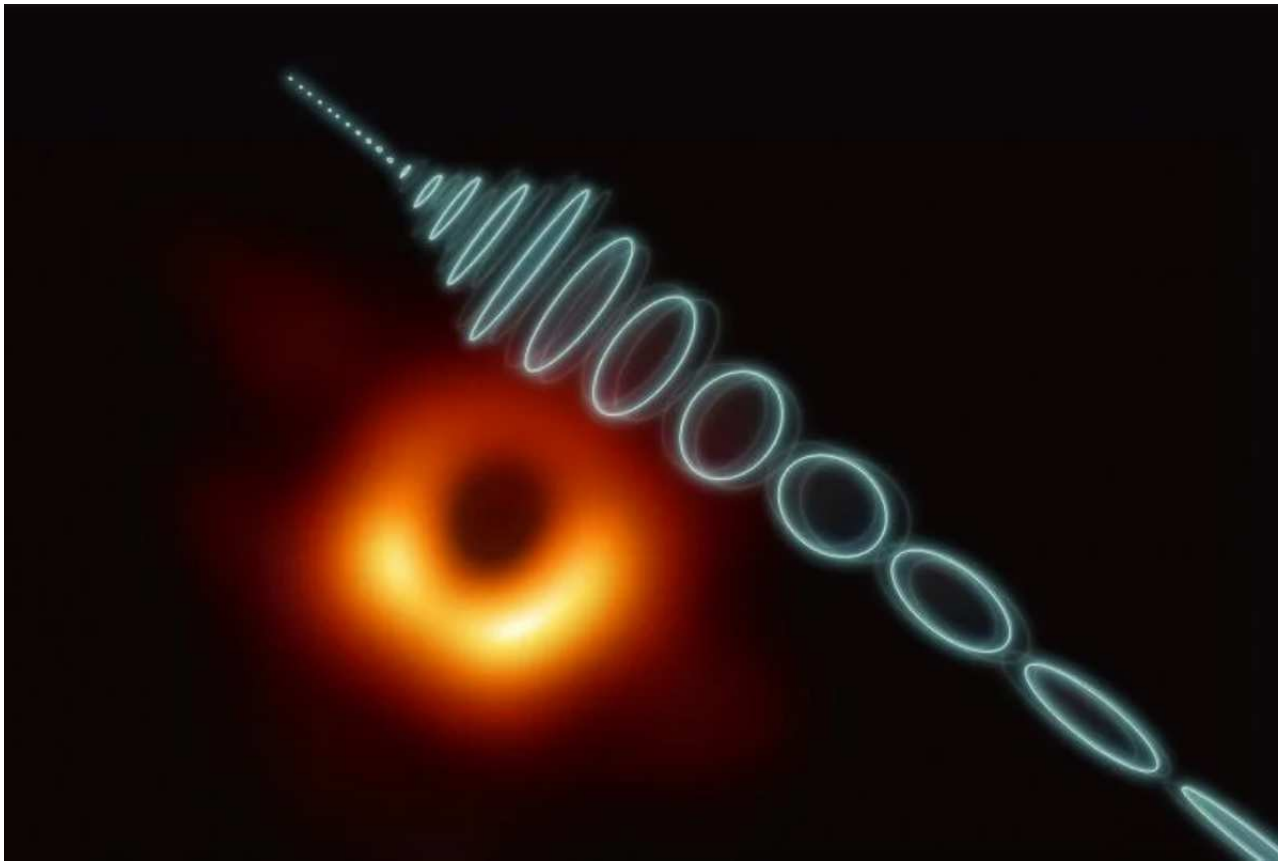
GEOMETRIA Z ALGEBRĄ LINIOWĄ II*

(notatki do wykładu i naprawdę długie opowieści o matematyce)

Arkadiusz Męcel



Wyobrażenie „struny" przelatującej niedaleko czarnej dziury. Jej deformacje można opisać w języku geometrii afinicznej.

https://mimuw.edu.pl/~amecel/alglin.html
ostatnia aktualizacja: 24 kwietnia 2024
Materiał realizowany według skryptów *Wykłady z Algebry Liniowej II* doc. dr. Tadeusza Koźniewskiego, rozszerzony na podstawie m.in. materiałów prof. A. Webera i prof. J. Wiśniewskiego.

Spis treści

Wprowadzenie do wykładu rozszerzonego z GAL II **4**

1 Endomorfizmy, macierze podobne. Wektory i wartości własne **5**
1.1 Uzupełnienie. Macierze endomorfizmu w różnych bazach 12
1.2 Dodatek. Dyskretny układ dynamiczny 15
1.3 Trivia. PCA, czyli w jakich kierunkach patrzeć na dane? 17
1.4 Coda. Wzbogacanie podstawowych struktur algebraicznych 18

2 Wielomian charakterystyczny. Diagonalizowalność **19**
2.1 Uzupełnienie. Probabilistyka na sześciennej kostce 28
2.2 Dodatek. Grafy i ich macierze sąsiedztwa 31
2.3 Trivia. Potęgi macierzy i rekurencje liniowe 34
2.4 Coda. Czym są reprezentacje? 35

3 Podprzestrzenie niezmiennicze. Twierdzenie Cayleya-Hamiltona **36**
3.1 Uzupełnienie. Wielomian minimalny kontra wyznaczniki 46
3.2 Dodatek. Macierze wielomianowe i podobieństwo 49
3.3 Trivia. Twierdzenie Cayleya-Hamiltona na konkursach 56

4 Endomorfizmy nilpotentne. Twierdzenie Jordana **58**
4.1 Uzupełnienie. Dowód przez endomorfizmy nierozkładalne 67
4.2 Dodatek. Dowód przez przestrzenie ilorazowe 70
4.3 Trivia. Dowód przez mnożenie macierzy 74

5 Podprzestrzenie pierwiastkowe. Dowód twierdzenia Jordana **78**
5.1 Uzupełnienie. Trudniejsze zadania o postaci i bazie Jordana 89
5.2 Trivia. Problem Collatza 92

6 Twierdzenie o rozkładzie prymarnym
Twierdzenie Jordana-Chevalleya **94**
6.1 Uzupełnienie. Rodziny endomorfizmów przemiennych
Wspólna diagonalizowalność i triangularyzowalność 100
6.2 Dodatek. Rozkład na podprzestrzenie cykliczne.
Postać kanoniczna Frobeniusa 104

7 Przestrzenie afiniczne w przestrzeniach liniowych **110**
7.1 Uzupełnienie. Wymiar sumy podprzestrzeni afinicznych 123
7.2 Dodatek. Współliniowość, współrzędne barycentryczne oraz intuicje i twierdzenia geometrii elementarnej 126
7.3 Trivia. SET, czyli przestrzeń afiniczna nad $\mathbb{Z}_3$ 131
7.4 Coda. Geometria wewnętrzna i zewnętrzna 133

8 Przekształcenia afiniczne **137**

9 Zbiory i kombinacje wypukłe **146**
9.1 Trivia. Spektrum macierzy stochastycznej 152

10 Wielościany, stożki i dualność **155**

11 Formy dwuliniowe i iloczyny skalarne **163**
11.1 Uzupełnienie. Ważne typy form dwuliniowych . . . 171
11.2 Dodatek. Iloczyn skalarny i funkcjonały . . . 172
11.3 Trivia. Dlaczego cosinus? . . . 173
11.4 Trivia. Zastosowania nierówności Cauchy-ego Schwarza . . . 174

12 Rzut prostopadły i ortogonalizacja
Przestrzenie euklidesowe afiniczne **175**
12.1 Uzupełnienie. Dlaczego ogólny wzór na odległość działa? . . . 185
12.2 Dodatek. Metoda najmniejszych kwadratów . . . 186
12.3 Trivia. Czy względna pierwszość to prostopadłość? . . . 188

13 Relacja kongruencji macierzy. Kryteria Sylvestera i Jacobiego **190**
13.1 Uzupełnienie. Dowód kryterium Sylvestera ze skryptu . . . 198
13.2 Dodatek. Grupa klas kwadratów ciała . . . 201
13.3 Trivia. Inne kątomierze, czyli o aksonometrii . . . 203

14 Macierz Grama układu wektorów. Miara, orientacja, iloczyn wektorowy **205**
14.1 Uzupełnienie. Orientowalność i transformacje ciągłe . . . 215
14.2 Dodatek. Miara Jordana i zasada Cavalieriego . . . 217
14.3 Trivia. Zastosowania macierzy Grama w zadaniach . . . 223

15 Izometrie liniowe i afiniczne. Macierze ortogonalne **227**
15.1 Uzupełnienie. Rozpoznawanie i zliczanie izometrii . . . 235

16 Składanie izometrii **238**
16.1 Uzupelnienie. Grupy izometrii i obrotów . . . 245
16.2 Dodatek. Punkty stałe i klasyfikowanie izometrii . . . 247

Wprowadzenie do wykładu rozszerzonego z GAL II

Wykład na potoku gwiazdkowym będzie realizował program równolegle z potokiem standardowym, do którego notatki (w wielu miejscach bardzo podobne lub identyczne) znaleźć można pod adresem `https://mimuw.edu.pl/~amecel/galII_mecel.pdf`. Poszczególne tematy omawiane będą w sposób pogłębiony – pojawią się dodatkowe definicje, większa będzie ilość dowodów, omawiane będą trudniejsze zadania. Pojawią się również dodatkowe tematy, ważne z punktu widzenia przyszłego udziału niektórych z Państwa w zaawansowanych kursach oferowanych przez Wydział MIM UW, np. geometrii algebraicznej, geometrii różniczkowej, teorii reprezentacji algebr czy analizy funkcjonalnej. Kurs ten jest więc po pierwsze dla osób, które myślą o tym, by w przyszłości realizować zaawansowaną ścieżkę rozwoju matematycznego i chcą mieć solidne podstawy. Oczywiście potok standardowy również stanowi solidne i wystarczające do nich przygotowanie – jednak z akcentami na kwestie i koncepcje bardziej podstawowe.

Potok gwiazdkowy jest dla tych z Państwa, dla których zajmowanie się matematyką jest wartością samą w sobie. Studia są trudne, a zajęcia gwiazdkowe – tym bardziej. Warto próbować pracować w swoim tempie i w poszanowaniu swojego czasu, możliwości, sił i zainteresowań. To jest najważniejsze – lepiej nauczyć się „mniej, ale dobrze", niż „więcej, ale płytko". Z pewnością jednak więcej pracy włożyć trzeba będzie w samodzielne opracowanie tematów podstawowych. Tu pracować będziemy nad ich pogłębieniem.

Gdyby chcieć uciec się do metafory górskiej (a nie ukrywam, że bardzo mi one pasują – patrz stosowna opowieść dla potoku standardowego: `https://mimuw.edu.pl/~amecel/notatkiawyklad.pdf`), potok gwiazdkowy przypomina chodzenie po Tatrach zimą. Nie są to Alpy czy Himalaje matematyki – uczymy się wciąż rzeczy bardzo fundamentalnych i przemierzamy szlaki dobrze znane każdemu solidnie wykształconemu matematykowi – dobrze znane, opisane i wielokrotnie przechodzone. Równoległy potok można porównać do grupy działającej latem – oczywiście pod opieką instruktorów dążących do wyrobienia u Państwa swoistej górskiej (a praktycznie – algebraicznej i geometrycznej) samodzielności. Nie przemierzamy już tylko szlaków „turystycznych" – studia to czas głębszego wtajemniczenia, wejście na szlaki zastrzeżone dla osób mających odpowiednie umiejętności i przygotowanie – umiejących reagować na zmienne warunki pogodowe i świadomych swoich możliwości. Co zatem powiedzieć o szlakach zimowych?

Pierwsza ważna różnica – to długość dnia. Zimą wycieczka standardowym choćby szlakiem planowana musi być z uwzględnieniem krótkiego czasu nasłonecznienia oraz obiektywnych trudności. Wiele czynności przygotowawczych do drogi wykonać trzeba zawczasu, a w trasie niewiele jest przestrzeni na przestoje. Trzeba wiedzieć co się robi, być skupionym i sprawnie przyswajać wiedzę – inaczej zabraknie czasu na realizację trasy i efekt będzie potencjalnie nawet gorszy niż podczas „standardowego" letniego kursu.

Druga różnica – choć poruszamy się szlakami podobnymi do letnich – ich przebieg nie jest zawsze oczywisty z uwagi na warunki. Pewne odcinki pokonać należy inaczej (czasem warto!), unikając lawin, groźnych wystaw, niekorzystnych czy zbędnych dłużyzn. Chcemy pokonywać wymagające trasy – poznać niektóre fakty z dowodami, kontekstem, w odniesieniu do innych. Czasem trzeba uniknąć „lawiny" rachunków czy komplikacji, jakie pojawiłyby się bez zastosowania odpowiednio zaawansowanego języka. Czasem w wykładzie pójdziemy nieco inną drogą – nie tylko z konieczności, ale i z chęci – po naukę wybierania ciekawych alternatyw. Być może nie ma królewskiej drogi do matematyki, ale to nie znaczy, że jest tylko jedna droga.

Trzecia i najważniejsza różnica – Tatry zimą oferują srogie warunki alpejskie i wielu słynnych „lodowych wojowników" zmagało się z nimi podczas przygotowań na wielomiesięczne wyprawy. Jest zapewne jasne co to znaczy – chcemy wspólnie stworzyć „laboratorium" poważnej pracy naukowej tak, by „wiążac się liną" w zaangażowanym zespole – doświadczyć przygody przybliżającej nas do matematyki profesjonalnej.

Rozdział 1

Endomorfizmy, macierze podobne. Wektory i wartości własne

W drugim semestrze kluczowe będzie dla nas rozważanie dodatkowych struktur na przestrzeniach liniowych, pozwalających na uzyskiwanie nowych rezultatów. Podstawowa jest struktura algebry, w której dodatkowym działaniem jest mnożenie wektorów.

Definicja 1.1: Algebra nad ciałem K

Przestrzeń liniową V nad ciałem K nazywamy ALGEBRĄ nad ciałem K, lub K-algebrą, jeśli istnieje łączne działanie $V \times V \to V$ postaci $(x, y) \mapsto x \cdot y$, zwane mnożeniem, liniowe ze względu na każdą współrzędną, tzn. dla każdych $\lambda, \mu \in K$ oraz $\alpha, \beta, \gamma \in V$ mamy:

$$(\lambda\alpha + \mu\beta) \cdot \gamma = \lambda(\alpha \cdot \gamma) + \mu(\beta \cdot \gamma) \quad \text{oraz} \quad \alpha(\lambda\beta + \mu\gamma) = \lambda(\alpha \cdot \beta) + \mu(\alpha \cdot \gamma).$$

Przypomnijmy podstawowe, znane nam już przykłady algebr.

- Jeśli $K \subseteq L$ są ciałami, to L ma strukturę K-algebry, gdzie mnożeniem jest mnożenie w L.
- Przestrzeń liniowa wielomianów $K[x]$, gdzie mnożeniem wektorów jest mnożenie wielomianów.
- Przestrzeń liniowa macierzy kwadratowych $M_n(K)$ rozmiaru n o wyrazach w ciele K, gdzie mnożeniem wektorów jest mnożenie macierzy.
- Przestrzeń liniowa przekształceń liniowych $L(V, V)$ nad ciałem K, z działaniem składania funkcji.

Powyższe przykłady są ważne z punktu widzenia algebry liniowej, choć rozważa się wiele innych, choćby algebrę kwaternionów Hamiltona (czterowymiarowa $\mathbb{R}$-algebra z odpowiednim mnożeniem).

Definicja 1.2: Endomorfizm i macierz endomorfizmu w bazie

Niech V będzie przestrzenią liniową nad ciałem K.

- Przekształcenie liniowe $\phi : V \to V$ nazywamy ENDOMORFIZMEM PRZESTRZENI V. Zbiór endomorfizmów przestrzeni V oznaczamy $\mathrm{End}(V)$. Innymi słowy $\mathrm{End}(V) = L(V, V)$.
- Izomorfizm $\phi \in \mathrm{End}(V)$ nazywamy AUTOMORFIZMEM PRZESTRZENI V. Zbiór automorfizmów przestrzeni V określamy jako $\mathrm{Aut}(V)$. Zbiór macierzy odwracalnych rozmiaru n o wyrazach w ciele K oznaczamy przez $\mathrm{GL}_n(K)$ (są to macierze automorfizmów, gdy $\dim V < \infty$).
- Niech $\phi \in \mathrm{End}(V)$. Przez n-krotną ITERACJĘ lub alternatywnie: n-tą POTĘGĘ tego endomorfizmu rozumiemy n-krotne złożenie ϕ ze sobą, czyli $\phi^n = \underbrace{\phi \circ \phi \circ \ldots \circ \phi}_{n}$, przy czym $\phi^0 := \mathrm{id}_V$.
- MACIERZĄ ENDOMORFIZMU $\phi \in \mathrm{End}(V)$ W BAZIE $\mathcal{A}$ przestrzeni V nazywamy macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$.

Załóżmy, że $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ oraz $\phi \in \mathrm{End}(V)$. Wówczas $M(\phi^2)^{\mathcal{A}}_{\mathcal{A}} = A^2$. Co nam to daje? Można stawiać pytanie: co o geometrycznej naturze działania endomorfizmu mówi fakt, że spełnia ono jakąś tożsamość typu wielomianowego? Wiemy chociażby, że równość $\phi^2 = \phi$ jest równoważna temu, że ϕ jest rzutem. W języku macierzy A endomorfizmu ϕ w (dowolnej) bazie tożsamość ta przekłada się na równość $A^2 = A$.

Z powyższym pytaniem wiąże się inne, nieco bardziej podstawowe: jak „najprościej" wyglądać mogą macierze endomorfizmów przestrzeni skończonego wymiaru w zależności od wyboru bazy?

Rozważmy następujący przykład „motywacyjny". Endomorfizm przestrzeni liniowej $\mathbb{R}^2$ zadany wzorem

$$\phi((x_1, x_2)) = (2x_1 + x_2, x_1 + 2x_2)$$

spełnia $\phi((1,-1)) = (1,-1)$ oraz $\phi((1,1)) = (3,3)$, a więc ϕ ma w bazie $\mathcal{A} = ((1,-1),(1,1))$ macierz

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},$$

Okazuje się, że w bazie $\mathcal{A}$ „widzimy" geometrię przekształcenia ϕ. Jest ono swego rodzaju *rozciąganiem* w kierunkach (osiach) $(1,-1)$ oraz $(1,1)$ w skali odpowiednio $1, 3$. Mówiąc bardziej precyzyjnie, dla dowolnych współrzędnych $a, b \in \mathbb{R}$ w bazie $\mathcal{A}$ widzimy nie tylko zachowanie ϕ, ale dowolnej jego iteracji.

$$\phi(a(1,-1) + b(1,1)) = 1 \cdot a(1,-1) + 3 \cdot b(1,1) \text{ oraz } \phi^n(a(1,-1) + b(1,1)) = 1^n \cdot a(1,-1) + 3^n \cdot b(1,1).$$

Przejdźmy do opisu macierzy ustalonego endomorfizmu. Zacznijmy od prostej obserwacji.

Obserwacja 1.1

Niech $\mathcal{A}, \mathcal{B}$ będą bazami przestrzeni liniowej V wymiaru n nad ciałem K oraz niech $\phi \in \mathrm{End}(V)$. Niech $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ oraz $B = M(\phi)^{\mathcal{B}}_{\mathcal{B}}$. Wówczas

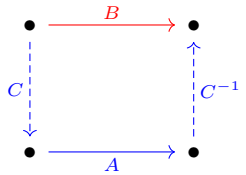
$$B = C^{-1}AC,$$

gdzie $C = M(\mathrm{id})^{\mathcal{A}}_{\mathcal{B}}$, tzn. $C \in M_n(K)$ jest macierzą zmiany współrzędnych od $\mathcal{B}$ do $\mathcal{A}$.

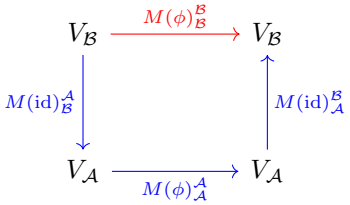
Dowód. Mamy $B = M(\phi)^{\mathcal{B}}_{\mathcal{B}} = M(\mathrm{id} \circ \phi \circ \mathrm{id})^{\mathcal{B}}_{\mathcal{B}} = M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}} \cdot M(\phi)^{\mathcal{A}}_{\mathcal{A}} \cdot M(\mathrm{id})^{\mathcal{A}}_{\mathcal{B}} = C^{-1}AC.$ □

Zatrzymamy się na moment, żeby spojrzeć na dowód powyższego rezultatu z nieco innej perspektywy. W poprzednim semestrze wspominaliśmy o interpretacji diagramowej złożeń przekształceń liniowych. Warto jest prześledzić sformułowanie i dowód powyższej obserwacji w ten właśnie sposób.

Macierz A reprezentuje endomorfizm $\phi \in \mathrm{End}(V)$, który czytamy we współrzędnych V w bazie $\mathcal{A}$, zaś macierz B reprezentuje ten sam endomorfizm, ale na V patrzymy inaczej — w bazie $\mathcal{B}$. Zapiszmy za pomocą diagramu równość macierzy $B = C^{-1}AC$. Pytamy czy istnieje taka macierz odwracalna C, że:



Zwróćmy uwagę na kolejność składania strzałek reprezentujących poszczególne macierze iloczynu $C^{-1}AC$. Macierz $C = M(\mathrm{id})^{\mathcal{A}}_{\mathcal{B}}$ znajdujemy patrząc na równość $M(\phi)^{\mathcal{B}}_{\mathcal{B}} = M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}} \cdot M(\phi)^{\mathcal{A}}_{\mathcal{A}} \cdot M(\mathrm{id})^{\mathcal{A}}_{\mathcal{B}}$.



Możemy też zupełnie uprościć zapis i wskazać tylko to, co w dowodzie wyżej istotne, czyli $\phi = \mathrm{id} \circ \phi \circ \mathrm{id}$.

$$\begin{array}{ccc} V & \xrightarrow{\phi} & V \\ {\scriptstyle \mathrm{id}}\downarrow & & \uparrow{\scriptstyle \mathrm{id}} \\ V & \xrightarrow[\phi]{} & V \end{array}$$

Poniższy diagram natomiast wyraża równość $B^3 = (C^{-1}AC)^3 = (C^{-1}AC)(C^{-1}AC)(C^{-1}AC) = C^{-1}A^3C$.

$$\begin{array}{ccccccc} V_{\mathcal{B}} & \xrightarrow{B} & V_{\mathcal{B}} & \xrightarrow{B} & V_{\mathcal{B}} & \xrightarrow{B} & V_{\mathcal{B}} \\ {\scriptstyle C}\downarrow & & {\scriptstyle C}\downarrow\uparrow{\scriptstyle C^{-1}} & & {\scriptstyle C}\downarrow\uparrow{\scriptstyle C^{-1}} & & \uparrow{\scriptstyle C^{-1}} \\ V_{\mathcal{A}} & \xrightarrow{A} & V_{\mathcal{A}} & \xrightarrow{A} & V_{\mathcal{A}} & \xrightarrow{A} & V_{\mathcal{A}} \end{array}$$

Powyższa obserwacja wyraźnie wskazuje na różnicę pomiędzy macierzami endomorfizmów w różnych bazach i macierzami w ustalonej bazie. Motywuje też wprowadzenie fundamentalnego dla nas pojęcia.

Definicja 1.3: Macierze podobne

Macierze $A, B \in M_n(K)$ nazywamy PODOBNYMI, jeśli istnieje macierz odwracalna $C \in M_n(K)$ taka, że:

$$B = C^{-1}AC.$$

Przykłady.

- Rozważmy macierze kwadratowe $A, B \in M_2(\mathbb{Q})$ oraz macierz odwracalną $P \in M_2(\mathbb{Q})$ postaci:

$$A = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -5 & -3 \\ 6 & 5 \end{bmatrix}, \quad P = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}.$$

  Macierze A, B są podobne, a dokładniej $B = P^{-1}AP$, ponieważ:

$$P^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \quad \text{oraz} \quad P^{-1}AP = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -5 & -3 \\ 6 & 5 \end{bmatrix}.$$

- Macierze

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{oraz} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

  są podobne nad dowolnym ciałem K, bowiem $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. W tym przypadku C jest macierzą operacji elementarnej zamiany wierszy/kolumn i mamy dodatkowo równość

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

- Dla dowolnej macierzy permutacyjnej $P \in S_n$, powstałej z macierzy identycznościowej $I \in M_n(K)$ przez permutację kolumn, mamy

$$P^{-1} = P^T.$$

  W szczególności, jeśli dane są dwie bazy przestrzeni n-wymiarowej V:

$$\mathcal{A} = (\alpha_1, \alpha_2, \ldots, \alpha_n), \quad \mathcal{B} = (\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)}),$$

  przy czym $\mathcal{B}$ jest bazą powstałą przez zamianę kolejności wektorów w bazie $\mathcal{A}$ za pomocą permutacji σ reprezentowanej przez macierz $P = M(\mathrm{id})^{\mathcal{A}}_{\mathcal{B}}$, to dla $\phi \in \mathrm{End}(V)$ mamy:

$$A = M(\phi)^{\mathcal{A}}_{\mathcal{A}} \iff P^{-1}AP = M(\phi)^{\mathcal{B}}_{\mathcal{B}}.$$

Obserwacja 1.2

Niech $A, B \in M_n(K)$. Następujące warunki są równoważne:

(a) macierze A, B są podobne,

(b) istnieje $\phi \in \mathrm{End}(K^n)$ oraz bazy $\mathcal{A}, \mathcal{B}$ przestrzeni K^n takie, że

$$A = M(\phi)^{\mathcal{A}}_{\mathcal{A}} \quad \text{oraz} \quad B = M(\phi)^{\mathcal{B}}_{\mathcal{B}}.$$

Dowód. Implikacja (b) $\Rightarrow$ (a) wynika z poprzedniej obserwacji. Dowodzimy zatem (a) $\Rightarrow$ (b). Przypuśćmy, że $B = C^{-1}AC$, dla pewnej macierzy odwracalnej C. Niech endomorfizm ϕ przestrzeni K^n będzie zadany warunkiem $M(\phi)^{st}_{st} = A$. Niech $\mathcal{B}$ będzie bazą przestrzeni K^n złożoną z kolumn macierzy C. Wówczas

$$C = M(\mathrm{id})^{st}_{\mathcal{B}}, \quad \text{więc} \quad C^{-1} = M(\mathrm{id})^{\mathcal{B}}_{st}.$$

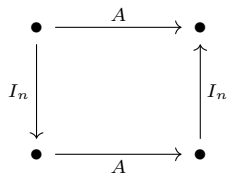
Stąd:

$$B = C^{-1}AC = M(\mathrm{id})^{\mathcal{B}}_{st} \cdot M(\phi)^{st}_{st} \cdot M(\mathrm{id})^{st}_{\mathcal{B}} = M(\phi)^{\mathcal{B}}_{\mathcal{B}}.$$

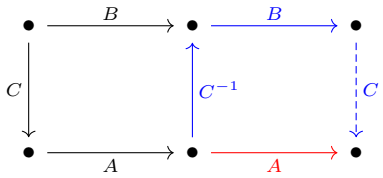
Otrzymaliśmy $A = M(\phi)^{st}_{st}$ oraz $B = M(\phi)^{\mathcal{B}}_{\mathcal{B}}$. □

Relacja podobieństwa w zbiorze $M_n(K)$ jest relacją równoważności. Czasem podobieństwo macierzy nad ciałem K oznacza się symbolem $A \sim_K B$, lub nawet $A \sim B$. Nie piszemy natomiast „macierze podobne nad ciałem K", bowiem jak się okazuje — i nie jest to łatwe — jeśli macierze o wyrazach w ciele K są podobne nad ciałem $L \supseteq K$, to muszą być też podobne nad K. Temat ten omówimy dokładniej w jednym z dodatków.

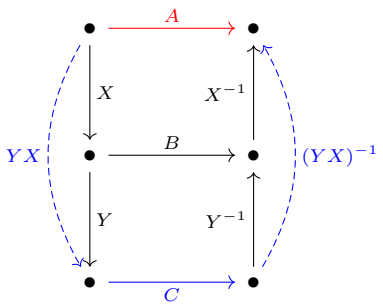
- **Zwrotność relacji podobieństwa**. Dla każdej $A \in M_n(K)$ mamy $A = I_n \cdot A \cdot I_n = (I_n)^{-1} \cdot A \cdot I_n$, a zatem A jest macierzą podobną do samej siebie.



- **Symetryczność relacji podobieństwa**. Jeśli dla macierzy $A, B \in M_n(K)$ macierz A jest podobna do macierzy B, to istnieje macierz odwracalna C, że $B = C^{-1}AC$. W rezultacie otrzymujemy $A = CBC^{-1} = (C^{-1})^{-1}BC^{-1}$, czyli także macierz B jest podobna do A.



- **Przechodniość relacji podobieństwa**. Załóżmy, że dla $A, B, C \in M_n(K)$ istnieją macierze odwracalne $X, Y \in M_n(K)$, że $A = X^{-1}BX$ i $B = Y^{-1}CY$, to $A = X^{-1}Y^{-1}CYX = (YX)^{-1}C(YX)$.



Kluczowe przykłady dla zrozumienia „geometrii" klas podobieństwa.

- Dla dowolnej macierzy odwracalnej $C \in M_n(K)$ zachodzi równość: $C^{-1} \cdot \lambda I \cdot C = \lambda C^{-1} C = \lambda I$. Homotetia o skali λ ma zatem tylko jedną „macierz w bazie".
- Niech $V = W \oplus U$ i niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ będzie taką bazą przestrzeni V, że $(\alpha_1, \ldots, \alpha_k)$ (dla pewnego $1 \leqslant k < n$) jest bazą przestrzeni W oraz $(\alpha_{k+1}, \ldots, \alpha_n)$ jest bazą przestrzeni U. Niech $\phi : V \to V$ będzie rzutem V na W oraz niech ψ będzie symetrią V względem W wzdłuż U. Wówczas macierz $R_k = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ ma w pierwszych k kolumnach pierwsze k wektorów bazy standardowej K^n, zaś dalej kolumny zerowe. Analogiczną macierz wskazać można dla symetrii $S_k = M(\psi)^{\mathcal{A}}_{\mathcal{A}}$:

$$R_k = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}, \quad S_k = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & -1 \end{bmatrix}.$$

Widzimy więc, że „geometrię jednokładności" można rozpoznać po klasie podobieństwa jej macierzy. Podobnie „rozpoznajemy" rzut lub symetrię, ale tu morał jest inny — macierz R_k jest podobna do macierzy KAŻDEGO rzutu przestrzeni n-wymiarowej (nad K) na podprzestrzeń wymiaru k.

Definicja 1.4: Ślad macierzy kwadratowej

ŚLADEM MACIERZY $A = [a_{ij}] \in M_n(K)$ nazywamy sumę wyrazów stojących na przekątnej macierzy A, to znaczy element $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} \in K$.

Ślad jest niezwykle istotnym obiektem i niezmiennikiem wielu ważnych konstrukcji, pomimo swojej pozornej prostoty. Oto kluczowa własność w kontekście podobieństwa macierzy. Jej dowód to łatwe ćwiczenie.

Obserwacja 1.3: Ćwiczenie

Dla dowolnych macierzy $A, B \in M_n(K)$ zachodzi równość $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Mówiąc o niezmiennikach podobieństwa, mamy na razie na myśli przede wszystkim niezmienniki liczbowe: rząd, wyznacznik i ślad. Wkrótce dołączą do nich bardziej skomplikowane. Oto podstawowe fakty.

Obserwacja 1.4

Jeśli macierze $A, B \in M_n(K)$ są podobne, to:

$$r(A) = r(B), \quad \det(A) = \det(B), \quad \operatorname{tr}(A) = \operatorname{tr}(B).$$

Dowód. Jeśli A i B są macierzami podobnymi, to na mocy Obserwacji 1.2 istnieje $\phi \in \operatorname{End}(K^n)$ oraz bazy $\mathcal{A}, \mathcal{B}$ przestrzeni K^n takie, że $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}, B = M(\phi)^{\mathcal{B}}_{\mathcal{B}}$. W szczególności $r(A) = r(B) = \dim \operatorname{im} \phi$ (co pokazywaliśmy już w poprzednim semestrze). Dalej, skoro macierze A, B są podobne, to istnieje macierz odwracalna C, że $B = C^{-1}AC$. Zatem z wzoru Cauchy'ego:

$$\det(B) = \det(C^{-1}AC) = \det(C^{-1}) \det(A) \det(C) = \frac{\det(A) \det C}{\det(C)} = \det(A).$$

Wreszcie, korzystając z Obserwacji 1.3 dla $C^{-1}A$ oraz C — $\operatorname{tr}(B) = \operatorname{tr}(C^{-1}AC) = \operatorname{tr}(CC^{-1}A) = \operatorname{tr}(A)$. □

Definicja 1.5: Wyznacznik i ślad endomorfizmu

Niech $\dim V < \infty$ oraz niech $\phi \in \mathrm{End}(V)$.

- Wyznacznikiem endomorfizmu ϕ, oznaczanym przez $\det \phi$, nazywamy taki element $\det(A) \in K$, gdzie $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$, dla pewnej bazy $\mathcal{A}$ przestrzeni V.
- Śladem endomorfizmu, oznaczanym przez $\mathrm{tr}\,\phi$, nazywamy taki element $\mathrm{tr}(A) \in K$, gdzie $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$, dla pewnej bazy $\mathcal{A}$ przestrzeni V.

Czytelnik zechce zauważyć, że poniższe dwie macierze mają takie same rzędy, wyznaczniki i ślady nad dowolnym ciałem:

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

Macierze te oczywiście nie są podobne, gdyż do macierzy I_2 podobna jest (jak już wiemy) tylko ona sama.

Kolejnym ważnym niezmiennikiem endomorfizmu są jego tzw. wartości własne.

Definicja 1.6: Wektor własny, podprzestrzeń własna

Niech V będzie przestrzenią liniową nad ciałem K oraz niech $\phi \in \mathrm{End}(V)$.

- Wektor $\alpha \in V$ nazywamy wektorem własnym endomorfizmu ϕ jeśli:
  - $\alpha \neq 0$,
  - istnieje $a \in K$ takie, że $\phi(\alpha) = a \cdot \alpha$.
- Element $a \in K$ spełniający powyższe warunki nazywamy wartością własną endomorfizmu ϕ, zaś o wektorze α mówimy, że jest wektorem własnym ϕ o wartości własnej a.
- Jeśli a jest wartością własną endomorfizmu $\phi \in \mathrm{End}(V)$, to zbiór

$$V_{(a)} = \{\alpha \in V \mid \phi(\alpha) = a \cdot \alpha\}$$

nazywamy podprzestrzenią[a] własną endomorfizmu ϕ odpowiadającą wart. własnej a.

[a] Podprzestrzeń ta składa się z wektorów własnych endomorfizmu ϕ o wartości własnej a oraz z wektora zerowego.

Oto przykłady.

- Jeśli $\phi : K^n \to K^n$ jest jednokładnością o skali a, to każdy niezerowy wektor K^n jest wektorem własnym o wartości własnej a. Zatem $V_{(a)} = K^n$.
- Endomorfizm $\phi \in \mathrm{End}(\mathbb{R}^2)$ o macierzy postaci $M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ ma dwie wartości własne: $1, 2$ oraz $V_{(1)} = \mathrm{lin}((1,0)), V_{(2)} = \mathrm{lin}((0,1))$, co można sprawdzić wypisując wzór na ϕ lub sprawdzając, że wektory $(1,0)$ oraz $(0,1)$ są wektorami własnymi macierzy $M(\phi)^{st}_{st}$:

$$\phi((1,0)) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \phi((0,1)) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

- Endomorfizm $\phi \in \mathrm{End}(\mathbb{R}^3)$ zadany wzorem $\phi((x_1, x_2, x_3)) = (x_1 + 2x_2 + 4x_3, 3x_1 + 2x_2 + 2x_3, x_1 + 6x_2)$ ma wektor własny $(1,1,1)$ o wartości własnej 7, tzn. $\phi((1,1,1)) = (7,7,7) = 7 \cdot (1,1,1)$.
- Choć na tym wykładzie interesować nas będą jedynie endomorfizmy przestrzeni skończenie wymiarowych, warto wspomnieć przykłady nieskończenie wymiarowe: weźmy $T \in \mathrm{End}(K^\infty)$ postaci:

$$T((a_1, a_2, a_3, \ldots)) = (a_2, a_3, a_4, \ldots).$$

Nietrudno sprawdzić, że dowolny element ciała K jest wartością własną tego endomorfizmu. Czy potraficie Państwo wskazać odpowiednie wektory własne?

- Branie pochodnej to endomorfizm przestrzeni $K[x]$. Łatwo widzieć, że jego jedyną wartością własną jest 0. A gdy zamiast $K[x]$ weźmiemy podprzestrzeń funkcji różniczkowalnych w $F(\mathbb{R},\mathbb{R})$?
- Endomorfizm $\phi:\mathbb{R}^2\to\mathbb{R}^2$ postaci $\phi((x_1,x_2))=(-x_2,x_1)$ nie ma rzeczywistych wartości własnych. Z geometrycznego punktu widzenia jest to obrót o $\pi/2$.

Określamy również pojęcie wektora własnego i wartości własnej macierzy kwadratowej.

Definicja 1.7: Wartość własna macierzy

Niech $A\in M_n(K)$. Mówimy, że $a\in K$ jest WARTOŚCIĄ WŁASNĄ MACIERZY A, jeśli a jest wartością własną endomorfizmu $\phi\in\mathrm{End}(K^n)$ zadanego wzorem $M(\phi)^{st}_{st}=A$. Niezerowy wektor $v\in K^n$ nazwiemy WEKTOREM WŁASNYM MACIERZY A o wartości własnej a, jeśli $A\cdot v=a\cdot v$. Zbiór wartości własnych macierzy oznaczamy przez $\mathrm{spec}(A)$ i nazywamy SPEKTRUM MACIERZY A.

Ogólnej teorii wektorów i podprzestrzeni własnych poświęcimy kolejny wykład. Tu odnotujmy jedynie podstawowe obserwacje dotyczące ich wyznaczania. W tym celu kluczowe jest przypomnienie fundamentalnego rezultatu poznanego przez nas w pierwszym semestrze.

Twierdzenie 1.1

Niech $A\in M_n(K)$. Następujące warunki są równoważne:

(a) układ równań postaci $Ax=0$ ma NIEZEROWE rozwiązanie $v\in K^n$,

(b) $\det(A)=0$.

Zacznijmy od przykładu. Niech $\phi:\mathbb{R}^3\to\mathbb{R}^3$ będzie endomorfizmem przestrzeni $\mathbb{R}^3$ zadanym macierzą:

$$M(\phi)^{st}_{st}=\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}.$$

Znalezienie wektorów i wartości własnych endomorfizmu ϕ sprowadza się do rozwiązania układu:

$$\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\lambda\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}\lambda x_1\\\lambda x_2\\\lambda x_3\end{bmatrix}.$$

Przekształcając równoważnie mamy zatem:

$$\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\lambda\cdot\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\iff\left(\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}-\lambda\cdot\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Należy więc rozstrzygnąć dla jakich wartości $\lambda\in\mathbb{R}$ jednorodny układ o macierzy $M(\phi)^{st}_{st}-\lambda I_3$ postaci:

$$\begin{bmatrix}1-\lambda&2&2\\2&1-\lambda&2\\2&2&1-\lambda\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

MA NIEZEROWE ROZWIĄZANIE, gdyż wektor własny ma być niezerowy. Korzystamy z Twierdzenia 1.1. Macierz $M(\phi)^{st}_{st}-\lambda I$ musi być rzędu <3, więc szukamy λ takich, że $\det(M(\phi)^{st}_{st}-\lambda I)=0$. Mamy:

$$\begin{vmatrix}1-\lambda&2&2\\2&1-\lambda&2\\2&2&1-\lambda\end{vmatrix}=-(\lambda-5)(\lambda+1)^2.$$

Stąd wartościami własnymi endomorfizmu ϕ są -1 oraz 5, a po rozwiązaniu układów o macierzach $M(\phi)^{st}_{st}+I$ oraz $M(\phi)^{st}_{st}-5I$ dostajemy podprzestrzenie własne:

$$V_{(-1)}=\mathrm{lin}((-1,1,0),(-1,0,1)),\quad V_{(5)}=\mathrm{lin}((1,1,1)).$$

Zauważmy, że układ $\mathcal{A}=((-1,1,0),(-1,0,1),(1,1,1))$ jest bazą $\mathbb{R}^3$ złożoną z wektorów własnych ϕ.

1.1 Uzupełnienie. Macierze endomorfizmu w różnych bazach

Przyglądając się izomorfizmom przestrzeni liniowych łatwo stwierdzimy, że macierze ich przekształceń w dowolnych (być może różnych) bazach są dowolnymi macierzami odwracalnymi. W szczególności dla dowolnego izomorfizmu istnieją bazy, w których ma on macierz identycznościową.

Obserwacja 1.5

Niech $\phi : K^n \to K^n$ będzie izomorfizmem. Istnieją takie bazy $\mathcal{A}$, $\mathcal{B}$ przestrzeni K^n, że:

$$M(\phi)^{\mathcal{B}}_{\mathcal{A}} = I_n.$$

Dowód. Niech $(\alpha_1, \ldots, \alpha_n)$ będzie bazą K^n. Wówczas $(\phi(\alpha_1), \ldots, \phi(\alpha_n))$ jest również bazą K^n, bo ϕ to izomorfizm. A zatem biorąc pierwszą z tych baz jako $\mathcal{A}$, a drugą jako $\mathcal{B}$ dostajemy, że $M(\phi)^{\mathcal{B}}_{\mathcal{A}} = I_n$. □

Geometrycznie obserwacja ta może być interpretowana tak, że izomorfizm przestrzeni liniowych jest jedynie swego rodzaju przejściem z jednego układu współrzędnych do drugiego.

Obserwacja 1.6

Każda macierz odwracalna w $M_n(K)$ jest macierzą izomorfizmu $\phi : K^n \to K^n$ w pewnych bazach.

Dowód. Niech $G \in M_n(K)$ będzie macierzą odwracalną. Niech $\mathcal{A}$, $\mathcal{B}$ będą takimi bazami przestrzeni K^n, że $M(\phi)^{\mathcal{B}}_{\mathcal{A}} = I_n$. Wystarczy pokazać, że istnieje baza $\mathcal{C}$ przestrzeni K^n, że $G = M(\mathrm{id})^{\mathcal{A}}_{\mathcal{C}}$, bo wówczas

$$M(\phi)^{\mathcal{B}}_{\mathcal{C}} = M(\phi)^{\mathcal{B}}_{\mathcal{A}} \cdot M(\mathrm{id})^{\mathcal{A}}_{\mathcal{C}} = I_n G = G.$$

Otóż szukaną bazę $\mathcal{C}$ można odczytać z kolumn macierzy AG, gdzie $A = M(\mathrm{id})^{st}_{\mathcal{A}}$. Wówczas mamy

$$G = A^{-1}(AG) = M(\mathrm{id})^{\mathcal{A}}_{st} \cdot M(\mathrm{id})^{st}_{\mathcal{C}} = M(\mathrm{id})^{\mathcal{A}}_{\mathcal{C}}.$$

□

A zatem dowolna macierz odwracalna rozmiaru n nad ciałem K jest macierzą dowolnego izomorfizmu przestrzeni wymiaru n nad ciałem K.

