}$ *oznacza prostą czasową. Wówczas* $L^\perp$ *z metryką* $-(l, m)$ *jest trójwymiarową przestrzenią euklidesową, a* $\mathcal{M} = L \oplus L^\perp$.

Dowód. Równość $\mathcal{M} = L \oplus L^\perp$ wynika ze stwierdzenia § 3, p. 2 wobec tego, że prosta czasowa jest oczywiście przestrzenią niezdegenerowaną. Ponieważ sygnatura metryki Minkowskiego na $\mathcal{M}$ jest równa $(1,3)$, więc na $L^\perp$ jest równa $(0,3)$, co kończy dowód.

Przejdźmy teraz do zbadania geometrycznego sensu iloczynu skalarnego. Nieokreśloność metryki Minkowskiego jest powodem istotnych różnic w porównaniu z sytuacją euklidesową, które mają ważne znaczenie fizyczne. Najbardziej jaskrawe fakty są związane z tym, że nierówność Cauchy'ego–Buniakowskiego–Schwarza dla wektorów czasowych jest „odwrócona w drugą stronę".

5. Stwierdzenie. *Niech* $(l_1, l_1) > 0$, $(l_2, l_2) > 0$, $l_i \in \mathcal{M}$. *Wówczas*

$$(l_1, l_2)^2 \geq (l_1, l_1)(l_2, l_2).$$

Równość zachodzi wtedy i tylko wtedy, gdy wektory l_1, l_2 *są liniowo zależne.*

Dowód. Przede wszystkim pokażemy, że trójmian kwadratowy $(tl_1+l_2,\ tl_1+l_2)$ ma zawsze pierwiastek rzeczywisty t_0. W realizacji macierzowej $\mathcal{M}$ warunek $(l_2,\ l_2) > 0$ oznacza, że $\det l_2 > 0$, tj. że l_2 ma rzeczywiste pierwiastki charakterystyczne z jednakowym znakiem, powiedzmy ε_2 ($+1$ albo -1). Analogicznie oznaczymy przez ε_1 wspólny znak pierwiastków charakterystycznych l_1. Wtedy dla $t \to -(\varepsilon_1\varepsilon_2)\infty$ macierz tl_1+t_2 ma wartości własne w przybliżeniu proporcjonalne do wartości własnych l_1 (bo $l_1 + t^{-1}l_2$ dąży do l_1), których znak jest równy $-\varepsilon_2$, a przy $t = 0$ wartości własne macierzy $0l_1+l_2 = l_2$ mają znak ε_2. Zatem gdy t przebiega od 0 do $-(\varepsilon_1\varepsilon_2)\infty$, wartości własne $tl_1 + l_2$ przechodzą przez zero, więc $\det(tl_1 + l_2)$ przyjmuje wartość zero. A to oznacza, że wyróżnik tego trójmianu jest nieujemny, czyli

$$(l_1,\ l_2)^2 \geq (l_1,\ l_1)(l_2,\ l_2).$$

Jeśli wyróżnik jest zerem, to trójmian ma pierwiastek dwukrotny, powiedzmy $t_0 \in \boldsymbol{R}$, a więc macierz $t_0l_1+l_2$, będąc macierzą hermitowską ze wszystkimi wartościami własnymi równymi zeru, jest macierzą zerową (bo jest diagonalizowalna!), co oznacza, że l_1, l_2 są liniowo zależne.

6. Wniosek („nierówność trójkąta w odwrotną stronę"). *Jeśli l_1, l_2 są czasowe i $(l_1,\ l_2) \geq 0$, to $l_1 + l_2$ jest też czasowy oraz*

$$|l_1 + l_2| \geq |l_1| + |l_2|$$

(gdzie $|l| = (l,\ l)^{1/2}$), a równość zachodzi wtedy i tylko wtedy, gdy l_1 i l_2 są liniowo zależne.

Dowód. Mamy

$$|l_1 + l_2|^2 = |l_1|^2 + 2(l_1,\ l_2) + |l_2|^2 \geq |l_1|^2 + 2|l_1||l_2| + |l_2|^2 = (|l_1| + |l_2|)^2.$$

Równość zachodzi tylko w przypadku, gdy $(l_1,\ l_2) = |l_1||l_2|$.

Przydyskutujemy teraz interpretację fizyczną tych faktów.

7. „Paradoks bliźniąt". Wektory czasowe l_1, l_2, dla których $(l_1,\ l_2) \geq 0$, będziemy nazywać *jednakowo zorientowanymi w czasie*. Ze stwierdzenia p. 5 wynika, że spełniają one $(l_1,\ l_2) > 0$. Rozważmy dwóch obserwatorów — braci bliźniaków: jeden jest obserwatorem inercjalnym i porusza się po swojej linii świata od punktu 0 do punktu $l_1 + l_2$, podczas gdy drugi dociera do tego samego punktu, wychodząc też z punktu 0, ale poruszając się inercjalnie najpierw do l_1, a następnie od l_1 do $l_1 + l_2$, natomiast w pobliżu zera i l_1 musi posłużyć się silnikami swojego pojazdu kosmicznego, aby najpierw oddalić się od brata, a potem do niego powrócić. Zgodnie z wnioskiem p. 6 czas własny podróżującego bliźniaka będzie istotnie krótszy od czasu, jaki zmierzy na swoim zegarze brat — domator.

8. Czynnik Lorentza. Jeśli l_1, l_2 są czasowe i jednakowo zorientowane w czasie, to na mocy stwierdzenia p. 5 $\dfrac{(l_1,\ l_2)}{|l_1||l_2|} \geq 1$ i nie można tej wielkości interpretować

jako kosinus kąta. Żeby wyjaśnić jej znaczenie, znowu poszukamy odpowiedzi w jej interpretacji fizycznej.

Załóżmy, że $|l_1| = |l_2| = 1$; w szczególności oznacza to, iż obserwator inercjalny l_1 przeżył jednostkę czasu własnego od chwili rozpoczęcia pomiaru czasu. W punkcie l_1 przestrzenią fizyczną jednoczesnych zdarzeń będzie dla niego $l_1 + (\boldsymbol{R}l_1)^\perp$. Linia świata obserwatora $\boldsymbol{R}l_2$ przecina wspomnianą przestrzeń w punkcie xl_2, gdzie x można wyznaczyć z warunku

$$(xl_2 - l_1,\, l_1) = 0,$$

co daje $x = (l_1,\, l_2)^{-1}$. Interwał od l_1 do xl_2 ma typ przestrzenny — dla obserwatora $\boldsymbol{R}l_1$ jest to odległość, na jaką $\boldsymbol{R}l_2$ oddalił się od niego w jednostce czasu, czyli względna prędkość $\boldsymbol{R}l_2$. Wynosi ona (trzeba uwzględnić, że dla metryki w $(\boldsymbol{R}l_1)^\perp$ należy zmienić znak!)

$$\begin{aligned} v &= [-(xl_2 - l_1,\, xl_2 - l_1)]^{1/2} = \\ &= [-(xl_2 - l_1,\, xl_2)]^{1/2} = [-x^2(l_2,\, l_2) + x(l_1,\, l_2)]^{1/2} = [-(l_1,\, l_2)^{-2} + 1]^{1/2}, \end{aligned}$$

skąd wynika

$$(l_1,\, l_2) = \frac{1}{\sqrt{1 - v^2}}.$$

To jest sławny *współczynnik Lorentza*, który jest także często zapisywany w postaci $\dfrac{1}{\sqrt{1 - v^2/c^2}}$, jawnie wskazującej na to, że prędkości mierzymy w odniesieniu do prędkości światła. W szczególności

$$x = \frac{1}{(l_1,\, l_2)} = \sqrt{1 - v^2},$$

tj. w momencie odmierzenia jednej jednostki czasu własnego przez pierwszego obserwatora drugi z obserwatorów znajdzie się w jego fizycznej przestrzeni, natomiast zegar drugiego pokaże tylko upływ $\sqrt{1 - v^2}$ jednostek. Jest to ilościowy opis tego efektu „zwolnienia zegarów" dla poruszającego się obserwatora, który przedyskutowaliśmy jakościowo w poprzednim punkcie.

9. Kąty euklidesowe. W przestrzeni $(\boldsymbol{R}l_0)^\perp$, gdzie l_0 jest wektorem czasowym, obowiązuje geometria euklidesowa i iloczyn skalarny ma tam zwykłe znaczenie. Wprowadźmy jeszcze dwa wektory czasowe l_1, l_2 o tej samej orientacji. Można skonstruować rzuty tych wektorów na $(\boldsymbol{R}l_0)^\perp$ i obliczyć kosinus kąta między nimi. Pozostawiamy czytelnikowi samodzielne przekonanie się o tym, że dla obserwatora $\boldsymbol{R}l_0$ ten kąt to kąt między kierunkami, w których w jego przestrzeni fizycznej oddalają się od niego obserwatorzy $\boldsymbol{R}l_1$ i $\boldsymbol{R}l_2$. Nie można temu kątowi nadawać absolutnego znaczenia — inny obserwator $\boldsymbol{R}l_0'$ wyznaczy inną jego wartość.

10. Cztery orientacje przestrzeni Minkowskiego. Niech $\{e_i\}$, $\{e'_i\}$, gdzie $i = 0, \ldots, 3$, będą dwiema bazami ortonormalnymi w $\mathcal{M}$: $(e_0, e_0) = (e'_0, e'_0) = 1$, $(e_i, e_i) = (e'_i, e'_i) = -1$ dla $i = 1, \ldots, 3$. Przez analogię do poprzednich definicji nazwiemy je *bazami jednakowo zorientowanymi*, gdy jedną z nich można przeprowadzić na drugą za pomocą ciągłej rodziny izometrii $f_t: \mathcal{M} \to \mathcal{M}$, $0 \leq t \leq 1$, $f_0 = \mathrm{id}$, $f_1(e_i) = e'_i$. Poniższe dwa warunki jednakowego zorientowania baz są w oczywisty sposób konieczne.

a) $(e_0, e'_0) > 0$.

Rzeczywiście, $(e_0, f_t(e_0))^2 \geq 1$ na mocy p. 5, a więc znak $(e_0, f_t(e_0))$ nie może się zmieniać przy zmianie t, a $(e_0, f_0(e_0)) = 1$. Wcześniej w takiej sytuacji mówiliśmy, że wektory e_0, e'_0 mają jednakową orientację czasową.

b) Wyznacznik odwzorowania rzutu ortogonalnego $\sum_{i=1}^{3} \boldsymbol{R}e_i \to \sum_{i=1}^{3} \boldsymbol{R}e'_i$, zapisanego w bazach $\{e_i\}$, $\{e'_i\}$, jest dodatni.

Rzeczywiście, rzut $\sum_{i=1}^{3} \boldsymbol{R}e_i \to \sum_{i=1}^{3} \boldsymbol{R}f_t(e_i)$ jest odwzorowaniem odwracalnym przy wszystkich wartościach t, bo w przeciwnym przypadku wektor przestrzenny należący do $\sum_{i=1}^{3} \boldsymbol{R}e_i$ byłby ortogonalny do $\sum_{i=1}^{3} \boldsymbol{R}f_t(e_i) = (f_t(e_0))^{\perp}$, więc proporcjonalny do wektora czasowego $f_t(e_0)$, co jest niemożliwe. Stąd wynika, że wyznaczniki tych rzutów mają dla wszystkich t jednakowe znaki, a dla $t = 0$ wyznacznik jest dodatni.

Pary baz mające własność b) można by nazwać bazami *jednakowo zorientowanymi przestrzennie*.

Odwrotnie, jeśli dwie bazy ortonormalne w $\mathcal{M}$ mają jednakowe czasowe i przestrzenne orientacje, to są jednakowo zorientowane, tj. można jedną z nich przeprowadzić na drugą za pomocą ciągłej rodziny izometrii f_t. Dla skonstruowania tych izometrii, określimy najpierw $f_t(e_0) = \dfrac{te'_0 + (1-t)e_0}{|te' + (1-t)e|}$. Z warunku $(e_0, e'_0) \geq 1$ wynika, że wektor $f_t(e_0)$ jest czasowy, a kwadrat jego długości jest równy jeden dla wszystkich $0 \leq t \leq 1$. Dalej jako $f_t(e_1, e_2, e_3)$ wybierzemy bazę ortonormalną $f_t(e_0)^{\perp}$, otrzymaną przez ortogonalizację Grama–Schmidta rzutu $\{e_1, e_2, e_3\}$ na $f_t(e_0)^{\perp}$. Oczywiście elementy tak skonstruowanej bazy zależą w sposób ciągły od t. Jest jasne, że $f_1(e_0) = e'_0$, a $\{f_1(e_1), f_1(e_2), f_1(e_3)\}$ i $\{e'_1, e'_2, e'_3\}$ są jednakowo zorientowanymi bazami ortonormalnymi w $(e'_0)^{\perp}$. Można przeprowadzić jedną na drugą za pomocą ciągłej rodziny euklidesowych obrotów w przestrzeni $(e'_0)^{\perp}$ zachowujących wektor e'_0. To kończy dowód.

Oznaczymy przez Λ *grupę Lorentza*, tj. grupę izometrii przestrzeni $\mathcal{M}$, lub inaczej grupę $\boldsymbol{O}(1,3)$. Dalej niech $\Lambda^{\uparrow}_{+}$ oznacza podgrupę Λ złożoną z izometrii zachowujących orientację pewnej bazy ortonormalnej, $\Lambda^{\uparrow}_{-}$ — podzbiór Λ odwzorowań zmieniających orientację przestrzenną, lecz nie zmieniających orientacji czasowej tej bazy, $\Lambda^{\downarrow}_{+}$ — podzbiór Λ odwzorowań zmieniających orientację czasową, lecz nie przestrzenną tej bazy, $\Lambda^{\downarrow}_{-}$ — podzbiór Λ odwzorowań zmieniających orientację przestrzenną i czasową wybranej bazy. Nietrudno przekonać się, że te podzbiory

są niezależne od wyboru wyjściowej bazy. W ten sposób wykazaliśmy następujące twierdzenie.

11. Twierdzenie. *Grupa Lorentza jest sumą czterech swoich składowych spójności, tj.*

$$\Lambda = \Lambda_+^\uparrow \cup \Lambda_-^\uparrow \cup \Lambda_+^\downarrow \cup \Lambda_-^\downarrow.$$

Odwzorowanie identycznościowe jest naturalnie elementem $\Lambda_+^\uparrow$. Następujący wynik jest odpowiednikiem twierdzenia § 11, p. 12.

12. Twierdzenie. *Zrealizujmy $\mathcal{M}$ jako przestrzeń macierzy Grama metryk hermitowskich przestrzeni $\mathcal{H}$ względem bazy $\{h_1, h_2\}$. Każdej macierzy $V \in \boldsymbol{SL}(2, \boldsymbol{C})$ odpowiada odwzorowanie $\mathcal{M}$ w siebie, które przyporządkowuje macierzy $l \in \mathcal{M}$ macierz $s(V)l$ określoną wzorem*

$$s(V)l = V^t l \overline{V}.$$

Tak zdefiniowane odwzorowanie s jest surjektywnym homomorfizmem $\boldsymbol{SL}(2, \boldsymbol{C})$ na $\Lambda_+^\uparrow$, a jego jądrem jest $\{\pm E_2\}$.

Dowód. Jest oczywiste, że $s(V)l$ jest liniowe względem l i zachowuje kwadraty odległości — $\det(V^t l V) = \det l$. Dlatego $s(V)l \in \Lambda$. Ponieważ grupa $\boldsymbol{SL}(2, \boldsymbol{C})$ jest spójna, dowolny jej element można przeprowadzić w sposób ciągły w jedynkę grupy, pozostając cały czas w obrębie $\boldsymbol{SL}(2, \boldsymbol{C})$ — przekształcenie Lorentza $s(V)$ można w sposób ciągły przeprowadzić w przekształcenie tożsamościowe, tak że $s(V) \in \Lambda_+^\uparrow$. Ponieważ $s(\mathrm{id}) = \mathrm{id}$ i $s(V_1 V_2) = s(V_1)s(V_2)$, więc s jest homomorfizmem grup. Jeśli $V^t l \overline{V} = l$ dla każdego $l \in \mathcal{M}$, to w szczególności $V^t \sigma_i \overline{V} = \sigma_i$, gdzie $\sigma_0 = E_2$, a $\sigma_1, \sigma_2, \sigma_3$ oznaczają macierze Pauliego. Z warunku $V^t \overline{V} = E_2$ wynika unitarność V, więc z kolei z warunków $V^t \sigma_i \overline{V} = V^t \sigma_i (V^t)^{-1} = \sigma_i$ wynika, jak to wykazaliśmy w p. 12, § 11, że $V = \pm E_2$. W ten sposób sprawdziliśmy, że $\operatorname{Ker} s = \{\pm E_2\}$.

Pozostaje do wykazania surjektywność s. Niech $f\colon \mathcal{M} \to \mathcal{M}$ oznacza przekształcenie Lorentza należące do $\Lambda_+^\uparrow$ i przprowadzające bazę ortonormalną $\{e_i\}$ na $\{e_i'\}$. Odpowiadające e_0 i e_0' metryki na $\mathcal{H}$ są określone, bo znaki obydwu wartości własnych zarówno e_0, jak i e_0' są jednakowe, wobec $\det e_0 = \det e_0' = 1$. Dalej, z $(e_0, e_0') > 0$ wynika, że te metryki są jednocześnie obie dodatnio określone lub obie ujemnie określone. Rzeczywiście, widzieliśmy powyżej, że łączący je odcinek $te_0' + (1-t)e_0$, $0 \leq t \leq 1$, zawiera tylko wektory czasowe. Stąd już wynika istnienie takiej macierzy $V \in \boldsymbol{SL}(2, \boldsymbol{C})$, że $s(V)$ przeprowadza e_0 w e_0', tj. $e_0' = V^t e_0 \overline{V}$, gdzie utożsamiliśmy e_0 i e_0' z ich macierzami Grama. Jako V możemy bowiem obrać macierz izometrii $(\mathcal{H}, e_0)$ z $(\mathcal{H}, e_0')$ — *a priori* jej wyznacznik mógłby równać się -1, ale to byłoby sprzeczne z możliwością połączenia V z E_2 w $\boldsymbol{SL}(2, \boldsymbol{C})$ krzywą ciągłą V_q, gdzie $e_0' = (V_q)^t f_q(e_0) \overline{V_q}$, a f_q oznacza odpowiadającą krzywą w $\Lambda_+^\uparrow$.

Tak więc $s(V)$ przeprowadza e_0 na e_0'. Pozostaje do pokazania, że obrót euklidesowy przeprowadzający $\{s(V)e_1, s(V)e_2, s(V)e_3\}$ na $\{e_1', e_2', e_3'\}$ można otrzymać za pomocą takiego $s(U)$, gdzie $U \in \boldsymbol{SL}(2, \boldsymbol{C})$, które zachowuje e_0. Możemy przyjąć, że e_0' ma macierz σ_0 w bazie $\{h_1, h_2\}$. Wtedy wystarcza znaleźć macierz unitarną U

spełniającą $U(s(V)e_i)U^{-1} = e'_i$ dla $i = 1,2,3$. Taka macierz istnieje na mocy twierdzenia p. 12, § 11, bo obie bazy $\{s(V)e_i\}$ i $\{e'_i\}$, $i = 1,2,3$, w $(e'_0)^\perp$ są ortonormalne i jednakowo zorientowane. Dowód zakończony.

13. Obroty euklidesowe i boosty. Niech e_0, e'_0 będą dwoma jednakowo zorientowanymi wektorami czasowymi o długości jeden, L_0, L'_0 — ich ortogonalnymi dopełnieniami. Istnieją standardowe przekształcenia Lorentza, przeprowadzające e_0 w e'_0, które w literaturze fizycznej są nazywane boostami. Dla $e_0 = e'_0$ takie przekształcenie jest tożsamością. Przy $e_0 \neq e'_0$ są one określone następująco. Rozważmy płaszczyznę $(L_0 \cap L'_0)^\perp$, która zawiera oba wektory e_0 i e'_0. Sygnatura metryki Minkowskiego obciętej do niej jest równa $(1,1)$. Istnieje więc para jednostkowych przestrzennych wektorów $e_1, e'_1 \in (L_0 \cap L'_0)^\perp$ ortogonalnych do e_0 i e'_0 odpowiednio. Szukane przekształcenie jest tożsamością na $L_0 \cap L'_0$ i przeprowadza e_0 na e_1 i e'_0 na e'_1. Dla obliczenia wyrazów macierzy przejścia $\{e_0, e_1\}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \{e'_0, e'_1\}$ zauważymy najpierw, że $a = (e, e'_0) = \dfrac{1}{\sqrt{1-v^2}}$, gdzie v jest prędkością względnego oddalania się od siebie obserwatorów odpowiadających e_0 i e'_0 odpowiednio. Co więcej, macierze Grama $\{e_0, e_1\}$ i $\{e'_0, e'_1\}$ są równe $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, a stąd otrzymujemy

$$a^2 - b^2 = 1, \quad ac - bd = 0, \quad c^2 - d^2 = -1.$$

Znając a, z pierwszego równania wyznaczamy $b = \dfrac{v}{\sqrt{1-v^2}}$, a następnie biorąc pod uwagę, że wyznacznik tego przekształcenia jest równy jeden, otrzymujemy $d = a$, $c = b$. Ostatecznie otrzymujemy macierz postaci

$$\begin{pmatrix} \dfrac{1}{\sqrt{1-v^2}} & \dfrac{v}{\sqrt{1-v^2}} & 0 & 0 \\ \dfrac{v}{\sqrt{1-v^2}} & \dfrac{1}{\sqrt{1-v^2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

jako macierz boostu w bazie $\{e_0, e_1, e_2, e_3\}$, gdzie $\{e_1, e_2\}$ oznacza ortonormalną bazę $L_0 \cap L'_0$. We współrzędnych czasoprzestrzennych mamy

$$x_0 = \frac{x'_0 + vx'_1}{\sqrt{1-v^2}}, \quad x_1 = \frac{vx'_0 + x'_1}{\sqrt{1-v^2}}, \quad x = x'_2 \quad x_3 = x'_3.$$

Podmacierz stojącą w lewym górnym rogu można zapisać jeszcze jako macierz „obrotu hiperbolicznego"

$$\begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix},$$

wyznaczając θ z warunków

$$\cosh\theta = \frac{e^{\theta}+e^{-\theta}}{2} = \frac{1}{\sqrt{1-v^2}}, \quad \sinh\theta = \frac{e^{\theta}+e^{-\theta}}{2} = \frac{v}{\sqrt{1-v^2}}.$$

Dla danych dwóch jednakowo zorientowanych baz ortonormalnych $\{e_0, e_1, e_2, e_3\}$ i $\{e'_0, e'_1, e'_2, e'_3\}$ przekształcenie Lorentza przeprowadzające jedną na drugą można przedstawić w postaci złożenia boostu odwzorowującego e_0 na e'_0 z euklidesowym obrotem w $(e'_0)^{\perp}$, który przeprowadza na bazę $\{e'_1, e'_2, e'_3\}$ obraz bazy $\{e_1, e_2, e_3\}$, pod działaniem boostu i nie zmienia e'_0.

14. Odbicia przestrzenne i czasowe. Dowolna trójwymiarowa podprzestrzeń $L \subset \mathcal{M}$, na której metryka Minkowskiego jest (anty)euklidesowa (tj. prosta $L^{\perp}$ jest czasowa) zadaje przekształcenie Lorentza, identycznościowe na L i zmieniające znak na $L^{\perp}$. Każdy taki operator nazywa się *odbiciem czasu.*

Dowolna trójwymiarowa podprzestrzeń $L \subset \mathcal{M}$, na której sygnatura metryki Minkowskiego jest $(1,2)$ (tj. prosta $L^{\perp}$ jest przestrzenna) także określa przekształcenie Lorentza, identycznościowe na L i zmieniające znak na $L^{\perp}$. Każde przekształcenie tego rodzaju nazywamy *odbiciem przestrzennym.*

Jeśli ustalić jakieś odbicie czasowe T i odbicie przestrzenne P, to wszystkie elementy ze zbiorów $\Lambda_{\downarrow}^{\downarrow}$, $\Lambda_{\downarrow}^{\uparrow}$, $\Lambda_{\uparrow}^{\downarrow}$ można otrzymać, mnożąc elementy z $\Lambda_{+}^{\uparrow}$ przez T, P i PT, odpowiednio.

§ 13. Przestrzenie symplektyczne

1. W tym paragrafie będziemy rozważać skończenie wymiarowe przestrzenie liniowe L nad ciałem $\boldsymbol{K}$ charakterystyki $\neq 2$, wyposażone w niezdegenerowany, antysymetryczny iloczyn skalarny $[\,,\,]: L \times L \to \boldsymbol{K}$. Nazywać je będziemy *przestrzeniami symplektycznymi.* Przypomnijmy własności przestrzeni symplektycznych, które były już wykazane w § 3.

Wymiar przestrzeni symplektycznej jest liczbą parzystą. Jeśli wynosi on $2r$, to w przestrzeni istnieje baza symplektyczna $\{e_1, \ldots, e_r; e_{r+1}, \ldots, e_{2r}\}$, tj. baza z macierzą Grama postaci

$$\begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix}.$$

W szczególności, każde dwie przestrzenie symplektyczne tego samego wymiaru nad tym samym ciałem są izomorficzne.

Podprzestrzeń $L_1 \subset L$ nazywamy *osobliwą*, jeśli obcięcie iloczynu skalarnego $[\,,\,]$ do niej jest identycznościowo równe zeru.([1]) Każda podprzestrzeń jednowymiarowa jest osobliwa.

([1]) W przypadku symplektycznym bardziej rozpowszechnione są terminy: podprzestrzeń zerowa lub podprzestrzeń izotropowa (*przyp. tłum.*).

2. Stwierdzenie. *Niech L będzie przęstrzenią symplektyczną o wymiarze $2r$, $L_1 \subset L$ — podprzestrzenią osobliwą o wymiarze r_1. Wtedy $r_1 \leq r$ i jeśli $r_1 < r$, to L_1 jest zawarta w podprzestrzeni osobliwej o maksymalnym możliwym wymiarze r.*

Dowód. Z założenia forma $[\ ,\ ]$ jest niezdegenerowana, więc zadaje ona izomorfizm $L \to L^*$, który wektorowi $l \in L$ przyporządkowuje funkcjonał liniowy $l' \mapsto [l, l']$. Stąd wynika, że dla dowolnej podprzestrzeni $L_1 \subset L$ mamy $\dim L^\perp = \dim L - \dim L_1$ (por. § 7, część 1). Jeśli ponadto przestrzeń L_1 jest osobliwa, to $L_1 \subset L^\perp$, skąd $r_1 = \dim L_1 \leq \dim L^\perp = \dim L - \dim L_1 = 2r - r_1$, a więc $r_1 \leq r$. Rozważmy teraz obcięcie formy $[\ ,\ ]$ do $L_1^\perp$. Ortogonalne dopełnienie $L_1^\perp$ względem całej przestrzeni L ma zgodnie z powyższym akapitem wymiar $\dim L - \dim L^\perp = \dim L_1$. Ponieważ L_1 jest także zawarta w tym dopełnieniu, musi być mu równa. Z tego wynika, że L_1 jest dokładnie jądrem obcięcia $[\ ,\ ]$ do $L_1^\perp$. Przestrzeń $L_1^\perp$ ma bazę symplektyczną

$$\{e_1, \ldots, e_{r-r_1}; e_{r-r_1+1}, \ldots, e_{2(r-r_1)}; e_{2(r-r_1)+1}, \ldots, e_{2r-r_1}$$

w tym znaczeniu, jakie było stosowane w § 3, gdzie rozważaliśmy także przypadek przestrzeni zdegenerowanej. Macierzą Grama tej bazy jest

$$\left(\begin{array}{c|c|c} 0 & E_r - r_1 & 0 \\ \hline -E_r - r_1 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

Wymiar klatki zawierającej macierz jednostkową jest równy $\frac{1}{2}(\dim L_1^\perp - \dim L_1) = r - r_1$. Wektory $e_{2(r-r_1)+1}, \ldots, e_{2r-r_1}$ rozpinają jądro formy na $L_1^\perp$, tj. przestrzeń L_1, a dołączając do nich np. wektory $e_1, \ldots, e_{r-r_1}$, otrzymamy r-wymiarową podprzestrzeń osobliwą zawierającą L_1.

3. Stwierdzenie. *Niech L będzie przestrzenią symplektyczną wymiaru $2r$, $L_1 \subset L$ — r-wymiarową podprzestrzenią osobliwą. Wówczas istnieje taka r-wymiarowa podprzestrzeń osobliwa $L_2 \subset L$, że $L = L_1 \oplus L_2$, a iloczyn skalarny indukuje izomorfizm $L_2 \to L_1^*$.*

Dowód. Wykażemy nieco silniejsze stwierdzenie, ważne ze względu na zastosowania, a mianowicie pokażemy, że taką podprzestrzeń można wybrać spośród skończonej liczby podprzestrzeni osobliwych związanych z ustaloną bazą symplektyczną $\{e_1, \ldots, e_r; e_{r+1}, \ldots, e_{2r}\}$ w przestrzeni L.

Rzeczywiście, rozważmy podział zbioru $\{1, \ldots, r\} = I \cup J$ na dwa rozłączne podzbiory. Wtedy podprzestrzeń rozpięta przez r wektorów $\{e_i, e_{r+j} \mid i \in I, j \in J\}$ jest r-wymiarową podprzestrzenią osobliwą w L, nazywaną *podprzestrzenią współrzędnościową* (względem wybranej bazy). Oczywiście w ten sposób można zbudować 2^r podprzestrzeni. Wykażemy, że jako L_2 można wybrać jedną z nich.

Oznaczmy przez M podprzestrzeń rozpiętą przez wektory $\{e_1, \ldots, e_r\}$ i niech $\dim(L_1 \cap M) = s$, $0 \leq s \leq r$. Można znaleźć taki podzbiór $I \subset \{1, \ldots, r\}$ złożony z $r - s$ elementów, że $L_1 \cap M$ jest transwersalna do podprzestrzeni N rozpiętej przez $\{e_i \mid i \in I\}$, tj. $L_1 \cap M \cap N = \{0\}$. Istotnie, ponieważ zbiór $\{\text{baza } L_1 \cap M\} \cup \{e_1, \ldots, e_r\}$ rozpina M, więc bazę $L_1 \cap M$ można uzupełnić do bazy w M poprzez dołączenie $r - s$ wektorów wybranych spośród $\{e_1, \ldots, e_r\}$, na mocy stwierdzenia p. 10, § 2, części 1. Wskaźniki odpowiadające tym wektorom tworzą szukany zbiór I, bowiem z $L_1 \cap M + N = M$ wynika $L_1 \cap M \cap N = \{0\}$.

Przyjmijmy teraz $J = \{1, \ldots, r\} \setminus I$; pokażemy, że podprzestrzeń osobliwa L_2 rozpięta przez $\{e_i, e_{r+j} \mid i \in I, j \in J\}$ jest przestrzenią dopełniającą do L_1. Wystarczy sprawdzić, że $L_1 \cap L_2 = \{0\}$. Rzeczywiście, z dowodu stwierdzenia p. 2 wynika, że $L_1^{\perp} = L_1$, $L_2^{\perp} = L_2$. Ponieważ $L_1 \cap M$ jest zawarte w L_1 i N jest zawarte w L_2, zatem suma $M = L_1 \cap M + N$ jest ortogonalna do $L_1 \cap L_2$. Ponieważ M jest podprzestrzenią osobliwą o wymiarze r, więc $M^{\perp} = M$ i $L_1 \cap L_2 \subset M$. Ostatecznie,

$$L_1 \cap L_2 = (L_1 \cap M) \cap (L_2 \cap M) = (L_1 \cap M) \cap N = \{0\}.$$

Odwzorowanie liniowe $L_2 \to L_1^*$ przyporządkowuje wektorowi $l \in L_2$ formę liniową $m \mapsto [l, m]$ na L_1. Jest to izomorfizm, co wynika z równości $\dim L_2 = \dim L^* = r$ i tego, że jego jądro jest zawarte w jądrze formy $[\ ,\ ]$, która z założenia jest niezdegenerowana. W ten sposób dowód został zakończony.

4. Wniosek. *Każde dwie pary wzajemnie dopełniających podprzestrzeni osobliwych przestrzeni L są w niej jednakowe położone: jeśli $L = L_1 \oplus L_2 = L_1' \oplus L_2'$, to istnieje izometria $f: L \to L$, taka że $f(L_1) = L_1'$ i $f(L_2) = L_2'$.*

Dowód. Wybierzmy bazę $\{e_1, \ldots, e_r\}$ w L_1 i dwoistą bazę $\{e_{r+1}, \ldots, e_{2r}\}$ w L_2 względem opisanego wyżej utożsamienia L_1^* z L_2. Jest oczywiste, że $\{e_1, \ldots, e_{2r}\}$ jest bazą symplektyczną w L. Utwórzmy w analogiczny sposób bazę symplektyczną $\{e_1', \ldots, e_{2r}'\}$, używając rozkładu $L = L_1' \oplus L_2'$. Wtedy odwzorowanie liniowe zdefiniowane przez $f: e_i \mapsto e_i'$, $i = 1, \ldots, 2r$, jest szukanym izomorfizmem.

Z tego wniosku i stwierdzeń p. p. 2, 3 wynika, że każde dwie podprzestrzenie osobliwe w L o tym samym wymiarze można przeprowadzić na siebie za pomocą odpowiedniej izometrii.

5. Grupa symplektyczna. Zbiór wszystkich izometrii $f: L \to L$ przestrzeni symplektycznej tworzy grupę. Zbiór macierzy reprezentujących elementy tej grupy w bazie symplektycznej $\{e_1, \ldots, e_{2r}\}$ nazywa się *grupą symplektyczną*, a przyjmując $2r = \dim L$, oznacza się ją przez $\boldsymbol{Sp}(2r, \boldsymbol{K})$. Warunek $A \in \boldsymbol{Sp}(2r, \boldsymbol{K})$ jest równoważny z tym, by macierz Grama bazy $\{e_1, \ldots, e_{2r}\}A$ była równa $I_{2r} = \begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix}$, tj. by $A^t I_{2r} A = I_{2r}$, skąd w szczególności wynika $\det A = \pm 1$. Niżej wykażemy, że w istocie $\det A = 1$ (por. p. 11). Ponieważ $I_{2r}^2 = -E_{2r}$, warunek powyższy można także zapisać w postaci $A = -I_{2r}(A^t)^{-1} I_{2r}$. Stąd wynika

6. Stwierdzenie. *Wielomian charakterystyczny* $P(t) = \det(tE_{2r} - A)$ *macierzy symplektycznej* A *jest zwrotny, tj.* $P(t) = t^{2r}P(t^{-1})$.

Dowód. Korzystając z tego, że $\det A = 1$, otrzymujemy

$$\det(tE_{2r} - A) = \det(tE_{2r} + I_{2r}(A^t)^{-1}I_{2r}) = \det(tE_{2r} - (A^t)^{-1}) =$$
$$= \det(tA^t - E_{2r}) = t^{2r}\det(t^{-1}E_{2r} - A^t) = t^{2r}\det(t^{-1}E_{2r} - A).$$

7. Wniosek. *Jeśli* $\boldsymbol{K} = \boldsymbol{R}$ *i* A *jest macierzą symplektyczną, to dla każdej wartości własnej* λ *macierzy* A *także* λ^{-1}, $\overline{\lambda}$, λ^{-1} *są wartościami własnymi* A.

Dowód. Ponieważ A jest nieosobliwa, więc $\lambda \neq 0$ i $P(\lambda^{-1}) = \lambda^{-2r}P(\lambda) = 0$, a ponieważ P ma współczynniki rzeczywiste, więc także $P(\overline{\lambda}) = P(\lambda) = 0$.

Sprzężenie zespolone jest symetrią względem osi rzeczywistej, a odwzorowanie $\lambda \mapsto \overline{\lambda}^{-1}$ jest symetrią względem okręgu jednostkowego. Stąd wynika, że zespolone wartości własne A występują czwórkami punktów symetrycznych względem osi rzeczywistej i okręgu jednostkowego, natomiast rzeczywiste wartości własne — parami.

8. Pfaffian. Niech $\boldsymbol{K}^{2r}$ będzie przestrzenią kartezjańską, A — niezdegenerowaną antysymetryczną macierzą rzędu $2r$ nad $\boldsymbol{K}$. Iloczyn skalarny $[\vec{x}, \vec{y}] = \vec{x}^{\,t}A\vec{y}$ w przestrzeni $\boldsymbol{K}^{2r}$ jest niezdegenerowany i antysymetryczny. Wobec możliwości przejścia od bazy wyjściowej do bazy symplektycznej w $\boldsymbol{K}^{2r}$ istnieje taka macierz nieosobliwa B, że

$$A = B^t \begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix} B,$$

skąd wnioskujemy, że $\det A = (\det B)^2$. Widzimy więc, że wyznacznik każdej macierzy antysymetrycznej jest pełnym kwadratem. To sugeruje, aby spróbować wyznaczyć pierwiastek tego wyznacznika, który powinien być *uniwersalnym wielomianem* wyrazów macierzy A. Taką konstrukcją można rzeczywiście wykonać.

9. Twierdzenie. *Istnieje jedyny wielomian względem wyrazów antysymetrycznej macierzy* A *o współczynnikach całkowitych,* $\operatorname{Pf} A$, *taki że* $\det A = (\operatorname{Pf} A)^2$ *i* $\operatorname{Pf}\begin{pmatrix} 0 & E_r \\ -E_r & 0 \end{pmatrix} = 1$. *Wielomian ten, nazywany pfaffianem, spełnia równość*

$$\operatorname{Pf}(B^tAB) = \det B \cdot \operatorname{Pf} A$$

dla każdej macierzy B. (*W przypadku* $\operatorname{char}\boldsymbol{K} \neq 0$ *współczynniki* Pf *są całkowite w tym sensie, że należą do prostego podciała ciała* $\boldsymbol{K}$, *tj. są sumami jedności.*)

Dowód. Rozważmy $r(2r-1)$ zmiennych niezależnych nad ciałem $\boldsymbol{K}$: $\{a_{ij} \mid 0 \leq i < j \leq 2r\}$ i oznaczmy przez $\mathcal{K}$ ciało funkcji wymiernych (ilorazów funkcji wielomianowych) zmiennych a_{ij} o współczynnikach z prostego podciała ciała $\boldsymbol{K}$. Określimy

macierz $A = (a_{ij})$, przyjmując $a_{ij} = -a_{ji}$ dla $i > j$, $a_{ii} = 0$, a następnie zdefiniujemy w przestrzeni kartezjańskiej $\mathcal{K}^{2r}$ niezdegenerowany antysymetryczny iloczyn skalarny $\vec{x}^t A\vec{y}$. Jeśli, tak jak powyżej, B jest macierzą przejścia do bazy symplektycznej, to otrzymamy jak poprzednio $\det A = (\det B)^2$. *A priori* $\det B$ jest jedynie funkcją wymierną od a_{ij} o współczynnikach z $\boldsymbol{Q}$ lub z podciała prostego skończonej charakterystyki. Ponieważ jednak $\det A$ jest wielomianem o współczynnikach całkowitych, więc jego pierwiastek kwadratowy także musi mieć całkowite współczynniki (tu korzystamy z twierdzenia o jednoznacznym rozkładzie na czynniki w pierścieniu wielomianów $\boldsymbol{Z}[a_{ij}]$ lub $F_p[a_{ij}]$). Znak $\sqrt{\det A}$ jest jednoznacznie wyznaczony przez żądanie, aby $\sqrt{\det I_{2r}}$ był równy jedności.

Ostatniej równości z podanych w tezie twierdzenia dowodzimy w następujący sposób. Przede wszystkim B^tAB jest antysymetryczna tak jak A, skąd wynika

$$\mathrm{Pf}^2(B^tAB) = \det(B^tAB) = (\det B)^2 \det A = (\det B)^2 \mathrm{Pf}^2 A.$$

Zatem

$$\mathrm{Pf}(B^tAB) = \pm \det B \,\mathrm{Pf}\, A.$$

Dla ustalenia znaku wystarczy wyznaczyć go w przypadku $A = I_{2r}$, $B = E_{2r}$, gdzie równość jest oczywiście spełniona ze znakiem plus.

10. Przykłady.

$$\mathrm{Pf}\begin{pmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{pmatrix} = a_{12},$$

$$\mathrm{Pf}\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix} = -a_{12}a_{34} + a_{13}a_{24} - a_{14}a_{23}.$$

11. Wniosek. *Wyznacznik dowolnej macierzy symplektycznej jest równy jedności.*

Dowód. Z warunku $A^tI_{2r}A = I_{2r}$ i twierdzenia p. 9 wynika, że

$$1 = \mathrm{Pf}\, I_{2r} = \mathrm{Pf}(A^tI_{2r}A) = \det A \cdot \mathrm{Pf}\, I_{2r}$$

Ten wniosek wykorzystaliśmy w dowodzie stwierdzenia p. 6.

12. Zależności między grupą ortogonalną, grupą unitarną a grupą symplektyczną. Niech $\boldsymbol{R}^{2n}$ będzie przestrzenią kartezjańską wyposażoną w dwa iloczyny skalarne — euklidesowy $(\ ,\ )$ i symplektyczny $[\ ,\ ]$:

$$(\vec{x}, \vec{y}) = \vec{x}^t\vec{y},$$

$$[\vec{x}, \vec{y}] = \vec{x}^t I_{2r}\vec{y} = (\vec{x}, I_{2r}\vec{y}).$$

Ponieważ $I_{2r}^2 = -E_{2r}$, więc operator I_{2r} zadaje na $\boldsymbol{R}^{2r}$ strukturę zespoloną (por. § 12, część 1) z bazą zespoloną $\{e_j + ie_{r+j} \mid j = 1, \ldots, r\}$, która umożliwia zdefiniowanie hermitowskiego iloczynu skalarnego za pomocą wzoru

$$\langle \vec{x}, \vec{y} \rangle = (\vec{x}, \vec{y}) - i[\vec{x}, \vec{y}]$$

(por. stwierdzenie § 6, p. 2).

Dla grup zdefiniowanych w terminach tych struktur mamy

$$\boldsymbol{U}(r) = \boldsymbol{O}(2r) \cap \boldsymbol{Sp}(2r) = \boldsymbol{GL}(r, \boldsymbol{C}) \cap \boldsymbol{Sp}(2r) = \boldsymbol{GL}(r, \boldsymbol{C}) \cap \boldsymbol{O}(2r).$$

Wyprowadzenie tych związków pozostawiamy czytelnikowi jako ćwiczenie.

§ 14. Twierdzenie Witta i grupa Witta

1. W tym paragrafie przedstawimy rezultaty Witta dotyczące teorii skończenie wymiarowych przestrzeni ortogonalnych nad dowolnymi ciałami. Precyzują one twierdzenie klasyfikacyjne z § 3 i mogą być uważane za daleko idące uogólnienie twierdzenia o bezwładności i pojęcia sygnatury. Zaczniemy od wprowadzenia pewnych definicji i jak zazwyczaj będziemy przyjmować, że charakterystyka ciała skalarów jest $\neq 2$.

Płaszczyzną hiperboliczną nazywamy dwuwymiarową przestrzeń L z niezdegenerowanym symetrycznym iloczynem skalarnym $(\ ,\ )$, zawierającą niezerowy wektor osobliwy.

Przestrzenią hiperboliczną nazywamy przestrzeń, która ma rozkład na sumę prostą wzajemnie ortogonalnych płaszczyzn hiperbolicznych.

Przestrzenią anizotropową nazywamy przestrzeń nie zawierającą niezerowych wektorów osobliwych.

Przestrzenie anizotropowe nad ciałem liczb liczb rzeczywistych mają sygnaturę $(n, 0)$ albo $(0, n)$, gdzie $n = \dim L$. Pokażemy teraz, że przestrzenie hiperboliczne można uważać za uogólnienie przestrzeni o sygnaturze (m, m).

2. Lemat. *W płaszczyźnie hiperbolicznej zawsze można znaleźć bazę* $\{e_1', e_2'\}$ *z macierzą Grama* $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ *i bazę* $\{e_1, e_2\}$ *z macierzą Grama* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Dowód. Niech $l \in L$, $(l, l) = 0$. Jeśli $l_1 \in L$ nie jest proporcjonalny do l, to $(l_1, l) \neq 0$, bo L jest niezdegenerowana. Możemy przyjąć, że $(l_1, l) = 1$. Zdefiniujmy $e_1 = l$, $e_2 = l_1 - (l_1, l_1)l$. Wówczas $(e_1, e_1) = (e_2, e_2) = 0$, $(e_1, e_2) = 1$. Przyjmijmy dalej $e_1' = \frac{1}{2}(e_1 + e_2)$, $e_2' = \frac{1}{2}(e_1 - e_2)$. Wtedy $(e_1', e_1') = 1$, $(e_2', e_2') = -1$, $(e_1', e_2') = 0$. To dowodzi lematu.

Bazę $\{e_1, e_2\}$ będziemy nazywać *bazą hiperboliczną*. Analogicznie, w przypadku ogólnej przestrzeni hiperbolicznej *bazą hiperboliczną* będziemy nazywać bazę o macierzy Grama złożonej z bloków $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ rozmieszczonych wzdłuż głównej diagonali.

3. Lemat. *Niech $L_0 \subset L$ będzie podprzestrzenią osobliwą niezdegenerowanej przestrzeni ortogonalnej L, a $\{e_1, \ldots, e_m\}$ — bazą w L_0. Wówczas istnieją takie wektory $e_1', \ldots, e_m' \in L$, że $\{e_1, e_1', \ldots, e_m', e_m\}$ jest bazą hiperboliczną swej powłoki liniowej.*

Dowód. Powłokę liniową $\{e_2, \ldots, e_m\}$ oznaczmy przez L_1. Ponieważ L_1 jest ściśle zawarta w L_0, więc $L_1^\perp$ jest ściśle zawarta w $L_0^\perp$ na mocy niezdegenerowania L. Niech $e_1'' \subset L_1^\perp \setminus L_0^\perp$. Wtedy $(e_1'', e_i) = 0$ dla $i \geq 2$, ale $(e_1'', e_1) \neq 0$, przy czym można przyjąć, że $(e_1'', e_1) = 1$, bo e_1'' nie jest proporcjonalny do e_1. Podobnie jak w dowodzie lematu p. 2 przyjmiemy $e_1' = e_1'' - \frac{1}{2}(e_1'', e_1'')e_1$. Zatem $\{e_1, e_1\}$ jest bazą hiperboliczną swej powłoki liniowej. Jej ortogonalne dopełnienie jest niezdegenerowane i zawiera podprzestrzeń osobliwą rozpiętą przez $\{e_2, \ldots, e_m\}$. Do tej pary podprzestrzeni można zastosować analogiczne rozumowanie, tak że dowód będzie zakończony po użyciu indukcji względem m.

4. Twierdzenie (Witt). *Niech L będzie niezdegenerowaną, skończenie wymiarową przestrzenią ortogonalną, $L', L'' \subset L$ — dwiema izometrycznymi podprzestrzeniami. Wtedy każdą izometrię $f': L' \to L''$ można przedłużyć do izometrii $f: L \to L$ równej f' na L'.*

Dowód. Rozważymy kolejno kilka możliwości.

a) *$L' = L''$ i obie przestrzenie są niezdegenerowane.* Wtedy $L = L' \oplus (L')^\perp$ i można przyjąć $f = f' \oplus id_{(L')^\perp}$.

b) *$L' \neq L''$, $\dim L' = \dim L'' = 1$ i obie przestrzenie są niezdegenerowane.* Izometrię $f': L' \to L''$ można przedłużyć do izometrii $f'': L' + L'' \to L' + L''$, przyjmując $f''(l) = f'(l)$ dla $l \in L'$, $f''(l) = (f')^{-1}(l)$ dla $l \in L''$. Jeśli $L' + L''$ jest niezdegenerowana, to f'' przedłuża się do f tak jak w poprzednim przypadku. Jeśli $L' + L''$ jest zdegenerowana, to jądro iloczynu skalarnego na $L' + L''$ jest jednowymiarowe. Niech e_1 rozpina to jądro, a e_2 rozpina L'. W dopełnieniu ortogonalnym do e_2 względem L znajdźmy taki wektor e_1', że $\{e_1, e_1'\}$ jest bazą hiperboliczną swojej powłoki liniowej. Taki wektor istnieje na podstawie lematu p. 3. Pokażemy teraz, że podprzestrzeń L_0 rozpięta przez $\{e_1, e_1', e_2\}$ jest niezdegenerowana, a izometrię $f': L'' \to L''$ można przedłużyć do izometrii $f: L_0 \to L_0$. W tej sytuacji można już będzie zastosować wynik punktu a).

Niezdegenerowanie L_0 wynika stąd, że macierzą Grama układu $\{e_1, e_1', e_2\}$ jest

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & (e_1, e_2) \end{pmatrix},$$

a $(e_2, e_2) \neq 0$. W celu przedłużenia f' zauważymy najpierw, że ortogonalne uzupełnienie do $f'(e_2)$ w L_0 jest dwuwymiarowe, niezdegenerowane i zawiera wektor osobliwy e_1. Jest ono więc płaszczyzną hiperboliczną, podobnie jak i ortogonalne uzupełnienie do e_2 w L_0. Z lematu p. 2 wynika istnienie izometrii drugiej z wy-

mienionych płaszczyzn na pierwszą. Suma prosta tej izometrii z f' jest szukanym przedłużeniem.

c) $\dim L' = \dim L'' > 1$ i L', L'' *są niezdegenerowane.* Przeprowadzimy indukcję względem $\dim L'$. Ponieważ L' zawiera bazę ortogonalną, istnieje rozkład $L' = L'_1 \oplus L'_2$ na ortogonalną sumę prostą podprzestrzeni niezerowego wymiaru. Ponieważ f' jest izometrią, zatem $L'' = L''_1 \oplus L''_2$, gdzie $L''_i = f'(L'_i)$, a suma jest też ortogonalna. Na mocy założenia indukcyjnego obcięcie f' do L'_1 przedłuża się do izometrii $f_1: L \to L$, która przeprowadza $(L'_1)^\perp \supset L'_2$ na $(L''_1)^\perp \supset L''_2$. Stosując jeszcze raz założenie indukcyjne, wnioskujemy o istnieniu izometrii $(L'_1)^\perp$ z $(L''_1)^\perp$, która na L'_2 jest równa obcięciu f'. Rozszerzając je o obcięcie f' do L'_1 otrzymujemy szukaną izometrię.

d) L' *jest zdegenerowana.* Sprowadzimy ten przypadek do omówionego wyżej. Niech $L'_0 \subset L'$ oznacza jądro obcięcia metryki do L'. Wybierając bazę ortogonalną w L' możemy skonstruować rozkład ortogonalny $L' = L'_1 \oplus L'_0$ z niezdegenerowanym L'_1. W ortogonalnym dopełnieniu do L'_1 względem L można, stosując lemat p. 3, wybrać taką podprzestrzeń $\overline{L}'_0$, dla której suma $\overline{L}'_0 \oplus L'_1 \oplus L'_0$ będzie sumą prostą, a przestrzeń $\overline{L}'_0 \oplus L'_0$ będzie przestrzenią hiperboliczną; w szczególności $\overline{L}'_0 \oplus L'_1 \oplus L'_0$ będzie niezdegenerowana. Analogicznie skonstruujemy $\overline{L}''_0 \oplus L''_1 \oplus L''_0$ dla przestrzeni L''. Jest jasne, że izometrię $f': L' \to L''$ można rozszerzyć do izometrii tych ortogonalnych sum prostych, bo wszystkie przestrzenie hiperboliczne jednakowych wymiarów są izometryczne. Możliwość dalszego rozszerzenia wynika teraz z przypadku a). Twierdzenie zostało wykazane.

5. Wniosek. *Niech L_1, L_2 będą izometrycznymi przestrzeniami niezdegenerowanymi, a L'_1, L'_2 — izometrycznymi podprzestrzeniami. Wówczas ich ortogonalne dopełnienia $(L'_1)^\perp$, $(L'_2)^\perp$ są także izometryczne.*

6. Wniosek. *Niech L będzie przestrzenią ortogonalną niezdegenerowaną. Wówczas każda podprzestrzeń osobliwa zawiera się w maksymalnej podprzestrzeni osobliwej, a dla każdych dwóch maksymalnych podprzestrzeni osobliwych L', L'' istnieje izometria $f: L \to L$ odwzorowująca L' na L''.*

Dowód. Pierwsza część jest oczywiście prawdziwa. Dla dowodu drugiej załóżmy, że $\dim L' \leq \dim L''$. Dowolna injekcja liniowa $f': L' \to L''$ jest izometrią L' z jej obrazem $\operatorname{Im} f'$ i dlatego można ją przedłużyć do izometrii $f: L \to L$. Wtedy $L' \subset f^{-1}(L'')$, a $f^{-1}(L'')$ jest osobliwa. Ponieważ L' jest maksymalna, $\dim L' = \dim f^{-1}(L'') = \dim L''$.

7. Wniosek. *Dla każdej przestrzeni ortogonalnej istnieje rozkład na ortogonalną sumę prostą $L_0 \oplus L_h \oplus L_d$, gdzie L_0 jest osobliwa, L_h — hiperboliczna, L_d — anizotropowa. Dla każdych dwóch takich rozkładów istnieje izometria $f: L \to L$, przeprowadzająca jeden rozkład w drugi.*

Dowód. Jako L_0 weźmiemy jądro iloczynu skalarnego i rozłożymy L w sumę prostą $L = L_0 \oplus L_1$. Wybierzemy w L_1 maksymalną podprzestrzeń osobliwą i wło-

żymy ją w podprzestrzeń hiperboliczną o podwojonym wymiarze, tak jak w lemacie p. 3. Jako L_d weźmiemy dopełnienie ortogonalne do L_h względem L_1 — nie zawiera ono wektorów osobliwych, w przeciwnym przypadku bowiem można byłoby taki wektor dołączyć do znalezionej na początku przestrzeni osobliwej, przecząc jej maksymalności. To dowodzi istnienia takiego rozkładu.

Odwrotnie, dla każdego takiego rozkładu $L_0 \oplus L_h \oplus L_d$ podprzestrzeń L_0 jest jądrem iloczynu skalarnego. Ponadto, maksymalna podprzestrzeń osobliwa w L_h jest jednocześnie maksymalną osobliwą podprzestrzenią w $L_h \oplus L_d$, a więc jej wymiar jest wyznaczony jednoznacznie. To oznacza, że dla dwóch takich rozkładów $L_0 \oplus L_h \oplus L_d$ i $L_0 \oplus L_h' \oplus L_d'$ istnieje izometria przeprowadzająca L_0 na L_0 i L_h na L_h'. Na mocy twierdzenia Witta można ją rozszerzyć przez izometrię L_d na L_d, co kończy dowód. Nazwiemy L_d *częścią anizotropową* przestrzeni L — jest ona wyznaczona z dokładnością do izometrii.

Ten wniosek jest uogólnieniem zasady bezwładności na przypadek dowolnego ciała skalarów, sprowadzającym klasyfikację przestrzeni ortogonalnych do klasyfikacji przestrzeni anizotropowych lub, w języku form kwadratowych, do klasyfikacji *form nie reprezentujących zera*, dla których z $q(l) = 0$, wynika, że $l = 0$.

8. Grupa Witta. Niech $\boldsymbol{K}$ oznacza ciało skalarów. Oznaczmy przez $W(\boldsymbol{K})$ zbiór klas izometrycznych anizotropowych przestrzeni ortogonalnych nad ciałem $\boldsymbol{K}$, z dołączoną klasą przestrzeni zerowej. W zbiorze $W(\boldsymbol{K})$ wprowadzimy następującą operację dodawania: jeśli L_1, L_2 są dwiema przestrzeniami anizotropowymi, $[L_1]$, $[L_2]$ — ich klasami w $W(\boldsymbol{K})$, to $[L_1] + [L_2]$ będzie oznaczać klasę części anizotropowej przestrzeni $L_1 \oplus L_2$ (jest to zewnętrzna ortogonalna suma prosta).

Nietrudno sprawdzić, że definicja jest poprawna. Co więcej, to działanie jest łączne, a klasa przestrzeni zerowej jest elementem neutralnym (zerowym) dla tego działania. Zachodzi następujące twierdzenie.

9. Twierdzenie. a) *$W(\boldsymbol{K})$ wraz z wprowadzonym powyżej działaniem dodawania jest grupą abelową, nazywaną grupą Witta ciała $\boldsymbol{K}$.*

b) *Oznaczmy przez L_a jednowymiarową przestrzeń kartezjańską nad $\boldsymbol{K}$ z iloczynem skalarnym axy, gdzie $a \in \boldsymbol{K} \setminus \{0\}$. Wówczas $[L_a]$ zależy tylko od warstwy $a(\boldsymbol{K}^*)^2$, a elementy $[L_a]$ są układem generatorów grupy $W(\boldsymbol{K})$.*

Dowód. Pozostaje nam sprawdzić, że dla każdego elementu $W(\boldsymbol{K})$ istnieje element odwrotny. Obierzmy w tym celu przestrzeń anizotropową L z metryką, która w bazie ortogonalnej $\{e_1, \ldots, e_n\}$ jest dana przez formę $\sum_{i=1}^{n} a_i x_i^2$. Oznaczając przez $\overline{L}$ przestrzeń L z metryką $-\sum_{i=1}^{n} a_i x_i^2$ wykażemy, że przestrzeń $L \oplus \overline{L}$ jest hiperboliczna, skąd wynika $[L] + [\overline{L}] = [0]$ w $W(\boldsymbol{K})$. Rzeczywiście, metryka w $L \oplus \overline{L}$ jest dana przez formę $\sum_{i=1}^{n} a_i(x_i^2 - y_i^2)$. Płaszczyzna z metryką $a(x - y^2)$ jest oczywiście hiperboliczna, bo forma jest niezdegenerowana i wektor $(1, 1)$ jest osobliwy. Ten fakt, że $[L_a]$ zależy tylko od $a(\boldsymbol{K}^*)^2$, został sprawdzony w p. 7 § 2. Poza tym, każda n-wymiarowa

przestrzeń ortogonalna rozkłada się na ortogonalną sumę prostą jednowymiarowych przestrzeni typu L_a, co kończy dowód.