Definicja 1.8: Macierze równoważne

Macierze $A, B \in M_{m\times n}(K)$ nazywamy RÓWNOWAŻNYMI, jeśli istnieją takie macierze odwracalne $C \in M_m(K)$ oraz $D \in M_n(K)$ takie, że $B = CAD$.

Obserwacja 1.7

Dla macierzy $A, B \in M_{m\times n}(K)$ następujące warunki są równoważne:

(i) macierze A oraz B są równoważne,

(ii) istnieje przekształcenie liniowe $\phi : K^n \to K^m$ oraz bazy $\mathcal{A}, \mathcal{A}'$ przestrzeni K^n i bazy $\mathcal{B}, \mathcal{B}'$ przestrzeni K^m, że $A = M(\phi)^{\mathcal{B}}_{\mathcal{A}}$ oraz $B = M(\phi)^{\mathcal{B}'}_{\mathcal{A}'}$,

(iii) B może być otrzymana z A ciągiem operacji elementarnych na wierszach i kolumnach,

(iv) $r(A) = r(B)$.

Dowód. Dowodzimy (i) $\Rightarrow$ (ii). Niech $M(\phi)_{st}^{st} = A$ oraz niech $B = CAD$. Przez $\mathcal{C}, \mathcal{D}$ okreslamy bazy złożone odpowiednio z kolumn macierzy odwracalnych C^{-1} i D. Wówczas $C = M(\mathrm{id})_{st}^{\mathcal{C}}$ oraz $D = M(\mathrm{id})_{\mathcal{D}}^{st}$. Stąd $B = CAD = M(\phi)_{\mathcal{D}}^{\mathcal{C}}$.

Implikacja (ii) $\Rightarrow$ (iii) wynika z tego, że macierz jest odwracalna wtedy i tylko wtedy, gdy jest iloczynem macierzy operacji elementarnych (I semestr). W szczególności skoro $B = M(\mathrm{id})_{\mathcal{B}}^{\mathcal{B}'} \cdot A \cdot M(\mathrm{id})_{\mathcal{A}'}^{\mathcal{A}}$, gdzie $M(\mathrm{id})_{\mathcal{B}}^{\mathcal{B}'}$ oraz $M(\mathrm{id})_{\mathcal{A}'}^{\mathcal{A}}$ są odwracalne, to mamy (iii).

Implikacja (iii) $\Rightarrow$ (iv) jest jasna, ponieważ wykonanie operacji elementarnej nie zmienia rzędu.

Natomiast jeśli zachodzi (iv) to znaczy, że za pomocą operacji elementarnych na wierszach i kolumnach każdą z macierzy A i B można sprowadzić za pomocą operacji elementarnych do macierzy $X = [x_{ij}]$, gdzie $r = r(A) = r(B)$ oraz

$$x_{ij} = \begin{cases} 1, & \text{dla } i = j = 1, ..., r, \\ 0, & \text{dla pozostałych } i, j \end{cases},$$

czyli do macierzy rozmiaru $m \times n$ o postaci blokowej

$$X = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Rzeczywiście, wiemy że macierze tego samego rzędu można sprowadzić za pomocą operacji elementarnych na wierszach do postaci A', B', które są w zredukowanej postaci schodkowej i mają $r = r(A) = r(B)$ niezerowych wierszy. Innymi słowy istnieją macierze odwracalne $E, G \in M_m(K)$, że $A' = EA$ oraz $B' = GB$, gdzie E, G — iloczyny odpowiednich macierzy operacji elementarnych na wierszach. Teraz zaczynamy wykonywać operacje kolumnowe. Po pierwsze porządkujemy kolumny tak, by wyrazy wiodące stojące w pierwszych r wierszach stały w pierwszych r kolumnach. Innymi słowy za pomocą operacji kolumnowych sprowadzamy macierze A' oraz B' do postaci A'' oraz B'' postaci blokowej:

$$A'' = \begin{bmatrix} I_r & * \\ 0 & 0 \end{bmatrix}, \quad B'' = \begin{bmatrix} I_r & ** \\ 0 & 0 \end{bmatrix}.$$

Widzimy zatem, że za pomocą pierwszych r kolumn można za pomocą operacji kolumnowych doprowadzić te macierze do postaci X. Innymi słowy istnieją macierze odwracalne E, F, G, H, że

$$X = EAF \quad \text{oraz} \quad X = GBH.$$

Zatem $EAF = GBH$, czyli

$$A = E^{-1}GBHF^{-1}.$$

Macierze $E^{-1}G$ oraz HF^{-1} są odwracalne, więc uzyskujemy (i). □

W świetle przytoczonego twierdzenia powyższa definicja mówi, że macierze są równoważne wtedy i tylko wtedy, gdy są macierzami tego samego przekształcenia liniowego. Równoważność macierzy nie jest na pozór pojęciem mówiącym za wiele o geometrii przekształcenia, choć jest ono przydatne z punktu widzenia samej teorii macierzy. W jednym z kolejnych dodatków powiemy jednak o macierzach wielomianowych i problemie charakteryzacji podobieństwa macierzy nad dowolnym ciałem. Wówczas pojęcie równoważności owych macierzy wielomianowych (odpowiednio rozumiane) okaże się zaskakująco pożyteczne.

Wniosek 1.1

Dla każdego przekształcenia liniowego $\phi : K^n \to K^m$ rzędu r istnieją bazy $\mathcal{A}$, $\mathcal{B}$ przestrzeni K^n, K^m takie, że $M(\phi)_{\mathcal{A}}^{\mathcal{B}} = A$, przy czym $A \in M_{m\times n}(K)$ oraz $A = [a_{ij}]$, gdzie

$$a_{ij} = \begin{cases} 1, & \text{dla } i = j = 1, ..., r, \\ 0, & \text{dla pozostałych } i, j \end{cases},$$

czyli do macierzy rozmiaru $m \times n$ o postaci blokowej $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

Nietrudno widzieć, że równoważność jest relacją równoważności w zbiorze $M_{m\times n}(K)$. Wniosek wyżej mówi, że jest $\min(m,n)+1$ klas równoważności macierzy rozmiaru $m\times n$ (nad dowolnym ciałem), bowiem każdą macierz rozmiaru $m\times n$ można sprowadzić, za pomocą operacji elementarnych na wierszach i kolumnach do jednej z $\min(m,n)+1$ macierzy zero-jedynkowych o postaci blokowej jak wyżej.

Oczywiście rozważano bardziej skomplikowane problemy. Załóżmy, że mamy macierz rozmiaru $m\times(r+s)$ o postaci blokowej z blokami $A\in M_{m\times r}(K)$ oraz $B\in M_{m\times s}$ postaci

$$\begin{bmatrix}A & B\end{bmatrix}. \qquad (\dagger)$$

Załóżmy też, że możemy wykonywać wszystkie operacje wierszowe (na całej macierzy blokowej), ale operacje kolumnowe mogą być wykonywane „w obrębie" bloku kolumn: od 1 do r lub „w obrębie" bloku kolumn od $r+1$ do $r+s$ (to znaczy: nie można na przykład dodać do kolumny r kolumny $r+s$, albo ich zamienić, ale można zamienić kolumny $r+1, r+2$ należące do bloku B).

Oto przykład. Rozważmy macierz M o blokach 3×2 oraz 3×1 postaci:

$$M=\left[\begin{array}{cc|c}1&2&3\\4&5&6\\7&8&9\end{array}\right].$$

Wówczas wykonać możemy serię operacji wierszowych i kolumnowych, które „nie mieszają" tych bloków, dostając:

$$\left[\begin{array}{cc|c}1&2&3\\4&5&6\\7&8&9\end{array}\right]\xrightarrow{w_2-4w_1}\left[\begin{array}{cc|c}1&2&3\\0&-3&-6\\7&8&9\end{array}\right]\xrightarrow{w_3-7w_1}\left[\begin{array}{cc|c}1&2&3\\0&-3&-6\\0&-6&-12\end{array}\right]\xrightarrow{(-3)w_2}$$

$$\left[\begin{array}{cc|c}1&2&3\\0&1&2\\0&-6&-12\end{array}\right]\xrightarrow{w_3+6w_2}\left[\begin{array}{cc|c}1&2&3\\0&1&2\\0&0&0\end{array}\right]\xrightarrow{w_1-2w_2}\left[\begin{array}{cc|c}1&0&-1\\0&1&2\\0&0&0\end{array}\right].$$

Uzyskaliśmy postać wierszową zredukowaną. Teraz czas na operacje kolumnowe, respektujące bloki.

$$\left[\begin{array}{cc|c}1&0&-1\\0&1&2\\0&0&0\end{array}\right]\xrightarrow{(-1)k_3}\left[\begin{array}{cc|c}1&0&1\\0&1&-2\\0&0&0\end{array}\right]\xrightarrow{w_2+2w_1}\left[\begin{array}{cc|c}1&0&1\\2&1&0\\0&0&0\end{array}\right]\xrightarrow{k_1-2k_2}\left[\begin{array}{cc|c}1&0&1\\0&1&0\\0&0&0\end{array}\right].$$

Zachęcam do wykazania, że problem macierzy o postaci (†) można rozwiązać sprowadzając każdą z nich do nieoczywistej postaci blokowej

$$\left[\begin{array}{ccc|ccc}I_a&0&0&0&I_a&0\\0&I_b&0&0&0&0\\0&0&0&I_c&0&0\\0&0&0&0&0&0\end{array}\right].$$

Czy za pomocą takich operacji elementarnych można uzyskać jedynie SKOŃCZENIE WIELE MACIERZY (zerojedynkowych)? Jaką mają postać? Mnożeniu przez jakie macierze odpowiada ten problem?

A jeśli mielibyśmy analogiczny problemy dotyczący macierzy o trzech blokach

$$\begin{bmatrix}A & B & C\end{bmatrix}?$$

A gdybyśmy mieli dowolną macierz blokową i dopuszczalne by było tylko „działanie w obrębie bloków wierszy" i „działanie w obrębie bloków kolumn" (oczywiście na całą macierz, nie tylko na bloki)?

Powyższe pytania prowadzą do tzw. problemów macierzowych i były rozważane od początku XX wieku. W drugiej jego połowie były jednym z nurtów rozwoju ważnej gałęzi algebry: teorii reprezentacji algebr skończenie wymiarowych. Więcej (wraz z przykładami) można przeczytać w rozdziale pierwszym tekstu: `https://www.matem.unam.mx/~christof/cursos/12_SAB/1111_RepThBook.pdf`.

1.2 Dodatek. Dyskretny układ dynamiczny

Teoria endomorfizmów, a w szczególności problem znajdowania macierzy diagonalnej podobnej do danej ma liczne zastosowania między innymi w tak zwanych dyskretnych układach dynamicznych.

Rozważmy następujący przykład[1] Weźmy „ekosystem" złożony z:

- m_n (milionów) much,
- z_n (tysięcy) żab.

Po roku liczba much/żab zmienia się zgodnie z (nierealnym) modelem liniowym:

$$\begin{cases} m_{n+1} &= -0,36z_n + 1,22m_n, \\ z_{n+1} &= 0,38z_n + 0,24m_n \end{cases}. \qquad (\spadesuit)$$

Jak rozumieć taki model? Muchy się rozmnażają, ale też są zjadane. Żaby gdy nie mają co jeść, wymierają. Mimo wszystko co roku również się rozmnażają, ale w zupełnie innym tempie. Oto nasze pytanie, **Jaka jest dynamika tego układu w zależności od wyboru** z_0, m_0**?**

Widzimy, że naśladując rozumowania dotyczące rekurencji mamy:

$$\begin{bmatrix} z_{n+1} \\ m_{n+1} \end{bmatrix} = \begin{bmatrix} 0,38 & 0,24 \\ -0,36 & 1,22 \end{bmatrix}^n \cdot \begin{bmatrix} z_0 \\ m_0 \end{bmatrix}. \qquad (\diamondsuit)$$

Niech $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ dany będzie macierzą

$$M(\phi)^{st}_{st} = \begin{bmatrix} 0,38 & 0,24 \\ -0,36 & 1,22 \end{bmatrix}.$$

Czytelnik zechce zauważyć, że $\phi((1,3)) = 1,1 \cdot (1,3)$ oraz $\phi((2,1)) = 0,5 \cdot (2,1)$. Mamy zatem, dla $a, b \in \mathbb{R}$:

$$\phi(a(1,3) + b(2,1)) = 1,1 \cdot a(1,3) + 0,5 \cdot b(2,1) \quad \Longleftrightarrow \quad M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} 1,1 & 0 \\ 0 & 0,5 \end{bmatrix}.$$

Dlaczego postać diagonalna rozważanego endomorfizmu jest tak wygodna? Przepisujemy równanie $(\diamondsuit)$ w bazie $\mathcal{A}$, przyjmując:

$$\begin{bmatrix} z_0 \\ m_0 \end{bmatrix} = c_0 \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d_0 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} z_{n+1} \\ m_{n+1} \end{bmatrix} = c_{n+1} \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d_{n+1} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

skąd dostajemy:

$$\begin{bmatrix} c_{n+1} \\ d_{n+1} \end{bmatrix} = \begin{bmatrix} 1,1 & 0 \\ 0 & 0,5 \end{bmatrix}^n \cdot \begin{bmatrix} c_0 \\ d_0 \end{bmatrix} = \begin{bmatrix} (1,1)^n c_0 \\ (0,5)^n d_0 \end{bmatrix}.$$

Zatem:

$$\begin{cases} z_{n+1} &= 1 \cdot (1,1)^n c_0 + 2 \cdot (0,5)^n d_0 \\ m_{n+1} &= 3 \cdot (1,1)^n c_0 + 1 \cdot (0,5)^n d_0 \end{cases}. \qquad (\heartsuit)$$

[1] Źródło. K. Behrend, *Dynamical Systems and Matrix Algebra*, `http://www.math.ubc.ca/~behrend/math223/DynSys.pdf`. Zamiast czytać można też obejrzeć:

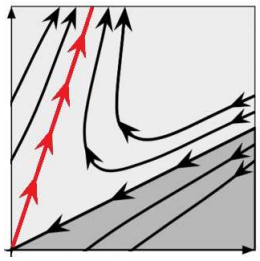
Po co nam potęgi macierzy? Film z czasów pandemii: `https://youtu.be/NkmoGML5uxs`

.

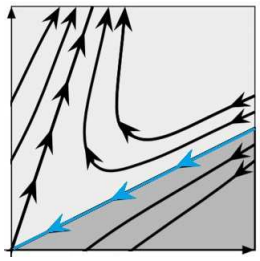
Zwróćmy uwagę, że uzyskany opis jest znacznie wygodniejszy od wyjściowego. Opisywanie „wektora much i żab" w standardowej bazie $\mathbb{R}^2$ układem (♠) zamiast w bazie z wektorów własnych, za pomocą układu (♡), nie pozwala poprawnie ocenić dynamiki układu, czyli tego jakie są skutki wyboru określonych wartości początkowych z_0, m_0.

Interpretacja graficzna uzyskanego wyniku w zależności od wyboru c_0, d_0.

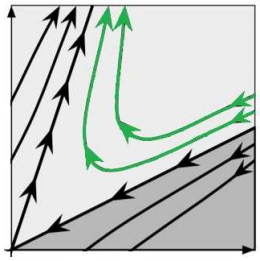
- Jeśli $c_0 = 1$, $d_0 = 0$, czyli $(z_0, m_0) = (1, 3)$ oraz $(z_{n+1}, m_{n+1}) = ((1, 1)^n, 3 \cdot (1, 1)^n)$, to populacje rosną w tym samym tempie: stosunek much do żab jest stały!



- Jeśli $c_0 = 0$, $d_0 = 1$, czyli $(z_0, m_0) = (2, 1)$ oraz $(z_{n+1}, m_{n+1}) = (2 \cdot (0, 5)^n, (0, 5)^n)$, to populacje maleją w tym samym tempie: stosunek much do żab jest stały!



- Jeśli $c_0 > 0$, $d_0 > 0$, to populacje być może pewien czas maleją, potem przyrastają. Stosunek żab do much zmierza „asymptotycznie" do $1 : 3$ (w istocie: $1 : 3000$). A jakie układy c_0, d_0 prowadzą do zagłady gatunków? Jest ona, jak się okazuje, możliwa.



Zagadnienia tu opisane są oczywiście jedynie drobną zabawką, ale opowiadają o jednym z ważnych zastosowań teorii endomorfizmów – rozwiązywaniu układów liniowych równań różniczkowych i różnicowych.

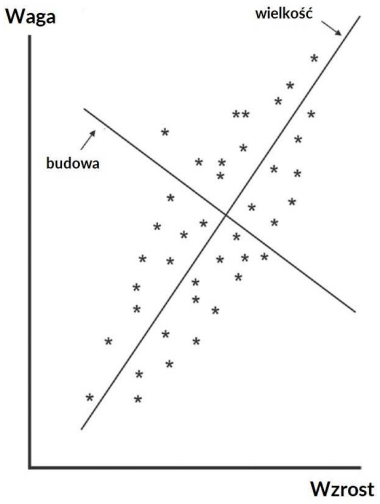
Na koniec warto jednak pomyśleć chwilę nad pewną komplikacją. Wyobraźmy sobie, że ktoś chce STEROWAĆ populacją much i żab mimo znajomości jej dynamiki. Jak to robi? Każdego roku wprowadza do ekosystemu pewną liczbę osobników, ale – uwaga – zachowując ich wyjściowe proporcje! Oto jak „wygląda" ekosystem po dwóch latach.

$$\begin{bmatrix} z_2' \\ m_2' \end{bmatrix} = \begin{bmatrix} 0,38 & 0,24 \\ -0,36 & 1,22 \end{bmatrix}^2 \begin{bmatrix} z_0 \\ m_0 \end{bmatrix} + a_1 \begin{bmatrix} 0,38 & 0,24 \\ -0,36 & 1,22 \end{bmatrix} \begin{bmatrix} z_0 \\ m_0 \end{bmatrix} + a_2 \begin{bmatrix} z_0 \\ m_0 \end{bmatrix}.$$

Jak teraz opisać liczbę osobników każdego roku w zależności od a_1, a_2? Czy można uzyskać dowolny rezultat? Jak zależy to od macierzy układu? Mowa tu o tzw. teorii sterowania. Zainteresowanych odsyłam do tekstu: `http://cse.lab.imtlucca.it/~bemporad/teaching/ac/pdf/01-Introduction.pdf`.

1.3 Trivia. PCA, czyli w jakich kierunkach patrzeć na dane?

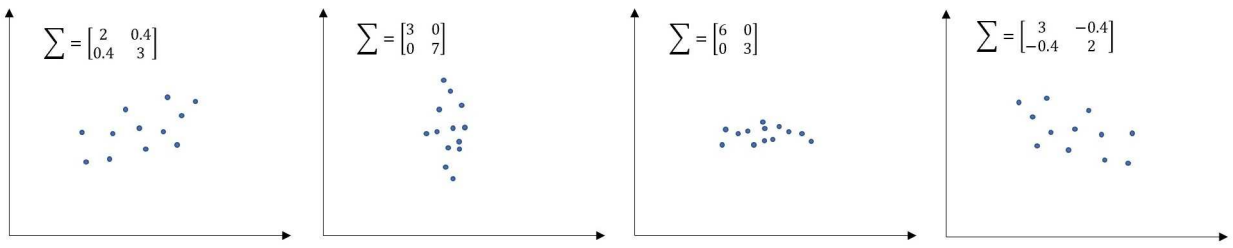
Analizując zbiory danych zainteresowani jesteśmy szukaniem w nich prawidłowości. Statystyka zna wiele metod uzyskiwania tego typu wyników. Jedną z nich jest PCA (Principal Components Anaiysis, czyli tzw. analiza składowych głównych), o której wypowiemy kilka intuicji, związanych z wektorami własnymi. Oto prosty przykład. Załóżmy, że zmierzyliśmy wzrost i wagę pewnej grupy osób. Chcielibyśmy zaprezentować wyniki naszych pomiarów na płaszczyźnie, umieszczając gwiazdkę w miejscu każdego z wyników. Współrzędna pozioma opisuje wzrost, a pionowa — wagę. Oto przykładowa prezentacja wyników.



Parametry wzrostu i wagi są oczywiście silnie skorelowane, czego używa się wprowadzając wielkość zwaną BMI. Oczywiście korelacja nie jest doskonała, ponieważ budowa ciała zależy od szeregu czynników. Przyglądając się jednak danym o bardzo silnej korelacji zmiennych widzimy, że istnieje kierunek, „wokół którego" oscylują silnie skorelowane dane. Kierunek ten nazwać można *osią główną*. Dlaczego to jest ciekawe? Najbardziej skorelowane dane „układają się" wzdłuż pewnej prostej i kierunek tej osi przyjąć za kierunek odniesienia i nadać mu nazwę rozmiaru (standardowego). Skąd wzięliśmy położenie tej prostej? Od każdego punktu reprezentującego pomiar prowadzimy odcinek prostopadły do tej prostej. Dalej liczymy średnią z kwadratów tych odległości. Okaże się, że dla wybranej prostej osiągamy zarówno „maksymalną" wariancję (rozrzut rzutów na prostej) i „minimalny" błąd, spośród wszystkich prostych[2].

Możemy też rozważać drugi kierunek (nie tylko w naszym przypadku) prostopadły do osi głównej, pokazujący kierunek „największych rozbieżności" pomiędzy wagą, a wzrostem. Czy taki kierunek można jakoś interpretować? Nazwijmy go „budową ciała" i zamiast reprezentować wyniki jako pary liczb (wzrost, waga) reprezentujmy je jako pary (rozmiar, budowa). Kluczowa jest prostopadłość — ponieważ w języku statystyki oznacza ona nieskorelowanie zmiennych (a jeśli są to zmienne losowe — to ich niezależność).

Opowieść powyższa może ma jakiś sens, ale czy ma jakiś związek z wektorami i wartościami własnymi? W istocie wyznaczone kierunki są wektorami własnymi rzeczywistej symetrycznej macierzy[3] kowariancji. Taka macierz ma rozmiar $d \times d$, gdzie d jest liczbą badanych parametrów. W naszym przypadku ma rozmiar 2×2. Nie wchodząc w szczegóły[4] zobaczmy kilka zestawów danych i ich macierzy kowariancji:



[2]Może poniższy GIF dobrze pokaże o co chodzi: `https://i.stack.imgur.com/Q7HIP.gif`. Patrz też: `https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues`.

[3]Takie macierze mają prostopadłe wektory własne, co będzie jednym z ważniejszych twierdzeń tego wykładu.

[4]Odsyłam do wykładów ze Statystyki II (paragraf 3.2.2, str. 18): `https://mst.mimuw.edu.pl/wyklady/st2/wyklad.pdf`. Niech to Państwa nie przerazi. Więcej intuicji: `https://www.neuraldesigner.com/blog/principal-components-analysis`. Można to tłumaczyć nawet bez matematyki: `https://kb.osu.edu/bitstream/handle/1811/69494/ASK_1998_83_102.pdf`.

1.4 Coda. Wzbogacanie podstawowych struktur algebraicznych

Drugi semestr naszego kursu kontynuuje teorię przestrzeni liniowych, ukazując zwłaszcza ich centralną rolę w procesie algebraizacji geometrii (przynajmniej tej elementarnej), ujętej pierwotnie przez Kartezjusza w ramach geometrii analitycznej, a później przez wielu autorów w strukturach afinicznych, rzutowych, analitycznych itd. Czytelnik śledzący podstawowe rezultaty uzyskane w pierwszym semestrze może wyrobić sobie pogląd, że mówimy o swego rodzaju „teorii wszystkiego" – przestrzenią liniową może być przestrzeń ciągów, wielomianów, funkcji, zbiór podzbiorów ustalonego zbioru, a przekształcenia liniowe pomiędzy nimi prowadzą do ważnych wyników głównie w tym sensie, że dają pewne poczucie unifikacji w dość „prymitywnym" języku pozwalającym przypisać rozmaitym „niedużym" obiektom skończony „wymiar" i pewien „układ współrzędnych", w którym wykonywać można czytelniej pewne operacje.

Podobne wrażenie można jednak odnieść choćby po wstępie do matematyki, gdzie formułowane były podstawy teorii mnogości i podstawowe pojęcia, takie jak relacja, funkcja, równoliczność, porządek, klasa równoważności, stanowiące swego rodzaju słownik pojęć matematycznych, obecnych we wszelkich jej dziedzinach. Dopiero ukonkretnienie tematyki tego kursu – choćby koncentracja bądź to na zbiorach skończonych, bądź to na liczbach porządkowych i kardynalnych, bądź to na teorii porządków, bądź to na teorii struktur ilorazowych – prowadzi do konkretnych teorii matematycznych, na przykład kombinatoryki, teorii nieskończoności, algebry uniwersalnej. Nie są to jednak teorie (łącznie z pierwszą) ujmowane w podstawowym kursie matematyki akademickiej, z uwagi na powszechne (jak dotąd) przekonanie, że prowadzą one do na tyle izolowanych od reszty matematyki teorii, że ich wykład nie byłby na tym etapie skuteczny z punktu widzenia budowania pewnej podstawowej świadomości matematycznej.

Nie jest to nic nowego w matematyce. Wiele ogólnych teorii powstaje przez swego rodzaju syntezę metod zgromadzonych wokół partykularnych ważnych problemów (algebra wyrastała chociażby z problemu równania Fermata), a gdy już powstaną – następuje proces eksploracji we wszystkich możliwych kierunkach. Nie wszystkie okazują się istotne z punktu widzenia całej matematyki. Można nawet powiedzieć, że zainteresowanie niektórymi kierunkami badań po pewnym okresie takich eksploracji zamiera – kończą się istotne pytania, istotne odniesienia, związki z innymi działami matematyki (czy nauki w ogóle).

W tej części kursu zawęzimy do pewnego stopnia obszar oglądu teorii przestrzeni liniowych, tak by uwypuklić ich geometryczny charakter, a tym samym – geometryczny charakter wielu (także kombinatorycznych, czy analitycznych) tematów. Uczynimy to przez rozważanie dodatkowych struktur na przestrzeniach liniowych. Będą to między innymi:

- dodatkowa struktura algebry (mnożenie wektorów),
- dodatkowa struktura afiniczna (parze wektorów przypisujemy „wektor łączący"),
- dodatkowa struktura wieloliniowa (n-tce wektorów przypisujemy skalar).

Dodatkowa struktura sprawia, że można wprowadzać nowe pojęcia i własności o charakterze geometrycznym (np. prostopadłość, miara). Można też uzasadnić ograniczenie się do pewnych konkretnych klas obiektów czy ich przekształceń (np. macierze symetryczne lub izometrie). Wyniki klasyfikacyjne nie dotyczą wówczas wprawdzie ogólnych przestrzeni liniowych, ale dają odpowiedzi na konkretne ważne zagadnienia. Tego typu podejście jest uniwersalne nie tylko dla algebry – teorię zbiorów wzbogacić można o strukturę przestrzeni topologicznej, uwzględniającą abstrakcyjne ujęcie zbiorów otwartych, domkniętych, a w szczególnych sytuacjach — pojęcia ciągłości, zbieżności, spójności, zwartości itd. Również w teorii funkcji struktury liniowe wzbogacamy strukturami analitycznymi. W ten sposób we współczesnej matematyce w ramach jednej działalności (choćby geometrycznej) uwzględniać trzeba cały szereg metod i teorii, a także towarzyszących danym obiektom struktur, wybieranych w zależności od potrzeb.

Gdy skonstruujemy pewną liczbę takich „ubogaconych struktur", czy może należałoby powiedzieć – różnych geometrii, przekonamy się, że możliwa jest niekiedy pewna ich hierarchizacja, osiągana poprzez rozważanie tzw. grup niezmienników, czyli przekształceń nie zmieniających nowej struktury. Tak, jak przekształcenia liniowe „nie naruszają" struktury liniowej (choć mogą mocno zmienić same własności przestrzeni liniowych, na które działają), tak i teoria endomorfizmów, automorfizmów, izometrii czy przekształceń afinicznych daje pewne możliwości porównywania uzyskiwanych struktur. Wreszcie – wzbogacać można nie tylko strukturę, ale i działania. Wkrótce powiemy w jednym z dodatków (i zapewne na ćwiczeniach) o tym, że wprowadzona dziś relacja podobieństwa jest przykładem tzw. działania grupy na zbiorze – tutaj: grupy automorfizmów na algebrze endomorfizmów. Język działania jednych struktur na inne stoi w centrum współczesnej matematyki, zwłaszcza tej ujmowanej w języku algebraicznym.

Rozdział 2

Wielomian charakterystyczny. Diagonalizowalność

Od tej pory zakładamy aż do odwołania, że rozważane przestrzenie liniowe są skończonego wymiaru. Pierwszą część wykładu poświęcimy problemowi znajdowania wartości własnych endomorfizmu.

Definicja 2.1: Wielomian charakterystyczny macierzy/endomorfizmu

Dla macierzy $A \in M_n(K)$ WIELOMIANEM CHARAKTERYSTYCZNYM A nazywamy:

$$w_A(\lambda) = \det(A - \lambda I) \in K[\lambda].$$

Niech $\phi \in \mathrm{End}(V)$, gdzie V jest przestrzenią skończenie wymiarową nad ciałem K. Niech $\mathcal{A}$ będzie dowolną bazą przestrzeni V. WIELOMIANEM CHARAKTERYSTYCZNYM endomorfizmu ϕ nazywamy

$$w_\phi(\lambda) = \det(M(\phi)^{\mathcal{A}}_{\mathcal{A}} - \lambda I) \in K[\lambda].$$

Obserwacja 2.1

Wielomian charakterystyczny endomorfizmu $\phi \in \mathrm{End}(V)$ nie zależy od wyboru bazy, to znaczy: dla dowolnych baz $\mathcal{A}$, $\mathcal{B}$ przestrzeni skończenie wymiarowej V mamy

$$\det(M(\phi)^{\mathcal{A}}_{\mathcal{A}} - \lambda I) = \det(M(\phi)^{\mathcal{B}}_{\mathcal{B}} - \lambda I).$$

Jeśli $A, B \in M_n(K)$ są podobne oraz $A = M(\phi)^{st}_{st}$, $B = M(\psi)^{st}_{st}$, to $w_\phi(\lambda) = w_\psi(\lambda)$.

Dowód. Niech C będzie macierzą odwracalną taką, że $C^{-1}AC = B$. Wówczas także

$$C^{-1}(A - \lambda I)C = C^{-1}AC - C^{-1}\lambda I \cdot C = B - \lambda I.$$

Teza wynika zatem ze wzoru Cauchy'ego:

$$\det(B - \lambda I) = \det(C^{-1}(A - \lambda I)C) = \det(C^{-1}) \det(A - \lambda I) \det C = \det(A - \lambda I).$$

□

Przykłady.

- Dla $\phi \in \mathrm{End}(\mathbb{R}^3)$ danego wzorem

$$\phi((x_1, x_2, x_3)) = (x_1 + 2x_2 + 2x_3, 2x_1 + x_2 + 2x_3, 2x_1 + 2x_2 + x_3)$$

mamy $w_\phi(\lambda) = \begin{vmatrix} 1-\lambda & 2 & 2 \\ 2 & 1-\lambda & 2 \\ 2 & 2 & 1-\lambda \end{vmatrix} = -\lambda^3 + 3\lambda^2 + 9\lambda + 5.$

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

$$A - \lambda I = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{bmatrix},$$

a zatem

$$w_A(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 4 - 6 = \lambda^2 - 5\lambda - 2.$$

- Niech $A \in M_n(K)$ będzie macierzą górnotrójkątną, dolnotrójkątną lub diagonalną, która na przekątnej ma kolejno wyrazy $a_{11}, \ldots, a_{nn}$. Wówczas $w_A(\lambda) = (a_{11} - \lambda) \cdot (a_{22} - \lambda) \cdot \ldots \cdot (a_{nn} - \lambda)$.

Zwróćmy uwagę, że w dowodzie pojawił się następujący ważny fakt, z którego jeszcze skorzystamy.

Wniosek 2.1

Niech $A, B \in M_n(K)$. Wówczas dla dowolnego $\lambda \in K$ równoważne są warunki:

- macierze A, B są podobne,
- macierze $A - \lambda I$ oraz $B - \lambda I$ są podobne.

Co więcej, jeśli zachodzi dowolny z powyższych warunków, to istnieje taka macierz odwracalna $C \in M_n(K)$, że $B = C^{-1}AC$ oraz $B - \lambda I = C^{-1}(A - \lambda I)C$.

Uwaga: macierze $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ mają równe wielomiany charakterystyczne $(1-\lambda)^2$, ale nie są podobne.

Obserwacja 2.2

Niech $A = [a_{ij}] \in M_n(K)$. Zachodzi równość:

$$w_A(\lambda) = \det(A - \lambda I) = (-1)^n \lambda^n + (-1)^{n-1} \operatorname{tr}(A) \lambda^{n-1} + \ldots + \det A.$$

Dowód. Dla $n = 1$ teza jest oczywista. Dla $n = 2$ mamy:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = \lambda^2 - \operatorname{tr} A \cdot \lambda + \det A.$$

Korzystamy ze wzoru permutacyjnego na wyznacznik. Wiemy, że $\det(A - \lambda I)$ jest sumą $n!$ elementów:

$$\operatorname{sgn}(\sigma) \cdot x_{\sigma(1)1} x_{\sigma(2)2} \cdot x_{\sigma(n)n},$$

indeksowanych permutacjami $\sigma \in S_n$. Każdy z nich jest iloczynem wyrazów macierzy A oraz tylu elementów postaci $(a_{kk} - \lambda)$, ile jest punktów stałych permutacji σ, czyli takich k, że $\sigma(k) = k$. Zauważmy, że każda permutacja zbioru n elementowego ma albo n punktów stałych, czyli jest identycznością, albo ma $\leqslant n - 2$ punkty stałe. Stąd

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda) + v(\lambda), \qquad (*)$$

gdzie v jest wielomianem stopnia co najwyżej $n - 2$.

W szczególności pokazaliśmy, że $w_A(\lambda)$ jest wielomianem stopnia równego n. Co więcej, na mocy $(*)$ widzimy, że aby znaleźć współczynniki wielomianu $w_A(\lambda)$ stojące przy λ^n oraz λ^{n-1} wystarczy znaleźć analogiczne współczynniki wielomianu $(a_{11} - \lambda)(a_{22} - \lambda) \ldots (a_{nn} - \lambda)$. Łatwo widzieć, że współczynniki te są równe odpowiednio

$$(-1)^n \quad \text{oraz} \quad (-1)^{n-1}(a_{11} + a_{22} + \ldots + a_{nn}) = (-1)^{n-1} tr(A).$$

Pozostaje zauważyć, że dla $\lambda = 0$ wartość funkcji $\det(A - \lambda I)$ równa jest $\det(A)$. □

Naszym kolejnym celem jest pokazanie, że wartościami własnymi endomorfizmu przestrzeni skończenie wymiarowej są dokładnie pierwiastki jego wielomianu charakterystycznego.

Obserwacja 2.3

Niech ϕ będzie endomorfizmem skończenie wymiarowej przestrzeni liniowej V nad ciałem K. Następujące warunki są równoważne dla $a \in K$:

(i) a jest wartością własną ϕ,

(ii) a jest pierwiastkiem wielomianu charakterystycznego $w_A(\lambda)$.

Niech $\mathcal{A}$ będzie bazą przestrzeni V. Załóżmy, że niezerowy wektor $\alpha \in V$ ma w bazie $\mathcal{A}$ współrzędne $x_1, \ldots, x_n$. Wówczas α jest wektorem własnym ϕ o wartości własnej a wtedy i tylko wtedy, gdy

$$(A - aI) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Dowód. Następujące warunki są równoważne dla wektora $v \in V$ oraz $\lambda \in K$:

$$\phi(v) = a \cdot v \iff \phi(v) - av = 0 \iff (\phi - a\,\mathrm{id})(v) = 0 \iff v \in \ker(\phi - a\,\mathrm{id}).$$

A zatem a jest wartością własną endomorfizmu ϕ wtedy i tylko wtedy, gdy $\ker(\phi - a\,\mathrm{id}) \neq 0$, bowiem wektor własny musi być niezerowy. Równoważnie: $\det(\phi - a\,\mathrm{id}) = 0$. Niech $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$. Wówczas dla każdego $a \in K$ mamy:

$$M(\phi - a \cdot \mathrm{id})^{\mathcal{A}}_{\mathcal{A}} = A - aI.$$

Zatem $\ker(\phi - \lambda\,\mathrm{id}) \neq 0$ wtedy i tylko wtedy, gdy jednorodny układ równań liniowych o macierzy $A - aI$ ma niezerowe rozwiązanie. Na mocy Twierdzenia 1.1 widzimy więc, że warunek (i) jest równoważny z (ii). □

Wniosek 2.2

Dla każdego $a \in K$ mamy $V_{(a)} = \ker(\phi - a \cdot \mathrm{id}_V)$, czyli $V_{(a)}$ składa się z wektorów, których współrzędne w bazie $\mathcal{A}$ są wektorami stanowiacymi przestrzeń rozwiązań układu równań liniowych o macierzy $A - aI$.

Z Wniosku 2.2 dostajemy klasyfikację macierzy diagonalnych, z dokładnością do podobieństwa.

Wniosek 2.3

Macierze diagonalne $D_1, D_2 \in M_n(K)$ są podobne wtedy i tylko wtedy, gdy mają identyczne wielomiany charakterystyczne, czyli gdy różnią się co najwyżej kolejnością wyrazów na przekątnej.

Z Obserwacji 2.2 i wzorów Viete'a dostajemy następujący bardzo istotny wniosek.

Wniosek 2.4

Jeśli wielomian charakterystyczny $w_A(\lambda)$ macierzy $A \in M_n(K)$ ma n pierwiastków w ciele K, to:

- suma wartości własnych macierzy A (uwzględniając krotności) równa jest $\mathrm{tr}\, A$,
- iloczyn wartości własnych macierzy A (uwzględniając krotności) równy jest $\det A$.

Przejdziemy teraz do rozważania najbardziej podstawowego problemu teorii endomorfizmów — które endomorfizmy mają w pewnej bazie macierze diagonalne. Odnotujmy najpierw zupełnie oczywiste obserwacje, wiążące ten problem z istnieniem bazy z wektorów własnych endomorfizmu.

Obserwacja 2.4

Niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ będzie bazą przestrzeni liniowej V i niech $\phi \in \mathrm{End}(V)$. Następujące warunki są równoważne

(i) macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest diagonalna

(ii) wektory $\alpha_1, \ldots, \alpha_n$ są wektorami własnymi endomorfizmu ϕ.

Dowód. Załóżmy (i). Niech $a_1, a_2 \ldots, a_n$ będą kolejnymi wyrazami na diagonali macierzy $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$, tzn.

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{bmatrix}.$$

Z definicji macierzy przekształcenia liniowego mamy: $\phi(\alpha_i) = a_i \cdot \alpha_i$, dla $1 \leqslant i \leqslant n$. W szczególności baza $\mathcal{A}$ jest bazą złożoną z wektorów własnych endomorfizmu ϕ, czyli dostajemy (ii). Druga implikacja jest równie oczywista. Jeśli baza $\mathcal{A}$ złożona jest z wektorów własnych, to współrzędne wektora $\phi(\alpha_j)$ w bazie $\mathcal{A}$ są zerowe, poza być może j-tą współrzędną (przy α_j). □

Definicja 2.2: Endomorfizm diagonalizowalny

Niech $\phi \in \mathrm{End}(V)$. Mówimy, że endomorfizm ϕ jest DIAGONALIZOWALNY, jeśli istnieje baza przestrzeni liniowej V złożona z wektorów własnych endomorfizmu ϕ.

Dobrze nam już znane (choć niejedyne) przykłady endomorfizmów diagonalizowalnych to identyczność, jednokładność, rzut oraz symetria. Podstawowym przykładem endomorfizmu niediagonalizowalnego jest obrót (o kąt różny od wielokrotnosci π) w przestrzeni rzeczywistej.

Obserwacja 2.5

Niech $\dim(V) < \infty$. Endomorfizm $\phi \in \mathrm{End}(V)$ jest diagonalizowalny wtedy i tylko wtedy, gdy ma w pewnej bazie macierz diagonalną.

Przeformułujmy pojęcia diagonalizowalności z języka endomorfizmów do języka relacji podobieństwa.

Definicja 2.3: Macierz diagonalizowalna nad ciałem K

Macierz $A \in M_n(K)$ nazywamy DIAGONALIZOWALNĄ NAD CIAŁEM K, jeśli istnieje taka macierz odwracalna $C \in M_n(K)$, że macierz $C^{-1}AC$ jest diagonalna. Innymi słowy, macierz A jest podobna do macierzy diagonalnej należącej do $M_n(K)$.

Powyższe definicje są oczywiście równoważne. Macierz A jest diagonalizowalna (nad ciałem K) wtedy i tylko wtedy, gdy $\phi \in \mathrm{End}(V)$ zadany przez $A = M(\phi)^{st}_{st}$ jest diagonalizowalny.

Zwróćmy uwagę na to, że macierz o wyrazach w ciele K może nie być diagonalizowalna nad K, natomiast może być diagonalizowalna nad pewnym ciałem zawierającym K. Klasycznym przykładem jest wielokrotnie już wspominany obrót w $\mathbb{R}^2$ o kąt $90°$. W bazie standardowej endomorfizm ten ma macierz

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

Macierz ta nie jest diagonalizowalna nad $\mathbb{R}$, ale jeśli potraktujemy ją jako macierz o wyrazach zespolonych (z których wszystkie są akurat rzeczywiste), to możemy wskazać podobną do niej macierz diagonalną:

$$\begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{i}{2} & \frac{1}{2} \\ -\frac{i}{2} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{1}{2} & \frac{i}{2} \end{bmatrix} \cdot \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}.$$

Diagonalizowalność jest niezwykle elegancką własnością. Podstawową jej motywacją jest problem badania w jaki sposób zachowują się kolejne iteracje endomorfizmu. Przypomnijmy ważną kwestię.

Obserwacja 2.6

Niech V będzie przestrzenią wymiaru n nad ciałem K. Dla dowolnego $\phi \in \mathrm{End}(V)$ oraz dla dowolnej bazy $\mathcal{A}$ endomorfizmu ϕ macierz n-krotnej iteracji endomorfizmu ϕ w bazie $\mathcal{A}$ jest równa $\left(M(\phi)^{\mathcal{A}}_{\mathcal{A}}\right)^n$.

Zobaczmy przykład ważnego zastosowań diagonalizowalności w kontekście następującego faktu.

Obserwacja 2.7

Niech $A = C^{-1}DC$, gdzie D jest macierzą diagonalną oraz C – macierzą odwracalną. Wówczas:

$$A^n = (C^{-1}DC)^n = C^{-1}DCC^{-1}DCC^{-1}\ldots CC^{-1}DC = C^{-1}D^nC.$$

Macierz D^n jest diagonalna i ma na przekątnej n-te potęgi odpowiednich elementów z przekątnej D.

Zobaczmy przykład zastosowania tej obserwacji. Niech (F_n) będzie ciągiem Fibonacciego. Rozważmy endomorfizm ϕ przestrzeni liniowej $\mathbb{R}^2$) dany wzorem $\phi(x_1, x_2) = (x_2, x_1 + x_2)$. Wówczas

$$\phi(F_{n-1}, F_n) = (F_n, F_{n+1}), \quad \text{czyli} \quad (F_n, F_{n+1}) = \phi^n(F_0, F_1) = \phi^n(0, 1).$$

Wartościami własnymi ϕ są

- $\lambda_+ = \frac{1+\sqrt{5}}{2}$ o wektorze własnym $\alpha_+ = \left(1, \frac{1+\sqrt{5}}{2}\right)$,
- $\lambda_- = \frac{1-\sqrt{5}}{2}$ o wektorze własnym $\alpha_- = \left(1, \frac{1-\sqrt{5}}{2}\right)$.