§ 15. Algebry Clifforda

1. *Algebrą* nad ciałem $\boldsymbol{K}$ będziemy nazywać pierścień łączny z jednością A, zawierający ciało $\boldsymbol{K}$ i taki, że $\boldsymbol{K}$ jest zawarte w centrum A, tj. komutuje ze wszystkimi elementami A. W szczególności, A jest przestrzenią liniową nad $\boldsymbol{K}$.

Rozważmy skończenie wymiarową przestrzeń ortogonalną L z metryką g. W tym paragrafie zbudujemy algebrę $C(L)$ i włożenie $\rho: L \to C(L)$ liniowe nad $\boldsymbol{K}$, że dla każdego $l \in L$ będzie spełniony warunek

$$\rho(l)^2 = g(l,l) \cdot 1,$$

tj. *kwadrat skalarny* każdego wektora z L będzie można otrzymać jako jego *kwadrat* względem mnożenia w $C(L)$. Ponadto, elementy $\rho(L)$ będą *multyplikatywnymi generatorami* $C(L)$, tj. każdy element z $C(L)$ można będzie zapisać jako kombinację liniową (nieprzemiennych) jednomianów względem elementów $\rho(L)$. Algebra $C(L)$ (wraz z odwzorowaniem ρ) o takich własnościach nazywa się *algebrą Clifforda* przestrzeni L.

2. Twierdzenie. a) *Dla każdej skończenie wymiarowej przestrzeni ortogonalnej L istnieje algebra Clifforda $C(L)$ i jej wymiar nad $\boldsymbol{K}$ jest równy 2^n, gdzie $n = \dim L$.*

b) *Niech $\sigma: L \to D$ będzie takim odwzorowaniem liniowym nad $\boldsymbol{K}$ przestrzeni L w algebrę D nad $\boldsymbol{K}$, dla którego $\sigma(l)^2 = g(l,l) \cdot 1$ dla każdego $l \in L$. Wówczas istnieje jedyny homomorfizm algebr nad $\boldsymbol{K}$ $\tau: C(L) \to D$, taki że $\sigma = \tau \circ \rho$. W szczególności $C(L)$ jest wyznaczona jednoznacznie z dokładnością do izomorfizmu.*

Dowód. a) Wybierzmy w L bazę ortogonalną $\{e_1, \ldots, e_n\}$, $(e_i, e_i) = a_i$. Na mocy definicji w $C(L)$ są spełnione relacje

$$\rho(e_i)^2 = a_i, \quad \rho(e_i)\rho(e_j) = -\rho(e_j)\rho(e_i), \quad i \neq j.$$

Drugą z nich otrzymujemy, zauważając, że $[\rho(e_i + e_j)]^2 = \rho(e_i)^2 + \rho(e_i)\rho(e_j) + \rho(e_j)\rho(e_i) + \rho(e_j)^2 = \rho(e_i)^2 + \rho(e_j)^2$. Możemy wyrazić elementy $l_1, \ldots, l_m \in L$ jako kombinacje liniowe elementów bazy $\{e_i\}$ i, korzystając z tego, że mnożenie w $C(L)$ jest $\boldsymbol{K}$-liniowe względem każdego z czynników (co wynika z faktu, że $\boldsymbol{K}$ jest zawarte w centrum), dowolny iloczyn $\rho(l_1) \cdots \rho(l_m)$ możemy zapisać w postaci kombinacji liniowej jednomianów względem $\rho(e_i)$. Zamieniając $\rho(e_i)^2$ na a_i, a $\rho(e_i)\rho(e_j)$ dla $i > j$ na $-\rho(e_j)\rho(e_i)$, możemy sprowadzić dowolny jednomian do postaci $a\rho(e_{i_1}) \cdots \rho(e_{i_m})$, gdzie $a \in \boldsymbol{K}$, $i_1 < i_2 < \ldots < i_m$. Innych relacji między tymi elementami nie ma, a liczba jednomianów $\rho(e_{i_1}) \cdots \rho(e_{i_m})$ wynosi 2^m (włączając trywialny jednomian 1 przy $m = 0$).

Strategia dowodu polega na uściśleniu przez bardziej formalną argumentację tych naprowadzających rozważań.

W tym celu każdemu podzbiorowi $S \subset \{1, \ldots, n\}$ przyporządkujemy symbol e_S (który następnie zinterpretujemy jako iloczyn $\rho(e_i) \cdots \rho(e_i)$, jeśli $S = \{i_1, \ldots, i_m\}$, $i_1 < \ldots < i_m$); przyjmiemy także $e_\emptyset = 1$ ($\emptyset$ oznacza zbiór pusty). Oznaczymy przez $C(L)$ przestrzeń liniową nad $\boldsymbol{K}$ z bazą $\{e_S\}$. Wprowadzimy mnożenie w $C(L)$ za pomocą następującej konstrukcji. Jeśli $1 \leq s, t \leq n$, to określimy

$$(s, t) = \begin{cases} 1, & \text{dla } s \leq t, \\ -1, & \text{dla } s > t. \end{cases}$$

Dla dwóch podzbiorów $S, T \subset \{1, \ldots, n\}$ przyjmiemy

$$a(S,T) = \prod_{\substack{s \in S \\ t \in T}} (s, t) \prod_{i \in S \cap T} a_i,$$

gdzie, przypomnijmy, $a_i = g(e_i, e_i)$. (Przyjmujemy, że iloczyny z pustym zbiorem indeksów są równe jedności.) I wreszcie, iloczyn kombinacji liniowych $\sum a_S e_S$, $\sum b_T e_T \in C(L)$; $a_S, b_T \in \boldsymbol{K}$ określimy wzorem

$$(\sum a_S e_S)(\sum b_T e_T) = \sum a_S b_T a(S,T) e_{S \nabla T},$$

gdzie przez $S \nabla T = (S \cup T) \setminus (S \cap T)$ oznaczyliśmy różnicę symetryczną zbiorów S, T. Trywialnie sprawdzamy wszystkie aksjomaty algebry nad $\boldsymbol{K}$, z wyjątkiem łączności mnożenia. Ten ostatni wystarczy sprawdzić dla elementów bazy, tj. wystarczy wykazać tożsamość

$$(e_S e_T) e_R = e_S (e_T e_R).$$

Ponieważ $e_S e_T = a(S,T) e_{S \nabla T}$, więc otrzymujemy

$$(e_S e_T) e_R = a(S,T) a(S \nabla T, R) e_{(S \nabla T) \nabla R},$$
$$e_S (e_T e_R) = a(S, T \nabla R) a(T, R) e_{S \nabla (T \nabla R)}.$$

Nietrudno sprawdzić, że

$$(S \nabla T) \nabla R = S \nabla (T \nabla R) = \{(S \cup T \cup R) \setminus [(S \cap T) \cup (S \cap R) \cup (T \cap R)]\} \cup (S \cap T \cap R).$$

Pozostaje więc przekonać się o równości czynników skalarnych. O ich znakach decyduje czynnik $a(S,T)a(S \nabla T, R)$ postaci

$$\prod_{\substack{s \in S \\ t \in T}} (s, t) \prod_{\substack{u \in S \nabla T \\ r \in R}} (u, r).$$

Przyjmując, że u w drugim iloczynie przebiega najpierw wszystkie elementy S, a następnie wszystkie elementy T (przy ustalonym r), wprowadzimy dodatkowe czynniki

$(u,r)^2$, które są równe jedności dla $u \in S \cap T$, a więc ten „znak" możemy zapisać w takiej postaci, która zależy w sposób symetryczny od S, T, R:

$$\prod_{\substack{s\in S\\ t\in T}}(s,t)\prod_{\substack{u\in S\\ r\in R}}(u,r)\prod_{\substack{u\in T\\ r\in R}}(u,r)$$

W analogiczny sposób otrzymamy ten sam rezultat dla znaku odnoszącego się do iloczynu $a(S, T\nabla R)a(T,R)$. Pozostają więc do zbadania czynniki zawierające kwadraty skalarne a_i. Dla $a(S,T)a(S\nabla T,R)$ mają one postać

$$\prod_{i\in S\cap T} a_i \prod_{j\in(S\nabla T)\cap R} a_i.$$

Jednakże $(S\nabla T)\cap R = (S\cap R)\nabla(T\cap R)$, $(S\cap T)$ ma puste przecięcie z tym zbiorem, a $(S\cap T)\cup[(S\cap R)\nabla(T\cap R)]$ składa się z tych elementów $S\cup T\cup R$, które są zawarte w więcej niż jednym z tych trzech zbiorów. Dlatego nasz czynnik zależy w sposób symetryczny od S, T, R. Analogiczne rozumowanie prowadzi do tego samego rezultatu dla rozważanej części współczynnika $a(S, T\nabla R)a(T,R)$, co kończy dowód łączności mnożenia w $C(L)$.

Zdefiniujmy teraz odwzorowanie liniowe nad $\boldsymbol{K}$ $\rho: L \to C(L)$, przyjmując $\rho(e_i) = e_i$. Zgodnie ze wzorami opisującymi mnożenie w $C(L)$, element $e_\emptyset$ jest jednością w $C(L)$, a ponadto

$$\rho(e_i)\rho(e_j) = e_ie_j = \begin{cases} a_ie_\emptyset & \text{dla } i=j,\\ -e_je_i & \text{dla } i\neq j.\end{cases}$$

Zatem $(\rho, C(L))$ jest algebrą Clifforda dla L.

b) Ostatnia część tezy dowodzona jest w sposób formalny. Niech $\sigma: L \to D$ będzie odwzorowaniem liniowym nad $\boldsymbol{K}$ spełniającym $\sigma(l)^2 = g(l,l)\cdot 1$. Istnieje jedyne odwzorowanie liniowe nad $\boldsymbol{K}$, $\tau: C(L) \to D$, które dla elementów bazy jest określone wzorem

$$\tau(e_{i_1,\dots,i_m}) = \sigma(e_{i_1})\cdots\sigma(e_{i_m}),$$
$$\tau(e_\emptyset) = 1_D.$$

Spełnia ono $\tau\circ\rho = \sigma$, ponieważ równość ta jest spełniona dla elementów bazy w L. Wreszcie τ jest homomorfizmem algebr, gdyż

$$\tau(e_Se_T) = \tau(a(S,T)e_{S\nabla T}) = a(S,T)\tau(e_{S\nabla T}),$$

i nietrudno sprawdzić, że $\tau(e_S)\tau(e_T)$ można sprowadzić do tej samej postaci, korzystając z zależności

$$\rho(e_i)^2 = a_i, \quad \rho(e_i)\rho(e_j) = -\rho(e_j)\rho(e_i), \quad i\neq j.$$

To kończy dowód.

3. Przykłady. a) Niech L oznacza dwuwymiarową płaszczyzną rzeczywistą z metryką $-(x^2+y^2)$. Bazą algebry Clifforda $C(L)$ jest zbiór $(1, e_1, e_2, e_1e_2)$, którego elementy spełniają relacje multyplikatywne

$$e_1^2 = e_2^2 = -1, \quad e_1e_2 = e_2e_1.$$

Nietrudno sprawdzić, że odwzorowanie $C(L) \to H$: $1 \mapsto \mathbf{1}$, $e_1 \mapsto \boldsymbol{i}$, $e_2 \mapsto \boldsymbol{j}$, $e_1e_2 \mapsto \boldsymbol{k}$ wyznacza izomorfizm $C(L)$ z algebrą kwaternionów H.

b) Niech L będzie przestrzenią liniową z metryką *zerową*. Algebra $C(L)$ jest w tym przypadku generowana przez elementy $\{e_1, \ldots, e_n\}$ i relacje

$$e^2 = 0, \quad e_ie_j = -e_je_i, \quad \text{przy } i \neq j.$$

Nazywamy ją *algebrą zewnętrzną* lub *algebrą Grassmanna* przestrzeni liniowej L. Wrócimy do niej w części 4.

c) Niech $L = \mathcal{M}^{\boldsymbol{C}}$ będzie kompleksyfikacją przestrzeni Minkowskiego z metryką $x_0^2 - \sum_{i=1}^{3} x_i^2$ względem bazy ortonormalnej $\{e_i\}$ w $\mathcal{M}$, będącej jednocześnie bazą $\mathcal{M}^{\boldsymbol{C}}$. Wykażemy, że algebra Clifforda $C(\mathcal{M}^{\boldsymbol{C}})$ jest izomorficzna z algebrą macierzy zespolonych 4×4. W tym celu rozważymy macierze Diraca zapisane w postaci blokowej:

$$\gamma_0 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \quad \gamma_j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad j = 1, 2, 3.$$

Posługując się własnościami macierzy Pauliego σ_j, nietrudno sprawdzić, że γ_j spełniają te same relacje w algebrze macierzy $M_4(\boldsymbol{C})$, co $\rho(e_j)$ w algebrze $C(\mathcal{M}^{\boldsymbol{C}})$:

$$\gamma_0^2 = -\gamma_1^2 = -\gamma_2^2 = -\gamma_3^2 = 1$$

(tutaj oczywiście $1 = E_4$) oraz

$$\gamma_i\gamma_j + \gamma_j\gamma_i = 0 \quad \text{dla } j \neq i.$$

To znaczy, że liniowe nad $\boldsymbol{C}$ odwzorowanie $\sigma\colon \mathcal{M}^{\boldsymbol{C}} \to M_4(\boldsymbol{C})$ indukuje homomorfizm algebr $\tau\colon C(\mathcal{M}^{\boldsymbol{C}}) \to M_4(\boldsymbol{C})$ spełniający $\tau \circ \rho(e_i) = \gamma_i$. Bezpośrednim rachunkiem można sprawdzić, że odwzorowanie τ jest surjektywne, a ponieważ obie algebry nad $\boldsymbol{C}$ są szesnastowymiarowe, więc stąd wynika, że τ jest izomorfizmem.

Część 3.

Geometria afiniczna i rzutowa

§ 1. Przestrzenie afiniczne, odwzorowania afiniczne i współrzędne afiniczne

1. Definicja. *Przestrzenią afiniczną nad ciałem* $\boldsymbol{K}$ *nazywamy trójkę* $(A, L, +)$ *składającą się z przestrzeni liniowej* L *nad ciałem* $\boldsymbol{K}$*, zbioru* A*, którego elementy nazywamy punktami, oraz zewnętrznej operacji dwuargumentowej* $A \times L \to A : (a, l) \mapsto a + l$*, spełniającej następujące aksjomaty:*

a) $(a + l) + m = a + (l + m)$ *dla każdych* $a \in A$, $l, m \in L$;

b) $a + 0 = a$ *dla każdego* $a \in A$;

c) *dla dowolnych dwóch punktów* $a, b \in A$ *istnieje jedyny wektor* $l \in L$*, taki że* $b = a + l$.

2. Przykład. Trójka $(L, L, +)$, gdzie L jest przestrzenią liniową, a operacja $+$ jest dodawaniem w L, jest przestrzenią afiniczną. W tej sytuacji wygodnie mówić, że ta trójka zadaje strukturę afiniczną w przestrzeni liniowej L. Przykład ten jest typowy — zobaczymy później, że każda przestrzeń afiniczna jest izomorficzna z przestrzenią tej postaci.

3. Terminologia. Będziemy często nazywać przestrzenią afiniczną parę (A, L) albo po prostu A, opuszczając jawne wskazanie działania $+$. Przestrzeń liniową L nazywamy *stowarzyszoną* z przestrzenią afiniczną A. Odwzorowanie $A \to A : a \mapsto a + l$ nazywamy *przesunięciem* o wektor l. Ponieważ wygodnie mieć do dyspozycji specjalne oznaczenie dla tego odwzorowania, będziemy je oznaczać przez t_l, a także będziemy pisać $a - l$ zamiast $t_{-l}(a)$ lub $a + (-l)$.

4. Stwierdzenie. *Odwzorowanie* $l \mapsto t_l$ *jest injektywnym homomorfizmem grupy addytywnej przestrzeni* L *w grupę permutacji punktów przestrzeni afinicznej* A lub, inaczej, *wyznacza efektywne działanie* L *na* A*. To działanie jest tranzytywne, tj. dla każdej pary punktów* $a, b \in A$ *istnieje takie* $l \in L$*, że* $t_l(a) = b$.

Odwrotnie, zadanie tranzytywnego i efektywnego działania grupy L *na zbiorze* A *wprowadza na* A *strukturę przestrzeni afinicznej związanej z przestrzenią* L.

Dowód. Z aksjomatów a) i b) wynika, że dla dowolnych $l \in L$ i $a \in A$ równanie

$t_l(x) = a$ ma rozwiązanie $x = a + (-l)$, i wobec tego każde t_l jest surjektywne. Jeśli $t_l(a) = t_l(b)$, to wyznaczając na podstawie aksjomatu c) taki wektor $m \in L$, że $b = a + m$, dostajemy $a + l = (a + m) + l = (a + l) + m$. Ponieważ $a + l = (a + l) + 0$, na podstawie warunku jednoznaczności aksjomatu c) wnioskujemy, że $m = 0$, więc $a = b$. Dlatego wszystkie t_l są injektywne.

Aksjomat a) oznacza, że $t_l \circ t_m = t_{l+m}$, a aksjomat b) — że $t_0 = \mathrm{id}_A$. Dlatego odwzorowanie $l \mapsto t_l$ jest homomorfizmem grupy L w grupę bijekcji A na A. Z aksjomatów b) i c) wynika, że jądro tego homomorfizmu jest zerowe.

Odwrotnie, niech $L \times A \to A : (l, a) \mapsto a + l$ oznacza efektywne i tranzytywne działanie L na A. Wówczas aksjomaty a) i b) otrzymuje się wprost z definicji, a aksjomat c) z połączenia warunków efektywności i tranzytywności działania.

5. Uwaga. Informacje o działaniach grup (niekoniecznie abelowych) na zbiorach można znaleźć we *Wstępie do algebry*, § 2, rozdz. 7. Zbiór, na którym grupa działa tranzytywnie i efektywnie nazywa się *główną przestrzenią jednorodną* tej grupy.

Zwróćmy uwagę na to, że w aksjomatach przestrzeni afinicznej nie pojawia się nigdzie w jawny sposób działanie mnożenia przez skalary w L. Wystąpi ono dopiero przy określeniu odwzorowań afinicznych oraz barycentrycznych kombinacji punktów przestrzeni A. Ale najpierw kilka słów o formalizmie.

6. Reguły rachunkowe. Ten jedyny wektor $l \in L$, dla którego $b = a + l$ wygodnie będzie oznaczyć przez $b - a$. Tak określona operacja „zewnętrznego odejmowania" $A \times A \to L : (b, a) \mapsto b - a$ ma następujące własności:

a) $(c - b) + (b - a) = c - a$ dla wszystkich $a, b, c \in A$; dodawanie po lewej stronie jest dodawaniem w L.

Rzeczywiście, niech $c = b + l$, $b = a + m$, wówczas $c = a + (l + m)$, tak że $c - a = l + m = (c - b) + (b - a)$.

b) $a - a = 0$ dla każdego $a \in A$.

c) $(a + l) - (b + m) = (a - b) + (l - m)$ dla każdych $a, b \in A$, $l, m \in L$.

W istocie, wystarczy sprawdzić, że $(b + m) + (a - b) + (l - m) = a + l$ lub $b + (a - b) = a$, a to jest określeniem $a - b$.

W ogólności użycie znaków $\pm$ dla rozmaitych operacji $L \times L \to L$, $A \times L \to L$, $A \times A \to L$ podlega następującym formalnym zasadom. Wyrażenie $\pm a_1 \pm a_2 \pm \cdots \pm a_m + l_1 + \cdots + l_n$ dla $a_j \in A$, $l_k \in L$ ma sens, jeśli m jest parzyste i wszystkie $\pm a_j$ można połączyć w pary postaci $a_i - a_j$ albo m jest nieparzyste i wszystkie punkty, poza jednym występującym w tym wyrażeniu ze znakiem $+$, można połączyć w takie pary. W pierwszym przypadku suma należy do L, a w drugim — do A. Poza tym zależy ona od swych składników łącznie i przemiennie — np. $a_1 - a_2 + l$ można obliczyć albo jako $(a_1 - a_2) + l$, albo $(a_1 + l) - a_2$, albo $a_1 - (a_2 - l)$. Pozwolimy sobie na pisanie zarówno $a + l$, jak też $l + a$.

Dowodzić tej zasady w pełnej ogólności nie będziemy. Za każdym razem, gdy będziemy się nią posługiwać, czytelnik będzie mógł bez trudności rozbić potrzebne rachunki na serię elementarnych kroków, z których każdy da się sprowadzić do za-

stosowania jednego z aksjomatów lub wzorów a) — c) wymienionych na początku tego punktu.

Zauważymy, że *suma* $a + b$ dla $a, b \in A$ na ogół *nie ma sensu*, podobnie jak *wyrażenie* xa, gdzie $x \in \boldsymbol{K}$ (wyjątek $A = L$). Niemniej jednak poniżej nadamy jednoznaczne znaczenie wyrażeniu $\frac{1}{2}a + \frac{1}{2}b$ (ale nie jego składnikom!).

Intuicyjnie można wyobrażać sobie przestrzeń afiniczną $(A, L, +)$ jako przestrzeń liniową L z „zatartym" początkiem układu współrzędnych 0. Pozostały jedynie operacje przesunięcia o wektory L, składania przesunięć i mnożenia wektora przesunięcia przez skalar.

7. Odwzorowania afiniczne. Niech (A_1, L_1), (A_2, L_2) będą dwiema przestrzeniami afinicznymi nad tym samym ciałem $\boldsymbol{K}$. *Odwzorowaniem afiniczno-liniowym* lub prościej *afinicznym* pierwszej z nich w drugą nazwiemy parę odwzorowań (f, Df), gdzie $f\colon A_1 \to A_2$, $Df\colon L_1 \to L_2$, spełniających następujące warunki:

a) Df jest odwzorowaniem liniowym;

b) dla dowolnych $a_1, a_2 \in A$ mamy

$$f(a_1) - f(a_2) = Df(a_1 - a_2).$$

(Oba wyrażenia należą do L_2.)

Df nazywamy *częścią liniową* odwzorowania afinicznego f. Ponieważ $a_1 - a_2$ przebiega wszystkie wektory L_1, kiedy $a_1, a_2 \in A_1$, więc część liniowa Df jest wyznaczona przez f jednoznacznie. To pozwala na oznaczanie odwzorowania afinicznego krótko przez $f\colon A_1 \to A_2$.

8. Przykłady. a) Dowolne odwzorowanie liniowe $f\colon L_1 \to L_2$ indukuje odwzorowanie afiniczne przestrzeni $(L_1, L_1, +) \to (L_2, L_2, +)$. W tym przypadku $Df = f$.

b) Dowolne przesunięcie $t_l\colon A \to A$ jest afiniczne i $D(t_l) = \mathrm{id}_L$. Rzeczywiście,

$$t_l(a_1) - t_l(a_2) = (a_1 + l) - (a_2 + l) = a_1 - a_2.$$

c) Jeśli $f\colon A_1 \to A_2$ jest odwzorowaniem afinicznym i $l \in L_2$, to odwzorowanie $t_l \circ f\colon A_1 \to A_2$ jest też afiniczne i $D(t_l \circ f) = D(f)$. W istocie,

$$t_l \circ f(a_1) - t_l \circ f(a_2) = (f(a_1) + l) - (f(a_2) + l) = f(a_1) - f(a_2) = Df(a_1 - a_2).$$

d) *Funkcja afiniczna* lub *afiniczno-liniowa* $f\colon A \to \boldsymbol{K}$ jest to odwzorowanie afiniczne A w $(\boldsymbol{K}^1, \boldsymbol{K}^1, +)$, gdzie $\boldsymbol{K}^1$ jest jednowymiarową przestrzenią kartezjańską. W takim przypadku f ma wartości w zbiorze $\boldsymbol{K}$, a Df jest funkcjonałem liniowym na L. Każda funkcja stała jest afiniczna i dla niej $Df = 0$.

9. Twierdzenie. a) *Przestrzenie afiniczne wraz z odwzorowaniami afinicznymi stanowią kategorię.*

b) *Odwzorowanie, przyporządkowujące przestrzeni afinicznej* (A, L) *przestrzeń liniową* L, *a odwzorowaniu afinicznemu* $f\colon (A_1, L_1) \to (A_2, L_2)$ *odwzorowanie liniowe*

$Df: L_1 \to L_2$, jest funktorem z kategorii przestrzeni afinicznych do kategorii przestrzeni liniowych.

Dowód. Spełnienie aksjomatów teorii kategorii (por. § 13, część 1) wynika z następujących faktów.

Odwzorowanie tożsamościowe $\mathrm{id}: A \to A$ jest afiniczne. Istotnie, $a_1 - a_2 = \mathrm{id}_L(a_1 - a_2)$. W szczególności $D(\mathrm{id}_A) = \mathrm{id}_L$.

Złożenie odwzorowań afinicznych $A \xrightarrow{f_1} A \xrightarrow{f_2} A$ jest odwzorowaniem afinicznym.

Istotnie, niech $a, b \in A$. Wówczas $f_1(a) - f_1(b) = Df_1(a - b)$, a następnie

$$f_2 f_1(a) - f_2 f_1(b) = Df_2[f_1(a) - f_1(b)] = Df_2 \circ Df_1(a - b).$$

Wykazaliśmy więc postulowany fakt, a wraz z nim dowiedliśmy wzoru $D(f_2 \circ f_1) = D(f_2) \circ D(f_1)$. Łącznie z $D(\mathrm{id}_A) = \mathrm{id}_L$ dowodzi to części b) tezy twierdzenia.

Następujący ważny wynik charakteryzuje izomorfizmy w naszej kategorii.

10. Stwierdzenie. *Następujące trzy własności odwzorowania afinicznego $f: A_1 \to A_2$ są równoważne:*

a) *f jest izomorfizmem,*

b) *Df jest izomorfizmem,*

c) *f jest bijekcją (zbiorów).*

Dowód. Zgodnie z ogólnym określeniem teorii kategorii, $f: A_1 \to A_2$ jest izomorfizmem wtedy i tylko wtedy, gdy istnieje takie odwzorowanie afiniczne $g: A_2 \to A_1$, że $gf = \mathrm{id}_{A_1}$, $fg = \mathrm{id}_{A_2}$. Jeśli takie odwzorowanie istnieje, to $D(fg) = \mathrm{id}_{L_2} = D(f)D(f)$ i $D(gf) = \mathrm{id}_{L_1} = D(g)D(f)$, skąd wynika, że $D(f)$ jest izomorfizmem.

Wykażemy teraz, że Df jest izomorfizmem wtedy i tylko wtedy, gdy f jest bijekcją. Ustalmy punkt $a_1 \in A_1$ i oznaczmy $a_2 = f(a_1)$. Każdy element A_i można jednoznacznie przedstawić w postaci $a_i + l_i$, $l_i \in L_i$, $i = 1, 2$. Na mocy podstawowej tożsamości

$$f(a_1 + l_1) - f(a_1) = Df[(a_1 + l_1) - a_1] = Df(l_1)$$

otrzymujemy $f(a_1 + l_1) = a_2 + Df(l_1)$. Zatem f jest bijekcją wtedy i tylko wtedy, gdy $Df(l_1)$ przyjmuje każdą wartość z L_2 jednokrotnie, gdy l_1 przebiega przestrzeń L_1, tj. gdy Df jest bijekcją. Wiadomo, że odwzorowanie liniowe jest bijekcją wtedy i tylko wtedy, gdy jest ono nieosobliwe, tj. gdy jest izomorfizmem.

Pokażemy wreszcie, że bijektywne odwzorowanie afiniczne jest izomorfizmem afinicznym. W tym celu musimy sprawdzić, że odwzorowanie odwrotne do f w sensie teoriomnogościowym jest afiniczne. W oznaczeniach z poprzedniego akapitu to odwzorowanie jest zadane wzorem

$$f^{-1}(a_2 + l_2) = a_1 + (Df)^{-1}(l_2), \quad l_2 \in L_2.$$

Dlatego na mocy liniowości $(Df)^{-1}$

$$f^{-1}(a_2 + l_2) - f^{-1}(a_2 + l_2') = (Df)^{-1}(l_2) - (Df)^{-1}(l_2') = (Df)^{-1}(l_2 - l_2').$$

A więc f^{-1} jest afiniczne i $(Df^{-1}) = (Df)^{-1}$. Podsumowując, dowiedliśmy implikacji a) $\Rightarrow$ b) $\Leftrightarrow$ c) $\Rightarrow$ a), co kończy dowód twierdzenia.

Konstrukcję konkretnych odwzorowań afinicznych często opiera się na następującym rezultacie.

11. Stwierdzenie. *Niech (A_1, L_1), (A_2, L_2) będą dwiema przestrzeniami afinicznymi. Dla dowolnej pary punktów $a_1 \in A_1$, $a_2 \in A_2$ i dowolnego odwzorowania liniowego $g: L_1 \to L_2$ istnieje jedyne odwzorowanie afiniczne $f: A_1 \to A_2$ spełniające $f(a_1) = a_2$ i $Df = g$.*

Istotnie, przyjmijmy

$$f(a_1 + l_1) = a_2 + g(l_1)$$

dla $l_1 \in L_1$. Ponieważ dowolny punkt A_1 może być jednoznacznie zapisany w postaci $a_1 + l_1$, ten wzór zadaje odwzorowanie (teoriomnogościowe) $f: A_1 \to A_2$. Jest ono afiniczne, $f(a_1) = a_2$ i $Df = g$, ponieważ

$$f(a_1 + l_1) - f(a_1 + l_1') = g(l_1) - g(l_1') = g(l_1 - l_1') = g[(a_1 + l_1) - (a_1 + l_1')].$$

To dowodzi istnienia f. Odwrotnie, jeśli f jest odwzorowaniem z postulowanymi własnościami, to

$$f(a_1 + l) - f(a_1) = g(l),$$

skąd $f(a_1 + l_1) = a_2 + g(l)$ dla każdego $l \in L$.

12. Ważny szczególny przypadek stwierdzenia p. 11 otrzymujemy w zastosowaniu do sytuacji (A, L), (L, L), $a \in A$, $0 \in L$ i $g = \mathrm{id}_L$. Wnioskujemy wtedy, że *dla dowolnego punktu $a \in A$ istnieje jedyny izomorfizm afiniczny $f: A \to L$ przeprowadzający ten punkt w początek układu i mający tożsamościową część liniową.* To właśnie precyzuje sens stwierdzenia o tym, że przestrzeń afiniczna jest „przestrzenią liniową z zatartym początkiem układu współrzędnych". W szczególności, dwie przestrzenie afiniczne są izomorficzne wtedy i tylko wtedy, gdy są izomorficzne stowarzyszone z nimi przestrzenie liniowe. Te ostatnie są klasyfikowane przez wymiar, możemy więc nazwać wymiarem przestrzeni afinicznej wymiar stowarzyszonej z nią przestrzeni liniowej.

13. Wniosek. *Niech f_1, $f_2: A_1 \to A_2$ będą dwoma odwzorowaniami afinicznymi. Ich części liniowe są równe wtedy i tylko wtedy, gdy f_2 jest złożeniem f_1 z przesunięciem o wektor z L_2. Wektor przesunięcia jest jednoznacznie wyznaczony.*

Dowód. Dostateczność została wykazana w przykładzie c) p. 8. Dla dowodu konieczności wybierzmy dowolny punkt $a \in A_1$ i przyjmijmy $f_2' = t_{f_2(a) - f_1(a)} \circ f_1$. Oczywiście, $f_2'(a) = f_2(a)$ i $D(f_2') = D(f_2)$. Na mocy stwierdzenia p. 11 $f_2' = f_2$. Odwrotnie, jeśli $f_2 = t_l \circ f_1$, to $l = f_2(a) - f_1(a)$, przy czym ten wektor jest niezależny od wyboru $a \in A$ ze względu na równość części liniowych f_1, f_2.

14. Współrzędne afiniczne. a) *Układem współrzędnych afinicznych* w przestrzeni afinicznej (A, L) jest para, złożona z punktu $a_0 \in A$ (zwanego początkiem

układu) i bazy $\{e_1, \ldots, e_n\}$ w stowarzyszonej przestrzeni liniowej L. *Współrzędne* punktu $a \in A$ względem tego układu stanowią wektor $(x_1, \ldots, x_n) \in \boldsymbol{K}^n$, jednoznacznie wyznaczony przez warunek $a = a_0 + \sum_{i=1}^{n} x_i e_i$.

Innymi słowy, dokonujemy utożsamienia A z L za pomocą odwzorowania afinicznego, które przeprowadza a_0 w 0 i ma tożsamościową część liniową; współrzędne $x_1, \ldots, x_n$ są współrzędnymi obrazu punktu a względem bazy $\{e_1, \ldots, e_n\}$.

Załóżmy, że w przestrzeniach A_1, A_2 zostały wybrane układy współrzędnych, pozwalające na utożsamienie ich z $\boldsymbol{K}^n$, $\boldsymbol{K}^m$ odpowiednio. Wówczas dowolne odwzorowanie afiniczne $f\colon A_1 \to A_2$ można zapisać w postaci

$$f(\vec{x}) = B\vec{x} + \vec{y},$$

gdzie przez B oznaczona została macierz odwzorowania Df w odpowiednich bazach L_1 i L_2, a $\vec{y}$ są współrzędnymi wektora $f(a_0') - a_0''$ względem bazy L_2; przez a_0' oznaczyliśmy tu początek układu w A_1, a przez a_0'' — początek układu w A_2. Istotnie, odwzorowanie $\vec{x} \mapsto B\vec{x} + \vec{y}$ jest afiniczne, przeprowadza a_0' na $f(a_0')$ i ma tę samą część liniową co f.

b) Inny wariant podanej definicji układu współrzędnych polega na tym, aby zastąpić wektory $\{e_1, \ldots, e_n\}$ punktami $\{a_0 + e_1, \ldots, a_0 + e_n\}$ przestrzeni A. Oznaczmy $a_i = a_0 + e_i$, $i = 1, \ldots, n$. Współrzędne punktu $a \in A$ wyznacza się wówczas z zależności $a = a_0 + \sum_{i=1}^{n} x_i(a_i - a_0)$. Przy tym zapisie może pojawić się pokusa, aby „zgrupować wyrazy podobne" i zapisać wyrażenie stojące po prawej stronie w postaci $\left(1 - \sum_{i=1}^{n} x_i\right)a_0 + \sum_{i=1}^{n} x_i a_i$. Chociaż osobno żaden z tych składników nie ma sensu, tym niemniej rozważanie sum podobnego typu okazuje się bardzo pożyteczne i nie pozbawione sensu.

15. Stwierdzenie. *Niech $a_0, \ldots a_s$, będą dowolnymi punktami przestrzeni afinicznej A. Dla dowolnych $y_0, \ldots, y_s \in \boldsymbol{K}$ spełniających warunek $\sum_{i=0}^{s} y_i = 1$ określimy formalną sumę $\sum_{i=0}^{s} y_i a_i$ przez warunek*

$$\sum_{i=0}^{s} y_i a_i = a + \sum_{i=0}^{s} y_i(a_i - a),$$

gdzie a oznacza dowolny punkt A, a wyrażenie po prawej stronie nie zależy od wyboru tego punktu. W ten sposób punkt $\sum_{i=0}^{s} y_i a_i$ jest określony jednoznacznie i jest nazywany barycentryczną kombinacją punktów $a_0, \ldots, a_s$ o współczynnikach $y_0, \ldots, y_s$.

Dowód. Zamieńmy punkt a na punkt $a + l$, $l \in L$. Dostaniemy wówczas

$$a + l + \sum_{i=0}^{s} y_i(a_i - a - l) = a + \sum_{i=0}^{s} y_i(a_i - a),$$

ze względu na $\left(1-\sum_{i=0}^{s} y_i\right)l = 0$ i po wykorzystaniu reguł sformułowanych w p. 6. Szczegółowe prześledzenie opuszczonych przez nas rachunków będzie dla czytelnika pożyteczne.

16. Wniosek. *Układ* $\{a_0; a_1 - a_0, \ldots, a_n - a_0\}$ *składający się z punktu* $a_0 \in A$ *i wektorów* $a_i - a_0 \in L$ *wtedy i tylko wtedy stanowi afiniczny układ współrzędnych, gdy dowolny punkt* A *można jednoznacznie przedstawić w postaci barycentrycznej kombinacji* $\sum_{i=0}^{n} x_i a_i$, $x_i \in \boldsymbol{K}$, $\sum_{i=0}^{n} x_i = 1$.

Jeśli ten warunek jest spełniony, to układ punktów $\{a_0, \ldots, a_n\}$ nazywany jest *barycentrycznym układem współrzędnych* w przestrzeni A, a liczby $x_0, \ldots, x_n$ — *współrzędnymi barycentrycznymi* punktu $\sum_{i=0}^{n} x_i a_i$.

Dowód. Wszystko wynika bezpośrednio z definicji, jeśli obliczyć $\sum_{i=0}^{n} x_i a_i$ za pomocą wzoru $a_0 + \sum_{i=1}^{n} x_i(a_i - a_0)$ Istotnie, ponieważ dowolny punkt A można jednoznacznie przedstawić w postaci $a_0 + l$, $l \in L$, więc układ $\{a_0; a_1 - a_0, \ldots a_n - a_0\}$ jest afinicznym układem współrzędnych wtedy i tylko wtedy, gdy każdy wektor $l \in L$ może być jednoznacznie zapisany w postaci kombinacji liniowej $\sum_{i=1}^{n} x_i(a_i - a_0)$, tj. jeśli $\{a_1 - a_0, \ldots a_n - a_0\}$ tworzą bazę L. Współrzędne $x_1, \ldots, x_n$ wektora l wyznaczają jednoznacznie współrzędne barycentryczne punktu $a_0 + l$; $x_0 = 1 - \sum_{i=1}^{n} x_i, x_1, \ldots, x_n$.

17. Z kombinacjami barycentrycznymi można postępować w wielu przypadkach tak samo jak z kombinacjami liniowymi w przestrzeni liniowej. Na przykład, składniki z zerowymi współczynnikami można opuszczać. Szczególnie użyteczne jest zauważenie takiego faktu, że kombinacja barycentryczna dowolnej ilości kombinacji barycentrycznych punktów $a_0, \ldots, a_n$ jest znowu kombinacją barycentryczną tych punktów, a jej współczynniki można obliczyć, posługując się łatwą do odgadnięcia formułą:

$$x_1 \sum_{i=0}^{s} y_{i1} a_i + x_2 \sum_{i=0}^{s} y_{i2} a_i + \cdots + x_m \sum_{i=0}^{s} y_{im} a_i = \sum_{i=0}^{s} \left(\sum_{k=1}^{m} x_k y_{ik} \right) a_i.$$

Rzeczywiście, z równości

$$\sum_{i=0}^{s} \sum_{k=1}^{m} x_k y_{ik} = \sum_{k=1}^{m} x_k \sum_{i=0}^{s} y_{ik} = \sum_{k=1}^{m} x_k = 1$$

wynika, że wyrażenie po prawej stronie w poprzednim wzorze jest kombinacją barycentryczną. Obliczając lewą i prawą stronę tego wzoru przy użyciu definicji sformułowanej w stwierdzeniu p. 15, przy czym obie strony odnosimy do tego samego punktu $a \in A$, oraz stosując formalizm p. 6, z łatwością przekonujemy się o ich równości.

Co więcej, kombinacje barycentryczne zachowują się względem odwzorowań afinicznych podobnie jak kombinacje liniowe względem odwzorowań liniowych.

18. Stwierdzenie. a) *Niech* $f: A_1 \to A_2$ *będzie odwzorowaniem afinicznym i* $a_0, \ldots, a_s \in A_1$. *Wówczas*

$$f\Big(\sum_{i=0}^{s} x_i a_i\Big) = \sum_{i=0}^{s} x_i f(a_i), \quad \textit{jeśli} \quad \sum_{i=0}^{s} x_i = 1.$$

b) *Niech* $a_0, \ldots, a_n$ *zadają współrzędna barycentryczne w* A_1. *Wówczas dla dowolnych punktów* $b_0, \ldots b_n \in A_2$ *istnieje jednoznacznie wyznaczone odwzorowanie afiniczne* f *przeprowadzające* a_i *na* b_i, $i = 1, \ldots, n$.

Dowód. Wybierając $a \in A_1$, otrzymujemy

$$\begin{aligned} f\Big(\sum_{i=0}^{s} x_i a_i\Big) &= f\Big(a + \sum_{i=0}^{s} x_i (a_i - a)\Big) = f(a) + Df\Big(\sum_{i=0}^{s} x_i (a_i - a)\Big) = \\ &= f(a) + \sum_{i=0}^{s} x_i Df(a_i - a) = f(a) + \sum_{i=0}^{s} x_i (f(a_i) - f(a)) = \\ &= \sum_{i=0}^{s} x_i f(a_i) \end{aligned}$$

na mocy stwierdzenia p. 15, co dowodzi części a).

Jeśli $a_0, \ldots, a_n$ stanowią barycentryczny układ współrzędnych w A_1, to wobec wniosku p. 16 każdy punkt A ma jednoznaczne przedstawienie w postaci kombinacji barycentrycznej $\sum_{i=0}^{n} x_i a_i$. Określimy odwzorowanie $f: A_1 \to A_2$ wzorem $f\big(\sum_{i=0}^{n} x_i a_i\big) = \sum_{i=0}^{n} x_i b_i$. Na mocy a) jest to jedyna możliwość zdefiniowania f i pozostaje jedynie do sprawdzenia, że jest to odwzorowanie afiniczne. Rzeczywiście, postępując tak jak w stwierdzeniu p. 15, otrzymujemy

$$\begin{aligned} f\Big(\sum_{i=0}^{n} x_i a_i\Big) - f\Big(\sum_{i=0}^{n} y_i a_i\Big) &= \sum_{i=0}^{n} x_i b_i - \sum_{i=0}^{n} y_i a_i = \\ &= b_0 + \sum_{i=0}^{n} x_i (b_i - b_0) - \Big[b_0 + \sum_{i=0}^{n} y_i (b_i - b_0)\Big] = \\ &= \sum_{i=0}^{n} (x_i - y_i)(b_i - b_0) = Df\Big(\sum_{i=0}^{n} x_i a_i - \sum_{i=0}^{n} y_i a_i\Big), \end{aligned}$$

gdzie przez $Df: L_1 \to L_2$ oznaczyliśmy odwzorowanie liniowe przeprowadzające $a_i - a_0$ na $b_i - b_0$ dla wszystkich $i = 1, \ldots, n$. Takie odwzorowanie istnieje na mocy założenia, że $a_1 - a_0, \ldots, a_n - a_0$ stanowią bazę L_1.

19. Uwagi. W przestrzeni afinicznej $\boldsymbol{R}^n$ kombinacja barycentryczna $\sum_{i=1}^{m} \frac{1}{m} a_i$ reprezentuje położenie „środka mas" układu punktów materialnych o jednostkowych masach rozłożonych w punktach a_i. To wyjaśnia użytą terminologię.

Jeśli $a_i = (0, \ldots, 1, \ldots, 0)$, gdzie 1 występuje na i-tym miejscu, to zbiór punktów o współrzędnych barycentrycznych $x_1, \ldots, x_n$, $0 \leq x_i \leq 1$, jest częścią wspólną rozmaitości liniowej $\sum_{i=1}^{n} x_i = 1$ i dodatniego oktantu (dokładniej „2^n-tantu"). W topologii ten zbiór jest nazywany *standardowym* $(n-1)$*-wymiarowym sympleksem*. Jednowymiarowy sympleks to odcinek na prostej, dwuwymiarowy — trójkąt, trójwymiarowy — czworościan (tetraedr). W ogólności zbiór $\left\{\sum_{i=1}^{n} x_i a_i \mid \sum_{i=0}^{n} x_i = 1,\ 0 \leq x_i \leq 1\right\}$ stanowi *domknięty sympleks* o wierzchołkach $a_1, \ldots, a_n$ w rzeczywistej przestrzeni afinicznej. Nazywamy go *sympleksem zdegenerowanym*, jeśli wektory $a_2 - a_1, \ldots, a_n - a_1$ są liniowo zależne.

§ 2. Grupy afiniczne

1. Niech A będzie przestrzenią afiniczną nad ciałem $\boldsymbol{K}$. Na mocy stwierdzenia § 1, p. 10 zbiór bijektywnych odwzorowań afinicznych $f\colon A \to A$ stanowi grupę, którą będziemy nazywać *grupą afiniczną* i oznaczać symbolem $\boldsymbol{AF}(A)$.

Odwzorowanie $D\colon \boldsymbol{AF}(A) \to \boldsymbol{GL}(L)$, gdzie przez $\boldsymbol{GL}(L)$ oznaczamy grupę automorfizmów liniowych stowarzyszonej z A przestrzeni liniowej, jest homomorfizmem. Na mocy stwierdzenia § 1, p. 11 jest on surjektywny, a jego jądrem jest grupa przesunięć $\{t_l \mid l \in L\}$ jak to wynika z wniosku § 1, p. 13. Na mocy stwierdzenia § 1, p. 4 grupa przesunięć jest z kolei izomorficzna z addytywną grupą przestrzeni L. Ostatecznie więc grupa $\boldsymbol{AF}(A)$ jest rozszerzeniem grupy $\boldsymbol{GL}(L)$ przez addytywną grupę L, która jest normalną podgrupą w $\boldsymbol{AF}(A)$.

W tym przypadku rozszerzenie jest iloczynem półprostym $\boldsymbol{GL}(L)$ z L. Aby przekonać się o tym, ustalmy dowolny punkt $a \in A$ i rozważmy podgrupę $G_a \subset \boldsymbol{AF}(A)$ składającą się z odwzorowań zachowujących a. Na mocy stwierdzenia § 1, p. 11 każdy element $f \in G_a$ jest jednoznacznie wyznaczony przez swą część liniową Df, a tę można wybrać dowolnie. Stąd wynika, że D indukuje izomorfizm G_a na $\boldsymbol{GL}(L)$. Dla dowolnego odwzorowania $f \in \boldsymbol{AF}(A)$ można znaleźć jedyne odwzorowanie $f_a \in G_a$ o tej samej części liniowej i $f = t_l \circ f_a$ dla odpowiednio dobranego $l \in L$ na mocy wniosku § 1, p. 13. Ustalając a będziemy zapisywać $t_l \circ f_a$ jako parę $[g; l]$, gdzie $g = Df = Df_a \in \boldsymbol{GL}(L)$. Reguły mnożenia w grupie $\boldsymbol{AF}(A)$ zapisane w języku takich par przybierają następującą postać.

2. Stwierdzenie. *Zachodzą wzory*

$$[g_1; l_1][g_2; l_2] = [g_1 g_2; g_1(l_2) + l_1],$$

$$[g; l]^{-1} = [g^{-1}; -g^{-1}(l)].$$

Dowód. Zgodnie z definicją $[g; l]$ przeprowadza punkt $a+m \in A$ na $a+g(m)+l$, skąd otrzymujemy

$$[g_1; l_1][g_2; l_2](a+m) = [g_1; l_1](a + g_2(m) + l_2) =$$

$$= a + g_1(g_2(m) + l_2) + l_1 = a + g_1g_2(m) + g_1(l_2) + l_1 =$$

$$= [g_1g_2; g_1(l_2) + l_1](a+m),$$

co dowodzi pierwszego z podanych wzorów. Jeśli zastosować go do obliczenia iloczynu $[g; l][g^{-1}; -g^{-1}(l)]$, to otrzymamy $[\mathrm{id}_L; 0]$, a ta para reprezentuje jedynkę grupy $\boldsymbol{AF}(A)$. To kończy dowód i pokazuje, że $\boldsymbol{AF}(A)$ jest rzeczywiście iloczynem półprostym.

3. Rozważmy teraz podgrupę $G \subset \boldsymbol{GL}(L)$. Zbiór tych elementów $f \in \boldsymbol{AF}(A)$, których części liniowe należą do G, jest oczywiście podgrupą $\boldsymbol{AF}(A)$ — przeciwobrazem G względem kanonicznego homomorfizmu $\boldsymbol{AF}(A) \to \boldsymbol{GL}(L)$. Będziemy ją nazywać *rozszerzeniem afinicznym grupy* G.

Przypadek, gdy stowarzyszona z A przestrzeń liniowa jest zaopatrzona w dodatkową strukturę — iloczyn skalarny, a G jest odpowiadającą tej strukturze grupą izometrii, zasługuje na szczególną uwagę. W ten sposób są bowiem skonstruowane dwie bardzo ważne dla zastosowań grupy: *grupa ruchów* euklidesowej przestrzeni afinicznej ($G = \boldsymbol{O}(L)$) i *grupa Poincarégo* (L — przestrzeń Minkowskiego, G — grupa Lorentza). Zbadajmy dokładniej grupę ruchów.

4. Definicja. a) *Afiniczną przestrzenią euklidesową* nazywamy parę złożoną ze skończenie wymiarowej przestrzeni afinicznej nad ciałem liczb rzeczywistych i określonej na niej metryki d (w znaczeniu definicji przyjętej w części 1, § 10, p. 1), o następującej własności: dla dowolnych punktów $a, b \in A$ odległość $d(a, b)$ zależy tylko od $a-b \in L$ i jest równa długości wektora $a-b$ w pewnej metryce euklidesowej przestrzeni L (niezależnej od wyboru a, b).

b) *Ruchem euklidesowej przestrzeni afinicznej* A nazywamy takie odwzorowanie $f: A \to A$, które zachowuje odległości, tj. $d(f(a), f(b)) = d(a, b)$ dla dowolnych $a, b \in A$.

5. Twierdzenie. *Ruchy afinicznej przestrzeni euklidesowej A stanowią grupę, która jest rozszerzeniem afinicznym grupy ortogonalnych izometrii $\boldsymbol{O}(L)$ przestrzeni euklidesowej L stowarzyszonej z A.*

Dowód. Najpierw sprawdzimy, że dowolne odwzorowanie afiniczne $f: A \to A$, którego część liniowa $Df \in \boldsymbol{O}(L)$, jest ruchem. W istocie, na mocy definicji

$$d(f(a), f(b)) = |f(a) - f(b)| = |Df(a-b)| = |a-b| = d(a, b),$$

przy tym w trzeciej równości wykorzystaliśmy to, że $Df \in \boldsymbol{O}(L)$.

Większego wysiłku wymaga dowód odwrotnej części twierdzenia.

Po pierwsze, jest oczywiste, że złożenie ruchów jest ruchem. Co więcej, już wykazaliśmy, że przesunięcia są ruchami. Wybierzmy więc dowolny punkt $a \in A$ i niech f będzie ruchem. Oznaczmy $g = t_{a-f(a)} \circ f$ — jest to więc ruch zachowujący

punkt a. Wystarczy pokazać, że jest to odwzorowanie afiniczne i jego część liniowa $Dg \in \boldsymbol{O}(L)$. Jak w § 1, p. 12 wprowadźmy utożsamienie A z L za pomocą odwzorowania z tożsamościową częścią liniową, które przeprowadza punkt a na $0 \in L$. To pozwoli nam traktować g jak odwzorowanie $g: L \to L$ spełniające $g(0) = 0$ i $|g(l) - g(m)| = |l - m|$ dla każdych $l, m \in L$, w następstwie czego wystarczy wykazać, że to odwzorowanie należy do $\boldsymbol{O}(L)$.

Z początku sprawdzimy, że g zachowuje iloczyn skalarny. Istotnie, dla dowolnych $l, m \in L$

$$|l|^2 - 2(l, m) + |m|^2 = |l - m|^2 = |g(l) - g(m)|^2 = |g(l)|^2 - 2(g(l), g(m)) + |g(m)|^2,$$

skąd wynika żądany fakt wobec $|g(l)|^2 = |l|^2$, $|g(m)|^2 = |m|^2$. Teraz wykażemy addytywność g, tj. $g(l + m) = g(l) + g(m)$. Oznaczając $l + m = n$ i posługując się wykazaną powyżej własnością, dostajemy

$$\begin{aligned} 0 &= |n - l - m|^2 = |n|^2 + |l|^2 + |m|^2 - 2(n, l) - 2(n, m) + 2(l, m) = \\ &= |g(n)|^2 + |g(l)|^2 + |g(m)|^2 - 2(g(n), g(l)) - 2(g(n), g(m)) + 2(g(l), g(m)) = \\ &= |g(n) - g(l) - g(m)|^2, \end{aligned}$$

skąd wynika $g(n) = g(l) + g(m)$. Wreszcie wykażemy jednorodność g, tj. $g(xl) = xg(l)$ dla każdego $x \in \boldsymbol{R}$, $l \in L$. Przyjmując $m = xl$, otrzymujemy

$$\begin{aligned} 0 &= |m - xl|^2 = |m|^2 - 2x(m, l) + x^2|l|^2 = \\ &= |g(m)|^2 - 2x(g(m), g(l)) + x^2|g(l)|^2 = |g(m) - xg(l)|^2. \end{aligned}$$

Pokazaliśmy w ten sposób, że g jest odwzorowaniem liniowym i zachowuje iloczyn skalarny, a więc $g \in \boldsymbol{O}(L)$, co kończy dowód twierdzenia.

6. Twierdzenie. *Niech $f: A \to A$ będzie ruchem afinicznej przestrzeni euklidesowej, Df — jego częścią liniową. Wówczas istnieje taki wektor $l \in L$, że $Df(l) = l$ i $f = t_l \circ g$, przy czym $g: A \to A$ jest ruchem mającym punkt stały $a \in A$.*

Dowód. Wyjaśnijmy na początek geometryczne znaczenie tezy twierdzenia. Utożsamiając A z L za pośrednictwem odwzorowania afinicznego z tożsamościową częścią liniową, przeprowadzającego a na 0, widzimy, że f jest złożeniem odwzorowania ortogonalnego g i przesunięcia o wektor l niezmienniczy dla g (korzystamy z równości $Df = Dg$). Innymi słowy, f jest „ruchem śrubowym" w przypadku $\det g = 1$ albo ruchem śrubowym połączonym z odbiciem, jeśli $\det g = -1$. Istotnie, g jest całkowicie wyznaczone przez swoje obcięcie g_0 do $l^\perp$: $g = g_0 \oplus \mathrm{id}_{\boldsymbol{R}l}$, a więc g jest obrotem wokół osi $\boldsymbol{R}l$ (być może z odbiciem).

Przejdźmy do dowodu. Oznaczmy $L_2 = \mathrm{Ker}(Df - \mathrm{id}_L)$, $L_1 = L_2^\perp$, tak że $L = L_1 \oplus L_2$. L_2 zawiera wektory stałe względem Df, podprzestrzeń L_1 jest niezmiennicza względem $Df - \mathrm{id}_L$, bo Df jest ortogonalne i obcięcie $Df - \mathrm{id}_L$ do niej jest odwracalne.

Wybierzemy teraz dowolny punkt $a' \in A$ i oznaczymy $g' = t_{a'-f(a')} \circ f$. Mamy oczywiście $a' = g'(a')$. Przedstawmy $f(a') - a' = l_1 + l_2$, gdzie $l_1 \in L_1$, $l_2 \in L_2$, a wtedy $f = t_{l_2} \circ t_{l_1} \circ g'$ i $Df(l_2) = l_2$ na mocy definicji L_2. Wyznaczymy teraz taki $m \in L_1$, że $a = a' + m$ jest punktem stałym dla $g = t_{l_1} \circ g'$. Mamy

$$t_{l_1} \circ g'(a' + m) = g'(a' + m) + l_1 = a' + Df(m) + l_1.$$

Prawa strona tej równości jest równa $a' + m$ wtedy i tylko wtedy, gdy $(Df - \mathrm{id}_L)m + l_1 = 0$. Zauważyliśmy już jednakże, że na przestrzeni L_1 operator $Df - \mathrm{id}_L$ jest odwracalny, a ponieważ $l_1 \in L_1$, więc istnieje m o żądanej własności. W ten sposób otrzymaliśmy poszukiwany rozkład $f = t_{l_2} \circ g$, co kończy dowód.

Takie ruchy f, dla których $\det Df = 1$, nazywane są często *ruchami właściwymi*, a pozostałe ($\det Df = -1$) — *niewłaściwymi*. Poniżej zestawimy w poglądowy sposób zawarte w twierdzeniu p. 6 wiadomości o ruchach afinicznych przestrzeni euklidesowych o wymiarach $n \leq 3$. Zachowujemy oznaczenia użyte w tym twierdzeniu.

7. Przykłady. a) $n = 1$. Ponieważ $O(1) = \{\pm 1\}$, więc właściwymi ruchami są tylko przesunięcia. Jeśli f jest niewłaściwy, to $Df = -1$, a z $Df(l) = l$ wynika, że $l = 0$. Dlatego każdy ruch niewłaściwy prostej ma punkt stały, a więc jest odbiciem względem tego punktu.

b) $n = 2$. Ruch właściwy f z częścią liniową $Df = \mathrm{id}$ jest przesunięciem. Jeśli $Df \neq \mathrm{id}$ i $\det Df = 1$, to Df jako obrót nie ma niezerowych punktów stałych, więc f ma punkt stały i f jest obrotem względem niego.

Jeśli f jest ruchem niewłaściwym, to Df jest odbiciem płaszczyzny względem prostej, a f jest połączeniem takiego odbicia i przesunięcia wzdłuż tej prostej. Jeśli więc ruch niewłaściwy płaszczyzny ma punkt stały, to istnieje cała prosta złożona z punktów stałych, a on sam jest odbiciem w tej prostej.

c) $n = 3$. Jeśli $\det Df = 1$, to jedna z wartości własnych Df jest równa jedności i Df ma wektor stały. Zatem każdy ruch właściwy trójwymiarowej przestrzeni euklidesowej jest ruchem śrubowym wzdłuż pewnej osi (włączając w to przesunięcia, tj. zdegenerowane ruchy śrubowe z obrotem o kąt zerowy). Jest to tzw. twierdzenie Chasles'a.