Przy tym $(0, 1) = \frac{1}{\sqrt{5}}(\alpha_+ - \alpha_-)$. Stąd:

$$(F_n, F_{n+1}) = \phi^n(0, 1) = \phi^n\left(\frac{1}{\sqrt{5}}(\alpha_+ - \alpha_-)\right) = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n \alpha_+ - \left(\frac{1-\sqrt{5}}{2}\right)^n \alpha_-\right).$$

Rachunek powyższy przełożyć można na język macierzy:

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix} \qquad (*).$$

A zatem jeśli uda nam się wyznaczyć dowolną potęgę macierzy

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

to będziemy mieli kontrolę nad wyrazem ogólnym ciągu Fibonacciego. Mamy $w_A(\lambda) = \lambda^2 - \lambda - 1$, stąd wartości własne macierzy A to

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

Macierz A jest natomiast podobna do macierzy diagonalnej:

$$A = S^{-1} \cdot \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \cdot S.$$

Wyznaczywszy S^{-1} (ma w kolumnach wektory własne A) oraz S uzyskujemy z $(*)$ słynny wzór Bineta:

$$F_n = \frac{\lambda_1^n - \lambda_2^n}{\sqrt{5}}.$$

W dalszej części wykładu przyjrzymy się ogólniej macierzom rekurencji liniowych.

Naszym celem na ten wykład będzie dowód następującego rezultatu.

Twierdzenie 2.1: Charakteryzacja endomorfizmów diagonalizowalnych

Niech $\phi \in \mathrm{End}(V)$, gdzie V – skończenie wymiarowa nad K. Niech $a_1, \ldots, a_k \in K$ będą wszystkimi parami różnymi wartościami własnymi endomorfizmu ϕ. Następujące warunki są równoważne:

(a) ϕ jest diagonalizowalny,

(b) $V_{(a_1)} \oplus V_{(a_2)} \oplus \ldots \oplus V_{(a_k)} = V$,

(c) $\dim V_{(a_1)} + \dim V_{(a_2)} + \ldots + \dim V_{(a_k)} = \dim V$.

Jeśli spełniony jest dowolny z warunków (a)-(c), to

$$w_\phi(\lambda) = (-1)^{\dim V} \cdot (\lambda - a_1)^{\dim V_{(a_1)}} \cdot \ldots \cdot (\lambda - a_k)^{\dim V_{(a_k)}}.$$

Skoro mówimy o bazach złożonych z wektorów własnych, spróbujmy najpierw zrozumieć zależność pomiędzy wektorami własnymi endomorfizmu odpowiadającymi parami różnym wartościom własnym.

Obserwacja 2.8

Niech V będzie przestrzenią liniową nad ciałem K, niech $\phi \in \mathrm{End}(V)$, oraz niech:

- $a_1, a_2, \ldots, a_k \in K$ – parami różne wartości własne ϕ,
- $\beta_1 \in V_{(a_1)}, \beta_2 \in V_{(a_2)}, \ldots, \beta_k \in V_{(a_k)}$ (czyli $\phi(\beta_i) = a_i \cdot \beta_i$, dla $1 \leqslant i \leqslant k$).

Wówczas jeśli $\beta_1 + \ldots + \beta_k = 0$, to $\beta_1 = \beta_2 = \ldots = \beta_k = 0$.

Dowód. Stosujemy indukcję po k. Dla $k = 1$ teza jest oczywista. Przechodzimy do kroku indukcyjnego, zakładając, że teza jest spełniona dla $k-1$. Niech $\beta_1, \ldots, \beta_k$ spełniają założenia i niech $\beta_1 + \ldots + \beta_k = 0$. Stosując ϕ do obydwu stron tej równości dostajemy:

$$0 = \phi(\beta_1 + \ldots + \beta_k) = \phi(\beta_1) + \ldots + \phi(\beta_k) = a_1\beta_1 + \ldots + a_k\beta_k.$$

Możemy także przemnożyć wyjściowe założenie przez a_1 i dostać

$$a_1\beta_1 + \ldots + a_1\beta_k = 0$$

A ta i poprzednia równość dają łącznie $a_1\beta_1 + a_1\beta_2 + \ldots + a_1\beta_k = a_1\beta_1 + a_2\beta_2 + \ldots + a_k\beta_k$, czyli:

$$(a_2 - a_1)\beta_2 + \ldots + (a_k - a_1)\beta_k = 0.$$

Składnik $(a_i - a_1)\beta_i$ należy do $V_{(a_i)}$, dla $i = 2, \ldots, k$, więc z założenia indukcyjnego wynika, że

$$(a_2 - a_1)\beta_2 = (a_3 - a_1)\beta_3 = \ldots = (a_k - a_1)\beta_k = 0.$$

Skoro $a_1, \ldots, a_k$ są parami różne[1], to mamy $\beta_2, \ldots, \beta_k = 0$. A zatem równość $\beta_1 + \ldots + \beta_k = 0$ redukuje się do $\beta_1 = 0$, co kończy krok indukcyjny i cały dowód. □

Rezultat ten prowadzi do fundamentalnego wniosku.

Obserwacja 2.9

Niech V będzie przestrzenią liniową nad ciałem K, niech $\phi \in \mathrm{End}(V)$ oraz niech:

- $a_1, a_2, \ldots, a_k$ – parami różne wartości własne ϕ,
- $\alpha_1, \ldots, \alpha_k$ – wektory własne odpowiadające (odpowiednio) wartościom własnym $a_1, \ldots, a_k$.

Wówczas układ $\alpha_1, \ldots, \alpha_k$ jest liniowo niezależny.

[1] Korzystamy tu z pierwszego faktu o przestrzeniach liniowych: jeśli $a \cdot \alpha = 0$, to $a = 0$ lub $\alpha = 0$, dla $a \in K$ oraz $v \in V$.

Dowód. Przypuśćmy, że $b_1\alpha_1+\ldots+b_k\alpha_k = 0$, dla pewnych skalarów $b_1, \ldots, b_k \in K$. Wówczas $b_i\alpha_i \in V_{(a_i)}$, dla $i = 1, 2, \ldots, k$. Na mocy poprzedniej Obserwacji mamy zatem, ze $b_i\alpha_i = 0$, dla $i = 1, 2, \ldots, k$. Stąd także $b_1 = b_2 = \ldots = b_k = 0$, bowiem $\alpha_i \neq 0$, dla $i = 1, \ldots, k$. □

W dowodzie wykorzystaliśmy założenie, że α_i są wektorami własnymi, więc muszą być niezerowe. Oto kluczowy wniosek.

Wniosek 2.5

Jeśli endomorfizm n wymiarowej przestrzeni liniowej ma n różnych wartości własnych, to jest on diagonalizowalny.

Oczywiście warunek posiadania n różnych wartości własnych nie jest konieczny do diagonalizowalności, bowiem macierz diagonalna może mieć te same wyrazy w różnych miejscach na przekątnej — jak choćby sama macierz identycznościowa. W sytuacji gdy wielomian charakterystyczny ma pierwiastki wielokrotne, diagonalizowalność zależy od wymiarów przestrzeni własnych odpowiadających parami różnym wartościom własnym. Temu zagadnieniu poświęcimy resztę wykładu.

Obserwacja 2.10

Niech $w_\phi(\lambda)$ będzie wielomianem charakterystycznym endomorfizmu ϕ przestrzeni V nad ciałem K oraz niech k będzie krotnością pierwiastka a wielomianu charakterystycznego tego endomorfizmu, czyli takim elementem K, że

$$w_\phi(\lambda) = (\lambda - a)^k \cdot g(\lambda),$$

przy czym a nie jest pierwiastkiem wielomianu g. Wówczas $k \geqslant \dim V_{(a)}$.

Dowód. Niech $\alpha_1, \ldots, \alpha_r$ będzie bazą przestrzeni $V_{(a)}$. Uzupełnijmy ją wektorami $\beta_{r+1}, \ldots, \beta_n$ do bazy $\mathcal{A}$ przestrzeni V. Wówczas macierz ϕ w bazie $\mathcal{A}$ ma postać blokową:

$$A = M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} aI_r & B \\ 0 & C \end{bmatrix},$$

dla pewnych $B \in M_{r\times(n-r)}(K)$, $C \in M_{n-r}(K)$. W szczególności

$$A - \lambda I = \begin{bmatrix} (a-\lambda)I_r & B \\ 0 & C - \lambda I_{n-r} \end{bmatrix}.$$

Zatem (patrz[2] Obserwacja 12.13 w notatkach z GAL I):

$$w_\phi = \det(A - \lambda I) = (a - \lambda)^r \cdot \det(C - \lambda I).$$

Skoro $\det(C - \lambda I)$ jest wielomianem zmiennej λ, to z definicji krotności[3] mamy $k \geqslant r = \dim V_{(a)}$. □

Definicja 2.4: Krotność algebraiczna i krotność geometryczna wartości własnej

Niech a będzie wartością własną endomorfizmu ϕ. Liczbę k występującą w poprzednim lemacie nazywamy KROTNOŚCIĄ ALGEBRAICZNĄ wartości własnej a endomorfizmu ϕ. Liczbę $\dim V_{(a)}$ nazywamy KROTNOŚCIĄ GEOMETRYCZNĄ wartości własnej a.

[2] Chodzi o wyznacznik macierzy blokowo-górnotrójkątnej. Dla $X \in M_m(K)$, $Y \in M_{m\times n}(K)$ oraz $Z \in M_n(K)$ mamy:

$$\begin{vmatrix} X & Y \\ 0 & Z \end{vmatrix} = \begin{vmatrix} I_m & 0 \\ 0 & Z \end{vmatrix} \cdot \begin{vmatrix} I_m & Y \\ 0 & I_n \end{vmatrix} \cdot \begin{vmatrix} X & 0 \\ 0 & I_n \end{vmatrix} = |Z| \cdot 1 \cdot |X| = |X| \cdot |Z|.$$

[3] Uważny Czytelnik może dokładniej zastanowić się nad tym krokiem. Jeśli zachodzi równość dwóch wielomianów z $K[\lambda]$, z których każdy jest podzielny przez ten sam czynnik liniowy, to równe są też ilorazy tych wielomianów przez ów wspólny czynnik. Kluczowe jest tu założenie, że współczynniki są w ciele (ogólniej: w pierścieniu będącym dziedziną całkowitości).

Kluczowy przykład rozróżniający te pojęcia: dla endomorfizmów danych macierzami:

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$$

mamy rowność wielomianów charakterystycznych $(\lambda-1)^2$. A zatem krotność algebraiczna wartości własnej 1 wynosi dla obydwu endomorfizmów 2. Różnicę wnosi krotnosć geometryczna. Dla endomorfizmu identycznościowego równa jest ona 2, a dla endomorfizmu danego macierzą po prawej — wynosi ona 1.

Obserwacja 2.11

Niech $a_1, \ldots, a_k$ będą wartościami własnymi endomorfizmu ϕ przestrzeni V. Wówczas:

$$\dim V_{(a_1)} + \dim V_{(a_2)} + \ldots + \dim V_{(a_k)} \leqslant \dim V.$$

Dowód. Wiemy, że $\dim V = \deg w_\phi(\lambda)$, zaś liczba pierwiastków (z krotnościami) wielomianu o współczynnikach z ciała nie przekracza jego stopnia. Teza wynika zatem z Obserwacji 2.10. □

Przejdźmy do dowodu Twierdzenia 2.1 charakteryzującego endomorfizmy diagonalizowalne.

Dowód. Zaczniemy od pokazania, że każdy układ wektorów przestrzeni V, który jest złożony z k liniowo niezależnych układów wektorów z podprzestrzeni własnych $V_{(a_i)}$ (dla $i = 1, \ldots, k$) jest liniowo niezależny.

Rozważmy więc k układów $\alpha_{i1}, \ldots, \alpha_{im_i}$ liniowo niezależnych w $V_{(a_i)}$ — każdy złożony z m_i wektorów. Przypuśćmy, że:

$$(a_{11}\alpha_{11} + \ldots + a_{1m_1}\alpha_{1m_1}) + (a_{21}\alpha_{21} + \ldots + a_{2m_2}\alpha_{2m_2}) + \ldots + (a_{k1}\alpha_{k1} + \ldots + a_{km_k}\alpha_{km_k}) = 0,$$

dla pewnych $a_{11}, \ldots, a_{km_k} \in K$. Oznaczmy każdą z sum częściowych należących do odpowiedniego $V_{(a_i)}$:

$$\beta_i = a_{i1}\alpha_{i1} + \ldots + a_{im_i}\alpha_{im_i} \in V_{(a_i)} \quad \text{dla} \quad i = 1, \ldots, k.$$

Skoro $\beta_1 + \ldots + \beta_k = 0$, to na mocy Obserwacji 2.8 mamy $\beta_i = 0$, dla wszystkich i. Stąd dostajemy

$$a_{i1}\alpha_{i1} + \ldots + a_{im_i}\alpha_{im_i} = 0.$$

Wobec liniowej niezależności każdego z układów $\alpha_{i1}, \ldots, \alpha_{im_i}$ mamy $a_{i1}, \ldots, a_{im_i} = 0$, dla wszystkich i. A zatem cały rozważany układ złożony z k podukładów pochodzących z V_{a_i} jest liniowo niezależny.

Z obserwacji tej wynika natychmiast, że jeśli zachodzi warunek

$$\dim V_{(a_1)} + \dim V_{(a_2)} + \ldots + \dim V_{(a_k)} = \dim V,$$

to układ

$$\alpha_{11}, \ldots \alpha_{1m_1}, \alpha_{21}, \ldots, \alpha_{2m_2}, \ldots, \alpha_{k1}, \ldots, \alpha_{km_k}$$

jest bazą przestrzeni V złożoną z wektorów własnych endomorfizmu ϕ. W szczególności endomorfizm ten jest diagonalizowalny.

Pozostaje więc wykazać, że jeśli przestrzeń V posiada bazę złożoną z wektorów własnych (czyli gdy ϕ jest endomorfizmem diagonalizowalnym), to zachodzi $\dim V_{(a_1)} + \dim V_{(a_2)} + \ldots + \dim V_{(a_k)} = \dim V$.

Niech $\mathcal{A}$ będzie bazą przestrzeni V złożonej z wektorów własnych endomorfizmu ϕ, mającego wartości własne $a_1, \ldots, a_k$. Dla każdego $i = 1, \ldots, k$ niech $\beta_{i1}, \ldots, \beta_{in_i}$ będą wszystkimi tymi spośród wektorów bazy $\mathcal{A}$, które należą do $V_{(a_i)}$. Wówczas

$$n_1 + \ldots + n_k = \dim V.$$

Ponadto wówczas dla każdego i układ $\beta_{i1}, \ldots, \beta_{in_i}$ jest liniowo niezależny, więc $n_i \leqslant \dim V_{(a_i)}$. Zatem

$$\dim V = n_1 + \ldots + n_k \leqslant \dim V_{(k_1)} + \ldots + \dim V_{(k_r)} \leqslant \dim V,$$

na mocy Obserwacji 2.11. Zatem z diagonalizowalności ϕ mamy $\dim V_{(a_1)} + \ldots + \dim V_{(a_k)} = \dim V$.

Zatem warunki (a) oraz (c) w Twierdzeniu 2.1 są równoważne. Oczywiście (c) natychmiast wynika z (b), a (b) jest oczywistą konsekwencją (a). □

Sklasyfikowaliśmy endomorfizmy diagonalizowalne w języku wymiarów podprzestrzeni własnych. Niestety baza z wektorów własnych nie zawsze istnieje nawet jeśli wielomian charakterystyczny rozkłada się na czynniki liniowe. Co możemy zatem powiedzieć o geometrycznych własnościach takiego endomorfizmu? Jak się okaże, kluczowe będzie rozważenie dodatkowej klasy endomorfizmów — tak zwanych endomorfizmów nilpotentnych. Mają one tę własność, że złożone ze sobą pewną liczbę razy stają się przekształceniem zerowym. Pozwolą one na zrozumienie centralnego wyniku tej części kursu — twierdzenia Jordana.

Na koniec zobaczmy kilka bezpośrednich przykładów zastosowań.

Zadanie. Znajdź wszystkie takie $s, t \in \mathbb{Z}_5$, że macierz

$$A = \begin{bmatrix} 1 & 0 & 0 \\ s & 2 & 0 \\ 1 & t & 1 \end{bmatrix} \in M_{3\times 3}(\mathbb{Z}_5)$$

jest diagonalizowalna nad ciałem pięcioelementowym $\mathbb{Z}_5$.

Rozwiązanie. Niech $\phi : \mathbb{Z}_5^3 \to \mathbb{Z}_5^3$ będzie endomorfizmem takim, że $M(\phi)_{st}^{st} = A$. Teza jest równoważna wskazaniu takich $s, t \in \mathbb{Z}_5$, dla których endomorfizm ϕ jest diagonalizowalny, czyli istnieje baza $\mathbb{Z}_5^3$ złożona z wektorów własnych ϕ. Wielomian charakterystyczny ϕ równy jest wielomianowi charakterystycznemu macierzy dolnotrójkątnej A, czyli $w_\phi(\lambda) = w_A(\lambda) = (1-\lambda)^2(2-\lambda) \in \mathbb{Z}_5[\lambda]$. A zatem ϕ ma wartości własne $1, 2$. Mają one odpowiednio krotności algebraiczne 2 oraz 1. A zatem podprzestrzeń własna $V_{(2)}$ jest wymiaru 1. Aby ϕ było diagonalizowalne potrzeba zatem i wystarcza, by $V_{(1)}$ była wymiaru 2. Mamy:

$$A - I = \begin{bmatrix} 0 & 0 & 0 \\ s & 1 & 0 \\ 1 & t & 0 \end{bmatrix}$$

Układ równań dany powyższą macierzą ma przestrzeń rozwiązań $V_{(1)}$ wymiaru 2 wtedy i tylko wtedy, gdy $r(A-I) = 1$. Jest tak wtedy i tylko wtedy, gdy $st = 1$. A zatem szukane pary $(s, t) \in \mathbb{Z}_5^2$, dla których macierz A jest diagonalizowalna to $(1,1), (2,3), (3,2), (4,4)$. ■

Zadanie. Macierz rzeczywista A rozmiarów 2×2 ma wartości własne 1 i 2. Wyznaczyć

$$\det(A^3 - 3A^2 + A + 5I).$$

Rozwiązanie. Traktujemy A jako macierz endomorfizmu ϕ przestrzeni dwuwymiarowej w bazie standardowej. Wówczas wartość $\det(\phi^3 - 3\phi^2 + \phi + 5\,\mathrm{id})$ nie zależy od wyboru macierzy przekształcenia ϕ. W szczególności, skoro endomorfizm ϕ przestrzeni dwuwymiarowej ma dwie różne wartości własne, to jest on diagonalizowalny (mamy $2 = V_{(1)} + V_{(2)}$) i ma w pewnej bazie macierz $J = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. A zatem

$$\det(A^3 - 3A^2 + A + 5I) = \det(J^3 - 3J^2 + J + 5I) = \begin{vmatrix} 1^3 - 3\cdot 1^2 + 1 + 5 & 0 \\ 0 & 2^3 - 3\cdot 2^2 + 2 + 5 \end{vmatrix} = 12.$$

■

Zadanie. Wyznacz macierz A^{2020}, gdzie $A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} \in M_2(\mathbb{R})$.

Rozwiązanie. Wielomian charakterystyczny macierzy A równy jest

$$\begin{vmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{vmatrix} = (7-\lambda)(-1-\lambda) - 9 = \lambda^2 - 6\lambda - 16 = (\lambda+2)(\lambda-8).$$

A zatem macierz A jest podobna do macierzy diagonalnej:

$$A = P^{-1} \cdot \begin{bmatrix} -2 & 0 \\ 0 & 8 \end{bmatrix} \cdot P,$$

przy czym macierz P^{-1} zawiera w kolumnach wektory własne o wartościach własnych $-2, 8$ — na przykład $(1,-3)$ oraz $(3,1)$. Istotnie, jeśli $\phi \in \mathrm{End}(\mathbb{R}^2)$ zadany jest warunkiem $A = M(\phi)_{st}^{st}$ oraz jeśli przyjmiemy $\mathcal{A} = ((1,-3),(3,1))$, wówczas $D = M(\phi)_{\mathcal{A}}^{\mathcal{A}}$ oraz $P = M(\mathrm{id})_{st}^{\mathcal{A}}$. Zatem $P^{-1} = \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}$. W rezultacie dostajemy:

$$A^{2020} = \left(P^{-1} \cdot \begin{bmatrix} -2 & 0 \\ 0 & 8 \end{bmatrix} \cdot P\right)^{2020} = P^{-1} \cdot \begin{bmatrix} 2^{2020} & 0 \\ 0 & 8^{2020} \end{bmatrix} \cdot P.$$

■

2.1 Uzupełnienie. Probabilistyka na sześciennej kostce

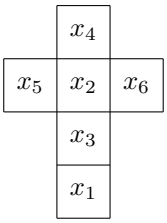
Diagonalizowalność macierzy służy także do badania przebiegu procesów „w dłuższej perspektywie". Jest to niezwykle rozbudowana teoria zawierająca wiele istotnych wyników. Zobaczmy ją na przykładzie.

> ZADANIE. Matematyk ma kostkę sześcienną, na której ścianach umieszczono liczby $N_1, \ldots, N_6$. Zapisuje on sobie, dla każdej ściany, średnią arytmetyczną liczb na sąsiednich czterech ścianach. Następnie zastępuje każdą z liczb N_i uzyskaną średnią, i powtarza tę procedurę. Jakie liczby pojawią się na poszczególnych ścianach po wykonaniu bardzo wielu takich zmian?

Operacja opisana w zadaniu będzie przez nas przetłumaczona na język endomorfizmu przestrzeni liniowych. Ale jakich przestrzeni? Będzie to przestrzeń $\mathbb{R}^6$ złożona z ciągów

$$(x_1, x_2, x_3, x_4, x_5, x_6),$$

gdzie x_i to są liczby wpisane na ścianach kostki jak na rysunku poniżej.



Rozważmy przekształcenie liniowe $f : \mathbb{R}^6 \to \mathbb{R}^6$ zadane macierzą A

$$A = \frac{1}{4}\begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Przemnożenie tej macierzy przez wektor $x = (x_1, \ldots, x_6)^T$ daje dokładnie operację, którą wykonuje się na ściankach kostki. Musimy tę operację wykonać wiele razy, więc interesuje nas jak wygląda, dla dużych n, wektor $A^n x$? Jest więc jasne, że trzeba diagonalizować macierz A. Można to zrobić i dostajemy sześć wektorów własnych:

- $v_1 = (1, 1, 1, 1, 1, 1)$, gdzie $f(v_1) = v_1$,
- $v_2 = (0, 0, 1, 1, -1, -1)$, gdzie $f(v_2) = -\frac{1}{2}v_2$,
- $v_3 = (-2, -2, 1, 1, 1, 1)$, gdzie $f(v_3) = -\frac{1}{2}v_3$,
- $v_4 = (0, 0, -1, 1, 0, 0)$, gdzie $f(v_4) = 0$,
- $v_5 = (0, 0, 0, 0, 1, -1)$, gdzie $f(v_5) = 0$,
- $v_6 = (-1, 1, 0, 0, 0, 0)$, gdzie $f(v_6) = 0$.

W bazie $\mathcal{A} = (v_1, v_2, \ldots, v_6)$ endomorfizm f dany jest macierzą B:

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Biorąc wektor $v = a_1 v_1 + \ldots + a_6 v_6$ widzimy, że

$$B \cdot (a_1 v_1 + \ldots + a_6 v_6) = a_1 v_1 - \frac{1}{2}a_2 v_2 - \frac{1}{2}a_3 v_3.$$

A zatem dla całkowitych dodatnich n mamy

$$B^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \left(-\frac{1}{2}\right)^n & 0 & 0 & 0 & 0 \\ 0 & 0 & \left(-\frac{1}{2}\right)^n & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Innymi słowy

$$B^n v \to a_1 v_1.$$

A jakie są współrzędne interesującego nas wektora $(N_1, \ldots, N_6)$ w bazie $\mathcal{A}$? Trzeba sprawdzić, że pierwszą z nich jest

$$\frac{1}{6}(N_1 + N_2 + \ldots + N_6).$$

Na pierwszy rzut oka może się wydawać, że policzenie tych współrzędnych jest trudne, ale za jakiś czas przekonamy się, że wektory v_i są prostopadłe i będzie nieco łatwiej. Natomiast konsekwencja jest taka:

$$B^n \cdot \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_6 \end{bmatrix} \to \frac{1}{6}(N_1 + N_2 + \ldots + N_6) \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} N_1 + N_2 + \ldots + N_6 \\ N_1 + N_2 + \ldots + N_6 \\ \vdots \\ N_1 + N_2 + \ldots + N_6 \end{bmatrix}.$$

A zatem dla dużych n stanie się to, czego się spodziewamy – nastąpi uśrednienie liczb wpisanych na każdej ścianie. A czy wiecie Państwo jak to pokazać bez takiej skomplikowanej matematyki?

Co to ma wspólnego z tytułową probabilistyką? Czytelnik zechce zauważyć, że wiersze i kolumny rozważanej przez nas macierzy przekształcenia f mają wyrazy nieujemne, których sumy są równe 1. Rozważmy następujący problem.

> Zadanie. Matematyk znalazł się na kostce sześciennej, na ścianie z jednym oczkiem. W każdym ruchu przechodzi losowo na jedną z czterech sąsiednich ścian. Jakie jest prawdopodobieństwo, że po dziesiątym przejściu stanie na ścianie z trzema oczkami? Czy jeśli Matematyk wykona *dużo* takich przejść, to trafi z równym prawdopodobieństwem na dowolną ścianę?

Rozwiązanie pierwszej części problemu polega na przemnożeniu wektora $(1, 0, 0, 0, 0, 0)$ przez macierz A^{10}. Dlaczego? Rozważmy ogólny kontekst, związany z dyskretnymi procesami Markowa, czyli ciągami zdarzeń, w których prawdopodobieństwo kolejnego zdarzenia zależy jedynie od wyniku poprzedniego.

Załóżmy, że mamy skończony zbiór E oraz ciąg $X = (X_n)_{n\in\mathbb{N}}$ przyjmujący wartości w „zbiorze stanów" E. Przykładowo: E może być zbiorem ścianek kostki, zaś X_i — położeniem Matematyka po i-tym skoku. Powiemy (starając się nie narazić zbytnio Probabilistom takim spłyceniem), że dyskretnym łańcuchem Markowa jest taki ciąg X (zmiennych losowych), dla którego dla każdego $n \in \mathbb{N}$ spełniona jest zależność (trzeba sobie przypomnieć prawdopodobieństwo warunkowe):

$$\mathbb{P}(X_{n+1} = s_{n+1} \,|\, X_n = s_n, \ldots, X_1 = s_1, X_0 = s_0) = \mathbb{P}(X_{n+1} = s_{n+1} \,|\, X_n = s_n).$$

Krótko mówiąc chodzi o sytuację, w której prawdopodobieństwo tego co się stanie w następnym kroku zależy wyłącznie od prawdopodobieństwa zajścia obecnego kroku, a bez uwzględnienia prawdopodobieństw zajścia wcześniejszych — jest to swego rodzaju „zapominanie o przeszłości". Czytelnik w przyszłości pozna te problemy dokładniej. Co to ma wspólnego z naszą grą, macierzami i diagonalizacją?

Załóżmy, że chcemy wyznaczyć prawdopodobieństwo tego, że na początku Matematyk jest na ściance s_0, po pierwszym przejściu trafia na s_1, a po drugim — na s_2, czyli:

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, X_2 = s_2).$$

Używając twierdzenia o prawdopodobieństwie całkowitym (szkolnego?) mamy:

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, X_2 = s_2) = \mathbb{P}(X_0 = s_0) \cdot \mathbb{P}(X_1 = s_1 \,|\, X_0 = s_0) \cdot \mathbb{P}(X_2 = s_2 \,|\, X_0 = s_0, X_1 = s_1).$$

Teraz kluczowy jest warunek Markowa, który upraszcza rachunek wyżej, bowiem:

$$\mathbb{P}(X_2 = s_2 \,|\, X_0 = s_0, X_1 = s_1) = \mathbb{P}(X_2 = s_2 \,|\, X_1 = s_1)$$

dając nam

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, X_2 = s_2) = q_0(s_0) \cdot p(s_0, s_1) \cdot p(s_1, s_2),$$

gdzie $q_0(s_i)$ to prawdopodobieństwo, że na początku jesteśmy w stanie s_i oraz

$$\mathbb{P}(X_{n+1} = s_{n+1} \,|\, X_n = s_n) = p(s_n, s_{n+1}).$$

Ogólnie można łatwo pokazać, że prawdopodobieństwo, iż w $n+1$-wszym ruchu procesu Markowa znajdziemy się w stanie s_{n+1} równe jest:

$$q_{n+1}(s_{n+1}) = \sum_{s \in E} p(s, s_{n+1}) \cdot q_n(s).$$

Czy Państwo widzą tu mnożenie macierzy? Rozważmy dowolny zbiór dopuszczalnych stanów $E = \{e_1, \ldots, e_k\}$. Wówczas

$$\begin{bmatrix} q_n(e_1) & \ldots & q_n(e_k) \end{bmatrix} \cdot \begin{bmatrix} p(e_1, e_1) & \ldots & p(e_1, e_n) \\ \vdots & \ddots & \vdots \\ p(e_n, e_1) & \ldots & p(e_n, e_n) \end{bmatrix} = \begin{bmatrix} q_{n+1}(e_1) & \ldots & q_{n+1}(e_k) \end{bmatrix}.$$

To oznacza, że do zrozumienia dynamiki procesu Markowa wystarczy znajomość wektora prawdopodobieństw zajścia odpowiednich stanów na początku procesu, oraz tzw. macierz przejścia, czyli:

$$\begin{bmatrix} q_0(e_1) & \ldots & q_0(e_k) \end{bmatrix} \cdot \begin{bmatrix} p(e_1, e_1) & \ldots & p(e_1, e_n) \\ \vdots & \ddots & \vdots \\ p(e_n, e_1) & \ldots & p(e_n, e_n) \end{bmatrix}^n = \begin{bmatrix} q_{n+1}(e_1) & \ldots & q_{n+1}(e_k) \end{bmatrix}.$$

Czy nasz proces skakania po kalendarzu jest procesem Markowa, dla $E = \{1, 2, 3, 4, 5, 6\}$? To zależy od naszego Matematyka, ale jeśli przyjmiemy, że błąka się on rzeczywiście w sposób „losowy" po naszej kostce, zapominając (lub nie przejmując się) gdzie był krok wcześniej — wówczas możemy działać.

Jeśli na przykład przyjmiemy, że Matematyk wylądował na ściance z jednym oczkiem, to jak policzyć dokąd może dojść? Mówi o tym macierz A. Z prawdopodobieństwem $1/4$ może przejść na pola z odpowiednio trzema, czterema, pięcioma lub sześcioma oczkami (patrz rysunek kostki na początku). A zatem po dziesiątym ruchu wektor prawdopodobieństw położeń Matematyka dla poszczególnych ścian na kostce to:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0,25 & 0,25 & 0,25 & 0,25 \\ 0 & 0 & 0,25 & 0,25 & 0,25 & 0,25 \\ 0,25 & 0,25 & 0 & 0 & 0,25 & 0,25 \\ 0,25 & 0,25 & 0 & 0 & 0,25 & 0,25 \\ 0,25 & 0,25 & 0,25 & 0,25 & 0 & 0 \\ 0,25 & 0,25 & 0,25 & 0,25 & 0 & 0 \end{bmatrix}^{10}.$$

Trudno to policzyć? Trudno. Ale jeśli umielibyśmy zapisać wektor $(1, 0, 0, 0, 0, 0)$ w bazie $\mathcal{A} = (v_1, v_2, \ldots, v_6)$ to historia by była całkiem inna, bowiem do policzenia mielibyśmy iloczyn pewnego wektora przez dziesiątą potęgę znanej nam już macierzy diagonalnej. To by nam ułatwiło zrozumienie losów Matematyka.

O rozważanym tu temacie mówić można naprawdę bardzo dużo. Macierze rzeczywiste rozważane przez nas wyżej nazywamy PODWÓJNIE STOCHASTYCZNYMI, a ich rozmaite własności oraz związane z nimi niezwykle głębokie i ważne twierdzenia poznać można na przykład poprzez lekturę artykułu

Ciesielska, Kowalczyk, *Twierdzenie Perrona-Frobeniusa i jego zastosowanie w algorytmie Page Rank*,
Prace Koła Matematyków Uniwersytetu Pedagogicznego w Krakowie
`https://pracekm.up.krakow.pl/article/view/2712/2379`

Algorytm Page Rank pojawia się nieprzypadkowo. W tym świetnym artykule popularyzującym zagadnienie macierzy o nieujemnych współczynnikach znajdziecie Państwo łańcuchy Markowa (z wieloma przykładami), wspominane w innym dodatku macierze sąsiedztwa, metody ustalania rankingów, tzw. problemy kontroli sił, modele przeżycia (muchy i żaby to tylko początek tej opowieści) oraz właśnie tytułowy Page Rank, opatentowany w w 1998 roku przez dwóch doktorantów Uniwersytetu Stanforda, służący określeniu „wartości" stron internetowych i sporządzanie ich rankingu. Zachęcam do lektury!

2.2 Dodatek. Grafy i ich macierze sąsiedztwa

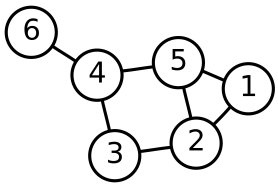
Algebra liniowa niewątpliwie pozwoli nam zrozumieć wiele obiektów geometrycznych i fenomenów w analizie. Chciałbym raz na jakiś czas zasygnalizować inny piękny kierunek jej zastosowań – kombinatorykę (czy mądrzej: matematykę dyskretną). Jednym z podstawowym narzędzi kombinatorycznych są grafy, a do badania ich własności przydają się, jak się okazuje, rozmaite związane z nimi macierze. Badanie własności tych macierzy, między innymi ich wektorów i wartości własnych, może służyć dowodzeniu pięknych twierdzeń, Interpretacja grafowa rozmaitych sytuacji kombinatorycznych jest taka: n osób i ich układ znajomości reprezentuje graficznie graf (nie)zorientowany o n wierzchołkach. Znajomość reprezentowana jest przez krawędź pomiędzy wierzchołkami. Nie ma krawędzi z wierzchołka do niego samego i nie ma krawędzi wielokrotnych pomiędzy dwoma wierzchołkami. Taki graf nazywamy czasem w kombinatoryce GRAFEM PROSTYM. Najpierw pokażmy jak związać z takim grafem pewną (jedną z wielu) macierz.

Definicja 2.5: Macierz sąsiedztwa grafu

Dla (niezorientowanego) grafu prostego G o wierzchołkach postaci $1, 2, \ldots, n$ definiujemy MACIERZ SĄSIEDZTWA $A(G) = [a_{ij}] \in M_n(\mathbb{C})$ warunkami:

- $a_{ij} = 1$, jeśli wierzchołki i oraz j są połączone w G,
- $a_{ij} = 0$, jeśli $i = j$ lub wierzchołki i, j nie są połączone w G.

Rozważmy następujący przykład grafu:



Macierz sąsiedztwa powyższego grafu to:

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

Macierz $A(G)$ (ang. *adjacency matrix*) jest dla grafów niezorientowanych macierzą symetryczną, to znaczy $A(G) = A(G)^T$. Jak się okazuje, rozmaite własności grafu G odczytać można z typowo algebraicznych własności $A(G)$. Osobom mającym doświadczenie w metodach teorii grafów polecam wykład *Metody algebraiczne w teorii grafów* zamieszczony na portalu `wazniak.mimuw.edu.pl`.

ŚCIEŻKA DŁUGOŚCI n w grafie niezorientowanym G to ciąg n krawędzi postaci $a_0 - a_1 - \ldots - a_n$, gdzie $a - b$ reprezentuje połączenie wierzchołków a oraz b. Wierzchołki mogą występować w ścieżce wiele razy. Jest jasne, ze jeśli w grafie jest cykl zamknięty, to żadna z potęg macierzy sąsiedztwa nie jest zerowa. Kluczowym powodem badania macierzy sąsiedztwa jest następująca, prosta obserwacja, której dowód (ładna motywacja dla mnożenia macierzy) pozostawiam Czytelnikowi.

Obserwacja 2.12

Niech G będzie skończonym niezorientowanym grafem prostym. Wyraz znajdujący się w i-tym wierszu i j-tej kolumnie macierzy $A(G)^n$ opisuje liczbę ścieżek długości n od wierzchołka i do wierzchołka j w grafie G.

Dla zilustrowania pewnych ciekawych własności macierzy sąsiedztwa zostawiam poniżej kilka zadań do samodzielnego rozwiązania[4]. Niektóre wymagają całkiem sporej pomysłowości.

[4] Więcej zadań będziemy mogli zrobić w połowie semestru, gdy wykażemy tzw. rzeczywiste twierdzenie spektralne

ZADANIE 1. Wektor stopni wierzchołków (niezorientowanego) grafu prostego G równy jest iloczynowi macierzy $A(G)$ przez wektor v postaci....?

ZADANIE 2. Jeśli G_1, G_2 są parą izomorficznych[5] grafów skończonych o macierzach sąsiedztwa $A(G_1)$, $A(G_2)$, to istnieje macierz permutacyjna P taka, że:

$$A(G_2) = P^{-1}A(G_1)P.$$

W szczególności macierze $A(G_1), A(G_2)$ mają ten sam zbiór wartości własnych (z krotnościami).

ZADANIE 3. Jeśli graf G jest prosty, ale nie jest spójny, wówczas można tak ponumerować jego wierzchołki, aby $A(G)$ było macierzą blokowo-diagonalną o liczbie bloków diagonalnych równej liczbie składowych spójności G.

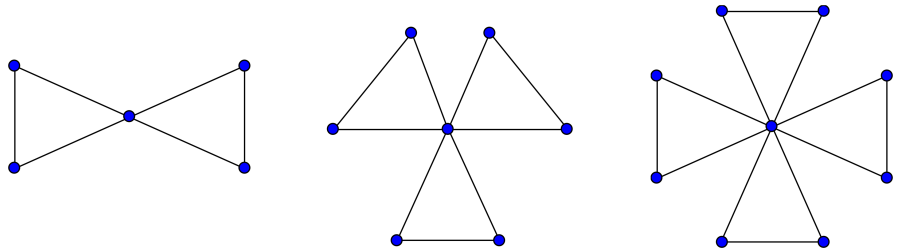
ZADANIE 4. Załóżmy, że graf G jest dwudzielny. Wówczas jeśli λ jest wartością własną $A(G)$, to $-\lambda$ również jest wartością własną $A(G)$. Czy zachodzi fakt odwrotny?

ZADANIE 5. Pokaż, że jeśli λ jest wartością własną macierzy $A(G)$ grafu G, to $|\lambda| \leqslant \Delta(G)$, gdzie $\Delta(G)$ to maksymalny stopień wierzchołka w grafie G.

To oczywiście jedynie bardzo proste własności $A(G)$ związane z wartościami własnymi. Jest ich znacznie więcej, a w zasadzie są całe podręczniki związane z badaniem metod algebraicznych w teorii grafów. Jednym z niezwykle pięknych wyników w tej tematyce jest następujące twierdzenie.

Twierdzenie 2.2: O przyjaźni (Erdős, Rényi, Sós), 1966

Załóżmy, że w gronie $n > 1$ osób każde dwie osoby mają dokładnie jednego wspólnego znajomego (A jest znajomym $B \iff B$ jest znajomym A). Wówczas istnieje w tym gronie osoba, która jest znajomym wszystkich pozostałych.



Przykłady tzw. grafów przyjaźni, czyli takich, które realizują tezę twierdzenia o przyjaźni. Źródło `wikipedia.org`.

To dość zaskakujące twierdzenie jest jednym z wielu pięknych zastosowań tzw. spektralnej teorii grafów. Istnieje też trudny czysto kombinatoryczny dowód tego rezultatu. Poniżej przedstawiam dowód algebraiczny. W tym momencie może on nie być do końca zrozumiały, ale proszę wrócić w to miejsce za jakieś dwa wykłady i być wszystko stanie się bardziej jasne. Już teraz jednak można z pożytkiem śledzić piękny pomysł dowodu. Niech G będzie grafem prostym opisującym układ przyjaźni n osób z twierdzenia.

Oto plan dowodu w czterech krokach[6].

1. Pokażemy, że w opisanej sytuacji dwie osoby nieznajome mają taką samą liczbę znajomych.
2. Pokażemy, że gdyby teza nie zachodziła, to wszyscy musieliby mieć taką samą liczbę znajomych.
3. Niech K będzie grafem k-regularnym (z każdego wierzchołka wychodzi k krawędzi) o n wierzchołkach, opisującym kontrprzykład do twierdzenia. Pokażemy, ze $n = k^2 - k + 1$.
4. Wyznaczymy wartości własne $A(G)^2$. Co to powie o wartościach własnych $A(G)$ oraz o n i k?

[5] To znaczy: G_1, G_2 mają po n wierzchołków, dla pewnego n, i istnieje permutacja $\sigma \in S_n$ taka, że (i, j) jest krawędzią G_1 wtedy i tylko wtedy, gdy $(\sigma(i), \sigma(j))$ jest krawędzią w G_2

[6] Na podstawie notatek dr R. Quinlan ze spektralnej teorii grafów: *Spectral Graph Theory MA500-1: Lecture Notes*, `https://maths.nuigalway.ie/~rquinlan/teaching/spectralgraphtheory.pdf`.

Dowód. Załóżmy, że sytuacja opisana w twierdzeniu reprezentowana jest przez pewien skończony niezorientowany graf prosty G (dla każdych dwóch wierzchołków G istnieje trzeci, z którym te dwa są połączone), ale wbrew tezie nie istnieje jeden wierzchołek G, który byłby połączony ze wszystkimi. Kroki 1. i 2. polegają na pokazaniu, że w opisanej sytuacji G musi być REGULARNY, to znaczy: z każdego wierzchołka wychodzi tyle samo krawędzi, powiedzmy k. Innymi słowy: jeśli nie zachodzi sytuacja opisana w zadaniu to wszyscy muszą mieć tyle samo przyjaciół.

Weźmy dwa niepołączone wierzchołki u, v grafu G reprezentujące dwóch nieznajomych. Niech $k = \deg u$ będzie stopniem u (liczbą połączeń z innymi wierzchołkami) i niech $x_1, \ldots, x_k$ będą sąsiadami u, gdzie x_1 jest owym unikatowym znajomym u oraz v. Dla każdego i od 1 do k, niech y_i będzie owym dokładnie jednym znajomym v oraz x_i. Wtedy wszystkie y_i są parami różne, bo gdyby któreś dwa z nich się pokrywały, to ów wierzchołek gdzie by się pokrywały miałby więcej niż jednego wspólnego znajomego z u. Zatem v ma stopień co najmniej k i $\deg u \leqslant \deg v$. Ten sam argument działa jednak, gdy zamienimy role u oraz v, więc dostajemy również $\deg u \leqslant \deg v$, a w rezultacie $\deg u = \deg v$. Punkt 1. w naszym planie jest zrealizowany. W opisanej sytuacji dwie osoby nieznajome mają taką samą liczbę znajomych.