Jeśli ruch $f = t_l \circ g$ jest niewłaściwy i $l \neq 0$, to obcięcie g do płaszczyzny ortogonalnej do l i przechodzącej przez punkt stały a tego ruchu jest ruchem niewłaściwym tej płaszczyzny. Jest ono zatem odbiciem w prostej położonej w tej płaszczyźnie. Oznaczmy przez P płaszczyznę rozpiętą przez tę prostą i wektor l. Wtedy $t_l \circ g$ jest złożeniem odbicia względem płaszczyzny P i przesunięcia o wektor l należący do P.

W końcu, jeśli $l = 0$, tj. f jest niewłaściwy i ma punkt stały, to utożsamiając ten punkt z zerem w L oraz f z Df i korzystając z istnienia prostej L_0 złożonej z wektorów własnych o wartości własnej -1, dochodzimy do geometrycznej charakteryzacji f jako złożenia obrotu w $L_0^{\perp}$ i odbicia w $L_0^{\perp}$.

Odwołując się do rozkładu biegunowego operatorów liniowych, możemy opisać także geometryczną strukturę dowolnego afinicznego odwzorowania odwracalnego afinicznej przestrzeni euklidesowej.

8. Twierdzenie. *Każde odwzorowanie afiniczne n-wymiarowej przestrzeni euklidesowej f można przedstawić jako złożenie trzech odwzorowań*: a) *dylatacji o dodatnich współczynnikach wzdłuż n parami ortogonalnych osi przechodzących przez pewien punkt $a_0 \in A$*, b) *ruchu z punktem stałym a_0*, c) *przesunięcia.*

Dowód. Składając f z odpowiednio dobranym przesunięciem i zastępując je tą superpozycją, tak jak w dowodzie twierdzenia p. 5, możemy przyjąć, że samo f ma punkt stały a_0. Utożsamiając następnie A z L i a_0 z zerem, możemy przedstawić $f = Df$ jako złożenie operatora symetrycznego dodatnio określonego z operatorem ortogonalnym. Sprowadzając pierwszy z tych operatorów na osie główne i przenosząc te osie do A, otrzymujemy tezę twierdzenia.

9. Na zakończenie odnotujemy to, że w tym paragrafie posługiwaliśmy się swobodnie podrozmaitościami liniowymi A (prostymi, płaszczyznami), określając je konstruktywnie jako przeciwobrazy podprzestrzeni liniowych w L przy rozmaitych utożsamieniach A z L zależnych od wyboru środka układu współrzędnych. Następny paragraf poświęcimy bardziej systematycznemu zbadaniu tych pojęć.

§ 3. Podprzestrzenie afiniczne

1. Definicja. Niech (A, L) oznacza przestrzeń afiniczną. Podzbiór $B \subset A$ nazywamy *podprzestrzenią afiniczną przestrzeni A*, jeśli jest pusty lub jeśli zbiór

$$M = \{b_1 - b_2 \in L \mid b_1, b_2 \in B\} \subset L$$

jest podprzestrzenią liniową w L i $t_m(B) \subset B$ dla każdego $m \in M$.

2. Uwagi. a) Jeśli warunki definicji są spełnione i B jest zbiorem niepustym, to para (B, M) jest przestrzenią afiniczną, co objaśnia użytą terminologię (przyjmujemy, że przesunięcie B o wektor z M otrzymamy, ograniczając do B to przesunięcie zadane na całym A). Istotnie, przejrzenie warunków nałożonych w definicji § 1, p. 1 od razu pokazuje, że są one spełnione dla pary (B, M). W szczególności, wybierając dowolny punkt $b \in B$, dostajemy $B = \{b + m \mid m \in M\}$.

b) Podprzestrzeń liniową $M = \{b_1 - b_2 \mid b_1, b_2 \in B\}$ przestrzeni L nazwiemy *przestrzenią kierunkową* podprzestrzeni afinicznej B. *Wymiarem B* nazywamy wymiar M. Jest jasne, że z warunku $B_1 \subset B_2$ wynika $M_1 \subset M_2$, a więc także $\dim B_1 \leq \dim B_2$. Podprzestrzenie afiniczne jednakowego wymiaru o tej samej przestrzeni kierunkowej nazwiemy *równoległymi*.

3. Stwierdzenie. *Podprzestrzenie afiniczne jednakowego wymiaru $B_1, B_2 \subset A$ są równoległe wtedy i tylko wtedy, gdy istnieje taki wektor $l \in L$, że $t_l(B_1) = B_2$. Dowolne dwa wektory o tej własności różnią się o wektor należący do (wspólnej) przestrzeni kierunkowej dla B_1 i B_2.*

Dowód. Jeśli $t_l(B_1) = B_2$ i M_1, M_2 są przestrzeniami kierunkowymi dla B_1 i B_2 odpowiednio, to

$$M_2 = \{a - b \mid a, b \in B_2\} = \{(a' + l) - (b' + l) \mid a', b' \in B\} = M_1,$$

tak że B_1 i B_2 są równoległe.

Odwrotnie, niech M oznacza wspólną przestrzeń kierunkową dla B_1 i B_2. Wybierzmy punkty $b_1 \in B_1$ i $b_2 \in B_2$. Ponieważ $B_1 = \{b_1 + l \mid l \in M\}$, $B_2 = \{b_2 + l \mid l \in M\}$, więc $B_2 = t_{b_2 - b_1}(B_1)$. A wreszcie, łatwo zobaczyć, że $t_{l_1}(B_1) = t_{l_2}(B_2)$ wtedy i tylko wtedy, gdy $l_1 - l_2 \in M$.

4. Wniosek. *Podprzestrzenie afiniczne w L (względem jej struktury afinicznej) to dokładnie podrozmaitości liniowe L w znaczeniu definicji części* 1, § 6, p. 1, *tj. przesunięcia podprzestrzeni liniowych.*

5. Wniosek. *Równoległe podprzestrzenie afiniczne jednakowego wymiaru albo są identyczne, albo nie przecinają się.*

Dowód. Jeśli $b \in B_1 \cap B_2$, to jak zauważyliśmy poprzednio $B_1 = \{b + m \mid m \in M\} = B_2$, gdzie M jest wspólną przestrzenią kierunkową B_1 i B_2.

6. Podprzestrzenie afiniczne B_1 i B_2 niekoniecznie jednakowego wymiaru nazywamy *równoległymi*, jeśli jedna z przestrzeni kierunkowych jest zawarta w drugiej. Niewiele zmieniając poprzedzające dowody łatwo wyprowadzić następujące fakty. Niech B_1 i B_2 będą równoległe i $\dim B_1 \leq \dim B_2$. Wówczas istnieje taki wektor $l \in L$, że $t_l(B_1) \subset B_2$, a dwa wektory o tej własności różnią się o wektor należący do M_1. Ponadto albo B_1 i B_2 nie mają punktów wspólnych, albo $B_1 \subset B_2$.

7. Stwierdzenie. *Niech (B_1, M_1), (B_2, M_2) będą dwiema podprzestrzeniami afinicznymi w A. Wówczas $B_1 \cap B_2$ albo jest puste, albo jest podprzestrzenią afiniczną o przestrzeni kierunkowej $M_1 \cap M_2$.*

Dowód. Niech $B_1 \cap B_2$ będzie niepuste i $b \in B_1 \cap B_2$, Wtedy $B_1 = \{b_1 + l_1 \mid l \in M_1\}$, $B_2 = \{b_1 + l_2 \mid l \in M_2\}$, skąd $B_1 \cap B_2 = \{b + l \mid l \in M_1 \cap M_2\}$, co trzeba było pokazać. (Oczywiście, stąd także wynika wniosek p. 5.)

8. Powłoki afiniczne. Niech $S \subset A$ będzie dowolnym zbiorem punktów przestrzeni afinicznej A. Najmniejszą podprzestrzeń afiniczną zawierającą S nazywamy *powłoką afiniczną* S. Taka podprzestrzeń istnieje i jest równa części wspólnej wszystkich podprzestrzeni afinicznych zawierających S. Można ją także scharakteryzować w terminach barycentrycznych kombinacji liniowych (stwierdzenie § 1, p. 11).

9. Stwierdzenie. *Powłoka afiniczna zbioru S jest równa zbiorowi kombinacji barycentrycznych elementów S:*

$$\widetilde{S} = \left\{ \sum_{i=1}^{n} x_i s_i \;\middle|\; \sum_{i=1}^{n} x_i = 1 \right\},$$

gdzie $\{s_1, \ldots, s_n\} \subset S$ przebiega wszystkie skończone podzbiory S.

Dowód. Najpierw wykażemy, że kombinacje barycentryczne stanowią podprzestrzeń afiniczną w A. W istocie, oznaczmy przez $M \subset L$ podprzestrzeń liniową rozpiętą przez wektory postaci $s - t, s, t \in S$. Każde dwie kombinacje barycentryczne punktów S można przedstawić w postaci kombinacji $\sum_{i=1}^{n} x_i s_i$, $\sum_{i=1}^{n} y_i s_i$ elementów tego samego zbioru $\{s_1, \ldots, s_n\}$, przyjmując jako ten zbiór sumę zbiorów odpowiadających wyjściowym kombinacjom, a dodatkowe współczynniki przyjmując równe zeru. Ponieważ $\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0$, różnicę tych kombinacji można przedstawić w postaci

$$\sum_{i=1}^{n}(x_i - y_i)(s_i - s_1),$$

która pokazuje, że należy ona do M. Ale i odwrotnie, dowolny element z M postaci $\sum_{i=1}^{n} x_i(s_i - t_i)$ można zapisać jako różnicę punktów $\sum_{i=1}^{n} x_i s_i + \left(1 - \sum_{i=1}^{n} x_i\right)s_1$ i $\sum_{i=1}^{n} x_i t_i + \left(1 - \sum_{i=1}^{n} x_i\right)s_1$ należących do $\tilde{S}$. A zatem $M = \{b_1 - b_2 \mid b_1, b_2 \in S\}$. Ten sam argument dowodzi, że $t_m(\tilde{S}) \subset \tilde{S}$ dla każdego $m \in M$, czyli $\tilde{S}$ jest istotnie podprzestrzenią afiniczną o przestrzeni kierunkowej M. Jest oczywiste, że $S \subset \tilde{S}$.

Odwrotnie, niech $B \supset S$ będzie dowolną podprzestrzenią, $\{s_1, \ldots, s_n\} \subset S$. Wówczas dla dowolnych $x_1, \ldots, x_n \in \boldsymbol{K}$, które spełniają $\sum_{i=1}^{n} x_i = 1$, mamy

$$\sum_{i=1}^{n} x_i s_i = s_1 + \sum_{i=1}^{n} x_i(s_i - s_1).$$

Ponieważ $s_1, \ldots, s_n \in B$, więc wektor $\sum_{i=1}^{n} x_i(s_i - s_1)$ należy do przestrzeni kierunkowej B, skąd wynika, że przesunięcie punktu s_1 o ten wektor należy do B. To oznacza, że $\tilde{S} \subset B$, więc $\tilde{S}$ jest rzeczywiście najmniejszą podprzestrzenią afiniczną zawierającą S.

10. Stwierdzenie. *Niech $f\colon A_1 \to A_2$ będzie odwzorowaniem afinicznym dwóch przestrzeni afinicznych, $B_1 \subset A_1$ i $B_2 \subset A_2$ — podprzestrzeniami afinicznymi. Wtedy $f(B_1) \subset A_2$, $f^{-1}(B_2) \subset A_1$ są podprzestrzeniami afinicznymi.*

Dowód. Niech $B_1 = \{b + l \mid l \in M_1\}$, gdzie M_1 oznacza przestrzeń kierunkową dla B_1. Wtedy $f(B_1) = \{f(b) + Df(l) \mid l \in M_1\} = \{f(b) + l' \mid l' \in Df(M_1)\}$, skąd wynika, że $f(B_1)$ jest podprzestrzenią afiniczną z przestrzenią kierunkową $Df(M_1)$.

W szczególności, $f(A_1)$ jest podprzestrzenią afiniczną w A_2, $B_2 \cap f(A_1)$ jest podprzestrzenią afiniczną i $f^{-1}(B_2) = f^{-1}(B_2 \cap f(A_1))$ na mocy znanych teoriomnogościowych zależności. Przyjmiemy, że f jest surjekcją, gdyż w przeciwnym przypadku można zastąpić A_2 podprzestrzenią $f(A_1)$ a B_2 $B_2 \cap f(A_1)$. Oznaczmy teraz przez M_2 przestrzeń kierunkową B_2. Wtedy $B_2 = \{b + l \mid l \in M_2\}$ i $f^{-1}(B_2) = \{b' + m' \mid f(b') = b,\ Df(m') \in M_2\}$. Po prawej stronie można ograniczyć się do jednej tylko

wartości $b' \in f^{-1}(b)$, bo pozostałe można otrzymać przesunięciami o wektory należące do Ker Df. Stąd wynika, że $f^{-1}(B_2)$ ma postać $\{b' + m \mid Df^{-1}(M_2)\}$, a więc jest podprzestrzenią afiniczną o przestrzeni kierunkowej $Df^{-1}(M_2)$.

11. Wniosek. *Poziomica dowolnej funkcji afinicznej jest podprzestrzenią afiniczną.*

Dowód. Istotnie, poziomice funkcji afinicznej $f: A \to \boldsymbol{K}^1$ są przeciwobrazami punktów $\boldsymbol{K}^1$, a punkt jest podprzestrzenią afiniczną dowolnej przestrzeni afinicznej (z przestrzenią kierunkową $\{0\}$).

12. Stwierdzenie. *Niech $f_1, \ldots, f_n$ będą funkcjami afinicznymi określonymi na przestrzeni afinicznej A. Wówczas zbiór $\{a \in A \mid f_1(a) = \cdots = f_n(a) = 0\}$ jest podprzestrzenią afiniczną A. Jeśli A jest skończenie wymiarowa, to każda jej podprzestrzeń afiniczna jest tej postaci.*

Dowód. Określony w tezie zbiór jest częścią wspólną skończonej liczby poziomic funkcji afinicznych i dlatego, na mocy wniosku p. 11 i stwierdzenia p. 7, jest podprzestrzenią afiniczną. Odwrotnie, niech $B \subset A$ będzie podprzestrzenią afiniczną w skończenie wymiarowej przestrzeni afinicznej A, a $M \subset L$ — odpowiadającą jej przestrzenią liniową. Jeśli B jest zbiorem pustym, to jest wyznaczony przez równanie $f = 0$ dla dowolnej niezerowej funkcji stałej f (oczywiście taka funkcja jest afiniczna i $Df = 0$). W przeciwnym przypadku niech $g_1 = \cdots = g_n = 0$ oznacza układ równań liniowych na przestrzeni L, dla którego M jest przestrzenią rozwiązań — jako $g_1, \ldots, g_n$ można wybrać bazę podprzestrzeni $M^\perp \subset L^*$. Wybierzmy punkt $b \in B$ i skonstruujmy funkcje afiniczne $f_i: A \to \boldsymbol{K}^1$ spełniające warunki $f_i(b) = 0$, $Df_i = g_i$, $i = 1, \ldots, n$. Mamy oczywiście $f_i(b + l) = g_i(l)$, a więc funkcja f_i zeruje się w punkcie $b + l \in A$ wtedy i tylko wtedy, gdy $l \in M$, tj. wtedy i tylko wtedy, gdy $b + l \in B$, co kończy dowód.

13. *Konfiguracją* w przestrzeni afinicznej A będziemy nazywać skończony układ uporządkowany podprzestrzeni afinicznych $\{B_1, \ldots, B_n\}$ w A. Dwie konfiguracje $\{B_1, \ldots, B_n\}$, $\{B'_1, \ldots, B'_n\}$ nazwiemy *afinicznie równoważnymi*, jeśli istnieje taki automorfizm afiniczny $f \in \boldsymbol{AF}(A)$, że $f(B_i) = B'_i$, $i = 1, \ldots, n$. Można także rozważać modyfikacje tej definicji, w których f wybiera się z pewnej podgrupy $\boldsymbol{AF}(A)$, np. grupy ruchów, gdy A jest euklidesowa. W takim przypadku konfiguracje przystające będziemy nazywać *metrycznie równoważnymi* (*przystającymi*). Pewne ważne pojęcia i rezultaty geometrii afinicznej są związane z wyznaczeniem niezmienników konfiguracji względem afinicznej równoważności. Odnotujmy, że to pojęcie jest afiniczną odmianą pojęcia „jednakowego rozmieszczenia", które badaliśmy w § 5 części 1.

Wykażemy niektóre z podstawowych wyników o afinicznej równoważności.

Niech A będzie n-wymiarową przestrzenią afiniczną. W zgodzie z wynikami p. p. 9–11, § 1 konfigurację $\{a_0, \ldots, a_n\}$ $n+1$ punktów nazwiemy *bazową* (*współrzędnościową*) lub *bazą punktową*), jeżeli A jest powłoką afiniczną zbioru $\{a_0, \ldots, a_n\}$.

14. Stwierdzenie. a) *Dowolne dwie konfiguracje bazowe są równoważne, a ponadto istnieje jedyne odwzorowanie* $f \in \boldsymbol{AF}(A)$, *które przeprowadza jedną z nich na drugą.*

b) *Konfiguracje bazowe* $\{a_0, \ldots, a_n\}$ *i* $\{a'_0, \ldots, a'_n\}$ *w przestrzeni euklidesowej* A *są izometryczne wtedy i tylko wtedy, gdy* $d(a_i, a_j) = d(a'_i, a'_j)$ *dla dowolnych* $i, j = 1, \ldots, n$.

Dowód. a) Oznaczmy $e_i = a_i - a_0$, $e'_i = a'_i - a'_0$. Układy $\{e_i\}$ i $\{e'_i\}$ stanowią bazy w L. Niech $g: L \to L$ będzie odwzorowaniem liniowym przeprowadzającym e_i na e'_i, a $f: A \to A$ — odwzorowaniem afinicznym spełniającym warunki $Df = g$ i $f(a_0) = a'_0$. Takie odwzorowanie istnieje na mocy stwierdzenia § 1, p. 11 i należy do $\boldsymbol{AF}(A)$, bo g jest nieosobliwe. Ponadto,

$$f(a_i) = f(a_0) + g(a_i - a_0) = a'_0 + e'_i = a'_0 + (a'_i - a'_0) = a'_i$$

dla wszystkich $i = 1, \ldots, n$. Ten wzór pokazuje także, że f jest jednoznacznie wyznaczone, jeśli Df ma przeprowadzać e_i na e'_i, a $f(a_0) = a'_0$.

b) Korzystając z już dowiedzionej części wystarczy wykazać, że f będzie ruchem wtedy i tylko wtedy, gdy $d(a_i, a_j) = d(a'_i, a'_j)$ dla wszystkich i, j. Rzeczywiście, $d(a_i, a_j) = |a_i - a_j| = |e_i - e_j|$, gdzie $e_0 = a_0 - a_0 = 0$, i podobnie $d(a'_i, a'_j) = |e'_i - e'_j|$, Jeśli f jest ruchem, to Df jest ortogonalne i zachowuje długości wektorów, więc warunek jest konieczny. Odwrotnie, przyjmijmy, że ten warunek jest spełniony. Wtedy $|e_i| = |e'_i|$ dla wszystkich $i = 1, \ldots, n$, a z równości $|e_i - e_j|^2 = |e'_i - e'_j|^2$ wnioskujemy, że $(e_i, e_j) = (e'_i, e'_j)$ dla wszystkich i, j. Wtedy odwzorowanie g, które przeprowadza $\{e_i\}$ na $\{e'_i\}$, jest izometrią, a stąd wynika, że f jest ruchem. Dowód zakończony.

Rozważmy teraz konfiguracje (b, B) składające się z punktu i podprzestrzeni afinicznej. W przypadku euklidesowym *odległością od* b *do* B nazwiemy liczbę

$$d(b, B) = \inf\{|l| \mid b + l \in B\}.$$

15. Stwierdzenie. a) *Konfiguracje* (b, B) *i* (b', B') *są afinicznie równoważne wtedy i tylko wtedy, gdy* $\dim B = \dim B'$ *oraz albo jednocześnie* $b \notin B$, $b' \notin B'$, *albo jednocześnie* $b \in B$, $b' \in B'$.

b) *Konfiguracje* (b, B) *i* (b', B') *są przystające wtedy i tylko wtedy, gdy* $\dim B = \dim B'$ *i* $d(b, B) = d(b', B')$.

Dowód. a) Warunki sformułowane w twierdzeniu są oczywiście konieczne. Załóżmy, że są spełnione i oznaczmy przez M, M' przestrzenie kierunkowe B i B' odpowiednio. Wybierzmy dalej taki automorfizm liniowy $g: L \to L$, że $g(M) = M'$. Jeśli $b \in B$ i $b' \in B'$, to oznaczymy przez $f: A \to A$ odwzorowanie afiniczne spełniające warunki $Df = g$ i $f(b) = b'$. Mamy oczywiście $f(b + l) = b + g(l)$, skąd wynika $f(B) = B'$.

W przypadku gdy $b \notin B$ i $b' \notin B'$, zażądamy spełnienia przez g dodatkowego warunku, a mianowicie, wybierając $a \in B$ i $a' \in B'$, zażądamy, aby g przeprowadzało

wektor $b-a$ na wektor $b'-a'$. Ponieważ oba te wektory są różne od zera, więc standardowa konstrukcja posługująca się bazami postaci {baza $M, b-a$, uzupełnienie} i {baza $M', b'-a'$, uzupełnienie} dowodzi istnienia g. Następnie skonstruujemy odwzorowanie afiniczne $f: A \to A$, dla którego $Df = g$ i $f(b) = b'$, oraz sprawdzimy, że także $f(B) = B'$. Po pierwsze mamy $f(a) = a'$, gdyż

$$f(a) = f(b-(b-a)) = f(b) - g(b-a) = b' - (b'-a') = a'.$$

Następnie $f(a+l) = f(a) + g(l)$, a ponieważ warunek $l \in M$ jest równoważny z warunkiem $g(l) \in M'$, więc $f(B) = B'$.

b) Znowu konieczność warunku jest oczywista. Dla wykazania dostateczności uzupełnimy konstrukcje z dowodu poprzedniego punktu dodatkowymi warunkami. Przede wszystkim utożsamimy A z L wybierając układ współrzędnych o środku w B. Wtedy B ulegnie utożsamieniu z M, a b stanie się pewnym wektorem L. Niech a oznacza rzut ortogonalny b na M. W kontekście przestrzeni liniowych wykazaliśmy już, że $d(b, B) = |b-a|$. W analogiczny sposób określimy punkt a' z przestrzeni M' lub, uwzględniając dokonane utożsamienie, z B'. Jako g weźmiemy izometrię L przeprowadzającą M na M' i b na b'. Taka izometria istnieje. Można się o tym przekonać, jeśli rozważyć uzupełnienia baz ortonormalnych w M i M' do baz ortonormalnych w L zawierających odpowiednio wektory $(b-a)/|b-a|$ i $(b'-a')/|b'-a'|$ i zdefiniować g jako izometrię przeprowadzającą pierwszą z tych baz na drugą. Odwzorowanie afiniczne $f: A \to A$, dla którego $Df = g$ i $f(b) = b'$, jest szukanym ruchem przeprowadzającym (b, B) na (b', B').

16. Jako ostatnie rozważymy konfiguracje składające się z dwóch podprzestrzeni B_1, B_2. Ich pełną klasyfikację względem afinicznej równoważności można otrzymać za pomocą analogicznego rezultatu dla podprzestrzeni liniowych, którego dowiedliśmy w § 5, p. 5, części 1. Pełna klasyfikacja metryczna jest dosyć skomplikowana i wymaga rozpatrzenia odległości między B_1 i B_2 oraz szeregu kątów. My ograniczymy się do rozpatrzenia tylko jednego niezmiennika metrycznego, a mianowicie odległości, którą jak zazwyczaj określimy za pomocą wzoru

$$d(B_1, B_2) = inf\{|b_1 - b_2| \mid b_1 \in B_1, b_2 \in B_2\}.$$

Wspólną prostopadłą do B_1, B_2 nazwiemy taką parę punktów $b_1 \in B_1$, $b_2 \in B_2$, że wektor $b_1 - b_2$ jest ortogonalny (prostopadły) do przestrzeni kierunkowych B_1 i B_2. Byłoby właściwiej nazywać tak odcinek $\{tb_1 + (1-t)b_2 \mid 0 \leq t \leq 1\}$.

17. Stwierdzenie. a) *Wspólna prostopadła do B_1 i B_2 zawsze istnieje. Zbiór takich wspólnych prostopadłych jest bijektywny z częścią wspólną przestrzeni kierunkowych B_1 i B_2.*

b) *Odległość B_1 od B_2 jest równa długości ich wspólnej prostopadłej $|b_1 - b_2|$.*

Dowód. a) Niech M_1, M_2 będą przestrzeniami kierunkowymi B_1 i odpowiednio B_2, i niech $b_1' \in B_1$, $b_2' \in B_2$. Zrzutujemy wektor $b_1' - b_2'$ prostopadle na $M_1 + M_2$

i zapiszmy ten rzut w postaci $m_1 + m_2$, $m_i \in M_i$. Dalej oznaczmy $b_1 = b_1' - m_1$, $b_2 = b_2' + m_2$. Mamy oczywiście $b_i \in B_i$ oraz

$$b_1 - b_2 = b_1' - b_2' - (m_1 + m_2) \in (M_1 + M_2)^\perp.$$

Pokazaliśmy więc, że $\{b_1, b_2\}$ jest wspólną prostopadłą do B_1, B_2.

Niech $\{b_1, b_2\}$, $\{b_1', b_2'\}$ będą dwiema wspólnymi prostopadłymi. Wtedy $b_1 - b_2 \in M_1$, $b_1' - b_2' \in M_2$ i ponadto

$$b_1 - b_2 \in (M_1 + M_2)^\perp, \quad b_1' - b_2' \in (M_1 + M_2)^\perp.$$

To oznacza, że różnica $(b_1 - b_1') - (b_2 - b_2')$ należy jednocześnie do $M_1 + M_2$ oraz do $(M_1 + M_2)^\perp$, a więc jest równa zeru. Stąd wynika, że $b_1 - b_1' = b_2 - b_2' \in M_1 \cap M_2$. Odwrotnie, jeśli $\{b_1, b_2\}$ jest ustaloną wspólną prostopadłą i $m \in M_1 \cap M_2$, to $\{b_1 + m, b_2 + m\}$ jest także wspólną prostopadłą. To kończy dowód pierwszej części stwierdzenia.

b) Niech $\{b_1, b_2\}$ będzie wspólną prostopadłą do B_1, B_2, a $b_1' \in B_1$, $b_2' \in B_2$ dowolną inną parą punktów. Wystarczy wykazać, że $|b_1 - b_2| \leq |b_1' - b_2'|$. Istotnie,

$$b_1' - b_2' = (b_1 - b_2) + (b_1' - b_1) + (b_2 - b_2').$$

Dalej mamy $(b_1' - b_1) + (b_2 - b_2') \in M_1 + M_2$, a wektor $b_1 - b_2$ jest ortogonalny do $M_1 + M_2$. Z twierdzenia Pitagorasa wynika

$$|b_1' - b_2'|^2 = |b_1 - b_2|^2 + |(b_1' - b_1) + (b_2 - b_2')|^2 \geq |b_1 - b_2|^2,$$

a to kończy dowód.

Na zakończenie wyprowadzimy pożyteczną charakteryzację podprzestrzeni afinicznych.

18. Stwierdzenie. *Podzbiór $S \subset A$ jest podprzestrzenią afiniczną wtedy i tylko wtedy, gdy wraz z każdymi dwoma punktami s, $t \in S$ zawiera on całą prostą przechodzącą przez te punkty, tj. ich powłokę afiniczną.*

Dowód. Prosta, przechodząca przez punkty s, $t \in S$, jest zbiorem $\{xs + (1-x)t \mid x \in \boldsymbol{K}\}$. Zatem konieczność sformułowanego warunku wynika ze stwierdzenia p. 9. Odwrotnie, niech ten warunek będzie spełniony. Ponieważ na mocy tego samego rezultatu powłoka afiniczna S składa się ze wszystkich barycentrycznych kombinacji punktów S, będziemy sprawdzać, że takie kombinacje $\sum_{i=1}^{n} x_i s_i$ należą do S. Przeprowadzimy indukcję względem n. Dla $n = 1, 2$ stwierdzenie jest oczywiste. Niech więc $n > 2$ i niech teza będzie spełniona dla liczb mniejszych od n. Przedstawimy $\sum_{i=1}^{n} x_i s_i$ w postaci

$$y_1 \sum_{i=1}^{n-2} \frac{x_i}{y_1} s_i + y_2 \sum_{i=n-1}^{n} \frac{x_i}{y_2} s_i,$$

gdzie $y_1 = \sum_{i=1}^{n-2} x_i$, $y_2 = x_{n-1} + x_n$ (możemy przyjąć, że obydwie sumy nie są równe zeru, bo w przeciwnym przypadku suma $\sum_{i=1}^{n} x_i s_i \in S$ na mocy założenia indukcyjnego). Mamy oczywiście

$$\sum_{i=1}^{n-2} \frac{x_i}{y_1} = \sum_{i=n-1}^{n} \frac{x_i}{y_2} = y_1 + y_2 = 1.$$

Stąd wynika, że obie sumy $\sum_{i=1}^{n-2} \frac{x_i}{y_1} s_i$, $\sum_{i=n-1}^{n} \frac{x_i}{y_2} s_i$ należą do S, a więc także ich barycentryczna kombinacja o współczynnikach y_1, y_2 należy do S. To kończy dowód.

Ćwiczenia

1. Określimy i-tą środkową układu punktów $a_1, \ldots, a_n \in A$ jako odcinek łączący punkt a_i ze środkiem ciężkości pozostałych punktów $\{a_j \mid i \neq j\}$. Dowieść, że wszystkie środkowe przecinają się w jednym punkcie — środku ciężkości układu $a_1, \ldots, a_n$.

2. Kąt między dwiema prostymi w afinicznej przestrzeni euklidesowej to kąt między ich przestrzeniami kierunkowymi. Dowieść, że dwie konfiguracje, złożone z dwóch prostych każda, są izometryczne wtedy i tylko wtedy, gdy kąty i odległości między prostymi w obu konfiguracjach są jednakowe.

3. Kąt między prostą a podprzestrzenią afiniczną wymiaru ≥ 1 to kąt między kierunkiem prostej a jego rzutem na przestrzeń kierunkową podprzestrzeni. Uogólnić wynik ćwiczenia 2 na konfiguracje złożone z prostej i podprzestrzeni.

§ 4. Wielościany wypukłe i programowanie liniowe

1. Sformułowanie zagadnienia. Podstawowe zadanie programowania liniowego można sformułować w następujący sposób. Jest dana skończenie wymiarowa przestrzeń afiniczna A nad ciałem liczb rzeczywistych $\boldsymbol{R}$ oraz $m+1$ funkcji afinicznych $f_1, \ldots, f_m$; $f: A \to \boldsymbol{R}$. Należy wyznaczyć punkt lub punkty $a \in A$, spełniające warunki $f_1(a) \geq 0, \ldots, f_m(a) \geq 0$, w których funkcja f przyjmuje największą możliwą przy tych warunkach wartość.

Do tego sformułowania można też sprowadzić ten wariant zagadnienia, w którym niektóre z nierówności są skierowane w przeciwną stronę, $f_i(a) \leq 0$, lub (i) poszukiwana jest najmniejsza dopuszczalna wartość funkcji — wystarczy zmienić znaki tej lub innych funkcji danych w zagadnieniu. Podobnie warunek $f_i(a) = 0$ jest równoważny z koniunkcją warunków $f_i(a) \geq 0$, $-f_i(a) \geq 0$. Można też przyjąć, że żadna z funkcji f_i nie jest równa stałej.

2. Motywacja. Rozważmy następujący matematyczny model produkcji. Załóżmy, że pewne przedsiębiorstwo wykorzystuje do produkcji n typów produktów m

rodzajów rozmaitych surowców. Ilości surowców i produktów w odpowiednich dla nich jednostkach wyrażane są nieujemnymi liczbami rzeczywistymi (pominiemy w naszych rozważaniach przypadek, gdy należy w tym celu używać liczb całkowitych, jak np. przy produkcji samochodów — jednakże przy dużych wielkościach produkcji i zapotrzebowania na surowce nawet taki przypadek można z dobrymi rezultatami przybliżać naszym „ciągłym" modelem).

Plan produkcji jest przedstawiony przez wektor $(x_1, \ldots, x_n) \in \boldsymbol{R}$, którego współrzędna x_j reprezentuje wyprodukowaną ilość j-tego produktu. Zastosujemy liniowy opis zużycia surowców w procesie produkcji, przyjmując, że do wyprodukowania jednostkowej ilości j-tego produktu zostaje zużyte a_{ij} jednostek ilości i-tego surowca, do wykonania planu produkcji $(x_1, \ldots, x_n)$ potrzeba $\sum_{j=1}^{n} a_{ij}x_j$ jednostek ilości i-tego surowca. Dostawy surowców do przedsiębiorstwa są opisane przez wektor $(b_1, \ldots, b_m)$ — i-ta współrzędna odpowiada otrzymaniu b_i jednostek i-tego surowca. Zatem plan może zostać tylko wtedy zrealizowany, gdy spełnione są warunki

$$f_i(x_1, \ldots, x_n) = b_i - \sum_{j=1}^{n} a_{ij}x_j \geq 0, \quad i = 1, \ldots, m.$$

Będziemy zawsze zakładać, że te warunki są niesprzeczne.

Przyjmijmy dalej, że przedsiębiorstwo pobiera cenę c_i za jednostkową ilość i-tego produktu. A zatem funkcja

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} c_i x_i$$

będzie wyrażać dochód tego przedsiębiorstwa. Plan produkcji $(x_1, \ldots, x_n)$ nazwiemy optymalnym, jeśli $f(x)$ osiąga największą wartość możliwą przy ograniczeniach $f_i \geq 0$, $x_i \geq 0$ $(i = 1, \ldots, n)$. Ten ostatni warunek oznacza, że przedsiębiorstwo nie uzupełnia produkowanych wyrobów z zewnątrz — przez zakupy lub z zapasów.

Widzimy, że zagadnienie wyboru optymalnego planu jest szczególnym przypadkiem zagadnienia sformułowanego w punkcie 1.

Jest zrozumiałe, że praktyczne zastosowania programowania liniowego wymagają opracowania konkretnych algorytmów wyszukiwania optymalnego planu, które można stosować bądź odręcznie, bądź przy użyciu maszyn liczących. Ograniczymy się tu do przedstawienia geometrycznych aspektów zagadnienia wspólnych dla konstrukcji takich algorytmów.

3. Podstawowe pojęcia geometryczne. Ustalmy skończenie wymiarową przestrzeń afiniczną A nad ciałem $\boldsymbol{R}$. Litera f z ewentualnymi indeksami będzie oznaczać funkcje afiniczne na A.

Półprzestrzenią nazwiemy zbiór punktów postaci $\{a \in A | f(a) \geq 0\}$, gdzie f jest funkcją afiniczną różną od stałej. *Wielościanem* nazywamy część wspólną skończonej liczby półprzestrzeni.

Przypomnijmy, że podzbiór $S \subset A$ nazywa się wypukłym, gdy $a_1, a_2 \in S$ i $0 \leq x \leq 1$ implikuje $xa_1 + (1-x)a_2 \in S$. Ponieważ $f(xa_1 + (1-x)a_2) = xf(a_1) + (1-x)f(a_2)$, więc każda półprzestrzeń jest wypukła. Dalej, każdy wielościan jest wypukły, co wynika stąd, że przecięcie dowolnej rodziny zbiorów wypukłych jest zbiorem wypukłym. Punkty $xa_1 + (1-x)a_2$ dla $0 < x < 1$ będziemy nazywać *punktami wewnętrznymi* odcinka o końcach a_1, a_2.

Niech S będzie zbiorem wypukłym. Podzbiór wypukły $T \subset S$ będziemy nazywać *ścianą* S, gdy dowolny odcinek o końcach należących do S, którego jakiś punkt wewnętrzny należy do T, jest cały zawarty w T. Cały zbiór S jest swoją własną ścianą. Ściana S składająca się z tylko jednego punktu nazywa się *wierzchołkiem* S. Czytelnik powinien odwołać się do wyobrażeń sześcianu, ośmiościanu czy kąta wielościennego w przestrzeni trójwymiarowej, aby nabrać poglądowego wyobrażenia o sytuacji rozważanej w zagadnieniach programowania liniowego. Dla tych figur ścianami — w znaczeniu naszej definicji — będą ściany, krawędzie i wierzchołki szkolnej geometrii oraz cała figura. Wierzchołkami kuli będą wszystkie punkty leżące na jej powierzchni.

Najważniejszym wynikiem tego paragrafu będzie wykazanie, że maksimum funkcji afinicznej na wielościanie ograniczonym (ten przypadek występuje najczęściej w zastosowaniach) jest osiągane na jednym z jego wierzchołków, których liczba jest skończona. Najpierw jednak zapoznamy się bliżej z budową wielościanów i ich ścian.

4. Lemat. *Część wspólna rodziny ścian i ściana ściany zbioru wypukłego S są ścianami S.*

Dowód. a) Niech $T = \cap T_i$, gdzie T_i są ścianami S. Dowolny odcinek o końcach w S, którego punkt wewnętrzny należy do T_i, jest zawarty w T_i. A zatem, jeśli jego punkt wewnętrzny należy do T, to i on sam należy do T.

b) Niech $T_1 \subset T \subset S$, gdzie T jest ścianą S. Dowolny odcinek o końcach należących do S, którego pewien punkt wewnętrzny należy do T_1, jest sam zawarty w T, bo T jest ścianą S. A więc jego końce należą do T, skąd wynika, że jest zawarty w T_1, bo T_1 jest ścianą T.

5. Lemat. *Niech S będzie wielościanem zdefiniowanym przez nierówności $f_i \geq 0$, $i = 1, \ldots, m$. Wówczas dla dowolnego i wielościan $S_i = S \cap \{a \mid f_i(a) = 0\}$ jest albo zbiorem pustym, albo ścianą S.*

Dowód. Niech S_i będzie niepusty, $a_1, a_2 \in S$ i niech wewnętrzny punkt odcinka $xa_1 + (1-x)a_2$ należy do S_i. Funkcja $x \mapsto f_i(xa_1 + (1-x)a_2)$ jest liniowa, zeruje się w punkcie $0 < x_0 < 1$, a ponadto jest nieujemna dla $x = 0$ i $x = 1$. Zatem musi być tożsamościowo zerem, skąd wynika, że cały odcinek należy do S_i.

6. Lemat. *Różna od stałej funkcja afiniczna f określona na wielościanie $S = \{a \mid f_i(a) \geq 0\}$ nie może przyjmować wartości maksymalnej w punkcie $a \in S$, w którym $f_i(a) > 0$ dla wszystkich i.*

Dowód. Ponieważ f nie jest stała, więc $Df \neq 0$. W przestrzeni wektorowej L stowarzyszonej z A wybierzmy wektor $l \in L$, taki że $Df(l) \neq 0$. Możemy przyjąć, że $Df(l) > 0$, zamieniając w razie potrzeby znak l. Jeśli liczba $\varepsilon > 0$ jest dostatecznie mała oraz $a \in S$, to $f_i(a + \varepsilon l) > 0$ dla wszystkich $i = 1, \ldots, m$ — wystarczy obrać $\varepsilon < \min_i \frac{f_i(a)}{|Df_i(l)|}$. Zatem dla takich ε mamy $a + \varepsilon l \in S$. Jednakże $f(a + \varepsilon l) = f(a) + \varepsilon Df(l) > f(a)$, więc $f(a)$ nie jest maksymalną wartością f.

Teraz możemy już wykazać nasz zasadniczy rezultat.

7. Twierdzenie. *Jeśli funkcja afiniczna f jest ograniczona z góry na wielościanie S, to osiąga swe wartości maksymalne w punktach należących do pewnej ściany S, która także jest wielościanem. Jeśli ponadto S jest ograniczony, to f osiąga wartość maksymalną w punkcie będącym wierzchołkiem S.*

Dowód. Przeprowadzimy indukcję względem wymiaru A. Przypadek $\dim A = 0$ jest oczywisty. Załóżmy więc, że $\dim A = n$ i że dla wymiarów mniejszych od n twierdzenie zostało wykazane. Niech S będzie zadany przez układ nierówności $f_1 \geq 0, \ldots, f_m \geq 0$. Ponieważ S jest zbiorem domkniętym, ograniczona z góry na tym zbiorze funkcja f osiąga wartość maksymalną w pewnym punkcie a. Jeśli $f_1(a) > 0, \ldots, f_m(a) > 0$, to na mocy lematu p. 6 f musi być funkcją stałą, a w szczególności swoją jedyną wartość przyjmuje na całym zbiorze S. Jeśli tak nie jest, to mamy $f_i(a) = 0$ dla pewnego i. To oznacza, że f osiąga swą wartość maksymalną w punkcie należącym do niepustego wielościanu S_i, będącego ścianą S i zawartego w podprzestrzeni afinicznej $\{a \mid f_i(a) = 0\}$ o wymiarze $n - 1$, gdyż f_i nie jest stała. Z założenia indukcyjnego wnioskujemy, że maksymalna wartość obcięcia f do S_i jest osiągana we wszystkich punktach należących do pewnej wielościennej ściany S_i. Na mocy lematu p. 4 jest to także ściana S, a to, że jest ona wielościanem, wynika stąd, iż jest wyznaczona przez układ nierówności, złożony z definiujących ją w S_i nierówności, po przedłużeniu ich lewych stron do A, oraz równoważnego z dwoma nierównościami równania $f_i = 0$.

Teraz przez indukcję względem wymiaru powłoki afinicznej S pokażemy, że dowolny wielościan ograniczony ma przynajmniej jeden wierzchołek. Dla wymiaru równego zeru to oczywiste. Rozważmy przypadek wymiaru większego od zera. Możemy uważać, że powłoka afiniczna S jest równa przestrzeni A. Rozpatrzmy dowolną różną od stałej funkcję afiniczną na A. Musi ona osiągać na S swą wartość maksymalną, bo S jest zbiorem domkniętym i ograniczonym, a zbiór punktów, w których to maksimum jest osiągane, jest niepustą ścianą S. Jest to ograniczony wielościan, o (ściśle) mniejszym wymiarze powłoki afinicznej. Na mocy założenia indukcyjnego zawiera wierzchołek, który jest także wierzchołkiem S wobec lematu p. 5.

Wreszcie, niech S będzie ograniczony, a T niech będzie wielościenną ścianą S, na której wyjściowa funkcja f osiąga swą wartość maksymalną. Wtedy dowolny wierzchołek T, którego istnienie właśnie wykazaliśmy, jest poszukiwanym wierzchołkiem S.

§ 5. Funkcje kwadratowe afiniczne i kwadryki

1. Definicja. Niech (A, L) będzie przestrzenią afiniczną nad ciałem $\boldsymbol{K}$. Odwzorowanie $Q\colon A \to \boldsymbol{K}$ nazywamy *funkcją kwadratową na* (A, L), jeśli istnieją: punkt $a_0 \in A$, forma kwadratowa $q\colon L \to \boldsymbol{K}$, forma liniowa $l\colon L \to \boldsymbol{K}$ oraz stała $c \in \boldsymbol{K}$, takie że dla wszystkich $a \in A$

$$Q(a) = q(a - a_0) + l(a - a_0) + c.$$

Forma q nazywana jest *częścią kwadratową* Q, a l — *częścią liniową* Q względem punktu a_0. Oczywiście, $c = Q(a_0)$. Przede wszystkim pokażemy, że „kwadratowość" Q nie zależy od wyboru punktu a_0. Dokładniej, niech symetryczna forma dwuliniowa g na L będzie formą biegunową q. Jak zazwyczaj zakładamy, że charakterystyka ciała K jest różna od 2.

2. Stwierdzenie. *Jeśli* $Q(a) = q(a-a_0)+l(a-a_0)+c$, *to dla dowolnego punktu* $a'_0 \in A$ *mamy*

$$Q(a) = q(a - a'_0) + l'(a - a'_0) + c',$$

gdzie

$$l'(m) = l(m) + 2g(m, a'_0 - a_0), \quad c' = Q(a'_0).$$

W szczególności, zamiana punktu a_0 *na* a'_0 *zmienia tylko część liniową i stałą w rozkładzie* Q.

Dowód. W istocie,

$$q(a - a_0) = q((a - a'_0) + (a'_0 - a_0)) =$$
$$= q(a - a'_0) + 2g(a_0 - a'_0, a'_0 - a_0) + q(a'_0 - a_0),$$

oraz

$$l(a - a_0) = l((a - a'_0) + (a'_0 - a_0)) = l(a - a'_0) + l(a'_0 - a_0),$$

czego należało dowieść.

3. Punkt a_0 będziemy nazywać *punktem środkowym* funkcji kwadratowej Q, jeśli część liniowa Q względem a_0 jest równa zeru. Wyjaśnieniem tej nazwy jest obserwacja, że punkt a_0 jest środkowy wtedy i tylko wtedy, gdy $Q(a) = Q(a_0 - (a - a_0))$ dla każdego a; istotnie, różnica między wyrażeniami po lewej i prawej stronie w ogólności jest równa $2l(a - a_0)$, bo $q(a - a_0) = q(-(a - a_0))$. Geometryczny sens tego warunku jest zawarty w obserwacji, że przy takim utożsamieniu A z L, przy którym punkt a_0 odpowiada środkowi układu, funkcja Q jest symetryczna względem odbicia $m \mapsto -m$.

Środkiem funkcji Q będziemy nazywać zbiór jej punktów środkowych.

4. Twierdzenie. a) *Jeśli część kwadratowa* q *funkcji* Q *jest niezdegenerowana, to środek* Q *zawiera tylko jeden punkt.*

b) *Jeśli q jest zdegenerowana, to albo środek Q jest pusty, albo jest podprzestrzenią afiniczną A o wymiarze* $\dim A - \mathrm{r}(q)$ ($\mathrm{r}(q)$ *oznacza rząd* q), *którego przestrzenią kierunkową jest jądro* q.

Dowód. Obierzmy dowolny punkt $a_0 \in A$ i zapiszmy Q w postaci $q(a-a_0)+l(a-a_0)+c$. Na mocy stwierdzenia p. 2 punkt $a_0' \in A$ będzie punktem środkowym Q wtedy i tylko wtedy, gdy równość

$$l(m) = -2g(m, a_0' - a_0)$$

jest spełniona dla wszystkich $m \in L$. Gdy a_0' przebiega (całą) przestrzeń A, wtedy wektor $a_0' - a_0$ przebiega całą przestrzeń L i wartościami funkcji liniowej argumentu $m \in M$ danej przez $-2g(m, a_0' - a_0)$ są wszystkie elementy przestrzeni L^* należące do obrazu kanonicznego odwzorowania $\tilde{g}: L \to L^*$ wyznaczonego przez formę g.

Jeśli q jest niezdegenerowana, to $\tilde{g}$ jest izomorfizmem. W szczególności, istnieje jednoznacznie wyznaczony wektor $a_0' - a_0 \in L$, którego obrazem jest $-\frac{1}{2}l \in L^*$, tj. $g(\cdot, a_0' - a_0) = -\frac{1}{2}l(\cdot)$. W tym przypadku punkt a_0' jest jedynym punktem środkowym funkcji Q.

Jeśli q jest zdegenerowana, to mamy dwie możliwości. Pierwsza, że $-\frac{1}{2}l$ nie należy do obrazu $\tilde{g}$, a wtedy Q nie ma punktów centralnych. Druga, że $-\frac{1}{2}l$ należy do obrazu $\tilde{g}$. W tym przypadku dla każdych dwóch punktów a_0', a_0'' spełniających

$$g(\cdot, a_0' - a_0) = g(\cdot, a_0' - a_0) = -\frac{1}{2}l(\cdot),$$

zachodzi $a_0' - a_0'' \in \operatorname{Ker}\tilde{g}$. Także odwrotnie, jeśli $g(\cdot, a_0' - a_0) = -\frac{1}{2}l(\cdot)$ oraz $a_0'' \in a_0' + \operatorname{Ker}\tilde{g}$, to

$$q(\cdot, a_0'' - a_0) = -\frac{1}{2}l(\cdot).$$

W ten sposób sprawdziliśmy, że środek Q jest podprzestrzenią afiniczną oraz że $\operatorname{Ker}\tilde{g}$, tj. jądro q, jest jego przestrzenią kierunkową, co kończy dowód.

Teraz jesteśmy w stanie wykazać twierdzenie o sprowadzeniu funkcji kwadratowej Q do postaci kanonicznej w odpowiednim afinicznym układzie współrzędnych $\{a_0; e_1, \ldots, e_n\}$, gdzie $\{e_i\}$ jest bazą L, $a_0 \in A$. Przypomnijmy, że punkt $a \in A$ jest w takim układzie reprezentowany przez wektor $(x_1, \ldots, x_n)$, gdy $a = a_0 + \sum_{i=1}^{n} x_i e_i$.

5. Twierdzenie. *Niech Q będzie funkcją kwadratową na przestrzeni afinicznej A. Wówczas istnieje taki afiniczny układ współrzędnych w A, w którym Q przyjmuje jedną z poniższych postaci*

a) *Jeśli q jest formą niezdegenerowaną, to*

$$Q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i x_i^2 + c, \quad \lambda_i,\ c \in \boldsymbol{K}.$$

b) *Jeśli q jest formą zdegenerowaną rzędu r oraz środek Q jest niepusty, to*

$$Q(x_1,\dots,x_n)=\sum_{i=1}^{r}\lambda_i x_i^2+c,\quad \lambda_i,\ c\in \boldsymbol{K}.$$

c) *Jeśli q jest formą zdegenerowaną rzędu r oraz środek Q jest pusty, to*

$$Q(x_1,\dots,x_n)=\sum_{i=1}^{r}\lambda_i x_i^2+x_{r+1},\quad \lambda_i\in \boldsymbol{K}.$$

Dowód. Jeśli q jest niezdegenerowana, to jako a_0 wybierzemy punkt środkowy Q. Wtedy $Q(a)=q(a-a_0)+c$. Jako $e_1,\dots,e_n$ wybierzemy bazę w L, w której q wyraża się sumą kwadratów ze współczynnikami. Taki sam wybór prowadzi do celu zawsze, jeśli tylko środek jest niepusty.

Jeśli środek Q jest pusty, to zaczniemy od wyboru dowolnego punktu a_0 i bazy $\{e_1,\dots,e_n\}$, w której część kwadratowa Q ma postać $\sum_{i=1}^{r}\lambda_i x_i^2$. Niech w tej bazie część liniowa wyraża się wzorem $l=\sum_{i=1}^{n}\mu_i x_i$. Twierdzimy, że $\mu_j\neq 0$ dla pewnego $j>r$. Rzeczywiście, w przeciwnym przypadku $l=\sum_{i=1}^{r}\mu_i x_i$ i możemy przedstawić Q w postaci

$$\sum_{i=1}^{r}\lambda_i x_i^2+\sum_{i=1}^{r}\mu_i x_i+c=\sum_{i=1}^{r}\lambda_i\Big(x_i+\frac{\mu_i}{2\lambda_i}\Big)^2+c'.$$

Stąd wynika, że $a_0-\sum_{i=1}^{r}\frac{\mu_i}{2\lambda_i}e_i$ jest punktem środkowym Q, co przeczy założeniu, że środek jest pusty.

Jeśli $\mu_j\neq 0$ dla pewnego $j>r$, to układ funkcjonałów $\{e^1,\dots,e^r,l\}$ jest liniowo niezależny. Można go uzupełnić do bazy L^* i wtedy w dualnej bazie przestrzeni L otrzymamy dla Q wyrażenie postaci $\sum_{i=1}^{r}\lambda_i x_i^2+x_{r+1}+c$, przy czym x_{r+1} jako funkcja na L jest równa l. Jest więc jasne, że istnieje punkt, w którym funkcja Q ma wartość zero, np. punkt, którego współrzędne w tym układzie są dane przez $x_1=\dots=x_r=0$, $x_{r+1}=-c$, $x_{r+2}=\dots=x_n=0$. Zaczynając konstrukcję od wyboru tego punktu, dochodzimy do wyrażenia Q w postaci $\sum_{i=1}^{r}\lambda_i x_i^2+x_{r+1}$.

6. Uzupełnienia. a) Zagadnienie jednoznaczności postaci kanonicznej sprowadza się do rozwiązanego już zadania o formach kwadratowych. Jeśli q jest niezdegenerowana i w pewnym układzie współrzędnych ma postać $\sum_{i=1}^{r}\lambda_i x_i^2+c$, to punkt $(0,\dots,0)$ jest środkiem, a więc jest wyznaczony jednoznacznie, stała c jest też wyznaczona jednoznacznie jako wartość Q w środku, a dowolność w wyborze osi i współczynników jest taka sama, jak dla form kwadratowych. W szczególności, nad $\boldsymbol{R}$ można przyjąć, że $\lambda_i=\pm 1$ i jedynym niezmiennikiem okaże się sygnatura. Nad $\boldsymbol{C}$ można przyjąć, że wszystkie $\lambda_i=1$.

W przypadku zdegenerowanym z niepustym środkiem, początek układu współrzędnych można wybrać w dowolnym punkcie środka, ale stała c jest mimo to wyznaczona jednoznacznie, bo wartość Q we wszystkich punktach środka jest ta sama. Istotnie, jeśli a, a_0 należą do środka, to $l(a-a_0)=0$ i $q(a-a_0)=0$, bo $a-a_0$ należy do jądra q. Do części kwadratowej można zastosować poprzednie uwagi.

W końcu, w zdegenerowanym przypadku i bez punktu środkowego początek układu można obrać w dowolnym punkcie, gdzie Q przyjmuje wartość zero, a do części kwadratowej można zastosować poprzednie uwagi.

b) Jeśli A jest euklidesową przestrzenią afiniczną, to Q przyjmuje postać kanoniczną w bazie ortonormalnej. Liczby $\lambda_1,\dots,\lambda_n$ są wyznaczone jednoznacznie. Dowolność w wyborze środka jest ta sama, co i w czysto afinicznym przypadku, a dowolność w wyborze osi jest ta sama, co dla form kwadratowych w przestrzeni liniowej euklidesowej.

7. Kwadryki afiniczne. *Kwadryką afiniczną* nazywamy zbiór postaci $\{a\in A \mid Q(a)=0\}$, gdzie Q jest dowolną funkcją kwadratową na A. Rzut oka na formy kanoniczne Q przekonuje, że do zbadania typów kwadryk można wykorzystać wszystkie wyniki § 10, części 2.

Rozpatrzmy problem jednoznaczności funkcji Q zadającej daną kwadrykę afiniczną nad ciałem $\boldsymbol{R}$. Po pierwsze, może się zdarzyć, że kwadryka jest podprzestrzenią afiniczną A (być może pustą): równanie $\sum_{i=1}^{r} x_i^2=0$ jest równoważne z układem równań $x_1=\dots=x_r=0$. Dla $r>1$ istnieje wiele nieproporcjonalnych funkcji kwadratowych zadających jedną i tę samą kwadrykę, np. $\sum_{i=1}^{r}\lambda_i x_i^2=0$ z dowolnymi $\lambda_i>0$. Pokażemy, że w pozostałych przypadkach odpowiedź jest prostsza.

8. Stwierdzenie. *Załóżmy, że kwadryka afiniczna X, nie będąca podprzestrzenią afiniczną, jest określona przez równania $Q_1=0$ oraz $Q_2=0$, gdzie Q_1, Q_2 są funkcjami kwadratowymi. Wówczas $Q_1=\lambda Q_2$ dla pewnego skalara $\lambda\in\boldsymbol{R}$.*

Dowód. Zauważmy przede wszystkim, że X nie redukuje się do punktu. Na mocy stwierdzenia § 3, p. 18 znajdziemy dwa punkty a_1, $a_2\in X$, których powłoka afiniczna, tj. prosta, nie zawiera się całkowicie w X.

Niech a_1, $a_2\in X$ i niech prosta przechodząca przez te punkty nie będzie zawarta w X. Wprowadźmy w A układ współrzędnych $\{a_1,e_1,\dots,e_n\}$, gdzie $e_n=a_2-a_1$, a następnie wyraźmy funkcję Q poprzez te współrzędne

$$Q_1(x_1,\dots,x_n)=\lambda x_n^2+l_1'(x_1,\dots,x_{n-1})x_n+l_1''(x_1,\dots,x_{n-1}),$$

gdzie l_1', l_1'' są funkcjami afiniczno-liniowymi, tj. wielomianami stopnia ≤ 1 zmiennych $x_1,\dots,x_{n-1}$. Ponieważ prosta przechodząca przez $a_1=(0,\dots,0)$ i przez $a_2=(0,\dots,0,1)$ nie zawiera się w całości w X, więc $\lambda\neq 0$ i $l_1'(0)^2-4\lambda l_1''(0)>0$. Dzieląc w razie potrzeby Q_1 przez λ możemy przyjąć, że $\lambda=1$. Analogicznie możemy przyjąć, że jako trójmian kwadratowy względem x_n

$$Q_2(x_1,\dots,x_n)=x_n^2+l_2'(x_1,\dots,x_{n-1})x_n+l_2''(x_1,\dots,x_{n-1})$$

oraz $l_2'(0)^2 - 4l_2''(0) > 0$. Wiemy więc teraz, że Q_1 i Q_2 mają jednakowe zbiory rzeczywistych pierwiastków i chcemy wykazać, że $Q_1 = Q_2$.

Ustalmy wektor $(c_1, \ldots, c_{n-1}) \in \boldsymbol{R}^{n-1}$ i rozpatrzmy wektory $(tc_1, \ldots, tc_{n-1})$, $t \in \boldsymbol{R}$. Dla małej wartości bezwzględnej t wyróżniki trójmianów względem x_n $Q_1(tc_1, \ldots, tc_{n-1}, x_n)$ i $Q_2(tc_1, \ldots, tc_{n-1}, x_n)$ pozostają dodatnie, a ich rzeczywiste pierwiastki, odpowiadające punktom przecięcia jednej i tej samej prostej z X, są identyczne. Stąd wynika, że $l_1' = l_2'$ i $l_1'' = l_2''$, w punktach $(tc_1, \ldots, tc_{n-1})$. Dlatego $l_1' \equiv l_2'$ i $l_1'' \equiv l_2''$, bo funkcje afiniczno-liniowe, które są identyczne na zbiorze otwartym, są tożsamościowo równe. W istocie, ich różnica zeruje się w pewnym otoczeniu początku układu i dlatego zbiór ich pierwiastków nie może być właściwą podprzestrzenią liniową. To kończy dowód.