Niech teraz w będzie dowolnym wierzchołkiem G, innym niż x_1. Skoro u oraz v mają tylko jednego wspólnego znajomego (właśnie x_1), to w nie jest połączony z jednym z nich. Skoro $\deg u = \deg v = k$, to także $\deg w = k$, na mocy punktu 1. A zatem wszystkie wierzchołki G mają stopień k, być może poza x_1. Jeśli, wbrew tezie, istnieje wierzchołek G, który nie ma połączenia z x_1, to ten wierzchołek ma stopień k (to pokazaliśmy), a więc i x_1 ma stopień k. A zatem jeśli graf spełnia założenia twierdzenia, ale nie jego tezę, musi być regularny, co daje nam krok 2.

Załóżmy teraz, że K jest grafem k-regularnym spełniającym warunki twierdzenia i policzmy ile K ma wierzchołków. Zliczamy ścieżki długości 2 grafu K startujące z ustalonego wierzchołka x. Ich liczba to k^2, bo każdy ma k znajomych, a każdy z nich dokładnie k kolejnych. Między każdym wierzchołkiem $y \neq x$ istnieje dokładnie jedna ścieżka długości 2 zaczynająca się w x (ta obserwacja jest kluczowa do pokazania kroku kombinatorycznego). A zatem licząc ile jest wszystkich wierzchołków grafu za pomocą ścieżek długości 2 każdy wierzchołek G policzyliśmy raz — oprócz x, który policzyliśmy k razy. Odejmujemy nadmiarowe $k-1$ razy i widzimy, że

$$n = k^2 - k + 1$$

gdzie n to liczba wierzchołków grafu K. Krok 3. jest zakończony i możemy przejść do wielkiego finału.

Załóżmy, że G jest grafem opisującym sytuację w twierdzeniu i nie spełnia on tezy. Zgodnie z krokiem 2., że G jest regularny stopnia k. Przejdźmy do macierzy sąsiedztwa $A = A(G)$. Co o niej wiemy? Nie za wiele. Ale wiemy coś o A^2. Skoro każde dwa wierzchołki grafu G mają dokładnie jednego wspólnego sąsiada, to macierz A^2 ma k na przekątnej (idziemy do jednego z k sąsiadów i wracamy) oraz 1 wszędzie indziej, bo 1 jest liczbą ścieżek długości 2 pomiędzy różnymi wierzchołkami w G.

Policzenie wielomianu charakterystycznego macierzy A^2 wyżej jest ładnym i pouczającym ćwiczeniem. Zapisujemy $A^2 = (k-1)I + J$, gdzie I jest macierzą identycznościową, zaś J jest macierzą, której wszystkie wyrazy to 1. Wartości własne A^2 to pierwiastki wielomianu

$$\det((k-1)I + J - \lambda I) = \det((k-1-\lambda)I + J).$$

A zatem liczba λ_0 jest wartością własną A^2 wtedy i tylko wtedy, gdy $\lambda_0 + k - 1$ jest wartością własną J. Czytelnik zechce sprawdzić, że wartościami własnymi macierzy J złożonej z samych jedynek są n (wektor własny z samych jedynek) oraz 0, będące pierwiastkiem $n-1$-krotnym. A zatem wartościami własnymi macierzy A^2 są $k-1$, która występuje $n-1$ razy oraz $n+k-1 = k^2$ (wzór wyżej na liczbę wierzchołków), która jest pierwiastkiem pojedynczym. Po co to wszystko?

„Crux" polega na tym, że wartości własne A to pierwiastki kwadratowe[7] wartości własnych A^2. A zatem wartości własne A mogą wynosić jedynie k lub $\pm\sqrt{k-1}$. I kluczowa myśl: wartości własne macierzy A dodają się (z krotnościami) do zera, zgodnie z Wnioskiem 2.1, a na przekątnej A są zera, czyli mamy $tr(A) = 0$. Zatem jeśli $\sqrt{k-1}$ ma (jako wartość własna A) krotność r, zaś $-\sqrt{k-1}$ ma krotność s, to

$$k + (r-s)\sqrt{k-1} = 0.$$

Analiza teorio-liczbowa tego warunku mówi, że $k-1$ dzieli k^2, czyli k musi być równe 1 lub 2. W tych przypadkach G jednak ma wspólnego „przyjaciela" innych wierzchołków. Dostaliśmy sprzeczność. □

[7] Jeśli $Av = \lambda v$, to $A^2 v = A(Av) = A \cdot \lambda v = \lambda^2 v$.

2.3 Trivia. Potęgi macierzy i rekurencje liniowe

Ci z Państwa, który zetknęli się z rekurencjami[8], mogli słyszeć o wielomianie/równaniu charakterystycznym rekurancji liniowej, pozwalającym na wyznaczanie wzoru ogólnego na ciąg będący jej rozwiązaniem. Okazuje się, że za równaniem tym stoi algebra liniowa i jego rozwiązania są wartościami własnymi.

Definicja 2.6: Liniowy ciąg rekurencyjny rzędu k

Niech K będzie ciałem. Przez LINIOWY CIĄG REKURENCYJNY rozumiem ciąg $(X_n) \in K^\infty$ rzędu k określony warunkami:

- $X_0, \ldots, X_{k-1} \in K$ – dane,
- dla $n \geqslant k$ mamy liczby $c_0, c_1, \ldots, c_{k-1} \in K$, dla których:

$$X_n - c_{k-1}X_{n-1} - \ldots - c_0 X_{n-k} = 0, \qquad \text{dla } n \geqslant k.$$

Zauważmy, że dowolne rozwiązanie (X_n) rekurencji rzędu k uzyskać można poprzez określenie pierwszych k wyrazów $X_0, \ldots, X_{k-1}$. Na ich podstawie pozostałe wyrazy wyznaczamy ze wzoru wyżej.
Nietrudno widzieć, że zbiór rozwiązań danej rekurencji liniowej jest podprzestrzenienią K^∞. Jak się okazuje, w niektórych sytuacjach rozwiązania rekurencji są bardzo eleganckie. Dzieję się tak wtedy, gdy wskazać można odpowiednią liczbę liniowo niezależnych ciągów geometrycznych, które ją spełniają.

Obserwacja 2.13

Niech $0 \neq \alpha \in K$. Ciąg geometryczny (a_k) dany wzorem $a_k = \alpha^n$ spełnia rekurencję (2.6) wtedy i tylko wtedy, gdy α jest rozwiązaniem równania:

$$\lambda^k - c_{k-1}\lambda^{k-1} - c_{k-2}\lambda^{k-2} - \ldots - c_1\lambda - c_0 = 0.$$

.

Dowód. Jeśli $\alpha^{n+k} = c_{k-1}\alpha^{n+k-1} + c_{k-2}\alpha^{n+k-2} + \ldots + c_0\alpha^n$. W zatem oczywiście po wyłączeniu $\alpha^n \neq 0$ widzimy, że α spełnia równanie wyżej. Dowód w drugą stronę jest analogiczny. □

Okazuje się, że równanie charakterystyczne przywołane wyżej to wielomian charakterystyczny pewnej macierzy. A jego pierwiastki to wartości własne. Rzeczywiście, mamy:

$$\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -c_0 & -c_1 & -c_2 & \ldots & -c_{k-1} \end{bmatrix} \cdot \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{k-1} \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \end{bmatrix} \implies \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -c_0 & -c_1 & -c_2 & \ldots & -c_{k-1} \end{bmatrix}^n \cdot \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{k-1} \end{bmatrix} = \begin{bmatrix} X_n \\ X_{n+1} \\ X_{n+2} \\ \vdots \\ X_{n+k-1} \end{bmatrix}.$$

Stąd rozwiązanie rekurencji wymaga podnoszenia macierzy do potęgi. Dlaczego tak interesuje nas podobieństwo do macierzy diagonalnej? Znajomość potęgi macierzy znajdującej się po lewej stronie równości jest kluczowa do ewentualnego określenia wzoru na n-ty wyraz ciągu rekurencyjnego. Łatwiej liczy się jednak potęgi macierzy diagonalnej, niż dowolnej. Jak się okazuje, podnoszenie do potęgi jest całkiem przyjemne, jeśli macierz jest podobna do diagonalnej. Problem w tym, że wypisana po lewej macierz $k \times k$ nie musi być zawsze podobna do macierzy diagonalnej (na przykład dla $X_n = 4X_{n-1} - 4X_{n-2}$), a to znacznie utrudnia wyznaczenie ogólnej formuły dla dowolnej liniowej rekurencji. Może się jednak zdarzyć klasyczna sytuacja, gdy takie podobieństwo można stwierdzić. Dzieje się to na przykład gdy wielomian charakterystyczny macierzy endomorfizmu ma n parami różnych pierwiastków (także w większym ciele, np. pierwiastki zespolone rzeczywistej rekurencji — ma to wpływ na wzór ogólny). Tak jest w przypadku ciągu Fibonacciego i dzięki temu mogliśmy choćby wyprowadzić wzór Bineta.

Wciąż nie wyjasniliśmy skąd się wzięła macierz wypisana wyżej. Sam rachunek jej wielomianu charakterystycznego nie jest trudny, ale wkrótce przekonamy się, że ma on znacznie poważniejsze zastosowania.

[8] Wszystkim, którzy chcieliby się dowiedzieć więcej o ciągach rekurencyjnych polecam bardzo ciekawe materiały dydaktyczne prof. W. Guzickiego: `https://www.mimuw.edu.pl/~guzicki/materialy/Rekurencja.pdf`.

2.4 Coda. Czym są reprezentacje?

Jedną z podstawowych koncepcji nauki jest metodologia badania skomplikowanego procesu w warunkach temu sprzyjających (np. sterylnych). W kontekście nauk empirycznych mówimy choćby o badaniach laboratoryjnych. W przypadku nauk dedukcyjnych koncepcja ta realizuje się zasadniczo w ten sposób, że badane są teorie aksjomatyczne, ich własności oraz ich tzw. modele – pewien zbiór obiektów matematycznych połączonych w pewien system, spełniający (cokolwiek to znaczy) określony system zdań (teorię). Podejście to miało wielkie znaczenie od II połowy XIX wieku, gdy zbudowane zostały modele geometrii nieeuklidesowych wewnątrz geometrii euklidesowej i odwrotnie, co stanowiło gigantyczny przełom – okazało się, że nie ma „jednej" geometrii. Geometria, w której zaprzeczymy V postulatowi Euklidesa jest nie mniej „realna" od dobrze nam znanej geometrii euklidesowej. Zainteresowanych odsyłam do tekstów prof. Kordosa w *Delcie* (`https://deltami.edu.pl/`), np. *Inne światy, inne geometrie* (2016), *Geometria Bolyaia–Łobaczewskiego – co to jest i jak ją poznawać* (2018), *Najregularniejsze i najpopularniejsze geometrie* (1981), *Beznadziejne marzenie o całej prawdzie* (1985). Polecam gorąco te teksty.

W XX wieku wielkie triumfy przeżywała logika matematyczna i teoria modeli, prowadząc do słynnych wyników, najpierw Tarskiego i Gödla, a później Łosia, Cohena, Morleya, Szelacha, gdzie najpierw doprecyzowane zostały pojęcia „zdania prawdziwego", „spełniania", wykazano klasyczne twierdzenia o istnieniu modelu (dla tzw. teorii pierwszego rzędu), a następnie słynne twierdzenia o niezupełności arytmetyczności czy o niedowodliwości niesprzeczności. Dziś język tej teorii rozwijany jest nie tylko w teorii mnogości.

Powyższe perspektywy mówią, w ogromnym uproszczeniu, o reprezentowaniu teorii za pomocą modeli, gdzie mamy rozmaite „bliskie nam" metody prowadzenia badań i próbujemy wyprowadzać w ten sposób twierdzenia ogólnej teorii. W algebrze również mówimy o reprezentacjach, istnieje nawet niezwykle istotny dział algebry – teoria reprezentacji. Jakie pytania zadają te teorie? Najczęściej chodzi o pytanie: czy pewna struktura algebraiczna może być zrealizowana jako podstruktura w dobrze znanej strukturze. Dlaczego nas to interesuje? Być może w dobrze znanej strukturze mamy narzędzia, które mogą być pomocne do badania ogólnych struktur? Możemy na przykład reprezentować struktury jako macierze i używać wyznacznika lub śladu. A jeśli niektóre z macierzy okażą się diagonalizowalne, to okaże się, że np. iteracje wykonuje się bardzo czytelnie (patrz wyżej). Możemy reprezentować niektóre struktury jako permutacje i używać znaku, cyklu itd. Oto przykładowe, elementarne rezultaty tego typu, które Państwo poznają.

- Dla przestrzeni liniowej V wymiaru n nad ciałem K i ustalonej bazy $\mathcal{A}$ tej przestrzeni można reprezentować dowolny wektor $\alpha \in V$ za pomocą ciągu n elementów ciała K – współrzędnych bazy $\mathcal{A}$. Ta (jedna z wielu) reprezentacja to w istocie stosownie wybrany izomorfizm V oraz K^n.
- Każdy endomorfizm przestrzeni V wymiaru n nad ciałem K można „reprezentować" za pomocą macierzy w ustalonej bazie i to w taki sposób, że macierz reprezentująca sumę lub złożenie jest odpowiednio: sumą lub złożeniem macierzy reprezentującej. Tu znowu mamy izomorfizm przestrzeni V oraz $M_n(K)$. Dowolny automorfizm V reprezentowany jest tu przez macierz odwracalną.
- Dla każdej n-wymiarowej algebry A nad ciałem K istnieje monomorfizm zachowujący mnożenie wektorów $\phi : A \to M_N(K)$, dla pewnego $N \in \mathbb{N}$. Innymi słowy – dowolną skończenie wymiarową algebrę nad ciałem K można „włożyć" w algebrę macierzy pewnego rozmiaru. Oto przykład: weźmy algebrę kwaternionów Hamiltona. Jest to czterowymiarowa algebra rzeczywista, której każdy element ma postać $a \cdot 1 + bx + cy + dz$, gdzie $a, b, c, d \in \mathbb{R}$ oraz $x^2 = y^2 = z^2 = xyz = -1$. Algebrę tę można włożyć w macierze rozmiaru 4 nad ciałem $\mathbb{R}$ (w algebrę wymiaru 16) za pomocą przekształcenia (liniowego zachowującego mnożenie) określonego na bazie w następujący sposób:

$$1 \mapsto \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad x \mapsto \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad y \mapsto \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad z \mapsto \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

  Zauważmy natomiast, że jeśli dopuścimy reprezentacje zespolone, wówczas możemy użyć znacznie mniejszych macierzy (zwanych macierzami Pauliego – przy czym mnożymy je przez i):

$$1 \mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad y = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad z = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

- Dowolna grupa skończona może być „włożona" w grupę permutacji. W szczególności np. grupy symetrii są podgrupami grupy permutacji. Tzw. teoria reprezentacji grup zajmuje się natomiast „reprezentowaniem" grup za pomocą macierzy. O tym jednak opowiemy osobno w innym odcinku.

Rozdział 3

Podprzestrzenie niezmiennicze. Twierdzenie Cayleya-Hamiltona

Każda podprzestrzeń własna endomorfizmu ϕ przestrzeni liniowej V ma tę własność, że nie zmienia się (jako cała przestrzeń) przy działaniu ϕ, to znaczy: $\phi(V_{(a)}) = V_{(a)}$. Uogólnieniem tej własności jest pojęcie podprzestrzeni niezmienniczej, któremu poświęcony jest ten wykład.

Definicja 3.1: Podprzestrzeń niezmiennicza względem endomorfizmu

Niech ϕ będzie endomorfizmem przestrzeni liniowej V nad ciałem K. Podprzestrzeń $W \subseteq V$ nazwiemy NIEZMIENNICZĄ WZGLĘDEM ϕ (lub ϕ-niezmienniczą) wtedy i tylko wtedy, gdy $\phi(W) \subseteq W$.

Przykłady.

- Niech $\phi \in \mathrm{End}(V)$. Przestrzeń V oraz podprzestrzeń zerowa są zawsze niezmiennicze względem ϕ.
- Niech $\phi \in \mathrm{End}(V)$. Podprzestrzenie $\ker(\phi)$ oraz $\mathrm{im}(\phi)$ są niezmiennicze względem ϕ. Ważne dla różnych rozumowań jest również to, że $\ker(\phi^k)$ oraz $\mathrm{im}(\phi^k)$ są ϕ-niezmiennicze, dla każdego $k \in \mathbb{N}$.
- Niech $0 \neq \alpha \in V$ będzie wektorem własnym $\phi \,\mathrm{End}(V)$. Wówczas $\mathrm{lin}(\alpha)$ jest podprzestrzenią ϕ-niezmienniczą. Jeśli a jest wartością własną ϕ, to $V_{(a)}$ jest podprzestrzenią ϕ-niezmienniczą.
- Niech $\phi \in \mathrm{End}(V)$. Jeśli $U, W \subseteq V$ są ϕ-niezmiennicze, to $U \cap W$ oraz $U + W$ — również.
- Jeśli $\phi, \psi \in \mathrm{End}(V)$ oraz $\phi \circ \psi = \psi \circ \phi$, to $\ker \psi$ oraz $\mathrm{im}\, \psi$ są ϕ-niezmiennicze (proste ćwiczenie).
- Niech $\phi \in \mathrm{End}(\mathbb{R}^4)$ będzie zadany w bazie standardowej macierzą

$$M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \qquad (*)$$

  Łatwo widzieć, że $\mathrm{lin}((1,0,0,0),(0,1,0,0))$ oraz $\mathrm{lin}((0,0,1,0),(0,0,0,1))$ są (nie jedynymi) podprzestrzeniami niezmienniczymi endomorfizmu ϕ. Nie są to też podprzestrzenie własne.

Definicja 3.2

Niech $\phi : V \to V$ będzie przekształceniem liniowym oraz niech $U \subseteq V$ będzie podprzestrzenią ϕ-niezmienniczą.[a] Przez $\phi|_U$ rozumieć będziemy endomorfizm przestrzeni U zadany wzorem:

$$\phi|_U(u) = \phi(u), \qquad \text{dla każdego } u \in U.$$

[a] Oznaczenie to sugeruje, że mowa jest o obcięciu przekształcenia f do podprzestrzeni U, ale obcięcie jest formalnie definiowane jako przekształcenie z U do V. Nam natomiast zależy na tym, żeby macierz endomorfizmu $\phi|_U$ była kwadratowa, co oczywiście ma sens w przypadku podprzestrzeni niezmienniczych.

Tak jak endomorfizm ma wektor własny wtedy i tylko wtedy, gdy jego wielomian charakterystyczny ma pierwiastek, tak istnienie nietrywialnych (różnych od zera i całości) podprzestrzeni niezmienniczych zależne jest od rozkładalności wielomianu ϕ w ciele K na czynniki oraz od możliwości przedstawienia macierzy przekształcenia ϕ w bazie w odpowiedniej postaci blokowej.

Obserwacja 3.1

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K, niech $\phi \in \mathrm{End}(V)$ oraz niech U będzie podprzestrzenią V, różną od 0 i V. Następujące warunki są równoważne:

(1) $U \subseteq V$ jest podprzestrzenią ϕ-niezmienniczą,

(2) istnieje baza $\mathcal{A}$ przestrzeni V oraz taka macierz $A \in M_{\dim U}(K)$, że

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix},$$

gdzie A jest macierzą $\phi|_U$ w pewnej bazie przestrzeni U.

Gdy spełniony jest jeden z warunków wyżej, wówczas istnieje $g \in K[\lambda]$, że:

$$w_\phi(\lambda) = w_{\phi|_U}(\lambda) \cdot g(\lambda).$$

Dowód. Załóżmy (1). Niech $\mathcal{A}' = (\alpha_1, \ldots, \alpha_k)$ będzie bazą U oraz niech $(\alpha_{k+1}, \ldots, \alpha_n)$ będzie dopełnieniem bazy $\mathcal{A}'$ do bazy $\mathcal{A}$ przestrzeni V. Skoro U jest ϕ-niezmiennicza, to obrazy wektorów bazowych $\alpha_1, \ldots, \alpha_k$ po zastosowaniu ϕ należą do $\mathrm{lin}(\alpha_1, \ldots, \alpha_k) = U$. A zatem pierwsze k kolumn macierzy $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ może mieć niezerowe wyrazy jedynie w pierwszych k wierszach. W szczególności istnieją takie macierze $A \in M_k(K)$, $B \in M_{k\times(n-k)}(K)$ oraz $C \in M_{n-k}(K)$, że:

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

Uzyskaliśmy zatem (2). Odwrotna implikacja (2) $\Rightarrow$ (1) jest teraz jasna. Co więcej,

$$w_\phi(\lambda) = \det \begin{bmatrix} A - \lambda I_k & B \\ 0 & C - \lambda I_{n-k} \end{bmatrix}.$$

A zatem $w_\phi(\lambda) = \det(A - \lambda I_k) \det(C - \lambda I_{n-k})$. Jednak $\det(A - \lambda I_k)$ to wielomian charakterystyczny endomorfizmu $\phi|_U$. Dowód jest zakończony. □

Niezwykle istotny wgląd w teorię endomorfizmów daje badanie ich nie tylko z perspektywy macierzy przekształcenia, ale wynikających z niej rozkładów przestrzeni liniowych, na których działają.

Definicja 3.3: Rozkład ϕ-niezmienniczy

Niech $\phi \in \mathrm{End}(V)$. Niech V będzie sumą prostą postaci

$$V = V_1 \oplus \ldots \oplus V_r \qquad (*).$$

Powiemy, że ROZKŁAD $(*)$ JEST ϕ-NIEZMIENNICZY, jeśli każda z podprzestrzeni V_i jest ϕ-niezmiennicza.

Przykłady.

- Jeśli $\phi \in \mathrm{End}(V)$ jest rzutem, to rozkład $V = \ker(\phi) \oplus \mathrm{im}(\phi)$ jest ϕ-niezmienniczy.
- Jeśli $\phi \in \mathrm{End}(V)$ jest diagonalizowalny, to rozkład V na sumę prostą podprzestrzeni własnych jest ϕ-niezmienniczy.
- Dla endomorfizmu $\phi \in \mathrm{End}(K^4)$ danego w bazie standardowej macierzą $(*)$ mamy rozkład na podprzestrzenie ϕ-niezmiennicze postaci $V = \mathrm{lin}(\epsilon_1, \epsilon_2) \oplus \mathrm{lin}(\epsilon_3, \epsilon_4)$.

KLUCZOWA IDEA. Jeśli znajdziemy rozkład ϕ-niezmienniczy postaci $V = V_1 \oplus \ldots \oplus V_r$, to badanie działania ϕ na V sprowadza się do badania *niezależnych działań* ϕ na każdym z V_i. Na kolejnych wykładach zobaczymy więcej typów rozkładów ϕ-niezmienniczych — zwłaszcza przedstawienie w postaci Jordana, a wcześniej — rozkład na podprzestrzenie pierwiastkowe (uogólnione podprzestrzenie własne).

Wniosek 3.1

Niech $V = V_1 \oplus \ldots \oplus V_k$ oraz niech $\phi \in \mathrm{End}(V)$ będzie taki, że $V_1, \ldots, V_k$ są podprzestrzeniami ϕ-niezmienniczymi w V. Wówczas ϕ ma w pewnej bazie $\mathcal{A}$ macierz blokowo-diagonalną postaci

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & A_k \end{bmatrix}, \quad \text{gdzie } A_i \in M_{\dim(V_i)}(K).$$

Co więcej, $w_\phi(\lambda) = w_{A_1}(\lambda) \cdot \ldots \cdot w_{A_r}(\lambda)$.

Prawdziwy *cud* polega na tym, że odpowiednia rozkładalność wielomianu charakterystycznego gwarantuje rozkłady ϕ-niezmiennicze. Przyjrzymy się tej sprawie mówiąc o tzw. rozkładzie prymarnym.

Niestety uzyskiwanie rozkładów przestrzeni liniowych na sumy proste podprzestrzeni niezmienniczych bywa bardzo trudne. Zauważmy, że kryterium diagonalizowalności macierzy dotyczy dokładnie sytuacji, gdy V jest sumą prostą podprzestrzeni własnych. Skąd jednak wiadomo, że taki warunek zachodzi? Niestety nie jest prawdą nad każdym ciałem, że jeśli $W \subseteq V$ jest podprzestrzenią niezmienniczą względem $\phi \in \mathrm{End}(V)$, to istnieje podprzestrzeń ϕ-niezmiennicza W' taka, ze $W \oplus W' = V$. Wybór odpowiednich podprzestrzeni (i ciała!), dla których takie dopełnienia da się uzyskać jest treścią głębokich twierdzeń.

Zanim przejdziemy do dalszej części, przyjrzymy się powyższym definicjom przywołując język przestrzeni ilorazowych. Przypomnijmy kilka podstawowych definicji.

Definicja 3.4: Układ liniowo niezależny modulo podprzestrzeń

Niech U będzie podprzestrzenią przestrzeni liniowej V nad ciałem K. Powiemy, że układ wektorów $v_1, \ldots, v_n \in V$ jest LINIOWO NIEZALEŻNY MODULO U, jeśli dla $a_1, \ldots, a_n \in K$ mamy

$$a_1 v_1 + \ldots + a_n v_n \in U \Rightarrow a_1 = \ldots = a_n = 0.$$

Powiemy, że układ liniowo niezależny modulo U jest BAZĄ V MODULO U, jeśli jest on maksymalnym liniowo niezależnym (ze względu na inkluzję) układem modulo U.

Układy liniowo niezależne i bazy modulo podprzestrzeń są istotne z punktu widzenia sum prostych. Jeśli zastanawiamy się jak mogą wyglądać wszystkie możliwe bazy modulo podprzestrzeń, to okazuje się, że muszą to być bazy dopełnień prostych tej podprzestrzeni do całej przestrzeni Mówi o tym poniższa uwaga, której łatwy dowód znajduje się w notatkach (dodatek) z GAL I.

Obserwacja 3.2

Niech $U \subsetneq V$ i niech $v_1, \ldots, v_n$ będzie układem liniowo niezależnym modulo U. Wówczas następujące warunki są równoważne:

(1) układ $v_1, \ldots, v_n$ jest bazą V modulo U,

(2) $U \oplus \mathrm{lin}(v_1, \ldots, v_n) = V$.

Widzimy zatem, że każda baza przestrzeni V modulo U odpowiada pewnemu dopełnieniu prostemu U do V, i odwrotnie. Konstrukcja przestrzeni ilorazowej wskazuje jeden „kanoniczny" obiekt, który będzie „izomorficzny" z każdym z tych dopełnień.

Definicja 3.5: Warstwa podprzestrzeni

Niech W będzie podprzestrzenią przestrzeni V nad ciałem K i niech $\alpha \in V$. Zbiór

$$\alpha + W = \{\alpha + \gamma \,|\, \gamma \in W\}$$

nazywamy WARSTWĄ PODPRZESTRZENI W w przestrzeni V.

Warstwa jest, jak widzimy, abstrakcyjnym odpowiednikiem zbioru rozwiązań niejednorodnego układu równań, którego zbiór rozwiązań traktujemy w konstrukcji przestrzeni ilorazowej jako pojedynczy wektor. Fundamentalne obserwacje dotyczące warstw zawarte są w następującej uwadze. Do dowodów odsyłamy ponowne do wspomnianych wyżej notatek.

Obserwacja 3.3

Niech W będzie podprzestrzenią przestrzeni V.

(i) Zachodzi równoważność $\alpha + W = \beta + W \iff \alpha - \beta \in W$,

(ii) Dla warstw $v + W$ oraz $v' + W$ określamy działania:

$$(v + W) + (v' + W) = (v + v') + W, \quad a \cdot (v + W) = av + W.$$

Działania te są dobrze określone, tzn. jeśli $v + W = v' + W$, to dla każdego $v'' \in V$ oraz dla każdego $a \in K$ mamy: $(v+W)+(v''+W) = (v'+W)+(v''+W)$ oraz $a\cdot(v+W) = a\cdot(v'+W)$.

(iii) Zbiór $V/W = \{\alpha + W \,|\, \alpha \in V\}$ z działaniami dodawania i mnożenia przez skalar określonymi wyżej oraz z warstwą $0 + W$ tworzy przestrzeń liniową nad ciałem K,

(iv) Jeśli $\mathcal{A} = \{v_1, \ldots, v_n\}$ jest bazą W oraz układ $\mathcal{B} = \{y_1, \ldots, y_m\}$ wektorów przestrzeni V ma tę własność, że $\{y_1 + W, \ldots, y_m + W\}$ to baza V/W, wówczas $\mathcal{A} \cap \mathcal{B} = \emptyset$ oraz $\mathcal{A} \cup \mathcal{B}$ jest bazą V. W szczególności, jeśli V jest skończenie wymiarowa, to V/W też jest skończenie wymiarowa i

$$\dim V = \dim W + \dim(V/W).$$

Definicja 3.6: Przestrzeń ilorazowa

Przestrzeń V/W określoną w poprzedniej uwadze nazywamy PRZESTRZENIĄ ILORAZOWĄ przestrzeni V przez podprzestrzeń W. Wymiar przestrzeni ilorazowej V/W nazywamy KOWYMIAREM W w V.

Przykład. Przestrzeń $\mathbb{R}^2/\operatorname{lin}(0,1)$ składa się z warstw

$$(x, y) + \operatorname{lin}(0, 1),$$

przy czym dla $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$:

$$(x_1, y_1) + \operatorname{lin}(0,1) = (x_2, y_2) + \operatorname{lin}(0,1) \iff (x_1 - x_2, y_1 - y_2) \in \operatorname{lin}(0,1) \iff x_1 = x_2.$$

Bijekcja $(x, y) + \operatorname{lin}(0, 1) \mapsto (x, 0)$ zadaje izomorfizm przestrzeni liniowych

$$\mathbb{R}^2/\operatorname{lin}(0,1) \simeq \operatorname{lin}(1,0).$$

Zauważmy, że jeśli $W \subseteq V$, to mamy naturalne przekształcenie $\pi : V \to V/W$ zadane wzorem:

$$\pi(\alpha) = \alpha + W$$

(przyporządkowujemy wektorowi jego warstwę). Jest to dobrze określone przekształcenie liniowe.

Definicja 3.7: Naturalne rzutowanie

Odwzorowanie $\pi : V \to V/U$ przypisujące każdemu elementowi $v \in V$ warstwę $v + U$ nazywane jest NATURALNYM RZUTOWANIEM na V/U. Obraz $v \in V$ jest często oznaczany w skrócie jako $\overline{v}$.

Przykład. Biorąc naturalne rzutowanie $\pi : \mathbb{R}^2 \to \mathbb{R}^2/\operatorname{lin}(0,1)$ mamy: $\overline{(1,0)} = \overline{(1,1)}$.

Z punktu widzenia przestrzeni niezmienniczych pojęcie przestrzeni ilorazowej pozwala na określanie naturalnych endomorfizmów na ilorazach, przydatnych w rozumowaniach indukcyjnych.

Definicja 3.8: Przekształcenie indukowane

Jeśli $\phi \in \operatorname{End}(V)$ oraz $U \subseteq V$ jest podprzestrzenią ϕ-niezmienniczą, to przekształcenie liniowe $\overline{\phi} : V/U \to V/U$ dane wzorem

$$\overline{\phi}(v + U) = \phi(v) + U$$

nazywamy ENDOMORFIZMEM INDUKOWANYM przez ϕ na V/U.

Wniosek 3.2

Niech V będzie n-wymiarową przestrzenią liniową nad ciałem K, niech $\phi \in \operatorname{End}(V)$ oraz niech U będzie podprzestrzenią niezmienniczą wymiaru $k > 0$. Istnieje wówczas baza $\alpha_1, \ldots, \alpha_n$ przestrzeni V, taka że $\alpha_1, \ldots, \alpha_k$ jest bazą U, zaś $\alpha_{k+1} + U, \alpha_{k+2} + U, \ldots, \alpha_n + U$ jest bazą V/U. Przekształcenie ϕ ma w tej bazie macierz blokowo-górnotrójkątną

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}, \qquad (\dagger)$$

gdzie A jest macierzą $\phi|_U$, a C jest macierzą $\overline{\phi}$. W szczególności $w_\phi(\lambda) = w_{\phi|_U}(\lambda) \cdot w_{\overline{\phi}}(\lambda)$.

Kolejny ważny wątek związany z podprzestrzeniami niezmienniczymi i przestrzeniami ilorazowymi dotyczy endomorfizmów, których macierze mają w pewnej bazie macierz górnotrójkątną.

Twierdzenie 3.1

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K oraz niech $\phi \in \operatorname{End}(V)$. Następujące warunki są równoważne:

(1) wielomian $w_\phi(\lambda)$ rozkłada się na czynniki stopnia nie większego niż 1 (liniowe),

(2) istnieje baza $\mathcal{A}$, w której macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest górnotrójkątna.

Dowód. Implikacja (2) $\Rightarrow$ (1) jest oczywista. Dowodzimy implikację przeciwną przez indukcję ze względu na $n = \dim V$. Przypadek $n = 1$ jest oczywisty. Niech $n > 1$ i niech $\lambda_1, \ldots, \lambda_n$ będą wartościami własnymi endomorfizmu ϕ Wybierzmy wektor własny v_1 odpowiadający wartości własnej λ_1 i niech $U = \operatorname{lin}(v_1)$. Rozważamy przekształcenie liniowe $\overline{\phi} : V/U \to V/U$ indukowane przez ϕ i zadane wzorem $\overline{\phi}(v + U) = \phi(v) + U$. Aby skorzystać z założenia indukcyjnego dla endomorfizmu $\overline{\phi}$, musielibyśmy wiedzieć, że $w_{\overline{\phi}}(\lambda)$ rozkłada się na czynniki liniowe.

Zauważmy, że dopełniając v_1 elementami $w_2, \ldots, w_n$ do dowolnej bazy $\mathcal{A}$ przestrzeni V wiemy, na mocy wniosku wyżej, że macierz ϕ w bazie $\mathcal{A}$ ma postać (†), gdzie C jest macierzą endomorfizmu $\overline{\phi}$ w bazie $(w_2 + U, w_3 + U, \ldots, w_n + U)$, czyli $w_\phi(\lambda) = (\lambda_1 - \lambda) \cdot w_{\overline{\phi}}(\lambda)$. Stąd $w_{\overline{\phi}}(\lambda)$ jest rozkładalny na czynniki liniowe, jako dzielnik wielomianu o tej samej własności. Możemy przejść do kroku indukcyjnego.

Z założenia indukcyjnego zatem możemy znaleźć bazę $\mathcal{B} = (v_2+U, \ldots, v_n+U)$ przestrzeni V/U taką, że $\overline{\phi}$ ma w tej bazie macierz górnotrójkątną. To jednak znaczy, z definicji macierzy przekształcenia liniowego, że dla $i \geqslant 2$ mamy

$$\overline{\phi}(v_i + U) = \phi(v_i) + U \subseteq \operatorname{lin}(v_2 + U, \ldots, v_i + U).$$

Rozważając teraz bazę $\mathcal{A} = (v_1, v_2, \ldots, v_n)$, złożoną z v_1 oraz reprezentantów warstw $v_2 + U, \ldots, v_n + U$ widzimy, że dla $i \geqslant 1$ mamy $\phi(v_i) \in \operatorname{lin}(v_1, \ldots, v_i)$, czyli macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest górnotrójkątna. □

Prześledźmy to samo rozumowanie, czyli dowód implikacji (1) ⇒ (2) bez wprowadzania przestrzeni ilorazowych. Wymaga ono sprytnego kroku, wykorzystującego macierze blokowe i teorię wyznaczników. Jest to również indukcja ze względu na $\dim(V)$. Jak wygląda zatem krok indukcyjny?

Z faktu, że $w_\phi(\lambda)$ rozkłada się na czynniki liniowe wynika, że co najmniej jeden z tych czynników jest stopnia 1, a więc zgodnie z twierdzeniem Bezout wielomian ten ma pierwiastek c. W szczególności, endomorfizm ϕ ma wektor własny $\alpha \neq 0$ o wartości własnej c. Dopełnijmy wektor α do dowolnej bazy $\mathcal{A}$ przestrzeni V wektorami $(\beta_1, \ldots, \beta_{n-1})$. Wówczas w bazie $\mathcal{A}$ macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ ma postać:

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} c & A \\ 0 & B \end{bmatrix},$$

gdzie $A \in M_{1\times(n-1)}(K)$ oraz $B \in M_{n-1}(K)$. Innymi słowy — pierwsza kolumna macierzy $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ ma, poza elementem c na przekątnej, same zerowe wyrazy.

Niech $B = M(\psi)^{st}_{st}$, dla pewnego endomorfizmu ψ przestrzeni $n-1$ wymiarowej $V' = \operatorname{lin}(\beta_1, \ldots, \beta_{n-1})$. Z rozwinięcia Laplace'a względem pierwszej kolumny wynika natychmiast, że

$$w_\phi(\lambda) = \det \begin{bmatrix} c-\lambda & A \\ 0 & B - \lambda I \end{bmatrix} = (c-\lambda) \cdot w_\psi(\lambda).$$

Skoro założyliśmy, że $w_\phi(\lambda)$ rozkłada się na czynniki liniowe, a $w_\psi(\lambda)$ jest dzielnikiem $w_\phi(\lambda)$, to także $w_\psi(\lambda)$ rozkłada się na czynniki liniowe (to jest nieco śliski moment, ale nie ma tu oszustwa). Z założenia indukcyjnego istnieje zatem baza $\mathcal{B}$ przestrzeni V' (wymiaru $n-1$), że endomorfizm ψ ma w niej macierz górnojtrójkątną B'. W języku macierzy oznacza to, że istnieje macierz odwracalna S rozmiaru $n-1$ taka, że $B' = S^{-1}BS$. Rozważmy macierz S' rozmiarów $n \times n$ postaci

$$\begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix},$$

gdzie w pierwszym wierszu i w pierwszej kolumnie poza przekątną stoją zera. Jest jasne, że:

$$S'^{-1} \cdot M(\phi)^{\mathcal{A}}_{\mathcal{A}} \cdot S' = \begin{bmatrix} 1 & 0 \\ 0 & S^{-1} \end{bmatrix} \cdot \begin{bmatrix} c & A \\ 0 & B \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} c & A \\ 0 & S^{-1}B \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & S \end{bmatrix} = \begin{bmatrix} c & AS \\ 0 & B' \end{bmatrix}.$$

Stąd macierz $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest podobna do macierzy górnotrójkątnej, co kończy dowód.

Definicja 3.9: Endomorfizm triangularyzowalny

Powiemy, że endomorfizm ϕ przestrzeni skończenie wymiarowej V nad ciałem K jest TRIANGULARYZOWALNY, jeśli w pewnej bazie tej przestrzeni ma on macierz górnotrójkątną.

Oczywiście każdy endomorfizm diagonalizowalny jest triangularyzowalny. Co więcej, kryterium triangularyzowalności jest znacznie słabsze niż diagonalizowalności. Wymaga ono jedynie, żeby wielomian charakterystyczny endomorfizmu rozkładał się na czynniki stopnia nie większego niż 1.

Wniosek 3.3

Każdy endomorfizm skończenie wymiarowej przestrzeni liniowej V nad ciałem algebraicznie domkniętym K jest triangularyzowalny. Każda macierz $A \in M_n(K)$ jest triangularyzowalna.

Na koniec zajmiemy się teraz wykorzystaniem następującej obserwacji. Jeśli α należy do podprzestrzeni ϕ-niezmienniczej U przestrzeni V, to do U należą również kolejne iteracje ϕ na α, czyli wektory $\phi^i(\alpha), \quad \ldots$.

Definicja 3.10: Podprzestrzeń cykliczna

Niech $\phi \in \mathrm{End}(V)$. Podprzestrzenią CYKLICZNĄ przestrzeni V względem ϕ rozpiętą przez wektor α nazywamy:

$$V_\alpha = \mathrm{lin}(\phi^n(\alpha) \,|\, n \in \mathbb{N}) = \mathrm{lin}(\alpha, \phi(\alpha), \phi^2(\alpha), \phi^3(\alpha), \ldots).$$

Przykłady.

- Niech ϕ będzie jednokładnością o skali λ na V (w tym identycznością, dla $\lambda = 1$.). Wówczas dla każdego $\alpha \in V$ mamy $V_\alpha = \mathrm{lin}(\alpha)$, gdyż dla dodatniej liczby całkowitej k mamy $\phi^k(\alpha) = \lambda^k \cdot \alpha$.
- Niech ϕ będzie symetrią $\mathbb{R}^2$ względem $\mathrm{lin}(1,0)$ wzdłuż $\mathrm{lin}(0,1)$. Wówczas z jednej strony mamy ciąg równości $(1,0) = \phi(1,0) = \phi^2(1,0) = \ldots$, czyli $V_{(1,0)} = \mathrm{lin}(1,0)$, $V_{(0,1)} = \mathrm{lin}(0,1)$, a z drugiej strony mamy dwie równości: $(1,1) = \phi((1,-1)) = \phi^2((1,1))$, skąd $V_{(1,1)} = \mathbb{R}^2$.
- Niech $\phi \in \mathrm{End}(K^n)$ dany będzie wzorem $\phi((x_1, x_2, \ldots, x_n)) = (x_2, x_3, \ldots, x_n, 0)$. Wówczas $\phi^n = 0$ oraz $V_{(1,1,1,\ldots,1,1)} = K^n$.

Kluczowe dla naszych dalszych rozważań jest zauważenie, ze podprzestrzeń cykliczna ϕ rozpięta przez wektor v składa się z wartości v na wszystkich wielomianach od endomorfizmu ϕ. O czym mowa?

Definicja 3.11: wielomian od endomorfizmu i od macierzy

Niech $w \in K[\lambda]$ będzie postaci $a_0 + a_1\lambda + \ldots + a_n\lambda^n$. Niech $\phi \in \mathrm{End}(V)$, gdzie V jest nad ciałem K oraz niech $A \in M_s(K)$. Definiujemy:

$$w(\phi) = a_0 \cdot \mathrm{id}_V + a_1 \cdot \phi + \ldots + a_n \cdot \phi^n \in \mathrm{End}(V)$$
$$w(A) = a_0 I + a_1 A + \ldots + a_n A^n \in M_s(K).$$

Kluczowe dla naszych rozważań jest to, że wielomiany od endomorfizmu są przemienne.

Obserwacja 3.4

Niech $\phi \in \mathrm{End}(V)$ Wówczas dla $w, v, z \in K[\lambda]$:

- jeśli $w = v + z$, to $w(\phi) = v(\phi) + z(\phi)$,
- jeśli $w = v \cdot z$, to $w(\phi) = v(\phi) \circ z(\phi) = z(\phi) \circ v(\phi) = vz(\phi) = zv(\phi)$.

Definicja 3.12: Algebra endomorfizmu

Niech $\phi \in \mathrm{End}(V)$, gdzie V jest przestrzenią liniową nad ciałem K. Podprzestrzeń

$$E(\phi) = \{w(\phi) \,|\, w \in K[\lambda]\} \subseteq \mathrm{End}(V)$$

nazywamy ALGEBRĄ ENDOMORFIZMU ϕ.

Możemy teraz napisać, że dla endomorfizmu ϕ podprzestrzeń cykliczna rozpięta przez wektor v złożona jest ze wszystkich wektorów postaci $V_v = \{f(v), f \in E(\phi)\}$. Zauważmy także, że jeśli $A = M(\phi)^{\mathcal{A}}_{\mathcal{A}}$, to dla dowolnego $w \in K[\lambda]$ mamy $w(A) = M(w(\phi))^{\mathcal{A}}_{\mathcal{A}}$. Wprowadzimy jeszcze jedno ważne pojęcie.