§ 6. Przestrzenie rzutowe

1. Przestrzenie afiniczne otrzymujemy z przestrzeni liniowych przez „zatarcie początku układu współrzędnych". Przestrzenie rzutowe można otrzymać z przestrzeni liniowych co najmniej dwoma sposobami:

a) dołączając do przestrzeni afinicznej „punkty w nieskończoności",

b) traktując przestrzeń rzutową jako zbiór prostych w przestrzeni liniowej.

Jako punkt wyjścia wybierzemy konstrukcję b) — pokazuje ona wyraźniej jednorodność przestrzeni rzutowej.

2. Definicja. Niech L będzie przestrzenią liniową nad ciałem $\boldsymbol{K}$. Zbiór $P(L)$ prostych (tj. jednowymiarowych podprzestrzeni liniowych) w L nazywamy *przestrzenią rzutową* wyznaczoną przez L, a same proste w L nazywane są *punktami* $P(L)$.

Liczbę $\dim L - 1$ nazywamy wymiarem przestrzeni rzutowej $P(L)$ i oznaczamy $\dim P(L)$. Jednowymiarowe i dwuwymiarowe przestrzenie rzutowe nazywamy odpowiednio *prostą rzutową* i *płaszczyzną rzutową*. Przestrzeń rzutowa wymiaru n nad ciałem $\boldsymbol{K}$ jest także oznaczana przez $\boldsymbol{K}P^n$, $P^n(\boldsymbol{K})$ lub po prostu P^n. Sens umowy $\dim P(L) = \dim L - 1$ wyjaśnimy za chwilę.

3. Współrzędne jednorodne. Wybierzmy bazę $\{e_0, \ldots, e_n\}$ przestrzeni L. Każdy punkt $p \in P(L)$ jest jednoznacznie wyznaczony przez dowolny niezerowy wektor na odpowiadającej prostej w L. Współrzędne tego wektora nazywamy *współrzędnymi jednorodnymi punktu* p. Są one wyznaczone z dokładnością do pomnożenia przez niezerowy skalar — punkt $(\lambda x_0, \ldots, \lambda x_n)$ należy do tej samej prostej p i wszystkie jej punkty otrzymujemy w ten sposób. Ze względu na to wektor współrzędnych jednorodnych punktu p tradycyjnie oznaczany jest przez $(x_0 : x_1 : \ldots : x_n)$.

Z tego punktu widzenia *n-wymiarowa kartezjańska przestrzeń rzutowa* $P(\boldsymbol{K}^{n+1})$ jest zbiorem orbit grupy multyplikatywnej $\boldsymbol{K}^* = \boldsymbol{K} \setminus \{0\}$, względem działania tej grupy na zbiorze niezerowych wektorów $\boldsymbol{K}^{n+1} \setminus \{0\}$ zadanego wzorem $\lambda(x_0, \ldots, x_n) = (\lambda x_0, \ldots, \lambda x_n)$, a $(x_0 : x_1 : \ldots : x_n)$ jest oznaczeniem orbity zawierającej punkt $(x_0, \ldots, x_n)$.

Posługując się współrzędnymi jednorodnymi można kilkoma różnymi sposobami przedstawić strukturę P^n jako zbioru.

a) *Pokrycie afiniczne* P^n. Przyjmijmy

$$U_i = \{(x_0 : x_1 : \ldots : x_n) \mid x_i \neq 0\}, \quad i = 0, \ldots, n.$$

Oczywiście, $P^n = \bigcup_{i=0}^{n} U_i$. W klasie wektorów współrzędnych jednorodnych dowolnego punktu $p \in U_i$ znajduje się jedyny wektor o i-tej współrzędnej równej 1: $(x_0 : \ldots : x_i : \ldots : x_n) = (x_0/x_i : \ldots : 1 : \ldots : x_n/x_i)$. Pomijając jedynkę widzimy, że U_i jest bijektywne ze zbiorem $\boldsymbol{K}^n$, który możemy z kolei interpretować jako n-wymiarową kartezjańską przestrzeń liniową lub afiniczną. Jednakże podkreślmy, że dotychczas nie mamy żadnych powodów, aby uważać, że na U_i jest jakakolwiek naturalna, niezależna od wyboru współrzędnych struktura liniowa czy afiniczna. Niżej wykażemy, że na U_i można niezmienniczym sposobem wprowadzić tylko całą klasę struktur afinicznych, ale związanych kanonicznymi izomorfizmami, tak że geometrie konfiguracji afinicznych w dowolnej z tych struktur będą jednakowe.

Zbiór $U_i \cong \boldsymbol{K}^n$ nazwiemy *i-tą mapą afiniczną* P^n (względem zadanego układu współrzędnych). Punkty $(y_1^{(i)}, \ldots, y_n^{(i)}) \in U_i$ oraz $(y_1^{(j)}, \ldots, y_n^{(j)}) \in U_j$ dla $i \neq j$ odpowiadają temu samemu punktowi P^n, należącemu do części wspólnej $U_i \cap U_j$, wtedy i tylko wtedy, gdy wektory otrzymane po zastąpieniu przez 1 i-tej współrzędnej wektora $(y_1^{(i)}, \ldots, y_n^{(i)})$ oraz j-tej współrzędnej wektorza $(y_1^{(j)}, \ldots, y_n^{(j)})$ są proporcjonalne.

W szczególności $P^1 = U_0 \cup U_1$, $U_0 \cong U_1 \cong \boldsymbol{K}$; punkt $y \in U_0$ przy $y \neq 0$ odpowiada punktowi $1/y \in U_1$; punkt $y = 0 \in U_0$ nie należy do U_1, a punkt $1/y = 0$ ze zbioru U_1 nie należy do U_0. Naturalnie będzie przyjąć, że P^1 jest otrzymane z $U_0 \cong \boldsymbol{K}$ przez dołączenie jednego punktu o współrzędnej $y = \infty$. Uogólniając tę konstrukcję otrzymujemy.

b) *Rozkład symplicjalny* P^n. Przyjmijmy

$$V_i = \{(x_0 : \ldots : x_n) \mid x_j = 0 \text{ dla } j < i,\ x_i \neq 0\}.$$

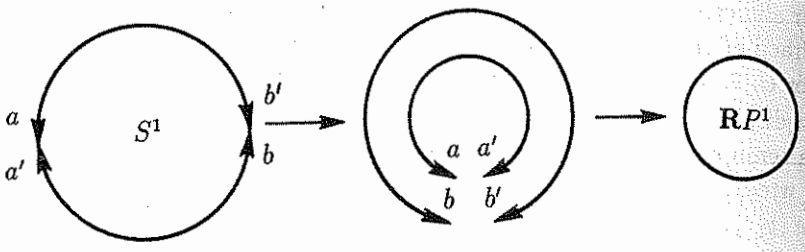
Oczywiście, $V_0 = U_0$ i $P^n = \bigcup_{i=0}^{n} V_i$, ale tym razem wszystkie V_i są parami rozłączne. W klasie współrzędnych jednorodnych dowolnego punktu $p \in V_i$ znajduje się jedyny reprezentant z jedynką na i-tym miejscu; pomijając tę jedynkę i poprzedzające ją zera otrzymujemy, bijekcję V_i z $\boldsymbol{K}^{n-i}$. Ostatecznie

$$P^n \cong \boldsymbol{K}^n \cup \boldsymbol{K}^{n-1} \cup \boldsymbol{K}^{n-2} \cup \ldots \boldsymbol{K}^0 \cong \boldsymbol{K}^n \cup P^{n-1}.$$

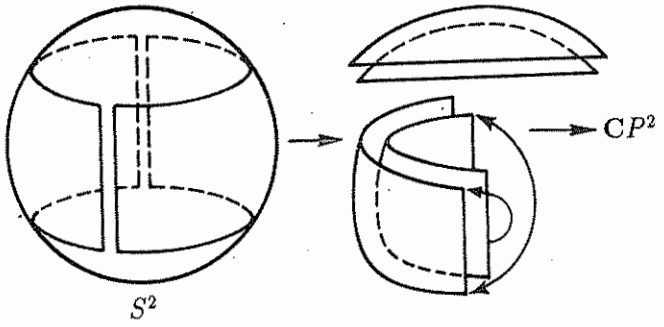
Inaczej mówiąc, przestrzeń P^n składa się z $U_0 \cong \boldsymbol{K}^n$ z dołączoną do niej nieskończenie oddaloną $(n-1)$-wymiarową przestrzeń rzutową, będącą zbiorem punktów o współrzędnych $(0 : x_1 : \ldots : x_n)$; ta ostatnia z kolei powstaje z podprzestrzeni afinicznej V_1 przez dołączenie do niej nieskończenie oddalonej (względem V_1) przestrzeni rzutowej P^{n-2}, itd.

c) *Przestrzenie rzutowe i sfery.* W przypadku $\boldsymbol{K} = \boldsymbol{R}$ lub $\boldsymbol{C}$ istnieje wygodny sposób normalizacji współrzędnych jednorodnych, nie wymagający wyboru niezerowej współrzędnej i dzielenia przez nią. Mianowicie dowolny punkt P^n można reprezentować poprzez współrzędne $(x_0 : \ldots : x_n)$ z dodatkowym warunkiem $\sum_{i=0}^{n} |x_i|^2 = 1$, tj. poprzez punkt n-wymiarowej (dla $\boldsymbol{K} = \boldsymbol{R}$) lub $(2n+1)$-wymiarowej (dla $\boldsymbol{K} = \boldsymbol{C}$) sfery euklidesowej. Pozostaje przy tym następująca niejednoznaczność: punkt $(\lambda x_0 : \ldots : \lambda x_n)$ należy do sfery jednostkowej wtedy i tylko wtedy, gdy $|\lambda| = 1$, tj. $\lambda = \pm 1$ dla $\boldsymbol{K} = \boldsymbol{R}$, $\lambda = e^{i\varphi}$, $0 \leq \varphi < 2\pi$ dla $\boldsymbol{K} = \boldsymbol{C}$.

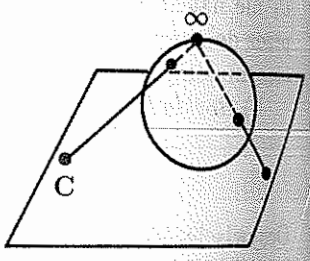
Inaczej mówiąc, n-wymiarowa rzeczywista przestrzeń rzutowa $\boldsymbol{R}P^n$ powstaje z n-wymiarowej sfery S^n poprzez utożsamienie par punktów antypodycznych (położonych na przeciwległych końcach jednej średnicy). W szczególności, $\boldsymbol{R}P^1$ wygląda jak okrąg (rys. 1), a $\boldsymbol{R}P^2$ — jak wstęga Mobiusa, do której na jej granicy doklejono koło (rys. 2).



Rys. 1



Rys. 2



Rys. 3

Trudniej „zobaczyć" $\boldsymbol{C}P^n$: całe koło wielkie sfery S^{2n+1}, złożone z punktów postaci $(x_0 e^{i\varphi} : \ldots : x_n e^{i\varphi})$ (gdzie φ parametryzuje punkty tego koła) zostaje sklejone w jeden punkt $\boldsymbol{C}P^n$. Porównując z przedstawieniem $\boldsymbol{C}P^1$ w postaci $C \cup \{\infty\}$ podanym w punkcie b), widzimy bez trudności, że $\boldsymbol{C}P^1$ można wyobrażać sobie jako dwuwymiarową sferę Riemanna, w której biegun północny odpowiada punktowi $\{\infty\}$, tak jak przy rzutowaniu stereograficznym (rys. 3). Zatem ten nowy model $\boldsymbol{C}P^1$ jako przestrzeni ilorazowej sfery S^3 pozwala odkryć interesujące odwzorowanie $S^3 \to S^2$, którego włóknami są okręgi S^1. Jest ono nazywane *projekcją Hopfa.*

W dyskusji tego punktu całkowicie pominęliśmy strukturę liniową, która była podstawą do konstrukcji $\boldsymbol{R}P^n$ i $\boldsymbol{C}P^n$, a za to wydobyliśmy własności topologiczne tych przestrzeni, a przede wszystkim ich zwartość. (Ściśle mówiąc, w definicji P^n nie było mowy o żadnej topologii — właśnie najwygodniej wprowadzać ją, używając odwzorowań sfer i przyjmując, że zbiorami otwartymi w $\boldsymbol{R}P^n$ i $\boldsymbol{C}P^n$ są te zbiory, których przeciwobrazy w S^n i S^{2n+1} są otwarte.) Dalej nie będziemy się posługiwać topologią i przejdziemy do badania geometrii liniowej przestrzeni rzutowych. Jednakże nie będzie przesadne stwierdzenie, że znaczenie $\boldsymbol{R}P^n$ i $\boldsymbol{C}P^n$ w dużej mierze pochodzi stąd, iż są one naturalnymi uzwarceniami $\boldsymbol{R}^n$ i $\boldsymbol{C}^n$, pozwalającymi na rozciągnięcie podstawowych cech struktury liniowej do nieskończoności. Nawet nad abstrakcyjnym ciałem $\boldsymbol{K}$, nie posiadającym żadnej topologii, ta „zwartość" przestrzeni rzutowych przejawia się szeregiem własności algebraicznych. Typowy przykład — na płaszczyźnie afinicznej dwie różne proste na ogół przecinają się w jednym punkcie, ale mogą także być równoległe. To oznacza, że punkt ich przecięcia „oddalił się do nieskończoności", a przy przejściu do płaszczyzny rzutowej przyjemnie zaobserwować jego powrót: dowolne dwie proste rzutowe w tej płaszczyźnie przecinają się.

Przejdźmy teraz do systematycznego zbadania geometrii przestrzeni P^n.

4. Podprzestrzenie rzutowe. Niech $M \subset L$ będzie dowolną podprzestrzenią liniową L. Wówczas $P(M) \subset P(L)$, bowiem każda prosta zawarta w M jest jednocześnie prostą zawartą w L.

Zbiory postaci $P(M)$ nazywamy *podprzestrzeniami rzutowymi w* $P(L)$. Oczywiście, $P(M_1 \cap M_2) = P(M_1) \cap P(M_2)$ i to samo zachodzi dla dowolnej rodziny podprzestrzeni. Stąd wynika, że rodzina podprzestrzeni rzutowych jest zamknięta względem operacji przecięcia. Zatem w zbiorze podprzestrzeni rzutowych $P(L)$ zawierających dany zbiór $S \subset P(L)$ istnieje najmniejsza — przecięcie wszystkich takich podprzestrzeni. Nazywamy ją *powłoką rzutową* S i oznaczamy $\overline{S}$; jest ona identyczna z $P(M)$, gdzie M jest powłoką liniową wszystkich prostych w L odpowiadających punktom $s \in S$. Przy przejściu od par $M \subset L$ do par $P(M) \subset P(L)$ wymiary zmniejszają się o jeden, a zatem *kowymiar* $\dim L - \dim M$ *jest równy* kowymiarowi $\dim P(L) - \dim P(M)$. Co więcej, jak już zauważyliśmy, $P(M_1 \cap M_2) = P(M_1) \cap P(M_2)$, a $P(M_1 + M_2)$ jest równe powłoce rzutowej $P(M_1) \cup P(M_2)$.

Wykorzystując te uwagi, możemy podać rzutowy wariant twierdzenia § 5, p. 3 części 1. Zauważmy jeszcze tylko, że w zgodzie z definicją w p. 2 jako wymiar pustej przestrzeni rzutowej należy przyjąć -1 — ten przypadek jest w pełni realistyczny, bo niepuste podprzestrzenie mogą mieć puste przecięcie.

5. Twierdzenie. *Niech* P_1, P_2 *będą dwiema skończenie wymiarowymi podprzestrzeniami rzutowymi w przestrzeni rzutowej* P. *Wówczas*

$$\dim P_1 \cap P_2 + \dim \overline{P_1 \cup P_2} = \dim P_1 + \dim P_2.$$

6. Przykłady. a) Niech P_1, P_2 będą różnymi punktami. Wtedy $\dim P_1 \cap P_2 = -1$, $\dim P_1 = \dim P_2 = 0$, skąd $\dim \overline{P_1 \cup P_2} = 1$, co oznacza, że powłoką rzutową dwóch punktów jest prosta. Zgodnie z definicją powłoki rzutowej, jest ona jedyną prostą rzutową, przychodzącą przez dwa punkty.

b) Załóżmy, że $\dim P_1 + \dim P_2 \geq \dim P$. Ze względu na $\dim \overline{P_1 \cup P_2} \leq \dim P$ mamy wtedy $\dim P_1 \cap P_2 \geq 0$. Mówiąc inaczej, dwie podprzestrzenie rzutowe, których suma ich wymiarów jest nie mniejsza niż wymiar przestrzeni je zawierającej, mają niepuste przecięcie. W szczególności, w płaszczyźnie rzutowej nie występują „równoległe" proste — dowolne dwie proste albo przecinają się w jednym punkcie, albo w dwóch, a wtedy (na podstawie przykładu a)) są identyczne. Analogicznie, dwie płaszczyzny rzutowe w trójwymiarowej przestrzeni rzutowej muszą albo przeciąć się wzdłuż prostej, albo być identyczne. Płaszczyzna rzutowa i prosta w trójwymiarowej przestrzeni rzutowej albo przecinają się w jednym punkcie, albo prosta leży w płaszczyźnie.

c) Warunek $P_1 \cap P_2 = \emptyset$ w przypadku $P_i = P(M_i)$ oznacza, że $M_1 \cap M_2 = \{0\}$, tj. że $M_1 + M_2$ jest sumą prostą.

7. Definiowanie podprzestrzeni rzutowych za pomocą równań. Funkcjonał liniowy $f: L \to \boldsymbol{K}$ na przestrzeni liniowej *nie definiuje* żadnej funkcji na $P(L)$ (poza przypadkiem $f \equiv 0$), bo zawsze znajdzie się prosta w L, na której funkcja nie jest stała i nie ma sposobu zadania jej wartości w odpowiadającym punkcie $P(L)$. Jednakże *równanie* $f = 0$ określa podprzestrzeń liniową w L, a więc także podprzestrzeń rzutową w $P(L)$. Jeśli L jest skończenie wymiarowa, to dowolną podprzestrzeń w L można zadać za pomocą układu równań

$$f_1 = \cdots = f_m = 0.$$

Używając współrzędnych jednorodnych w P^n możemy to zjawisko wyrazić następująco — jednorodny układ równań liniowych

$$\sum_{j=0}^{n} a_{ij} x_j = 0, \quad i = 1, \ldots, m,$$

definiuje podprzestrzeń rzutową w P^n, składającą się z punktów o współrzędnych jednorodnych $(x_0 : \ldots : x_n)$, które są rozwiązaniami tego układu. Pomnożenie współrzędnych przez λ przeprowadza rozwiązanie tego układu na rozwiązanie.

8. Podprzestrzenie afiniczne i hiperpłaszczyzny. Niech $M \subset L$ będzie podprzestrzenią liniową kowymiaru 1. Wówczas $P(M) \subset P(L)$ też jest podprzestrzenią kowymiaru 1 — takie podprzestrzenie będziemy nazywać *hiperpłaszczyznami*.

Pokażemy teraz jak wprowadzić na uzupełnieniu A_M hiperpłaszczyzny $P(M)$ strukturę przestrzeni afinicznej $(A_M, M, +)$. Wybierzmy w L rozmaitość liniową $M' = m' + M$, nie przechodzącą przez początek układu współrzędnych. Ma ona naturalną strukturę przestrzeni afinicznej — przesunięcie o $m \in M$ w M' jest indukowane przez przesunięcie o m w L, tj. polega na dodawaniu m.

Z drugiej strony, istnieje naturalna bijekcja A_M z M': punkty A_M są prostymi nie leżącymi w M, a więc każda z nich przecina M' w jednoznacznie wyznaczonym punkcie, który przyporządkowujemy wyjściowemu punktowi A_M. W ten sposób możemy otrzymać każdy punkt M' i to tylko jednokrotnie. Strukturę afiniczną rozmaitości M' przenosimy za pomocą tej bijekcji na A_M. Jednakże niejednoznaczność wyboru M' prowadzi do niejednoznaczności tak określonej struktury afinicznej na A_M. Dla porównania dwóch takich struktur wykażemy, że odwzorowanie tożsamościowe A_M na siebie jest izomorfizmem afinicznym względem tych struktur.

9. Stwierdzenie. *Niech* $(A_M, M, +')$ *i* $(A_M, M, +'')$ *będą dwiema strukturami afinicznymi na* A_M, *otrzymanymi za pomocą powyższej konstrukcji. Wówczas odwzorowanie tożsamościowe* A_M *na siebie jest izomorfizmem afinicznym, którego część liniowa jest pewną homotetią* M.

Dowód. Załóżmy, że dwie dane struktury afiniczne odpowiadają podrozmaitościom $m' + M$ i $m'' + M$. Warstwy $m' + M$ i $m'' + M$ należące do jednowymiarowej przestrzeni ilorazowej L/M są proporcjonalne, można więc przyjąć, że $m'' = am'$, $a \in K$. Mnożenie przez a w przestrzeni L przeprowadza $m' + M$ na $m'' + M$ i indukuje tożsamościowe odwzorowanie $P(L)$ w siebie, więc także i A_M w siebie. Z drugiej strony przesunięcie o wektor $m \in M$ w warstwie $m' + M$ homotetia przeprowadza na przesunięcie o wektor $am \in M$ w warstwie $m'' + M$. To dowodzi stwierdzenia.

10. Wniosek. *Zbiór podprzestrzeni afinicznych w* A_M *z ich relacjami incydencji, a także zbiory odwzorowań afinicznych* A_M *w inne przestrzenie afiniczne nie zależą od dowolności w wyborze struktury afinicznej* A_M.

Ten wniosek wprowadza możliwość traktowania dopełnienia dowolnej hiperpłaszczyzny w przestrzeni rzutowej po prostu jako przestrzeni afinicznej bez konieczności dalszego precyzowania.

Rozważmy teraz, jak wygląda przestrzeń rzutowa $P(M)$ „z punktu widzenia" przestrzeni afinicznej A_M.

11. Stwierdzenie. *Między punktami* $P(M)$ *a klasami prostych równoległych w* A_M *istnieje bijektywna odpowiedniość. Inaczej mówiąc, każdy punkt* $P(M)$ *jest „kierunkiem odchodzenia do nieskończoności" w* A_M.

Dowód. Utożsamimy A_M z $m' + M$. Klasa prostych równoległych w $m' + M$ jest wyznaczona przez swoją przestrzeń kierunkową w M, tj. przez punkt w $P(M)$, i ta odpowiedniość jest bijekcją.

12. W rzeczywistości zachodzi mocniejsza własność: każda prosta l w A_M jednoznacznie wyznacza zawierającą ją prostą w $P(L)$, a mianowicie swoją powłokę rzutową $\bar{l}$. Powłokę rzutową otrzymujemy dołączając do l jeden jedyny punkt, ten właśnie, który należy do $P(M)$ i jest „punktem w nieskończoności" tej prostej. Cała klasa prostych równoległych w A_M ma wspólny „punkt w nieskończoności" należący do $P(M)$. Przy utożsamieniu A_M z $m'+M$ powłoka $\bar{l}$ odpowiada wszystkim prostym

tej płaszczyzny w L, która przechodzi przez l i przestrzeń kierunkową l, a punkt w nieskończoności $\overline{l}$ — to sama przestrzeń kierunkowa.

Ogólniej, niech $A \subset A_M$ będzie dowolną podprzestrzenią afiniczną. Wówczas jej powłoka rzutowa $\overline{A}$ w $P(L)$ ma następujące własności:

a) $\overline{A} \setminus A \subset P(M)$ — dołączone zostają tylko punkty w nieskończoności;

b) $\dim A = \dim \overline{A}$;

c) $\overline{A} \setminus A$ jest podprzestrzenią rzutową w $P(M)$ wymiaru $\dim A - 1$. (Z tego względu $\overline{A}$ nazywa się także rzutowym domknięciem A.)

Utożsamienie A_M z $m' + M$ sprowadza sprawdzenie tych własności do prostego zastosowania definicji. Istotnie, $\overline{A}$ składa się z prostych należących do powłoki liniowej $A \subset m' + M$. Ta powłoka liniowa jest rozpięta przez przestrzeń kierunkową L_0 podprzestrzeni A i dowolny wektor należący do A. Zatem jej wymiar jest równy $\dim L_0 + 1 = \dim A + 1$, skąd otrzymujemy $\dim A = \dim \overline{A}$. Wszystkie proste tej powłoki liniowej przecinają się z $m' + M$, tj. odpowiadają punktom A, z wyjątkiem prostych należących do przestrzeni kierunkowej L_0. Te ostatnie należą do $P(M)$ i tworzą przestrzeń rzutową o wymiarze $\dim L_0 - 1 = \dim A - 1$.

§ 7. Dwoistość rzutowa i kwadryki rzutowe

1. Niech L będzie przestrzenią liniową nad ciałem $\boldsymbol{K}$, L^* — przestrzenią dwoistą do L, tj. przestrzenią funkcjonałów liniowych na L. Przestrzeń rzutową $P(L^*)$ nazywamy *dwoistą do przestrzeni rzutowej* $P(L)$.

Każdy punkt $P(L^*)$ jest prostą $\{\lambda f\}$ w przestrzeni funkcjonałów liniowych na L. Hiperpłaszczyzna $f = 0$ w $P(L)$ nie zależy od wyboru funkcjonału f z tej prostej i jednoznacznie określa całą prostą. Można to wyrazić mówiąc, że *punktami dwoistej przestrzeni rzutowej są hiperpłaszczyzny wyjściowej przestrzeni rzutowej.*

Jeśli w L i L^* zostały wybrane bazy dwoiste i odpowiednie układy współrzędnych jednorodnych w $P(L)$ i $P(L^*)$, to ta odpowiedniość przybiera prostą postać: hiperpłaszczyźnie w $P(L)$ opisanej równaniem

$$\sum_{i=0}^{n} a_i x_i = 0$$

odpowiada punkt przestrzeni $P(L^*)$ o współrzędnych jednorodnych $(a_0 : \ldots : a_n)$. Izomorfizm kanoniczny $L \to L^{**}$ dowodzi symetrii relacji dwoistości między tymi dwiema przestrzeniami rzutowymi. Ogólniej, formułując wyniki § 7 części 1 w języku rzutowym, otrzymamy następującą relację dwoistości między układami podprzestrzeni rzutowych w $P(L)$ i $P(L^*)$ (przyjmujemy w dalszym ciągu, że przestrzeń L jest skończenie wymiarowa).

a) Podprzestrzeni $P(M) \subset P(L)$ odpowiada dwoista do niej podprzestrzeń $P(M^\perp) \subset P(L^*)$, a przy tym

$$\dim P(M) + \dim P(M^\perp) = \dim P(L) - 1.$$

b) Przecięciu podprzestrzeni rzutowych odpowiada powłoka rzutowa dwoistych do nich, a powłoce rzutowej — przecięcie. W szczególności, relacji incydencji dwóch podprzestrzeni (tj. zawieraniu się jednej w drugiej) odpowiada relacja incydencji.

To pozwala na sformułowanie następującej *zasady dwoistości rzutowej*, będącej, właściwie mówiąc, *zasadą metamatematyczną*, ponieważ jest ona stwierdzeniem dotyczącym języka geometrii rzutowej.

2. Zasada dwoistości rzutowej. *Załóżmy, że jest prawdziwe twierdzenie o konfiguracjach podprzestrzeni rzutowych w przestrzeniach rzutowych, którego sformułowanie zawiera jedynie własności wymiarów, incydencji, tworzenia przecięć i powłok rzutowych. Wówczas twierdzenie dwoiste, w którym wszystkie te terminy zostały zastąpione terminami do nich dwoistymi zgodnie z zasadami poprzedniego punktu, jest także prawdziwym twierdzeniem geometrii rzutowej.*

Prosty przykład: twierdzeniem dwoistym do twierdzenia „dwie różne płaszczyzny w trójwymiarowej przestrzeni rzutowej przecinają się wzdłuż jednej prostej" jest następujące „przez dwa różne punkty w trójwymiarowej przestrzeni rzutowej przechodzi jedna prosta". (W paragrafie 9 zapoznamy się z o wiele więcej mówiącymi twierdzeniami o konfiguracjach rzutowych.)

3. Dwoistość rzutowa i kwadryki. Izomorfizm przestrzeni liniowej L z przestrzenią dwoistą L^* pozwala utożsamić $P(L^*)$ z $P(L)$, a wówczas odwzorowanie dwoistości pomiędzy podprzestrzeniami rzutowymi $P(L)$ i $P(L^*)$ można traktować jako *odwzorowanie dwoistości pomiędzy podprzestrzeniami* $P(L)$.

Zadanie izomorfizmu $L \to L^*$ jest równoważne z zadaniem niezdegenerowanego iloczynu skalarnego $g: L \times L \to \boldsymbol{K}$. Rozpatrzymy bardziej szczegółowo geometrię dwoistości rzutowej w przypadku, gdy iloczyn skalarny g jest symetryczny. Jak zazwyczaj, będziemy zakładać, że charakterystyka ciała $\boldsymbol{K}$ jest różna od 2. Wówczas g jest jednoznacznie wyznaczone przez formę kwadratową $q(l) = g(l, l)$.

Równanie $q(l) = 0$ wyznacza kwadrykę Q_0 w L. Jej obraz Q w $P(L)$ będziemy także nazywać kwadryką, a w zastosowaniu do teorii dwoistości — *kwadryką biegunową*. Zauważmy, że Q_0 jest stożkiem o środku w początku układu: jeśli $l \in Q_0$, to cała prosta $\boldsymbol{K}l$ zawiera się w Q_0. Utożsamienie $P(L)$ z nieskończenie oddalonymi punktami przestrzeni L prowadzi do utożsamienia Q z podstawą stożka Q_0.

Zgodnie z ogólną teorią g i q określają odwzorowanie dwoistości zbioru podprzestrzeni rzutowych $P(L)$ w siebie; hiperpłaszczyzna w $P(L)$ dwoista do punktu $p \in P(L)$ jest nazywana *biegunową punktu* p (względem q lub Q). Aby zaznajomić się z geometrycznym znaczeniem tego odwzorowania, wyprowadzimy najpierw równanie hiperpłaszczyzny biegunowej we współrzędnych jednorodnych. Zaczniemy rozważania od przestrzeni L. Niech równanie Q_0 ma postać

$$q(x_0, \ldots, x_n) \equiv \sum_{i,j=0}^{n} a_{ij} x_i x_j = 0, \quad a_{ij} = a_{ji}.$$

Związany z q izomorfizm $L \to L^*$ przyporządkowuje punktowi $(x_0^0, \ldots, x_n^0) \in L$

funkcję liniową $(x_0, \ldots, x_n) \mapsto \sum_{i,j=0}^{n} a_{ij} x_i^0 x_j$ na L. Zatem równanie hiperpłaszczyzny biegunowej ma postać

$$\sum_{i,j=0}^{n} a_{ij} x_i^0 x_j = 0.$$

W szczególności, jeśli $(x_0^0, \ldots, x_n^0) \in Q$, to *hiperpłaszczyzna biegunowa do danego punktu zawiera ten punkt.* Co więcej, w takim przypadku jej równanie można zapisać w postaci

$$\sum_{j=1}^{n} \frac{\partial q}{\partial x_j}(x_0^0, \ldots, x_n^0)(x_j - x_j^0) = \sum_{i,j=0}^{n} a_{ij} x_i^0 (x_j - x_j^0) = 0.$$

W elementarnej geometrii analitycznej (nad $\boldsymbol{R}$) to równanie określa *hiperpłaszczyznę styczną do* Q_0 w należącym do niej punkcie $(x_0^0, \ldots, x_n^0)$. To uzasadnia wprowadzenie następującego określenia.

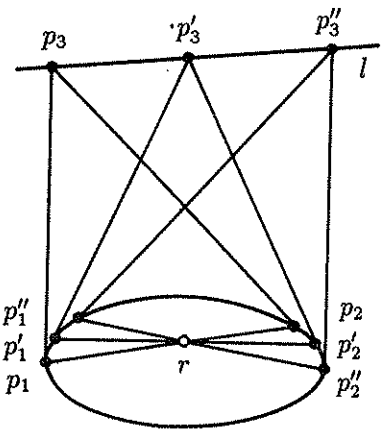
4. Definicja. *Hiperpłaszczyzną styczną do kwadryki niezdegenerowanej* $Q \subset P(L)$ *w punkcie* $p \in Q$ nazywamy hiperpłaszczyznę biegunową do p względem formy kwadratowej q wyznaczającej Q.

Korzystając z ogólnych własności dwoistości rzutowej możemy teraz natychmiast odtworzyć geometrycznie znaczącą część relacji dwoistości i wyprowadzić szereg pięknych i nieoczekiwanych twierdzeń geometrycznych, jednocześnie ilustrując je przykładami. Przez Q będziemy w dalszym ciągu oznaczać (niezdegenerowaną) kwadrykę w P^2 lub w P^3.

a) Niech $Q \subset P^2$, a p_1, p_2 niech będą dwoma punktami kwadryki, p_3 — punktem przecięcia stycznych do Q w p_1 i p_2. Zgodnie z ogólną zasadą dwoistości punkt p_3 odpowiada wtedy prostej p_1p_2 przechodzącej przez p_1 i p_2, tj. powłoce rzutowej p_1 i p_2.

Przyjmijmy, że punkt p_3 *przemieszcza się wzdłuż prostej* l; *przeprowadźmy z każdego punktu prostej dwie styczne do* Q *i połączmy punkty styczności. Wówczas wszystkie w ten sposób otrzymane „cięciwy" Q przetną się w jednym punkcie* r, *który odpowiada prostej* l *na mocy dwoistości.* Jeszcze raz podkreślimy, że dla dowodu nie potrzebne są nam żadne rachunki: wynika on wprost z tego, że zgodnie z ogólną zasadą dwoistości powłoka rzutowa punktów $p_3, p_3', p_3'', \ldots$ jest dwoista do przecięcia dwoistych z nimi prostych, którymi są właśnie rozważane cięciwy.

Jeden szczegół zasługuje na specjalną uwagę. Przecięcia par stycznych do punktów Q mogą nie zapełniać całej płaszczyzny. Na przykład dla elipsy w $\boldsymbol{R}P^2$, pokazanej na rys. 4 (narysowaliśmy, oczywiście, tylko część afinicznej mapy $\boldsymbol{R}P^2$), otrzymamy jedynie *zewnętrze* elipsy. Jak znaleźć te proste, które odpowiadają wewnętrznym punktom elipsy? Rysunek 4 podpowiada rozwiązanie: na mocy symetrii dwoistości należy przeprowadzić przez wewnętrzny punkt r pęk cięciw Q, a następnie wyznaczyć punkty przecięcia stycznych do Q w przeciwległych końcach tych cięciw — to właśnie te punkty zakreślą dwoistą z punktem r prostą l.



Rys. 4

Opis odwzorowania dwoistości otrzymany w ten sposób staje się niejednorodny. Mamy bowiem dwa sposoby konstrukcji prostej biegunowej l dla punktu r.

1) Jeśli punkt r leży zewnątrz elipsy Q (albo na niej), to należy poprowadzić z punktu r dwie styczne do Q (albo jedną) i połączyć punkty styczności prostą l (albo wziąć styczną l).

2) Jeśli punkt r leży wewnątrz elipsy Q, to należy poprowadzić wszystkie proste przechodzące przez r i znaleźć punkty przecięcia stycznych do Q wyprowadzonych z punktów przecięcia z Q prostych wychodzących z r. Miejsce geometryczne tych punktów jest prostą dwoistą dla punktu r.

Jak się okazuje, przyczyną tej trudności jest to, że ciało podstawowe $\boldsymbol{R}$ nie jest algebraicznie domknięte. Gdybyśmy rozważali przypadek $\boldsymbol{CP^2}$, *można byłoby zastosować oba sposoby*, i to jeszcze dla *każdego punktu* $r \in \boldsymbol{CP^2}$. Prosta rzeczywista l położona zewnątrz rzeczywistej elipsy Q, w istocie także *przecina się z nią, jednakże w dwóch punktach zespolonych wzajemnie sprzężonych*, i dwie sprzężone w sensie zespolonym styczne w tych punktach do Q przecinają się już w *rzeczywistym punkcie* r, leżącym wewnątrz Q. Z rzeczywistego punktu r leżącego wewnątrz Q, można również poprowadzić *dwie zespolone wzajemnie sprzężone styczne do* Q, których punkty styczności leżą na prostej rzeczywistej — szukanej prostej l.

W tym sensie rzeczywista geometria rzutowa $\boldsymbol{RP^2}$ jest jedynie cząstką geometrii $\boldsymbol{CP^2}$. Prawdziwie prosta i symetryczna teoria dwoistości ma miejsce dopiero w $\boldsymbol{CP^2}$, a $\boldsymbol{RP^2}$ odbija tylko jej część rzeczywistą.

Klasyczna geometria rzutowa była z znacznej mierze poświęcona objaśnieniu szczegółów tego pociągającego świata konfiguracji składających się z kwadryk, cięciw i stycznych oraz „niedostrzegalnych" zespolonych punktów styczności i przecięcia. W istocie cała kwadryka może nie mieć punktów rzeczywistych, na przykład $x_0^2 + x_1^2 + x_2^2 = 0$. Niemniej jednak widoczna część dwoistości rozgrywa się w $\boldsymbol{RP^2}$.

b) Podamy jeszcze jedną ilustrację w przypadku trójwymiarowym. Rozpatrzymy niezdegenerowaną kwadrykę rzutową Q w trójwymiarowej przestrzeni rzutowej i przeprowadzimy z punktu r położonego zewnątrz Q płaszczyzny styczne do Q. Wówczas wszystkie punkty styczności leżą w jednej płaszczyźnie, a mianowicie w płaszczyźnie dwoistej z r. Przyczyna jest znów ta sama: dwoistą do przecięcia płaszczyzn styczności jest powłoka rzutowa punktów styczności, i jeśli wszystkie płaszczyzny styczne przecinają się dokładnie w punkcie r (to należy i można dowieść w przypadku, gdy Q ma dostatecznie wiele punktów), to ta powłoka rzutowa musi być dwuwymiarowa.

Komentarz dotyczący zespolonych punktów styczności i przecięcia jest taki sam, jak w przypadku dwuwymiarowym.

Precyzyjne określenie dołączenia brakujących punktów przestrzeni rzutowej i kwadryki w przypadku $\boldsymbol{K} = R$ wykorzystuje pojęcie kompleksyfikacji (por. § 12, część 1).

5. a) *Kompleksyfikacją przestrzeni rzutowej* $P(L)$ nad $\boldsymbol{R}$ nazywamy przestrzeń rzutową $P(L^{\boldsymbol{C}})$ nad $\boldsymbol{C}$. Kanoniczne włożenie $L \subset L^{\boldsymbol{C}}$ pozwala przyporządkować każdej prostej nad $\boldsymbol{R}$ w L jej kompleksyfikację — prostą $L^{\boldsymbol{C}}$ nad $\boldsymbol{C}$, co zadaje włożenie $P(L) \subset P(L^{\boldsymbol{C}})$. Punkty $P(L^{\boldsymbol{C}})$ są „punktami zespolonymi" rzeczywistej przestrzeni rzutowej $P(L)$.

b) Izomorfizm $L \to L^*$ określony przez iloczyn skalarny g na L indukuje skompleksyfikowany izomorfizm $L^{\boldsymbol{C}} \to (L^{\boldsymbol{C}})^*$. Jest on określony przez symetryczny iloczyn skalarny $g^{\boldsymbol{C}}$ na $L^{\boldsymbol{C}}$, który wyznacza kwadrykę $Q^{\boldsymbol{C}}$ i dwoistość rzutową w $P(L^{\boldsymbol{C}})$. W $L^{\boldsymbol{C}}$ i $P(L^{\boldsymbol{C}})$ jest określona operacja sprzężenia zespolonego, indukowana przez antyliniowy izomorfizm $L^{\boldsymbol{C}} \to \overline{L}^{\boldsymbol{C}}$, będący tożsamością na $L \subset L^{\boldsymbol{C}}$. Punkty należące do $P(L)$ są punktami stałymi $P(L^{\boldsymbol{C}})$ względem sprzężenia zespolonego — nazywamy je punktami rzeczywistymi. Ogólniej, podprzestrzenie rzutowe w $P(L^{\boldsymbol{C}})$ przeprowadzane w siebie przez sprzężenie zespolone są w bijektywnej odpowiedniości z podprzestrzeniami rzutowymi w $P(L)$. Podprzestrzenie o tej własności nazywamy rzeczywistymi. Możemy zatem te dwa odwzorowania, będące wzajemnie odwrotnymi bijekcjami, opisać następująco:

(rzeczywista podprzestrzeń rzutowa w $P(L^{\boldsymbol{C}})$*)* $\longrightarrow$ *(zbiór jej punktów rzeczywistych w* $P(L)$*)*;

(podprzestrzeń rzutowa w $P(L)$*)* $\longrightarrow$ *(jej kompleksyfikacja w* $P(L^{\boldsymbol{C}})$*)*.

c) Relacja dwoistości w $P(L)$ określona przez g jest otrzymana z relacji dwoistości w $P(L^{\boldsymbol{C}})$, określonej przez $g^{\boldsymbol{C}}$, za pośrednictwem obcięcia tej ostatniej do układu podprzestrzeni rzeczywistych w $P(L^{\boldsymbol{C}})$, utożsamionego z układem podprzestrzeni w $P(L)$, jak w punkcie b) powyżej. Sprawdzenie wszystkich tych stwierdzeń, po uwzględnieniu rezultatów § 12 części 1 nie przedstawia trudności, a w rzeczywistym układzie współrzędnych $L^{\boldsymbol{C}}$, pochodzącym z L jest całkowicie tautologiczne. Jedyna znacząca strona tej sytuacji, zilustrowana poprzedzającymi przykładami,

polega na możliwości wystąpienia punktów rzeczywistych na niewidocznych konfiguracjach zespolonych podobnych do leżącego wewnątrz elipsy punktu przecięcia nierzeczywistych stycznych do dwóch sprzężonych w zespolonym sensie punktów tej elipsy.

W przypadku gdy ciało podstawowe $\boldsymbol{K}$ jest różne od $\boldsymbol{R}$, należy zamiast komplekszyfikacji posłużyć się ogólnym funktorem rozszerzenia ciała podstawowego (np. do jego domknięcia algebraicznego). Jednakże sytuacja komplikuje się z tego powodu, że zamiast jednego odwzorowania sprzężenia zespolonego należy używać całej grupy Galois dla wyróżnienia obiektów określonych nad wyjściowym ciałem (rzeczywistych w przypadku $\boldsymbol{K} = \boldsymbol{R}$).

§ 8. Grupy rzutowe i rzutowania

1. Niech L, M będą dwiema przestrzeniami liniowymi, $f: L \to M$ — odwzorowaniem liniowym. Jeśli $\operatorname{Ker} f = \{0\}$, to f przeprowadza dowolną prostą w L na jednoznacznie określoną prostą w M, indukując w ten sposób odwzorowanie $P(f)$: $P(L) \to P(M)$, nazywane *projektywizacją* f. W szczególności, jeśli f jest izomorfizmem, to $P(f)$ nazywa się *izomorfizmem rzutowym*. W przypadku $\operatorname{Ker} f \neq \{0\}$ sytuacja się komplikuje, bo proste zawarte w $\operatorname{Ker} f$, tj. stanowiące podprzestrzeń rzutową $P(\operatorname{Ker} f) \subset P(L)$, przeprowadzane są na zero, które nie określa żadnego punktu $P(M)$. Z tego powodu projektywizacja $P(f)$ jest określona jedynie na dopełnieniu $U_f = P(L) \setminus P(\operatorname{Ker} f)$. Obydwa przypadki są ważne, ale prowadzą w różnych kierunkach i z tego powodu rozważymy je osobno. Najbardziej istotne geometryczne cechy tej sytuacji pojawiają się już dla $L = M$.

2. Grupa rzutowa. Niech $M = L$, a f, g niech oznaczają dowolne elementy grupy automorfizmów liniowych przestrzeni L. Następujące fakty są oczywiste:

a) $P(\mathrm{id}_L) = \mathrm{id}_{P(L)}$;

b) $P(fg) = P(f)P(g)$.

W szczególności, każde odwzorowanie $P(f)$ jest bijektywne i $P(f^{-1}) = P(f)^{-1}$. Wobec tego $P(f)$ tworzą grupę przekształceń przestrzeni $P(L)$, zwaną *grupą rzutową przestrzeni* $P(L)$ i oznaczaną symbolem $\boldsymbol{PGL}(L)$, a odwzorowanie $P: \boldsymbol{GL}(L) \to \boldsymbol{PGL}(L)$, $f \mapsto P(f)$ jest surjektywnym homomorfizmem grup.

Każde odwzorowanie $P(f)$ odwzorowuje podprzestrzenie rzutowe $P(L)$ w podprzestrzenie rzutowe, zachowując wymiary i wszystkie relacje incydencji.

Zamiast $\boldsymbol{PGL}(\boldsymbol{K})^{n+1}$ będziemy pisać $\boldsymbol{PGL}(n)$.

3. Stwierdzenie. *Zbiór wszystkich homotetii przestrzeni L jest jądrem odwzorowania kanonicznego $P: \boldsymbol{GL}(L) \to \boldsymbol{PGL}(L)$. Zatem $\boldsymbol{PGL}(L)$ jest izomorficzna z grupą ilorazową $\boldsymbol{GL}(L)/\boldsymbol{K}^*$, gdzie $\boldsymbol{K}^* = \{a \cdot \mathrm{id}_L \mid a \in \boldsymbol{K} \setminus \{0\}\}$.*

Dowód. Na mocy definicji $\operatorname{Ker} P = \{f \in \boldsymbol{GL}(L) \mid P(f) = \mathrm{id}_{P(L)}\}$. Każda homotetia przeprowadza każdą prostą na siebie, a zatem $\boldsymbol{K}^* \subset \operatorname{Ker} P$. Odwrotnie,

każdy element $\operatorname{Ker} P$ przeprowadza dowolną prostą w siebie i dlatego jest diagonalny w dowolnej bazie L. Wtedy wszystkie jego wartości własne muszą być równe. Istotnie, niech $f(e_1) = \lambda_1 e_1$, $f(e_2) = \lambda_2 e_2$, gdzie e_1, e_2 są liniowo niezależne. Zatem z warunku $f(e_1 + e_2) = \mu(e_1 + e_2) = \lambda_1 e_1 + \lambda_2 e_2$ otrzymujemy, że $\lambda_1 = \mu = \lambda_2$. To oznacza, że f jest homotetią, czego należało dowieść.

4. Odwzorowania $P(f)$ we współrzędnych. Jeśli odwzorowanie liniowe wyraża się we współrzędnych przez macierz A:

$$f(x_0, \ldots, x_n) = A \cdot [x_0, \ldots, x_n]$$

(iloczyn macierzy A przez wektor kolumnowy $[x_0, \ldots, x_n]$), to $P(f)$ w odpowiadających współrzędnych jednorodnych wyraża się tą samą macierzą albo proporcjonalną do niej:

$$P(f)(x_0 : \ldots : x_n) \subset (\lambda A) \cdot [x_0, \ldots, x_n], \quad \lambda \in \boldsymbol{K}^*.$$

Jeśli ograniczyć się tylko do rozważania punktów z $x_0 \neq 0$, dla których współrzędne jednorodne można wybrać w postaci $(1 : y_1 : \ldots : y_n)$ i tak samo zapisywać współrzędne obrazu, to wynik można wyrazić za pomocą wzorów

$$\begin{aligned} P(f)(1 : y_1 : \ldots : y_n) &= (1 : y_1' : \ldots : y_n') = \\ &= \left(a_{00} + \sum_{i=1}^{n} a_{i0} y_i : a_{01} + \sum_{i=1}^{n} a_{i1} y_i : \ldots : a_{0n} + \sum_{i=1}^{n} a_{in} y_i \right) = \\ &= \left(1 : \frac{a_{01} + \sum_{i=1}^{n} a_{i1} y_i}{a_{00} + \sum_{i=1}^{n} a_{i0} y_i} : \ldots : \frac{a_{0n} + \sum_{i=1}^{n} a_{in} y_i}{a_{00} + \sum_{i=1}^{n} a_{i0} y_i} \right). \end{aligned}$$

(Naturalnie, $P(f)$ wyraża się analogicznym wzorem na zbiorze tych punktów, gdzie $x_i \neq 0$ dla dowolnego i.) Te wyrażenia nie mają sensu tam, gdzie mianownik zeruje się, tj. w tych punktach dopełnienia do hiperpłaszczyzny $x_0 = 0$, które przeprowadzane są w tę hiperpłaszczyznę przez odwzorowanie $P(f)$. Jeśli takich punktów nie ma, to, wyrażając $P(f)$ poprzez współrzędne afiniczne $(y_1, \ldots, y_n)$ na $P(L) \setminus \{x_0 = 0\}$, otrzymujemy odwzorowanie afiniczne. Niezmiennicze objaśnienie tego faktu przynosi następujący rezultat.

5. Stwierdzenie. *Niech $M \subset L$ będzie podprzestrzenią o kowymiarze 1, $P(M) \subset P(L)$ — odpowiadającą jej hiperpłaszczyzną, A_M — jej uzupełnieniem wyposażonym w strukturę afiniczną opisaną w § 6. Przyporządkowując dowolnemu automorfizmowi rzutowemu $P(f)\colon P(L) \to P(L)$ spełniającemu warunek $f(M) \subset M$ jego obcięcie do A_M, otrzymamy izomorfizm podgrupy $\boldsymbol{PGL}(L)$, przeprowadzającej $P(M)$ w siebie,*

$z\ \boldsymbol{Af}(A_M)$. *Część liniowa obcięcia* $P(f)$ *do* A_M *jest proporcjonalna do obcięcia* f *do* M.

Dowód. Wprowadzimy strukturę afiniczną na A_M utożsamiając A_M z podrozmaitością liniową $m' + M \subset L$, tj. każdemu punktowi A_M przyporządkowujemy przecięcie odpowiadającej mu prostej w L z $m' + M$. Jeśli $f(M) \subset M$, to w klasie f odpowiadającej jednemu $P(f)$ można dobrać jedyne odwzorowanie f_0 spełniające $f_0(m' + M) = m' + M$. Obcięcia wszystkich takich odwzorowań f_0 stanowią grupę odwzorowań afinicznych $m' + M$, ponieważ $m' + M$ jest podprzestrzenią afiniczną w L względem jej struktury afinicznej, a $f_0 : L \to L$ jest liniowe, a więc i afiniczne. Część liniowa takiego odwzorowania f_0 oczywiście jest równa obcięciu f_0 do M. Dla każdej części liniowej można znaleźć odpowiadające f_0, a przy ustalonej części liniowej można dobrać f_0 przeprowadzające dowolny punkt $m' + M$ w dowolny inny. Żeby się o tym przekonać, wystarczy wybrać bazę w L składającą się z bazy M i wektora m', a następnie skorzystać ze wzorów p. 3. W końcu, jeśli f działa tożsamościowo na $m' + M$ i M, to $P(f) = \mathrm{id}_{P(L)}$, bo f przeprowadza każdą prostą w L w nią samą. To kończy dowód.

6. Działanie grupy rzutowej na konfiguracje rzutowe. *Konfiguracją rzutową* nazwiemy skończony uporządkowany układ podprzestrzeni rzutowych w $P(L)$. Powiemy też, że dwie konfiguracje są rzutowo równoważne wtedy i tylko wtedy, gdy można przeprowadzić jedną z nich na drugą za pomocą rzutowego izmomorfizmu $P(L)$. Oczywiście, w tym celu jest konieczne i wystarczające, aby odpowiadające konfiguracje podprzestrzeni liniowych w L były jednakowo położone w znaczeniu § 5 części 1. Możemy więc od razu przełożyć wykazane tam rezultaty na język rzutowy, a otrzymamy następujące fakty.

a) Grupa $\boldsymbol{PGL}(L)$ działa tranzytywnie na zbiorze podprzestrzeni rzutowych ustalonego wymiaru w $P(L)$, tj. wszystkie takie podprzestrzenie są równoważne (por. § 5, p. 5, część 1).

b) Grupa $\boldsymbol{PGL}(L)$ działa tranzytywnie na zbiorze uporządkowanych par podprzestrzeni rzutowych w $P(L)$ o ustalonych wymiarach składników pary i ich przecięcia, tj. wszystkie takie pary są równoważne (por. § 5, p. 5, część 1)

c) Grupa $\boldsymbol{PGL}(L)$ działa tranzytywnie na zbiorze uporządkowanych n-tek podprzestrzeni rzutowych $(P_1, \ldots, P_n)$ w $P(L)$ o ustalonych wymiarach $\dim P_i$, które spełniają następujący warunek: dla każdego i podprzestrzeń P_i nie przecina powłoki rzutowej $(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$, tj. najmniejszej podprzestrzeni rzutowej zawierającej ten układ.

Rzeczywiście, niech $P_i = P(L_i)$, $L_i \subset L$. Nietrudno spostrzec, że powłoka rzutowa $(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$ jest równa $P(L_1 + \cdots + L_{i-1} + L_{i+1} + \cdots + L_n)$, a warunek rozłączności z $P(L_i)$ oznacza, że $L_i \cap \sum_{j \neq i}^{n} L_j = \{0\}$. Na mocy warunku a) twierdzenia § 5, p. 8, część 1, suma $L_1 \oplus \cdots \oplus L_n$ jest sumą prostą , a $\boldsymbol{GL}(L)$ działa tranzytywnie na takich n-elementowych układach podprzestrzeni (należy wybrać bazę w L uzupełniający sumę baz wszystkich L_i i wykorzystać tranzytywność

$\boldsymbol{GL}(L)$ na zbiorze baz L). Jako szczególny przypadek ($\dim P_i = 0$ dla wszystkich i) otrzymujemy następujący wynik: wszystkie takie układy n punktów w $P(L)$, że żaden z tych punktów nie należy do powłoki rzutowej pozostałych, są rzutowo równoważne.

d) Grupa $\boldsymbol{PGL}(L)$ działa tranzytywnie na zbiorze flag rzutowych $P_1 \subset P_2 \subset \cdots \subset P_n$ w $P(L)$ o ustalonej długości n oraz ustalonych wymiarach P_i.

Istotnie, każda taka flaga jest obrazem flagi $L_1 \subset L_2 \subset \cdots \subset L_n$ w L. Wybierzmy taką bazę w L, że pierwsze $\dim P_i + 1$ jej elementów generuje podprzestrzeń L_i dla każdego i, a następnie raz jeszcze odwołajmy się do tranzytywności działania $\boldsymbol{GL}(L)$ na zbiorze baz.

Poza tymi rezultatami, będącymi prostymi wnioskami z odpowiadających im twierdzeń dla przestrzeni liniowych, rozważymy interesujący nowy przypadek, w którym po raz pierwszy pojawia się nietrywialny niezmiennik przystawania rzutowego — klasyczny „dwustosunek czwórki punktów na prostej rzutowej". Większą część rozważań można prowadzić w przypadku dowolnego wymiaru, zaczniemy więc od ogólnej definicji.

7. Definicja. Układ punktów $p_1, \ldots, p_N$ należących do n-wymiarowej przestrzeni rzutowej P *znajduje się w położeniu ogólnym*, jeśli dla każdego $m \leq \min\{N, n+1\}$ i każdego m-elementowego podzbioru $S \subset \{1, \ldots, N\}$ wymiar powłoki rzutowej punktów $\{p_i \mid i \in S\}$ wynosi $m-1$. Przypadki $N = n+1$, $n+2$, $n+3$ będą dla nas szczególnie interesujące.

a) $n+1$ *punktów w położeniu ogólnym.* Ponieważ żaden z tych punktów nie należy do powłoki rzutowej pozostałych (w przeciwnym przypadku wymiar powłoki rzutowej całego układu byłby równy $n-1$, a nie n), są to konfiguracje rozpatrywane w p. 6, c), gdzie w szczególności pokazaliśmy, że grupa rzutowa działa na zbiorze tych konfiguracji tranzytywnie. Teraz zwrócimy uwagę na ten fakt, że odwzorowanie rzutowe, przeprowadzające jeden układ $n+1$ punktów w położeniu ogólnym na drugi, nie jest wyznaczone jednoznacznie.

Rzeczywiście, jeśli przez $e_1, \ldots, e_{n+1}$ oznaczyć niezerowe wektory należące do $p_1, \ldots, p_{n+1}$ odpowiednio, to $\{e_1, \ldots, e_{n+1}\}$ jest bazą L (tutaj $P = P(L)$), a grupa przekształceń rzutowych, zachowujących na miejscu wszystkie punkty p_i zawiera dokładnie te i tylko te przekształcenia postaci $P(f)$, dla których f jest diagonalne w bazie $\{e_1, \ldots, e_{n+1}\}$. Ten dodatkowy stopień swobody pozwoli nam wykazać tranzytywność działania $\boldsymbol{PGL}(L)$ na układach $n+2$ punktów w położeniu ogólnym.

b) $n+2$ *punkty w położeniu ogólnym.* Jeśli punkty $\{p_1, \ldots, p_{n+2}\}$ znajdują się w położeniu ogólnym, to punkty $\{p_1, \ldots, p_{n+1}\}$ też znajdują się w położeniu ogólnym. Jak w poprzedzającym fragmencie wybierzmy bazę $\{e_1, \ldots, e_{n+1}\}$, $e_i \in p_i$. Zadaje ona układ współrzędnych jednorodnych w P. Niech $(x_1 : \ldots : x_{n+1})$ będą współrzędnymi punktu p_{n+2} względem tej bazy. Żadna z tych współrzędnych nie jest zerem, bo wtedy wektor $(x_1, \ldots, x_{n+1})$ należący do prostej p_{n+2} wyrażałby się liniowo przez wektory e_j, $0 \leq j \leq n+1$, $j \neq i$, skąd otrzymalibyśmy, że wymiar powłoki rzutowej $n+1$ punktów $\{p_i \mid i \neq j\}$ wynosiłby $n-1$, a nie n. Przekształcenie $P(f)$ dla

$f = \operatorname{diag}(\lambda_1, \ldots, \lambda_{n+1})$ (w bazie $\{e_1, \ldots, e_{n+1}\}$) przeprowadza punkt $(x_1 : \ldots : x_{n+1})$ na punkt $(\lambda_1 x_1 : \ldots : \lambda_{n+1} x_{n+1})$, nie zmieniając $p_1, \ldots, p_{n+1}$. Stąd wynika, że dowolny punkt $(x_1 : \ldots : x_{n+1})$, gdzie wszystkie $x_i \neq 0$, można przeprowadzić na dowolny inny punkt $(y_1 : \ldots : y_{n+1})$ (wszystkie $y_i \neq 0$) jednoznacznie wyznaczonym przekształceniem rzutowym, nie zmieniającym punktów $p_1, \ldots, p_{n+1}$.