Definicja 3.13: Wielomian minimalny endomorfizmu

Niech $\phi \in \mathrm{End}(V)$, gdzie V – skończenie wymiarowa nad ciałem K. Niech $\phi \in \mathrm{End}(V)$. Wówczas wielomian $m \in K[\lambda]$ nazwiemy WIELOMIANEM MINIMALNYM endomorfizmu ϕ, ozn. $m_\phi(\lambda)$, jeśli:

- $m(\phi) = 0$,
- nie istnieje niezerowy wielomian $r \in K[\lambda]$ taki, że $r(\phi) = 0$ i $\deg(r) < \deg(m)$,
- wyraz wiodący m to 1 (czyli jest to tzw. wielomian UNORMOWANY).

Kilka uwag

- Niech $\dim V < \infty$ oraz $\phi \in \mathrm{End}(V)$. Wówczas $\dim \mathrm{End}(\phi) = \dim \mathrm{lin}\{\phi^k \mid k \geqslant 0\} \leqslant (\dim V)^2$. Stąd $m_\phi(\lambda)$ istnieje.
- Jeśli $\phi \in \mathrm{End}(V)$ jest (nietrywialnym) rzutem ($\neq 0, \mathrm{id}$) to $m_\phi(\lambda) = \lambda^2 - \lambda$.
- **Ćwiczenie.** Niech $\phi \in \mathrm{End}(V)$, gdzie V – skończenie wymiarowa nad K. Wówczas dowolny wielomian $p \in K[\lambda]$ o własności $p(\phi) = 0$ jest podzielny przez wielomian minimalny $m_\phi(\lambda)$.

Ostatnia własność, nazwana ćwiczeniem, jest bezpośrednim wnioskiem z twierdzenia o dzieleniu z resztą, zastosowanego do wielomianów $K[x]$. Jak się okaże, wielomian minimalny jest dzielnikiem wielomianu charakterystycznego, a jeśli $V = V_\alpha$ jest podprzestrzenią cykliczną, wówczas wielomian minimalny i wielomian charakterystyczny endomorfizmu ϕ różnią się co najwyżej znakiem.

Jesteśmy gotowi sformułować twierdzenie Cayleya-Hamiltona. Twierdzenie nosi nazwisko wielkich matematyków XIX wieku. Pokazali oni w połowie tamtego stulecia jedynie pewne szczególne przypadki. Co znamienne, Cayley nie widział powodu ani sensu szukania dowodu dla macierzy dowolnego rozmiaru. Zostało ono udowodnione przez Frobeniusa w 1878 roku, który uświadomił sobie dopiero w 1896 roku, że ktoś już o tym problemie myślał przed nim. Zgodnie z ówczesnym zwyczajem przyznał pierwszeństwo swoim słynnym poprzednikom.

Twierdzenie 3.2: Twierdzenie Cayleya-Hamiltona

Jeśli $A \in M_n(K)$ oraz $w = w_A(\lambda)$ jest wielomianem charakterystycznym macierzy A, to $w(A) = 0$.

Zanim przejdziemy do dowodu, udowodnimy następującą obserwację pomocniczą.

Obserwacja 3.5: Własności podprzestrzeni ϕ-cyklicznej V_α, dla $\phi \in \mathrm{End}(V)$

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K oraz niech $\phi \in \mathrm{End}(V)$. Wówczas V_α jest najmniejszą względem inkluzji podprzestrzenią ϕ-niezmienniczą zawierającą α. Jeśli $0 \neq \alpha$, to bazą V_α jest układ $\mathcal{B} = (\alpha, \phi(\alpha), \phi^2(\alpha), \phi^3(\alpha), \ldots, \phi^{k-1}(\alpha))$, gdzie $k = \dim V_\alpha$.

Niech $k = \dim V_\alpha$ oraz niech $c_0, \ldots, c_{k-1} \in K$ spełniają

$$c_0\alpha + c_1\phi(\alpha) + \ldots + c_{k-1}\phi^{k-1}(\alpha) + \phi^k(\alpha) = 0.$$

Wówczas

$$M(\phi|_{V_\alpha})^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -c_{k-1} \end{bmatrix}. \qquad (**)$$

Co więcej: $w_{\phi|_{V_\alpha}}(\lambda) = (-1)^k \cdot (c_0 + c_1\lambda + \ldots + c_{k-1}\lambda^{k-1} + \lambda^k)$.

Dowód. Dowolny element przestrzeni V_α jest kombinacją liniową postaci $v = a_1\phi^{i_1} + \ldots + a_k\phi^{i_k}$, dla pewnych $a_i \in K$ oraz $i_k \in \mathbb{N}$. Oczywiście $\phi(v) = a_1\phi^{i_1+1} + \ldots + a_k\phi^{i_k+1} \in V_\alpha$, co dowodzi, że V_α jest ϕ-niezmiennicza. Oczywiście każda podprzestrzeń ϕ-niezmiennicza zawierająca α zawiera też $\phi^i(\alpha)$, dla $i \geqslant 0$, stąd V_α jest minimalną podprzestrzenią ϕ-niezmienniczą zawierającą α.

Niech $\alpha \neq 0$. Rozważmy maksymalne k takie, że układ $\alpha, \phi(\alpha), \phi^2(\alpha), \phi^3(\alpha), \ldots, \phi^{k-1}(\alpha)$ jest liniowo niezależny (skoro $\alpha \neq 0$, to k istnieje). Twierdzimy, że $k = \dim V_\alpha$. Oczywiście układ ten rozpina V_α, ponieważ wobec maksymalności k mamy $\phi^k(\alpha) \in \operatorname{lin}(\alpha, \phi(\alpha), \phi^2(\alpha), \phi^3(\alpha), \ldots, \phi^{k-1}(\alpha))$, co oznacza, że dla $s \geqslant k$ mamy $\phi^s(\alpha) \in \operatorname{lin}(\alpha, \phi(\alpha), \phi^2(\alpha), \phi^3(\alpha), \ldots, \phi^{k-1}(\alpha))$. Zatem rozważany układ rozpina V_α.

Niech $k = \dim V_\alpha > 0$ oraz niech $c_0, \ldots, c_{k-1} \in K$ spełniają $c_0\alpha + c_1\phi(\alpha) + \ldots + c_{k-1}\phi^{k-1}(\alpha) + \phi^k(\alpha) = 0$. Jest jasne, że $M(\phi|_{V_\alpha})_{\mathcal{B}}^{\mathcal{B}}$ ma postać taką, jak w $(**)$. Policzmy jej wielomian charakterystyczny przez indukcję względem k, stosując rozwinięcie względem pierwszego wiersza. Oczywiście dla $k = 1$ teza jest spełniona, gdyż dla $c_0\alpha + \phi(\alpha) = 0$, mamy wielomian charakterystyczny $w_{\phi|_{V_\alpha}}(\lambda) = \det[-c_0 - \lambda] = (-1) \cdot (c_0 + \lambda)$. Również dla $k = 2$, gdy $c_0\alpha + c_1\phi(\alpha) + \phi^2(\alpha) = 0$, mamy

$$w_{\phi|_{V_\alpha}}(\lambda) = \det \begin{bmatrix} -\lambda & -c_0 \\ 1 & -\lambda - c_1 \end{bmatrix} = (-\lambda)(-\lambda - c_1) + c_0 = c_0 + c_1\lambda + \lambda^2 = (-1)^2(c_0 + c_1\lambda + \lambda^2).$$

Wreszcie, mamy (rozwinięcie względem pierwszego wiersza):

$$\begin{vmatrix} -\lambda & 0 & \ldots & 0 & -c_0 \\ 1 & -\lambda & \ldots & 0 & -c_1 \\ 0 & 1 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -\lambda - c_{k-1} \end{vmatrix} = -\lambda \cdot \begin{vmatrix} -\lambda & \ldots & 0 & -c_1 \\ 1 & \ldots & 0 & -c_2 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 1 & -\lambda - c_{k-1} \end{vmatrix} - c_0 \cdot (-1)^{k+1} + \begin{vmatrix} 1 & -\lambda & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{vmatrix}$$

$$\overset{\text{zał. ind.}}{=} -\lambda \cdot (-1)^{k-1} \cdot (c_1 + c_2\lambda + \ldots + c_{k-1}\lambda^{k-2} + \lambda^{k-1}) + (-1)^{k+2} \cdot c_0$$
$$= (-1)^k(c_0 + c_1\lambda + c_2\lambda^2 + \ldots + c_{k-1}\lambda^{k-1} + \lambda^k).$$

□

Przechodzimy do dowodu twierdzenia Cayleya-Hamiltona. Główna idea jest prosta: wykażemy, że dla każdego $\alpha \neq 0$ istnieje wielomian $w \in K[x]$, który jest dzielnikiem wielomianu charakterystycznego oraz $w(\phi)(\alpha) = 0$. Oczywiście istnienie takiego wielomianu zagwarantuje nam fakt wyżej.

Dowód. Niech $\phi \in \operatorname{End}(V)$ oraz $A = M(\phi)_{st}^{st}$, przy czym $\dim(V) = n$. Aby pokazać, że $w_A(A) = 0$ wystarczy pokazać, że dla dowolnego $\alpha \in V$ zachodzi równość

$$w_\phi(\phi)(\alpha) = 0.$$

Weźmy więc dowolny niezerowy wektor $\alpha \in V$. Rozważamy podprzestrzeń cykliczną V_α wymiaru k. Mamy więc liczby $a_0, a_1, \ldots, a_{k-1} \in K$, że

$$0 = a_0\alpha + a_1\phi(\alpha) + \ldots + a_{k-1}\phi^{k-1}(\alpha) + \phi^k(\alpha),$$

bo baza V_α złożona jest z $\alpha, \ldots, \phi^{k-1}(\alpha)$ (patrz ćwiczenie wyżej). Niech

$$q(\lambda) = a_0 + a_1\lambda + \ldots + a_{k-1}\lambda^{k-1} + \lambda^k \in K[\lambda].$$

Rozszerzamy bazę V_α do bazy $\mathcal{A}$ całego V. Wówczas, zgodnie z twierdzeniem o macierzy przekształcenia mającego podprzestrzeń ϕ-niezmienniczą mamy macierze $X \in M_k(K)$, $Y \in M_{k \times n-k}(K)$, $Z \in M_{n-k}(K)$, że:

$$M(\phi)_{\mathcal{A}}^{\mathcal{A}} = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}.$$

Opiszmy blok X tej macierzy wiedząc co robi ϕ z pierwszymi k wektorami $\alpha, \phi(\alpha), \phi^2(\alpha), \ldots, \phi^{k-1}(\alpha)$ bazy $\mathcal{A}$:

$$X = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_{k-1} \end{bmatrix}.$$

Rzeczywiście

$$\phi(\alpha) = \phi(\alpha), \quad \phi(\phi(\alpha)) = \phi^2(\alpha), \quad \ldots \quad , \phi(\phi^{k-1}(\alpha)) = \phi^k(\alpha) = -(a_0\alpha + a_1\phi(\alpha) + \ldots + a_{k-1}\phi^{k-1}(\alpha)).$$

Cóż nam to wszystko dało? Otóż wiemy, że wielomian charakterystyczny $\phi|_{V_\alpha}$ jest dzielnikiem wielomianu charakterystycznego ϕ, bo V_α jest niezmiennicza. Ale macierz $\phi|_{V_\alpha}$ w bazie $\alpha, \ldots, \phi^{k-1}(\alpha)$ ma dokładnie postać X. A więc wielomian charakterystyczny X jest dzielnikiem wielomianu charakterystycznego endomorfizmu ϕ. Wielomian charakterystyczny tej macierzy to $q(\lambda)$, czyli (jak wiemy z innego z ćwiczeń) $(-1)^k(a_0 + a_1\lambda + \ldots + a_{k-1}\lambda^{k-1} + \lambda^k)$. A zatem wiemy, że $w_\phi(\lambda) = q(\lambda) \cdot h(\lambda) = h(\lambda) \cdot q(\lambda)$, dla pewnego $h \in K[\lambda]$. Jednocześnie

$$q(\phi)(\alpha) = (-1)^k(a_0\alpha + a_1\phi(\alpha) + \ldots + a_{k-1}\phi^{k-1}(\alpha) + \phi^k(\alpha)) = 0.$$

W szczególności mamy $w_\phi(\phi)(\alpha) = hq(\phi)(\alpha) = (h(\phi) \circ q(\phi))(\alpha) = h(\phi)(0) = 0$. Zatem $w_\phi(\phi) = 0$. □

Alternatywny dowód twierdzenia Cayleya-Hamiltona, oparty o przestrzenie ilorazowe jest znacznie prostszy, jeśli przyjmiemy niebanalny fakt: każde ciało K zawarte jest w ciele algebraicznie domkniętym $\overline{K}$, lub ewentualnie nieco bardziej elementarny: dla każdego wielomianu $f \in K[x]$ istnieje ciało $L \subseteq K$ (zwane ciałem rozkładu f), w którym f rozkłada się na czynniki liniowe. Ten dowód poznają Państwo na Algebrze II. W ten sposób endomorfizm ϕ przestrzeni skończenie wymiarowej V nad ciałem K możemy rozpatrywać jako endomorfizm przestrzeni V nad ciałem L. W tym przypadku jednak ϕ jest triangularyzowalny i ma w pewnej bazie $\alpha_1, \ldots, \alpha_n$ macierz górnotrójkątną, gdzie na przekątnej stoją wyrazy $\lambda_i \in L$. Niech $V_i = \mathrm{lin}(\alpha_1, \ldots, \alpha_i)$. Wtedy V_i są podprzestrzeniami niezmienniczymi i $(\phi - \lambda_i \,\mathrm{id})(V_i) \subseteq V_{i-1}$. Stąd:

$$\begin{aligned} W_\phi(\phi)(V) &= (\phi - \lambda_1 \,\mathrm{id})(\phi - \lambda_2 \,\mathrm{id}) \ldots (\phi - \lambda_{n-1} \,\mathrm{id})(\phi - \lambda_n \,\mathrm{id})(V_n) \subseteq \\ &\subseteq (\phi - \lambda_1 \,\mathrm{id})(\phi - \lambda_2 \,\mathrm{id}) \ldots (\phi - \lambda_{n-1} \,\mathrm{id})(V_{n-1}) \subseteq \\ &\subseteq \ldots \subseteq (\phi - \lambda_1 \,\mathrm{id})(V_1) \subseteq V_0 = \{0\}. \end{aligned}$$

* * *

Poza głębokimi zastosowaniami, twierdzenie Cayleya-Hamiltona służy również do rozwiązywania prostych problemów dotyczących potęg i odwrotności macierzy, niekoniecznie diagonalizowalnych. Mówi ono chociażby, że dla każdej macierzy rozmiaru n macierz A^n jest kombinacją liniową swoich niższych potęg.

Dla przykładu: jeśli A to macierz 3×3 o wielomianie charakterystycznym $-\lambda^3 + 1$, to na mocy twierdzenia Cayleya-Hamiltona mamy $-A^3 + I = 0$. Stąd $A^3 = I$. A zatem chcąc wyliczyć na przykład A^{1000} mamy po prostu

$$A^{1000} = (A^3)^{333}A = A.$$

Istnieją też bardziej wyrafinowane zastosowania tej metody[1].

Twierdzenie to pozwala także w ciekawy sposób odwracać niektóre macierze. Jeśli wiemy, dla przykładu, że macierz odwracalna A ma wielomian charakterystyczny $\lambda^5 + \lambda - 1$, to oczywiście $A^5 + A - I = 0$. Mnożąc tę równość z dowolnej strony przez A^{-1} dostajemy warunek: $A^4 + I - A^{-1} = 0$. A zatem $A^{-1} = A^4 + I$.

Jeszcze jeden przykład dotyczy macierzy diagonalizowalnej i podnoszenia do potęgi, pozwalający na uniknięcie dużych rachunków. Rozważmy macierz

$$B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} \in \mathrm{M}_3(\mathbb{C}).$$

Ma ona wartości własne $1, i, -i$. A zatem

$$B^{100} = C^{-1} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix}^{100} \cdot C,$$

gdzie C jest macierzą odwracalną. Zaważmy, że już czwarta potęga macierzy diagonalnej mającej na przekątnej $1, i, -i$ równa jest macierzy identycznościowej I. A zatem $B^{100} = C^{-1} \cdot I \cdot C = I$. Można jednak stwierdzić także, że $w_B = (\lambda^2 + 1)(1 - \lambda) = -\lambda^3 + \lambda^2 - \lambda + 1$, co na mocy twierdzenia Cayleya-Hamiltona oznacza, że $-B^3 + B^2 - B + I = 0 \Rightarrow B^4 = I$. A zatem $B^{100} = I$.

[1] Więcej przykładów i agitacji w filmie z czasów pandemicznych: `https://youtu.be/wSGW9EO6Wg8`.

3.1 Uzupełnienie. Wielomian minimalny kontra wyznaczniki

Wielomian minimalny jest ważnym niezmiennikiem podobieństwa, pozwalającym dokonywać rozróżnień między macierzami, których nie rozpoznają inne niezmienniki. Widzimy to choćby biorąc endomorfizmy dane macierzami

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$$

mającymi identyczne rzędy, ślady i wyznaczniki. Mają takie same wielomiany charakterystyczne, ale ich wielomiany minimalne są różne! Identyczność ma oczywiście wielomian minimalny $\lambda - 1$, a drugi z wypisanych endomorfizmów – $(\lambda - 1)^2$. W dalszej części uzupełnienia wielomian minimalny endomorfizmu ϕ oznaczamy przez $m_\phi(\lambda)$, zaś wielomian minimalny macierzy A – przez $m_A(\lambda)$.

Jak uzasadnić, że macierze podobne mają identyczny wielomian charakterystyczny? Do szukania wielomianu minimalnego służy świadomość, postulowana już na wykładzie, że wielomian minimalny dzieli każdy wielomian, od którego endomorfizm jest zerowy.

Obserwacja 3.6

Niech $\phi \in \mathrm{End}(V)$. Wówczas $w \in K[\lambda]$ spełnia $w(\phi) = 0$ wtedy i tylko wtedy, gdy $m_\phi(\lambda) \,|\, w(\lambda)$.

Dowód. Jeśli $m_\phi(\lambda)|w(\lambda)$, to $w(\phi) = 0$, ponieważ $w = m_\phi(\lambda) \cdot v(\lambda)$, czyli $w = v(m_\phi(\phi)) = v(0) = 0$. Z drugiej strony, jeśli $w(\phi) = 0$. Dzielimy $w(\lambda)$ z resztą przez $m_\phi(\lambda)$, dostając $w(\lambda) = q(\lambda)m_\phi(\lambda) + r(\lambda)$, przy czym $r(\lambda) = 0$ lub $\deg r < \deg m_\phi$. W drugim przypadku jednak $0 = m_\phi q(\phi) + r(\phi) = r(\phi)$. Jednak $r(\phi) = 0$ przeczy minimalności $m_\phi(\lambda)$, o ile $r(\lambda) \neq 0$. □

Przykład, niech

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

Wielomian charakterystyczny tego wielomianu to $-\lambda^3 + 2\lambda^2 - \lambda = -\lambda(\lambda - 1)^2$. Jeśli wielomian charakterystyczny nie jest minimalnym, to wielomian minimalny dzieli jeden z czynników kwadratowych, czyli $\lambda(\lambda - 1)$ lub $(\lambda - 1)^2$. Rachunek pokazuje, że $A(A - I) = 0$, więc wielomian minimalny dzieli $\lambda(\lambda - 1)$. Skoro A oraz $A - I$ nie są zerowe, to wielomianem minimalnym A jest $\lambda(\lambda - 1)$.

Dokonajmy odpowiedniego rachunku.

Obserwacja 3.7

Macierze podobne mają ten sam wielomian minimalny

Dowód. Załóżmy, że dla macierzy $A, B, C \in M_n(K)$ mamy $C^{-1}AC = B$. Jest jasne, że dla dowolnego $w \in K[\lambda]$ mamy $w(B) = C^{-1}w(A)C$. Stąd także:

$$C^{-1}m_B(A)C = m_B(B) = 0, \quad \text{oraz} \quad C^{-1}m_A(A)C = m_A(B) = 0.$$

Z pierwszej równości i obserwacji wyżej wynika, ze $m_A(\lambda) \,|\, m_B(\lambda)$. Z drugiej równości natomiast wynika, e $m_B(\lambda) \,|\, m_A(\lambda)$. Skoro wielomiany minimalne są unormowane, to $m_B(\lambda) = m_A(\lambda)$. □

Obserwacja 3.8

Niech ϕ będzie endomorfizmem skończenie wymiarowej przestrzeni liniowej V nad ciałem K. Niech $W_1, \ldots, W_k$ będą ϕ-niezmienniczymi podprzestrzeniami V takimi, że $V = W_1 + \ldots + W_k$. Niech m_i będą wielomianami minimalnymi endomorfizmów $\phi|_{W_i}$ przestrzeni W_i, dla $i = 1, \ldots, k$. Wówczas wielomian minimalny ϕ równy jest najmniejszej wspólnej wielokrotności $m_1, \ldots, m_k$.

Dowód. Niech m będzie najmniejszą wspólną wielokrotnością wielomianów minimalnych $m_1, \ldots, m_n$ endomorfizmów ϕ_i. A zatem m jest wielokrotnością każdego z tych wielomianów. Zapiszmy $m = g_i m_i$, dla każdego i, dla pewnych $g_i \in K[\lambda]$. Dla każdego $v \in W_i$ mamy $m(\phi)(v) = g_i(\phi)(m_i(\phi))v = g_i(\phi)(0) = 0$, a zatem $m(\phi)$ jest endomorfizmem zerowym na każdym z W_i. Skoro V jest sumą W_i, to $m(\phi) = 0$ na całym V. A zatem wielomian minimalny ϕ jest dzielnikiem m.

Teraz pokażemy, że wielomian minimalny ϕ jest podzielny przez m. Skoro wielomian minimalny m_ϕ zeruje się na ϕ, to w szczególności jądro $m_\phi(\phi)$ zawiera każdą z W_i. A zatem ograniczając $m_\phi(\phi)$ do W_i widzimy, że również tu jest to przekształcenie zerowe. W szczególności $m_i(\lambda)$ są dzielnikami $m_\phi(\lambda)$, dla każdego i, czyli ich najmniejsza wspólna wielokrotność – wielomian m, również dzieli $m_\phi(\lambda)$. □

Do problemu wyznaczania wielomianu minimalnego można podejść w następujący sposób. Dla $\phi \in \mathrm{End}(V)$ wybieramy $v \neq 0$ i rozważamy podprzestrzeń cykliczną V_v. Wówczas za m_1 można wziąć wielomian charakterystyczny ϕ obciętej do tej podprzestrzeni. Następnie kładziemy $v_1 = v$, $W_1 = V_v$ i wybieramy $v_2 \notin W_1$. Za W_2 bierzemy podprzestrzeń cykliczną rozpiętą przez v_2. I wielomian charakterystyczny ϕ obciętego do W_2 to jest, jak się okazuje, m_2. Kontynuując w ten sposób i korzystając z faktu, że V jest skończenie wymiarowa dostajemy zbiór podprzestrzeni ϕ-niezmienniczych W_i, których sumą jest całe V (i nie wnikamy czy to jest suma prosta). Wielomian minimalny ϕ jest natomiast najmniejszą wspólną wielokrotnością m_i. Oczywiście stosowanie takiego algorytmu może być uciążliwe.

* * *

Teoria wielomianów minimalnych oferuje alternatywne podejście do teorii wartości własnych, bez użycia pojęcia wyznacznika (i wielomianu charakterystycznego). Upraszcza to szereg problemów dotyczących problemu istnienia wartości własnych. Wyłożymy tu podejście przedstawione przez prof. Szymiczka, opierające się na słynnym artykule Axlera Down with determinants! *Amer. Math. Monthly* 102 (1995).

Twierdzenie 3.3

Niech $\phi \in \mathrm{End}(V)$, gdzie V jest przestrzenią liniową nad ciałem K. Jeśli wielomian minimalny $m_\phi(\lambda)$ endomorfizmu ϕ rozkłada się nad ciałem K na iloczyn czynników liniowych, to endomorfizm ϕ ma przynajmniej jedną wartość własną. Dokładniej, co najmniej jeden z pierwiastków wielomianu $m_\phi(\lambda)$ jest wartością własną endomorfizmu ϕ.

Dowód. Wielomian $m_\phi(\lambda)$ rozkłada się nad ciałem K na czynniki liniowe, czyli dla pewnych elementów $a_1, \ldots, a_k \in K$ istnieje rozkład:

$$w_\phi(\lambda) = (\lambda - a_1)(\lambda - a_2) \cdots (\lambda - a_k).$$

W takim razie dla każdego wektora $\alpha \in V$ mamy:

$$0 = m_\phi(\phi)(\alpha) = ((\phi - a_1 \,\mathrm{id})(\phi - a_2 \,\mathrm{id}) \cdots (\phi - a_k \,\mathrm{id}))(\alpha).$$

Złożenie $(\phi - a_1 \,\mathrm{id})(\phi - a_2 \,\mathrm{id}) \cdots (\phi - a_k \,\mathrm{id})$ jest więc endomorfizmem zerowym. Oznacza to, że co najmniej jeden z endomorfizmów $\phi - a_i \,\mathrm{id}$ ma niezerowe jądro. Inaczej, gdyby wszystkie miały zerowe jądra, to również ich złożenie miałoby tę własność, co jest niemożliwe, bo ich złożenie jest zerowe. W takim razie a_i jest wartością własną endomorfizmu ϕ □

Argument wykorzystany wyżej wykorzystamy do drobnego wzmocnienia tezy. Uzasadnimy, że jeśli $m_\phi(\lambda)$ ma choć jeden pierwiastek w ciele K, to jest on wartością własną endomorfizmu ϕ.

Obserwacja 3.9: proste ćwiczenie

Niech $\phi \in \mathrm{End}(V)$, gdzie V jest przestrzenią liniową nad ciałem K. Niech α będzie niezerowym wektorem w V oraz niech $g \in K[\lambda]$ będzie dowolnym wielomianem. Wtedy

$$\phi(v) = av \quad \Longrightarrow \quad g(\phi)(v) = g(a)v.$$

Innymi słowy, jeśli a jest wartością własną endomorfizmu ϕ, to $g(a)$ jest wartością własną $g(\phi)$.

Poniższe twierdzenie daje charakteryzacje wartości własnych dla endomorfizmów przestrzeni liniowych dowolnego wymiaru, także nieskończonego (o ile dla danego endomorfizmu istnieje wielomian minimalny, co nie jest gwarantowane).

Twierdzenie 3.4

Niech V będzie przestrzenią liniową nad ciałem K. Dla $a \in K$ oraz endomorfizmu $\phi \in \mathrm{End}(V)$ następujące warunki są równoważne:

(a) a jest wartością własną endomorfizmu ϕ,

(b) a jest pierwiastkiem wielomianu minimalnego $m_\phi(\lambda)$ endomorfizmu ϕ.

Dowód. Jeśli zachodzi (a), to niech α będzie wektorem własnym o wartości własnej a. Zatem $\phi(\alpha) = a\alpha$ oraz $\alpha \neq 0$. Na podstawie lematu wyżej mamy $m_\phi(\phi)(\alpha) = m_\phi(a) \cdot \alpha$. Ponieważ $m_\phi(\phi) = 0$, z definicji wielomianu minimalnego, to $m_\phi(a)\alpha = 0$. Skoro jednak $\alpha \neq 0$, to $m_\phi(a) = 0$, co daje (b).

Odwrotnie, jeśli $m_\phi(a) = 0$ oraz $a \in K$, to na mocy twierdzenia Bezout istnieje wielomian $q \in K[\lambda]$, taki że $m_\phi(\lambda) = (x-a)q(\lambda)$. Zatem $0 = m_\phi(\phi) = (\phi - a\,\mathrm{id})q(\phi)$. Przy tym jednak $q(\phi)$ nie jest endomorfizmem zerowym, gdyż stopień wielomianu q jest mniejszy od stopnia wielomianu minimalnego m_ϕ. Istnieje więc wektor $\alpha \in V$, że $\beta = q(\phi)(\alpha) \neq 0$. Zatem

$$(\phi - a\,\mathrm{id})(\beta) = (\phi - a\,\mathrm{id})q(\phi)(\alpha) = m_\phi(\phi)(\alpha) = 0.$$

Stąd a jest wartością własną endomorfizmu ϕ. □

Wielomian minimalny pozwala elegancko dowodzić liniową niezależność wektorów własnych $\alpha_1, \ldots, \alpha_k$, które odpowiadają parami różnym wartościom własnym $a_1, \ldots, a_k$ endomorfizmu ϕ. Jest on nieco krótszy i mniej „błyskotliwy" niż ten podany na wykładzie. Otóż gdyby $\alpha_1, \ldots, \alpha_k$ były liniowo zależne, to istniałyby skalary $c_1, \ldots, c_k$, nie wszystkie równe 0, takie że $c_1\alpha_1 + \ldots + c_k\alpha_k = 0$. Rozpatrzmy obraz obu stron tej przez endomorfizm $(\phi - a_2\,\mathrm{id})(\phi - a_3\,\mathrm{id}) \cdots (\phi - a_k\,\mathrm{id})$. Wykorzystując fakt, że endomorfizmy $\phi - a_i\,\mathrm{id}$ są przemienne, otrzymujemy:

$$c_1(a_1 - a_2)\ldots(a_1 - a_k)\alpha_1 = 0.$$

Stąd $c_1 = 0$. Podobnie dowodzimy, że $c_j = 0$, dla $j = 2, \ldots, k$ (usuwając ze złożenia k endomorfizmów $\phi - a_i\,\mathrm{id}$ czynnik $\phi - a_j\,\mathrm{id}$).

Zachęcam Czytelnika do zajrzenia do skryptu prof. Szymiczka[2], aby przekonać się, że całą budowaną przez nas teorię można uzyskać opierając się na wielomianie minimalnym. Przytoczę na koniec dwa twierdzenia, charakteryzujące znane nam już klasy endomorfizmów diagonalizowalnych i triangularyzowalnych.

Twierdzenie 3.5: Charakteryzacja endomorfizmów diagonalizowalnych

Niech V będzie przestrzenią liniową skończonego wymiaru nad ciałem K. Endomorfizm $\phi \in \mathrm{End}(V)$ jest diagonalizowalny nad ciałem K wtedy i tylko wtedy, gdy wielomian minimalny endomorfizmu ϕ rozkład się nad ciałem K na iloczyn parami różnych czynników liniowych.

Twierdzenie 3.6: Charakteryzacja endomorfizmów triangularyzowalnych

Niech V będzie przestrzenią liniową skończonego wymiaru nad ciałem K. Endomorfizm $\phi \in \mathrm{End}(V)$ jest triangularyzowalny nad ciałem K wtedy i tylko wtedy, gdy wielomian minimalny endomorfizmu ϕ rozkład się nad ciałem K na iloczyn czynników liniowych (niekoniecznie różnych).

Również inne klasy endomorfizmów można charakteryzować w języku wielomianu minimalnego, np. w języku rozkładalności na parami różne czynniki nierozkładalne (są to tzw. endomorfizmy półproste). Zainteresowanych odsyłam do notatek Keitha Conrada[3].

[2] http://www.math.us.edu.pl/zatl/szymiczek/referaty/AlgebraLiniowa3.pdf
[3] https://kconrad.math.uconn.edu/blurbs/linmultialg/semisimple.pdf

3.2 Dodatek. Macierze wielomianowe i podobieństwo

Celem tego dodatku będzie przeformułowanie problemu podobieństwa macierzy nad dowolnym ciałem do problemu równoważności tzw. macierzy wielomianowych, który z kolei ma do pewnego stopnia posmak znany z eliminacji Gaussa. Zanim powiemy czym są owe macierze wielomianowe, potrzebujemy kilku wstępnych uwag motywacyjnych, osadzających nas w kontekście wykładu.

Badanie podobieństwa macierzy, które nie są triangularyzowalne, a więc nie są podobne do żadnej macierzy górnotrójkątnej, jest możliwe. Chciałbym tu dwa podejścia do tego problemu.

Jedno podejście mówi, że można próbować określać inne typy „postaci kanonicznych", do których można by próbować sprowadzać dowolną macierz, niekoniecznie triangularyzowalną. Postać Jordana, którą poznamy na kolejnym wykładzie, jest typem „postaci kanonicznej" dla macierzy triangularyzowalnej – za jej pomocą można po prostu zliczać wszystkie klasy podobieństwa macierzy tego typu. Jest wiele innych form, ogólniejszych – w szczególności niegórnotrójkątnych. Najsłynniejszą jest zapewne postać kanoniczna (normalna) Frobeniusa, którą opiszemy w innym dodatku, w oparciu o wyniki uzyskane w niniejszym tekście. Drugie podejście mówi coś bardzo zaskakującego. Mianowicie — podobieństwo macierzy nie zależy od ciała, nad którym je rozważamy. Co to znaczy?

Twierdzenie 3.7

Jeśli $K \subseteq L$ są ciałami oraz $A, B \in M_n(K)$, to jeśli A, B są podobne nad L, to są podobne nad K. Innymi słowy, jeśli istnieje macierz odwracalna $C \in GL_n(L)$, taka że $C^{-1}AC = B$, to istnieje macierz odwracalna $D \in GL_n(K)$, taka że $D^{-1}AD = B$.

Przyjmijmy, że $K = \mathbb{R}$ oraz $L = \mathbb{C}$. Powyższy wynik mówi, że do sprawdzenia czy dwie macierze rzeczywiste, potencjalnie nietriangularyzowalne, są podobne nad $\mathbb{R}$, wystarczy sprawdzić czy są one podobne jako macierze zespolone, a jako macierze zespolone są triangularyzowalne. Jeśli są podobne jako macierze zespolone, to muszą być podobne również nad $\mathbb{R}$. Podejście to więc zapewni jedynie istnienie macierzy podobieństwa, bez wskazywania jej.

Dowód tego rezultatu jest stosunkowo nietrudny, gdy K jest ciałem nieskończonym. Oto szkic odpowiedniego rozumowania.

- Niech $K \subseteq L$ oraz niech $A, B \in M_n(K)$ będą podobne nad L. Zatem istnieje macierz odwracalna $C \in M_n(L)$, że $B = C^{-1}AC$.
- Inaczej mówiąc $CB - AC = 0$, czyli wyrazy C są rozwiązaniem jednorodnego układu równań liniowych o współczynnikach w K. A zatem przestrzeń F rozpięta przez wyrazy C jest skończenie wymiarowa nad K.
- Z poprzedniego punktu mamy $C = C_1e_1 + \ldots + C_re_r$, gdzie $C_i \in M_n(K)$ oraz $\{e_i\}$ – liniowo niezależny układ wektorów z F (nad K).
- Rozważmy wielomian r zmiennych postaci $P(t_1, \ldots, t_r) = \det(t_1C_1 + \ldots + t_rC_r) \in K[t_1, \ldots, t_r]$.
- Skoro $\det C \neq 0$, to $P(e_1, \ldots, e_r) \neq 0$, a więc $P \neq 0$.
- Skoro K jest nieskończone (!), to istnieją $\lambda_1, \ldots, \lambda_r \in K$, że $P(\lambda_1, \ldots, \lambda_r) \neq 0$.
- Zatem $D = \lambda_1C_1 + \ldots \lambda_rC_r \in M_n(K)$ jest odwracalna i łatwo widzieć, że $B = D^{-1}AD$.

Jest jasne, że powyższy dowód nie może być zastosowany dla ciał skończonych. Wprowadzimy teraz technologię pozwalającą uzyskać ten rezultat w pełnej ogólności. Na pierwszy rzut oka wygląda ona na skomplikowaną i nie mającą zastosowań. Jest wprost przeciwnie. Czytelnika zainteresowanego szerszym kontekstem odsyłam do podręcznika[4] prof. Tadeusza Kaczorka, pt. *Zastosowanie macierzy wielomianowych i wymiernych w teorii układów dynamicznych*. W kolejnych dodatkach uzyskane tu wyniki posłużą do dowodu istnienia tzw. postaci kanonicznej Frobeniusa, uogólniającej na dowolne ciała postać kanoniczną Jordana, dostępną jedynie dla endomorfizmów triangularyzowalnych.

[4] Dostępny online: `https://pbc.biaman.pl/Content/77/zast_mac_wiel_i_wym.pdf`.

Definicja 3.14: Macierz wielomianowa

Przez MACIERZ WIELOMIANOWĄ ROZMIARU $m \times n$ NAD CIAŁEM K nazywamy macierz o m wierszach i n kolumnach, której wyrazami są wielomiany o współczynnikach w ciele K. Każdą taką macierz można zapisać w formie $P(\lambda) = [p_{ij}(\lambda)]$, gdzie $p_{ij} \in K[\lambda]$ oraz

$$P(\lambda) = \begin{bmatrix} p_{11}(\lambda) & \dots & p_{1n}(\lambda) \\ \vdots & \ddots & \vdots \\ p_{m1}(\lambda) & \dots & p_{mn}(\lambda) \end{bmatrix},$$

lub grupując odpowiednio potęgi λ w postaci:

$$P(\lambda) = P_0 + \lambda P_1 + \ldots + \lambda^d P_d,$$

gdzie $P_0, \ldots, P_d \in M_{m\times n}(K)$. Największe k takie, że $P_k \neq 0$, nazywamy STOPNIEM macierzy wielomianowej $P(\lambda)$, ozn. $\deg P(\lambda)$. Zbiór wszystkich macierzy wielomianowych rozmiaru $m \times n$ nad K oznaczamy $M_{m\times n}(K[\lambda])$.

Przykład. Niech $p_{11} = \lambda^2 + \lambda + 1$, $p_{12} = \lambda + 3$, $p_{21} = 2\lambda^2 + \lambda + 3$, $p_{22} = 3\lambda^2 - 3$. Wtedy

$$P(\lambda) = \begin{bmatrix} \lambda^2 + \lambda + 1 & \lambda + 3 \\ 2\lambda^2 + \lambda + 3 & 3\lambda^2 - 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \lambda^2 + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \lambda + \begin{bmatrix} 1 & 3 \\ 3 & -3 \end{bmatrix}.$$

Zbiór $M_n(K[\lambda])$ traktować można jako podzbiór $M_n(K(\lambda))$, gdzie $K(\lambda)$ jest ciałem funkcji wymiernych, czyli zbiorem elementów postaci

$$\frac{f}{g}, \text{ gdzie } f, g \in K[\lambda], g \neq 0.$$

Stąd mają sens definicje wprowadzone dla macierzy nad ciałem: dodawanie i mnożenie macierzy, operacje elementarne (i ich macierze), postaci: schodkowa i zredukowana, rząd, wyznacznik (dla elementów $M_n(K[\lambda])$ będący wielomianem!), macierz odwrotna, macierz dołączona.

Nie każda macierz wielomianowa $P(\lambda)$ ma element odwrotny w $M_n(K(\lambda))$, to znaczy taki $Q(\lambda)$, że $P(\lambda)Q(\lambda) = Q(\lambda)P(\lambda) = 1$. Jeśli bowiem $\det P(\lambda) \neq 0$, to

$$P(\lambda)^{-1} = \frac{Adj(P(\lambda))}{\det P(\lambda)} \in M_n(K(\lambda)),$$

skąd łatwo wynika, że $P(\lambda)^{-1} \in M_n(K[\lambda]) \Leftrightarrow \det P(\lambda) \in K \setminus \{0\}$.

Naszym celem jest dowód następującego twierdzenia.

Twierdzenie 3.8: Frobenius, 1878

Niech K będzie ciałem oraz niech $A, B \in M_n(K)$. Następujące warunki są równoważne.

- Macierze A, B są podobne nad K.
- Istnieją odwracalne macierze wielomianowe $P(\lambda), Q(\lambda) \in M_n(K[\lambda])$ takie, że

$$\lambda I - B = P(\lambda)^{-1} \cdot (\lambda I - A) \cdot Q(\lambda).$$

Ogólna filozofia powyższego wyniku jest następująca. Macierz A jest podobna do macierzy B nad K wtedy i tylko wtedy, gdy macierz wielomianową $\lambda I - A$ można sprowadzić przy pomocy elementarnych operacji wierszowych i kolumnowych „ograniczonych do $M_n(K[\lambda])$" do $\lambda I - B$. Inaczej mówiąc: podobieństwo macierzy A i B nad ciałem K jest tym samym, co równoważność (por. pierwsze uzupełnienie) macierzy wielomianowych $\lambda I - A$, $\lambda I - B$ nad $K[\lambda]$.

Aby udowodnić twierdzenie Frobeniusa, wprowadzimy kilka pojęć pomocniczych – przede wszystkim namiastkę dzielenia macierzy wielomianowych.

Definicja 3.15: Lewostronna wartość macierzy wielomianowej

Rozważmy macierz wielomianową $P(\lambda) = P_0 + \lambda P_1 + \ldots + \lambda^d P_d$, gdzie $P_i \in M_n(K)$. Dla każdej $A \in M_n(K)$ definiujemy wówczas:

$$P(A) = P_0 + AP_1 + \ldots + A^d P_d$$

i nazywamy LEWOSTRONNĄ WARTOŚCIĄ $P(\lambda)$ dla $\lambda = A$.

Oto rezultat, który motywuje powyższa definicję i pozwala na uzyskiwanie „wartości" macierzy wielomianowych, mimo tego, że w przypadku pierścieni nieprzemiennych ewaluacja jest skomplikowanym obiektem (w pierwszym semestrze pisaliśmy o tym przy okazji tw. Bezout dla wielomianów).

Obserwacja 3.10: Lemat 1 – uogólnienie tw. Bezout

Niech $P(\lambda) \in M_n(K[\lambda])$ oraz niech $A \in M_n(K)$. Wówczas następujące warunki są równoważne:

- $P(A) = 0$,
- istnieje $F(\lambda) \in M_n(K[\lambda])$ taki, że $P(\lambda) = (\lambda I - A)F(\lambda)$.

Dowód. Niech $P(\lambda) = P_0 + \lambda P_1 + \ldots + \lambda^d P_d$, dla pewnych $P_i \in M_n(K[\lambda])$. Jeśli $P(A) = 0$, to

$$\begin{aligned} P(\lambda) =& P(\lambda) - P(A) = \\ =& (P_0 + \lambda P_1 + \ldots + \lambda^d P_d) - (P_0 + AP_1 + \ldots + A^d P_d) = \\ =& (\lambda I - A)P_1 + (\lambda^2 I - A^2)P_2 + \ldots + (\lambda^d I - A^d)P_k = \\ =& (\lambda I - A)(...) \end{aligned}$$

Jeśli zaś $P(\lambda) = (\lambda I - A)F(\lambda)$, gdzie $F(\lambda) = F_0 + \lambda F_1 + \ldots + \lambda^s F_s$, to:

$$P(\lambda) = \lambda F_0 + \lambda^2 F_1 + \ldots + \lambda^{s+1} F_s - AF_0 - \lambda AF_1 - \ldots - \lambda^s AF_s,$$

czyli

$$P(A) = AF_0 + A^2 F_1 + \ldots + A^{s+1} F_s - AF_0 - A^2 F_1 + \ldots - A^s AF_s = 0.$$

□

Przechodzimy do dowodu Twierdzenia Frobeniusa. Jedna implikacja jest trywialna. Jeśli $B = T^{-1}AT$, gdzie $T \in M_n(K)$ jest odwracalna, to $\lambda I - B = T^{-1}(\lambda I - A)T$, czyli można położyć $P(\lambda) = Q(\lambda) = T$.

Załóżmy, że $\lambda I - B = P(\lambda)^{-1}(\lambda I - A)Q(\lambda)$, dla pewnych odwracalnych macierzy wielomianowych $P(\lambda), Q(\lambda)$. Równoważnie:

$$P(\lambda)(\lambda I - B) = \lambda P(\lambda) - P(\lambda)B = (\lambda I - A)Q(\lambda),$$

czyli z Lematu 1 mamy

$$AP(A) - P(A)B = 0,$$

równoważnie: $AP(A) = P(A)B$. Wystarczy więc pokazać, że $P(A)$ jest macierzą odwracalną.

Wiadomo, że istnieje $R(\lambda)$, że $P(\lambda)R(\lambda) = 1$. Pokażemy, że $P(A)R(B) = 1$. Zgodnie z dowodem Lematu 1 istnieje $G(\lambda) \in M_n(K[\lambda])$ taka, że $R(\lambda) - R(B) = (\lambda I - B)G(\lambda)$. Stąd

$$\begin{aligned} 1 - P(\lambda)R(B) &= P(\lambda)R(\lambda) - P(\lambda)R(B) = P(\lambda)(R(\lambda) - R(B)) = \\ &= P(\lambda)(\lambda I - B)G(\lambda) = (\lambda I - A)Q(\lambda)G(\lambda). \end{aligned}$$

Wstawiając $\lambda = A$ mamy z Lematu 1 równość $I = P(A)R(B)$, koniec.

Czy sprowadziliśmy problem do łatwiejszego? Jak się okazuje, mamy metody sprawdzania czy macierze wielomianowe są równoważne. Poniżej opowiemy o równoważności z „diagonalnymi" macierzami wielomianowymi. Jaki będzie nasz cel? Zacznijmy od sformułowania ogólnego rezultatu dającego ideę kierunku naszych dalszych rozważań.

Twierdzenie 3.9: postać quasi-diagonalna

Niech $0 \neq A(\lambda) \in M_{m\times n}(K[\lambda])$. Wówczas za pomocą operacji elementarnych[a] zamiany wierszy (kolumn) oraz dodania do wiersza (kolumny) wielokrotności (przez wielomian!) innego wiersza (innej kolumny) można przeprowadzić $A(\lambda)$ do macierzy wielomianowej o postaci blokowej:

$$B(\lambda) = \begin{bmatrix} C(\lambda) & 0 \\ 0 & 0 \end{bmatrix},$$

przy czym $C(\lambda)$ jest diagonalną macierzą wielomianową postaci

$$C(\lambda) = \begin{bmatrix} c_1(\lambda) & 0 & 0 & \dots & 0 \\ 0 & c_2(\lambda) & 0 & \dots & 0 \\ 0 & 0 & c_3(\lambda) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & c_r(\lambda) \end{bmatrix}, \qquad (*),$$

gdzie $c_i(\lambda) \in K[\lambda]$ są niezerowe oraz $c_i(\lambda)$ dzieli $c_{i+1}(\lambda)$.

Inaczej: istnieją odwracalne $P(\lambda) \in M_m(K[\lambda])$, $Q(\lambda) \in M_n(K[\lambda])$, że $C(\lambda) = P(\lambda)^{-1}A(\lambda)Q(\lambda)$.

[a] Nie dopuszczamy mnożenia wiersza/kolumny przez niezerowy wielomian niestały, gdyż macierz takiej operacji elementarnej ma macierz odwrotną, która nie jest wielomianowa.

Warto przemyśleć analogiczny fakt dla macierzy o współczynnikach całkowitych. Dowód polega, na poziomie intuicji, na tym, by wybierać element najniższego stopnia, umieszczać go w pierwszym wierszu i pierwszej kolumnie (przed odpowiednie zamiany wierszy i kolumn), a pozostałe dzielić przez niego z resztą elementy z pierwszego wiersza i kolumny (odejmując pierwszy wiersz/kolumnę przemnożony przez odpowiednie czynniki od pozostałych wierszy/kolumn, modyfikując pierwszy wiersz/kolumnę). Następnie powtarzać ten proces aż w pierwszym wierszu i kolumnie zostaną poza przekątną zera. Oto przykład. Niech

$$A(\lambda) = \begin{bmatrix} \lambda^2 & \lambda - 1 \\ \lambda & \lambda^2 \end{bmatrix}.$$

Za pomocą zamian wierszy i kolumn przenosimy wyraz $\lambda - 1$ na miejsce p_{11}, a dalej przedstawiamy wielomiany w pierwszym wierszu i kolumnie w postaci $p_{ij}(\lambda) = p_{11}(\lambda)q_{ij}(\lambda) + r_{ij}(\lambda)$. Mamy:

$$A_1(\lambda) = \begin{bmatrix} \lambda - 1 & \lambda^2 \\ \lambda^2 & \lambda \end{bmatrix} = \begin{bmatrix} \lambda - 1 & (\lambda-1)(\lambda+1)+1 \\ (\lambda-1)(\lambda+1)+1 & \lambda \end{bmatrix}.$$

Odejmujemy $(\lambda + 1)$ razy pierwszy wiersz od drugiego, czyli otrzymujemy macierz:

$$A_1(\lambda) - \begin{bmatrix} 0 & 0 \\ (\lambda+1)(\lambda-1) & (\lambda+1)\lambda^2 \end{bmatrix} = \begin{bmatrix} \lambda - 1 & (\lambda-1)(\lambda+1)+1 \\ 1 & -\lambda^3 - \lambda^2 + \lambda \end{bmatrix}.$$

Następnie ponownie przenosimy wielomian najniższego stopnia na miejsce w pierwszym wierszu i kolumnie. Następnie powtarzamy ten proces dążąc do tego, aby wszystkie wyrazy w pierwszym wierszu i pierwszej kolumnie były podzielne przez ten wielomian. W przypadku wyżej, zamieniamy wiersze i mamy wielomian stały równy 1, czyli

$$A_2(\lambda) = \begin{bmatrix} 1 & 1\cdot(-\lambda^3 - \lambda^2 + \lambda) \\ 1\cdot(\lambda-1) & (\lambda-1)(\lambda+1)+1 \end{bmatrix}.$$

Teraz odejmujemy od drugiego wiersz pierwszy pomnożony przez $(\lambda - 1)$:

$$A_3(\lambda) = \begin{bmatrix} 1 & -\lambda^3 - \lambda^2 + \lambda \\ 0 & \lambda^4 - \lambda^2 + \lambda \end{bmatrix}.$$

Wreszcie, odejmujemy od drugiej pierwszą kolumnę przemnożoną przez $-\lambda^3 - \lambda^2 + \lambda$, otrzymując:

$$B(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & \lambda^4 - \lambda^2 + \lambda \end{bmatrix}.$$

Gdyby macierz była większego rozmiaru, to cały proces powtórzylibyśmy dla podmacierzy powstającej po usunięciu pierwszego wiersza i pierwszej kolumny. Zauważmy, że każdy wyraz tej podmacierzy jest dzielnikiem wyrazu $c_1(\lambda) = 1$ w macierzy $B(\lambda)$. Ogólny dowód indukcyjny prowadzimy analogicznie.

Analogiem wyniku wyżej jest wynik dla macierzy całkowitoliczbowych, ujęty w dwa zadania niżej.

Zadanie 3.1. *Niech $A = (a_{ij}) \in M_{m\times n}(\mathbb{Z})$ będzie macierzą o wyrazach całkowitych i niech d będzie minimalnym niezerowym elementem ze zbioru $\{|a_{ij}|, 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$. Określamy „$\mathbb{Z}$-operacje elementarne" na wierszach i kolumnach macierzy A, czyli:*

- *dodanie do wiersza innego wiersza przemnożonego przez liczbę całkowitą,*
- *zamiana wierszy,*
- *przemnożenie wiersza przez niezerową liczbę całkowitą,*
- *dodanie do kolumny innej kolumny przemnożonej przez liczbę całkowitą,*
- *zamiana kolumn,*
- *przemnożenie kolumny przez niezerową liczbę całkowitą,*

Pokaż, ze jeśli nie można obniżyć d przez „$\mathbb{Z}$-operacje elementarne", to d dzieli wszystkie wyrazy A.

Zadanie 3.2. *Pokaż, że z pomocą „$\mathbb{Z}$-operacji elementarnych" na wierszach i kolumnach każdą niezerową macierz całkowitoliczbową rozmiaru d sprowadzić można do pewnej macierzy $A = (a_{ij})$, gdzie*

$$a_{ij} = \begin{cases} d_i, & i = j, \\ 0, & i \neq j. \end{cases},$$

przy czym d_i jest dzielnikiem d_{i+1}, dla $i = 1, 2, \ldots, d-1$.

Przykład zastosowania, ponownie realizujący pomysł przedstawiony dla macierzy wielomianowych.

$$\begin{bmatrix} 2 & 4 & 4 \\ -6 & 6 & 12 \\ 10 & 4 & 16 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 0 \\ -6 & 18 & 24 \\ 10 & -16 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 18 & 24 \\ 0 & -16 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 20 \\ 0 & -16 & -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 20 \\ 0 & 0 & 156 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 156 \end{bmatrix}.$$

Istnienie formy opisanej przez $(*)$ nie jest jeszcze wystarczające. W istocie naszym celem jest pokazanie, że wielomiany $c_i(\lambda)$ występujące w $(*)$ są, z dokładnością do współczynnika wiodącego, jednoznacznie wyznaczone. W tym celu potrzebne jest wprowadzenie dodatkowych definicji.

Definicja 3.16: Minory

Dla macierzy $A \in M_{m\times n}(K)$ oraz dla każdego $k \leqslant \min(m, n)$

- przez $A^{i_1,\ldots,i_k}_{j_1,\ldots,j_k}$ oznaczamy macierz powstałą z A przez wykreślenie $m-k$ wierszy i $n-k$ kolumn o indeksach różnych odpowiednio od i_s oraz j_l,
- MINORAMI STOPNIA k nazywamy wyznaczniki macierzy $A^{i_1,\ldots,i_k}_{j_1,\ldots,j_k}$.

Definicja 3.17: Dzielniki wyznacznikowe

Niech $A(\lambda) \in M_{m\times n}(K[\lambda])$. Dla $1 \leqslant j \leqslant \min(m.n)$ przez $d_j(A(\lambda))$ określamy

- unormowany NWD wszystkich minorów stopnia j macierzy $A(\lambda)$, jeśli istnieje,
- 0, jeśli wszystkie minory stopnia j są zerowe.

Wielomiany $d_j(A(\lambda))$ nazywamy DZIELNIKAMI WYZNACZNIKOWYMI $A(\lambda)$.

Powyższe definicje pozwalają na sformułowanie wyniku, do którego dążymy.

Twierdzenie 3.10: O postaci kanonicznej Smitha

Niech $0 \neq A(\lambda) \in M_{m\times n}(K[\lambda])$ oraz niech istnieją odwracalne $P(\lambda) \in M_m(K[\lambda])$, $Q(\lambda) \in M_n(K[\lambda])$ takie, że $P(\lambda)^{-1}A(\lambda)Q(\lambda)$ ma postać:

$$P(\lambda)^{-1}A(\lambda)Q(\lambda) = \begin{bmatrix} c_1(\lambda) & & & \\ & \ddots & & 0 \\ & & c_r(\lambda) & \\ & 0 & & 0 \end{bmatrix}, \qquad (*)$$

gdzie $c_i(\lambda) \in K[\lambda]$ są unormowane, niezerowe oraz $c_i(\lambda)$ dzieli $c_{i+1}(\lambda)$. Wówczas kładąc $d_0(A(\lambda)) = 1$ mamy, dla $1 \leqslant i \leqslant r$:

$$c_i(\lambda) = \frac{d_i(A(\lambda))}{d_{i-1}(A(\lambda))}.$$

Wielomiany c_i nazywamy wówczas CZYNNIKAMI NIEZMIENNICZYMI $A(\lambda)$.

Rozpoczynamy dowód od następującego lematu.

Obserwacja 3.11: Lemat 2

Niech $A(\lambda) \in M_{m\times n}(K[\lambda])$. Wówczas:

(a) Jeśli $P(\lambda) \in M_{l\times m}(K[\lambda])$ oraz $Q(\lambda) \in M_{n\times p}(K[\lambda])$, i jeśli $j \leqslant \min(m, n)$, to $d_j(A(\lambda))$ jest dzielnikiem $d_j(P(\lambda)A(\lambda))$ oraz $d_j(A(\lambda)Q(\lambda))$.

(b) Jeśli $P(\lambda) \in M_m(K[\lambda])$ oraz $Q(\lambda) \in M_n(K[\lambda])$ są odwracalne oraz $B(\lambda) = P(\lambda)^{-1}A(\lambda)Q(\lambda)$, to dla wszystkich j mamy:

$$d_j(B(\lambda)) = d_j(A(\lambda)).$$

Dowód wymaga następującego ważnego wyniku.

Twierdzenie 3.11: Cauchy-Binet

Jeśli $A \in M_{l\times m}(K)$, $B \in M_{m\times n}(K)$ oraz $r \leqslant \min(l, m, n)$, to

$$\det(AB)^{i_1,\ldots,i_r}_{j_1,\ldots,j_r} = \sum_{1\leqslant k_1<\ldots<k_r\leqslant m} \left(\det A^{i_1,\ldots,i_r}_{k_1,\ldots,k_r}\right)\left(\det B^{k_1,\ldots,k_r}_{j_1,\ldots,j_r}\right).$$

Idea dowodu twierdzenia Cauchy'ego-Bineta jest następująca. Wystarczy pokazać, że jeśli $A \in M_{n\times N}(K)$ oraz $B \in M_{N\times n}(K)$ oraz jeśli $\mathcal{S}$ jest rodziną n elementowych podzbiorów $\{1, \ldots, n\}$ to

$$\det(AB) = \sum_{S\in\mathcal{S}} \det A_s \det B_s,$$

gdzie A_S powstaje z A przez usunięcie kolumn o indeksach spoza S, oraz B_S powstaje przez usunięcie wierszy o indeksach spoza S. Niech $A_1, \ldots A_n$ będą wierszami A oraz $B_1, \ldots, B_n$ będą kolumnami B, traktowanymi jako wektory w K^N. Rozważamy funkcje

$$f(A, B) = \det(AB) = f(A_1, \ldots, A_n, B_1, \ldots, B_n) \text{ oraz } g(A, B) = \sum_{S\in\mathcal{S}} \det A_s \det B_s = g(A_1, \ldots, A_n, B_1, \ldots, B_n).$$

Jak się okazuje, funkcje f, g są jednorodne i addytywne względem każdej współrzędnej (sprawdzamy to poprzez własności wyznacznika). Aby pokazać, że są identyczne wystarczy zatem rozważać przypadek, gdy A_i, B_j są wektorami bazy standardowej. I dalej działamy jak w dowodzie wzoru permutacyjnego.

Przechodzimy do dowodu Lematu 2.

Dowód (a). Skoro każdy minor d stopnia j macierzy $P(\lambda)A(\lambda)$ jest sumą iloczynów $f(P(\lambda))g(A(\lambda))$ odpowiednich minorów stopnia j macierzy $P(\lambda)$ oraz $A(\lambda)$, to skoro każdy $g(A(\lambda))$ jest podzielny przez $d_j(A(\lambda))$, to również d jest podzielny przez $d_j(A(\lambda))$. Zatem $d_j(P(\lambda)A(\lambda))$ jako NWD wszystkich takich d również jest podzielny przez $d_j(A(\lambda))$. Analogicznie pokazujemy, że $d_j(A(\lambda)Q(\lambda))$ dzieli się przez $d_j(A(\lambda))$.

Dowód (b). Z (a) widzimy, że

- $d_j(A(\lambda))$ to dzielnik $d_j(P^{-1}(\lambda)A(\lambda)Q(\lambda)) = d_j(B(\lambda))$,
- $d_j(B(\lambda))$ to dzielnik $d_j(P(\lambda)B(\lambda)Q^{-1}(\lambda)) = d_j(A(\lambda))$.

Zatem $d_j(A(\lambda))$, $d_j(B(\lambda))$ dzielą się wzajemnie. Z twierdzenia o jednoznaczności rozkładu w $K[\lambda]$ wynika stąd, że różnią się one co najwyżej o skalar. Ale dzielniki wyznacznikowe są unormowane, więc dostajemy (b).

Dowód twierdzenia o postaci Smitha macierzy wielomianowej. Mamy $d_j(A(\lambda)) = d_j(P(\lambda)^{-1}A(\lambda)Q(\lambda))$, dla wszystkich j. Patrząc na postać:

$$C(\lambda) = P(\lambda)^{-1}A(\lambda)Q(\lambda) = \begin{bmatrix} c_1(\lambda) & & & \\ & \ddots & & 0 \\ & & c_r(\lambda) & \\ & 0 & & 0 \end{bmatrix},$$

bez trudu widać, że:

$$d_j(C(\lambda)) = \begin{cases} c_1(\lambda)\dots c_j(\lambda), & \text{jeśli } j \leqslant r, \\ 0, & \text{jeśli } j > r. \end{cases}.$$

Stąd $c_1(\lambda) = d_1(A(\lambda))$ oraz $c_{j+1}(\lambda) = d_{j+1}(A(\lambda))/d_j(A(\lambda))$, dla $1 \leqslant j \leqslant r-1$.

Wniosek 3.4: rozwinięcie twierdzenia Frobeniusa

Niech K będzie ciałem oraz niech $A, B \in M_n(K)$. Następujące warunki są równoważne.

- Macierze A, B są podobne nad K.
- Istnieją odwracalne macierze wielomianowe $P(\lambda), Q(\lambda) \in M_n(K[\lambda])$ takie, że $\lambda I - B = P(\lambda)^{-1} \cdot (\lambda I - A) \cdot Q(\lambda)$.
- $\lambda I - A$ oraz $\lambda I - B$ mają identyczne dzielniki wyznacznikowe,
- $\lambda I - A$ oraz $\lambda I - B$ mają identyczne czynniki niezmiennicze.

Wniosek 3.5

Jeśli $A \in M_n(K)$, to A jest podobna do A^T.

Dowód. Macierze wielomianowe $\lambda I - A$ oraz $\lambda I - A^T$ mają identyczne czynniki niezmiennicze/dzielniki wyznacznikowe. □

Powyższy wniosek z Twierdzenia Frobeniusa pozwala wreszcie na uzasadnienie, że jeśli $K \subseteq L$ są ciałami oraz $A, B \in M_n(K)$, to jeśli A, B są podobne nad L, to są podobne nad K. Wystarczy zauważyć, że unormowany $\mathrm{NWD}(p_1, \dots, p_m)$ w $L[\lambda]$ jest równy unormowanemu $\mathrm{NWD}(p_1, \dots, p_m)$ w $K[\lambda]$. Stąd dzielniki wyznacznikowe $\lambda I - A$ oraz $\lambda I - B$ nad K i L są takie same.

3.3 Trivia. Twierdzenie Cayleya-Hamiltona na konkursach

Warto wspomnieć, że studenci mogą brać udział w wielu konkursach na poziomie pierwszych lat studiów matematycznych. Do najstarszych i najsłynniejszych należy amerykański Putnam, największy zasięg i globalny prestiż ma IMC (International Mathematics Competition for University Students), a nie brakuje także konkursów lokalnych. W niektórych krajach np. w Rumunii zadania z analizy matematycznej oraz algebry liniowej i abstrakcyjnej występują na najwyższych poziomach krajowej olimpiady. Zachęcam do pracy z ciekawymi problemami z tych konkursów. Potrafi to dać wiele satysfakcji — bywa i frustrujące. Zacząć można od forum AOPS: `https://artofproblemsolving.com/community/c15_undergraduate_contests`, ale przydatne w tym kontekście będą również książki[5] Titu Andreescu lub Fuzhena Zhanga.

Na zawodach bardzo często korzystać należy z następującej, prostej wersji twierdzenia Cayleya-Hamiltona związanej z Obserwacją 2.2 i dotyczącej macierzy $X \in M_2(K)$, która mówi, że zachodzi równość

$$X^2 - \operatorname{tr}(X)X + \det(X)I = 0.$$

Dowód to po prostu bezpośredni rachunek. Jeśli

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

to macierz $X^2 - \operatorname{tr}(X)X + \det(X)I$ równa jest:

$$\begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} - \begin{bmatrix} (a+d)a & (a+d)b \\ (a+d)c & (a+d)d \end{bmatrix} + \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Oto trzy przykładowe zadania, gdzie wykorzystywany jest powyższy fakt. Są to niestandardowe problemy konkursowe, zwłaszcza zadanie drugie, ale warto docenić jak sprytnie korzysta się ze znanych nam faktów (nie tylko z tego semestru).

Zadanie 1. Niech $A, B \in M_2(K)$, gdzie $\det A = \det B = 1$. Pokaż, że

$$\operatorname{tr}(AB) - \operatorname{tr}(A)\operatorname{tr}(B) + \operatorname{tr}(AB^{-1}) = 0.$$

Rozwiązanie. Z twierdzenia wyżej mamy

$$B^2 - (\operatorname{tr} B)B + I_2 = 0,$$

co po przemnożeniu przez AB^{-1} z lewej daje

$$AB - \operatorname{tr}(B)A + AB^{-1} = 0. \qquad (\dagger)$$

Teraz trzeba zauważyć, że dla dowolnych macierzy kwadratowych $X, Y \in M_n(K)$ oraz $\lambda \in K$ mamy

$$\operatorname{tr}(X+Y) = \operatorname{tr}(X) + \operatorname{tr}(Y) \quad \text{oraz} \quad \operatorname{tr}(\lambda X) = \lambda \operatorname{tr}(X).$$

Zatem obkładając równość (†) funkcją tr otrzymujemy tezę. ■

Zadanie 2. Pokaż, że $\det(X) = 0$, gdzie $X \in M_2(\mathbb{R})$ spełnia, dla pewnego całkowitego $n > 2$, równanie:

$$X^n + X^{n-2} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Pokaż, że istnieje tylko jedna macierz $X \in M_2(\mathbb{Z}_5)$ taka, że:

$$X^5 = \begin{bmatrix} 4 & 2 \\ 4 & 1 \end{bmatrix}.$$

Rozwiązanie. Pierwsza część zadania dotyczy teorii wyznaczników, ale jest bardzo interesująca i użyjemy w niej twierdzenia Cayleya-Hamiltona. Aby pokazać, że $\det(X) = 0$ obłóżmy wyjściową równość wyznacznikiem. Mamy:

$$\det(X^n + X^{n-2}) = \det(X)^{n-2} \det(X^2 + I) = 0.$$

[5] T. Andreescu, *Essential Linear Algebra with Applications. A Problem-Solving Approach*, Birkhauser, 2014; F. Zhang, Linear Algebra. *Challenging Problems for Students* (2nd ed.), Johns Hopkins Studies in the Mathematical Sciences, 1996.

Jeśli jednak $\det(X^2+I)=0$, to przechodząc do wielomianów zespolonych mamy: $\det(X-iI)\det(X+iI)$. A zatem i lub $-i$ są wartościami własnymi X. To jednak oznacza, ze obydwie te liczby są pierwiastkami wielomianu charakterystycznego X, który jest rzeczywisty. Wielomian ten ma stopień 2, a więc musi mieć postać λ^2+1. Stąd jednak mielibyśmy, na mocy twierdzenia Cayleya-Hamiltona, $X^2+I=0$, co znaczyłoby, że lewa strona wyjściowego równania to macierz 0, co jest niemożliwe.

Uwaga. Teraz możliwe jest rozwiązanie wyjściowego równania (zadanie pochodzi z Olimpiady w Rumunii). Otóż stwierdzamy, że wobec $\det(X)=0$ wielomian charakterystyczny macierzy X ma postać $\lambda^2-\mathrm{tr}(X)\lambda$, czyli na mocy tw. C-H mamy $X^2=\mathrm{tr}(X)X$. Zatem mamy: $X^n+X^{n-2}=(\mathrm{tr}(X)^{n-1}-\mathrm{tr}(X)^{n-3})X$. Przechodzimy do śladu macierzy po obydwu stronach. Po lewej ślad wynosi 2 (patrz wyjściowe równanie), a po prawej jest to oczywiście $\mathrm{tr}(X)^n-\mathrm{tr}(X)^{n-2}$. A zatem liczba $\mathrm{tr}(X)$ jest pierwiastkiem wielomianu $x^n+x^{n-2}-2$. Teraz już przechodzimy do analizy. Dla n parzystego okazuje się, że $\mathrm{tr}(X)$ może być równe jedynie -1 oraz 1 (dlaczego?). Dla n nieparzystych jedynym zerem rozważanego wielomianu jest 1. A zatem wyznaczyliśmy ślad i wyznacznik macierzy X, co daje trzy rozwiązania wyjściowego równania.

Przechodzimy do części drugiej. Rozwiążemy zadanie na trzy sposoby.

Sposób 1. Zauważmy, że ślad i wyznacznik macierzy $A=\begin{bmatrix}4 & 2\\ 4 & 1\end{bmatrix}$ to odpowiednio: 0 oraz 1. Szukamy macierzy rozmiaru 2×2, która w piątej potędze ma taki ślad i wyznacznik. To oznacza, na mocy twierdzenia Cayleya-Hamiltona, że

$$(X^5)^2+I=0.$$

Zatem $A^2+I=0$. Czyli $A^5=A$. Zatem A jest rozwiązaniem postawionego równania. Zauważmy dalej, że $\det X^5=1$, a równanie $x^5=1$ ma w $\mathbb{Z}^5$ jedynie rozwiązanie $x=1$. Zatem równanie Cayleya-Hamiltona dla macierzy X ma postać: $X^2-\mathrm{tr}(X)X+I=0$. Podnosimy całość do potęgi piątej:

$$X^{10}=(\mathrm{tr}(X)-I)^5,$$

co nad ciałem $\mathbb{Z}_5$ jest równe po prostu $\mathrm{tr}(X)^5X-I$ (bo I jest przemienna z X oraz zachodzi cudowna tożsamość $(x+y)^5=x^5+y^5$). Co więcej, w ciele $\mathbb{Z}^5$ mamy $x^5=x$, więc tak naprawdę $X^{10}=\mathrm{tr}(X)X-I$. Ale przecież $X^{10}=A^2=-I$, czyli $\mathrm{tr}(X)X=0$, co oznacza $\mathrm{tr}(X)=0$. A zatem $X^2=-I$, czyli $X^5=X=A$.

Sposób 2. Biorąc $m=\mathrm{tr}(X)$, iterujemy trzykrotnie znane nam równanie $X^2-mX+I=0$ dostając:

$$X^5=X^3(mX-I)=mX^4-X^3=mX^2(mX-I)-X(mX-I)=\ldots=(m^4-3m^2+1)X-(m^3-2m)I. \quad (\dagger)$$

Przechodzimy do śladu macierzy po obydwu stronach. Mamy

$$\mathrm{tr}(X^5)=m(m^4-3m^2+1)-2(m^3-2m)=m^5=m,$$

skąd $m=0$ (wiemy ile wynosi ślad X^5). Zatem wracając do (†) mamy $X^5=X$, czyli $X=A$.

Sposób 3. Skoro A spełnia wyjściowe równanie to można pokazać, że dowolne inne rozwiązanie X musi mieć te same wartości własne, co macierz A (stwierdzamy najpierw, że X ma te same wartości własne, co X^5). Ale A ma dwie różne wartości własne: $2, 3$. A zatem X musi być również diagonalizowalna. Natomiast dla macierzy diagonalizowalnej X nad $\mathbb{Z}_5$ mamy $X^5=X$, co daje znowu $X=A$. ■

Zadanie 3. Niech $A, B\in M_2(\mathbb{R})$ spełniają warunek

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

Wyznacz $((BA)^{-1}+BA)^{2019}$.

Rozwiązanie. Korzystając z faktu, że dla dowolnych macierzy kwadratowych X, Y zachodzą równości $\mathrm{tr}(XY)=\mathrm{tr}(YX), \det(XY)=\det(YX)$ widzimy, że

$$\mathrm{tr}(BA)=\mathrm{tr}(AB)=3, \quad \det(BA)=\det(AB)=1.$$

Zatem na mocy twierdzenia Cayleya-Hamiltona:

$$(BA)^2-3(BA)+I=0.$$

Mnożąc uzyskaną równość przez $(BA)^{-1}$ dostajemy:

$$BA+(BA)^{-1}=3I.$$

■

Rozdział 4

Endomorfizmy nilpotentne. Twierdzenie Jordana

Sklasyfikowaliśmy endomorfizmy diagonalizowalne w języku wymiarów podprzestrzeni własnych, a także zbudowaliśmy grunt do rozważania ogólniejszych klas endomorfizmów w języku podprzestrzeni niezmienniczych i macierzy blokowo-diagonalnych. Niestety baza z wektorów własnych nie zawsze istnieje nawet jeśli wielomian charakterystyczny rozkłada się na czynniki liniowe. Co możemy zatem powiedzieć o geometrycznych własnościach takiego endomorfizmu? Jak się okaże, kluczowe będzie rozważenie dodatkowej klasy endomorfizmów — tak zwanych endomorfizmów nilpotentnych. Mają one tę własność, że złożone ze sobą pewną liczbę razy stają się przekształceniem zerowym. Pozwolą one na zrozumienie centralnego wyniku tej części kursu — twierdzenia Jordana.

Definicja 4.1: Klatki Jordana

Niech $a \in K$. Klatką Jordana o wartości własnej a i rozmiarze n nazywamy macierz postaci

$$M = \begin{bmatrix} a & 1 & 0 & \dots & 0 \\ 0 & a & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & a \end{bmatrix} \in M_n(K),$$

przy czym $M = [a]$, gdy $n = 1$. Innymi słowy wyrazy m_{ij} macierzy M spełniają warunki:

$$m_{ij} = \begin{cases} a, & \text{gdy } i = j, \\ 1, & \text{gdy } j - i = 1, \\ 0, & \text{w pozostałych przypadkach.} \end{cases}$$

Powyższe macierze są górnotrójkątne, ale nie są one diagonalizowalne nad ciałem K, dla $n > 1$. Oczywiście $w_M(\lambda) = (a - \lambda)^n$, Widzimy więc, że krotność algebraiczna wartości własnej a wynosi n, podczas gdy jej krotność geometryczna wynosi 1, czyli równa jest wymiarowi przestrzeni rozwiązań układu o macierzy:

$$M - aI = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Jedyną wartością własną macierzy wyżej jest zero. Jest ona macierzą w bazach standardowych endomorfizmu $\phi \in \mathrm{End}(K^n)$ postaci:

$$\phi((x_1, \dots, x_n)) = (x_2, x_3, \dots, x_n, 0).$$

Endomorfizm ϕ zachowuje się jak taśma, która przy każdej iteracji „zrzuca” jedną zmienną. Mamy:

$$\phi^2((x_1, \dots, x_n)) = (x_3, \dots, x_n, 0, 0), \dots, \phi^{n-1}((x_1, \dots, x_n)) = (x_n, 0, \dots, 0), \phi^n((x_1, \dots, x_n)) = (0, 0, \dots, 0).$$

Definicja 4.2: Macierz w postaci Jordana

Mówimy, że macierz $A \in M_n(K)$ jest w POSTACI JORDANA, jeśli istnieją takie liczby całkowite dodatnie $n_1, n_2, \ldots, n_k$ oraz macierze

$$J_1 \in M_{n_1}(K), \quad J_2 \in M_{n_2}(K), \quad \ldots, \quad J_k \in M_{n_k}(K)$$

będące klatkami Jordana takimi, że $n = n_1 + n_2 + \ldots + n_k$ oraz

$$A = \begin{bmatrix} J_1 & 0 & 0 & \ldots & 0 \\ 0 & J_2 & 0 & \ldots & 0 \\ 0 & 0 & J_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & J_k \end{bmatrix}, \qquad (*)$$

czyli A jest macierzą blokową diagonalną o blokach diagonalnych $J_1, \ldots, J_k$. Dla danej wartości własnej λ macierzy A te klatki Jordana J_i, które na swojej przekątnej mają wartość λ nazywamy KLATKAMI JORDANA ODPOWIADAJĄCYMI WARTOŚCI WŁASNEJ λ w postaci $(*)$.

Oto przykład sytuacji, w której rozważamy endomorfizm ϕ przestrzeni wymiaru 16 mający (parami różne) wartości własne $\lambda_1 = 1, \lambda_2 = 3$ o krotnościach algebraicznych $a_1 = 8$, $a_2 = 8$, gdzie:

$$M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \in M_{16}(\mathbb{R}).$$

Macierz $M(\phi)^{st}_{st}$ ma zatem postać Jordana o:

- trzech klatkach odpowiadających wartości własnej 1; klatki te mają rozmiary 4, 3, 1,
- trzech klatkach odpowiadających wartości własnej 3; klatki te mają rozmiary 3, 3, 2.

Główny rezultat dotyczący macierzy w postaci Jordana jest następujący.

Twierdzenie 4.1: Twierdzenie Jordana

Niech V będzie przestrzenią skończenie wymiarową nad ciałem K oraz niech $\phi \in \mathrm{End}(V)$. Jeżeli wielomian charakterystyczny endomorfizmu ϕ rozkłada się nad ciałem K na czynniki liniowe to istnieje baza $\mathcal{A}$ przestrzeni V zwana BAZĄ JORDANA, w której macierz ϕ ma postać Jordana.

W szczególności jeśli ciało K jest algebraicznie domknięte, wówczas każdy endomorfizm $\phi \in \mathrm{End}(V)$ ma w pewnej bazie przestrzeni liniowej V macierz w postaci Jordana. Macierze $A, B \in M_n(K)$ będące w postaci Jordana są podobne nad K wtedy i tylko wtedy, gdy różnią się co najwyżej kolejnością klatek w postaci $(*)$.

Zaczniemy od pokazania, że macierze rozmiaru n o wyrazach w ciele K będące w postaci Jordana są – podobnie jak macierze diagonalne – podobne nad K wtedy i tylko, gdy różnią się kolejnością klatek.

Twierdzenie 4.2

Niech $A \in M_n(K)$ i niech a będzie wartością własną macierzy A. Dla każdego $m = 1, 2, \ldots$ niech

$$q_m = r(A - aI)^{m-1} - r(A - aI)^m.$$

Przyjmujemy też $(A-aI)^0 = I$. Wówczas jeśli $B \in M_n(K)$ jest macierzą w postaci Jordana podobną do A, to w macierzy B:

(1) suma rozmiarów klatek odpowiadających wartości własnej a jest krotnością algebraiczną a.

(2) Liczba klatek Jordana rozmiaru $\geqslant m$ odpowiadających a wynosi q_m.

(3) Liczba klatek Jordana rozmiaru m odpowiadających a wynosi $q_m - q_{m+1}$.

Dowód. Macierze A i B są podobne, więc mają ten sam wielomian charakterystyczny. Stwierdzenie (1) wynika od razu z tego, że wszystkimi pierwiastkami wielomianu charakterystycznego macierzy (górnotrójkątnej) w postaci Jordana są elementy z przekątnej. Suma rozmiarów klatek odpowiadających wartości własnej a równa jest sumie wystąpień a na przekątnej macierzy B. Dowodzimy (2), (3). Zacznijmy od sprawdzenia, że jeśli A jest macierzą rozmiaru n nad K, oraz B jest podobną do niej macierzą w postaci Jordana, to dla każdego $a \in K$ mamy

$$r(A - aI)^k = r(B - aI)^k.$$

Rzeczywiście, skoro $B = C^{-1}AC$, dla pewnej macierzy odwracalnej C, to:

$$C^{-1}(A - aI)C = C^{-1}AC - C^{-1}aIC = B - aI.$$

W szczególności macierze $A - aI$ oraz $B - aI$ są podobne. Mamy więc:

$$\begin{aligned}(B - aI)^k &= (C^{-1}(A - aI)C)^k = \\ &= \underbrace{C^{-1}(A - aI)C \cdot C^{-1}(A - aI)C \cdot \ldots \cdot C^{-1}(A - aI)C}_{k} \\ &= C^{-1}(A - aI)^k C.\end{aligned}$$

Skoro macierze $(B - aI)^k$ oraz $(A - aI)^k$ są podobne, to oczywiście ich rzędy są równe. Zajmiemy się zatem wyznaczeniem rzędów potęg odpowiednich macierzy powstałych z macierzy w postaci Jordana.

Zauważmy, że dla macierzy blokowo-diagonalnej B mamy:

$$B = \begin{bmatrix} J_1 & 0 & 0 & \ldots & 0 \\ 0 & J_2 & 0 & \ldots & 0 \\ 0 & 0 & J_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & J_k \end{bmatrix} \Longrightarrow B^m = \begin{bmatrix} J_1^m & 0 & 0 & \ldots & 0 \\ 0 & J_2^m & 0 & \ldots & 0 \\ 0 & 0 & J_3^m & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & J_k^m \end{bmatrix}.$$

W szczególności:

$$r(B^m) = r(J_1^m) + r(J_2^m) + \ldots + r(J_k^m).$$

Po drugie, zauważmy, że jeśli B jest blokowo-diagonalna, to również $B-aI$ jest blokowo diagonalna. Skoro B jest w postaci Jordana z klatkami $J_1, J_2, \ldots, J_k$ odpowiadającymi wartościom własnym $a_1, \ldots, a_k$, to $B - aI$ ma klatki Jordana $J_1 - aI, J_2 - aI, \ldots, J_k - aI$ o wartościach własnych $a_1 - a, a_2 - a, \ldots, a_n - a$ (i tych samych rozmiarach). Stąd:

$$r(B - aI)^m = r(J_1 - aI)^m + r(J_2 - aI)^m + \ldots + r(J_k - aI)^m.$$

Chcemy wyznaczyć rząd potęgi macierzy postaci:

$$J_i - aI = \begin{bmatrix} a_i - a & 1 & 0 & \ldots & 0 & 0 \\ 0 & a_i - a & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_i - a & 1 \\ 0 & 0 & 0 & \ldots & 0 & a_i - a \end{bmatrix} \in M_{n_i}(K).$$

Oczywiście jeśli $a_i - a \neq 0$, to rząd każdej potęgi tej klatki równy jest jej rozmiarowi n_i. Jeśli jednak $a_i - a = 0$, to dostajemy macierz mającą jedynki nad przekątną samych zer. A zatem jeśli (po ewentualnym przenumerowaniu) przyjmiemy, że $a_i - a = 0$ dla $i = 1, 2, 3, \ldots, s$, gdzie s to liczba klatek Jordana macierzy B odpowiadających wartości własnej a, to mamy:[1]

$$r(B - aI)^{m-1} - r(B - aI)^m = \sum_{i=1}^{s} r(J_i - aI)^{m-1} - r(J_i - aI)^m.$$

Ile wynosi składnik opisany wyżej? Przypomnijmy, że

$$J_i - aI = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \end{bmatrix} \in M_{s_i}(K).$$

Innymi słowy $J_i - aI$ jest macierzą w bazach standardowych (znanego nam już) endomorfizmu $\phi \in \mathrm{End}(K^{s_i})$ postaci $\phi(x_1, \ldots, x_{s_i}) = (x_2, x_3, \ldots, x_{n_i}, 0)$. Jest zatem jasne, że

$$\dim \operatorname{im} \phi^m = r(J_i - aI)^m = \begin{cases} n_i - m, & m \leqslant n_i \\ 0, & m \geqslant n_i. \end{cases}.$$

Stąd dla $i = 1, 2, \ldots, s$ mamy:

$$r(J_i - aI)^{m-1} - r(J_i - aI)^m \in \{0, 1\},$$

przy czym różnica ta równa jest

- 0 dokładnie wtedy, gdy $m > n_i$, czyli gdy klatka J_i ma rozmiar $< m$,
- 1 dokładnie wtedy, gdy $m \leqslant n_i$, czyli gdy klatka J_i ma rozmiar $\geqslant m$.

W szczególności mamy:

- $r(B - aI)^0 - r(B - aI)^1$ to liczba kl. Jordana macierzy B odp. wart. własnej a rozmiaru $\geqslant 1$,
- $r(B - aI)^1 - r(B - aI)^2$ to liczba kl. Jordana macierzy B odp. wart. własnej a rozmiaru $\geqslant 2$,
- $r(B - aI)^2 - r(B - aI)^3$ to liczba kl. Jordana macierzy B odp. wart. własnej a rozmiaru $\geqslant 3$, itd.

□

Powyższe rozważania pozwalają sformułować wniosek, wynikający oczywiście z twierdzenia Jordana.

Wniosek 4.1

Niech K będzie ciałem algebraicznie domkniętym oraz niech $A, B \in M_n(K)$. Następujące warunki są równoważne:

- macierze A, B są podobne nad K,
- dla każdego $a \in K$ oraz dla każdego $k \geqslant 1$ mamy: $r(A - aI)^k = r(B - aI)^k$.

Ustaliwszy podstawowe obserwacje spróbujemy teraz zrozumieć na co składa się uzasadnienie twierdzenia Jordana. Dowód ma bowiem dwie części. Jedna dotyczy sytuacji, gdy wszystkie wartości własne endomorfizmu są równe 0, a druga — to rozważenie sytuacji, gdy endomorfizm ma różne niezerowe wartości własne. W tej części wykładu zajmiemy się pierwszym wątkiem, dotyczącym tak zwanych endomorfizmów nilpotentnych, które w pewnej iteracji są przekształceniami zerowymi.

[1] Dla $i > s$, czyli dla klatek J_i nie odpowiadających wartości własnej a mamy

$$r(J_i - aI)^{m-1} = r(J_i - aI)^m = n_i.$$

Dla dowodu twierdzenia Jordana kluczowe jest zrozumienie endomorfizmu $\phi - a\,\mathrm{id}$ przestrzeni n wymiarowej V w sytuacji, gdy ϕ ma w bazie Jordana $\mathcal{J} = (\beta_1, \ldots, \beta_n)$ jedną klatkę odpowiadającą wartości własnej a postaci:

$$M(\phi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} a & 1 & 0 & \ldots & 0 & 0 \\ 0 & a & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a & 1 \\ 0 & 0 & 0 & \ldots & 0 & a \end{bmatrix}.$$

Wywnioskowaliśmy, że mamy:

$$(\phi - a\,\mathrm{id})^n(\beta_n) = 0, \quad (\phi - a\,\mathrm{id})^{n-1}(\beta_{n-1}) = 0, \quad (\phi - a\,\mathrm{id})^{n-2}(\beta_{n-2}) = 0, \quad \ldots \quad (\phi - a\,\mathrm{id})(\beta_1) = 0.$$

Oczywiście nietrudno wnioskować stąd także, że dla dowolnego wektora $v \in V$ mamy $(\phi - a\,\mathrm{id})^n(v) = 0$. Tego typu własność ma kluczowe znaczenie dla zrozumienia działań endomorfizmów.

Definicja 4.3: Endomorfizm nilpotentny

Endomorfizm ϕ skończenie wymiarowej przestrzeni liniowej V nazywamy NILPOTENTNYM, jeśli istnieje liczba całkowita $r > 0$ taka, ze

$$\phi^r = 0. \qquad (\heartsuit)$$

Najmniejsze k spełniające $(\heartsuit)$ nazywamy STOPNIEM NILPOTENTNOŚCI ϕ. Macierz $A \in M_n(K)$ nazywamy NILPOTENTNĄ STOPNIA k, jeśli $A^r = 0$ oraz $A^{r-1} \neq 0$.