W ten sposób wykazaliśmy, że wszystkie uporządkowane układy $n+2$ punktów w położeniu ogólnym w P, gdzie $\dim P = n$, są równoważne, a co więcej, stanowią główną przestrzeń jednorodną dla grupy $\boldsymbol{PGL}(L)$.

Używając „biernego" zamiast „aktywnego" punktu widzenia, możemy powiedzieć, że dla dowolnego uporządkowanego układu punktów $\{p_1, \ldots, p_{n+2}\}$ istnieje jedyny układ współrzędnych jednorodnych w P, w którym współrzędne punktów $p_1, \ldots, p_{n+2}$ mają następującą postać:

$$p_1 = (1 : 0 : \ldots : 0),\ p_2 = (0 : 1 : 0 : \ldots : 0),\ \ldots,\ p_{n+1} = (0 : \ldots : 0 : 1),$$
$$p_{n+2} = (1 : \ldots : 1).$$

Można by nazwać ten układ dostosowanym do $\{p_1, \ldots, p_{n+2}\}$.

c) $n+3$ *punkty w położeniu ogólnym.* Dwie takie konfiguracje na ogół nie są równoważne. Dla danych $\{p_1, \ldots, p_{n+3}\}$ i $\{p'_1, \ldots, p'_{n+3}\}$ można jednoznacznie wyznaczyć przekształcenie rzutowe, przeprowadzające p_i na p'_i dla każdego $1 \leq i \leq n+2$, ale przy tym p_{n+3}, w zależności od sytuacji, może być odwzorowany na p'_{n+3} lub nie.

Nietrudno opisać niezmienniki rzutowe układu $n+3$ punktów. Wybierzmy układ współrzędnych jednorodnych w P, w którym pierwsze $n+2$ punkty zadane są współrzędnymi podanymi w punkcie b). Punkt p_{n+3} opisany jest w tym układzie przez współrzędne $(x_1 : \ldots : x_{n+1})$, wyznaczone z dokładnością do proporcjonalności. Dowolny automorfizm rzutowy, zastosowany jednocześnie do konfiguracji $\{p_1, \ldots, p_{n+2}\}$ i do dostosowanego do niej układu współrzędnych, przeprowadzi tę konfigurację na inną, a dostosowany do niej układ współrzędnych — na układ dostosowany do nowej konfiguracji. W ten sposób współrzędne $(x_1 : \ldots : x_{n+1})$ ostatniego punktu pozostaną niezmienione.

Wszystkie poprzedzające rozważania z oczywistymi zmianami przenoszą się na przypadek dwóch przestrzeni rzutowych P i P', konfiguracji $\{p_1, \ldots, p_N\} \subset P$ i $\{p'_1, \ldots, p'_N\} \subset P'$ oraz izomorfizmów rzutowych $P \to P'$ przeprowadzających jedną konfigurację na drugą.

Zbierzemy rezultaty powyższych rozważań w następującym twierdzeniu.

8. Twierdzenie. a) *Niech P, P' będą n-wymiarowymi przestrzeniami rzutowymi, $\{p_1, \ldots, p_{n+2}\} \subset P$ i $\{p'_1, \ldots, p'_{n+2}\} \subset P'$ — dwoma układami punktów w położeniu ogólnym. Wówczas istnieje jedyny izomorfizm rzutowy $P \to P'$, przeprowadzający jedną konfigurację na drugą.*

b) *Analogiczny rezultat jest wtedy i tylko wtedy prawdziwy dla układów $n+3$ punktów w położeniu ogólnym, gdy współrzędne $(n+3)$-go punktu w układzie dostosowanym do pierwszych $n+2$ punktów w obydwóch konfiguracjach są jednakowe (oczywiście, z dokładnością do czynnika skalarnego).*

9. Dwustosunek. Zastosujmy twierdzenie p. 8 w przypadku $n = 1$. Jest jasne, że jeśli na dwóch prostych rzutowych zadane są uporządkowane trójki parami różnych punktów (do tego sprowadza się tutaj warunek położenia ogólnego), to istnieje jedyny izomorfizm rzutowy prostych, przeprowadzający jedną z tych trójek na drugą.

Rozważmy uporządkowaną czwórkę parami różnych punktów $\{p_1, p_2, p_3, p_4\} \subset P^1$ zadaną w dostosowanym do niej układzie współrzędnymi $(1:0)$, $(0:1)$, $(1:1)$ oraz $(x_1 : x_2)$. Wtedy $x_2 \neq 0$. Zdefiniujemy liczbę nazywaną *dwustosunkiem* lub *stosunkiem anharmonicznym* czwórki punktów $\{p_i\}$ za pomocą wzoru

$$[p_2, p_3, p_1, p_4] = x_1 x_2^{-1}.$$

Niezwykła kolejność punktów jest uwarunkowana pragnieniem zachowania zgodności z klasyczną definicją: w mapie afinicznej, gdzie $p_2 = \infty$, $p_3 = 0$, $p_1 = 1$, współrzędne punktów występujących w nawiasie kwadratowym są rozłożone następująco: $[0, 1, \infty, x]$, gdzie x jest właśnie dwustosunkiem tej czwórki.

Sama nazwa „dwustosunek" pochodzi od następującego jawnego wzoru dla obliczenia niezmiennika $[x_1, x_2, x_3, x_4]$, gdzie $x_i \in \boldsymbol{K}$ tym razem mają sens współrzędnych punktów p_i w dowolnej mapie afinicznej P^1. Zgodnie z wynikami p. 4 grupa $\boldsymbol{PGL}(L)$ w tej mapie jest wyrażona wzorami opisującymi odwzorowania homograficzne postaci $x \mapsto \dfrac{ax+b}{cx+d}$, $ad - bc \neq 0$. Takie odwzorowanie przeprowadzające (x_1, x_2, x_3) na $(0, 1, \infty)$ ma postać

$$x \mapsto \frac{x_1 - x}{x_3 - x} : \frac{x_1 - x_2}{x_3 - x_2}.$$

Podstawiając tu $x = x_4$ otrzymujemy

$$[x_1, x_2, x_3, x_4] = \frac{x_1 - x_4}{x_3 - x_4} : \frac{x_1 - x_2}{x_3 - x_2}.$$

Jeszcze jedna klasyczna konstrukcja związana z częścią a) tezy twierdzenia p. 8 dla $n = 1$, opisuje realizację grupy symetrycznej S_3 za pomocą odwzorowań homograficznych. Zgodnie z tym twierdzeniem dowolna permutacja $\{p_1, p_2, p_3\} \mapsto \{p_{\sigma(1)}, p_{\sigma(2)}, p_{\sigma(3)}\}$ trzech punktów na prostej rzutowej jest indukowana jedynym przekształceniem rzutowym tej prostej.

W mapie afinicznej, gdzie $\{p_1, p_2, p_3\} = \{0, 1, \infty\}$, odpowiednie odwzorowania rzutowe dane są przez homografie

$$x \mapsto \left\{ x,\ \frac{1}{x},\ 1 - x,\ \frac{1}{1-x},\ \frac{x-1}{x},\ \frac{x}{x-1} \right\}.$$

Przejdziemy teraz do zbadania rzutowań.

10. Załóżmy, że przestrzeń liniowa L jest przedstawiona w postaci sumy prostej dwóch podprzestrzeni o wymiarach ≥ 1: $L = L_1 \oplus L_2$. Oznaczmy $P = P(L)$, $P_i = P(L_i)$, $i = 1, 2$.

Jak to zostało pokazane w p. 1 rzut $f: L \to L_2$, $f(l_1 + l_2) = l_2$, $l_i \in L_i$, indukuje odwzorowanie

$$P(f): P \setminus P_1 \longrightarrow P_2,$$

które będziemy nazywać *rzutowaniem ze środka P_1 na P_2*. Dla opisania tej sytuacji w czysto rzutowych terminach odnotujmy następujące fakty.

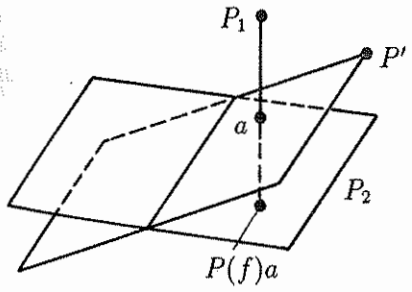
a) $\dim P_1 + \dim P_2 = \dim P - 1$ i $P_1 \cap P_2 = \emptyset$. Odwrotnie, każda konfiguracja z takimi własnościami pochodzi od jednoznacznie wyznaczonego rozkładu na sumę prostą $L = L_1 \oplus L_2$.

b) jeśli $a \in P_2$, to $P(f)a = a$: jeśli $a \in P \setminus (P_1 \cup P_2)$, to $P(f)a$ jest zdefiniowane jako punkt przecięcia z P_2 jedynej prostej rzutowej w P przechodzącej przez a i przecinającej P_1 i P_2.

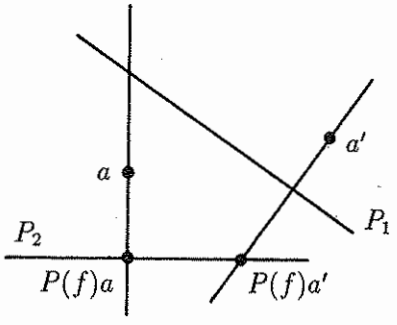
Istotnie, przypadek $a \in P_2$ jest oczywisty. Jeśli $a \notin P_1 \cup P_2$, to w terminach przestrzeni L oczekiwany przez nas rezultat może być sformułowany następująco: przez dowolną prostą $L_0 \subset L$, nie należącą do L_1 i L_2, przechodzi jedyna płaszczyzna przecinająca L_1 i L_2 wzdłuż prostych, a jej przecięcie z L_2 jest równe jej rzutowi na L_2. W istocie, jedna płaszczyzna o takich własnościach istnieje — jest to płaszczyzna rozpięta przez rzuty L_0 na L_1 i L_2 odpowiednio. Istnienie dwóch różnych płaszczyzn o tych własnościach pociągałoby możliwość przedstawienia wektora $l_0 \in L_0$ jako sumy wektorów z L_1B i L_2 na dwa różne sposoby, co jest niemożliwe, bo $L = L_1 \oplus L_2$.

Ponieważ opisana rzutowa konstrukcja odwzorowania $P \to P \setminus P_1$ podnosi się do liniowej, więc natychmiast widać, że dla dowolnej podprzestrzeni $P' \subset P \setminus P_1$ obcięcie rzutowania $P' \to P_1$ jest odwzorowaniem rzutowym, tj. ma postać $P(g)$ dla pewnego odwzorowania liniowego g odpowiadających porzestrzeni liniowych.

W ważnym szczególnym przypadku, gdy jako P_1 obieramy punkt, a jako P_2 — hiperpłaszczyznę, odwzorowanie rzutowania ze środka P_1 na P_2 przeprowadza punkt a na punkt z P_2, będący jego obrazem dla obserwatora patrzącego z punktu P_1. Z tego względu relacja pomiędzy figurą a jej rzutem w takim przypadku nazywana jest nieraz stosunkiem perspektywicznym. Intuicyjnie mniej oczywiste jest, na przykład, rzutowanie z prostej na prostą w P^3 (rys. 5, 6).



Rys. 5



Rys. 6

Ważną własnością rzutowania, o której nie należy zapominać, jest następujący fakt: *jeśli* $P' \subset P \setminus P_1$, *to rzutowanie ze środka* P_1 *zadaje izomorfizm rzutowy* P' *z jego obrazem w* P_2. Istotnie, w terminach przestrzeni liniowych oznacza to, że rzut $f: L_1 \oplus L_2 \to L_2$ indukuje izomorfizm M z $f(M)$ dla dowolnej podprzestrzeni $M \subset L$, takiej że $L_1 \cap M = \{0\}$. A ten fakt wynika z równości $L_1 = \operatorname{Ker} f$.

11. Zachowanie rzutowania w pobliżu środka. W dalszym ciągu ograniczymy się do rozważania rzutowania z punktu $p_1 = a \in P$ i spróbujemy wyjaśnić, co dzieje się z punktami położonymi w pobliżu środka. W przypadku $\boldsymbol{K} = \boldsymbol{R}$ i $\boldsymbol{C}$, gdy rzeczywiście można mówić o bliskości punktów, sytuacja jest następująca: w punkcie a rzutowanie przestaje być ciągłe, bo punkty b, dowolnie bliskie a, lecz zbliżające się do a z „różnych stron", są rzutowane na punkty bardzo od siebie oddalone. Właśnie ta własność rzutowania jest wykorzystywana w jej zastosowaniach do rozmaitych zagadnień o „rozwiązywaniu osobliwości". Jeśli w P położona jest jakaś „figura" (rozmaitość algebraiczna, pole wektorowe), o skomplikowanej wewnętrznej budowie w pobliżu punktu a, to przez rzutowanie jej z punktu a możemy rozciągnąć otoczenie tego punktu i zobaczyć w powiększonej skali, co się w nim dzieje, przy czym współczynnik powiększenia nieograniczenie rośnie w miarę zbliżania się do a.

Chociaż takie rozważania znajdują zastosowania w istotnie nieliniowych sytuacjach (trywializujących się w liniowych modelach), to warto rozpatrzyć strukturę rzutowania w pobliżu jego środka nieco dokładniej, na ile to jest możliwe przy pozostawaniu w ramach geometrii liniowej.

12. Wprowadźmy w $P(L)$ układ współrzędnych jednorodnych, w którym środkiem rzutowania jest punkt $(0 : \ldots, : 0 : 1)$, a $P_2 = P^{n-1}$ składa się z punktów o współrzędnych $(x_0 : x_1 : \ldots : x_{n-1} : 0)$; żeby to osiągnąć, należy wybrać w L bazę będącą sumą (teoriomnogościową) baz L_1 i L_2 (ponieważ środek jest punktem, więc $\dim L_1 = 1$).

Nietrudno spostrzec, że wówczas punkt $(x_0 : \ldots : x_n)$ jest rzutowany na punkt $(x_0 : \ldots : x_{n-1} : 0)$. Dopełnienie A do P_2 jest wyposażone w układ współrzędnych afinicznych

$$(y_0, \ldots, y_{n-1}) = (x_0/x_n, \ldots, x_{n-1}/x_n)$$

z początkiem O w środku rzutowania. Rozważmy iloczyn kartezjański $A \times P_2 = A \times P^{n-1}$ i zawarty w nim wykres Γ_0 odwzorowania rzutowania, które, przypomnijmy to, jest określone tylko na $A \setminus \{0\}$. Ten wykres składa się z par punktów o współrzędnych

$$\big((x_0/x_n, \ldots, x_{n-1}/x_n), (x_0 : \ldots : x_{n-1})\big),$$

gdzie nie wszystkie $x_0, \ldots, x_{n-1}$ są jednocześnie równe zeru. Powiększmy wykres Γ_0, dołączając do niego nad punktem $O \in A$ zbiór $\{0\} \times P^{n-1} \subset A \times P^{n-1}$,

$$\Gamma = \Gamma_0 \cup \{0\} \times P^{n-1},$$

w zgodzie z geometryczną intuicją sugerującą, że przy rzutowaniu z 0 środek „przechodzi na całą przestrzeń P^{n-1}". Zbiór Γ ma cały szereg dobrych własności.

a) Γ zawiera dokładnie te pary punktów

$$\big((y_0,\dots,y_{n-1}),\,(x_0:\dots:x_{n-1})\big),$$

których współrzędne spełniają układ równań algebraicznych

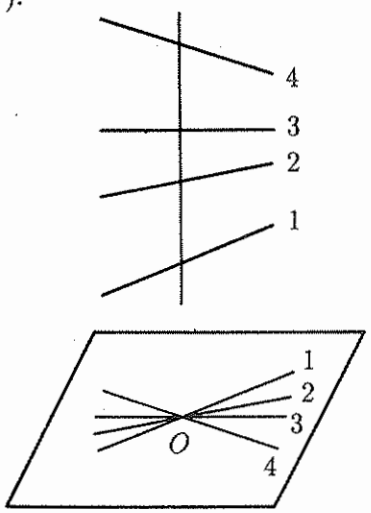
$$y_i x_j - y_j x_i = 0,\quad i,j = 0,\dots,n-1.$$

Istotnie, te równania oznaczają, że wszystkie minory macierzy $\begin{pmatrix} x_1 & \dots & x_{n-1} \\ y_1 & \dots & y_{n-1} \end{pmatrix}$ są równe zeru, tak że jej rząd jest równy jeden (bo pierwszy wiersz jest niezerowy), a więc drugi wiersz jest proporcjonalny do pierwszego. Jeśli współczynnik proporcjonalności nie jest zerem, to otrzymujemy punkt z Γ_0, a w przeciwnym przypadku — punkt $\{0\} \times P^{n-1}$.

b) Odwzorowanie $\Gamma \to A : \big((y_0,\dots,y_{n-1}),(x_0:\dots:x_{n-1})\big) \mapsto (y_0,\dots,y_{n-1})$ jest bijekcją wszędzie, z wyjątkiem włókna nad punktem $\{0\}$.

Innymi słowy, Γ jest otrzymane z A za pomocą „doklejania" całej przestrzeni rzutowej P^{n-1} na miejsce jednego punktu. Mówi się, że Γ jest otrzymane z A „rozdmuchaniem" punktu albo σ-procesem ze środkiem w punkcie O. Przeciwobraz w Γ każdej prostej z A, przechodzącej przez punkt O, przecina doklejoną przestrzeń rzutową P^{n-1} także w jednym punkcie, ale innym dla każdej prostej.

W przypadku $\boldsymbol{K} = \boldsymbol{R}$ i $n = 2$, można wyobrazić sobie mapę afiniczną doklejonej przestrzeni rzutowej P^1 jako oś korbki Γ, wykonującej półobroty na całej swej nieskończonej długości (rys. 7).



Rys. 7

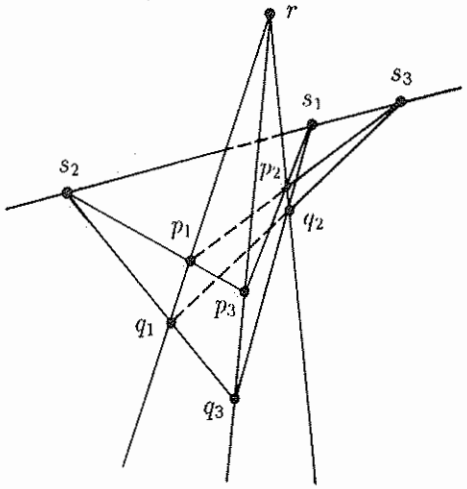
Rzeczywiste zastosowanie rzutowania do badania osobliwości w punkcie $O \in A$ jest związane z przeniesieniem interesującej nas figury do Γ i rozpatrzeniem geometrii jej przeciwobrazu w pobliżu wklejonej przestrzeni P^{n-1}. Wnosimy stąd na przykład, że przeciwobraz rozmaitości algebraicznej jest rozmaitością algebraiczną, a to dzięki temu, że Γ jest zadane równaniami algebraicznymi.

§ 9. Konfiguracje Desargues'a i Pappusa oraz klasyczna geometria rzutowa

1. Klasyczna syntetyczna geometria rzutowa poświęcona była w znacznej mierze badaniu rodziny podprzestrzeni przestrzeni rzutowej wraz z relacją incydencji — własności tej relacji można przyjąć jako podstawę aksjomatyki i dojść do współczesnego określenia przestrzeni $P(L)$ i ciała skalarów $\boldsymbol{K}$, gdzie L i $\boldsymbol{K}$ pojawią się jako struktury pochodne. W takim podejściu dużą rolę grają dwie konfiguracje — Desargues'a i Pappusa. Wprowadzimy je i zbadamy, wychodząc od naszych definicji, a następnie pokrótce omówimy ich rolę w geometrii syntetycznej.

2. Konfiguracja Desargues'a. Niech S oznacza układ punktów w przestrzeni rzutowej. Symbolem $\overline{S}$ będziemy oznaczać powłokę rzutową tego układu. Rozważmy w trójwymiarowej przestrzeni rzutowej uporządkowaną szóstkę punktów $(p_1, p_2, p_3; q_1, q_2, q_3)$. Zakładamy, że punkty są parami różne i że $\overline{p_1p_2p_3}$ i $\overline{q_1q_2q_3}$ są płaszczyznami. Dalej, niech proste $\overline{p_1q_1}$, $\overline{p_2q_2}$ i $\overline{p_3q_3}$ przecinają się w jednym punkcie r, różnym od p_i i q_j. Innymi słowy, „trójkąty" $p_1p_2p_3$ i $q_1q_2q_3$ są „perspektywiczne" i jeśli należą do różnych płaszczyzn, to każdy z nich jest rzutem drugiego ze środka r. Wówczas dla dowolnej pary różnych wskaźników $\{i,j\} \subset \{1,2,3\}$ proste $\overline{p_ip_j}$ i $\overline{q_iq_j}$ nie są identyczne, gdyż w przeciwnym przypadku mielibyśmy $p_i = q_i$, bowiem p_i i q_i są punktami przecięcia tych prostych z prostą $\overline{p_iq_i}$. Poza tym proste $\overline{p_ip_j}$ i $\overline{q_iq_j}$ leżą we wspólnej płaszczyźnie $\overline{rp_ip_j}$. Dlatego te proste przecinają się w jednym punkcie, który oznaczymy przez s_k, gdzie $\{i,j,k\} = \{1,2,3\}$; jest to punkt przecięcia przedłużeń pary odpowiednich boków trójkątów $p_1p_2p_3$ i $q_1q_2q_3$.

Twierdzenie Desargues'a, które udowodnimy w następnym punkcie, stwierdza, że trzy punkty s_1, s_2, s_3 leżą na jednej prostej. Konfiguracja złożona z dziesięciu punktów p_i, q_j, s_k, r i dziesięciu łączących je prostych, pokazanych na rys. 8, nazywa się *konfiguracją Desargues'a*.



Rys. 8

Każda z jej prostych zawiera dokładnie trzy punkty należące do niej i przez każdy z jej punktów przechodzą dokładnie trzy z jej prostych. Pozostawiamy czytelnikowi do samodzielnego sprawdzenia, że ma ona pewną własność symetrii, a mianowicie, że grupa permutacji jej punktów i prostych, zachowująca związki incydencji, jest tranzytywna zarówno na punktach, jak i na prostych.

3. Twierdzenie Desargues'a. *Przy uczynionych powyżej założeniach punkty* s_1, s_2, s_3 *należą do jednej prostej.*

Dowód. Rozważymy dwa przypadki, w zależności od tego, czy płaszczyzny $\overline{p_1p_2p_3}$ i $\overline{q_1q_2q_3}$ są identyczne czy też nie.

a) $\overline{p_1p_2p_3} \neq \overline{q_1q_2q_3}$ („przestrzenne twierdzenie Desargues'a"). W tym przypadku płaszczyzny $\overline{p_1p_2p_3}$ i $\overline{q_1q_2q_3}$ przecinają się wzdłuż prostej i nietrudno zobaczyć, że punkty s_1, s_2 i s_3 leżą na tej prostej. Rzeczywiście, np. punkt s_1 leży na prostych $\overline{p_2p_3}$ i $\overline{q_2q_3}$, które z kolei leżą w płaszczyznach $\overline{p_1p_2p_3}$ i $\overline{q_1q_2q_3}$, a więc należy on także do ich przecięcia.

b) $\overline{p_1p_2p_3} = \overline{q_1q_2q_3}$ („płaskie twierdzenie Desargues'a"). W tym przypadku wybierzemy w przestrzeni punkt r', nie należący do płaszczyzny $\overline{p_1p_2p_3}$, i połączymy go prostymi z punktami r, p_i, p_j. Prosta $\overline{p_iq_i}$ zawarta jest w płaszczyźnie $\overline{r'p_iq_i}$, więc i punkt r należy do tej płaszczyzny. Przeprowadzimy w niej prostą przez punkt r tak, aby nie przechodziła przez r' i p_i i oznaczymy jej przecięcie z prostymi $\overline{r'p_i}$ i $\overline{r'q_i}$ odpowiednio przez p' i q'. Trójki (p', p_2, p_3) i (q', q_2, q_3) leżą już w innych płaszczyznach — w przeciwnym przypadku ich wspólna płaszczyzna zawierałaby proste $\overline{p_2p_3}$ i $\overline{q_2q_3}$, a zatem musiała by być identyczna z płaszczyzną $\overline{p_1p_2p_3}$, a to jest niemożliwe, bo p', i q' do tej początkowej płaszczyzny nie należą. Poza tym proste $\overline{p'q'}$, $\overline{p_2q_2}$ i $\overline{p_3q_3}$ przechodzą przez punkt r. Na mocy przestrzennego twierdzenia Desargues'a punkty $\overline{p'p_2}\cap\overline{q'q_2}$, $\overline{p'p_3}\cap\overline{q'q_3}$ i $\overline{p_2p_3}\cap\overline{q_2q_3}$ należą do jednej prostej. W wyniku rzutowania z punktu r' na płaszczyznę $\overline{p_1p_2p_3}$ tych punktów otrzymamy odpowiednio s_3, s_2, s_1, ponieważ r' przeprowadza (p', p_2, p_3) na (p_1, p_2, p_3) i (q', q_2, q_3) na (q_1, q_2, q_3), a więc boki każdego z tych trójkątów na odpowiednie boki wyjściowych trójkątów.

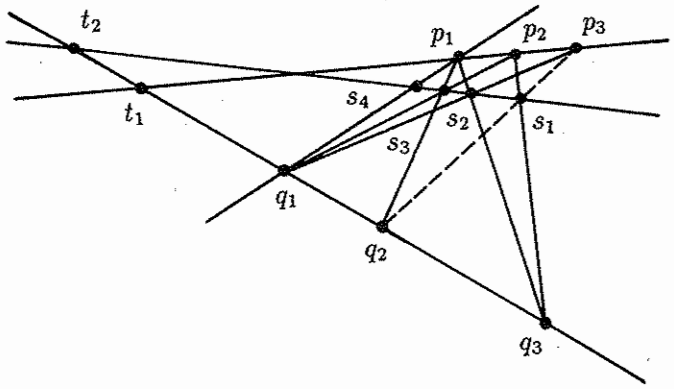
To kończy dowód.

4. Konfiguracja Pappusa. Rozważmy dwie różne proste położone w płaszczyźnie rzutowej i dwie trójki p_1, p_2, p_3 i q_1, q_2, q_3 parami różnych punktów należących do tych prostych. Dla każdej pary różnych wskaźników $\{i, j\} \subset \{1, 2, 3\}$ oznaczmy $s_k = \overline{p_iq_j}\cap\overline{q_ip_j}$, gdzie $\{i, j, k\} = \{1, 2, 3\}$.

5. Twierdzenie Pappusa. *Punkty* s_1, s_2, s_3 *leżą na wspólnej prostej.*

Dowód. Przeprowadźmy prostą przez punkty s_3, s_2 i oznaczmy przez s_4 jej przecięcie z prostą $\overline{p_1q_1}$. Naszym celem jest wykazanie, że punkt s_1 leży na tej prostej.

Zbudujemy dwa odwzorowania rzutowe f_1, f_2: $\overline{p_1p_2p_3} \to \overline{q_1q_2q_3}$. Jako pierwsze z nich, f_1, weźmiemy złożenie rzutowania $\overline{p_1p_2p_3}$ na $\overline{s_2s_3}$ z punktu q_1 z rzutowaniem $\overline{s_3s_2}$ na $\overline{q_1q_2q_3}$ z punktu p_1. Mamy oczywiście $f_1(p_i) = q_i$ dla każdego $i = 1, 2, 3$, a ponadto $f_1(l_1) = l_2$, gdzie $l_1 = \overline{p_1p_2p_3}\cap\overline{q_1q_2q_3}$, $l_2 = \overline{s_4s_3s_2}\cap\overline{q_1q_2q_3}$ (rys. 9).



Rys. 9

Odwzorowanie f_2 określimy jako złożenie rzutowania $\overline{p_1p_2p_3}$ na $\overline{s_3s_2}$ z punktu q_2 z rzutowaniem $\overline{s_3s_2}$ na $\overline{q_1q_2q_3}$ z punktu p_2. To złożenie przeprowadza p_1 w q_1, p_2 w q_2 i t_1 w t_2.

Odwzorowania f_1 i f_2 są sobie równe, gdyż jednakowo działają na trójkę punktów (t_1, p_1, p_2). W szczególności $f_1(p_3) = f_2(p_3)$, a wobec $f_1(p_3) = a_3$, mamy $f_2(p_3) = q_3$. Geometryczne znaczenie tej równości jest następujące: jeśli oznaczyć przez s' przecięcie $\overline{q_2p_3} \cap \overline{s_3s_2}$, to prosta $\overline{p_2q_3}$ przechodzi przez s'. Wówczas $s' = \overline{q_2p_3} \cap \overline{p_2q_3} = s_1$. To oznacza, że s_1 leży na prostej $\overline{s_2s_3}$, czego należało dowieść.

6. Klasyczna aksjomatyka trójwymiarowej przestrzeni rzutowej i płaszczyzny rzutowej. Klasyczną trójwymiarową przestrzeń rzutową definiuje się jako zbiór, którego elementy nazywane są punktami, z wyróżnionymi dwiema rodzinami podzbiorów, których elementy nazywane są odpowiednio prostymi i płaszczyznami. Przy tym zakłada się wypełnienie następujących aksjomatów.

T_1. Dwa różne punkty należą do jednej prostej.

T_2. Trzy różne punkty, nie należące do jednej prostej, należą do jednej płaszczyzny.

T_3. Prosta i płaszczyzna mają punkt wspólny.

T_4. Przecięcie prostej i płaszczyzny zawiera prostą.

T_5. Istnieją cztery punkty, nie leżące we wspólnej płaszczyźnie i takie, że każde trzy spośród nich nie leżą na wspólnej prostej.

T_6. Każda prosta zawiera nie mniej niż trzy punkty.

Klasyczna płaszczyzna rzutowa jest zdefiniowana jako zbiór, którego elementy nazywane są punktami, z wyróżnioną rodziną podzbiorów zwanych prostymi. Przy tym zakłada się spełnienie następujących aksjomatów.

Π_1. Dwa różne punkty należą do jednej prostej.

Π_2. Przecięcie dwóch prostych jest niepuste.

Π_3. Istnieją trzy punkty nie leżące na jednej prostej.

Π_4. Każda prosta zawiera nie mniej niż trzy punkty.

Zbiory $P(L)$, gdzie L oznacza przestrzeń liniową nad ciałem $\boldsymbol{K}$ o wymiarze 4 lub 3, wraz z rodzinami płaszczyzn i prostych rzutowych w nich zawartych, określonych tak jak wyżej, spełniają aksjomaty T_1–T_6 i Π_1–Π_4 odpowiednio, co wynika wprost ze standardowych własności przestrzeni liniowych wykazanych w części 1. Jednakże nie każda klasyczna przestrzeń lub płaszczyzna rzutowa jest izomorficzna (w oczywistym znaczeniu tego słowa) z jedną z naszych przestrzeni $P(L)$. Następująca fundamentalna konstrukcja dostarcza wielu przykładów.

7. Przestrzenie liniowe i rzutowe nad nieprzemiennymi ciałami. *Ciałem nieprzemiennym* (zwanym także *pierścieniem z dzieleniem*) nazywamy taki pierścień łączny K, w którym zbiór elementów niezerowych tworzy grupę względem mnożenia (niekoniecznie przemienną). Każde ciało jest ciałem nieprzemiennym, ale nie na odwrót. Na przykład, klasyczny pierścień kwaternionów jest ciałem nieprzemiennym, lecz nie jest ciałem.

Addytywna grupa L wraz z dwuargumentową operacją mnożenia $\boldsymbol{K} \times L \to L$: $(a, l) \mapsto al$ nazywa się (lewostronną) *przestrzenią liniową nad ciałem nieprzemiennym* $\boldsymbol{K}$, jeśli są spełnione warunki definicji § 1, p. 2 części 1. Znaczną część teorii przestrzeni liniowych daje się przenieść prawie bez zmian na przypadek przestrzeni liniowych nad ciałami nieprzemiennymi. Dotyczy to w szczególności teorii wymiaru i bazy oraz teorii podprzestrzeni, włącznie z twierdzeniem o wymiarze przecięcia podprzestrzeni. To pozwala zbudować dla każdego ciała nieprzemiennego $\boldsymbol{K}$ i przestrzeni liniowej nad tym ciałem przestrzeń rzutową $P(L)$, składającą się z prostych w L oraz rodzinę jego podprzestrzeni rzutowych $P(M)$, gdzie $M \subset L$ jest podprzestrzenią liniową L dowolnego wymiaru. Jeśli $\dim_{\boldsymbol{K}} L = 4$ albo 3, to te obiekty spełniają wszystkie aksjomaty T_1–T_6 i Π_1–Π_4 odpowiednio.

8. Znaczenie twierdzenia Desargues'a. Jak się jednak okazuje, istnieją klasyczne płaszczyzny rzutowe, nie izomorficzne z żadną płaszczyzną rzutową postaci $P(L)$, gdzie L jest trójwymiarową przestrzenią liniową nad dowolnym, niekoniecznie przemiennym, ciałem. Przyczyną tego zjawiska jest to, że w płaszczyznach rzutowych typu $P(L)$ twierdzenie Desargues'a jest, jak poprzednio wykazaliśmy spełnione, a tymczasem istnieją płaszczyzny rzutowe, dla których to twierdzenie nie zachodzi. Podamy bez dowodu następujący wynik.

9. Twierdzenie. *Następujące trzy własności klasycznej płaszczyzny rzutowej są równoważne*:

a) *Spełnione jest dla niej twierdzenie Desargues'a.*

b) *Istnieje jej zanurzenie w klasyczną przestrzeń rzutową.*

c) *Istnieje taka trójwymiarowa przestrzeń liniowa L nad niekoniecznie przemiennym ciałem K, określonym jednoznacznie z dokładnością do izomorfizmu, że ta płaszczyzna jest izomorficzna z $P(L)$.*

Implikacja b) $\Rightarrow$ a) wynika z nieskomplikowanego sprawdzenia tego faktu, że dowód przestrzennego twierdzenia Desargues'a wykorzystuje tylko aksjomaty T_1–T_6.

Implikacja c) $\Rightarrow$ b) wynika stąd, że L można zanurzyć w czterowymiarową przestrzeń liniową nad tym samym ciałem nieprzemiennym.

I wreszcie implikacja a) $\Rightarrow$ c), będąca najsubtelniejszą częścią dowodu, wykazywana jest za pomocą bezpośredniej konstrukcji ciała nieprzemiennego bazującej na własnościach płaszczyzny, w której jest spełnione twierdzenie Desargues'a. Istotnie, najpierw przy użyciu geometrycznej konstrukcji rzutowania środkowego wprowadza się pojęcie przekształceń rzutowych prostych rzutowych na płaszczyźnie. Dalej dowodzi się, że dla dwóch uporządkowanych trójek punktów leżących na dwóch prostych istnieje jednoznacznie wyznaczone przekształcenie rzutowe jednej prostej w drugą. A w końcu, ustalając prostą D z trójką punktów p_0, p_1, p_2, określa się zbiór $\boldsymbol{K}$ jako $D \setminus \{p_2\}$ z p_0 jako zerem i p_1 jako jedynką, a operacje dodawania i mnożenia w $\boldsymbol{K}$ wprowadza się geometrycznie za pomocą przekształceń rzutowych. Dla sprawdzenia aksjomatów ciała w istotny sposób jest wykorzystywane twierdzenie Desargues'a, w tym kontekście traktowane jako *aksjomat Desargues'a* Π_5.

10. Znaczenie twierdzenia Pappusa. Nawet jeśli w płaszczyźnie zachodzi twierdzenie Desargues'a, może nie być spełnione twierdzenie Pappusa. Nazywając odpowiednie stwierdzenie *aksjomatem Pappusa*, Π_6 możemy sformułować następujące twierdzenie, które także podamy bez dowodu.

11. Twierdzenie. a) *Jeśli w klasycznej płaszczyźnie rzutowej zachodzi aksjomat Pappusa, to zachodzi twierdzenie Desargues'a.*

b) *Klasyczna płaszczyzna rzutowa, w której zachodzi twierdzenie Desargues'a, spełnia aksjomat Pappusa wtedy i tylko wtedy, gdy związane z nią ciało jest przemienne, tj. gdy ta płaszczyzna jest izomorficzna z $P(L)$, gdzie L jest trójwymiarową przestrzenią liniową nad ciałem (przemiennym).*

Czytelnik może znaleźć dalsze szczegóły i opuszczone przez nas dowody w książce R. Hartshorne, *Podstawy geometrii rzutowej* (w jęz. ang.).

§ 10. Metryka Kählera

1. Jeśli L jest przestrzenią liniową unitarną nad $\boldsymbol{C}$, to na przestrzeni rzutowej $P(L)$ można wprowadzić specjalną metrykę, metrykę Kählera, nazwaną tak dla podkreślenia zasług E. Kählera, który odkrył jej ważne uogólnienia. Jej oryginalny wariant był wprowadzony jeszcze w poprzednim wieku przez Fubiniego i Study'ego. Odgrywa ona szczególnie ważną rolę w zespolonej geometrii algebraicznej i niejawnie także w mechanice kwantowej, gdyż przestrzenie takie jak $P(L)$ są, jak to podkreśliliśmy w części 2, przestrzeniami stanów układów kwantowo-mechanicznych.

Metryka ta jest niezmiennicza względem tych przekształceń rzutowych $P(L)$, które są postaci $P(f)$ dla odwzorowania unitarnego f przestrzeni L w siebie. Wprowadza się ją w taki sposób. Niech $p_1, p_2 \in P(L)$. Punktom p_1, p_2 odpowiadają dwa koła wielkie na sferze jednostkowej $S \subset L$, jak to pokazaliśmy w § 6, p. 3 c). Wówczas odległością Kählera $d(p_1, p_2)$ tych punktów nazywamy odległość pomiędzy tymi ko-

:ami w sferycznej metryce euklidesowej na S, tj. długość najkrótszego odcinka koła wielkiego na S, łączącego dwa punkty należące do przeciwobrazów p_1 i p_2.

Podstawowym celem tego paragrafu jest dowód następujących dwóch wzorów dla $d(p_1, p_2)$.

2. Twierdzenie. a) *Niech* $l_1, l_2 \in L$, $|l_1| = |l_2| = 1$ *i* $p_1, p_2 \in P(L)$ *będą prostymi* Cl_1, Cl_2. *Wówczas*

$$d(p_1, p_2) = \arccos |(l_1, l_2)|,$$

gdzie (l_1, l_2) *oznacza iloczyn skalarny w* L.

b) *Niech w* L *będzie wybrana baza ortonormalna i w* $P(L)$ *wprowadzony układ współrzędnych jednorodnych wyznaczony przez tę bazę. Niech dwa bliskie punkty* p_1, $p_2 \in P(L)$ *będą wyznaczone przez współrzędne* $(y_1, \ldots, y_n)$ *i* $(y_1 + dy_1, \ldots, y_n + dy_n)$ *względem mapy afinicznej* U_0 (por. § 6, p. 3 a). *Wówczas kwadrat odległości między tymi punktami z dokładnością do wyrazów trzeciego rzędu względem* dy_i *jest równy*

$$\frac{\sum_{i=1}^{n} |dy_i|^2}{1 + \sum_{i=1}^{n} |y_i|^2} - \frac{\left|\sum_{i=1}^{n} y_i \overline{dy_i}\right|^2}{\left(1 + \sum_{i=1}^{n} |y_i|^2\right)}$$

Dowód. a) W przestrzeni euklidesowej odległość między dwoma punktami na sferze jednostkowej jest równa długości tego z odcinków koła wielkiego łączącego te punkty, który jest nie dłuższy niż π, tj. kątowi euklidesowemu lub jeszcze inaczej „arkus-kosinusowi" euklidesowego iloczynu skalarnego wektorów wodzących tych punktów. Struktura euklidesowa na L odpowiadająca wyjściowej strukturze unitarnej jest zadana przez iloczyn skalarny $\mathrm{Re}(l_1, l_2)$. Ponieważ poszukujemy minimalnej odległości pomiędzy punktami kół wielkich $(e^{i\varphi} l_1)$, $(e^{i\psi} l_2)$, a arccos jest funkcją malejącą, więc należy tak dobrać φ i ψ, aby przy ustalonych l_1, l_2 wartość $\mathrm{Re}(e^{i\varphi} l_1, e^{i\psi} l_2)$ była możliwie największa. Ta wartość jednakże nie przekracza $|(l_1, l_2)|$ i przy odpowiednio dobranych φ, ψ osiąga tę wartość: jeśli $\varphi = -\arg(l_1, l_2)$, to $(e^{i\varphi} l_1, l_2) = |(l_1, l_2)|$. Dlatego ostatecznie

$$d(p_1, p_2) = \arccos |(l_1, l_2)|.$$

b) Oznaczmy

$$R = \left(1 + \sum_{i=1}^{n} |y_i|^2\right)^{1/2}, \quad R + dR = \left(1 + \sum_{i=1}^{n} |y_i + dy_i|^2\right)^{1/2}$$

Wówczas przeciwobrazami punktów $(y_1, \ldots, y_n)$ i $(y_1 + dy_1, \ldots, y_n + dy_n)$ na S są punkty

$$l_1 = \left(\frac{1}{R}, \frac{y_1}{R}, \ldots, \frac{y_n}{R}\right), \quad l_2 = \left(\frac{1}{R + dR}, \frac{y_1 + dy_1}{R + dR}, \ldots, \frac{y_n + dy_n}{R + dR}\right).$$

Stąd wynika

$$(l_1, l_2) = \frac{1}{R(R+dR)}\Big(1 + \sum_{i=1}^{n} y_i(\overline{y_i} + \overline{dy_i})\Big) = \frac{R^2 + \sum_{i=1}^{n} y_i \overline{dy_i}}{R(R+dR)}$$

oraz

$$|(l_1, l_2)|^2 = \frac{R^4 + R^2 \sum_{i=1}^{n} (y_i \overline{dy_i} + \overline{y_i} dy_i + \Big|\sum_{i=1}^{n} y_i \overline{dy_i}\Big|^2}{R^2(R+dR)^2}$$

Dalej

$$(R+dR)^2 = 1 + \sum_{i=1}^{n} |y_i + dy_i|^2 = R^2 + \sum_{i=1}^{n} (y_i \overline{dy_i} + \overline{y_i} dy_i + |dy_i|^2).$$

Dlatego z dokładnością do wyrazów trzeciego rzędu względem dy_i

$$|(l_1, l_2)|^2 = 1 - \frac{\sum_{i=1}^{n} |dy_i|^2}{R^2} + \frac{\Big|\sum_{i=1}^{n} y_i \overline{dy_i}\Big|^2}{R^4} + \cdots$$

Z drugiej strony, jeśli $\varphi = \arccos |(l_1, l_2)|$, to z dokładnością do φ^4 mamy dla małych wartości φ

$$|(l_1, l_2)|^2 = (\cos \varphi)^2 = \Big(1 - \frac{\varphi^2}{2} + \cdots\Big)^2 = 1 - \varphi^2 + \ldots$$

Porównanie tych wzorów kończy dowód.

§ 11. Rozmaitości algebraiczne i wielomiany Hilberta

1. Niech $P(L)$ będzie n-wymiarową przestrzenią rzutową nad ciałem $\boldsymbol{K}$ z ustalonym układem współrzędnych jednorodnych. Wielokrotnie rozważaliśmy już podprzestrzenie rzutowe i kwadryki, które są określone układami równań, odpowiednio

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

lub

$$\sum_{j,i=0}^{n} a_{ij} x_i x_j = 0, \quad a_{ij} = a_{ji}.$$

Ogólniej, rozważmy dowolny *wielomian jednorodny*, inaczej *formę*, stopnia $m \geq 1$:

$$F(x_0, \ldots, x_n) = \sum_{i_0 + \ldots + i_n = m} a_{i_0 \ldots i_n} x_0^{i_0} \cdots x^{i_n}.$$

Jakkolwiek nie określa on funkcji na $P(L)$, zbiór punktów, których współrzędne jednorodne $(x_0 : \ldots : x_n)$ spełniają równanie $F = 0$, jest wyznaczony jednoznacznie. Nazywa się go *hiperpowierzchnią algebraiczną* (stopnia m) wyznaczoną przez równanie $F = 0$.

W ogólności, zbiór punktów należących do $P(L)$ i spełniających układ równań

$$F_1 = F_2 = \cdots = F_k = 0,$$

gdzie F_i są formami, być może różnych stopni, nazywamy *rozmaitością algebraiczną* określoną przez ten układ równań.

Badanie rozmaitości algebraicznych w przestrzeni rzutowej stanowi jedno z podstawowych zadań geometrii algebraicznej. Jest jasne, że rozmaitość algebraiczna jest w ogólności obiektem wysoce nieliniowym i z tego względu, jak w innych działach geometrii tak i tu ważne miejsce zajmują metody linearyzacji problemów nieliniowych.

W tym paragrafie wprowadzimy jedną z takich metod, pochodzącą jeszcze od Hilberta i dającą przy minimum technicznych przygotowań ważne informacje o rozmaitościach algebraicznych $V \subset P(L)$. W skrócie, polega ona na tym, aby z rozmaitością algebraiczną V powiązać przeliczalną rodzinę przestrzeni liniowych $\{I_m(V)\}$ i zbadać zależność ich wymiaru jako funkcję m. A mianowicie, określimy $I_m(V)$ jako przestrzeń form stopnia m przyjmujących wartość zero na V. Wykażemy, że istnieje wielomian $Q_V(m)$ o współczynnikach wymiernych, taki że $\dim I_m(V) = Q_V(m)$ dla wszystkich dostatecznie dużych m. W istocie wykażemy o wiele dalej idący wynik, ale do jego sformułowania i dowodu będziemy potrzebować szeregu nowych pojęć.

2. Przestrzenie liniowe z gradacją. Ustalimy do końca rozważań podstawowe ciało skalarów $\boldsymbol{K}$. *Przestrzenią liniową nad* $\boldsymbol{K}$ *z gradacją* będziemy nazywać przestrzeń liniową L wraz z ustalonym rozkładem tej przestrzeni na sumę prostą podprzestrzeni: $L = \bigoplus_{i=0}^{\infty} L_i$. Wprawdzie ta suma jest nieskończona, ale każdy element $l \in L$ ma jednoznaczne przedstawienie w postaci skończonej sumy: $l = \sum_{i=0}^{\infty} l_i$, $l_i \in L_i$, w której tylko skończona liczba składników l_i, na ogół różnych dla każdego wektora, jest różna od zera. Wektor l_i nazywa się *składową jednorodną l stopnia i*; jeśli $l \in L_i$, to l nazywamy *elementem jednorodnym stopnia i*.

Oto przykład: pierścień wielomianów $A^{(n)}$ zmiennych $x_0, \ldots, x_n$ traktowany jako przestrzeń liniowa ma rozkład na sumę prostą $\bigoplus_{i=0}^{\infty} A_i^{(n)}$, gdzie $A_i^{(n)}$ składa się z wielomianów jednorodnych stopnia i. Zauważmy, że jeśli interpretować x_i jako funkcjonały współrzędnych na przestrzeni liniowej L, a elementy $A^{(n)}$ jak funkcje wielomianowe na tej przestrzeni, to liniowe nieosobliwe zamiany współrzędnych zachowują jednorodność i stopień wielomianu.

Inny przykład: $I = \bigoplus_{m=0}^{\infty} I_m(V)$, gdzie V jest dowolną rozmaitością algebraiczną. Oczywiście, $I \subset A^{(n)}$ oraz $I_m(V) = A_m^{(n)} \cap I$.

Ogólniej, będziemy mówić, że M jest *podprzestrzenią z gradacją* przestrzeni z gradacją $L = \bigoplus_{i=0}^{\infty} L_i$, gdy M jest podprzestrzenią liniową i $M = \bigoplus_{i=0}^{\infty}(M \cap L_i)$. Jest oczywiste, że równoważną definicją jest też następująca: każda składowa jednorodna dowolnego elementu M sama należy do M.

Jeśli $M \subset L$ jest parą składającą się z przestrzeni z gradacją i podprzestrzeni z gradacją, to przestrzeń ilorazowa L/M także ma naturalną gradację. Istotnie, rozważmy naturalne odwzorowanie liniowe

$$\bigoplus_{i=0}^{\infty} L_i/M_i \longrightarrow L/M : \sum_{i=0}^{\infty}(l_i + M_i) \mapsto \left(\sum_{i=0}^{\infty} l_i\right) + M$$

(skończone sumy na prawej stronie). Jest ono surjektywne, bo dowolny element $\left(\sum_{i=0}^{\infty} l_i\right) + M$, $l_i \in L_i$, jest obrazem elementu $\sum_{i=0}^{\infty}(l_i + M_i)$. Jest także injektywne, bo jeśli $\sum_{i=0}^{\infty} l_i + M = M$, to $\sum_{i=0}^{\infty} l_i \in M$, gdzie $l_i \in M_i$ na mocy założonej jednorodności M. Zatem jest to izomorfizm i możemy określić gradację L/M, przyjmując $(L/M)_i = L_i/M_i$.

Rodzina podprzestrzeni z gradacją przestrzeni L jest zamknięta względem przecięć i sum, a wszystkie izomorfizmy zwykłej algebry liniowej mają oczywiste analogony dla przestrzeni z gradacją.

3. Pierścienie z gradacją. Niech A będzie przestrzenią liniową nad $\boldsymbol{K}$ z gradacją i jednocześnie przemienną $\boldsymbol{K}$-algebrą z jedynką, w której mnożenie spełnia warunek

$$A_i A_j \subset A_{i+j}.$$

Takie A nazywamy *pierścieniem z gradacją* (a dokładniej $\boldsymbol{K}$-algebrą z gradacją). Ponieważ $\boldsymbol{K}A_i \subset A_i$, więc $\boldsymbol{K} \subset A_0$. Najważniejszy przykład stanowią pierścienie wielomianów $A^{(n)}$; dla nich mamy oczywiście $A_0^{(n)} = \boldsymbol{K}$.

4. Ideały z gradacją. *Ideałem* I dowolnego pierścienia przemiennego A nazywamy podzbiór będący addytywną podgrupą i zamknięty względem mnożenia przez elementy z A — jeśli $f \in I$ oraz $a \in A$, to $af \in I$. *Ideałem z gradacją pierścienia z gradacją* A nazywamy ideał, który jako podprzestrzeń $\boldsymbol{K}$-przestrzeni A ma gradację, tj. $I = \bigoplus_{m=0}^{\infty} I_m$, $I_m = I \cap A_m$.

Podstawowy przykład: ideały $I_m(V)$ rozmaitości algebraicznych w pierścieniach wielomianów $A^{(n)}$. Standardowy sposób konstrukcji ideałów jest następujący. Jeśli $S \subset A$ jest dowolnym podzbiorem, to zbiór wszystkich kombinacji liniowych $\left\{\sum_{s_i \in S} a_i s_i \mid a_i \in A\right\}$ jest ideałem A, *generowanym przez zbiór* S, a zbiór S nazywa się *układem generatorów* tego ideału. Jeśli ideał ma skończony zbiór generatorów, to nazywamy go *skończenie generowanym*. W przypadku z gradacją wystarczy rozważać zbiory S składające się tylko z elementów jednorodnych, a wtedy generowane przez

nie ideały są automatycznie ideałami z gradacją. Rzeczywiście, jednorodna składowa stopnia j dowolnej kombinacji liniowej $\sum a_i s_i$ jest także kombinacją liniową postaci $\sum a_i^{(k_i)} s_i$, gdzie $a_i^{(k_i)}$ oznacza jednorodną składową a_i stopnia $k_i = j - \deg s_i$ ($\deg s_i$ — stopień s_i). A zatem należy ona do ideału generowanego przez S. Jeśli ideał z gradacją jest skończenie generowany, to ma skończony układ jednorodnych generatorów, który składa się ze składowych jednorodnych elementów wyjściowego układu.

Dla uproszczenia dowodów musimy jeszcze uogólnić pojęcie ideału z gradacją, a także rozpatrzyć moduły z gradacją. To będą ostatnie z listy potrzebnych nam pojęć.

5. Moduły z gradacją. *Modułem* M nad pierścieniem przemiennym A albo A-modułem nazywamy grupę addytywną wyposażoną w operację $A \times M \to M$: $(a, m) \mapsto am$, która jest łączna ($(ab)m = a(bm)$ dla każdych $a, b \in A$, $m \in M$) i rozdzielna względem obu argumentów:

$$(a + b)m = am + bm, \quad a(m + n) = am + an.$$

Ponadto zakładamy, że $1m = m$ dla każdego $m \in M$, gdzie 1 oznacza jedność pierścienia A.

Jeśli A jest ciałem, to M jest po prostu przestrzenią liniową nad A — można powiedzieć, że pojęcie modułu jest uogólnieniem pojęcia przestrzeni liniowej na przypadek, gdy skalary tworzą tylko pierścień (por. *Wstęp do algebry*, rozdz. 9, § 3).

Jeśli A jest $\boldsymbol{K}$-algebrą z gradacją, to *A-modułem z gradacją* nazwiemy taki A-moduł, który jest przestrzenią liniową nad $\boldsymbol{K}$ z gradacją, $M = \bigoplus_{i=0}^{\infty} M_i$ i ponadto

$$A_i M_j \subset M_{i+j}$$

dla wszystkich $i, j \geq 0$.

Przykłady: a) A jest A-modułem z gradacją.

b) Każdy ideał z gradacją w A jest A-modułem z gradacją.

Jeśli M jest A-modułem z gradacją, to każda podprzestrzeń z gradacją $N \subset M$, zamknięta względem mnożenia przez elementy z A, jest sama A-modułem z gradacją. Dla dowolnego układu elementów jednorodnych $S \subset M$ można skonstruować generowany przezeń moduł z gradacją, składający się ze wszystkich skończonych kombinacji liniowych $\sum a_i s_i$, $a_i \in A$, $s_i \in S$. Jeśli jest on równy M, to S jest nazywane jednorodnym układem generatorów A. Moduł nazywany jest skończenie generowanym, jeśli ma skończony układ generatorów. Jeśli moduł z gradacją ma jakikolwiek skończony układ generatorów, to ma także skończony układ jednorodnych generatorów — składowe jednorodne elementów układu wyjściowego.

Rozpatrywanie w naszym problemie wszystkich modułów, a nie tylko ideałów prowadzi do większej swobody działania. Mnożenie przez $a \in A$ w module ilorazowym jest dane wzorem

$$a(m + N) = am + N.$$

Przy zadanej gradacji na M można jak poprzednio wprowadzić gradację na M/N wzorem $(M/N)_i = M_i/N_i$, a poprawność definicji daje się sprawdzić bez trudu.

Elementy teorii sum prostych, podmodułów i modułów ilorazowych formalnie nie różnią się od odpowiadających rezultatów dla przestrzeni liniowych.

Teraz możemy przejść do dowodu podstawowych wyników tego paragrafu.

6. Twierdzenie. *Niech M będzie dowolnym skończenie generowanym modułem nad pierścieniem wielomianów $A^{(n)} = \boldsymbol{K}[x_0, \ldots, x_n]$ skończonej liczby zmiennych. Wówczas każdy podmoduł $N \subset M$ jest skończenie generowany.*

D o w ó d. Rozbijemy go na kilka etapów. Wprowadzimy standardową terminologię: *moduł, w którym każdy podmoduł jest skończenie generowany*, nazywa się *modułem noetherowskim* (na cześć Emmy Noether).

a) *Moduł M jest noetherowski wtedy i tylko wtedy, gdy każdy wstępujący ciąg podmodułów $M_1 \subset M_2 \subset \cdots$ modułu M stabilizuje się, tj. istnieje takie a_0, że $M_a = M_{a+1}$ dla wszystkich $a \geq a_0$.*

W istocie, niech M będzie noetherowski. Oznaczmy $N = \bigcup_{i=0}^{\infty} M_i$. Niech $n_1, \ldots, n_k$ będzie skończonym układem generatorów N. Dla każdego $1 \leq j \leq k$ istnieje taki $i(j)$, że $n_j \subset M_{i(j)}$. Oznaczmy $a_0 = \max\{i(j) \mid j = 1, \ldots, k\}$. Wtedy M_a zawiera $n_1, \ldots, n_k$ dla każdego $a \geq a_0$, a więc $M_a = N$.