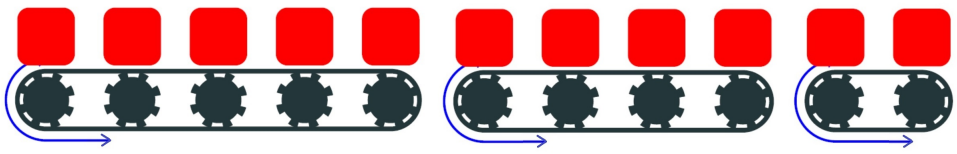
Zauważmy, że dla dowolnego endomorfizmu nilpotentnego stopnia r postaci $\phi \in \mathrm{End}(V)$ istnieje niezerowy wektor $v \in V$ taki, że $\phi^r(v) = 0$ oraz $\phi^{r-1}(v) \neq 0$. Gdyby okazało się, że sama przestrzeń jest wymiaru r, to ciąg wektorów $\phi^{r-1}(v), \phi^{r-2}(v), \ldots, \phi(v), v$ staje się bazą przestrzeni V, w której endomorfizm ϕ ma macierz

$$\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{bmatrix} \in M_r(K). \qquad (*)$$

Zauważmy, że nilpotentność może działać w bardziej subtelny sposób. Biorąc endomorfizm nilpotentny stopnia 5 przestrzeni 11-wymiarowej dany w bazie standardowej macierzą

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

widzimy, że w istocie działa on „niezależnie" na podprzestrzeniach $\mathrm{lin}(\epsilon_1, \ldots, \epsilon_5)$, $\mathrm{lin}(\epsilon_6, \ldots, \epsilon_9)$ oraz $\mathrm{lin}(\epsilon_{10}, \epsilon_{11})$, co można ilustrować za pomocą trzech działających jednocześnie taśmociągów.



Źródło: `http://qirui.li/` – polecam slajdy z bardzo pomysłowymi poglądowymi ilustracjami tematów z algebry liniowej.

Twierdzenie 4.3: Twierdzenie Jordana - przypadek nilpotentny

Niech $\phi \in \mathrm{End}(V)$, gdzie V jest n wymiarowa nad ciałem K. Następujące warunki są równoważne:

(1) ϕ jest endomorfizmem nilpotentnym,

(2) istnieje baza przestrzeni V, w której ϕ ma macierz blokowo-diagonalną o k klatkach Jordana odpowiadających wartości własnej 0 postaci:

$$\begin{bmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix} \in M_{n_i}(K),$$

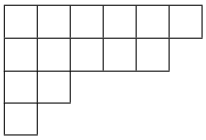
gdzie $n_i > 0$ oraz $n_1 + n_2 + \ldots + n_k = n$.

Jedyną nietrywialną implikacją jest (1) $\Rightarrow$ (2). Aby ją udowodnić, musimy doprecyzować to jakiej bazy szukamy. W tym celu potrzebujemy dodatkowych pojęć kombinatorycznych.

Definicja 4.4: Podział (partycja) liczby całkowitej, diagramy Younga

Przez PODZIAŁ (lub PARTYCJĘ) λ nieujemnej liczby całkowitej n rozumiemy ciąg $(n_1, \ldots, n_k)$ dodatnich liczb całkowitych n_i, które nazywamy **blokami**, spełniającymi warunki: $n = n_1 + \ldots + n_k$ oraz $n_1 \geqslant \ldots \geqslant n_k$.

Geometryczną reprezentacją podziału λ jest tzw. DIAGRAM YOUNGA (lub tableaux Younga) złożony z k wierszy, gdzie i-ty wiersz ma n_i elementów, np. dla podziału $(6, 5, 2, 1)$ liczby 14 jest to:



Twierdzenie 4.4: Baza Jordana endomorfizmu nilpotentnego

Niech V będzie przestrzenią skończenie wymiarową i niech $\phi \in \mathrm{End}(V)$ będzie endomorfizmem nilpotentnym. Wówczas istnieją liczby dodatnie $n_1, \ldots, n_k$, których suma jest równa $\dim V$ oraz baza V postaci:

$$\underbrace{\alpha_1^1, \ldots, \alpha_{n_1}^1}_{n_1}, \underbrace{\alpha_1^2, \ldots, \alpha_{n_2}^2}_{n_2}, \ldots, \underbrace{\alpha_1^k, \ldots, \alpha_{n_k}^k}_{n_k},$$

taka że na jej elementach endomorfizm ϕ działa następująco[a], dla każdego $1 \leqslant i \leqslant k$:

$$\alpha_{r_i}^i \xrightarrow{\phi} \alpha_{r_i-1}^i \xrightarrow{\phi} \ldots \xrightarrow{\phi} \alpha_1^i \xrightarrow{\phi} 0.$$

Jeśli $V_i = \mathrm{lin}(\alpha_1^i, \ldots, \alpha_{n_i}^i)$, to macierz $\phi|_{V_i}$ jest w bazie $(\alpha_1^i, \ldots, \alpha_{n_i}^i)$ klatką Jordana rozmiaru r_i.

[a] Innymi słowy układ $\alpha_1^i, \ldots, \alpha_{n_i}^i$ jest bazą przestrzeni cyklicznej wymiaru n_i rozpiętej na $\alpha_{n_i}^i$.

W jaki sposób twierdzenie to wiąże się z podziałami i diagramami Younga? Postulujemy wyżej podział liczby $\dim V$ postaci $(n_1, \ldots, n_k)$, a z nim — diagram o k wierszach, z których każdemu odpowiada jedna klatka Jordana. Do i-tego wiersza diagramu wpisujemy elementy $\alpha_1^i, \ldots, \alpha_{n_i}^i$. Są to elementy bazy przestrzeni cyklicznej rozpiętej przez wektor $\alpha_{n_i}^i$. Zauważmy, że i-ty element tego ciągu należy do zbioru $\ker \phi^i \setminus \phi^{i-1}$, a pokażemy, że i-ta kolumna stanowi dopełnienie bazy $\ker \phi^{i-1}$ do $\ker \phi^i$.

Rozważmy kilka przykładów. Oto postaci Jordana macierzy endomorfizmów nilpotentnych przestrzeni K^4 i odpowiadające im diagramy Younga z wpisanymi wektorami z bazy Jordana.

- Macierz endomorfizmu o czterech zerowych klatkach Jordana: $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

  diagram Younga z wpisaną bazą Jordana

α_1^1
α_1^2
α_1^3
α_1^4

  oraz rozkład na sumę prostą podprzestrzeni cyklicznych $V = V_{\alpha_1^1} \oplus V_{\alpha_1^2} \oplus V_{\alpha_1^3} \oplus V_{\alpha_1^4}$.

- Macierz endomorfizmu o trzech klatkach Jordana: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

  diagram Younga z wpisaną bazą Jordana

α_1^1	α_2^1
α_1^2	
α_1^3	

  oraz rozkład na sumę prostą podprzestrzeni cyklicznych $V = V_{\alpha_2^1} \oplus V_{\alpha_1^2} \oplus V_{\alpha_1^3}$.

- Macierz endomorfizmu o dwóch klatkach Jordana: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

  diagram Younga z wpisaną bazą Jordana

α_1^1	α_2^1
α_1^2	α_2^2

  oraz rozkład na sumę prostą podprzestrzeni cyklicznych $V = V_{\alpha_2^1} \oplus V_{\alpha_2^2}$.

- Macierz endomorfizmu o dwóch klatkach Jordana: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

  diagram Younga z wpisaną bazą Jordana

α_1^1	α_2^1	α_3^1
α_1^2		

  oraz rozkład na sumę prostą podprzestrzeni cyklicznych $V = V_{\alpha_3^1} \oplus V_{\alpha_1^2}$.

- Macierz endomorfizmu o jednej klatce Jordana: $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

  diagram Younga z wpisaną bazą Jordana

α_1^1	α_2^1	α_3^1	α_4^1

  oraz rozkład na sumę prostą podprzestrzeni cyklicznych $V = V_{\alpha_4^1}$.

Przechodzimy do dowodu o istnieniu bazy Jordana endomorfizmu nilpotentnego. Zaczniemy od prostej obserwacji dotyczących obrazów i jąder kolejnych iteracji endomorfizmu nilpotentnego.

Obserwacja 4.1: Łańcuchy jąder i obrazów endomorfizmu nilpotentnego

Niech V będzie przestrzenią liniową wymiaru n, niech $\phi \in \mathrm{End}(V)$ będzie endomorfizmem nilpotentnym. Rozważmy najmniejszą liczbę całkowitą r, taką że $\phi^r = 0$. Wówczas:

$$0 \subsetneq \ker\phi \subsetneq \ker\phi^2 \subsetneq \cdots \subsetneq \ker\phi^{r-1} \subsetneq \ker\phi^r = V, \quad V \supsetneq \operatorname{im}\phi \supsetneq \operatorname{im}\phi^2 \supsetneq \cdots \supsetneq \operatorname{im}\phi^{r-1} \supsetneq \operatorname{im}\phi^r = 0.$$

Dowód. Skoro dla dowolnego $v \in V$ mamy $\phi^m(v) = 0 \Rightarrow \phi^{m+1}(v) = 0$ oraz dla dowolnych podprzestrzeni $V_1 \subseteq V_2$ przestrzeni V mamy $\phi(V_1) \subseteq \phi(V_2)$, to jest jasne, że zachodzą postulowane inkluzje. Należy jedynie wykazać, że są one właściwe. Startujemy od zauważenia, dla dowolnej iteracji ϕ^m endomorfizmu ϕ mamy $\dim\ker\phi^m + \dim\operatorname{im}\phi^m = n$, co oznacza, że wystarczy wykazanie, że jeden z ciągów inkluzji zawiera same inkluzje właściwe. Pokażemy tezę dla ciągu obrazów.

Jeśli dla pewnego $m < r$ mamy $\operatorname{im}\phi^m = \operatorname{im}\phi^{m+1}$, to wobec $\phi(\phi^m(V)) = \phi^{m+1}(V)$ widzimy, że endomorfizm ϕ obcięty do $\operatorname{im}\phi^m$ jest izomorfizmem na ϕ^{m+1}, a zatem $\ker(\phi|_{\operatorname{im}\phi^m}) = 0$. Stąd, skoro $\operatorname{im}(\phi|_{\operatorname{im}\phi^m}) = \operatorname{im}(\phi^{m+1})$ mamy $\operatorname{im}\phi^m = \operatorname{im}\phi^{m+1} = \operatorname{im}\phi^{m+2} = \cdots = \operatorname{im}\phi^r = 0$. Zatem ϕ^m jest przekształceniem zerowym, co przeczy minimalności r. □

Przechodzimy do dowodu. Dla $\dim V = 0, 1$ teza jest jasna. Przechodzimy do kroku indukcyjnego. Niech $\dim V = n$ i załóżmy, że teza jest prawdziwa dla endomorfizmów nilpotentnych przestrzeni wymiaru mniejszego od n. Niech r będzie najmniejszą liczbą taką, że $\phi^r = 0$. Jeśli $r = 1$, to dowolna baza przestrzeni V spełnia warunki twierdzenia (przekształcenie jest zerowe). Jeśli $r > 1$, to zgodnie z poprzednią obserwacją $0 \neq \operatorname{im}\phi \subsetneq V$, czyli $0 < \dim\operatorname{im}\phi < \dim V$. Podprzestrzeń $\operatorname{im}\phi$ jest ϕ-niezmiennicza i można zdefiniować endomorfizm

$$\overline{\phi} : \operatorname{im}\phi \to \operatorname{im}\phi \quad \text{dany wzorem} \quad \overline{\phi(v)} = \phi(v).$$

Jest to endomorfizm przestrzeni wymiaru mniejszego niż $\dim V$, więc z założenia indukcyjnego istnieje baza Jordana przestrzeni $\operatorname{im}\phi$, postaci (uwaga – wprowadzamy górny indeks, nie zaś potęgę wektora):

$$\mathcal{B} = \underbrace{\alpha_1^1, \ldots, \alpha_{r_1}^1}_{r_1}, \underbrace{a_1^2, \ldots, \alpha_{r_2}^2}_{r_2}, \ldots, \underbrace{\alpha_1^s, \ldots, \alpha_{r_s}^s}_{r_s},$$

w której endomorfizm $\overline{\phi}$ ma postać Jordana. Możemy wpisać te wektory do odpowiadającego tej bazie diagramu Younga o s wierszach długości r_i (jego kształt nie musi być oczywiście „prostokątny"):

α_1^1	$\ldots$	$\alpha_{r_1}^1$
α_1^2	$\ldots$	$\alpha_{r_2}^2$
$\vdots$	$\ddots$	$\vdots$
α_1^s	$\ldots$	$\alpha_{r_s}^s$

Zauważmy najpierw, że jądro $\overline{\phi}$ rozpinają elementy pierwszej kolumny, tzn. $\ker(\overline{\phi}) = \ker(\phi) \cap \operatorname{im}(\phi) = \operatorname{lin}(\alpha_1^1, \ldots, \alpha_1^s)$. Istotnie, mamy z założenia indukcyjnego o postaci $\mathcal{B}$ założenie

$$\phi(\alpha_i^j) = \alpha_{i-1}^j,$$

przy czym przyjmujemy $\alpha_0^j = 0$. Stąd warunek $\operatorname{im}(\phi) \ni v = \sum a_{ij} \cdot \alpha_i^j$ implikuje $\phi(v) = \sum a_{ij} \cdot \alpha_{i-1}^j$. Skoro układ $\mathcal{B}$ jest liniowo niezależny, to $\overline{\phi}(v) = \phi(v) = 0$ tylko wtedy, gdy $a_{ij} = 0$, dla $i > 1$, czyli gdy $v \in \operatorname{lin}(\alpha_1^1, \ldots, \alpha_1^s)$. Zatem $\ker\overline{\phi} = \operatorname{lin}(\alpha_1^1, \ldots, \alpha_1^s)$ i $\dim\ker\overline{\phi} = s$.

Rozważmy wektory $\alpha_{r_1}^1, \alpha_{r_2}^2, \ldots, \alpha_{r_s}^s$ bazy $\mathcal{B}$ rozpinające, zgodnie z założeniem indukcyjnym, podprzestrzenie cykliczne wymiarów $r_1, r_2, \ldots, r_s$. Skoro są to elementy $\operatorname{im}\phi$, to istnieją taki układ wektorów $\alpha_{r_1+1}^1, \ldots, \alpha_{r_s+1}^s \in V$, że (jakoby przedłużających z prawej strony kolejne wiersze diagramu wyżej):

$$\phi(\alpha_{r_1+1}^1) = \alpha_{r_1}^1, \quad \phi(\alpha_{r_2+1}^2) = \alpha_{r_2}^2, \quad \ldots, \quad \phi(\alpha_{r_s+1}^s) = \alpha_{r_s}^s.$$

Dodany układ liczy $\dim\ker\overline{\phi} = \dim\ker\phi \cap \operatorname{im}\phi = s$ elementów.

W zależności od tego, czy zachodzi równość $\ker\overline{\phi} = \ker\phi$, czy ostre zawieranie $\ker\overline{\phi} \subsetneq \ker\phi$, uzupełniamy (lub nie) układ wektorów $\alpha_1^1, \ldots, \alpha_1^s$ do bazy $\ker\phi \supseteq \ker\overline{\phi}$ układem $k - s$ wektorów $\alpha_1^{s+1}, \ldots, \alpha_1^k$.

Rozważamy jedynie drugi przypadek (pierwszy wynika z niego bezpośrednio). Twierdzimy, że poniższy układ jest bazą V:

$$\mathcal{A} = \mathcal{B} \cup \{\alpha^1_{r_1+1}, \ldots, \alpha^s_{r_s+1}\} \cup \{\alpha^{s+1}_1, \ldots, \alpha^k_1\}$$

Istotnie, układ $\mathcal{B}$ ma $\dim \operatorname{im} \phi$ elementów, a dołączony do niego układ wektorów jest równoliczny z bazą $\ker \phi$. Łącznie więc układ $\mathcal{A}$ ma $\dim \ker \phi + \dim \operatorname{im} \phi = \dim V$ wektorów. Wystarczy więc pokazać, że $\mathcal{A}$ jest układem liniowo niezależnym.

Każda kombinacja liniowa elementów układu $\mathcal{A}$ jest postaci $v + w + z$, gdzie

$$v \in \operatorname{lin}(\mathcal{B}), \quad w \in \operatorname{lin}\{\alpha^1_{r_1+1}, \ldots, \alpha^s_{r_s+1}\}, \quad z \in \operatorname{lin}\{\alpha^{s+1}_1, \ldots, \alpha^k_1\},$$

i możemy przyjąć dla pewnych $b_{ij}, c_j, d_j \in K$, że:

$$v = \sum b_{ij} \cdot \alpha^j_i, \quad w = \sum c_j \cdot \alpha^j_{r_j+1}, \quad z = \sum d_j \cdot \alpha^t_{r_1}.$$

Wówczas jeśli $v + w + z = 0$, to

$$\phi(v + w + z) = \sum b_{ij} \cdot \alpha^j_{i-1} + c_j \cdot \alpha^j_{r_j} = 0.$$

Skoro $\mathcal{B}$ jest bazą $\operatorname{im} \phi$, to $a_{ij} = 0$, dla $i > 1$ oraz $c_j = 0$. Zatem kombinacja $v + w + z$ równa jest w istocie kombinacji liniowej wektorów z bazy $\ker \phi$. A zatem i jej współczynniki są zerowe, co kończy dowód liniowej niezależności układu $\mathcal{A}$.

Jest jasne, że $\mathcal{A}$ to baza Jordana o diagramie Younga złożonym z k wierszy (pierwsze s ma długość $r_i + 1$, a kolejne $k - s$ ma długość 1). Dowód twierdzenia Jordana w przypadku nilpotentnym jest zakończony.

Wniosek 4.2: Liczba klatek Jordana endomorfizmu nilpotentnego

Niech V będzie przestrzenią skończenie wymiarową oraz $\phi \in \operatorname{End}(V)$. Załóżmy, że macierz T endomorfizmu ϕ w pewnej bazie ma postać Jordana. Liczba klatek Jordana macierzy T rozmiaru $\geqslant k$ równa jest

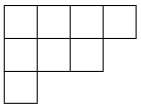
$$q_k = r(T^{k-1}) - r(T^k) = \dim \operatorname{im} \phi^{k-1} - \dim \ker \phi^k = \dim \ker \phi^k - \dim \ker \phi^{k-1}.$$

Oto przykłady bardzo prostych ilustracji pojęć wprowadzonych wyżej.

Zadanie 4.1. *Znajdź postać Jordana i kształt diagramu Younga operatora nilpotentnego ϕ, spełniającego:*

$$\dim(\ker(\phi^2)) = 5, \quad \dim(\ker(\phi^4)) = 8, \quad \dim(\ker(\phi^5)) = 8.$$

Rozwiązanie. Odpowiedź (dlaczego?)



■

Zadanie 4.2. *Wyznacz postać Jordana endomorfizmu ϕ^2, przy czym endomorfizm ϕ przestrzeni wymiaru 9 jest nilpotentny i ma bazę Jordana, której elementy ustawić można do diagramu Younga postaci:*

α^1_1	α^1_2	α^1_3	α^1_4	α^1_5	α^1_6
α^2_1	α^2_2	α^2_3			

.

Rozwiązanie. Z diagramu wyżej odczytujemy, że $\ker \phi^2$ jest przestrzenią wymiaru 4 (pierwsze dwie kolumny), $\dim \ker(\phi^4) = 7$ oraz $\dim \ker(\phi^6) = 9$. Stąd baza Jordana i odpowiedni diagram Younga dla endomorfizmu ϕ^2 ma postać:

α^1_1	α^1_3	α^1_5
α^1_2	α^1_4	α^1_6
α^2_1	α^2_3	
α^2_2		

.

■

4.1 Uzupełnienie. Dowód przez endomorfizmy nierozkładalne

Dowód twierdzenia Jordana jest na tyle istotny, że warto zapoznać się z różnymi na niego spojrzeniami. W głównym wykładzie podajemy dowód indukcyjny odwołujący się najpierw do wydzielenia części nilpotentnej, a potem do uzyskania rozkładu na podprzestrzenie pierwiastkowe. W tym uzupełnieniu pokażemy, że może być on też przepisany w języku rozkładów na tzw. podprzestrzenie nierozkładalne, które okazują się być przestrzeniami pierwiastkowymi. Podejście to pochodzi z notatek K. Ziemiańskiego[2]

Definicja 4.5: Endomorfizm nierozkładalny

Niech V będzie przestrzenią liniową nad ciałem K i niech $\phi \in \mathrm{End}(V)$. Powiemy, że przestrzeń V jest ϕ-nierozkładalna, jeśli nie istnieje nietrywialny rozkład ϕ-niezmienniczy przestrzeni V. Innymi słowy, jeśli $V = U \oplus W$ jest rozkładem ϕ-niezmienniczym, to $U = 0$ lub $V = 0$.

Nietrudno widzieć, że dla dowolnej skończenie wymiarowej przestrzeni liniowej V oraz jej endomorfizmu ϕ istnieje rozkład ϕ-niezmienniczy postaci

$$V = W_1 \oplus \dots W_k,$$

taki że W_i jest $\phi|_{W_i}$-nierozkładalna. Idea dowodu twierdzenia Jordana jest następująca: dla endomorfizmu triangularyzowalnego można wskazać taki rozkład jak wyżej, gdzie $\phi|_{W_i}$ są przesunięciami endomorfizmu nilpotentnego. Przekonajmy się najpierw, że endomorfizmy te w istocie są nierozkładalne.

Obserwacja 4.2

Niech V będzie przestrzenią liniową wymiaru n oraz niech $\psi \in \mathrm{End}(V)$ będzie endomorfizmem nilpotentnym stopnia n. Wówczas

(a) V jest ψ-nierozkładalna,

(b) istnieje baza V, w której ψ jest zadane klatką Jordana rozmiaru n odpowiadającą wartości własnej 0; w szczególności $\dim \ker \psi = 1$.

Dowód. Część (b) uzasadniliśmy już w zasadniczej części wykładu. Mamy bowiem dla $k \leqslant n$ inkluzje $\mathrm{im}(\psi^k) \subsetneq \mathrm{im}(\psi^{k-1})$, skoro ψ jest nilpotentny stopnia n. Stąd $\dim \mathrm{im}\, \phi^{n-1} = 1$ i bierzemy wektor β_1 rozpinający tę przestrzeń. Wtedy istnieje β_2, takie że $\psi(\beta_2) = \beta_1$ itd. W ten sposób konstruujemy szukaną bazę Jordana $(\beta_1, \dots, \beta_n)$ przestrzeni V. Jest jasne, że $\dim \ker \psi = 1$.

Załóżmy, że $V = U \oplus W$ jest ψ-niezmienniczym rozkładem. Wówczas jedna z tych przestrzeni zawiera

$$\beta_n + \sum_{i<n} c_i \beta_i,$$

więc musi być równa V (jako podprzestrzeń ϕ-cykliczna tego wektora). □

Główny rezultat, stanowiący alternatywę dla dowodu z wykładu, jest następujący.

Twierdzenie 4.5: Klasyfikacja triangularyzowalnych endomorfizmów nierozkładalnych

Niech V będzie n-wymiarową przestrzenią liniową nad ciałem K i niech $\phi \in \mathrm{End}(V)$. Załóżmy, że:

- ϕ jest endomorfizmem triangularyzowalnym,
- V jest przestrzenią ϕ-nierozkładalną.

Wówczas $\phi = \psi + \lambda\,\mathrm{id}$, gdzie $\psi \in \mathrm{End}(V)$ jest endomorfizmem nilpotentnym stopnia n.

[2] Dostępne pod adresem `https://www.mimuw.edu.pl/~ziemians/gax2015w/galx15.pdf`.

Dlaczego ten fakt zapewnia nam twierdzenie Jordana? Mówi on bowiem, że triangularyzowalny endomorfizm nierozkładalny czyni w istocie z przestrzeni, na której działa, przestrzeń cykliczną endomorfizmu nilpotentnego $\phi - \lambda$, czyli taką, która w pewnej bazie ma macierz będącą pojedynczą klatką Jordana.

Bierzemy zatem dowolny endomorfizm triangularyzowalny ϕ i dowolny jego nietrywialny rozkład na podprzestrzenie nierozkładalne postaci:

$$V = W_1 \oplus \ldots \oplus W_k.$$

Wielomian charakterystyczny $\phi|_{W_i}$ ma czynniki liniowe, więc na mocy twierdzenia każda z W_i jest w istocie taką przestrzenią, dla której istnieje baza Jordana będąca pojedynczą klatką. Widzimy, że dowód ten w zasadzie całkowicie pomija przestrzenie pierwiastkowe.

Do dowodu części potrzebujemy jeszcze jednego drobnego faktu.

Obserwacja 4.3

Niech V będzie przestrzenią liniową nad ciałem K i niech $\phi \in \mathrm{End}(V)$. Jeśli $V = V_1 \oplus \ldots \oplus V_k$ jest ϕ-niezmienniczym rozkładem, to

$$\ker \phi = (V_1 \cap \ker \phi) \oplus \ldots \oplus (V_k \cap \ker \phi)$$

Dowód. Oczywiście zachodzi zawieranie $\supseteq$. Z drugiej strony jeśli $v = v_1 + \ldots + v_k$ jest elementem $\ker \phi$, to skoro V_i są ϕ-niezmiennicze, to $\phi(v_i) \in V_i$. Z jednoznaczności rozkładu $0 = \phi(v)$ w $V_1 \oplus \ldots \oplus V_k$ mamy zatem $\phi(v_i) = 0$. Zatem $v_i \in V_i \cap \ker \phi$. □

Dowód klasyfikacji triangularyzowalnych endomorfizmów nierozkładalnych

Niech ϕ będzie endomorfizmem triangularyzowalnym przestrzeni n-wymiarowej V. Zatem ϕ ma wartość własną λ. Niech $\psi = \phi - \lambda\,\mathrm{id}$. Łatwo widzieć, że ϕ-nierozkładalność i ψ-nierozkładalność są równoważne, więc wystarczy rozważyć przypadek, gdy $\lambda = 0$.

Załóżmy zatem, że 0 jest wartością własną endomorfizmu ϕ. Dowód nilpotentności ϕ jest indukcją ze względu na wymiar n przestrzeni V. Dla $n = 1$ mamy $\phi = 0$, bo 0 jest wartością własną. Wówczas teza jest spełniona.

Załóżmy, że $n > 1$. Wtedy podobnie jak w dowodzie twierdzenia Jordana w przypadku nilpotentnym stwierdzamy, że $\ker \phi \neq 0$, czyli $\dim \mathrm{im}\, \phi < n$. Niech

$$\mathrm{im}\, \phi = V_1 \oplus \ldots \oplus V_k$$

będzie ϕ-niezmienniczym rozkładem na składniki ϕ-nierozkładalne. Oznaczmy $n_i = \dim V_i$. Po ewentualnym przenumerowaniu możemy zakładać, że 0 jest wartością własną ϕ ograniczonego do składników $V_0, \ldots, V_s$, a nie jest wartością własną tego endomorfizmu po ograniczeniu do składników $V_{s+1}, \ldots, V_k$ (pamiętajmy, że w_ϕ rozkłada się na czynniki liniowe, więc także $w_{\phi|_{W_i}}$ się rozkłada, tzn. każde ograniczenie ϕ do W_i jest triangularyzowalne).

Dla każdego składnika o wartości własnej 0 wybieramy wektor $\alpha_i \in V_i$, że $\phi^{n_i}(\alpha_i) = 0$ oraz $\phi^{n_i - 1}(\alpha_i) \neq 0$. Taki wektor istnieje z założenia indukcyjnego względem podprzestrzeni $\mathrm{im}\, f$. Wybieramy następnie wektor $\beta_i \in V$ taki, że $\phi(\beta_i) = \alpha_i$. To jest dokładnie zabieg wykonany w klasycznym dowodzie. Dalej też rozumowanie wygląda podobnie. To, co w klasycznym dowodzie nazwaliśmy wydłużeniem wiersza diagramu Younga z prawej, teraz formalizujemy określając:

$$V_i = \mathrm{lin}(V_i \cup \{\beta_i\}).$$

Mamy rzecz jasna $\phi(V_i') \subseteq V_i \subseteq V_i'$, czyli V_i' jest ϕ-niezmiennicza. Określamy również przestrzeń rozpiętą przez wektory z jądra ϕ, które dopełniają do bazy przecięcie jądra i obrazu tego endomorfizmu – tak jak w oryginalnym dowodzie dokładaliśmy wiersze o jednej klatce. Innymi słowy, definiujemy podprzestrzeń $W \subseteq \ker \phi$ jako

$$\ker \phi = (\ker \phi \cap \mathrm{im}\, \phi) \oplus W.$$

Oczywiście W jest ϕ-niezmiennicza, jako podprzestrzeń w jądrze.

Naszym celem jest ponownie wykazać, że $V = \lin(V_1' \cup \ldots V_s' \cup V_{s+1} \cup \ldots V_k \cup W)$. Niech $\alpha \in V$. Wówczas $\phi(\alpha)$ jako element $\im \phi$ rozkłada się jednoznacznie (założenie indukcyjne) jako:

$$\phi(\alpha) = \gamma_1 + \ldots + \gamma_k,$$

gdzie $\gamma_i \in V_i$. Dla każdego $i \leqslant s$ mamy przedstawienie

$$\gamma_i = c_i \alpha_i + \phi(\delta_i),$$

gdzie $\delta_i \in V_i$, gdyż $\alpha_i, \ldots, \phi^{n_i - 1}(\alpha_i)$ jest bazą V_i. Tu znowu korzystamy z założenia indukcyjnego i z postaci bazy Jordana dla pojedynczej klatki – czyli postaci bazy dla pojedynczego wiersza diagramu Younga. Weźmy też:

$$\alpha' = (c_1\beta_1 + \delta_1) + \ldots + (c_s\beta_s + \delta_s) + (\phi|_{V_{s+1}})^{-1}(\lambda_{s+1}) + \ldots + (\phi|_{V_k})^{-1}(\lambda_k).$$

Określenie to jest poprawne, gdyż z założenia $\phi|_{V_i}$ jest endomorfizmem odwracalnym na składnikach $V_{s+1}, \ldots, V_k$, skoro wartości własne tego endomorfizmu na tych podprzestrzeniach są niezerowe. Co więcej, mamy:

$$f(\alpha') = \sum_{i=1}^{s} (\phi(c_i\beta_i) + \phi(\delta_i)) + \sum_{i=s+1}^{k} \phi((\phi|_{V_i})^{-1}(\lambda_i)(\lambda_i)) = \sum_{i=1}^{s} (c_i\alpha_i + \phi(\delta_i)) + \sum_{i=s+1}^{k} \delta_i = \phi(\alpha).$$

Wobec tego $\phi(\alpha - \alpha') = 0$, czyli $\alpha \in \lin(\{\alpha'\} \cup \ker \phi)$. Ostatecznie więc

$$\alpha \in \lin\left(\bigcup_{i \leqslant s} V_i' \cup \bigcup_{i > s} V_i \cup \im \phi \cup W\right).$$

Stad dowolny element V należy do odpowiedniej sumy podprzestrzeni V_i', dla $1 \leqslant i \leqslant s$, V_j, dla $s+1 \leqslant j \leqslant k$ oraz W. Pozostaje odnotować, podobnie jak w oryginalnym dowodzie, równość wymiarów:

$$\sum_{i \leqslant s} \dim V_i' + \sum_{j > s} \dim V_j + \dim V = \dim \im \phi + s + \dim \ker \phi - \dim(\ker \phi \cap \im \phi).$$

Z Obserwacji 5.2 wnioskujemy, że:

$$\im \phi \cup \ker \phi = (V_1 \cap \ker \phi) \cap \ldots \cap (V_k \cap \ker \phi).$$

Ponadto z Obserwacji 5.1 mamy $\dim(V_i \cap \ker \phi) = 1$, dla $i \leqslant s$ oraz $\dim(V_j \cap \ker \phi) = 0$, dla $j > s$, gdyż w tym przypadku $\phi|_{V_j}$ jest izomorfizmem. Wobec tego $\dim(\ker \phi \cap \im \phi) = s$. Zatem ostatecznie

$$\sum_{i \leqslant s} \dim V_i' + \sum_{j > s} \dim V_j + \dim W = \dim \im \phi + \dim \ker \phi = n.$$

W szczególności mamy rozkład ϕ-niezmienniczy

$$V = V_1' \oplus \ldots \oplus V_s' \oplus V_{s+1} \oplus \ldots \oplus V_k \oplus W.$$

Tu w zasadzie powtórzyliśmy wcześniejszy dowód. Skoro jednak V jest nierozkładalna (a nie nilpotentna, jak w oryginalnym dowodzie w przypadku nilpotentnym), to tylko jeden z powyższych składników jest niezerowy. Mamy więc trzy przypadki do rozważenia.

- Nie może zachodzić $V = W$, gdyż wtedy $\dim W > 1$ i $f|_W = 0$, więc V jest rozkładalna jako podprzestrzeń własna wymiaru większego niż 1.
- Nie może być $V = V_i$, gdyż wtedy 0 nie byłoby wartością własną V.
- A zatem $V = V_i' = \lin(V_i \cup \{\beta_i\})$. Wtedy $\phi^n = 0$ oraz $\phi^{n-1}(\beta_i) = \phi^{n-2}(\alpha_i) \neq 0$. Teza indukcyjna jes więc spełniona.

Dowód jest zakończony. Czy Czytelnik widzi, że pozwala on na wyznaczanie baz Jordana za pomocą obrazów obcięć pewnych endomorfizmów? Jeśli λ jest wartością własną ϕ, to co daje badanie obrazu endomorfizmu $\psi = \phi - \lambda \id$ i to ograniczonego do podprzestrzeni $\im \psi$?

4.2 Dodatek. Dowód przez przestrzenie ilorazowe

Przyjrzymy się w tym uzupełnieniu uzasadnieniu przypadku nilpotentnego twierdzenia Jordana, używając przestrzeni ilorazowej. Będzie tu mocniej widoczne wykorzystanie Obserwacji 4.1. W przeciwieństwie do dowodu z wykładu, nasze rozumowanie będzie miało charakter do pewnego stopnia konstruktywny, a nie egzystencjalny (dołączanie układu do abstrakcyjnego układu, istniejącego z założenia). Innymi słowy – będziemy starali się kontrolować skąd są elementy konstruowanej bazy Jordana w poszczególnych segmentach odpowiadającego jej diagramu Younga. Stąd też (być może) zawiły język całej konstrukcji.

Załóżmy, że ϕ jest endomorfizmem przestrzeni skończenie wymiarowej i jest nilpotentny stopnia q. Niech $V_i = \ker \phi^i$, dla $1 \leqslant i \leqslant q$. Jak już wiemy $\{0\} \subsetneq \ker \phi \subsetneq \ker \phi^2 \ldots \subsetneq \ker \phi^q = V$. Określamy przestrzenie ilorazowe

$$\overline{V_k} = V_k/V_{k-1}.$$

Szukana baza Jordana endomorfizmu ϕ wypełni diagram Younga o q kolumnach, gdzie w k-tej kolumnie wpisani będą reprezentanci bazy $\overline{V_k}$ (złożonej z warstw – klas abstrakcji). Jaka to będzie baza?

Wiemy, że $v \in V_k$ wtedy i tylko wtedy gdy $\phi(v) \in V_{k-1}$. Zatem przekształcenie $\phi' : \overline{V_k} \to \overline{V_{k-1}}$ dane wzorem

$$v + V_{k-1} \mapsto \phi(v) + V_{k-2}$$

to monomorfizm (intuicja: nie sklejamy elementów przy *uzupełnianiu w lewo*). Dla $0 < k < q$ określamy przestrzeń ilorazową przestrzeni ilorazowej (korzystamy z trzeciego twierdzenia o izomorfizmie):

$$\overline{\overline{V_k}} = \overline{V_k}/\phi'(\overline{V_{k+1}}) = (V_k/V_{k-1})/(\phi(V_{k+1})/V_{k-1}).$$

Intuicja jest następująca: w kolumnie k-tej diagramu Younga mają stać reprezentanci bazy $\overline{V_k}$, spośród których wykluczyć chcemy reprezentantów pochodzących (przez ϕ') z dalszych kolumn $(k+1, k+2, \ldots, q)$. Pozostali reprezentanci reprezentantów bazy $\overline{\overline{V_k}}$ generować będą poszukiwane przez nas podprzestrzenie cykliczne.

Niech $e_1^{(q)}, \ldots, e_{n_q}^{(q)} \in V_q$ będą reprezentantami bazy V_q/V_{q-1} oraz dla $k < q$ niech

$$\overline{\overline{V_k}} = \overline{V_k}/\phi'(\overline{V_{k+1}}) = (V_k/V_{k-1})/(\phi(V_{k+1})/V_{k-1}) \simeq V_k/\phi(V_{k+1}).$$

Niech $n_k = \dim \overline{\overline{V_k}}$ i rozważmy układ liniowo niezależny

$$e_1^{(k)}, \ldots, e_{n_k}^{(k)} \in V_k$$

taki, że $\overline{\overline{e_1^{(k)}}}, \ldots, \overline{\overline{e_{n_k}^{(k)}}}$ jest bazą $\overline{\overline{V_k}}$. Równoważnie, warstwy $\overline{e_1^{(k)}}, \ldots, \overline{e_{n_k}^{(k)}} \in \overline{V_k}$ nie są w $\phi'(V_{k+1})$, czyli nie są postaci $\phi(v) + V_{k-1}$, gdzie $v \in V_{k+1} \setminus V_k$.

Twierdzimy, że dla każdych $1 \leqslant k \leqslant q, 1 \leqslant l \leqslant n_k$ układ $\mathcal{B} = (\mathcal{B})_{lk}$, gdzie

$$\mathcal{B}_{lk} = \{\phi^{k-1}(e_l^{(k)}), \phi^{k-2}(e_l^{(k)}), \ldots, \phi(e_l^{(k)}), e_l^{(k)}\}$$

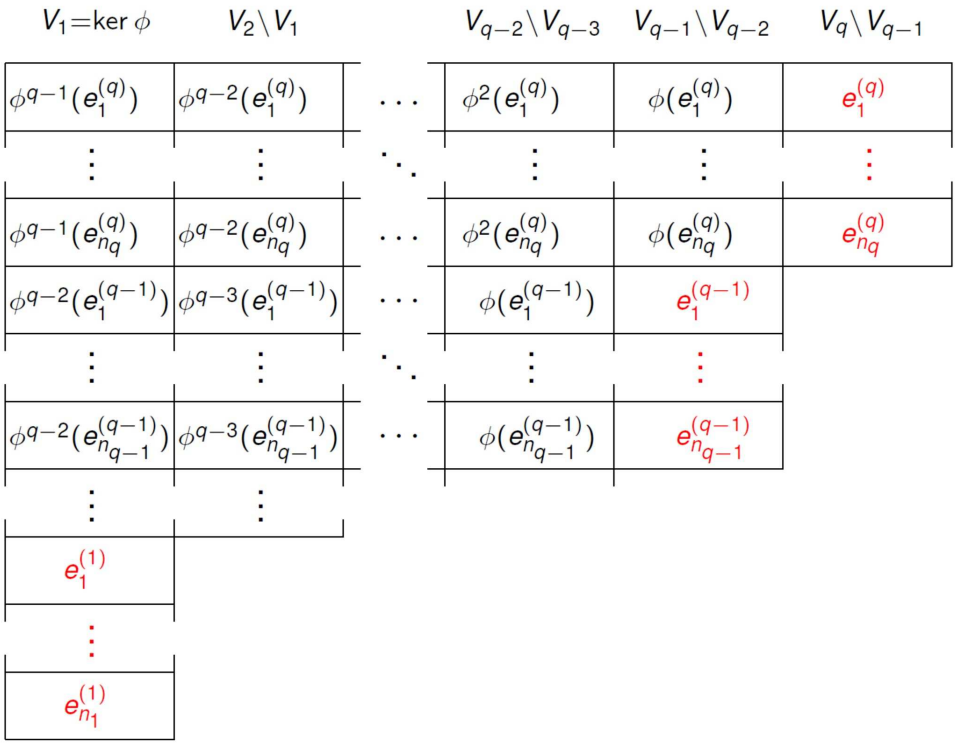
jest bazą przestrzeni V o podziale $\lambda = (\underbrace{q, \ldots, q}_{n_q}, \underbrace{q-1, \ldots, q-1}_{n_{q-1}}, \ldots, \underbrace{1, \ldots, 1}_{n_1})$, przy czym

$$V = \bigoplus_{1 \leqslant k \leqslant q, 1 \leqslant l \leqslant n_k} \operatorname{lin}(\mathcal{B}_{lk}) = \bigoplus_{1 \leqslant k \leqslant q, 1 \leqslant l \leqslant n_k} V_{e_l^{(k)}}$$

oraz macierz ϕ ograniczonego do $V_{e_l^{(k)}}$ to klatka Jordana rozmiaru k.

Rozmieszczenie układu $\mathcal{B}$ w diagramie Younga:

- w kolumnie k są reprezentanci baz V_k/V_{k-1}
- na lewo od pola z elementem $x \in V_k$ znajduje się $\phi(x) \in V_{k-1}$
- w ostatnich polach wierszy są reprezentanci baz $\overline{\overline{V_k}}$.



Podstawowe uwagi.

- Jest jasne, że jeśli $\mathcal{B}$ jest bazą, to jest to baza Jordana ϕ.
- W bazie tej ϕ ma n_i klatek rozmiaru i.
- Wykażemy, że układ $\mathcal{B}$ jest liniowo niezależny oraz, że $|B| = \dim V$.
- Kluczowa idea: wykonanie ϕ^i zeruje elementy z i kolumn.

Punktem wyjścia dowodu jest rozważenie kombinacji liniowej:

$$\sum_{1 \leqslant k \leqslant q,\, 0 \leqslant r \leqslant k-1,\, 1 \leqslant n \leqslant n_k} \lambda_{k,r,n} \cdot \phi^r(e_n^{(k)}) = 0, \qquad (*)$$

dla pewnych $\lambda_{k,r,n} \in K$ (k - kolumna, r - potęga, n - numer).

Chcemy pokazywać, że kolejne współczynniki tej kombinacji są zerowe. Wykonujemy to kolejno dla współczynników przy wektorach

- wpisanych w ostatnie n_q wierszy kolumny q,
- wpisanych w ostatnie n_{q-1} wierszy kolumny $q-1$,
- wpisanych w pierwsze n_q wierszy kolumny $q-1$,
- wpisanych w ostatnie n_{q-2} wierszy kolumny $q-2$,
- wpisanych we wcześniejsze n_{q-1} wierszy kolumny $q-2$,
- wpisanych w pierwsze n_q wierszy kolumny $q-2$,
- wpisanych w ostatnie n_{q-3} wierszy kolumny $q-3$,
- ... (krótko mówiąc to podwójna indukcja).

Przechodzimy do dowodu.

- Krok q.q. Eliminacja $\lambda_{k,r,n}$ dla $e_1^{(q)}, \ldots, e_{n_q}^{(q)}$.

  Obkładamy obie strony $(*)$ przekształceniem ϕ^{q-1}. Po prawej zeruje się wszystko poza kombinacją elementów z ostatniej kolumny, czyli z $V_q \setminus V_{q-1}$ (które zeruje dopiero ϕ^q), czyli mamy:

$$\phi^{q-1}(\underbrace{\lambda_{q,0,1} \cdot e_1^{(q)} + \ldots + \lambda_{q,0,n_q} \cdot e_{n_q}^{(q)}}_{\zeta}) = \phi^{q-1}(\zeta) = 0.$$

  Zatem $\zeta \in \ker \phi^{q-1} = V_{q-1}$, czyli $\overline{\zeta} = \overline{0}$ w $V_q/V_{q-1} = \overline{V_q}$, czyli $\overline{\overline{\zeta}} = \overline{\overline{0}}$ w $\overline{\overline{V_q}}$. A zatem mamy:

$$\lambda_{q,0,1} \cdot \overline{\overline{e_1^{(q)}}} + \ldots + \lambda_{q,0,n_q} \cdot \overline{\overline{e_{n_q}^{(q)}}} = \overline{\overline{0}}.$$

  Ale $\overline{\overline{e_1^{(q)}}}, \ldots, \overline{\overline{e_{n_q}^{(d)}}}$ to baza $\overline{\overline{V_q}}$, czyli

$$\lambda_{q,0,1} = \ldots = \lambda_{q,0,n_q} = 0.$$

  A zatem kombinacja $(*)$ nie zawiera elementów $\mathcal{B}$ wpisanych do kolumny q.