Odwrotnie, załóżmy, że każdy wstępujący ciąg podmodułów M urywa się. Skonstruujemy układ generatorów podmodułu $N \subset M$ przez indukcję. Jako $n_1 \in N$ weźmiemy dowolny element; jeśli $n_1, \ldots, n_i \in N$ są już skonstruowane, to oznaczymy przez $M_i \subset N$ generowany przez nie podmoduł; jeśli $N \neq M_i$, to wybierzemy $n_{i+1} \in N \setminus M_i$. Ten proces musi się zakończyć, bo w przeciwnym przypadku ciąg $M_1 \subset \cdots \subset M_i \subset \cdots$ nie byłby od pewnego miejsca stały.

b) *Jeśli podmoduł $N \subset M$ i moduł ilorazowy M/N są noetherowskie, to M też jest noetherowski; odwrotne twierdzenie jest także spełnione.*

Rzeczywiście, niech $M_1 \subset M_2 \subset \cdots$ będzie ciągiem podmodułów M. Niech a_0 oznacza taki wskaźnik, że obydwa ciągi $M_1 \cap N \subset M_2 \cap N \subset \cdots$ i $(M_1 + N)/N \subset (M_2 + N)/N \subset \cdots$ są stałe dla $a \geq a_0$. Wówczas ciąg $M_1 \subset M_2 \subset \cdots$ jest też stały dla $a \geq a_0$. Odwrotna implikacja jest trywialna.

c) *Suma prosta skończonej liczby modułów noetherowskich jest modułem noetherowskim.*

Rzeczywiście, niech $M = \bigoplus_{i=1}^{n} M_i$, gdzie M_i są noetherowskie. Przeprowadzimy indukcję względem n. W przypadku $n = 1$ teza jest oczywista, załóżmy więc, że $n \geq 2$. W takim przypadku M zawiera podmoduł izomorficzny z M_n, a odpowiadający mu moduł ilorazowy jest izomorficzny z $\bigoplus_{i=1}^{n-1} M_i$. Oba te moduły są noetherowskie, więc M jest także noetherowski na mocy b).

d) *Pierścień $A^{(n)}$ jest noetherowski jako moduł nad samym sobą. Innymi słowy, każdy ideał $A^{(n)}$ jest skończenie generowany.*

Ten podstawowy przypadek szczególny naszego twierdzenia został wykazany już

przez Hilberta(1). Dowodzi się go przez indukcję względem n. W przypadku $n = -1$, tj. dla $A^{(-1)} = \boldsymbol{K}$ teza jest oczywista. Istotnie, każdy ideał I ciała $\boldsymbol{K}$ jest równy albo $\{0\}$, albo $\boldsymbol{K}$: jeśli $a \in I$, $a \neq 0$, to $b = (ba^{-1})a \in I$ dla każdego $b \in \boldsymbol{K}$. Krok indukcyjny opiera się o przedstawienie $A^{(n)}$ jako $A^{(n-1)}[x_n]$. Niech $I^{(n)} \subset A^{(n)}$ będzie ideałem. Zapiszmy każdy element z $I^{(n)}$ w postaci wielomianu względem potęg x_n o współczynnikach z $A^{(n-1)}$. Zbiór wszystkich współczynników stojących przy najwyższych potęgach elementów z $I^{(n)}$ stanowi ideał $I^{(n-1)}$ w $A^{(n-1)}$. Na mocy założenia indukcyjnego ma on skończony układ generatorów $\varphi_1, \ldots, \varphi_m$. Dla każdego z tych generatorów dobierzemy należący do $I^{(n)}$ wielomian $f_i = \varphi_i x_n^{d_i} + \ldots$, gdzie wielokropek zastępuje człony niższego stopnia względem x_n. Oznaczmy $d = \max\limits_{1 \leq i \leq m} \{d_i\}$ i niech $I \subset I^{(n)}$ oznacza ideał w $A^{(n)}$ generowany przez wielomiany $f_1, \ldots, f_m$.

Dowolny element f z $I^{(n)}$ możemy przedstawić w postaci $f = \varphi x_n^s + (\textit{wyrazy niższych stopni})$. Z definicji $\varphi \in I^{(n-1)}$, więc $\varphi = \alpha_1 \varphi_1 + \cdots + \alpha_m \varphi_m$. Jeśli $s \geq d$, to wielomian $f - \sum \alpha_i f_i x_n^{s-d_i}$ należy do $I^{(n)}$ i jego stopień $< s$. Postępując w analogiczny sposób, otrzymamy wyrażenie $f = g + h$, gdzie $h \in I$, a g jest wielomianem stopnia niższego niż d, należącym do $I^{(n)}$.

Wszystkie wielomiany z $I^{(n)}$ stopnia $< d$ tworzą podmoduł J $A^{(n-1)}$-modułu generowanego przez zbiór skończony $\{1, x_n, \ldots, x_n^{d-1}\}$. Na mocy założenia indukcyjnego $A^{(n-1)}$ jest modułem noetherowskim, więc z c) powyżej wynika, że podmoduł J jest skończenie generowany.

Wykazaliśmy zatem, że $I^{(n)} = I + J$ jest sumą dwóch modułów skończenie generowanych, więc jest też skończenie generowany.

Teraz możemy już bez trudności zakończyć dowód twierdzenia.

Załóżmy, że moduł M nad $A^{(n)}$ ma skończony układ generatorów $m_1, \ldots, m_k$. Zatem istnieje surjektywny homomorfizm $A^{(n)}$-modułów

$$\underbrace{A^{(n)} \oplus \cdots \oplus A^{(n)}}_{k \text{ razy}} \longrightarrow M : (f_1, \ldots, f_k) \mapsto \sum_{i=1}^{k} f_i m_i.$$

Ponieważ moduł $A^{(n)} \oplus \cdots \oplus A^{(n)}$ jest noetherowski na mocy d) i c), więc także moduł ilorazowy M jest noetherowski.

7. Wniosek. *Każdy nieskończony układ równań $F_i = 0$, $i \in S$, gdzie F_i są wielomianami z $A^{(n)}$, jest równoważny z pewnym swoim podukładem skończonym.*

Dowód. Oznaczmy przez I ideał generowany przez wszystkie F_i i niech $\{G_i\}$ będzie jego skończonym układem generatorów. Rozważmy taki skończony podzbiór $S_0 \subset S$, że wszystkie G_i wyrażają się w sposób liniowy przez F_i, $i \in S_0$. Wówczas układ równań $F_i = 0$, $i \in S_0$, jest równoważny z układem wyjściowym, tj. oba mają ten sam zbiór rozwiązań.

(1) i w tym przypadku nosi on nazwę *twierdzenia Hilberta o bazie* (*przyp. tłum.*).

8. Wielomiany Hilberta i szereg Poincarégo modułu z gradacją. Niech teraz M oznacza skończenie generowany moduł nad pierścieniem $A^{(n)}$ z gradacją. Wówczas każda przestrzeń liniowa $M_k \subset M$ ma wymiar skończony nad $\boldsymbol{K}$. Rzeczywiście, jeśli $\{a_i\}$ oznacza $\boldsymbol{K}$-bazę jednorodną $A^{(n)} + \cdots + A^{(n)}$ i $\{m_i\}$ oznacza skończony układ generatorów M, to jako przestrzeń liniowa M_k jest generowana przez skończony zbiór elementów $a_i m_j$ spełniających $\deg a_i + \deg m_j = k$.

Oznaczmy $d_k(M) = \dim M_k$. *Szeregiem Poincarégo* modułu M nazywamy szereg formalny zmiennej t

$$H_M(t) = \sum_{k=0}^{\infty} d_k(M) t^k.$$

9. Twierdzenie. a) *Przy założeniach poprzedzającego punktu istnieje taki wielomian $f(t)$ ze współczynnikami całkowitymi, że*

$$H_M(t) = \frac{f(t)}{(1-t)^{n+1}}.$$

b) *Przy tych samych założeniach istnieje taki wielomian $P(k)$ o współczynnikach wymiernych i taka liczba N, że*

$$d_k(M) = P(k) \quad \text{dla każdego } k \geq N.$$

Dowód. Najpierw wyprowadzimy drugą część twierdzenia z pierwszej. Zapiszmy $f(t) = \sum_{i=0}^{N} a_i t^i$ i porównajmy współczynniki przy t^k po obu stronach tożsamości

$$H_M(t) = f(t)(1-t)^{-(n+1)}.$$

Uwzględniając, że $(1-t)^{-(n+1)} = \sum_{k=0}^{\infty} \binom{n+k}{n} t^k$, otrzymujemy

$$d_k(M) = \sum_{i=0}^{\min(k,N)} a_i \binom{n+k-i}{n}.$$

Dla $k \geq N$ wyrażenie po prawej jest wielomianem zmiennej k o współczynnikach wymiernych.

Teraz za pomocą indukcji względem n wykażemy tezę z punktu a). Wygodnie przyjąć $A^{(-1)} = \boldsymbol{K} = A_0^{(-1)}$; $A_i^{(-1)} = \{0\}$ dla $i \geq 1$. Skończenie generowany moduł nad $A^{(-1)}$ z gradacją to po prostu skończenie wymiarowa przestrzeń wymiarowa nad $\boldsymbol{K}$, zapisana w postaci sumy prostej $\sum_{k=0}^{N} M_k$. Szeregiem Poincarégo tej przestrzeni jest wielomian $\sum_{k=0}^{N} \dim M_k t^k$, a więc teza jest trywialnie spełniona.

Przyjmiemy teraz, że teza jest wykazana dla $A^{(n-1)}$, $n \geq 0$, i wykażemy ją dla $A^{(n)}$. Niech M będzie skończenie generowanym modułem nad $A^{(n)}$ z gradacją. Oznaczmy

$$K = \{m \in M \mid x_n m = 0\}, \quad C = M/x_n M.$$

Jest jasne, że K i $x_n M$ są podmodułami z gradacją modułu M i dlatego C ma także strukturę $A^{(n)}$-modułu z gradacją. Jednakże mnożenie przez x_n przeprowadza zarówno K, jak i C na zero. A zatem, jeśli rozważać K i C jako moduły nad podpierścieniem $A^{(n-1)} = \boldsymbol{K}[x_0, \ldots, x_{n-1}] \subset A^{(n)} = \boldsymbol{K}[x_0, \ldots, x_n]$, to każdy ich układ generatorów nad $A^{(n)}$ będzie jednocześnie układem generatorów nad $A^{(n-1)}$. Na mocy twierdzenia p. 6 K jest skończenie generowany nad $A^{(n)}$ jako podmoduł skończenie generowanego modułu. Z drugiej strony C jest skończenie generowany nad $A^{(n)}$, gdyż jeśli $m_1, \ldots, m_p$ generują M, to $m_1 + x_n M, \ldots, m_p + x_n M$ generują C. Stąd wynika, że K i C są skończenie generowane nad $A^{(n-1)}$ i można do nich stosować założenie indukcyjne. Na mocy dokładności ciągów przestrzeni liniowych nad $\boldsymbol{K}$

$$0 \longrightarrow K_m \longrightarrow M_m \xrightarrow{x_n} M_{m+1} \longrightarrow C_{m+1} \longrightarrow 0, \quad m \geq 0,$$

wnioskujemy, że

$$\dim M_{m+1} - \dim M_m = \dim C_{m+1} - \dim K_m.$$

Mnożąc otrzymaną równość przez t^{m+1} i sumując względem m od 0 do ∞, dostajemy

$$H_M(t) - \dim M_0 - tH_M(t) = H_C(t) - \dim C_0 - tH_K(t),$$

a następnie wykorzystując założenie indukcyjne dla K i C

$$(1-t)H_M(t) = \dim M_0 - \dim C_0 + \frac{f_C(t)}{(1-t)^n} + \frac{tf_K(t)}{(1-t)^n},$$

gdzie $f_C(t)$ i $f_K(t)$ są wielomianami o współczynnikach całkowitych. Stąd wynika teza twierdzenia.

10. Wymiar i stopień rozmaitości algebraicznej. Niech teraz $V \subset P^n$ będzie rozmaitością algebraiczną, której odpowiada ideał $I(V)$. Rozpatrzmy wielomian Hilberta $P_V(k)$ modułu ilorazowego $A^{(n)}/I(V)$:

$$P_V(k) = \dim A_k^{(n)}/I_k(V) \quad \text{dla każdego } k \geq k_0.$$

Łatwo przekonać się, że $P_{P^n}(k) = \binom{n+k}{n}$, tak że $\deg P_n(k) = n$. Zatem $\deg P_V \leq n$. Liczbę $d = \deg P_V$ nazywa się *wymiarem* rozmaitości V. Zapiszmy najwyższy wyraz $P_V(k)$ w postaci $e\frac{k^d}{d!}$. Można wykazać, że e jest liczbą całkowitą — nazywa się ją *stopniem* rozmaitości V. Wymiar i stopień są najważniejszymi charakterystykami „rozmiarów" rozmaitości algebraicznej. Można także podać czysto geometryczną

definicję tych wielkości: jeśli ciało $\boldsymbol{K}$ jest algebraicznie domknięte, to d-wymiarowa rozmaitość stopnia e przecina „dostatecznie ogólną" przestrzeń rzutową o dopełniającym wymiarze $P^{n-d} \subset P^n$ dokładnie w e różnych punktach. Nie będziemy dowodzić tego twierdzenia.

Na zakończenie zauważmy, że przez około pół wieku po odkryciu Hilberta pozostawało nierozwiązane zagadnienie znalezienia interpretacji wartości wielomianu Hilberta $P_V(k)$ dla tych liczb k, dla których $P_V(k) \neq \dim I_k(V)$, a w szczególności dla ujemnych k. Został on rozwiązany dopiero w latach pięćdziesiątych wraz z rozwojem teorii snopów koherentnych, kiedy okazało się, że dla dowolnego k wartość $P_V(k)$ jest sumą alternującą wymiarów pewnych przestrzeni kohomologii rozmaitości V. Analogiczna interpretacja wielomianów Hilberta została także podana dla dowolnych skończenie generowanych modułów z gradacją.

Ćwiczenia

1. Wykazać, że wielomian Hilberta przestrzeni rzutowej P^n nie zależy od wymiaru m przestrzeni rzutowej P^m, w którą P^n została zanurzona.

2. Wyznaczyć wielomian Hilberta dla modułu $A^{(n)}/FA^{(n)}$, gdzie F oznacza formę stopnia e.

Algebra wieloliniowa

§ 1. Iloczyny tensorowe przestrzeni wektorowych

1. Ostatnia część naszej książki jest poświęcona systematycznemu zbadaniu wieloliniowych konstrukcji algebry liniowej. Podstawą aparatu algebraicznego jest pojęcie iloczynu tensorowego, wprowadzone w tym paragrafie i szczegółowo badane w paragrafach następnych. Jest nieco rozczarowujące, że główne zastosowania tego formalizmu leżą poza granicami algebry liniowej i odnoszą się do geometrii różniczkowej, teorii reprezentacji grup i mechaniki kwantowej. Będziemy więc mogli jedynie pobieżnie wspomnieć o nich w ostatnich paragrafach.

2. Konstrukcja. Rozważmy skończoną rodzinę przestrzeni liniowych $L_1, \ldots, L_p$ nad jednym i tym samym ciałem $\boldsymbol{K}$. Przypomnijmy, że odwzorowanie $L_1 \times \cdots \times L_p \longrightarrow L$, gdzie L jest jeszcze jedną przestrzenią liniową nad $\boldsymbol{K}$, nazywa się wieloliniowe, gdy jest ono liniowe względem każdego z argumentów $l_i \in L_i$, $i = 1, \ldots, p$, przy ustalonych pozostałych.

Naszym najbliższym celem jest konstrukcja *uniwersalnego* wieloliniowego odwzorowania przestrzeni $L_1, \ldots, L_p$. Jego obraz nazwiemy iloczynem tensorowym tych przestrzeni. Precyzyjny sens stwierdzenia o uniwersalności będzie wyjaśniony w dalszym ciągu przy sformułowaniu twierdzenia 3. Konstrukcja przebiega w trzech etapach.

a) Określimy *przestrzeń* $\mathcal{M}$ jako zbiór wszystkich funkcji zadanych na $L_1 \times \cdots \times L_p$ i przyjmujących wartości w $\boldsymbol{K}$, które są równe zeru dla prawie wszystkich elementów $L_1 \times \cdots \times L_p$, tj. dla wszystkich poza skończoną liczbą. Zbiór ten stanowi przestrzeń liniową względem zwykłych operacji punktowego dodawania i mnożenia przez skalary należące do $\boldsymbol{K}$. Bazę tej przestrzeni tworzą funkcje delta $\delta(l_1, \ldots, l_p)$ równe jedności w jedynym punkcie $(l_1, \ldots, l_p) \in L_1 \times \cdots \times L_p$ i zeru we wszystkich pozostałych. Opuszczając symbol δ możemy uważać, że $\mathcal{M}$ składa się z formalnych kombinacji liniowych układów $(l_1, \ldots, l_p) \in L_1 \times \cdots \times L_p$:

$$\mathcal{M} = \left\{ \sum a_{i_1 \ldots i_p} (l_1, \ldots, l_p) \mid a_{i_1 \ldots i_p} \in \boldsymbol{K} \right\}.$$

Zauważmy, że jeśli ciało $\boldsymbol{K}$ jest nieskończone i choćby jedna z przestrzeni L_i nie jest zerowymiarowa, to $\mathcal{M}$ ma wymiar nieskończony.

b) *Podprzestrzeń* $\mathcal{M}_0$. Na mocy definicji jest ona rozpięta przez wektory postaci

$$(l_1,\dots,l'_j+l''_j,\dots,l_p)-(l_1,\dots,l'_j,\dots,l_p)-(l_1,\dots,l''_j,\dots,l_p),$$
$$(l_1,\dots,al_j,\dots,l_p)-a(l_1,\dots,l_j,\dots,l_p),\quad a\in\boldsymbol{K}.$$

c) *Iloczyn tensorowy* $L_1\otimes\cdots\otimes L_p$. Z definicji

$$L_1\otimes\cdots\otimes L_p=\mathcal{M}/\mathcal{M}_0,$$
$$l_1\otimes\cdots\otimes l_p=(l_1,\dots,l_p)+\mathcal{M}_0\in L_1\otimes\cdots\otimes L_p,$$
$$t\colon L_1\times\cdots\times L_p\longrightarrow L_1\otimes\cdots\otimes L_p,\quad t(l_1,\dots,l_p)=l_1\otimes\cdots\otimes l_p.$$

Tutaj $\mathcal{M}/\mathcal{M}_0$ jest przestrzenią ilorazową w zwykłym sensie. Elementy $L_1\otimes\cdots\otimes L_p$ nazywają się *tensorami*, $l_1\otimes\dots\otimes l_p$ — *tensorami rozkładalnymi* lub *prostymi*. Ponieważ układy $(l_1,\dots,l_p)$ tworzą bazę $\mathcal{M}$, więc tensory rozkładalne $l_1\otimes\cdots\otimes l_p$ rozpinają przestrzeń $L_1\otimes\cdots\otimes L_p$ iloczynu tensorowego, ale w żadnej mierze nie stanowią bazy, bo zachodzi między nimi wiele zależności liniowych.

Podstawową własność iloczynów tensorowych opisuje następujące twierdzenie.

3. Twierdzenie. a) *Odwzorowanie kanoniczne*

$$t\colon L_1\times\cdots\times L_p\longrightarrow L_1\otimes\cdots\otimes L_p,\quad (l_1,\dots,l_p)\mapsto l_1\otimes\cdots\otimes l_p,$$

jest odwzorowaniem wieloliniowym.

b) *Odwzorowanie wieloliniowe t jest uniwersalne w następującym znaczeniu: dla dowolnej przestrzeni liniowej M nad ciałem* $\boldsymbol{K}$ *i każdego odwzorowania wieloliniowego* $s: L_1\times\cdots\times L_p\longrightarrow M$ *istnieje dokładnie jedno odwzorowanie liniowe* $f: L_1\otimes\cdots\otimes L_p\to M$ *spełniające* $s=f\circ t$. Będziemy mówić krótko, że *można przeprowadzić s poprzez f*.

Dowód. a) Musimy sprawdzić następujące wzory:

$$l_1\otimes\cdots\otimes(l'_j+l''_j)\otimes\cdots\otimes l_p=l_1\otimes\cdots\otimes l'_j\otimes\cdots\otimes l_p+l_1\otimes\cdots\otimes l''_j\otimes\cdots\otimes l_p,$$
$$l_1\otimes\cdots\otimes al_j\otimes\cdots\otimes l_p=a(l_1\otimes\cdots\otimes l_j\otimes\cdots\otimes l_p),$$

co, na przykład, dla pierwszego wzoru oznacza

$$(l_1,\dots,l'_j+l''_j,\dots,l_p)+\mathcal{M}_0=\big[(l_1,\dots,l'_j,\dots,l_p)+\mathcal{M}_0\big]+\big[(l_1,\dots,l''_j,\dots,l_p)+\mathcal{M}_0\big].$$

Na mocy definicji przestrzeni ilorazowej (część 1, § 6, p. 2 i 3) i po uwzględnieniu opisu elementów rozpinających podprzestrzeń $\mathcal{M}_0$ danego w p. 2 b) tego paragrafu dostajemy żądane równości wprost z definicji.

b) Jeśli w ogóle f istnieje, to warunek $s=f\circ t$ jednoznacznie określa wartości f na tensorach rozkładalnych:

$$f(l_1\otimes\cdots\otimes l_p)=f\circ t(l_1,\dots,l_p)=s(l_1,\dots,l_p).$$

Wobec tego, że te ostatnie rozpinają $L_1 \otimes \cdots \otimes L_p$ odwzorowanie f jest co najwyżej jedno.

Dla wykazania istnienia f, rozważymy odwzorowanie liniowe $g\colon \mathcal{M} \to M$, określone na elementach bazowych wzorem

$$g(l_1, \ldots, l_p) = s(l_1, \ldots, l_p),$$

inaczej mówiąc

$$g(\sum a_{i_1,\ldots,i_p}(l_1, \ldots, l_p)) = \sum a_{i_1\ldots i_p} s(l_1, \ldots, l_p).$$

Jak łatwo sprawdzić mamy $\mathcal{M}_0 \subset \operatorname{Ker} g$. Istotnie,

$$g\left[(l_1, \ldots, l'_j + l''_j, \ldots, l_p) - (l_1, \ldots, l'_j, \ldots, l_p) - (l_1, \ldots, l''_j, \ldots, l_p)\right] =$$
$$= s(l_1, \ldots, l'_j + l''_j, \ldots, l_p) - s(l_1, \ldots, l'_j, \ldots, l_p) - s(l_1, \ldots, l''_j, \ldots, l_p) = 0$$

na mocy wieloliniowości s. W analogiczny sposób można sprawdzić, że g przeprowadza na zero tworzące $\mathcal{M}_0$ drugiego typu, związane z pomnożeniem jednego z czynników przez skalar.

Wynika stąd, że g indukuje odwzorowanie liniowe

$$f\colon \mathcal{M}/\mathcal{M}_0 = L_1 \otimes \cdots \otimes L_p \longrightarrow M$$

(por. część 1, stwierdzenie § 6, p. 8), które spełnia

$$f(l_1 \otimes \cdots \otimes l_p) = s(l_1, \ldots, l_p).$$

To kończy dowód.

Wyprowadzimy teraz kilka bezpośrednich wniosków z tego twierdzenia i przedstawimy pewne jego zastosowania.

4. Niech $\mathcal{L}(L_1, \ldots, L_p; M)$ oznacza zbiór odwzorowań wieloliniowych $L_1 \times \cdots \times L_p$ w M. Na podstawie twierdzenia p. 3 zbudujemy odwzorowanie zbiorów

$$\mathcal{L}(L_1, \ldots, L_p; M) \longrightarrow \mathcal{L}(L_1 \otimes \cdots \otimes L_p, M),$$

przyporządkowujące odwzorowaniu wieloliniowemu s odwzorowanie liniowe f, takie że $s = f \circ t$. Przestrzenie po lewej i prawej stronie są przestrzeniami liniowymi nad $\boldsymbol{K}$ (jako przestrzenie funkcji o wartościach w przestrzeni wektorowej M — dodawanie i mnożenie przez skalary wykonuje się punktowo). Uwzględniając ten fakt, wprost z konstrukcji wnioskujemy, że nasze odwzorowanie jest liniowe, a co więcej jest izomorfizmem przestrzeni liniowych. Rzeczywiście, jego surjektywność wynika stąd, że dla każdego odwzorowania liniowego $f\colon L_1 \otimes \cdots \otimes L_p \to M$ odwzorowanie $s = f \circ t$ jest wieloliniowe na mocy części a) twierdzenia p. 3. Injektywność wynika

stąd, że jeśli $s \neq 0$, to $f \circ t \neq 0$, więc także $f \neq 0$. W wyniku otrzymaliśmy kanoniczne utożsamienie przestrzeni liniowych

$$\mathcal{L}(L_1, \ldots, L_p; M) = \mathcal{L}(L_1 \otimes \cdots \otimes L_p, M).$$

W ten sposób za pomocą konstrukcji iloczynu tensorowego przestrzeni można sprowadzić badanie odwzorowań wieloliniowych do badania odwzorowań liniowych poprzez zdefiniowanie nowej operacji w kategorii przestrzeni liniowych.

5. Wymiar i bazy. a) *Jeśli choć jedna z przestrzeni $L_1, \ldots, L_p$ jest zerowa, to iloczyn tensorowy tych przestrzeni jest przestrzenią zerową.*

Istotnie, powiedzmy, że $L_j = \{0\}$. Dowolne odwzorowanie wieloliniowe $f\colon L_1 \times \cdots \times L_p \to M$ przy ustalonych $l_i \in L_i$, $i \neq j$, jest liniowe na przestrzeni L_j. Jednakże jedynym odwzorowaniem liniowym przestrzeni zerowej jest odwzorowanie zerowe, a to oznacza, że $f = 0$ dla każdych wartości argumentów. W szczególności, uniwersalne odwzorowanie wieloliniowe $t\colon L_1 \times \cdots \times L_p \to L_1 \otimes \cdots \otimes L_p$ jest odwzorowaniem zerowym, a że jego obraz rozpina przestrzeń iloczynu tensorowego, więc ta ostatnia jest też zerowymiarowa.

b) $\dim(L_1 \otimes \cdots \otimes L_p) = \dim L_1 \ldots \dim L_p$.

Jeśli choćby jedna z przestrzeni jest zerowa, to równość wynika z poprzedniej obserwacji. W przeciwnym przypadku równości wymiarów dowiedziemy w następujący sposób. Zaobserwujmy, że wymiar przestrzeni $L_1 \otimes \cdots \otimes L_p$ jest równy wymiarowi przestrzeni dwoistej $\mathcal{L}(L_1 \otimes \cdots \otimes L_p, \boldsymbol{K})$. Tę ostatnią przestrzeń utożsamiliśmy w p. 4 z przestrzenią odwzorowań wieloliniowych $\mathcal{L}(L_1, \ldots, L_p; \boldsymbol{K})$. Wybierzmy w przestrzeniach L_i bazy $\{e_1^{(i)}, \ldots, e_{n_i}^{(i)}\}$. Każdemu odwzorowaniu wieloliniowemu

$$f\colon L_1 \times \cdots \times L_p \longrightarrow \boldsymbol{K}$$

przyporządkujemy układ $n_1 \cdots n_p$ liczb

$$f(e_{i_1}^{(i)}, \ldots, e_{i_p}^{(p)}), \quad 1 \leq i_j \leq n_j,\ 1 \leq j \leq p.$$

Na mocy wieloliniowości ten układ jednoznacznie wyznacza f:

$$f\Big(\sum_{i_1=1}^{n_1} x_{i_1}^{(1)} e_{i_1}^{(1)}, \ldots, \sum_{i_p=1}^{n_p} x_{i_p}^{(p)} e_{i_p}^{(p)}\Big) = \sum_{i_1, \ldots, i_p} x_{i_1}^{(1)} \cdots x_{i_p}^{(p)} f(e_{i_1}^{(1)}, \ldots, e_{i_p}^{(p)}),$$

a poza tym jest dowolny: prawa strona ostatniego wzoru określa wieloliniowe odwzorowanie wektorów $\left(\vec{x}^{(1)}, \ldots, \vec{x}^{(p)}\right)$ przy każdych wartościach współczynników. A to oznacza, że przestrzeń wieloliniowych odwzorowań $L_1 \times \cdots \times L_p \to \boldsymbol{K}$ ma wymiar $n_1 \cdots n_p = \dim L_1 \cdots \dim L_p$, co kończy dowód.

c) *Baza tensorowa* $L_1 \otimes \cdots \otimes L_p$. Z poprzedzających rozważań wynika także, że iloczyny tensorowe $e_{i_1}^{(1)} \otimes \cdots \otimes e_{i_p}^{(p)}$ tworzą bazę przestrzeni $L_1 \otimes \cdots \otimes L_p$ (przy tym przyjmujemy, że wszystkie rozważane przestrzenie mają wymiar ≥ 1 i, dla prostoty,

są skończenie wymiarowe). W istocie, te iloczyny tensorowe rozpinają $L_1 \otimes \cdots \otimes L_p$, bo każdy tensor rozkładalny wyraża się w postaci ich kombinacji liniowych. Co więcej, liczba tych iloczynów jest równa wymiarowi $L_1 \otimes \cdots \otimes L_p$.

6. Iloczyny tensorowe przestrzeni funkcji. Rozważmy zbiory skończone $S_1, \ldots, S_p$ i niech $F(S_i)$ oznacza przestrzeń funkcji na S_i o wartościach w $\boldsymbol{K}$. Wprowadzimy kanoniczne utożsamienie

$$F(S_1 \times \cdots \times S_p) = F(S_1) \otimes \cdots \otimes F(S_p),$$

przyporządkowując funkcji $\delta_{(s_1, \ldots, s_p)}$ iloczyn tensorowy $\delta_{s_1} \otimes \cdots \otimes \delta_{s_p}$ (por. § 1, p. 7, część 1). Funkcje tego rodzaju tworzą bazy w odpowiednich przestrzeniach, więc to przyporządkowanie rzeczywiście wyznacza izomorfizm. Jeśli $f_i \in F(S_i)$, to

$$f_1 \otimes \cdots \otimes f_p = \Big(\sum_{s_1 \in S_1} f_1(s_1)\delta_{s_1} \Big) \otimes \cdots \otimes \Big(\sum_{s_p \in S_p} f_p(s_p)\delta_{s_p} \Big)$$

odpowiada przy tym izomorfizmie funkcji

$$(s_1, \ldots, s_p) \mapsto f_1(s_1) \cdots f_p(s_p),$$

tj. tensory rozkładalne odpowiadają „zmiennym rozseparowanym".

Jeśli $S_1 = \cdots = S_p = S$, to iloczyn tensorowy funkcji na S odpowiada zwyczajnemu iloczynowi ich wartości w „niezależnych punktach S".

Właśnie w takim kontekście iloczyny tensorowe pojawiają się najczęściej w analizie funkcjonalnej i fizyce. Jednakże ta algebraiczna definicja iloczynu tensorowego wymaga istotnych modyfikacji w analizie funkcjonalnej wobec konieczności uwzględnienia topologii rozważanych przestrzeni — w szczególności, często należy go uzupełniać w rozmaitych topologiach.

7. Rozszerzenie ciała skalarów. Niech L będzie przestrzenią liniową nad ciałem $\boldsymbol{R}$, $L^{\boldsymbol{C}}$ jej kompleksyfikacją (por. § 12, część 1). Ponieważ ciało $\boldsymbol{C}$ można uważać za przestrzeń liniową nad $\boldsymbol{R}$ (z bazą 1, i), więc można utworzyć przestrzeń liniową $\boldsymbol{C} \otimes L$ rozpiętą nad $\boldsymbol{R}$ przez bazę $1 \otimes e_1, \ldots, 1 \otimes e_n, i \otimes e_1, \ldots, i \otimes e_n$, gdzie $\{e_1, \ldots, e_n\}$ jest bazą L. Jest jasne, że odwzorowanie $\boldsymbol{R}$-liniowe

$$\boldsymbol{C} \otimes L \to L^{\boldsymbol{C}} : 1 \otimes e_j \mapsto e_j, \quad i \otimes e_j \mapsto ie_j$$

zadaje izomorfizm $\boldsymbol{C} \otimes L$ z $L^{\boldsymbol{C}}$.

Ogólniej, niech $\boldsymbol{K} \subset \mathcal{K}$ będzie podciałem ciała $\mathcal{K}$, L — przestrzenią liniową nad $\boldsymbol{K}$. Traktując początkowo $\mathcal{K}$ jako przestrzeń liniową nad $\boldsymbol{K}$ skontruujemy iloczyn tensorowy $\mathcal{K} \otimes L$, a następnie wprowadzimy na nim strukturę przestrzeni liniowej nad $\mathcal{K}$, określając mnożenie przez skalary $a \in \mathcal{K}$ wzorem

$$a(b \otimes l) = ab \otimes l, \quad a, b \in \mathcal{K}, \ l \in L.$$

Dla sprawdzenia poprawności tej definicji skonstruujemy tak jak w p. 2 przestrzeń $\mathcal{M}$, rozpiętą przez (swobodne) elementy $\mathcal{K} \times L$ i jej podprzestrzeń $\mathcal{M}_0$, tak aby $\mathcal{K} \otimes L = \mathcal{M}/\mathcal{M}_0$. Określimy w $\mathcal{M}$ mnożenie przez skalary należące do $\mathcal{K}$, przyjmując dla elementów bazowych

$$a(b, l) = (ab, l), \quad a, b \in \mathcal{K}, \; l \in L,$$

i rozszerzając tę regułę przez $\boldsymbol{K}$-liniowość na pozostałe elementy $\mathcal{K} \times L$. Bezpośrednie sprawdzenie pokazuje, że to działanie wprowadza na $\mathcal{M}$ strukturę $\mathcal{K}$-liniowej przestrzeni, a na $\mathcal{M}_0$ — jej podprzestrzeni, tak że $\mathcal{M}/\mathcal{M}_0 = \mathcal{K} \otimes L$ staje się przestrzenią liniową nad $\mathcal{K}$. To właśnie jest tą ogólną konstrukcją rozszerzenia ciała skalarów, która była wspomniana w § 11, p. 15 część 1.

Ważny przypadek szczególny: dla $\mathcal{K} = \boldsymbol{K}$ przestrzeń liniowa $\boldsymbol{K} \otimes L$ jest kanonicznie izomorficzna z L, a ten izomorfizm przeprowadza $a \otimes l$ na al.

§ 2. Izomorfizmy kanoniczne i odwzorowania liniowe iloczynów tensorowych

1. Mnożenie tensorowe ma niektóre algebraiczne własności operacji, które występując w innych kontekstach są nazywane mnożeniem, tak jak np. łączność. Jednakże w sformułowaniach dotyczących tych własności pojawiają się pewne cechy specyficzne wynikające z faktu, że mnożenie tensorowe jest operacją nad obiektami kategorii. Na przykład, przestrzenie $(L_1 \otimes L_2) \otimes L_3$ i $L_1 \otimes (L_2 \otimes L_3)$ nie są identyczne, co jest widoczne przy porównaniu ich konstrukcji, a są jedynie związane *kanonicznie zadanym izomorfizmem.*

W tym paragrafie opiszemy szereg takich „elementarnych" izomorfizmów, bardzo użytecznych przy rozważaniach iloczynów tensorowych. Jednakże uprzedzamy czytelnika, że jesteśmy zmuszeni do ograniczenia się jedynie do wstępu do teorii izomorfizmów kanonicznych. Głównym zagadnieniem spośród opuszczonych przez nas jest problem zgodności tych izomorfizmów. Na przykład, załóżmy, że są dane dwa naturalne izomorfizmy między pewnymi iloczynami tensorowymi, w różny sposób złożone z pewnych „elementarnych" naturalnych izomorfizmów. Czy takie izomorfizmy będą identyczne? Można przeprowadzać bezpośrednie sprawdzenie w każdym konkretnym przypadku albo próbować skonstruować ogólną teorię — ta ostatnia okazuje się jednak dosyć zawikłana.

Analogiczne problemy pojawiają się w związku z naturalnymi odwzorowaniami nie będącymi izomorfizmami, takimi jak symetryzacja czy zwężenie.

2. Łączność. Niech $L_1, \ldots, L_p$ będą przestrzeniami liniowymi nad $\boldsymbol{K}$. Pokażemy jak skonstruować izomorfizmy kanoniczne pomiędzy przestrzeniami postaci $(L_1 \otimes L_2)(\cdots \otimes L_p)$ otrzymywanymi w rezultacie mnożenia tensorowego przestrzeni $L_1, \ldots, L_p$ grupami w kolejności odpowiadającej różnym sposobom rozstawienia nawiasów. Najwygodniejszy sposób, automatycznie zapewniający zgodność, polega na

tym, aby dla każdego rozstawienia nawiasów skonstruować odwzorowanie liniowe $L_1 \otimes \cdots \otimes L_p \to (L_1 \otimes L_2)(\cdots \otimes L_p)$ za pomocą własności uniwersalności z § 1, p. 3 b) i pokazać, że jest ono izomorfizmem.

Przedstawimy szczegółowo przypadek $L_1 \otimes L_2 \otimes L_3 \to (L_1 \otimes L_2) \otimes L_3$; w sytuacji ogólnej rozumowanie jest w pełni analogiczne.

Odwzorowanie $L_1 \times L_2 \to L_1 \otimes L_2 : (l_1, l_2) \mapsto l_1 \otimes l_2$ jest dwuliniowe, skąd wynika, że $L_1 \times L_2 \times L_3 \to (L_1 \otimes L_2) \otimes L_3 : (l_1, l_2, l_3) \mapsto (l_1 \otimes l_2) \otimes l_3$ jest trójliniowe. Można je przeprowadzić przez jedyne odwzorowanie liniowe $L_1 \otimes L_2 \otimes L_3 \to (L_1 \otimes L_2) \otimes L_3$. Z samej konstrukcji wynika, że obrazem tensora $l_1 \otimes l_2 \otimes l_3$ jest tensor $(l_1 \otimes l_2) \otimes l_3$. Wybierając w przestrzeniach L_1, L_2, L_3 bazy i wykorzystując rezultat § 1, p. 5, widzimy, że to odwzorowanie przeprowadza bazę na bazę, a więc jest izomorfizmem.

Podsumowując: *Iloczyn tensorowy* $(L_1 \otimes L_2)(\cdots \otimes L_p)$ *z dowolnym rozstawieniem nawiasów można utożsamić z* $L_1 \otimes L_2 \otimes \cdots \otimes L_p$, *po prostu opuszczając nawiasy; obraz elementu* $(l_1 \otimes l_2)(\cdots \otimes l_p)$ *przy tym utożsamieniu znajdujemy zgodnie z tą samą zasadą.* Z tego względu będziemy pisać $(l_1 \otimes l_2) \otimes l_3 = l_1 \otimes l_2 \otimes l_3 = l_1 \otimes (l_2 \otimes l_3)$ itp.

3. Przemienność. Niech σ oznacza dowolną permutację liczb $1, \ldots, p$. Określimy rodzinę izomorfizmów

$$f_\sigma \colon L_1 \otimes \cdots \otimes L_p \longrightarrow L_{\sigma(1)} \otimes \cdots \otimes L_{\sigma(p)}$$

spełniających relację $f_{\sigma\tau} = f_\sigma \circ f_\tau$ dla każdej pary σ, τ. W tym celu zauważmy, że odwzorowanie

$$L_1 \times \cdots \times L_p \longrightarrow L_{\sigma(1)} \otimes \cdots \otimes L_{\sigma(p)} : (l_1, \ldots, l_p) \mapsto l_{\sigma(1)} \otimes \cdots \otimes l_{\sigma(p)}$$

jest wieloliniowe, a zatem przechodzi przez $f_\sigma \colon L_1 \otimes \cdots \otimes L_p \to L_{\sigma(1)} \otimes \cdots \otimes L_{\sigma(p)}$ na mocy twierdzenia § 1, p. 3 b). Na iloczyny wektorów działa ono w oczywisty sposób przestawiając czynniki, a rozważenie jego działania na iloczynie tensorowym baz $L_1, \ldots, L_p$ pokazuje, że jest izomorfizmem. Własność $f_{\sigma\tau} = f_\sigma \circ f_\tau$ jest oczywista.

W ten sposób określiliśmy działanie grupy symetrycznej S_p na $L_1 \otimes \cdots \otimes L_p$. W przypadku gdy wszystkie L_i są różne, można posłużyć się izomorfizmami f_σ dla jednoznacznego ustalenia identyfikacji $L_1 \otimes \cdots \otimes L_p$ z $L_{\sigma(1)} \otimes \cdots \otimes L_{\sigma(p)}$. Właśnie w takim znaczeniu mnożenie tensorowe jest przemienne. Jednakże zapisywanie tego utożsamienia w postaci równości, bez jawnego odwołania do f_σ (tak jak to zrobiliśmy w przypadku łączności), może być mylące, jeśli niektóre z przestrzeni $L_1, \ldots, L_p$ są równe.

4. Dwoistość. Istnieje izomorfizm kanoniczny

$$L^* \otimes \cdots \otimes L^* \longrightarrow (L_1 \otimes \cdots \otimes L_p)^*.$$

Aby go skonstruować, przyporządkujemy każdemu elementowi $(f_1, \ldots, f_p) \in L^* \times \cdots \times L^*$ funkcję wieloliniową $f_1(l_1) \cdots f_p(l_p)$ argumentu $(l_1, \ldots, l_p) \in L_1 \times \cdots \times L_p$.

Na mocy twierdzenia § 1, p. 3 b), to odwzorowanie przechodzi przez odwzorowanie $L^* \otimes \cdots \otimes L^* \to$ (*przestrzeń funkcji wieloliniowych na* $L_1 \times \cdots \times L_p$). Jeśli poprzez konstrukcję z § 1, p. 4 utożsamić tę ostatnią przestrzeń z przestrzenią $\mathcal{L}(L_1 \otimes \cdots \otimes L_p, \boldsymbol{K}) = (L_1 \otimes \cdots \otimes L_p)^*$, to otrzymamy poszukiwane odwzorowanie. Dla wykazania, że jest to izomorfizm, zauważymy, iż wymiary przestrzeni $(L_1 \otimes \cdots \otimes L_p)^*$ i $L^* \otimes \cdots \otimes L^*$ są jednakowe (tu ograniczamy się do przypadku skończenie wymiarowych L_i) i dlatego wystarczy sprawdzić, że nasze odwzorowanie jest surjektywne. Istotnie, jego obraz zawiera funkcje $f_1(l_1) \cdots f_p(l_p)$, gdzie f_i przebiegają pewną bazę L_i, i tak jak w § 1, p. 5 nietrudno przekonać się, że te funkcje tworzą bazę $\mathcal{L}(L_1 \otimes \cdots \otimes L_p, \boldsymbol{K}) = (L_1 \otimes \cdots \otimes L_p)^*$.

Identyfikacja $(L_1 \otimes \cdots \otimes L_p)^*$ z $L^* \otimes \cdots \otimes L^*$ za pomocą opisanego izomorfizmu zazwyczaj nie stwarza problemów.

5. Izomorfizm $\mathcal{L}(L, M)$ z $L^* \otimes M$. Rozważmy odwzorowanie dwuliniowe

$$L \times M \longrightarrow \mathcal{L}(L, M) : (f, m) \mapsto [l \mapsto f(l)m],$$

gdzie $f \in L^*$, $l \in L$, $m \in M$. Zarówno dwuliniowość wyrażenia $f(l)m$ względem f, m, jak i jego liniowość względem l są oczywiste. Na mocy wielokrotnie przywoływanej własności uniwersalności odpowiada mu odwzorowanie liniowe

$$L \otimes M \longrightarrow \mathcal{L}(L, M).$$

Wybierzmy w L i M bazy $\{l_1, \ldots, l_a\}$, $\{m_1, \ldots, m_b\}$ i w L^* bazę dwoistą $\{l^1, \ldots, l^a\}$. Element $l^i \otimes m_j$ iloczynu tensorowego baz L^*, M przechodzi na na odwzorowanie liniowe, które przeprowadza wektor $l_k \in L$ na $l^i(l_k)m_j = \delta_{ik}m_j$. Elementami macierzy o wymiarach $b \times a$ odpowiadającej temu odwzorowaniu są jedynka (na miejscu (i, j)) i zera na pozostałych miejscach. Ponieważ takie macierze tworzą bazę w przestrzeni macierzy $b \times a$, skonstruowane odwzorowanie przeprowadza bazę na bazę, a więc jest izomorfizmem.

Rozważmy szczególny, lecz ważny przypadek $L = M$, w którym

$$\mathcal{L}(L, L) = L^* \otimes L.$$

Przestrzeń endomorfizmów $\mathcal{L}(L, L)$ zawiera wyróżniony element — odwzorowanie tożsamościowe id_L. Jego obraz jest, jak to widać z poprzedzających rozważań, równy

$$\sum_{k=1}^{a} l^k \otimes l_k,$$

gdzie $\{l_k\}$, $\{l^k\}$ jest parą baz dwoistych w L i L^*. Stąd wniosek, że wzór wyrażający ten element ma jednakową postać dla wszystkich par baz dwoistych.

Ponadto, przestrzeń endomorfizmów $\mathcal{L}(L, L)$ ma kanonicznie wyznaczony funkcjonał liniowy — ślad Tr: $\mathcal{L}(L, L) \to \boldsymbol{K}$. Z poprzednich rozważań wynika, że ślad odwzorowania odpowiadającego elementowi $l^i \otimes l_j$ jest równy δ_{ij} (przyjrzyjmy się

macierzy!), tak że dowolnemu elementowi iloczynu tensorowego $L^* \otimes L$ przyporządkowana jest liczba

$$\sum_{i,j=1}^{a} a_i^j l^i \otimes l_j \mapsto \sum_{i=1}^{a} a_i^i.$$

Ten funkcjonał liniowy $L^* \otimes L \to \boldsymbol{K}$ jest nazywany *zwężeniem* (*kontrakcją*). Poniżej podamy definicję zwężenia w ogólniejszym kontekście.

6. Izomorfizm $\mathcal{L}(L \otimes M, N)$ z $\mathcal{L}(L, \mathcal{L}(M, N))$. Przestrzeń $\mathcal{L}(L \otimes M, N)$ jest izomorficzna z przestrzenią odwzorowań dwuliniowych $L \times M \to N$. Każde odwzorowanie dwuliniowe $f\colon (l, m) \to f(l, m)$ wyznacza, przez ustalenie wektora l jako pierwszego argumentu, odwzorowanie liniowe $M \to N$, przy czym to odwzorowanie zależy liniowo od pierwszego argumentu l. W ten sposób otrzymujemy kanonicznie określone odwzorowanie liniowe

$$\mathcal{L}(L \otimes M, N) = \mathcal{L}(L, M; N) \longrightarrow \mathcal{L}(L, \mathcal{L}(M, N)).$$

Argument z użyciem baz w L, M, N, analogiczny do podanego w poprzednim punkcie, dowodzi, że jest to izomorfizm (jak wszędzie zakładamy skończoność wymiaru przestrzeni).

(Ta identyfikacja jest ważnym przykładem dla rozważanego w ogólnej teorii kategorii „funktora sprzężonego w sensie Kahna".)

7. Iloczyny tensorowe odwzorowań liniowych. Niech $L_1, \ldots, L_p$, $M_1, \ldots, M_p$ będą dwoma układami przestrzeni liniowych, $f_i\colon L_i \to M_i$ — odwzorowaniami liniowymi. Wówczas można określić odwzorowanie liniowe

$$f_1 \otimes \cdots \otimes f_p\colon L_1 \otimes \cdots \otimes L_p \longrightarrow M_1 \otimes \cdots \otimes M_p,$$

nazywane iloczynem tensorowym f_i i jednoznacznie wyznaczone przez następującą prostą własność:

$$(f_1 \otimes \cdots \otimes f_p)(l_1 \otimes \cdots \otimes l_p) = f_1(l_1) \otimes \cdots \otimes f_p(l_p)$$

dla wszystkich $l_i \in L_i$. Istnienie tego odwzorowania wynika przez takie jak poprzednio standardowe zastosowanie twierdzenia p. 3 b) z § 1, jeśli zauważyć, że odwzorowanie

$$L_1 \times \cdots \times L_p \longrightarrow M_1 \otimes \cdots \otimes M_p\colon (l_1, \ldots, l_p) \mapsto f_1(l_1) \otimes \cdots \otimes f_p(l_p)$$

jest wieloliniowe. Jeśli wszystkie f_i są izomorfizmami, to także $f_1 \otimes \cdots \otimes f_p$ jest izomorfizmem.

8. Zwężenie i podniesienie indeksów. Za pomocą tej konstrukcji możemy podać ogólną definicję zwężenia „względem pary lub kilku par wskaźników". Rozważmy iloczyn tensorowy $L_1 \otimes \cdots \otimes L_p$ i załóżmy, że dla pewnej pary wskaźników

$i, j \in \{1, \cdots p\}$ zachodzi $L_i = L^*$, $L_j = L$. *Zwężeniem względem pary wskaźników* i, j nazywamy odwzorowanie liniowe

$$L_1 \otimes \cdots \otimes L_p \longrightarrow \bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k,$$

zdefiniowane jako złożenie następujących odwzorowań liniowych:

a) Odwzorowania f_σ, gdzie σ jest taką permutacją wskaźników $\{1, \cdots p\}$, że $\sigma(1) = i$, $\sigma(2) = j$, a porządek pozostałych wskaźników się nie zmienia:

$$f_\sigma \colon L_1 \otimes \cdots \otimes L_p \longrightarrow L_i \otimes L_j \otimes \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big) = L^* \otimes L \otimes \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big).$$

b) Zwężenia względem dwóch pierwszych wskaźników pomnożonego tensorowo przez odwzorowanie tożsamościowe pozostałych argumentów:

$$L^* \otimes L \otimes \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big) \colon \longrightarrow \boldsymbol{K} \otimes \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big).$$

c) Utożsamienia

$$\boldsymbol{K} \otimes \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big) \longrightarrow \Big(\bigotimes_{\substack{k=1 \\ k \neq i,j}}^{p} L_k \Big).$$

Jeśli występuje więcej takich par wskaźników $(i_1, j_1), \ldots, (i_r, j_r)$, dla których $L_{i_k} = M^*$, $L_{j_k} = M_k$, to tę konstrukcję można zastosować kolejno do każdej z tych par. Otrzymane w ten sposób odwzorowanie liniowe nazywa się zwężeniem względem zadanych par wskaźników. Zależy ono tylko od tych par, a nie od kolejności, w jakiej przeprowadzane są zwężenia względem poszczególnych par. W przypadku gdy $\{1, \ldots, p\} = \{i_1, j_1, \ldots, i_r, j_r\}$, otrzymujemy *pełne* zwężenie.

Rozważmy jeszcze raz iloczyn tensorowy $L_1 \otimes \cdots \otimes L_p$ i załóżmy, że dla i-tej przestrzeni jest zadany izomorfizm $g \colon L_i \to L^*$ (w zastosowaniach jest on najczęściej zadany za pomocą niezdegenerowanej formy dwuliniowej symetrycznej na L_i). Wówczas odwzorowanie liniowe

$$\mathrm{id} \otimes \cdots \otimes g \otimes \cdots \otimes \mathrm{id} \colon L_1 \otimes \cdots \otimes L_i \otimes \cdots \otimes L_p \longrightarrow L_1 \otimes \cdots \otimes L^* \otimes \cdots \otimes L_p$$

nazywa się „opuszczeniem i-tego wskaźnika", a odwrotne do niego — „podnoszeniem i-tego wskaźnika". Pochodzenie tej nazwy zostanie objaśnione w następnym paragrafie.

Obie konstrukcje, zwężenie i podniesienie/opuszczanie wskaźnika, najczęściej stosowane są w przypadku $L_i = L$ albo L^*, gdy na L jest wyróżniona struktura ortogonalna. Pojawia się wtedy cała masa odwzorowań liniowych wiążących między sobą przestrzenie $L^{*\otimes p} \otimes L^{\otimes q}$, otrzymywanych jako złożenia podniesień i opuszczań wskaźników i zwężeń. Odgrywają one bardzo istotną rolę w geometrii riemanowskiej,

gdzie są wykorzystywane (w połączeniu z operacjami analitycznymi typu różniczkowania) do konstrukcji najważniejszych niezmienników geometryczno-różniczkowych.

9. Mnożenie tensorowe jako funktor dokładny. Ustalmy przestrzeń liniową M i rozważmy odwzorowanie kategorii skończenie wymiarowych przestrzeni liniowych w siebie zadane na obiektach jako $L \mapsto L \otimes M$, a jako $f \mapsto f \otimes \mathrm{id}_M$ na morfizmach. Posługując się samą definicją, łatwo sprawdzić, że $\mathrm{id}_L \mapsto \mathrm{id}_{L \otimes M}$ oraz

$$f \circ g \mapsto f \circ g \otimes \mathrm{id}_M = (f \otimes \mathrm{id}_M) \circ (g \otimes \mathrm{id}_M).$$

Wynika stąd, że to odwzorowanie jest *funktorem*, który nazywamy funktorem mnożenia tensorowego przez M.

Dowiedziemy, że jeśli ciąg $0 \longrightarrow L_1 \longrightarrow L \longrightarrow L_2 \longrightarrow 0$ jest dokładny, to także ciąg

$$0 \longrightarrow L_1 \otimes M \xrightarrow{f \otimes \mathrm{id}} L \otimes M \xrightarrow{g \otimes \mathrm{id}} L_2 \otimes M \longrightarrow 0$$

jest dokładny. Ta własność jest nazywana *dokładnością funktora mnożenia tensorowego*. Nie przysługuje ona kategorii modułów, podobnie jak dokładność funktora $\mathcal{L}$, a ten niedostatek jest ważnym obiektem badań w algebrze homologicznej — por. dyskusję w § 14, część 1.

Najprostszym sposobem wykazania tej dokładności jest wprowadzenie baz w L_1, L, L_2, dostosowanych do f i g w ten sposób, że $\{e_1, \ldots, e_a\}$ jest bazą L_1, $\{f(e_1), \ldots, f(e_a); e'_{a+1}, \ldots, e'_{a+b}\}$ jest bazą L, a $\{g(e'_{a+1}), \ldots, g(e'_{a+b})\}$ jest bazą L_2. Wybrawszy jeszcze bazę $\{e''_1, \ldots, e''_c\}$ przestrzeni M widzimy, że iloczyny tensorowe baz

$$\{e_i \otimes e''_j\}, \quad \{f(e_i) \otimes e''_j, e'_k \otimes e''_j\}, \quad \{g(e'_k) \otimes e''_j\}$$

są dostosowane do $f \otimes \mathrm{id}_M$, $g \otimes \mathrm{id}_M$ w powyższym znaczeniu.

§ 3. Algebra tensorowa przestrzeni liniowej

1. Niech L będzie skończenie wymiarową przestrzenią liniową nad ciałem $\boldsymbol{K}$. Elementy iloczynu tensorowego

$$\mathcal{T}_p^q(L) = \underbrace{L^* \otimes \cdots \otimes L^*}_{p} \otimes \underbrace{L \otimes \cdots \otimes L}_{q}$$

nazywa się *tensorem nad L typu (p, q) wagi* (lub *rzędu*) $p+q$. Mówi się także, że jest to tensor mieszany, p razy *kowariantny* i q razy *kontrawariantny*. Dwie pierwsze części tej książki były w istocie poświęcone badaniu opisanych niżej tensorów niskiego rzędu.

a) Wygodnie przyjąć $\mathcal{T}_0^0(L) = \boldsymbol{K}$, tj. nazywać skalary tensorami rzędu zerowego.

b) $\mathcal{T}_1^0(L) = L^*$, tj. tensorami typu $(1, 0)$ są funkcjonały liniowe na L. Tensorami typu $(0, 1)$ są same wektory przestrzeni L.

c) $\mathcal{T}_1^1(L) = L^* \otimes L$. W paragrafie 2, p. 5 utożsamiliśmy $L^* \otimes L$ z przestrzenią $\mathcal{L}(l, L)$. A zatem tensory typu $(1,1)$ „są" operatorami liniowymi w L.

d) $\mathcal{T}_2^0(L) = L^* \otimes L^*$. W paragrafie 2, p. 4 utożsamiliśmy $L^* \otimes L^*$ z $(L \otimes L)^*$ albo z przestrzenią odwzorowań dwuliniowych $L \times L \to \boldsymbol{K}$. W ten sposób można uważać, że tensory typu $(0,2)$ „są" iloczynami skalarnymi na L. Natomiast w § 2, p. 5 utożsamiliśmy $L^* \otimes L^*$ z $\mathcal{L}(L^{**}, L^*) \cong \mathcal{L}(L, L^*)$. Przy takim utożsamieniu iloczynowi skalarnemu na L przyporządkowujemy odwzorowanie liniowe $L \to L^*$ odpowiadające traktowaniu tego iloczynu skalarnego jako funkcji jednego z argumentów przy ustalonym drugim. W takim razie konstrukcje tensorowe rozpatrywane w § 2 są uogólnieniem konstrukcji z części 2.

e) Podamy jeszcze jeden przykład — tensor strukturalny $\boldsymbol{K}$-algebry. Tutaj jako $\boldsymbol{K}$-algebrę przyjmiemy przestrzeń liniową L z dwuliniową operacją mnożenia $L \times L \to L : (l, m) \mapsto lm$, niekoniecznie przemienną ani nawet łączną, tak że na przykład algebry Liego spełniają te warunki.

Zgodnie z twierdzeniem § 1, p. 3 b) mnożenie można określić także jako odwzorowanie liniowe $L \otimes L \to L$. W paragrafie 2, p. 5 dokonaliśmy utożsamienia przestrzeni $\mathcal{L}(L \otimes L, L)$ z $(L \otimes L)^* \otimes L$ albo, wykorzystując jeszcze raz § 2, p. 4 oraz łączność, z $L^* \otimes L^* \otimes L$. Tak więc wprowadzić na przestrzeni L strukturę $\boldsymbol{K}$-algebry jest tym samym co zadać tensor typu $(2,1)$ nad tą przestrzenią. Nazywa się on *tensorem strukturalnym* tej algebry.

2. Mnożenie tensorowe. Zgodnie z § 2, p. 4 możemy utożsamić przestrzeń $\mathcal{T}_p^q(L)$ z $(L^{\otimes p} \otimes (L^*)^{\otimes q})^*$, a następnie z przestrzenią odwzorowań wieloliniowych

$$f\colon \underbrace{L \times \cdots \times L}_{p} \times \underbrace{L^* \times \cdots \times L^*}_{q} \longrightarrow \boldsymbol{K}.$$

Dwa takie wieloliniowe odwzorowania typu (p, q) i (p', q') można pomnożyć tensorowo, w wyniku czego dostajemy wieloliniowe odwzorowanie typu $(p + p', q + q')$:

$$(f \otimes g)(l_1, \ldots, l_p; l'_1, \ldots, l'_{p'}; l_1^*, \ldots, l_q^*; l_1'^*, \ldots, l_{q'}'^*) =$$
$$= f(l_1, \ldots, l_p; l_1^*, \ldots, l_q^*)\, g(l'_1, \ldots, l'_{p'}; l_1'^*, \ldots, l_{q'}'^*),$$

gdzie $l_i, l'_j \in L$, $l^*, l'^* \in L^*$. Z tej definicji bezpośrednio wynika dwuliniowość mnożenia tensorowego względem jego argumentów:

$$(af_1 + bf_2) \otimes g = a(f_1 \otimes g) + b(f_2 \otimes g);$$
$$f \otimes (ag_1 + bg_2) = a(f \otimes g_1) + b(f \otimes g_2),$$

jak również łączność:

$$(f \otimes g) \otimes h = f \otimes (g \otimes h).$$

Jednakże nie jest ono przemienne, gdyż w ogólności $f \otimes g$ nie równa się $g \otimes f$.

Jeśli nie korzystać z interpretacji tensorów jako odwzorowań wieloliniowych, to mnożenie tensorowe można zdefiniować za pomocą odwzorowań permutacji z § 2, p. 3, z uwzględnieniem łączności, jako odwzorowanie

$$f_\sigma: \underbrace{L^*\otimes\cdots\otimes L^*}_{p}\otimes\underbrace{L\otimes\cdots\otimes L}_{q}\otimes\underbrace{L^*\otimes\cdots\otimes L^*}_{p'}\otimes\underbrace{L\otimes\cdots\otimes L}_{q'}\longrightarrow$$

$$\longrightarrow \underbrace{L^*\otimes\cdots\otimes L^*}_{p+p'}\otimes\underbrace{L\otimes\cdots\otimes L}_{q+q'},$$

gdzie σ przestawia trzecią grupę złożoną z p' wskaźników na miejsca bezpośrednio po pierwszej grupie złożonej z p wskaźników, zachowując ich względną kolejność i względną kolejność pozostałych wskaźników. W tym kontekście dwuliniowość mnożenia jest podobnie oczywista, a jego łączność jest równoważna z pewną tożsamością dla permutacji, którą czytelnikowi będzie łatwiej zauważyć samemu niż śledzić długie, choć banalne objaśnienia.

3. Algebra tensorowa przestrzeni L. Określmy

$$\mathcal{T}(L)=\bigoplus_{p,q=1}^{\infty}\mathcal{T}_p^q(L)$$

(suma prosta przestrzeni liniowych). Jest to przestrzeń nieskończenie wymiarowa, która wraz z określoną w poprzednim punkcie operacją mnożenia tensorowego jest nazywana *algebrą tensorową przestrzeni L*.

Zauważmy, że dla przypadku przestrzeni nad ciałem liczb zespolonych jest nieraz potrzebne rozważanie rozszerzonej algebry tensorowej, będącej sumą prostą przestrzeni $L^{\otimes p}\otimes L^{*\otimes q}\otimes\overline{L}^{\otimes p'}\otimes\overline{L}^{*\otimes q'}$. Na przykład, forma półtoraliniowa traktowana jako tensor należy do $L^*\otimes\overline{L}^*$. Ze względu na brak miejsca nie będziemy szczegółowo przedstawiać tej konstrukcji.

§ 4. Notacja klasyczna

1. W klasycznej analizie tensorowej formalizm tensorowy jest stosowany w oznaczeniach jawnie wykorzystujących współrzędne. Do dziś taką jego postać wykorzystuje się w geometrycznej i fizycznej literaturze, oceniając ją sprawiedliwie, należy uznać jej giętkość i zwartość. Tę postać wprowadzimy w obecnym paragrafie i pokażemy, jak wyrażają się w niej rozliczne konstrukcje opisane poprzednio.

2. Bazy i współrzędne. Niech L będzie skończenie wymiarową przestrzenią liniową. Ustalimy bazę $\{e_1,\ldots,e_n\}$ w tej przestrzeni i będziemy zadawać wektory z L poprzez współrzędne $(a^1,\ldots,a^n)$ względem tej bazy: $\sum_{i=1}^{n}a^ie_i$.

W L^* wybierzemy bazę dualną $\{e^1,\ldots,e^n\}$, $(e^i,e_j)=\delta^i_j=0$ dla $i\neq j$; 1 dla $i=j$, a wektory z L^* będziemy wyrażać poprzez współrzędne $(b_1,\ldots,b_n)$, $\sum_{i=1}^{n}b_ie^i$.

Położenie wskaźników w obydwu przypadkach wybiera się w taki sposób, aby sumowanie rozciągało się na pary jednakowych wskaźników, z których jeden zapisuje się w góry, a drugi u dołu.

W $L^{*\otimes p}\otimes L^{\otimes q}$ wprowadzimy iloczyn tensorowy rozważanych baz

$$\left\{e^{i_1}\otimes\cdots\otimes e^{i_p}\otimes e_{j_1}\otimes\cdots\otimes e_{j_q}\;\middle|\;1\leq i_k\leq n,\ 1\leq j_l\leq n\right\}.$$

Dowolny tensor $T\in\mathcal{T}_p^q(L)$ jest wyznaczony przez swoje współrzędne $T_{i_1\ldots i_p}^{j_1\ldots j_q}$ w tej bazie:

$$T=\sum T_{i_1\ldots i_p}^{j_1\ldots j_q}e^{i_1}\otimes\cdots\otimes e^{i_p}\otimes e_{j_1}\otimes\cdots\otimes ej_q.$$

Zauważymy, że sumowanie znowu prowadzi się względem par jednakowych wskaźników, z których jeden jest zapisany u góry, a drugi u dołu. Jest to tak charakterystyczna cecha klasycznego formalizmu, że często przyjmuje się umowę ([1]) o opuszczaniu znaku sumy we wszystkich takich przypadkach, gdzie występuje tego rodzaju sumowanie.

W szczególności, przy tej umowie wektory są zapisywane w postaci a^je_j, a funkcjonały liniowe w postaci b_ie^i. Iloczyn skalarny (dwoistość) między L^* a L, tj. wartość funkcjonału b_ie^i na wektorze a^je_j jest zapisywany wówczas jako a^ib_i lub b_ia^i.

Można posunąć się jeszcze dalej na tej drodze oszczędności i ominąć w zapisie nawet wektory e_i czy e^i. Wówczas elementy L oznacza się po prostu przez a^i, elementy L^* przez b_i, a ogólny tensor $T\in\mathcal{T}_p^q(L)$ zapisuje się w postaci $T_{i_1\ldots i_p}^{j_1\ldots j_q}$. Innymi słowy, w klasycznym zapisie tensora T

są jawnie uwzględnione: współrzędne, inaczej *składowe* T względem bazy tensorowej $L^{*\otimes p}\otimes L^{\otimes q}$, indeksowane w ten sam sposób jak elementy bazy tensorowej przez grupy wskaźników (wielowskaźniki) — część kowariantna wielowskaźnika $(i_1,\ldots,i_p)$ jest zapisywana u dołu, a część kontrawariantna $(j_1,\ldots,j_q)$ u góry;

nie są uwzględnione: wybór bazy $\{e_1,\ldots,e_n\}$ w L, wraz z dobraną do niej bazą dwoistą $\{e^1,\ldots,e^n\}$ w L^* oraz bazami tensorowymi we wszystkich przestrzeniach $\mathcal{T}_p^q(L)$.

Czasami wygodnie rozpatrywać tensory należące do przestrzeni, w których czynniki L i L^* są rozłożone w innym porządku niż ten przyjęty przez nas, np. $L\otimes L^*$ zamiast $L^*\otimes L$ lub $L\otimes L^*\otimes L\otimes L$. Ten fakt jest wskazywany poprzez zapis używający „blokowego rozkładu" wielowskaźników przy składowych tensora. Na przykład, składowe tensora $T\in L\otimes L^*$ są oznaczane T_i^j, a $T\in L\otimes L^*\otimes L\otimes L$ — przez $T_{i_1}{}^{j_1}{}_{i_2i_3}$.

3. Pewne ważne tensory. Do nich należą:

a) *Tensor metryczny* g_{ij}. Zgodnie z naszymi oznaczeniami należy on do $\mathcal{T}_2^0(L)$, a na mocy § 3, p. 1 d) odpowiada on iloczynowi skalarnemu na L. Wartość tego

([1]) nazywaną często konwencją Einsteina (*przyp. tłum.*).

iloczynu dla pary wektorów a^i, b^j jest równa $\sum g_{ij}a^ib^j$ lub prościej, $g_{ij}a^ib^j$. Stąd wynika, że składowymi tensora metrycznego są wyrazy macierzy Grama wyjściowej bazy przestrzeni L względem odpowiadającego mu iloczynu skalarnego.

b) *Macierz* A^i_j. Jest to element $T^1_1(L)$, tj. na mocy § 3, p. 1 c) odwzorowanie liniowe przestrzeni L w siebie. Przeprowadza ono wektor a^j na wektor, którego i-tą współrzędną jest $\sum A^i_j a^j$ lub prościej, $A^i_j a^j$. Tensor o walencji $p+q$ można interpretować jako „$p+q$-wymiarową macierz", a zwykłe macierze są płaskie.

c) *Tensor Kroneckera* δ^i_j. Jest to element $T^1_1(L)$ reprezentujący odwzorowanie tożsamościowe przestrzeni L.

d) *Tensor strukturalny algebry.* Zgodnie z § 3, p. 1 e) należy on do $T^1_2(L)$ i z tego względu w notacji współrzędnościowej zapisuje się go jako γ^k_{ij}. Wyznacza on dwuliniowe mnożenie w L zadane wzorem

$$a^i \cdot b^j = c^k = \gamma^k_{ij} a^i b^j,$$

którego pełny zapis ma postać

$$\Big(\sum_i a^i e_i\Big)\Big(\sum_j b^j e_j\Big) = \sum_k \Big(\sum_{i,j} \gamma^k_{ij} a^i b^j\Big) e_k.$$

4. Zmiana współrzędnych tensora przy zamianie bazy w L. Niech A^i_j będzie macierzą zamiany bazy w L: $e'_k = A^i_k e_i$, a B^i — macierzą przejścia od bazy $\{e^k\}$, dwoistej do bazy $\{e_k\}$, do bazy $\{e'^k\}$, dwoistej do $\{e'_k\}$. Łatwo sprawdzić, że $B = (A^t)^{-1}$ — tę macierz nazywamy *kontragredientą* do A.

Współrzędne a'^j względem bazy $\{e'_j\}$ wektora, zadanego w bazie $\{e_i\}$ współrzędnymi a^i, są równe $B^j_k a^k$.

Analogicznie, współrzędne b'_i względem bazy $\{e'^i\}$ funkcjonału (czyli „kowektora") o współrzędnych b_i w początkowej bazie $\{e^i\}$ są równe $A^k_j b_k$.

Dla wyznaczenia składowych $T'^{j_1 \dots j_q}_{i_1 \dots i_p}$ względem „primowanej" bazy tensorowej tensora o składowych $T^{j_1 \dots j_q}_{i_1 \dots i_p}$ w wyjściowej bazie, wystarczy zauważyć, że przekształcają się one tak jak współrzędne iloczynu tensorowego q wektorów i p kowektorów, tj.

$$T'^{j_1 \dots j_q}_{i_1 \dots i_p} = A^{l_1}_{i_1} \cdots A^{l_p}_{i_p} B^{j_1}_{k_1} \cdots B^{j_q}_{k_q} \cdot T^{k_1 \dots k_q}_{l_1 \dots l_p}.$$

Nie należy zapominać, że zakłada się przeprowadzenie sumowania po wszystkich parach jednakowych wskaźników po prawej stronie wzoru.

W tradycyjnym wykładzie tę formułę przyjmuje się jako podstawę *definicji* tensora.

A mianowicie, *tensorem typu* (p,q) *na* n-wymiarowej *przestrzeni* nazywa się odwzorowanie T, przyporządkowujące każdej bazie L układ n^{p+q} składowych — skalarów $T^{j_1 \dots j_q}_{i_1 \dots i_p}$, przy tym takie, że przy zamianie bazy zadanej macierzą A składowe tensora zmieniają się zgodnie z podanym powyżej wzorem.

5. Konstrukcje tensorowe we współrzędnych. a) *Kombinacje liniowe tensorów jednakowego typu.* Tu wzory są oczywiste:

$$(aT + bT')_{i_1 \dots i_p}^{j_1 \dots j_q} = aT_{i_1 \dots i_p}^{j_1 \dots j_q} + bT'^{j_1 \dots j_q}_{i_1 \dots i_p}.$$

b) *Iloczyn tensorowy.* Na podstawie definicji podanej w § 3, p. 2 mamy

$$(T \otimes T')_{i_1 \dots i_p i'_1 \dots i_{p'}}^{j_1 \dots j_q j'_1 \dots j'_{q'}} = T_{i_1 \dots i_p}^{j_1 \dots j_q} T'^{j'_1 \dots j'_{q'}}_{i'_1 \dots i_{p'}}.$$

W szczególności, składowymi tensora rozkładalnego są $T_{i_1} \cdots T_{i_p} \cdot T^{j_1} \dots T^{j_q}$.

c) *Permutacje.* Niech σ oznacza permutację wskaźników $1, \dots, p$, τ — permutację wskaźników $1, \dots, q$; $f_{\sigma,\tau}$: $\mathcal{T}_p^q(L) \to \mathcal{T}_p^q(L)$ — odwzorowanie liniowe odpowiadające tym permutacjom, opisane w § 2, p. 3. Wówczas dla każdego $T \in \mathcal{T}_p^q(L)$ mamy

$$[f_{\sigma,\tau}(T)]_{i_1 \dots i_p}^{j_1 \dots j_q} = T_{i_{\sigma^{-1}(1)} \dots i_{\sigma^{-1}(p)}}^{j_{\tau^{-1}(1)} \dots j_{\tau^{-1}(q)}}.$$

d) *Zwężenie.* Niech $a \in \{1, \dots, p\}$, $b \in \{1, \dots, q\}$. Jak to pokazaliśmy w § 2, p. 8 istnieje odwzorowanie $\mathcal{T}_p^q(L) \to \mathcal{T}_{p-1}^{q-1}(L)$ „usuwające" a-ty czynnik L^* i b-ty czynnik L za pomocą odwzorowania zwężenia $L^* \otimes L \to \boldsymbol{K}$, które jest niczym innym jak tylko zwyczajnym iloczynem skalarnym wektorów i funkcjonałów liniowych: $(b_i) \otimes (a^j) \mapsto b_i a^i$. A zatem, oznaczając przez T' tensor T zwężony względem pary wskaźników (dolny a-ty, górny b-ty), otrzymujemy

$$T'^{j_1 \dots j_{b-1} j_{b+1} \dots j_q}_{i_1 \dots i_{a-1} i_{a+1} \dots i_p} = T^{j_1 \dots j_{b-1} k j_{b+1} \dots j_q}_{i_1 \dots i_{a-1} k i_{a+1} \dots i_p}$$

(po prawej sumowanie względem k). Iterując tę konstrukcję, otrzymamy odwzorowanie zwężenia względem kilku par wskaźników.

Przekonaliśmy się więc, że liczne wzory algebry tensorowej zapisują się w terminach iloczynu tensorowego i następującego po nim zwężenia względem jednej lub kilku par wskaźników. Powtórzmy je dla utrwalenia:

$g_{ij} a^i b^j$ — zwężenie $((g_{ij}) \otimes (a^k) \otimes (b^l))$.

Iloczyn skalarny:

$b_i a^i$ — zwężenie $((b_i) \otimes (a^j))$.

Składowe tensora w nowej bazie lub, używając „aktywnego punktu widzenia", obraz tensora przy odwzorowaniu liniowym przestrzeni podstawowej:

$T'^{j_1 \dots j_q}_{i_1 \dots i_p}$ — zwężenie $((A \otimes \cdots \otimes A) \otimes (B \otimes \cdots \otimes B) \otimes T)$.

Mnożenie w algebrze:

$a^i \cdot b^j$ — zwężenie $((\text{tensor strukturalny}) \otimes (a^i) \otimes (b^j))$.

Jeszcze jeden przykład — iloczyn macierzy;

$$(A^i_j)(B^j_k) = (A^i_j B^j_k) \quad \text{— zwężenie } (A^i_j \otimes B^j_k).$$

Podkreślmy raz jeszcze, że dla pełnego określenia zwężenia należy wskazać, względem jakiej pary wskaźników jest ono przeprowadzone — w zestawionych wyżej przykładach jest to albo oczywiste, albo wynika z podanych wcześniej pełnych wzorów.

Można stwierdzić ogólnie, że w klasycznym sformułowaniu algebry tensorowej operacja zwężenia odgrywa taką samą unifikującą rolę, jak operacja mnożenia macierzy w sformułowaniu algebry liniowej. W paragrafie 4 części 1 podkreślaliśmy, że teoriomnogościowe operacje o bardzo zróżnicowanym charakterze można jednolicie opisać przy użyciu iloczynu macierzowego. Ta obserwacja odnosi się w jeszcze większym stopniu do algebry tensorowej i zwężenia połączonego z iloczynem tensorowym.

e) *Podnoszenie i opuszczanie wskaźników.* Zgodnie z definicją podaną w § 2, p. 8 c) podniesienie a-tego i opuszczenie b-tego wskaźnika są odwzorowaniami liniowymi

$$T^q_p(L) \longrightarrow T^{q+1}_{p-1}(L), \quad T^q_p(L) \longrightarrow T^{q-1}_{p+1}(L),$$

indukowanymi przez izomorfizmy $g: L^* \to L$ lub $g^{-1}: L \to L^*$: „zamieniają" one w iloczynie $L^{*\otimes p} \otimes L^{\otimes q}$ odpowiednio a-ty czynnik L^* na L lub b-ty czynnik L na L^*.

Zgodnie z konwencjami przyjętymi na końcu p. 2 tego paragrafu składowe otrzymanych w ten sposób tensorów należałoby zapisać w postaci

$$T_{i_1 \dots i_{a-1}}{}^{i_a}{}_{i_{a+1} \dots i_p}{}^{j_1 \dots j_q} \quad \text{oraz} \quad T_{i_1 \dots i_p}{}^{j_1 \dots j_{b-1}}{}_{j_b}{}^{j_{b+1} \dots j_q}.$$

Jeśli jednakże umówić się, że po zastosowaniu odwzorowania podniesienia (opuszczenia) wskaźnika stosuje się także odwzorowanie permutacji przenoszące nowo wprowadzony czynnik L na prawo, a L^* — na lewo, dopóki nie będzie sąsiadować z już obecnymi podobnymi sobie czynnikami, to można zachować poprzedni sposób zapisu składowych.

Jak już wspominaliśmy, izomorfizmy $g: L^* \to L$ i $g^{-1}: L \to L^*$ w zastosowaniach pochodzą najczęściej od formy dwuliniowej symetrycznej i niezdegenerowanej g_{ij} na L. Ponieważ sama ta forma jest tensorem, można do niej zastosować operacje podnoszenia i opuszczania wskaźników. Opiszemy ten formalizm dokładniej.

Forma g_{ij} przyporządkowuje wektorowi a^i funkcjonał liniowy

$$b^j \mapsto \sum g_{ij} a^i b^j.$$

Współrzędnymi tego funkcjonału w bazie dwoistej przestrzeni L^* są $g_{ij}a^i$ (sumowanie względem i) lub ze względu na symetrię $g_{ij}a^j$. Innymi słowy, *opuszczanie (jedynego) górnego wskaźnika tensora a^i za pomocą tensora metrycznego g_{ij}* prowadzi do tensora

$$a_i = g_{ij} a^j.$$

Stąd natychmiast otrzymujemy ogólny wzór dla opuszczania dowolnej liczby wskaźników tensora rozkładalnego, a następnie przez liniowość dla dowolnego tensora:

$$T_{i_1 \dots i_p j_1 \dots j_r}{}^{j_{r+1} \dots j_q} = g_{j_1 j_1'} \cdots g_{j_r j_r'} T_{i_1 \dots i_p}{}^{j_1' \dots j_r' j_{r+1} \dots j_q}.$$

W szczególności możemy go wykorzystać do znalezienia tensora g^{ij}, otrzymanego z g_{ij} przez podniesienie wskaźników. Istotnie,

$$g_{ij} = g_{ik} g_{jl} g^{kl}.$$

Odczytajmy teraz prawą stronę wzoru jako wyrażenie dla (i,j)-go wyrazu macierzy otrzymanej przez pomnożenie macierzy (g_{ik}) przez $\left(\sum_l g_{jl} g^{kl}\right)$. Ponieważ po lewej stronie mamy także macierz (g_{ik}), więc oczywiście

$$g_{jl} g^{kl} = \delta^k,$$

tj. *macierz* (g^{kl}) *jest macierzą odwrotną do* (g_{ij}) (*uwzględnić symetrię*). Ten sam rachunek pokazuje, że g_i^j jest tensorem Kroneckera.

To dowodzi, że ogólny wzór dla podnoszenia wskaźników ma postać

$$T_{i_1 \dots i_b}{}^{i_{b+1} \dots i_p j_1 \dots j_q} = g_{i_1 i_1'} \cdots g_{i_p i_p'} T_{i_1 \dots i_b i_{b+1}' \dots i_p'}{}^{j_1 \dots j_q}.$$

Dla podnoszenia (lub opuszczania) innych kombinacji wskaźników należy posłużyć się zmienionymi w oczywisty sposób wzorami.

§ 5. Tensory symetryczne

1. Niech L będzie ustaloną przestrzenią liniową i $\mathcal{T}_0^q(L) = L^{\otimes q}$, $q \geq 1$. W paragrafie 2, p. 3 pokazaliśmy, że każdej permutacji liczb $1, \dots, q$ można przyporządkować odwzorowanie liniowe $f_\sigma \colon \mathcal{T}_0^q(L) \to \mathcal{T}_0^q(L)$, działające na tensory rozkładalne zgodnie ze wzorem

$$f_\sigma(l_1 \otimes \cdots \otimes l_q) = l_{\sigma(1)} \otimes \cdots \otimes l_{\sigma(q)}.$$

Tensor $T \in \mathcal{T}_0^q(L)$ nazwiemy *symetrycznym*, jeśli $f_\sigma(T) = T$ dla każdej permutacji $\sigma \in S_q$. Oczywiście tensory symetryczne tworzą podprzestrzeń liniową w $\mathcal{T}_0^q(L)$. Wygodnie uważać skalary za tensory symetryczne. Tensory symetryczne z $\mathcal{T}_0^2(L^*)$ odpowiadają przy utożsamieniu z § 3, p. 1 d) formom dwuliniowym symetrycznym na L.

Przez $S^q(L)$ oznaczymy podprzestrzeń $\mathcal{T}_0^q(L)$ złożoną z tensorów symetrycznych. Skonstruujemy teraz projektor $S \colon \mathcal{T}_0^q(L) \to \mathcal{T}_0^q(L)$, z obrazem równym $S^q(L)$, zakładając, że charakterystyka ciała podstawowego jest równa zeru lub co najmniej nie jest dzielnikiem $q!$. Nazywa się on *odwzorowaniem symetryzującym* (*symetryzacją*). W klasycznej notacji zamiast $S(T)$ zapisuje się $T^{(i_1 \dots i_q)}$.

2. Stwierdzenie. *Określmy*

$$S = \frac{1}{q!} \sum_{\sigma \in S_q} f_\sigma : \mathcal{T}_0^q(L) \longrightarrow \mathcal{T}_0^q(L).$$

Wówczas $S^2 = S$ *oraz* $\operatorname{Im} S = S^q(L)$.

Dowód. Jest oczywiste, że w wyniku symetryzacji dowolnego tensora otrzymujemy tensor symetryczny, więc $\operatorname{Im} S \subset S^q(L)$. Odwrotnie, operacja symetryzacji nie zmienia tensorów symetrycznych, więc jeśli $T \in S^q(L)$, to $T = S(T)$. To pokazuje jednocześnie, że $\operatorname{Im} S = S^q(L)$ i $S^2 = S$.

3. Niech $\{e_1, \ldots, e_n\}$ będzie bazą przestrzeni L. Wówczas tensory rozkładalne $e_{i_1} \otimes \cdots \otimes e_{i_q}$ tworzą bazę przestrzeni $\mathcal{T}_0^q(L)$, a ich symetryzacje $S(e_{i_1} \otimes \cdots \otimes e_{i_q})$ rozpinają $S^q(L)$. Oznaczymy

$$S(e_i \otimes \cdots \otimes e_i) = e_i \cdots e_i.$$

Formalny iloczyn $e_{i_1} \cdots e_{i_q}$ nie zmienia się przy permutacjach indeksów, więc można w drodze umowy wprowadzić kanoniczny sposób zapisu takich tensorów symetrycznych w postaci $e_1^{a_1} \cdots e_n^{a_n}$, gdzie $a_i \geq 0$, $a_1 + \cdots + a_n = q$, a liczba a_i wskazuje, ile razy wektor e_i występuje w $e_{i_1} \otimes \cdots \otimes e_{i_q}$.

4. Stwierdzenie. *Tensory* $e_1^{a_1} \cdots e_n^{a_n} \in S^q(L)$, $a_1 + \cdots + a_n = q$, *stanowią bazę w przestrzeni* $S^q(L)$, *dzięki czemu można utożsamić tę przestrzeń z przestrzenią wielomianów jednorodnych stopnia* q *względem elementów bazy* L.

Dowód. Musimy jedynie sprawdzić, że tensory $e_1^{a_1} \cdots e_n^{a_n}$ są liniowo niezależne w $\mathcal{T}_0^q(L)$. Jeśli

$$\sum c_{a_1 \ldots a_n} e_1^{a_1} \cdots e_n^{a_n} = 0,$$

to

$$S\Big(\sum c_{a_1 \ldots a_n} \underbrace{e_1 \otimes \cdots \otimes e_1}_{a_1} \otimes \cdots \otimes \underbrace{e_n \otimes \cdots \otimes e_n}_{a_n}\Big) = 0.$$

Grupując po lewej stronie podobne składniki nietrudno przekonać się, że współczynnikami przy elementach bazy tensorowej przestrzeni $\mathcal{T}_0^q(L)$ są skalary $c_{a_1 \ldots a_n}$, pomnożone przez liczby całkowite będące iloczynami liczb pierwszych $\leq q$. Ponieważ z założenia charakterystyka $\boldsymbol{K}$ jest większa niż $q!$, więc z równości zeru tych współczynników wynika, że wszystkie $c_{a_1 \ldots a_n}$ są równe zeru.

5. Wniosek. $\dim S^q(L) = \binom{n+q-1}{q}$.

6. Oznaczmy $S(L) = \bigoplus_{q=0}^{\infty} S^q(L)$. Na mocy stwierdzenia p. 4 można utożsamić $S(L)$ z przestrzenią wszystkich wielomianów względem elementów bazy L. Ta przestrzeń ma strukturę algebry, w której mnożeniem jest zwykłe mnożenie wielomianów.

Jednakże być może nie jest od razu jasne, czy to mnożenie nie zależy od wyboru bazy wyjściowej. Dlatego wprowadzimy je w sposób niezmienniczy. Ze względu na to, że niżej rozpatrujemy wszystkie $S^q(L)$ jednocześnie, założymy, że charakterystyka $\boldsymbol{K}$ jest równa zeru.

7. Stwierdzenie. *Wprowadźmy w przestrzeni $S(L)$ dwuliniowe mnożenie wzorem*

$$T_1T_2 = S(T_1 \otimes T_2), \quad T_1 \in S^p(L),\ T_2 \in S^q(L).$$

Zadaje ono w $S(L)$ strukturę łącznej, przemiennej algebry nad ciałem $\boldsymbol{K}$. Jeśli tensory symetryczne są przedstawione w postaci wielomianów względem elementów bazy L, to mnożenie odpowiada zwykłemu mnożeniu wielomianów.

Dowód. Sprawdzimy najpierw, że dla dowolnych tensorów $T_1 \in T_0^p(L)$, $T_2 \in T_0^q(L)$ zachodzi wzór

$$S(S(T_1) \otimes T_2) = S(T_1 \otimes S(T_2)) = S(T_1 \otimes T_2).$$

W istocie

$$S(T_1) \otimes T_2 = \frac{1}{p!} \sum_{\sigma \in S_p} f_\sigma(T_1) \otimes T_2,$$

skąd

$$S(S(T_1) \otimes T_2) = \frac{1}{p!} \sum_{\sigma \in S_p} f_\sigma(T_1) \otimes T_2.$$

Dalej $S(f_\sigma(T_1) \otimes T_2) = S(T_1 \otimes T_2)$ dla dowolnych $\sigma \in S_p$. Jest to oczywiste w przypadku rozkładalnych tensorów T_1, T_2, a dla pozostałych wynika z liniowości. Dlatego suma po prawej stronie zawiera $p!$ składników $S(T_1 \otimes T_2)$, tak że

$$S(S(T_1) \otimes T_2) = S(T_1 \otimes T_2).$$

Analogicznie dowodzi się drugiej równości.

Stąd łatwo wywnioskować, że dla tensorów symetrycznych operacja $(T_1, T_2) \mapsto S(T_1 \otimes T_2) = T_1T_2$ jest łączna. Istotnie,

$$(T_1T_2)T_3 = S(S(T_1 \otimes T_2) \otimes T_3) = S(T_1 \otimes T_2 \otimes T_3)$$

i analogicznie

$$T_1(T_2T_3) = S(T_1 \otimes S(T_2 \otimes T_3)) = S(T_1 \otimes T_2 \otimes T_3).$$

Jest ona także przemienna: wzór $S(T_1 \otimes T_2) = S(T_2 \otimes T_1)$ jest oczywisty dla tensorów rozkładalnych, a dla pozostałych wynika z liniowości.

Z powyższych rozważań wnioskujemy, że

$$\left(e_1^{a_1} \dots e_n^{a_n}\right)\left(e_1^{b_1} \dots e_n^{b_n}\right) = e_1^{a_1 + b_1} \dots e_n^{a_n + b_n},$$

co kończy dowód.

8. Skonstruowana powyżej algebra $S(L)$ jest nazywana *algebrą symetryczną* przestrzeni L.

Elementy algebry $S(L^*)$ można traktować jako funkcje wielomianowe na przestrzeni L o wartościach w $\boldsymbol{K}$: elementowi $f \in L^*$ przyporządkowuje się jego samego traktowanego jako funkcjonał liniowy na L, a iloczynowi elementów z $S(L^*)$ i ich kombinacji liniowej — iloczyn i kombinację liniową odpowiadających im funkcji. Nie jest bez reszty oczywiste, że różne elementy $S(L^*)$ różnią się także jako funkcje na przestrzeni L. To zagadnienie pozostawiamy czytelnikowi jako ćwiczenie. Dla algebry symetrycznej przestrzeni nad skończonym ciałem, którą wprowadzimy poniżej, sytuacja jest już inna, np. funkcja $x^p - x$ jest tożsamościowo równa zeru w ciele $\boldsymbol{K}$ o p elementach.

9. Inne określenie algebry symetrycznej. W przyjętym przez nas określeniu algebry symetrycznej za pomocą operatora symetryzacji S wykonywane jest dzielenie przez $p!$. Jest to niemożliwe dla ciał o skończonej chrakterystyce i w teorii modułów nad pierścieniami, gdzie także jest wykorzystywany z wielkim pożytkiem formalizm algebry tensorowej. Z tego względu pokrótce przedyskutujemy inne określenie algebry symetrycznej przestrzeni L, w którym jest ona zrealizowana nie jako podprzestrzeń, a jako przestrzeń ilorazowa $\mathcal{T}_0(T) = \bigoplus_{p=0}^{\infty} \mathcal{T}_0^p(L)$.

W tym celu rozważymy ideał dwustronny I w algebrze tensorowej $\mathcal{T}_0(T)$, generowany przez wszystkie elementy postaci

$$T - f_\sigma(T),\ T \in \mathcal{T}_0^p(L),\ \sigma \in S_p, \quad p = 1, 2, 3, \ldots$$

Składa się on ze wszystkich możliwych sum tensorów takiego typu, pomnożonych z lewej i prawej strony przez dowolne elementy $\mathcal{T}_0(T)$. Nietrudno spostrzec, że $I = \bigoplus_{p=1}^{\infty} I^p$, gdzie $I^p = I \cap \mathcal{T}_0^p(L)$, tj. że ten ideał jest ideałem z gradacją.

Określimy

$$\tilde{S}(L) = \mathcal{T}_0(L)/I$$

tymczasem tylko ze strukturą przestrzeni ilorazowej. Te same rozważania, co w § 11 części 3 pokazują, że

$$\tilde{S}(L) = \bigoplus_{p=0}^{\infty} \tilde{S}^p(L), \quad \tilde{S}^p(L) = \mathcal{T}_0^p(L)/I^p.$$

Ze względu na fakt, że I jest ideałem, w $\tilde{S}(L)$ można zadać mnożenie za pomocą wzoru

$$(T_1 + I)(T_2 + I) = T_1 \otimes T_2 + I.$$

Jest ono dwuliniowe i łączne, ponieważ te własności ma mnożenie tensorowe. Ponadto jest ono przemienne, bo jeśli T_1, T_2 są rozkładalne, to $T_2 \otimes T_1 = f_\sigma(T_1 \otimes T_2)$ dla odpowiednio dobranej permutacji σ, skąd wynika $T_1 \otimes T_2 - T_2 \otimes T_1 \in I$. W ten

sposób $\tilde{S}(L)$ otrzymuje strukturę łącznej przemiennej $\boldsymbol{K}$-algebry. Można następnie dowieść, że naturalne odwzorowanie $L \to \tilde{S}(L)\colon l \mapsto l + I$ jest włożeniem oraz elementy $\tilde{S}(L)$ jednoznacznie wyrażają się jako wielomiany względem elementów dowolnej bazy przestrzeni L. Elementy $\tilde{S}^p(L)$ odpowiadają wielomianom jednorodnym stopnia p.

Jeśli charakterystyka $\boldsymbol{K}$ jest zerem, to odwzorowanie

$$S(L) \longrightarrow \mathcal{T}_0(L) \longrightarrow \tilde{S}(L)$$

jest izomorfizmem algebr zachowującym gradację. Wobec tego, że $\tilde{S}(L)$ można zdefiniować w tej ogólniejszej sytuacji, dla potrzeb zastosowań algebraicznych wygodniej określić algebrę symetryczną właśnie tym sposobem.

§ 6. Tensory antysymetryczne i algebra zewnętrzna przestrzeni liniowej

1. Wracając do sytuacji rozpatrywanej w § 5, p. 1 tensor $T \in \mathcal{T}_0^q(L)$ nazwiemy *antysymetrycznym* (*skośnie symetrycznym* lub, rzadziej, *alternującym*), jeśli $f_\sigma(T) = \varepsilon(\sigma)T$, gdzie dla dowolnej permutacji $\sigma \in S_q$ przez $\varepsilon(\sigma)$ oznaczyliśmy znak σ. Jest oczywiste, że tensory antysymetryczne w $\mathcal{T}_0^q(L)$ stanowią podprzestrzeń liniową, natomiast skalary wygodnie jest traktować zarówno jako tensory symetryczne, jak i antysymetryczne. Przy utożsamieniu opisanym w § 3 p. 1 d) tensorom antysymetrycznym z $\mathcal{T}_0^2(L^*)$ odpowiadają formy dwuliniowe symplektyczne na L.

Przez $\bigwedge^q(L)$ oznaczymy podprzestrzeń $\mathcal{T}_0^q(L)$ złożoną z tensorów antysymetrycznych. Analogicznie do § 5 skonstruujemy projektor liniowy $A\colon \mathcal{T}_0^q(L) \to \mathcal{T}_0^q(L)$, którego obrazem jest $\bigwedge^q(L)$. Jak i tam założymy na razie, że charakterystyka ciała skalarów nie jest dzielnikiem $q!$. Projektor A będziemy nazywać *antysymetryzacją* albo *alternowaniem*. W klasycznych oznaczeniach zamiast $A(T)$ pisze się $T^{[i_1 \ldots i_q]}$.

2. Stwierdzenie. *Określmy*

$$A = \frac{1}{q!} \sum_{\sigma \in S_q} \varepsilon(q) f_\sigma\colon \mathcal{T}_0^q(L) \longrightarrow \mathcal{T}_0^q(L).$$

Wówczas $A^2 = A$ *oraz* $\operatorname{Im} A = \bigwedge^q(L)$.

Dowód. Najpierw sprawdzimy, że po zastosowaniu operacji antysymetryzacji z dowolnego tensora otrzymujemy tensor antysymetryczny. Rzeczywiście, wobec tego, że f_σ i $\varepsilon(\sigma)$ są multyplikatywne względem σ oraz $\varepsilon(\sigma)^2 = 1$ otrzymujemy

$$f_\sigma(AT) = f_\sigma\Big(\frac{1}{q!}\sum_{\tau\in S_q}\varepsilon(\tau)f_\tau(T)\Big) = \frac{1}{q!}\sum_{\tau\in S_q}\varepsilon(\tau)f_{\sigma\tau}(T) =$$

$$= \varepsilon(\sigma)\frac{1}{q!}\sum_{\tau\in S_q}\varepsilon(\sigma\tau)f_{\sigma\tau}(T) = \varepsilon(\sigma)AT.$$

Co więcej, mamy

$$A^2 = \frac{1}{(q!)^2}\sum_{\sigma,\tau\in S_q}\varepsilon(\sigma\tau)f_{\sigma\tau} = \frac{1}{q!}\sum_{\rho\in S_q}\varepsilon(\rho)f_\rho = A,$$

skąd wynika, że A jest projektorem.

Istotnie, dowolny element $\rho \in S_q$ można przedstawić w postaci iloczynu $\sigma\tau$ dokładnie $q!$ sposobami: wybieramy σ dowolnie, a τ wyznaczamy z równości $\tau = \sigma^{-1}\rho$.

Stąd tak jak w stwierdzeniu § 5, p. 1 wynika, że $\operatorname{Im} A = \bigwedge^q(L)$.

3. Niech $\{e_1, \ldots, e_n\}$ oznacza bazę przestrzeni L. Wtedy tensory proste $e_{i_1} \otimes \cdots \otimes e_{i_q}$ tworzą bazę $T_0^q(L)$, a ich obrazy po antysymetryzacji $A(e_{i_1} \otimes \cdots \otimes e_{i_q})$ rozpinają $\bigwedge^q(L)$. Wprowadzimy oznaczenie

$$A(e_{i_1} \otimes \cdots \otimes e_{i_q}) = e_{i_1} \wedge \ldots \wedge e_{i_q}$$

(znak $\wedge$ nazywany jest symbolem „mnożenia zewnętrznego").

Zauważmy teraz, że w odróżnieniu od przypadku symetrycznego, przestawienie dowolnych dwóch wektorów w iloczynie $e_{i_1} \wedge \ldots \wedge e_{i_q}$ zmienia jego znak, ze względu na antysymetrię tego tensora. Stąd wynikają następujące dwa wnioski:

a) $e_{i_1} \wedge \ldots \wedge e_{i_q} = 0$, *jeśli* $i_a = i_b$ *dla pewnych* a, b, *pod warunkiem, że* $\operatorname{char} \boldsymbol{K} \neq 2$.

b) *Przestrzeń* $\bigwedge^q(L)$ *jest rozpięta przez tensory postaci* $e_{i_1} \wedge \ldots \wedge e_{i_q}$, *gdzie* $1 \leq i_1 < i_2 < \ldots < i_q \leq n$.

W szczególności wynika stąd natychmiast, że $\bigwedge^m(L) = \{0\}$ dla $m > n = \dim L$.

Następujący rezultat jest odpowiednikiem stwierdzenia § 5, p. 4.

4. Stwierdzenie. *Tensory* $e_{i_1} \wedge \ldots \wedge e_{i_q} \in \bigwedge^q(L)$ *dla* $q \leq n$, $1 \leq i_1 < i_2 < \ldots < i_q \leq n$ *stanowią bazę przestrzeni* $\bigwedge^q(L)$.

D o w ó d. Musimy jedynie sprawdzić, że w $T_0^q(L)$ ten układ tensorów jest liniowo niezależny. Jeśli

$$\sum c_{i_1 \ldots i_q} e_{i_1} \wedge \ldots \wedge e_{i_q} = 0,$$

to

$$A(\sum c_{i_1 \ldots i_q} e_{i_1} \otimes \ldots \otimes e_{i_q}) = 0.$$

Ponieważ wśród wskaźników $i_1, \ldots, i_q$ żadne dwa nie powtarzają się, a one same są uporządkowane rosnąco, więc po wykonaniu permutacji indeksów po lewej stronie

otrzymamy kombinację liniową parami różnych elementów bazy tensorowej $T_0^q(L)$ ze współczynnikami postaci $\pm\frac{1}{q!}c_{i_1\dots i_q}$. Ta kombinacja będzie tylko wtedy równa zeru, gdy wszystkie $c_{i_1\dots i_q}$ będą zerami.

5. Wniosek. $\dim\bigwedge^q(L)=\binom{n}{q}$, $\dim\bigoplus_{q=0}^{n}\bigwedge^q(L)=2^n$.

6. Wprowadźmy oznaczenie $\bigwedge(L)=\bigoplus_{q=0}^{n}\bigwedge^q(L)$. W przestrzeni tensorów antysymetrycznych, analogicznie do przypadku tensorów symetrycznych, określimy operację zewnętrznego mnożenia i wykażemy, że wprowadza ona do $\bigwedge(L)$ strukturę algebry łącznej, zwanej *algebrą zewnętrzną* lub *algebrą Grassmanna* przestrzeni L.

7. Stwierdzenie. *Operacja dwuliniowa*

$$T_1\wedge T_2=A(T_1\otimes T_2),\quad T_1\in\bigwedge^p(L),\ T_2\in\bigwedge^q(L),$$

na $\bigwedge(L)$ jest łączna, $T_1\wedge T_2\in\bigwedge^{q+p}(L)$, $T_2\wedge T_1=(-1)^{pq}T_1\wedge T_2$ (tę własność nazywa się czasem antykomutatywnością).

W szczególności, podprzestrzeń $\bigwedge^+(L)=\bigoplus_{q=0}^{[n/2]}\bigwedge^{2q}(L)$ jest centralną podalgebrą algebry $\bigwedge(L)$.

Dowód. Podobnie jak w przypadku symetrycznym, sprawdzimy najpierw, że dla każdych $T_1\in T_0^p(L)$, $T_2\in T_0^q(L)$ zachodzą wzory

$$A(A(T_1)\otimes T_2)=A(T_1\otimes A(T_2))=A(T_1\otimes T_2).$$

W istocie

$$A(T_1)\otimes T_2=\frac{1}{p!}\sum_{\sigma\in S_p}\varepsilon(\sigma)f_\sigma(T_1)\otimes T_2,$$

skąd

$$A(A(T_1)\otimes T_2)=\sum_{\sigma\in S_p}\varepsilon(\sigma)A(f_\sigma(T_1)\otimes T_2).$$

Rozważmy włożenie $S_p\to S_{p+q}$, $\sigma\mapsto\tilde{\sigma}$, gdzie

$$\tilde{\sigma}(i)=\begin{cases}\sigma(i) & \text{dla } 1\le i\le p,\\ i & \text{dla } i>p.\end{cases}$$

Oczywiście, $f_\sigma(T_1)\otimes T_2=f_{\tilde{\sigma}}A(T_1\otimes T_2)$, a ponadto A i $f_{\tilde{\sigma}}$ komutują ze sobą, tak że $Af_{\tilde{\sigma}}(T_1\otimes T_2)=f_{\tilde{\sigma}}A(T_1\otimes T_2)=\varepsilon(\tilde{\sigma})A(T_1\otimes T_2)=\varepsilon(\sigma)A(T_1\otimes T_2)$. Dlatego

$$A(A(T_1)\otimes T_2)=\frac{1}{p!}\sum_{\sigma\in S_p}\varepsilon^2(\sigma)A(T_1\otimes T_2)=A(T_1\otimes T_2).$$

Analogicznie można wykazać drugą równość, a następnie sprawdzić łączność mnożenia zewnętrznego tak samo jak w przypadku symetrycznym:

$$(T_1 \wedge T_2) \wedge T_3 = A(A(T_1 \otimes T_2) \otimes T_3) = A(T_1 \otimes T_2 \otimes T_3),$$
$$T_1 \wedge (T_2 \wedge T_3) = A(T_1 \otimes A(T_2 \otimes T_3)) = A(T_1 \otimes T_2 \otimes T_3).$$

Wzór $A(T_1 \otimes T_2) = (-1)^{pq} A(T_2 \otimes T_1)$ dla $T_1 \in \mathcal{T}_0^p(L)$, $T_2 \in \mathcal{T}_0^q(L)$ wynika z obserwacji, że $T_2 \otimes T_1 = f_\sigma(T_1 \otimes T_2)$, gdzie σ jest permutacją będącą iloczynem pq transpozycji — czynniki T_2 należy kolejno przeprowadzać na lewą stronę T_1, za każdym razem przestawiając je na miejsce ich lewego sąsiada z T_1.

8. Inne określenie algebry zewnętrznej. Podobnie jak w przypadku symetrycznym, wadą przyjętego przez nas określenia algebry zewnętrznej jest konieczność wykonywania dzielenia przez $p!$. Inną, pozbawioną tej wady definicję, dzięki której $\bigwedge(L)$ jest zrealizowana jako przestrzeń ilorazowa, a nie podprzestrzeń $\mathcal{T}_0(L)$, wprowadzamy w pełnej analogii do przypadku algebry symetrycznej.

Rozważmy ideał dwustronny J w algebrze $\mathcal{T}_0(T)$, generowany przez wszystkie elementy postaci

$$T - \varepsilon(\sigma) f_\sigma(T), \quad T \in \mathcal{T}_0^p(L), \quad \sigma \in S_p, \quad p = 1, 2, 3, \ldots$$

Nietrudno przekonać się, że $J = \bigoplus_{p=1}^{\infty} J^p$, gdzie $J^p = J \cap \mathcal{T}_0^p(L)$, a więc jest to ideał z gradacją. Określimy przestrzeń ilorazową $\widetilde{\bigwedge}(L) = \mathcal{T}_0(L)/J$. Wtedy

$$\widetilde{\bigwedge}(L) = \bigoplus_{p=0}^{\infty} \widetilde{\bigwedge}^p(L), \quad \widetilde{\bigwedge}^p(L) = \mathcal{T}_0^p(L)/J^p.$$

Ponieważ J jest ideałem, więc można zdefiniować mnożenie w $\widetilde{\bigwedge}(L)$ za pomocą wzoru

$$(T_1 + J) \wedge (T_2 + J) = T_1 \otimes T_2 + J.$$

Jest ono dwuliniowe i łączne, ponieważ te własności przysługują mnożeniu tensorowemu. Jest ono ponadto antykomutatywne, gdyż dla $T_1 \in \mathcal{T}_0^p(L)$, $T_2 \in \mathcal{T}_0^q(L)$ zachodzi $T_1 \otimes T_2 - (-1)^{pq} T_2 \otimes T_1 \in J$.

Nietrudno sprawdzić, że *skonstruowana w ten sposób algebra jest izomorficzna z wprowadzoną w* § 15 *części* 2 *algebrą Clifforda przestrzeni* L *wyposażonej w zerowy iloczyn skalarny.*

Rzeczywiście, odwzorowanie $\sigma: L \to \widetilde{\bigwedge}(L)$, $\sigma(l) = l + J$, spełnia warunek $\sigma(l)^2 = \sigma(l) \wedge \sigma(l) = 0$ dla każdego l, gdyż $\sigma(l) \wedge \sigma(l) = -\sigma(l) \wedge \sigma(l)$. A zatem na mocy twierdzenia § 15, p. 2 części 2 istnieje jedyny homomorfizm $\boldsymbol{K}$-algebr $C(L) \longrightarrow \bigwedge(L)$, taki że σ jest superpozycją $L \xrightarrow{\rho} C(L) \longrightarrow \widetilde{\bigwedge}(L)$, gdzie ρ oznacza odwzorowanie kanoniczne. Ponieważ L generuje $\mathcal{T}_0(L)$ jako algebrę, a $\sigma(L)$ generuje $\widetilde{\bigwedge}(L)$, więc $C(L) \to \widetilde{\bigwedge}(L)$ jest surjekcją. Z drugiej strony wobec $\dim C(L) = 2^n$ wystarczy

sprawdzić, że $\dim \widetilde{\bigwedge}(L) = 2^n$. Można to zrobić sprawdzając, że elementy postaci $e_{i_1} \wedge \ldots \wedge e_{i_q}$, $1 \leq i_1 < i_2 < \ldots < i_q \leq n$, gdzie $\{e_1, \ldots, e_n\}$ oznacza bazę L, stanowią bazę $\widetilde{\bigwedge}^q(L)$. To sprawdzenie opuścimy.

Podobnie jak w przypadku symetrycznym, jeśli charakterystyka $\boldsymbol{K}$ jest zerem, to złożenie odwzorowań

$$\bigwedge(L) \longrightarrow \mathcal{T}_0(L) \longrightarrow \widetilde{\bigwedge}(L)$$

jest także izomorfizmem algebr zachowującym gradację.

Wobec tego, że $\widetilde{\bigwedge}(L)$ jest zdefiniowana w ogólniejszej sytuacji, w ten właśnie sposób wprowadza się algebrę zewnętrzną dla potrzeb samej algebry. W zastosowaniach do geometrii różniczkowej lub analizy, gdzie $\boldsymbol{K} = \boldsymbol{R}$ albo $\boldsymbol{C}$, można posługiwać się naszym wyjściowym określeniem.

9. Iloczyn zewnętrzny i wyznaczniki. Niech L oznacza n-wymiarową przestrzeń liniową. Zgodnie z wnioskiem p. 5 przestrzeń $\bigwedge^n(L)$ jest jednowymiarowa — jest to maksymalna niezerowa potęga zewnętrzna przestrzeni L.

Na mocy § 2, p. 7 dowolny endomorfizm $f: L \to L$ indukuje endomorfizm iloczynów tensorowych

$$f^{\otimes p} = f \otimes \cdots \otimes f: \mathcal{T}_0^p(L) \longrightarrow \mathcal{T}_0^p(L).$$

Łatwo sprawdzić, że $f^{\otimes p}$ komutuje z operatorem alternowania A i z tego powodu odwzorowuje $\bigwedge^p(L)$ w $\bigwedge^p(L)$. Obcięcie do $\bigwedge^p(L)$ odwzorowania $f^{\otimes p}$ naturalnie będzie oznaczyć przez $f^{\wedge p}$. W szczególności, dla $p = n$ odwzorowanie $f^{\wedge n}: \bigwedge^n(L) \to \bigwedge^n(L)$ powinno być mnożeniem przez skalar $d(f)$, gdyż $\bigwedge^n(L)$ jest jednowymiarowa.

10. Twierdzenie. *W podanych oznaczeniach* $d(f) = \det f$.

Dowód. Obierzmy bazę $\{e_1, \ldots, e_n\}$ przestrzeni L i wyraźmy f za pomocą macierzy w tej bazie:

$$f(e_j) = \sum_{i=1}^{n} a_j^i e_i.$$

Iloczyn zewnętrzny $e_1 \wedge \ldots \wedge e_n$ jest bazą $\bigwedge^n(L)$, a zatem liczbę $d(f)$ można wyznaczyć z równości

$$f^{\wedge n}(e_1 \wedge \ldots \wedge e_n) = d(f) e_1 \wedge \ldots \wedge e_n.$$

Jednakże

$$f^{\wedge n}(e_1 \wedge \ldots \wedge e_n) = A(f(e_1) \otimes \cdots \otimes f(e_n)) = f(e_1) \wedge \ldots \wedge f(e_n) = \\ = \left(\sum_{i_1=1}^{n} a_1^{i_1} e_{i_1}\right) \wedge \ldots \wedge \left(\sum_{i_n=1}^{n} a_n^{i_n} e_{i_n}\right).$$

Zgodnie z tabelką mnożenia w algebrze zewnętrznej

$$a_1^{i_1} e_{i_1} \wedge a_2^{i_2} e_{i_2} \wedge \ldots \wedge a_n^{i_n} e_{i_n} = \\ = \begin{cases} \varepsilon(\sigma) a_1^{i_1} \ldots a_n^{i_n} e_1 \wedge \ldots \wedge e_n, & \text{jeśli } \{i_1, \ldots, i_n\} = \{1, \ldots, n\}, \\ 0, & \text{w pozostałych przypadkach,} \end{cases}$$

gdzie przez σ została oznaczona permutacja przeprowadzająca i_k na k dla $1 \leq k \leq n$. Zatem suma wszystkich współczynników $\varepsilon(\sigma)a_1^{i_1} \ldots a_n^{i_n}$ jest identyczna ze standardowym wyrażeniem dla wyznacznika $\det(a_j^i)$, co kończy dowód.

11. Wniosek. *Układ wektorów $e'_1, \ldots, e'_n \in L$ jest liniowo zależny wtedy i tylko wtedy, gdy $e'_1 \wedge \ldots \wedge e'_n = 0$.*

Rzeczywiście, niech $f: L \to L$ będzie izomorfizmem przeprowadzającym e_i na e'_i, gdzie $\{e_1, \ldots, e_n\}$ jest pewną bazą L. Wówczas liniowa zależność układu $\{e'_i\}$ jest równoważna z tym, że $\det f = 0$, tj. $e'_1 \wedge \ldots \wedge e'_n = 0$.

12. Proste p-wektory Przyjęło się nazywać elementy $T \in \bigwedge^p(L)$ *p-wektorami.* T nazywamy *p-wektorem prostym* (*rozkładalnym*), jeśli istnieją takie wektory $e_1, \ldots, e_p \in L$, że $T = e_1 \wedge \ldots \wedge e_p$. Dla dowolnego p-wektora T zbiór

$$\operatorname{Ann} T = \{e \in L \mid e \wedge T = 0\}$$

nazwiemy jego *anihilatorem*. Oczywiście, $\operatorname{Ann} T$ jest podprzestrzenią L.

13. Twierdzenie. *Niech T_1 będzie rozkładalnym p-wektorem i T_2 — rozkładalnym q-wektorem, a L_1, odpowiednio, L_2 — ich anihilatorami. Wówczas*

a) *$L_1 \supset L_2$ wtedy i tylko wtedy, gdy T_1 dzieli się przez T_2, tj. $T_1 = T \wedge T_2$ dla pewnego $T \in \bigwedge^{p-q}(L)$;*

b) *$L_1 \cap L_2 = \{0\}$ wtedy i tylko wtedy, gdy $T_1 \wedge T_2 \neq 0$;*

c) *jeśli $L_1 \cap L_2 = \{0\}$, to $L_1 + L_2 = \operatorname{Ann}(T_1 \wedge T_2)$.*

Dowód. a) Jeśli $x \wedge T_2 = 0$, to $x \wedge (T \wedge T_2) = \pm T \wedge (x \wedge T_2) = 0$, a więc z podzielności T_1 przez T_2 wynika, że $L_2 \subset L_1$.

Dla dowodu odwrotnej implikacji wyznaczymy anihilator p-wektora $e_1 \wedge \ldots \wedge e_p$. Jeśli układ $e_1, \ldots, e_p$ jest liniowo zależny, to jeden z wektorów e_i, powiedzmy e_1, wyraża się liniowo poprzez pozostałe, a wówczas

$$e_1 \wedge \ldots \wedge e_p = \left(\sum_{i=2}^{n} a^i e_i\right) \wedge e_2 \ldots \wedge e_n = 0.$$

Przyjmiemy, że $e_1 \wedge \ldots \wedge e_p$ nie jest zerem i pokażemy, że wówczas $\operatorname{Ann}(e_1 \wedge \ldots \wedge e_p)$ jest powłoką liniową wektorów $e_1, \ldots, e_p$. Jest oczywiste, że ta powłoka liniowa jest zawarta w anihilatorze, bo

$$e_j \wedge (e_1 \wedge \ldots \wedge e_p) = \pm(e_j \wedge e_j) \wedge (e_1 \wedge \ldots \wedge e_{j-1} \wedge e_{j+1} \wedge \ldots \wedge e_p) = 0.$$

Uzupełnimy liniowo niezależny układ wektorów $\{e_1, \ldots, e_p\}$ do bazy $\{e_1, \ldots, e_p, e_{p+1}, \ldots, e_n\}$ przestrzeni L oraz wykażemy, że jeśli $\sum_{i=1}^{n} a^i e_i \in \operatorname{Ann}(e_1 \wedge \ldots \wedge e_p)$, to $a^i = 0$ dla $i > p$. W istocie,

$$\left(\sum_{i=1}^{n} a^i e_i\right) \wedge (e_1 \wedge \ldots \wedge e_p) = \sum_{i=p+1}^{n} a^i e_i \wedge (e_1 \wedge \ldots \wedge e_p),$$

a $(p+1)$-wektory $e_i \wedge e_1 \wedge \ldots \wedge e_p$, dla $p+1 \leq i \leq n$ są liniowo niezależne.

Załóżymy teraz, że $L_1 \supset L_2$ i niech $T_1 = e_1 \wedge \ldots \wedge e_p$, $T_2 = e'_1 \wedge \ldots \wedge e'_q$. Powłoka liniowa $\{e_1, \ldots, e_p\}$ zawiera powłokę liniową $\{e'_1, \ldots, e'_q\}$, a więc ma bazę postaci $\{e'_1, \ldots, e'_q, e'_{q+1}, \ldots, e'_p\}$. Możemy zatem wyrazić wektory e_j jako kombinacje liniowe wektorów tej bazy, co pozwoli zapisać T_1 w postaci $ae'_1 \wedge \ldots \wedge e'_q \wedge e'_{q+1} \wedge \ldots \wedge e'_p$, gdzie a jest wyznacznikiem macierzy przejścia od bazy primowanej do bazy nieprimowanej. To pokazuje, że T_1 jest podzielne przez T_2.

b), c). Jeśli $(e_1 \wedge \ldots \wedge e_p) \wedge (e'_1 \wedge \ldots \wedge e'_q) \neq 0$, to układ $\{e_1, \ldots, e_p, e'_1, \ldots, e'_q\}$ jest liniowo niezależny. Stąd wynika, że przecięcie powłok liniowych $\{e_1, \ldots, e_p\}$ i $\{e'_1, \ldots, e'_q\}$, tj. anihilatorów T_1 i T_2, zawiera tylko wektor zerowy. Ten argument daje się oczywiście odwrócić. Z charakteryzacji anihilatora p-wektora prostego, wykazanej w trakcie dowodu punktu a), wynika teraz ostatni punkt tezy.

14. Wniosek. *Odwzorowanie*

$$\mathrm{Ann}: \left(\begin{array}{l} \textit{niezerowe } p\textit{-wektory proste} \\ \textit{z dokładnością do czynnika skalarnego} \end{array} \right) \longrightarrow$$

$$\longrightarrow (p\textit{-wymiarowe podprzestrzenie w } L).$$

jest bijekcją.

Dowód. Jest widoczne, że anihilatory dwóch różniących się o czynnik skalarny niezerowych wektorów prostych są jednakowe, a więc określenie powyższego odwzorowania jest poprawne. Każda p-wymiarowa podprzestrzeń $L_1 \subset L$ jest zawarta w obrazie tego odwzorowania, bowiem jeśli $\{e_1, \ldots, e_p\}$ jest bazą L_1, to $L_1 = \mathrm{Ann}\,(e_1 \wedge \ldots \wedge e_p)$. I wreszcie, injektywność wynika z tezy a) twierdzenia p. 13 — jeśli $\mathrm{Ann}\,T_1 = \mathrm{Ann}\,T_2$, to $T_1 = T \wedge T_2$, a T jest 0-wektorem, tj. skalarem.

15. Rozmaitości Grassmanna. *Rozmaitością Grassmanna* lub *grassmannianem* $\mathrm{Gr}\,(p, L)$ nazywamy zbiór wszystkich p-wymiarowych podprzestrzeni przestrzeni L. Dla $p = 1$ otrzymujemy szczegółowo zbadaną poprzednio przestrzeń rzutową $P(L)$. Wniosek p. 14 umożliwia także w przypadku dowolnego p zrealizowanie $\mathrm{Gr}\,(p, L)$ jako podzbioru przestrzeni rzutowej $P(\bigwedge^p(L))$.

W istocie, odwzorowanie odwrotne do Ann wyznacza włożenie

$$\mathrm{Ann}^{-1}: \mathrm{Gr}\,(p, L) \longrightarrow P(\overset{p}{\bigwedge}(L)).$$

Zapiszemy je w jawnej postaci. Wybierzmy bazę $\{e_1, \ldots, e_n\}$ w L i rozważmy powłokę liniową układu p wektorów

$$\sum_{i=1}^{n} a^i_j e_i, \quad j = 1, \ldots, p.$$

Bazę $\bigwedge^p(L)$ stanowi układ $\binom{n}{p}$ p-wektorów $\{e_{i_1} \wedge \ldots \wedge e_{i_p} \mid 1 \leq i_1 < \ldots < i_p \leq n\}$. Odwzorowanie Ann^{-1} przyporządkowuje tej powłoce liniowej prostą w $\bigwedge^p(L)$

generowaną przez p-wektor

$$\Big(\sum_{i_1=1}^{n} a_1^{i_1} e_{i_1}\Big) \wedge \ldots \wedge \Big(\sum_{i_p=1}^{n} a_p^{i_p} e_{i_p}\Big).$$

Współrzędnymi jednorodnymi odpowiadającego jej punktu w $P(\bigwedge^p(L))$ są współrzędne tego p-wektora względem bazy $\{e_{i_1} \wedge \ldots \wedge e_{i_p}\}$:

$$\bigwedge_{j=1}^{p}\Big(\sum_{i=1}^{n} a_j^i e_i\Big) = \sum_{1\le i_1<\ldots<i_p\le n} \Delta^{i_1\ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

Dokładnie taki sam rachunek, jak przy dowodzie twierdzenia p. 10 pokazuje, że $\Delta^{i_1\ldots i_p}$ jest minorem macierzy (a_j^i) złożonym z wierszy o wskaźnikach $i_1, \ldots, i_p$. Przynajmniej jeden z tych minorów jest różny od zera dokładnie wtedy, gdy rząd macierzy (a_j^i) ma maksymalną wartość p, tj. gdy powłoka liniowa wybranych p-wektorów jest rzeczywiście p-wymiarowa.

Wektor $(\ldots, : \Delta^{i_1\ldots i_p} : \ldots)$ jest nazywany wektorem *grassmannowskich współrzędnych* p-wymiarowej podprzestrzeni rozpiętej przez

$$\sum_{j=1}^{n} a_j^i e_i, \quad j = 1, \ldots, p.$$

Z tej konstrukcji można wywnioskować, że dla scharakteryzowania obrazu $\mathrm{Gr}\,(p, L)$ w $P(\bigwedge^p(L))$ potrzebne jest kryterium prostoty p-wektorów. Temu zagadnieniu jest poświęcone następujące twierdzenie.

16. Twierdzenie. a) *Niezerowy p-wektor T jest prosty wtedy i tylko wtedy, gdy* $\dim \mathrm{Ann}\, T = p$. *Dla pozostałych p-wektorów niezerowych* $\dim \mathrm{Ann}\, T < p$.

b) *Ustalmy bazę $\{e_1, \ldots, e_n\}$ przestrzeni L i niech p-wektory T będą zadane poprzez swoje współrzędne względem bazy $\{e_{i_1} \wedge \ldots \wedge e_{i_p} \mid 1 \le i_1 < i_2 < \ldots < i_p \le n\}$ w $\bigwedge^p(L)$:*

$$T = \sum T^{i_1\ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

Wówczas istnieje taki układ równań wielomianowych o współczynnikach całkowitych względem $T^{i_1\ldots i_p}$, zależny tylko od n i p, że p-wektor T jest prosty wtedy i tylko wtedy, gdy $T^{i_1\ldots i_p}$ jest rozwiązaniem tego układu.

Dowód. Równość $\dim \mathrm{Ann}\, T = p$ dla p-wektorów prostych wykazaliśmy już przy dowodzie twierdzenia p. 10.

Niech $\dim \mathrm{Ann}\, T = r$ i wektory $e_1, \ldots, e_r$ rozpinają $\mathrm{Ann}\, T$. Uzupełnimy ten układ do bazy $\{e_1, \ldots, e_n\}$ w L i przyjmiemy

$$T = \sum T^{i_1\ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}.$$

Z warunku $e_i \wedge T = 0$ dla każdego $i = 1, \ldots, r$ wynika, że $T^{i_1 \ldots i_p} = 0$, gdy tylko $\{1, \ldots, r\} \not\subset \{i_1, \ldots, i_p\}$. Dalej jest też jasne, że jeśli $T \neq 0$, to $r \leq p$ i że T jest podzielne przez $e_1 \wedge \ldots \wedge e_r$. Dlatego jeśli $r = p$, to p-wektor T jest proporcjonalny do $e_1 \wedge \ldots \wedge e_r$, a więc jest prosty.

b) Korzystając z powyższego, możemy jako kryterium prostoty p-wektora T podać warunek, aby przestrzeń rozwiązań następującego układu równań liniowych z niewiadomymi $x^1, \ldots, x^n \in \boldsymbol{K}$ miała wymiar p:

$$\left(\sum_{i=1}^{n} x^i e_i\right) \wedge \left(\sum T^{i_1 \ldots i_p} e_{i_1} \wedge \ldots \wedge e_{i_p}\right) = 0.$$

Układ ma n niewiadomych i składa się z $\binom{n}{p+1}$ równań, a wyrazami macierzy układu są całkowitoliczbowe kombinacje liniowe $T^{i_1 \ldots i_p}$. Rząd tej macierzy jest zawsze $\geq n - p$, bo $\dim \mathrm{Ann} \leq p$. Stąd wynika, że prostota T jest równoważna z żądaniem, by rząd tej macierzy był $\leq n - p$, tj. by zerowały się wszystkie minory rzędu $(n - p + 1)$ tej macierzy. To właśnie jest szukanym układem równań dla współrzędnych grasmannowskich tensora prostego.

Rozpatrzmy kilka przykładów i szczególnych przypadków.

17. Stwierdzenie. *Dowolny $(n-1)$-wektor T jest prosty.*

Dowód. Oczywiście, $x \wedge T = f(x) e_1 \wedge \ldots \wedge e_n$, gdzie $f(x)$ jest jakąś funkcją liniową na L, a $\{e_1, \ldots, e_n\}$ jest ustaloną bazą L. A więc $\dim \mathrm{Ann}\, T = \dim \mathrm{Ker}\, f \geq n - 1$. Jeśli jednak $T \neq 0$, to $f \neq 0$, a więc także $\dim \mathrm{Ker}\, f = n - 1$. Na mocy części a) twierdzenia p. 16 T jest prosty.

W odniesieniu do rozmaitości Grassmannna stwierdzenie to oznacza istnienie bijekcji

$$(\textit{hiperpłaszczyzny w } L) \longrightarrow P(\textstyle\bigwedge^{n-1}(L)).$$

Ponieważ hiperpłaszczyzny w L odpowiadają punktom $P(L^*)$, otrzymujemy izomorfizm kanoniczny

$$P(L^*) \cong P(\textstyle\bigwedge^{n-1}(L)).$$

Poniżej podamy uogólnienie tego rezultatu.

18. Stwierdzenie. *Niezerowy dwuwektor $T \in \bigwedge^2(T)$ jest wtedy i tylko wtedy prosty, gdy $T \wedge T = 0$.*

Dowód. Konieczność warunku jest oczywista. Dowód dostateczności przeprowadzimy przez indukcję względem wymiaru n, zaczynając od trywialnego przypadku $n = 2$. Niech $\{e_1, \ldots, e_{n+1}\}$ oznacza bazę L. Wyrażając T przez $e_i \wedge e_j$ możemy zapisać T w postaci $T = e_{n+1} \wedge T_1 + T_2$, gdzie T_1 jest kombinacją e_i, T_2 — kombinacją $e_i \wedge e_j$, gdzie $1 \leq i, j \leq n$. Z warunku $T \wedge T = 0$ wynika, że

$$T_2 \wedge T_2 + 2 e_{n+1} \wedge T_1 \wedge T_2 = 0,$$

bo $(e_{n+1} \wedge T_1) \wedge (e_{n+1} \wedge T_1) = 0$ i T_2 należy do centrum $\bigwedge(L)$. Iloczyny zawierające e_{n+1} nie mogą występować w $T_2 \wedge T_2$ i dlatego

$$T_2 \wedge T_2 = e_{n+1} \wedge T_1 \wedge T_2 = 0.$$

Wobec $T_2 \wedge T_2 = 0$, na mocy założenia indukcyjnego T_2 jest prosty. Ponieważ $T_1 \wedge T_2$ nie zawiera iloczynów z e_{n+1}, wnioskujemy, że $T_1 \wedge T_2 = 0$. To oznacza, że T_1 należy do dwuwymiarowego anihilatora T_2 i $T_2 = T_1' \wedge T_1$. Ostatecznie

$$T = e_{n+1} \wedge T_1 + T_1' \wedge T_1 = (e_{n+1} + T_1') \wedge T_1,$$

co kończy dowód.

Ten rezultat także dostarcza informacji o rozmaitościach Grassmanna, tym razem o $\mathrm{Gr}(2, L)$.

19. Wniosek. *Kanoniczne odwzorowanie* $\mathrm{Gr}(2, L) \to P(\bigwedge^2(L))$ *pozwala utożsamić przy* $n \geq 3$ *grassmannian płaszczyzn w* L *z przecięciem kwadryk w* $P(\bigwedge^2(L))$.

Dowód. Płaszczyzny w L odpowiadają prostym rozpiętym przez dwuwektory proste w $\bigwedge^2(L)$. Warunek prostoty dwuwektora $\sum_{i_1<i_2} T^{i_1 i_2} e_{i_1} \wedge e_{i_2}$ ma na podstawie stwierdzenia p. 18 postać

$$\Big(\sum_{i_1<i_2} T^{i_1 i_2} e_{i_1} \wedge e_{i_2}\Big)\Big(\sum_{j_1<j_2} T^{j_1 j_2} e_{j_1} \wedge e_{j_2}\Big) = 0$$

lub inaczej

$$\sum_{i_1<i_2} T^{i_1 i_2} T^{j_1 j_2} \varepsilon(i_1, i_2, j_1, j_2) = 0,$$

gdzie każdy składnik po lewej odpowiada czwórce wskaźników $1 \leq k_1 < k_2 < k_3 < k_4 \leq n$, a $\varepsilon(i_1, i_2, j_1, j_2)$ jest znakiem permutacji rozmieszczającej elementy zbioru $\{i_1, i_2, j_1, j_2\} = \{k_1, k_2, k_3, k_4\}$ w porządku rosnącym.

W szczególności przy $n = 4$ otrzymujemy tylko jedno równanie

$$T^{12} T^{34} - T^{13} T^{24} + T^{14} T^{23} = 0,$$

z którego wynika, że $\mathrm{Gr}(2, \boldsymbol{K}^4)$ jest czterowymiarową kwadryką w $P(\bigwedge^2(\boldsymbol{K}^4)) = P(\boldsymbol{K}^6)$. Nazywamy ją kwadryką Plückera.

20. Mnożenie zewnętrzne i dwoistość. Niech $\dim L = n$. Zgodnie z wnioskiem p. 5 i znaną symetrią współczynników dwumiennych mamy dla każdego $1 \leq p \leq n$

$$\dim \bigwedge^p(L) = \binom{n}{p} = \binom{n}{n-p} = \dim \bigwedge^{n-p}(L).$$

To może sugerować istnienie kanonicznego izomorfizmu bądź kanonicznej dwoistości między tymi przestrzeniami. Z dokładnością do niewielkiej komplikacji realizuje się ta druga możliwość.

Rozważmy operację mnożenia zewnętrznego

$$\bigwedge^p(L) \times \bigwedge^{n-p}(L) \longrightarrow \bigwedge^n(L) : (T_1, T_2) \mapsto T \wedge T_2.$$

Z jej dwuliniowości wynika istnienie odwzorowania liniowego

$$\bigwedge^p(L) \longrightarrow \mathcal{L}(\bigwedge^{n-p}(L), \bigwedge^n(L)) \cong (\bigwedge^{n-p}(L))^* \otimes \bigwedge^n(L)$$

(ostatni izomorfizm jest szczególnym przypadkiem izomorfizmu opisanego w § 2, p. 5). Jądrem tego odwzorowania jest przestrzeń zerowa. Istotnie, niech $\{e_1, \ldots, e_n\}$ będzie bazą L. Przyjmijmy $T_1 \wedge T_2 = (T_1, T_2)e_1 \wedge \ldots \wedge e_n$, gdzie $T_1 \in \bigwedge^p(L)$, $T_2 \in \bigwedge^{n-p}(L)$. Jest jasne, że (T_1, T_2) jest dwuliniowym iloczynem skalarnym między $\bigwedge^p(L)$ a $\bigwedge^{n-p}(L)$. Zbudujmy w $\bigwedge^p(L)$ i $\bigwedge^{n-p}(L)$ bazy złożone z p-wektorów, odpowiednio $(n-p)$-wektorów prostych $\{e_i \wedge \ldots \wedge e_i\}$, $\{e_j \wedge \ldots \wedge e_j\}$, $1 \leq i_1 < \ldots << i_p \leq n$, $1 \leq j_1 < \ldots < j_{n-p} \leq n$. Utożsamimy $\bigwedge^p(L)$ z $\bigwedge^{n-p}(L)$ za pomocą odwzorowania liniowego, przyporządkowującego p-wektorowi $e_{i_1} \wedge \ldots \wedge e_{i_p}$ taki $(n-p)$-wektor $e_{j_1} \wedge \ldots \wedge e_{j_{n-p}}$, że $\{i_1, \ldots, i_p, j_1 \ldots j_{n-p}\} = \{1, \ldots, n\}$. Wówczas możemy traktować (T_1, T_2) jako iloczyn skalarny w $\bigwedge^p(L)$ z diagonalną macierzą Grama postaci $\operatorname{diag}(\pm 1, \ldots, \pm 1)$. Jest on niezdegenerowany i dlatego jego lewe jądro jest zerowe.

W ten sposób zakończyliśmy konstrukcję izomorfizmu kanonicznego

$$\bigwedge^p(L) \longrightarrow (\bigwedge^{n-p}(L))^* \otimes \bigwedge^n(L).$$

W przypadku $p = n-1$ otrzymujemy $\bigwedge^{n-1}(L) \longrightarrow L^* \otimes \bigwedge^n(L)$, co wyjaśnia istnienie izomorfizmu $P(\bigwedge^{n-1}(L)) \longrightarrow P(L^*)$ z p. 18: pomnożenie tensorowo przez jednowymiarową przestrzeń $\bigwedge^n(L)$ nie zmienia zbioru prostych.

W następnym paragrafie będziemy kontynuować badanie związku pomiędzy mnożeniem zewnętrznym a dwoistością, wprowadzając do rozważań algebrę zewnętrzną $\bigwedge(L^*)$.

§ 7. Formy zewnętrzne

1. Niech L będzie skończenie wymiarową przestrzenią liniową nad ciałem $\boldsymbol{K}$, a L^* — przestrzenią do niej dwoistą.

Elementy p-tej zewnętrznej potęgi $\bigwedge^p(L^*)$ nazywamy *p-formami zewnętrznymi na przestrzeni L*. W szczególności 1-formami są po prostu funkcjonały liniowe na przestrzeni L. Tę obserwację można uogólnić na przypadek dowolnego p na dwa nieco różne sposoby.

2. Twierdzenie. *Przestrzeń $\bigwedge^p(L^*)$ jest kanonicznie izomorficzna:*

a) *$(\bigwedge^p(L))^*$, tj. przestrzeni funkcjonałów liniowych na przestrzeni p-wektorów;*

b) *przestrzeni skośnie symetrycznych (alternujących) p-liniowych odwzorowań F: $L \times \cdots \times L \to \boldsymbol{K}$, tj. odwzorowań spełniających dla każdego $\sigma \in S_p$ warunek*

$$F(l_{\sigma(1)}, \ldots, l_{\sigma(p)}) = \varepsilon(\sigma) F(l_1, \ldots, l_p)$$

Dowód. Zgodnie z przyjętą przez nas definicją

$$\bigwedge^p(L^*) \subset L^* \otimes \cdots \otimes L^* = \mathcal{T}_0^p(L^*).$$

W paragrafie 2, p. 4 utożsamiliśmy $\mathcal{T}_0^p(L^*)$ z przestrzenią wszystkich p-liniowych funkcji na L^*. Za pomocą tego utożsamienia formy zewnętrzne można interpretować jako p-liniowe, skośnie symetryczne odwzorowania L. Istotnie, wystarczy to sprawdzić dla form prostych. Mamy w tym przypadku

$$(f_1 \wedge \ldots \wedge f_p)(l_1, \ldots, l_p) = A(f_1 \otimes \cdots \otimes f_p)(l_1, \ldots, l_p) = \\ = \frac{1}{p!} \sum_{\tau \in S_p} \varepsilon(\tau) f_{\tau(1)}(l_1) \ldots f_{\tau(p)}(l_p).$$

Zatem

$$(f_1 \wedge \ldots \wedge f_p)(l_{\sigma(1)}, \ldots, l_{\sigma(p)}) = \tfrac{1}{p!} \sum_{\tau \in S_p} \varepsilon(\tau) f_{\tau(1)}(l_{\sigma(1)}) \ldots f_{\tau(p)}(l_{\sigma(p)}) = \\ = \tfrac{1}{p!} \sum_{\tau \in S_p} \varepsilon(\tau\sigma) f_{\tau\sigma(1)}(l_{\sigma(1)}) \ldots f_{\tau\sigma(p)}(l_{\sigma(p)}) = \\ = \varepsilon(\sigma)(f_1 \wedge \ldots \wedge f_p)(l_1, \ldots, l_p).$$

W ten sposób skonstruowaliśmy włożenie liniowe $\bigwedge^p(L^*) \longrightarrow$ (skośnie symetryczne p-liniowe formy na L). Dla wykazania, że jest to izomorfizm, wystarczy sprawdzić, iż wymiar przestrzeni po prawej stronie jest równy $\dim \bigwedge^p(L^*) = \binom{n}{p}$. Wybierając w L bazę $\{e_1, \ldots, e_n\}$ widzimy, że dowolna skośnie symetryczna forma p-liniowa F na L jest jednoznacznie wyznaczone przez swoje wartości $F(e_{i_1}, \ldots, e_{i_p})$, $1 \leq i_1 < \ldots < i_p \leq n$, które można wybierać dowolnie. Zatem wymiar przestrzeni takich form jest rzeczywiście równy $\binom{n}{p}$. W ten sposób wykazaliśmy część b) tezy twierdzenia.

Dla wykazania części a) utożsamimy $L^* \otimes \cdots \otimes L^*$ z $(L \otimes \cdots \otimes L)^*$, używając raz jeszcze konstrukcji § 2, p. 4, a następnie ograniczymy każdy element $\bigwedge^p(L^*)$ (traktowany jako funkcja liniowa na $L \otimes \cdots \otimes L$) na podprzestrzeń p-wektorów $\bigwedge^p(L)$. Otrzymamy w ten sposób odwzorowanie $\bigwedge^p(L^*) \to (\bigwedge^p(L))^*$, a ponieważ wymiary obu przestrzeni są jednakowe, wystarczy sprawdzić jego surjektywność. Przestrzeń funkcjonałów liniowych na $\bigwedge^p(L)$ jest rozpięta przez funkcjonały postaci $\frac{1}{p!}\delta_I$, gdzie $I = \{i_1, \ldots, i_p\} \subset \{1, \ldots, n\}$ oraz

$$\delta_I(e_{i_1} \wedge \ldots \wedge e_{i_p}) = 1,$$
$$\delta_I(e_{j_1} \wedge \ldots \wedge e_{j_p}) = 0, \quad \text{jeśli } \{j_1, \ldots, j_p\} \neq I.$$

Można się przekonać, że taki funkcjonał jest obrazem p-formy $e^{i_1} \wedge \ldots \wedge e^{i_p} \in \bigwedge^p(L^*)$, gdzie jak zazwyczaj przez $\{e^i\}$ oznaczyliśmy bazę dwoistą do $\{e_i\}$. Istotnie, wartość

$e^{i_1} \wedge \ldots \wedge e^{i_p}$ na p-wektorze $e_{j_1} \wedge \ldots \wedge e_{j_p}$ wynosi

$$A(e^{i_1} \otimes \cdots \otimes e^{i_p})(A(e_{j_1} \otimes \cdots \otimes e_{j_p})) =$$

$$= \frac{1}{(p!)^2} \sum_{\sigma,\tau \in S_p} \varepsilon(\sigma\tau) e^{i_{\sigma}(1)}(e_{j_{\tau}(1)}) \ldots e^{i_{\sigma}(1)}(e_{j_{\tau}(p)}).$$

Pośród składników stojących po prawej stronie różne od zera są te tylko, dla których $i_{\sigma(1)} = j_{\tau(1)}, \ldots, i_{\sigma(p)} = j_{\tau(p)}$, a więc cała suma jest równa zeru, jeśli $\{i_1, \ldots, i_p\} \neq \{j_1, \ldots, j_p\}$. Jeśli natomiast te zbiory są jednakowe, to wartość sumy wynosi

$$\frac{1}{(p!)^2} \sum_{\sigma \in S_p} \varepsilon(\sigma)^2 = \frac{1}{p!},$$

gdy zbiory $\{i_1, \ldots, i_p\} = \{j_1, \ldots, j_p\}$ są uporządkowane rosnąco. Dlatego $e^{i_1} \wedge \ldots \wedge e^{i_p}$ traktowany jako funkcjonał na $\bigwedge^p(L)$ jest równy $\frac{1}{p!}\delta_I$, co kończy dowód.

3. Uwagi. a) Przyjęty przez nas sposób utożsamienia $\bigwedge^p(L^*)$ z $(\bigwedge^p(L))^*$ odpowiada odwzorowaniu dwuliniowemu

$$\bigwedge^p(L^*) \times \bigwedge^p(L) \longrightarrow \boldsymbol{K},$$

które na dowolnych parach p-wektorów prostych zadane jest wzorem

$$(f^1 \wedge \ldots \wedge f^p, l_1 \wedge \ldots \wedge l_p) = \frac{1}{p!} \det(f^i(l_j)),$$

gdzie $f^i \in L^*$, $l_j \in L$. W istocie obie strony są wieloliniowe i skośnie symetryczne oddzielnie względem f^i i l_j, a ponadto przyjmują jednakowe wartości dla

$$(f^1, \ldots, f^p) = (e^{i_1}, \ldots, e^{i_p}), \quad (l_1, \ldots, l_p) = (e_{j_1}, \ldots, e_{j_p}),$$

co było sprawdzone w trakcie poprzedniego dowodu.

Czasami w definicji tego iloczynu skalarnego pomija się czynnik $\frac{1}{p!}$.

b) Nieraz przyjmuje się jedno z utożsamień ustanowionych w twierdzeniu w charakterze *definicji* potęgi zewnętrznej. Na przykład dosyć często, a zwłaszcza w geometrii różniczkowej wprowadza się $\bigwedge^p(L)$ jako przestrzeń skośnie symetrycznych p-liniowych funkcjonałów na L^*. Ta konstrukcja pod względem ogólności zajmuje pośrednią pozycję w stosunku do naszego pierwszego i drugiego określenia potęgi zewnętrznej z § 6. Ponieważ nie korzysta się w niej z dzielenia przez silnię, nadaje się do stosowania w przypadku przestrzeni liniowych nad ciałami skończonej charakterystyki, a także dla modułów wolnych nad pierścieniami przemiennymi. Jednakże przy zastosowaniu do przypadku dowolnych modułów przeszkadza dodatkowa dualizacja, i druga definicja jest korzystniejsza.

Analogiczny wynik do twierdzenia p. 2 jest prawdziwy także dla potęg symetrycznych, a nasze poprzedzające uwagi również się do nich stosują. W szczególności, można określić $S^p(L)$ jako przestrzeń p-liniowych symetrycznych funkcjonałów na L^* dla przestrzeni nad dowolnymi ciałami i modułów wolnych nad pierścieniami przemiennymi. W ogólnym przypadku bardziej prawidłowe podejście do określenia $S^p(L)$ prowadzi przez definicję z § 5, p. 9.

4. Iloczyn wewnętrzny. *Iloczynem wewnętrznym* nazywamy odwzorowanie dwuliniowe

$$L \times \bigwedge^p(L^*) \longrightarrow \bigwedge^{p-1}(L^*) : (l, F) \mapsto i(l)F,$$

określone w następujący sposób. Interpretujemy $F \in \bigwedge^p(L^*)$ jako skośnie symetryczną p-liniową formę na L oraz analogicznie $i(l)F$, i przyjmujemy

$$i(l)F(l_1, \ldots, l_{p-1}) = F(l, l_1, \ldots, l_{p-1}).$$

Jest jasne, że prawa strona jest $(p-1)$-liniową, skośnie symetryczną funkcją $l_1, \ldots, l_{p-1}$ oraz dwuliniową funkcją F, l, a więc definicja jest poprawna. Przy $p = 0$ wygodnie przyjąć, że $i(l)F = 0$. Często oznacza się $i(l)F$ także przez $l \rfloor F$.

§ 8. Pola tensorowe

1. W tym paragrafie opiszemy pokrótce typowe zastosowania algebry tensorowej w geometrii różniczkowej.

Rozważmy obszar $U \subset \boldsymbol{R}^n$ rzeczywistej przestrzeni kartezjańskiej i pierścień C nieskończenie różniczkowalnych funkcji rzeczywistych na U. W szczególności funkcje współrzędnych x^i, $i = 1, \ldots, n$, należą do C.

2. Definicja. *Wektorem stycznym* X_a do U w punkcie $a \in U$ nazywamy odwzorowanie liniowe $X_a : C \to \boldsymbol{R}$ spełniające warunki:

$X_a f = 0$, jeśli f jest stała w jakimś otoczeniu a;

$$X_a(fg) = X_a f \cdot g(a) + f(a) \cdot X_a g.$$

Jeśli X_a, Y_a są dwoma wektorami stycznymi w punkcie a, to dowolna rzeczywista kombinacja liniowa tych wektorów jest także wektorem stycznym w tym punkcie:

$$\begin{aligned}(cX_a + dY_a)(fg) &= cX_a(fg) + dY_a(fg) = \\ &= cX_a f \cdot g(a) + cf(a) \cdot X_a g + dY_a f \cdot g(a) + df(a) \cdot Y_a g = \\ &= (cX_a + dY_a)f \cdot g(a) + f(a) \cdot (cX_a + dY_a)g.\end{aligned}$$

Tak więc wektory styczne tworzą przestrzeń liniową, oznaczaną T_a i nazywaną przestrzenią styczną do U w punkcie a. Wartość $X_a f$ nazywamy pochodną f w kierunku wektora X_a. Można dowieść, że przestrzeń T_a ma wymiar n.

3. Definicja. *Polem wektorowym* w obszarze U nazywamy taką rodzinę wektorów stycznych $X = \{X_a \in T_a \mid a \in U\}$, że dla dowolnej funkcji $f \in C$ określona na U funkcja

$$a \mapsto X_a f$$

jest elementem C.

Oznaczymy tę funkcję przez Xf. Jest oczywiste, że każde pole wektorowe wyznacza odwzorowanie liniowe $X\colon C \to C$, zerujące się na zbiorze funkcji stałych i spełniające([1]) $X(fg) = Xf \cdot g + f \cdot Xg$ dla każdych $f, g \in C$. Takie odwzorowania nazywają się *różniczkowaniami pierścienia* C.

Odwrotnie, każdemu różniczkowaniu $X\colon C \to C$ i punktowi $a \in U$ odpowiada wektor styczny X_a w tym punkcie: $X_a f = (Xf)(a)$. W ten sposób powstaje bijekcja różniczkowań pierścienia C i pól wektorowych na U.

Suma pól wektorowych $X + Y$, określona żądaniem, aby $(X + Y)f = Xf + Yf$ dla każdego $f \in C$, jest znowu polem wektorowym. Iloczyn fX, określony wzorem $(fX)g = f(Xg)$, gdzie X jest polem wektorowym, a $f, g \in C$, jest polem wektorowym. W szczególności każda kombinacja liniowa pól wektorowych, $\sum_{i=1}^{m} f^i X_i$, $f^i \in C$, jest polem wektorowym.

4. Przykład. Niech $\frac{\partial}{\partial x^i}$, $i = 1, \ldots, n$, oznaczają klasyczne operatory pochodnych cząstkowych. Każdy z nich jest polem wektorowym na U. Następujący podstawowy rezultat podajemy bez dowodu.

5. Twierdzenie. *Każde pole wektorowe X w spójnym obszarze $U \subset \boldsymbol{R}^n$ ma jednoznaczną reprezentację w postaci kombinacji $\sum_{i=1}^{n} f^i \frac{\partial}{\partial x^i}$, gdzie $x^1, \ldots, x^n$ są funkcjami współrzędnych na $\boldsymbol{R}^n$.*

6. W języku algebraicznym ta własność oznacza, że zbiór wszystkich pól wektorowych T w obszarze spójnym U jest modułem wolnym rzędu n nad przemiennym pierścieniem łącznym C nieskończenie różniczkowalnych funkcji na U.

Moduły wolne skończonego rzędu nad przemiennymi pierścieniami tworzą kategorię o niezwykle bliskich własnościach do kategorii skończenie wymiarowych przestrzeni liniowych nad ciałem. W szczególności, w niej są prawdziwe wszystkie wyniki teorii dwoistości i wszystkie konstrukcje algebry tensorowej z tej części wykładu.

Inny wariant, nie korzystający z przeniesienia algebry tensorowej na pierścienie i moduły, ale wymagający w zamian rozwinięcia pewnej techniki geometrycznej, opiera się na rozważaniu dowolnego pola wektorowego X jako rodziny wektorów $\{X_a \mid a \in U\}$, należących do rodziny skończenie wymiarowych przestrzeni $\{T_a\}$. Wtedy wszystkie potrzebne nam operacje algebry tensorowej można budować „punktowo", określając na przykład $X \otimes Y$ jako $\{X_a \otimes Y_a \mid a \in U\}$.

([1]) Tę tożsamość nazywa się często *regułą Leibniza (przyp. tłum.)*.

Oba warianty konstrukcji algebry tensorowej są w istocie w pełni równoważne — w poniższych definicjach będziemy stosować pierwszy z nich.

7. Oznaczmy przez T^* C-moduł C-liniowych odwzorowań

$$T^* = \mathcal{L}_C(T, C).$$

Składa się on z odwzorowań $\omega: T \to C$ o własności

$$\omega\left(\sum_{i=1}^{m} f^i X_i\right) = \sum_{i=1}^{m} f^i \omega(X_i)$$

dla każdego $X_i \in T$, $f^i \in C$. Dodawanie i mnożenie przez elementy C określa się standardowymi wzorami. C-moduł T^* często bywa oznaczany przez Ω^1 albo $\Omega^1(U)$ i nazywany jest modułem (różniczkowych) 1-*form* w obszarze U.

Każda funkcja $f \in C$ wyznacza element $df \in \Omega^1$ za pomocą wzoru

$$(df)(X) = Xf, \quad X \in T.$$

Nazywa się go *różniczką funkcji* f. W szczególności można skonstruować różniczki funkcji współrzędnych $dx^1, \ldots, dx^n \in \Omega^1$. Z twierdzenia p. 5 wynika bez trudności

8. Stwierdzenie. *Każda* 1-*forma* $\omega \in \Omega^1$ *ma jednoznaczną reprezentację w postaci kombinacji liniowej* $\sum_{i=1}^{n} f_i dx^i$.

9. Elementy iloczynu tensorowego C-modułów $T^* \otimes \ldots \otimes T^* \otimes T \otimes \ldots \otimes T$ nazywane są *polami tensorowymi* typu (p, q) albo p razy kowariantnymi i q razy kontrawariantnymi polami tensorowymi w obszarze U. W geometri różniczkowej najczęściej opuszcza się słowo „pole" i mówi się o polach tensorowych po prostu „tensory".

Z twierdzenia p. 5 i stwierdzenia p. 8 wynika, że każdy tensor typu (p, q) jest wyznaczony jednoznacznie przez swoje współrzędne $T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$ za pomocą wzoru

$$T = \sum T^{j_1 \ldots j_q}_{i_1 \ldots i_p} dx^{i_1} \otimes \cdots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_q}},$$

gdzie i_k, j_l przebiegają niezależnie od siebie wartości od 1 do n. W klasycznych oznaczeniach wszystkie symbole poza $T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$ są opuszczane, a sam ten znak jest używany jako oznaczenie tensora. Podkreślmy raz jeszcze, że tutaj $T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$ nie są liczbami, ale nieskończenie różniczkowalnymi funkcjami na U.

10. Zamiana współrzędnych. Pierwsze zastosowanie analizy do badania pól tensorowych polega na umożliwieniu dokonywania ogólniejszych niż liniowe zamian zmiennych na U. Możemy przechodzić od zmiennych $x^1, \ldots, x^n$ do $y^1, \ldots, y^n$, gdzie $y^i = y^i(x^1, \ldots, x^n)$ są nieskończenie różniczkowalnymi funkcjami mającymi odwrotne

$x^i = x^i(y^1, \ldots, y^n)$, które też są nieskończenie różniczkowalne. Rzecz w tym, że współrzędne pól wektorowych i form różniczkowych także przy takiej zamianie przekształcają się liniowo, tyle że macierze zamiany zmiennych zależą od punktu: zgodnie ze wzorem na pochodną funkcji złożonej

$$\frac{\partial}{\partial x^j} = \frac{\partial y^k}{\partial x^j}\frac{\partial}{\partial y^k}$$

(po prawej *implicite* występuje sumowanie względem k — tzw. konwencja sumacyjna), a także

$$dx^j = \frac{\partial x^j}{\partial y^k} dy^k$$

(również z uwzględnieniem sumowania). Zatem współczynniki tensora $T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$ w nowych współrzędnych wyrażają się wzorami

$$(T')^{j'_1 \ldots j'_q}_{i'_1 \ldots i'_p} = \frac{\partial x^{i_1}}{\partial y^{i'_1}} \cdots \frac{\partial x^{i_p}}{\partial y^{i'_p}} \frac{\partial y^{j'_1}}{\partial x^{j_1}} \cdots \frac{\partial y^{j'_q}}{\partial x^{j_q}} T^{j_1 \ldots j_q}_{i_1 \ldots i_p}$$

(z zastosowaniem tej samej konwencji sumacyjnej).

Wszystkie konstrukcje algebraiczne i ustalenia terminologiczne można teraz stosować do pól tensorowych.

Na zakończenie tego paragrafu podamy kilka przykładów pól tensorowych, odgrywających szczególnie ważną rolę w geometrii i fizyce.

11. Tensor metryczny. Ten tensor, oznaczany przez g_{ij} lub w pełniejszym zapisie przez $\sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j$, jest z definicji symetryczny i niezdegenerowany w każdym punkcie $a \in U$, tj. $\det(g_{ij}(a)) \neq 0$. Zadaje on strukturę ortogonalną w każdej przestrzeni stycznej T_a, a pary (U, g_{ij}), jak również uogólnienia na przypadek rozmaitości otrzymanych przez „sklejenie" wielu obszarów U stanowią podstawowy obiekt badania w geometrii riemannowskiej, a w przypadku $n = 4$ i metryki o sygnaturze $(1,3)$ — w ogólnej teorii względności.

Metryki używa się do mierzenia długości krzywych różniczkowalnych $\{x^1(t), \ldots, x^n(t) \mid t_0 \leq t \leq t_1\}$, których długość określa się wzorem $\int_{t_0}^{t_1} \sqrt{g_{ij}(x^k(t)) \frac{dx^i}{dt}\frac{dx^j}{dt}}\, dt$, a także dla podnoszenia i opuszczania indeksów pól tensorowych.

12. Formy zewnętrzne i forma objętości. Elementy przestrzeni $\bigwedge^p(\Omega^1)$, tj. skośnie symetryczne tensory typu $(p,0)$, nazywane są p-formami zewnętrznymi na U, a n-formy zewnętrzne nazywane są formami objętości. Ta nazwa jest odzwierciedleniem możliwości określenia „całek krzywoliniowych" $\int_V f(x^1, \ldots x^n)\, dx^1 \wedge \ldots \wedge dx^n$ względem dowolnego obszaru $V \subset U$, mających własności miary. W przypadku

$f = 1$ wartością takiej całki jest objętość euklidesowa obszaru V, której własności opisaliśmy w § 5 części 2.

Dla $p < n$ można zdefiniować całkę dowolnej formy $\omega \in \bigwedge^p(\Omega^1)$ po „gładkich p-wymiarowych hiperpowierzchniach" w U. Moduły form zewnętrznych wszystkich stopni są powiązane ważnym operatorem „różniczkowania zewnętrznego" $d^p: \bigwedge^p \to \bigwedge^{p+1}$, który we współrzędnych jest dany wzorem

$$d^p\left(\sum f_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}\right) = \sum \frac{\partial f_{i_1 \ldots i_p}}{\partial x^{i_{p+1}}} dx^{i_{p+1}} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Te operatory spełniają warunek $d^{p+1} \circ d^p = 0$ i pojawiają się w sformułowaniu uogólnionego twierdzenia Stokesa, wiążącego całkę po p-wymiarowej hiperpowierzchni z brzegiem V^p z całką po jej brzegu ∂V^p:

$$\int_{V^p} d\omega^{p-1} = \int_{\partial V^p} \omega^{p-1}.$$

Szczególną rolę odgrywają dwuformy zewnętrzne ω^2, spełniające warunek $d\omega^2 = 0$, dzięki którym można w niezmienniczy sposób sformułować formalizm hamiltonowski.

§ 9. Iloczyny tensorowe w mechanice kwantowej

1. Układy złożone. Rolę odgrywaną w mechanice kwantowej przez iloczyny tensorowe przybliża następujące podstawowe założenie, będące kontynuacją serii postulatów sformułowanych w § 6, p. 8 i § 9, p. p. 1–6, część 2.

Niech $\mathcal{H}_1, \ldots, \mathcal{H}_n$ będą przestrzeniami stanów pewnych układów kwantowych. Wówczas przestrzeń stanów układu otrzymanego przez połączenie w jeden tych układów jest pewną podprzestrzenią $\mathcal{H} \subset \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$.

Ściśle mówiąc, w nieskończenie wymiarowym przypadku zamiast iloczynu tensorowego po prawej stronie powinno być uzupełnienie iloczynu tensorowego przestrzeni hilbertowskich, ale my, jak zazwyczaj, będziemy pomijać takie subtelności, zajmując się skończenie wymiarowymi przestrzeniami.

To, jaka konkretnie podprzestrzeń $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ odpowiada układom połączonym, musi być rozstrzygnięte na podstawie dalszych reguł, którymi zajmiemy się poniżej. Tutaj, na podstawie przypadku $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ spróbujemy pokazać, że już pierwszy postulat mechaniki kwantowej — zasada superpozycji — prowadzi do powstania istotnie różnych niż klasyczne związków między układami. W tym celu rozważmy, w jakich stanach może się znaleźć układ złożony. Jeśli $\psi_i \in \mathcal{H}_i$ oznaczają stany podukładów, to tensor prosty $\psi_1 \otimes \cdots \otimes \psi_n$ jest jednym z możliwych stanów układu i można uważać, że odpowiada on sytuacji, w której każdy z podukładów znajduje się w stanie ψ_i. Ale takie rozkładalne stany nie stanowią wszystkich wektorów w $\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$, gdzie występują także dowolne ich kombinacje liniowe. Jednak

jeśli taki układ złożony znajduje się w jednym z takich stanów nierozkładalnych, traci sens rozważanie jego podukładów, gdyż ani te podukłady, ani ich stany nie mogą być jednoznacznie wyróżnione. Mówiąc inaczej, w przygniatającej większości stanów układu złożonego takie podukłady istnieją tylko w sposób wirtualny.

Trzeba podkreślić, że to rozumowanie w najmniejszym stopniu nie wykorzystuje klasycznego pojęcia oddziaływania podukładów, opartego na wymianie energii między nimi. Einstein, Rosen i Podolsky zaproponowali eksperyment myślowy, w którym po rozpadzie układu złożonego tworzące go dwa poukłady są oddzielone w przestrzeni, a obserwacja jednego z tych podukładów pozwala na momentalne „przeprowadzenie do określonego stanu" drugiego z nich, choć klasyczne oddziaływanie potrzebowałoby na to skończonego czasu. Ten wniosek z postulatów superpozycji i iloczynu tensorowego jest w ostrej sprzeczności z intuicją klasyczną. Niemniej jednak przyjęcie tych postulatów doprowadziło do ogromnej ilości teoretycznych modeli, które prawidłowo wyjaśniają rzeczywistość, tak że nie pozostaje nic innego, jak rozwijanie opartych na nich, nowych intuicji.

Zauważmy jeszcze, że opisanie oddziaływania wymaga wprowadzenia hamiltonianu układu złożonego. W najprostszym przypadku ma on postać „swobodną"

$$H_1 \otimes \mathrm{id} \otimes \cdots \otimes \mathrm{id} + \mathrm{id} \otimes H_2 \otimes \cdots \otimes \mathrm{id} + \cdots + \mathrm{id} \otimes \cdots \otimes H_n,$$

gdzie $H_i \colon \mathcal{H}_i \to \mathcal{H}_i$ oznacza hamiltonian i-tego układu, a id — odwzorowanie tożsamościowe. W takim przypadku mówimy, że układy nie oddziałują ze sobą. Pewnym wyjaśnieniem tego jest obserwacja, że jeśli układ złożony ma taki właśnie hamiltonian i w chwili początkowej znajdował się w stanie zadanym przez wektor prosty $\psi_1 \otimes \cdots \otimes \psi_n$, to w dowolnej chwili t będzie znajdował się także w stanie odpowiadającym wektorowi prostemu $e^{-itH_1}\psi_1 \otimes \cdots \otimes e^{-itH_n}\psi_n$, tj. ewolucja tego układu przebiega tak, jak gdyby składające się nań podukłady ewoluowały niezależnie od siebie. W ogólnym przypadku hamiltonian jest sumą części swobodnej i operatora wyrażającego oddziaływanie.

2. Nierozróżnialność. Wyróżniamy dwa podstawowe przypadki, gdy przestrzeń stanów układu złożonego nie jest równa całej przestrzeni $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. W obu tych przypadkach łączone ze sobą układy są identyczne, jak to się mówi — nierozróżnialne między sobą — i mogą być, np. cząstkami elementarnymi jednego typu; w szczególności $\mathcal{H}_1 = \cdots = \mathcal{H}_n = \mathcal{H}$.

a) *Bozony.* Układ, którego przestrzenią stanów jest $\mathcal{H}$ nazywamy *bozonem*, jeśli przestrzenią stanów układu złożonego z n takich układów jest n-ta potęga symetryczna $S^n(\mathcal{H})$.

Na podstawie eksperymentu do bozonów zaliczamy fotony i cząstki alfa (jądra helu).

b) *Fermiony.* Układ, którego przestrzenią stanów jest $\mathcal{H}$ nazywamy *fermionem*, jeśli przestrzenią stanów układu złożonego z n takich układów jest n-ta potęga zewnętrzna $\bigwedge^n(\mathcal{H})$.

Zgodnie z eksperymentem do fermionów zaliczamy elektrony, protony i neutrony.

3. Liczby obsadzenia i zasada Pauliego. Niech $\{\psi_1, \ldots, \psi_m\}$ będzie bazą przestrzeni stanów układu bozonowego czy fermionowego. Wówczas elementy symetryzowanej (lub antysymetryzowanej) bazy tensorowej $S^n(\mathcal{H})$ (lub $\bigwedge^n(\mathcal{H})$) fizycy zapisują w postaci

$$|a_1, \ldots, a_m\rangle = \begin{cases} \underbrace{S(\psi_1 \otimes \cdots \otimes \psi_1}_{a_1} \otimes \cdots \otimes \underbrace{\psi_m \otimes \cdots \otimes \psi_m)}_{a_m} & \text{w } S^n(\mathcal{H}), \\ A\underbrace{(\psi_1 \otimes \cdots \otimes \psi_1}_{a_1} \otimes \cdots \otimes \underbrace{\psi_m \otimes \cdots \otimes \psi_m)}_{a_m} & \text{w } \bigwedge^n(\mathcal{H}). \end{cases}$$

W obu przypadkach $a_1 + \cdots + a_m = n$, ale dla bozonów a_i mogą przyjmować dowolne nieujemne całkowitoliczbowe wartości, podczas gdy dla fermionów tylko 0 lub 1, bo w przeciwnym przypadku odpowiednie antysymetryzacje zerują się i nie określają stanu kwantowego.

Liczby a_i nazywają się „liczbami obsadzeń" odpowiadającego stanu. Uważa się, że można umownie przyjąć, że w stanie $|a_1, \ldots, a_m\rangle$ układu złożonego, a_i podukładów znajduje się w stanie ψ_i. Jednakże ani w przypadku fermionowym, ani w bozonowym układ złożony nie może znajdować się w stanie opisywanym przez tensor prosty $\psi_1 \otimes \cdots \otimes \psi_m$, poza przypadkiem (dla bozonów), gdy wszystkie ψ_i są jednakowe. Stąd wnioskujemy, że nawet w stanach bazowych $|a_1, \ldots, a_m\rangle$ nie można powiedzieć, który z podukładów znajduje się, na przykład, w stanie ψ_i. Podukłady są nierozróżnialne.

Warunek $a_i = 0$ lub 1 w przypadku fermionowym interpretujemy jako stwierdzenie, że dwa podukłady nie mogą znaleźć się w jednakowym stanie. To jest właśnie sławną „regułą zakazu" Pauliego.

Gdy n jest bardzo duże, szereg fizycznie ważnych wniosków o przestrzeniach $S^n(\mathcal{H})$ i $\bigwedge^n(\mathcal{H})$ formułuje się w terminach teorii prawdopodobieństwa, na przykład, poprzez określenie udziału stanów $|a_1, \ldots, a_m\rangle$ o takich czy innych warunkach na liczby obsadzeń. Dlatego często mówi się, że bozony i fermiony podlegają różnej *statystyce* — odpowiednio Bosego–Einsteina i Fermiego.

4. Przypadek zmiennej liczby cząstek. W trakcie ewolucji układu kwantowego składające się nań „elementarne podukłady" albo cząstki mogą pojawiać się lub znikać. Dla opisania takich efektów, używa się odpowiednio w przypadkach bozonowym i fermionowym przestrzeni stanów $\bigoplus_{i=0}^{\infty} S^i(\mathcal{H})$ (a dokładniej uzupełnienia tej przestrzeni) oraz $\bigoplus_{i=0}^{\infty} \bigwedge^i(\mathcal{H})$, tj. pełnej algebry symetrycznej lub zewnętrznej przestrzeni jednocząstkowej $\mathcal{H}$.

Operator, mnożący wektory z $S^n(\mathcal{H})$ (odpowiednio, z $\bigwedge^n(\mathcal{H})$) przez n, gdzie $n = 0, 1, 2, 3, \ldots$, nazywa się *operatorem liczby cząstek*. Jego jądro, będące odpowiednio podprzestrzenią $C = S^0(\mathcal{H})$ lub $\bigwedge^0(\mathcal{H})$, nazywa się *stanem próżni* — w tym stanie nie ma cząstek.

Zasadniczą rolę odgrywają tu także specjalne operatory *kreacji i anihilacji cząstek*. Operator $a^-(\psi_0)$ anihilacji bozonu w stanie $\psi_0 \in \mathcal{H}$ działa na stan reprezento-

wany przez wektor $S(\psi_1 \otimes \cdots \otimes \psi_n)$ zgodnie ze wzorem

$$a^-(\psi_0)S(\psi_1 \otimes \cdots \otimes \psi_n) = \sqrt{n+1} \cdot \frac{1}{n!} \sum_{\sigma \in S_n} (\psi_0, \psi_{\sigma(1)}) \otimes \psi_{\sigma(2)} \otimes \cdots \otimes \psi_{\sigma(n)},$$

gdzie $(\psi_0, \psi_{\sigma(1)})$ jest iloczynem skalarnym w $\mathcal{H}$. Operator $a^+(\psi_0)$ kreacji bozonu w stanie ψ_0 jest określony jako operator sprzężony z $a^-(\psi_0)$ w sensie geometrii hermitowskiej. Analogiczne wzory można podać także w przypadku fermionowym. Znaczenie tych standardowych operatorów algebry tensorowej pochodzi stąd, że za ich pomocą można w wygodny sposób zapisać operatory ważnych obserwabli, przede wszystkim hamiltonianu.

Ćwiczenia

W następującej serii ćwiczeń przedstawione są podstawowe fakty teorii rzędu tensorowego, ważnej dla oszacowań złożoności obliczeniowej. Podstawy tej teorii sformułował F. Strassen.

1. Niech $L_1, \ldots, L_n$ będą skończenie wymiarowymi przestrzeniami nad ciałem $\boldsymbol{K}$, $t \in L_1 \otimes \cdots \otimes L_n$, $t \neq 0$. Rzędem $\operatorname{rk} t$ tensora t nazywa się taką najmniejszą liczbę r, że dla odpowiednich wektorów $l_i^{(j)} \in L_i$, $j = 1, \ldots, r$,

$$t = \sum_{j=1}^{r} l_1^{(j)} \otimes \cdots \otimes l^{(j)}.$$

Oczywiście, dla $n = 1$ mamy $\operatorname{rk} t = 1$ dla dowolnego $t \neq 0$.

Niech $t \in L_1^* \otimes L_2 = \mathcal{L}(L_1, L_2)$ (por. § 2, p. 5). Dowieść, że

$$\operatorname{rk} t = \dim \operatorname{Im} t, \quad t: L_1 \to L_2.$$

Wyprowadzić stąd następujące fakty:

a) dla $n = 2$ rząd t nie zmienia się przy rozszerzeniu ciała skalarów;

b) dla $n = 2$ zbiór $\{t \mid \operatorname{rk} t \leq r\}$ jest wyznaczony poprzez skończony układ równań $P_{j,r}(t^{i_1 \cdots i_n})$, gdzie $P_{j,r}$ są wielomianami względem współrzędnych.

Obydwa te fakty nie są prawdziwe dla przypadku $n = 3$, będącego zasadniczym punktem zainteresowania teorii złożoności obliczeniowej; zob. ćwiczenia 4–9 poniżej.

2. Niech $L = \bigoplus_{i,j=1}^{2} \boldsymbol{C} a_{ij}$ będzie przestrzenią zespolonych macierzy kwadratowych stopnia 2. Dowieść, że

$$\operatorname{rk}\Big(\sum_{i,j,k=1}^{2} a_{ij} \otimes a_{jk} \otimes a_{ki} \Big) = 7.$$

Wskazówka. Skorzystać z ćwiczenia 12 do § 4 cz. 1. Ta sama wskazówka dotyczy następnego ćwiczenia.

3. Dowieść, że dla pewnej stałej c

$$\operatorname{rk}\Big(\sum_{i,j,k=1}^{N} a_{ij} \otimes a_{jk} \otimes a_{ki} \Big) \leq c N^{\log_2 7}.$$

(Tutaj $L = \bigoplus_{i,j,k=1}^{N} Ca_{ij}$, a rozważany przez nas tensor to $\operatorname{Tr} A \otimes A \otimes A$, $A = (a_{ij})$ jest dowolną macierzą stopnia N.)

4. Niech L będzie skończenie wymiarową $\boldsymbol{K}$-algebrą, $L \underset{K}{\otimes} L \to L : a \otimes b \mapsto ab$ mnożeniem w tej algebrze. Rozważmy je jako tensor $t \in L^* \otimes L \otimes L$ (współrzędnymi tego tensora są stałe strukturalne algebry). Obliczyć $\operatorname{rk} t$ dla przypadku $\boldsymbol{K} = \boldsymbol{R}$, $L = \boldsymbol{C}$.

5. W oznaczeniach poprzedniego ćwiczenia niech $L = \boldsymbol{K}^n$, a mnożenie wykonuje się po współrzędnych:

$$(a_1, \ldots, a_n)(b_1, \ldots, b_n) = (a_1 b_1, \ldots, a_n b_n).$$

Obliczyć $\operatorname{rk} t$.

6. Wykorzystując rezultaty zadań 4 i 5 wykazać, że rząd tensora stałych strukturalnych algebry $\boldsymbol{C}$ nad $\boldsymbol{R}$ zmniejsza się przy rozszerzeniu ciała skalarów do $\boldsymbol{C}$.

Wskazówka. $\boldsymbol{C} \underset{R}{\otimes} \boldsymbol{C}$ jest izomorficzna z $\boldsymbol{C}^2$ jako $\boldsymbol{C}$-algebra.

7. Niech $L = \boldsymbol{C}e_1 \oplus \boldsymbol{C}e_2$. Wykazać, że tensor

$$t = e_1 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_2 + e_2 \otimes e_1 \otimes e_2$$

ma rząd 3.

8. Wykazać, że tensor t z poprzedniego zadania jest granicą pewnego ciągu tensorów rzędu 2.

Wskazówka.

$$t + \varepsilon e_2 \otimes e_2 \otimes e_2 = \frac{1}{\varepsilon}[e_1 \otimes e_1 \otimes (-e_2 + \varepsilon e_1) + (e_1 + \varepsilon e_2) \otimes (e_1 + \varepsilon e_2) \otimes e_2].$$

9. Z zadań 7 i 8 wyprowadzić, że zbiór tensorów rzędu ≤ 2 w $L \otimes L \otimes L$ nie jest dany przez układ równań postaci

$$P_j(t^{i_1 i_2 i_3}) = 0,$$

gdzie P_j są wielomianami.

10. Rzędem granicznym $\operatorname{brk}(t)$ tensora t nazwiemy taką najmniejszą liczbę s, że t można przedstawić jako granicę ciągu tensorów rzędu $\leq s$. Dowieść, że dla ogólnej macierzy A stopnia 3

$$\operatorname{brk}(\operatorname{Tr} A \otimes A \otimes A) \leq 21.$$

11. Ile wynoszą

$$\operatorname{rk}(\operatorname{Tr} A \otimes A \otimes A), \quad \operatorname{brk}(\operatorname{Tr} A \otimes A \otimes A)$$

dla dowolnej macierzy A stopnia N? (W chwili pisania tych słów odpowiedź nie jest znana nawet dla $N = 3$.)

12. Niech L będzie n-wymiarową przestrzenią liniową nad ciałem $\boldsymbol{K}$, $M \subset \bigwedge^2(L)$ dowolną podprzestrzenią. Załóżmy, że dla każdego $v \in L$, $v \neq 0$, istnieje $w \in L$, takie że $0 \neq v \wedge w \in M$. Dowieść, że można znaleźć w L taką bazę $\{e_1, \ldots, e_n\}$, że

$$M + \bigwedge^2(L_i) = \bigwedge^2(L), \quad 1 \leq i \leq n,$$

gdzie $L_i = \bigoplus_{i \neq j} \boldsymbol{K} e_j$.

W przypadku ciała $\boldsymbol{K} = F_p$ o p elementach znany jest skomplikowany dowód kombinatoryczny tego faktu (zob. M. R. Vaughan-Lee — J. Algebra, **32**(1974), p. 278–285), dopuszczający teoriogrupową interpretację. Byłoby interesujące znaleźć prostsze podejście.

Skorowidz

Spis treści

Do nabycia w księgarniach

A.V. Balakrishnan
Analiza funkcjonalna stosowana

R. Bittner
Rachunek operatorów w grupach

J.B. Czerminski, A. Iwasiewicz, Z. Paszek, A. Sikorski
Metody statystyczne dla chemików

R. Frankiewicz, P. Zbierski
Granice i luki

H. Koczyk
Geometria wykreślna

J. Uchmański
Klasyczna ekologia matematyczna

Książki PWN są do nabycia w księgarni PWN Warszawa, ul. Miodowa 10 (poniedziałki–piątki, 9.00–16.00) oraz w księgarniach promocyjnych PWN na terenie całego kraju.