  Dalej obkładamy obie strony $(*)$ przekształceniem ϕ^{q-2}. Po prawej zeruje się wszystko poza elementami $\mathcal{B}$ wpisanymi do $q-1$-kolumny diag. Younga, czyli

$$\phi^{q-2}\left(\sum_i \lambda_{q-1,1,i} \cdot \phi(e_i^{(q)}) + \sum_i \lambda_{q-1,0,i} \cdot e_i^{(q-1)}\right) = \phi^{q-2}(\phi(\zeta^{(0)}) + \zeta^{(1)}) = 0. \quad (**)$$

  Z definicji $\phi(\zeta^{(0)}) + \zeta^{(1)}$ należy do V_{q-1}, ale skoro zeruje się na ϕ^{q-2}, to w istocie należy do V_{q-2}. Zatem w $\overline{V_{q-1}}$ mamy $\overline{0} = \phi'(\overline{\zeta^{(0)}}) + \overline{\zeta^{(1)}}$. Zatem $\overline{\zeta^{(1)}} \in \phi'(V_q)$, czyli $\overline{\overline{\zeta^{(1)}}} = \overline{\overline{0}}$. Ale $\overline{\overline{\zeta^{(1)}}}$ to kombinacja liniowa elementów $\overline{\overline{e_1^{(q-1)}}}, \ldots, \overline{\overline{e_{n_{q-1}}^{(q-1)}}}$ z bazy $\overline{\overline{V_{q-1}}}$, stąd $\lambda_{q-1,0,i}$ są zerami. A zatem kombinacja $(*)$ nie zawiera elementów $\mathcal{B}$ wpisanych do kolumny q.

- Krok q-1.q-1. Eliminacja $\lambda_{k,r,n}$ dla $e_1^{(q-1)}, \ldots, e_{n_{q-1}}^{(q-1)}$. W kombinacji $(*)$ nie ma elementów $\mathcal{B}$ wpisanych do kolumny q.

  Obkładamy obie strony $(*)$ przekształceniem ϕ^{q-2}. Po prawej zeruje się wszystko poza elementami $\mathcal{B}$ wpisanymi do $q-1$-kolumny diag. Younga, czyli

$$\phi^{q-2}\left(\sum_i \lambda_{q-1,1,i} \cdot \phi(e_i^{(q)}) + \sum_i \lambda_{q-1,0,i} \cdot e_i^{(q-1)}\right) = \phi^{q-2}(\phi(\zeta^{(0)}) + \zeta^{(1)}) = 0. \quad (**)$$

  Z definicji $\phi(\zeta^{(0)}) + \zeta^{(1)}$ należy do V_{q-1}, ale skoro zeruje się na ϕ^{q-2}, to w istocie należy do V_{q-2}. Zatem w $\overline{V_{q-1}}$ mamy

$$\overline{0} = \phi'(\overline{\zeta^{(0)}}) + \overline{\zeta^{(1)}}.$$

  Zatem $\overline{\zeta^{(1)}} \in \phi'(V_q)$, czyli $\overline{\overline{\zeta^{(1)}}} = \overline{\overline{0}}$. Ale $\overline{\overline{\zeta^{(1)}}}$ to kombinacja liniowa elementów $\overline{\overline{e_1^{(q-1)}}}, \ldots, \overline{\overline{e_{n_{q-1}}^{(q-1)}}}$ z bazy $\overline{\overline{V_{q-1}}}$, stąd $\lambda_{q-1,0,i}$ są zerami. Stąd w $(**)$ mamy $\phi^{q-1}(\zeta^{(0)}) = 0$, gdzie $\zeta^{(0)}$ to kombinacja reprezentantów $e_1^{(q)}, \ldots, e_{n_q}^{(q)}$, czyli postępując jak w kroku q.q dostajemy: $\lambda_{q-1,1,i} = 0$. Wyeliminowaliśmy zatem z $(*)$ wyrazy z $q-1$-wszej kolumny.

- Krok q-k.q-j. Eliminacja $\lambda_{k,r,n}$ dla $\phi^{k-j}(e_1^{(q-j)}), \ldots, \phi^{k-j}(e_{n_{q-j}}^{(q-j)})$ z kolumny $q-k$.

  Dalej podwójna indukcja dla k, j. W kombinacji $(*)$ nie ma elementów $\mathcal{B}$ wpisanych do kolumn $q-k+1, \ldots, q$. Zaczynamy od kroku q-k.q-k.

  Obkładamy $(*)$ ϕ^{q-k-1}, które zeruje wszystko poza kolumną $q-k$ dając:

$$\phi^{q-k-1}\left(\phi^k(\zeta^{(0)}) + \ldots + \phi(\zeta^{(k-1)}) + \zeta^{(k)}\right) = 0,$$

gdzie $\zeta^{(j)} \in \operatorname{lin}(e_1^{(q-j)}, \ldots, e_{n_{q-j}}^{(q-j)})$ (ale współczynniki są z kolumny $q-k$).

Stąd wywodzi się, że najpierw, że $\overline{\overline{\zeta^{(k)}}} = \overline{\overline{0}}$ w $\overline{\overline{V_{q-k}}}$, czyli $\lambda_{q-k,0,i} = 0$.

Następnie w kroku q-k, q-k+1 $(*)$ obkładamy ϕ^{q-k}, co sprawia, że

$$\phi^{q-k}\left(\phi^{k-1}(\zeta^{(1)}) + \ldots + \phi(\zeta^{(k-1)})\right) = 0,$$

skąd wywodzi się że $\overline{\overline{\zeta^{(k-1)}}} = \overline{\overline{0}}$ w $\overline{\overline{V_{q-k+1}}}$, czyli $\lambda_{q-k,1,i} = 0$...

Postępujemy tak dalej: w q-k.q-j-tym kroku eliminujemy współczynniki $\lambda_{q-k,j,i} = 0$, co ostatecznie eliminuje $q-k$-tą kolumnę. Dostajemy liniową niezależność $\mathcal{B}$.

Aby pokazać, że $\mathcal{B}$ jest bazą wykażemy, że $|\mathcal{B}| = \dim V$. Mamy:

$$\begin{aligned}
|\mathcal{B}| &= q \cdot n_q + (q-1) \cdot n_{q-1} + \ldots + 2 \cdot n_2 + n_1 = \\
&= \sum_{k=1}^{q} k \cdot \dim \overline{\overline{V_k}} = \sum_{k=1}^{q} k \cdot \dim \overline{V_k}/\phi'(\overline{V_{k+1}}) = \\
&= \sum_{k=1}^{q} k \cdot (\dim \overline{V_k} - \dim \phi'(\overline{V_{k+1}})) = \sum_{k=1}^{q} k \cdot (\dim \overline{V_k} - \dim \overline{V_{k+1}}).
\end{aligned}$$

przy czym w ostatniej równości korzystaliśmy z faktu, że ϕ' to monomorfizm.

Nietrudno sprawdzić, że zachodzi równość:

$$\sum_{k=1}^{q} k(a_k - a_{k+1}) = \sum_{k=1}^{q} a_k - q \cdot a_{q+1},$$

zatem skoro w naszej konwencji $a_{q+1} = 0$ (bo $\overline{\overline{V_q}} = \overline{V_q}$), to

$$|\mathcal{B}| = \sum_{k=1}^{q} \dim \overline{V_k} = \sum_{k=1}^{q} \dim(V_k/V_{k-1}) = \sum_{k-1}^{q} \dim V_k - \dim V_{k-1} = \dim V_q - \dim V_0 = \dim V.$$

A zatem $\mathcal{B}$ to układ liniowo niezależny w V oraz $|\mathcal{B}| = \dim V$, czyli jest to baza. Twierdzenie Jordana jest udowodnione.

Wykażmy również w języku przestrzeni ilorazowych, że liczba klatek Jordana rozmiaru $\geqslant m$ to suma

$$\dim \overline{\overline{V_m}} + \ldots + \dim \overline{\overline{V_q}}.$$

Mamy:

$$\begin{aligned}
\sum_{k=m}^{q} \dim \overline{\overline{V_m}} &= \dim V - \dim V_{q-1} + \sum_{k=m}^{q-1} \dim \overline{V_k}/\phi'(\overline{V_{k+1}}) = \\
&= \dim V - \dim V_{q-1} + \sum_{k=m}^{q-1} \dim \overline{V_k} - \dim \overline{V_{k+1}} = \\
&= \dim V - \dim V_{q-1} + \dim \overline{V_m} - \dim \overline{V_q} = \\
&= \dim V - \dim V_{q-1} + \dim V_m - \dim V_{m-1} - \dim V + \dim V_{q-1} = \\
&= \dim V_m - \dim V_{m-1} = \\
&= \dim \ker(\phi - a\,\mathrm{id})^m - \dim \ker(\phi - a\,\mathrm{id})^{m-1} = \\
&= (\dim V - \dim \operatorname{im}(\phi - a\,\mathrm{id})^m) - (\dim V - \dim \operatorname{im}(\phi - a\,\mathrm{id})^{m-1}) = \\
&= \dim \operatorname{im}(\phi - a\,\mathrm{id})^{m-1} - \dim \operatorname{im}(\phi - a\,\mathrm{id})^m = \\
&= r((A - aI)^{m-1}) - r((A - aI))^m = q_m.
\end{aligned}$$

4.3 Trivia. Dowód przez mnożenie macierzy

Dowodów twierdzenia Jordana można przedstawić wiele. O ile trzy (już trzy!) przedstawione wyżej mają wszystkie posmak kontaktu z endomorfizmami nilpotentnymi, bazami i skomplikowanymi strukturami, o tyle jak zwykle gdy chodzi o uzyskanie odpowiedniej postaci blokowej przez (de facto) mnożenie macierzy, można próbować działać na macierzach blokowych. Przedstawimy tu podejście z książki *Matrix Analysis* autorstwa R. Horna i C. Johnsona.

W języku macierzy chodzi o pokazanie, że każda macierz górnotrójkątna $A \in M_n(K)$ jest podobna do macierzy w postaci Jordana. Istotnie, na mocy twierdzenia 3.1 endomorfizm triangularyzowalny ma w pewnej bazie macierz górnotrójkątną. Przyjmijmy, że macierz A ma postać blokowo górnotrójkątną:

$$A = \begin{bmatrix} A_1 & * & \dots & * \\ 0 & A_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix},$$

gdzie $A_i \in M_{n_i}(K)$ są macierzami górnotrójkątnymi mającymi na przekątnych identyczny wyraz a_i, tzn. $A_i = aI + U_i$, gdzie U_i są macierzami ściśle górnotrójkątnymi. Możemy tak zakładać po ewentualnym wykonaniu iloczynu $P^{-1}AP$, gdzie P jest macierzą permutacyjną (tak, by na przekątnej kolejne wartości własne były „po kolei"). Pierwszy krok dowodu polega na wykazaniu, że istnieje macierz odwracalna $T \in M_n(K)$, taka że macierz $T^{-1}AT$ jest blokowo diagonalna postaci:

$$B = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{bmatrix}.$$

W tym celu przypomnijmy, że przez jedynki macierzowe $E_{ij} \in M_n(K)$ rozumiemy macierze, które mają wszystkie wyrazy zerowe. Przypomnijmy też, że zachodzi równość

$$E_{ij} \cdot E_{kl} = \begin{cases} E_{il}, & j = l \\ 0, & j \neq l. \end{cases}.$$

Zauważmy też, że jeśli $X = [x_{ij}] \in M_n(K)$, to mamy $X = \sum_{i,j} x_{ij}E_{ij}$. W szczególności:

$$E_{kl}X = \sum_{i,j} x_{ij}E_{kl}E_{ij} = x_{l1}E_{k1} + x_{l2}E_{k2} + \ldots + x_{ln}E_{kn},$$

czyli mnożenie $E_{kl}X$ zwraca macierz, której jedyny niezerowy wiersz jest k-ty i zawiera wyrazy z l-tego wiersza X. Podobnie

$$XE_{kl} = \sum_{i,j} x_{ij}E_{ij}E_{kl} = x_{1k}E_{1l} + x_{2k}E_{2l} + \ldots + x_{nk}E_{nl},$$

czyli mnożenie XE_{kl} zwraca macierz, której jedyna niezerowa kolumna jest l-ta i zawiera wyrazy z k-tej kolumny X.

Niech $A = [a_{ij}]$, przy czym $a_{ii} = \lambda_i$, jak wyżej. Dla $i < j$ określamy $T_{ij} = I - \theta_{ij}E_{ij}$, tak że mamy $T_{ij}^{-1} = I_n + \theta_{ij}E_{ij}$, gdzie

$$\theta_{ij} = \begin{cases} \frac{a_{ij}}{\lambda_i - \lambda_j}, & \text{gdy } \lambda_i \neq \lambda_j \\ 0, & \text{gdy } \lambda_i = \lambda_j. \end{cases}$$

Istotnie

$$(I - \theta_{ij}E_{ij})(I + \theta_{ij}E_{ij}) = I - \theta_{ij}^2 E_{ij}E_{ij} = \begin{cases} I - \theta_{ij}^2 \cdot 0, & \text{gdy } i \neq j \\ I - 0 \cdot E_{ii}^2, & \text{gdy } i = j. \end{cases}.$$

Jeśli $\lambda_i = \lambda_j$, wówczas $T_{ij}^{-1}AT_{ij} = A$. Jeśli jednak $\lambda_i \neq \lambda_j$, to

$$\begin{aligned} T_{ij}^{-1}AT_{ij} &= (I_n + \theta_{ij}E_{ij})A(I_n - \theta_{ij}E_{ij}) = (A + \theta_{ij}E_{ij}A)(I_n - \theta_{ij}E_{ij}) = \\ &= A - \theta_{ij}AE_{ij} + \theta_{ij}E_{ij}A + \theta_{ij}^2 E_{ij}AE_{ij} = A - \theta_{ij}AE_{ij} + \theta_{ij}E_{ij}A \end{aligned}$$

Zauważmy, że wyraz ij-ty uzyskanej macierzy to

$$a_{ij} - \frac{a_{ij}}{\lambda_i - \lambda_j}\lambda_i + \frac{a_{ij}}{\lambda_i - \lambda_j}\lambda_j = a_{ij} - a_{ij} = 0.$$

Wyraz ij-ty macierzy A, znajdujący się wobec założenia $\lambda_i \neq \lambda_j$ poza jednym z bloków diagonalnych A_{ij}, zamieniliśmy więc na 0. Poza tym macierz $T_{ij}^{-1}AT_{ij}$ zmienia jedynie wyrazy A postaci a_{is}, dla $s > j$ oraz a_{tj}, dla $t < j$. Krótko mówiąc transformacja macierzy A przez sprzęgnięcie macierzą T_{ij} zeruje wyraz ij-ty oraz transformuje wyrazy w i-tym wierszu na prawo od uzyskanego zera oraz w j-tej kolumnie — powyżej uzyskanego zera.

Możemy zatem wykonać serię transformacji postaci $T_{ij}^{-1}AT_{ij}$, dla wszystkich wartości $i < j$, wybierając jednak porządek od wiersza $n-1$ w górę, czyli biorąc kolejno za (i,j) pary

$$(n-1,n),(n-2,n-1),(n-2,n),(n-3,n-2),(n-3,n-1),(n-3,n),\dots(1,2),\dots,(1,n).$$

Każda z tych transformacji zmieni na zerowy wyraz $ij-$ty poprzedniej macierzy, o ile $\lambda_i \neq \lambda_j$ — czyli gdy a_{ij} nie jest w bloku diagonalnym A_{ij}, i żadne z wcześniej uzyskanych zer nie zostaną naruszone. W ten sposób macierz A przetransformowana będzie w B. Uzyskaliśmy pierwszy krok dowodu.

Chcemy teraz doprowadzić macierz blokowo-diagonalną A do postaci Jordana. Każdy z bloków górnotrójkątnych $A_i \in M_{n_i}(K)$ tej macierzy ma na przekątnej ten sam wyraz, nazwijmy go a_i. Zapiszmy teraz

$$A_i = a_i I_{n_i} + U_i,$$

gdzie U_I jest ściśle górnotrójkątną macierzą rozmiaru n_i, mającą zera na przekątnej. Jeśli pokażemy, że istnieją macierze odwracalne S_i, takie że $S_i^{-1}U_iS_i$ mają postać Jordana o klatkach odpowiadających wartości własnej 0, to $S_i^{-1}A_iS_i$ będą miały postać Jordana o klatkach odpowiadających wartości własnej A. Wtedy biorąc macierz blokowo diagonalną $S \in M_n(K)$ o blokach diagonalnych rozmiarów n_i postaci $S_1,\dots,S_n$ dostaniemy, że $S^{-1}AS$ ma postać Jordana.

Sprowadziliśmy zatem problem do wykazania, że macierz ściśle górnotrójkątna (czyli nilpotentna!) A rozmiaru n jest podobna do macierzy w postaci Jordana. Wynik ten uzyskamy przez odpowiednią serię podobieństw, konstruowanych wprost (przydatne do programowania).

Dowodzimy fakt przez indukcję ze względu na n. Dla $n=1$ macierz A jest zerowa. Dla $n=2$ mamy jeden potencjalnie niezerowy wyraz a macierzy A i piszemy:

$$A = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}, \quad T = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}.$$

Jeśli $a=0$, to nie ma czego dowodzić. Jeśli $a \neq 0$, to mamy:

$$T^{-1}AT = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

co kończy dowód. Przechodzimy do kroku indukcyjnego w naszym twierdzeniu. Załóżmy, że macierz A ma postać blokową:

$$A = \begin{bmatrix} 0 & a^T \\ 0 & A_1 \end{bmatrix},$$

gdzie $a \in K^{n-1}$ oraz $A_1 \in M_{n-1}(K)$ jest ściśle górnotrójkątna. Z założenia indukcyjnego istnieje macierz odwracalna $S_1 \in M_{n-1}$, że $S_1^{-1}A_1S_1$ ma żądaną postać Jordana, czyli

$$S^{-1}A_1S_1 = \begin{bmatrix} J_{k_1} & & \\ & \ddots & \\ & & J_{k_s} \end{bmatrix} = \begin{bmatrix} J_{k_1} & 0 \\ 0 & J \end{bmatrix},$$

gdzie $k_1 \geqslant k_2 \geqslant \dots \geqslant k_s$, przy czym $k_1 + \dots + k_s = n-1$. Ważne jest przy tym, że zaden z bloków diagonalnych macierzy Jordana J nie ma klatki większej niż k_1, czyli $J^{k_1} = 0$. Nietrudno sprawdzić, że

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a^TS_1 \\ 0 & S_1^{-1}A_1S_1 \end{bmatrix} \quad (*).$$

Rozbijmy iloczyn $a^T S_1$ do postaci $[a_1^T\, a_2^T]$, gdzie $a_1 \in K^{k_1}$ oraz $a_2 \in K^{n-k_1-1}$ i zapiszmy $(*)$ jako:

$$\begin{bmatrix} 1 & 0 \\ 0 & S_1^{-1} \end{bmatrix} A \begin{bmatrix} 1 & 0 \\ 0 & S_1 \end{bmatrix} = \begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}.$$

Rozważamy kolejne dość niebywałe macierze będące wzajemnymi odwrotnościami. Mamy mianowicie:

$$\begin{bmatrix} 1 & -a_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} 1 & a_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} = I_n$$

i możemy rozważyć podobieństwo

$$\begin{bmatrix} 1 & -a_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} 0 & a_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \cdot \begin{bmatrix} 1 & a_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} = \begin{bmatrix} 0 & a_1^T(I - J_{k_1}^T J_{k_1}) & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}.$$

Czym jest $a_1^T(I - J_{k_1}^T J_{k_1})$? Zauważmy, że dla każdego k mamy:

$$J_k^T J_k = \begin{bmatrix} 0 & 0 \\ 0 & I_{k-1} \end{bmatrix} \quad J_k J_k^T = \begin{bmatrix} I_{k-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Stąd biorąc $a_1^T = (x_1, \ldots, x_n) \in M_{1\times n}(K)$, mamy

$$a_1^T(I - J_{k_1}^T J_{k_1}) = (I - J_{k_1}^T J_{k_1})^T a_1 = (I - J_{k_1}^T J_{k_1}) a_1 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Wygodnie dla dalszej notacji jest wziąć bazę standardową $e_1, \ldots, e_n$ przestrzeni K^n i patrzeć na uzyskany wektor $a_1^T(I - J_{k_1}^T J_{k_1})$ jako na:

$$\begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \left(\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = (a_1^T e_1) e_1^T$$

A zatem nasza macierz B jest podobna do macierzy postaci:

$$\begin{bmatrix} 0 & (a_1^T e_1) e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix}. \qquad (**).$$

Rozważmy dwa przypadki. Jeśli $(a_1^T e_1) = x_1$ równy jest 0, to przez macierz permutacji macierz powyższa jest podobna do macierzy

$$\begin{bmatrix} J_{k_1} & 0 & 0 \\ 0 & 0 & a_2^T \\ 0 & 0 & J \end{bmatrix}. \qquad (\dagger)$$

Z założenia indukcyjnego istnieje wówczas macierz odwracalna $S_2 \in M_{n-k_1}(K)$, że:

$$S_2^{-1} \begin{bmatrix} 0 & a_2^T \\ 0 & J \end{bmatrix} S_2 = \widehat{J} \in M_{n-k_1}(K),$$

gdzie $\widehat{J}$ jest macierzą Jordana z wyrazami zerowymi na przekątnej. Stąd dana macierz $(\dagger)$, czyli też B, jest podobna do macierzy blokowej

$$\begin{bmatrix} J_{k_1} & 0 \\ 0 & \widehat{J} \end{bmatrix},$$

co kończy dowód.

Przypadek $x_1 = (a_1^T e_1) \neq 0$ jest nieco bardziej skomplikowany. Rozważamy w jego ramach kolejne podobieństwo poprzez macierz diagonalną postaci:

$$\begin{bmatrix} 1/a_1^T e_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (1/a_1^T e_1) I \end{bmatrix} \cdot \begin{bmatrix} 0 & (a_1^T e_1) e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} \cdot \begin{bmatrix} a_1^T e_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & a_1^T e_1 I \end{bmatrix} = \begin{bmatrix} 0 & e_1^T & a_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{bmatrix} = \begin{bmatrix} \widetilde{J} & e_1 a_2^T \\ 0 & J \end{bmatrix},$$

gdzie

$$\widetilde{J} = \begin{bmatrix} 0 & e_1^T \\ 0 & J_{k_1} \end{bmatrix}$$

jest klatką Jordana rozmiaru $k_1 + 1$. Przypomnijmy, że $e_1, \ldots, e_n$ jest bazą standardową przestrzeni K^n. Skoro $\widetilde{J} e_{i+1} = e_i$, dla $i = 1, 2, \ldots, k_1$, dostajemy następne podobieństwo

$$\begin{bmatrix} I & e_2 a_2^T \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} \widetilde{J} & e_1 a_2^T \\ 0 & J \end{bmatrix} \cdot \begin{bmatrix} I & -e_2 a_2^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} \widetilde{J} & -\widetilde{J} e_2 a_2^T + e_1 a_2^T + e_2 a_2^T J \\ 0 & J \end{bmatrix} = \begin{bmatrix} \widetilde{J} & e_2 a_2^T J \\ 0 & J \end{bmatrix}.$$

Dalsze podobieństwa uzyskujemy rekursywnie stosując ciąg analogicznych podobieństw postaci:

$$\begin{bmatrix} I & e_{i+1} a_2^T J^{i-1} \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} \widetilde{J} & e_i a_2^T J^{i-1} \\ 0 & J \end{bmatrix} \begin{bmatrix} I & e_{i+1} a_2^T J^{i-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} \widetilde{J} & e_{i+1} a_2^T J^i \\ 0 & J \end{bmatrix},$$

dla $i = 2, 3, \ldots$. Skoro $J^{k_1} = 0$, to po co najwyżej k_1 krotkach w tym ciągu podobieństw bloki pozadiagonalne znikną. Stąd macierz B jest podobna do macierzy Jordana postaci:

$$\begin{bmatrix} \widetilde{J} & 0 \\ 0 & J \end{bmatrix}.$$

W ten sposób uzyskaliśmy czysto macierzowy dowód twierdzenia Jordana.

* * *

Przedstawiony dowód miał zupełnie inny charakter niż wcześniejsze i czerpał mocno z patrzenia na klatki Jordana jako na macierze operatorów. Odnotujmy kilka dodatkowych faktów. Jeśli przez $J_k \in M_n(K)$ oznaczać będziemy klatkę Jordana o wartości własnej 0, będącą jak już wiemy macierzą nilpotentną stopnia k, to

$$J_k^T J_k = \begin{bmatrix} 0 & 0 \\ 0 & I_{k-1} \end{bmatrix}, \quad J_k J_k^T = \begin{bmatrix} I_{k-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Można to sprawdzić bezpośrednio lub zrozumieć, że J_k jest w istocie macierzą przesuwania współrzędnych do przodu, a jej transpozycja J_k^T – do tyłu, tzn.

$$J_k \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_k \\ 0 \end{bmatrix} \quad \text{oraz} \quad J_k^T \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix} = \begin{bmatrix} 0 \\ x_1 \\ \vdots \\ x_{k-2} \\ x_{k-1} \end{bmatrix}$$

Co ciekawe również operatorom danym macierzami odwracalnymi $I - J_k^T$ oraz $(I - J_k^T)^{-1}$ można przypisać czytelną interpretację: pierwsza przeprowadza wektor na wektor różnic współrzędnych, a druga – na wektor sum częściowych:

$$(I - J_k^T)x = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_3 \\ \vdots \\ x_k - x_{k-1} \end{bmatrix}$$

$$(I - J_k^T)^{-1}x = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \\ \vdots \\ x_1 + \ldots + x_n \end{bmatrix}$$

Rozdział 5

Podprzestrzenie pierwiastkowe. Dowód twierdzenia Jordana

Opis endomorfizmów nilpotentnych jest w istocie kluczowym punktem w dowodzie twierdzenia Jordana. Wyjaśnieniem tego faktu jest rozważanie tzw. podprzestrzeni pierwiastkowych.

Definicja 5.1: Podprzestrzeń pierwiastkowa

Niech $\phi \in \mathrm{End}(V)$. Dla każdego $a \in K$ przez $V_{[a]}$ oznaczamy podprzestrzeń złożoną ze wszystkich $\alpha \in V$, dla których istnieje n takie, że

$$(\phi - a\,\mathrm{id})^n(\alpha) = 0.$$

Przestrzeń $V_{[a]}$ nazywamy PODPRZESTRZENIĄ PIERWIASTKOWĄ endomorfizmu ϕ lub UOGÓLNIONĄ PODPRZESTRZENIĄ WŁASNĄ odpowiadającą wartości własnej a.

Przykład 1. Jeśli $\phi \in \mathrm{End}(\mathbb{C}^4)$ oraz $w_\phi(\lambda) = (\lambda-1)^2(\lambda-4)^2$, to w celu opisu przestrzeni $V_{[1]}$ rozważamy endomorfizm $\phi - \mathrm{id}$, który ograniczony do $V_{[1]}$ jest w kwadracie zerowy, czyli $(\phi - \mathrm{id})^2|_{V_{[1]}} = 0$, np. dla $M(\phi)^{st}_{st} = (x_1 + x_2, x_2, 4x_3 + x_4, 4x_4)$ mamy $V_{[1]} = \mathrm{lin}(\epsilon_1, \epsilon_2)$ oraz:

$$M(\phi - \mathrm{id})^{st}_{st} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad M((\phi - \mathrm{id})^2)^{st}_{st} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 9 & 6 \\ 0 & 0 & 0 & 9 \end{bmatrix}.$$

Przykład 2. Niech $\phi \in \mathrm{End}(K^{16})$ ma bazie $\mathcal{B} = (\beta_1, \ldots, \beta_{16})$ macierz:

$$M(\phi)^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 0 & 0 & & & & & & & & & & & & \\ 0 & 1 & 1 & 0 & & & & & & & & & & & & \\ 0 & 0 & 1 & 1 & & & & & & & & & & & & \\ 0 & 0 & 0 & 1 & & & & & & & & & & & & \\ & & & & 1 & 1 & 0 & & & & & & & & & \\ & & & & 0 & 1 & 1 & & & & & & & & & \\ & & & & 0 & 0 & 1 & & & & & & & & & \\ & & & & & & & 1 & & & & & & & & \\ & & & & & & & & 3 & 1 & 0 & & & & & \\ & & & & & & & & 0 & 3 & 1 & & & & & \\ & & & & & & & & 0 & 0 & 3 & & & & & \\ & & & & & & & & & & & 3 & 1 & & & \\ & & & & & & & & & & & 0 & 3 & & & \\ & & & & & & & & & & & & & 3 & 1 & \\ & & & & & & & & & & & & & 0 & 3 & \end{bmatrix}$$

Mamy $w_\phi(\lambda) = (\lambda - 1)^8 \cdot (\lambda - 3)^8$, ale $\dim V_{(1)} = 3$ oraz $\dim V_{(3)} = 3$. Nie jest to więc endomorfizm diagonalizowalny. Tymczasem możemy łatwo stwierdzić, że: $V_{[1]} = \mathrm{lin}(\beta_1, \ldots, \beta_8)$ oraz $V_{[3]} = \mathrm{lin}(\beta_9, \ldots, \beta_{16})$.

Jaki jest nasz cel? Okazuje się, co jest niezwykle istotnym wynikiem, że jeśli wielomian charakterystyczny endomorfizmu rozkłada się na czynniki liniowe, to przestrzeń V jest sumą prostą podprzestrzeni pierwiastkowych. Innymi słowy macierz $\phi|_{V_{[a_i]}}$ ma mieć w bazie $\mathcal{J}_i$ wszystkie klatki Jordana odpowiadające wartości własnej a_i w postaci Jordana macierzy całego endomorfizmu ϕ – i tylko te klatki. Oczywiście może być ich wiele. A więc można powiedzieć, że twierdzenie Jordana orzeka, że dla dowolnego endomorfizmu triangularyzowalnego przestrzeni V, możemy napisać, że V jest sumą prostą przestrzeni stanowiących podprzestrzenie pierwiastkowe odpowiadające wartościom własnym ϕ (które podzielić można w mniejsze sumy proste przestrzeni cyklicznych odpowiadających poszczególnym klatkom).

Twierdzenie 5.1: Twierdzenie o rozkładzie na podprzestrzenie pierwiastkowe

Niech $\phi \in \mathrm{End}(V)$ i niech $a_1, \ldots, a_k$ będą parami różnymi wartościami własnymi ϕ. Następujące warunki są równoważne.

(i) Endomorfizm ϕ jest triangularyzowalny.

(ii) $\dim V = \dim V_{[a_1]} + \dim V_{[a_2]} + \ldots + \dim V_{[a_k]}$.

(iii) $V = V_{[a_1]} \oplus V_{[a_2]} \oplus \ldots \oplus V_{[a_k]}$.

Jeśli zachodzi dowolny z warunków wyżej oraz

$$w_\phi(\lambda) = (a_1 - \lambda)^{t_1} \cdot \ldots \cdot (a_k - \lambda)^{t_k},$$

to

(1) $V_{[a_i]} = \ker(\phi - a_i \,\mathrm{id})^{t_i}$.

(2) $\mathrm{im}(\phi - a_i \,\mathrm{id})^{t_i} = \bigoplus_{j \neq i} V_{[a_j]}$

(3) $w_{\phi|_{V_{[a_i]}}}(\lambda) = (a_i - \lambda)^{t_i}$,

(4) $\dim V_{[a_i]} = t_i$.

Widzimy, że pierwsze trzy równoważne warunki w powyższym twierdzeniu nawiązują bezpośrednio do twierdzenia o endomorfizmach diagonalizowalnych. Mówiąc inaczej, dla endomorfizmu triangularyzowalnego podprzestrzenie pierwiastkowe odgrywają analogiczną rolę do podprzestrzeni własnych. Dowód ma kilka części. Zaczniemy od wskazania rozkładu niezmienniczego na podprzestrzenie, na których endomorfizm ten działa odpowiednio jako: nilpotentny i jako automorfizm.

Twierdzenie 5.2: Rozkład na sumę z częścią nilpotentną

Jeśli V jest przestrzenią skończenie wymiarową nad ciałem K, wówczas:

- $\phi|_{V_{[0]}}$ jest przekształceniem nilpotentnym,
- $\dim V_{[0]}$ jest krotnością algebraiczną pierwiastka 0 wielomianu $w_\phi(\lambda)$
- ma miejsce rozkład ϕ-niezmienniczy

$$V = V_{[0]} \oplus V', \qquad \text{gdzie} \qquad V' = \bigcap_{r=1}^{\infty} \mathrm{im}\, \phi^r,$$

jest podprzestrzenią ϕ-niezmienniczą oraz $\phi|_{V'} \in \mathrm{Aut}(V')$.

Dowód. Skoro zachodzą inkluzje $0 \subseteq \ker \phi \subseteq \ker \phi^2 \subseteq \ker \phi^3 \subseteq \ldots$ oraz $V \supseteq \mathrm{im}\, \phi \supseteq \mathrm{im}\, \phi^2 \supseteq \mathrm{im}\, \phi^3 \supseteq \ldots$, to jest jasne, że podprzestrzenie

$$\bigcup_{r=1}^{\infty} \ker \phi^r, \qquad \bigcap_{r=1}^{\infty} \mathrm{im}\, \phi^r$$

są ϕ-niezmiennicze. Ponadto dla każdego $r \geqslant 1$ mamy

$$\dim V = \dim \ker \phi^r + \dim \mathrm{im}\, \phi^r.$$

Skoro $\dim V < \infty$, to istnieje liczba r_0, dla której $\dim \operatorname{im} \phi^{r_0} = \dim \operatorname{im} \phi^{r_0+1}$, a skoro $\operatorname{im} \phi^{r_0} \supseteq \operatorname{im} \phi^{r_0+1}$, to $\operatorname{im} \phi^{r_0} = \operatorname{im} \phi^{r_0+1}$. Z równości $\dim V = \dim \ker \phi^{r_0} + \dim \operatorname{im} \phi^{r_0} = \ker \phi^{r_0+1} + \dim \operatorname{im} \phi^{r_0+1}$ wynika zatem również, że $\dim \ker \phi^{r_0} = \dim \ker \phi^{r_0+1}$, czyli $\ker \phi^{r_0} = \ker \phi^{r_0+1}$.

Wniosek z tej części dowodu jest następujący. Mamy:

$$\ker \phi^{r_0} = \ker \phi^{r_0+1} = \ker \phi^{r_0+2} = \ldots = \bigcup_{r=1} \ker \phi^r = V_{[0]},$$
$$\operatorname{im} \phi^{r_0} = \operatorname{im}^{r_0+1} = \operatorname{im}^{r_0+2} = \ldots = \bigcap_{r=1} \operatorname{im} \phi^r = V'.$$

Przed nami kluczowy argument całego dowodu, a także dowodu twierdzenia o rozkładzie na podprzestrzenie pierwiastkowe. Przekształcenie liniowe

$$\phi|_{\operatorname{im} \phi^{r_0}} : \operatorname{im} \phi^{r_0} \to \operatorname{im} \phi^{r_0+1}$$

jest izomorfizmem, a stąd

$$\ker \phi^{r_0} \cap \operatorname{im} \phi^{r_0} = \{0\}. \qquad (\dagger).$$

Istotnie, jeśli $v \in \ker \phi^{r_0} \cap \operatorname{im} \phi^{r_0}$ to $\phi^{r_0}(v) = 0$. Z drugiej strony $\phi|_{\operatorname{im} \phi^{r_0}}$ jest izomorfizmem więc jeśli $v \in \operatorname{im} \phi^{r_0}$ mamy $\phi^{r_0}(v) = 0$, to $v = 0$. Stąd wynika $(\dagger)$.

Wobec $\dim V = \dim \ker \phi^{r_0} + \dim \operatorname{im} \phi^{r_0} = \dim V_{[0]} + \dim V'$, uzyskujemy

$$V = \ker \phi^{r_0} \oplus \operatorname{im} \phi^{r_0} = \bigcup_{r=1} \ker \phi^r \oplus \bigcap_{r=1} \operatorname{im} \phi^r = V_{[0]} \oplus V'.$$

Oczywiście mamy też $\phi^{r_0}|_{V_{[0]}} = 0$, skąd $\phi|_{V_{[0]}}$ jest przekształceniem nilpotentnym i jego wielomian charakterystyczny to $\pm\lambda^{\dim V_{[0]}}$ (stopień $w_{\phi|_{V_{[0]}}}(\lambda)$ równy jest wymiarowi przestrzeni, na której działa $\phi|_{V_{[0]}}$). Skoro $\phi|_{V'}$ jest izomorfizmem, to nie ma on wartości własnej 0. Skoro zaś

$$w_\phi(\lambda) = w_{\phi|_{V_{[0]}}}(\lambda) \cdot w_{\phi|_{V'}}(\lambda)$$

więc algebraiczna krotność 0 jako krotność pierwiastka $w_\phi(\lambda)$ jest równa $\dim V_{[0]}$. □

Powyższe twierdzenie daje nam rozkład ϕ-niezmienniczy $V = V_{[0]} \oplus V'$ i natychmiast wnioskujemy z niego, że każdą podprzestrzeń pierwiastkową można wydzielić jako składnik prosty rozkładu ϕ-niezmienniczego.

Wniosek 5.1: Rozkład związany z wartością własną

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K i niech $\phi \in \operatorname{End}(V)$. Niech $a \in \operatorname{spec}(\phi)$ będzie pierwiastkiem wielomianu $w_\phi(\lambda)$ o krotności algebraicznej δ, tzn.

$$w_\phi(\lambda) = (a - \lambda)^\delta \cdot w(\lambda), \quad \text{gdzie } w \in K[x] \text{ oraz } w(a) \neq 0.$$

Wówczas:

- istnieje rozkład ϕ-niezmienniczy $V = V_{[a]} \oplus W'$, taki że
$$w_{\phi|_{V_{[a]}}} = (a - \lambda)^\delta \quad \text{oraz} \quad w_{\phi|_{W'}}(\lambda) = w(\lambda),$$
- $\dim V_{[a]} = \delta$,
- istnieje taka baza przestrzeni $V_{[a]}$, że macierz $\phi|_{V_{[a]}}$ ma postać Jordana z klatkami Jordana odpowiadającymi wyłącznie wartości własnej a.

Dowód. Korzystamy z poprzedniego twierdzenia, stosując je dla endomorfizmu $\phi - a\,\mathrm{id}$. Wówczas podprzestrzeń pierwiastkowa $V_{[a]}$ względem endomorfizmu ϕ jest podprzestrzenią $V_{[0]}$ względem endomorfizmu $\phi - a\,\mathrm{id}$. Zatem mamy rozkład

$$V = V_{[a]} \oplus W',$$

który jest $(\phi - a\,\mathrm{id})$-niezmienniczy, a więc także ϕ-niezmienniczy (dla podprzestrzeni $U \subseteq V$ oraz dowolnego $u \in U$ mamy $(\phi - a\,\mathrm{id})(u) \in U \iff \phi(u) \in U$). Ponadto, skoro $\phi - a\,\mathrm{id}$ ograniczony do W' jest automorfizmem, więc a nie jest pierwiastkiem $w_{\phi|_{W'}}(\lambda)$.

Krotność algebraiczna a endomorfizmu ϕ jest równa krotności algebraicznej 0 endomorfizmu $\phi - a\,\mathrm{id}$, gdyż $w_\phi(\lambda) = w_{\phi - a\,\mathrm{id}}(\lambda - a)$, więc $\dim V_{[a]} = \delta$.

Trzecia część tezy wynika z tego, że $V_{[a]}$ jest podprzestrzenią, na której endomorfizm $\phi - a\,\mathrm{id}$ jest nilpotentny. Zatem istnieje baza $\mathcal{B}_a$, w której macierz $M((\phi - a\,\mathrm{id})|_{V_{[a]}})^{\mathcal{B}_a}_{\mathcal{B}_a}$ ma postać Jordana o klatkach odpowiadających wartości własnej 0. Zatem macierz $M(\phi|_{V_{[a]}})^{\mathcal{B}_a}_{\mathcal{B}_a}$ ma postać Jordana o klatkach odpowiadających wartości własnej a, gdyż:

$$M((\phi - a\,\mathrm{id})|_{V_{[a]}})^{\mathcal{B}_a}_{\mathcal{B}_a} + aI = M(\phi|_{V_{[a]}})^{\mathcal{B}_a}_{\mathcal{B}_a}.$$

□

Przykład ilustrujący ostatni argument. Weźmy $A \in M_{16}(\mathbb{R})$ postaci:

$$\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2
\end{bmatrix}$$

Jeśli przyjmiemy, że jest to macierz endomorfizmu $\phi \in \mathrm{End}(\mathbb{R}^{16})$ w bazie Jordana $(\beta_1, \ldots, \beta_{16})$, to dla każdej wartości własnej rozważyć można diagramy Younga opisujące działania odpowiednich endomorfizmów na podprzestrzeniach pierwiastkowych :

- endomorfizm $\phi - 1 \cdot \mathrm{id}$ ograniczony do przestrzeni $V_{[1]}$ rozpiętej na $\{\beta_1, \ldots, \beta_8\}$ jest nilpotentny i przekształca te wektory zgodnie ze schematem

$$\beta_4 \mapsto \beta_3 \mapsto \beta_2 \mapsto \beta_1 \mapsto 0, \quad \beta_7 \mapsto \beta_6 \mapsto \beta_5 \mapsto 0, \quad \beta_8 \mapsto 0,$$

- endomorfizm $\phi - 2 \cdot \mathrm{id}$ ograniczony do przestrzeni $V_{[2]}$ rozpiętej na $\{\beta_{15}, \beta_{16}\}$ jest nilpotentny przeprowadza te wektory zgodnie ze schematem

$$\beta_{16} \mapsto \beta_{15} \mapsto 0,$$

- endomorfizm $\phi - 3 \cdot \mathrm{id}$ ograniczony do przestrzeni $V_{[3]}$ rozpiętej na $\{\beta_9, \ldots, \beta_{14}\}$ jest nilpotentny i przeprowadza te wektory zgodnie ze schematem

$$\beta_{11} \mapsto \beta_{10} \mapsto \beta_9 \mapsto 0, \quad \beta_{14} \mapsto \beta_{13} \mapsto \beta_{12} \mapsto 0.$$

Innymi słowy, naśladując terminologię dowodu mamy:

$$\mathcal{B}_1 = \{\beta_1, \ldots, \beta_8\}, \quad \mathcal{B}_2 = \{\beta_{15}, \beta_{16}\}, \quad \mathcal{B}_3 = \{\beta_9, \ldots, \beta_{14}\}.$$

W dowodzie jest odwrotnie: dla każdego z endomorfizmów $(\phi - 1 \cdot \mathrm{id})|_{V_{[1]}}$, $(\phi - 2 \cdot \mathrm{id})|_{V_{[2]}}$, $(\phi - 3 \cdot \mathrm{id})|_{V_{[3]}}$, konstruujemy – na mocy twierdzenia Jordana w przypadku nilpotentnym – bazy $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ i z nich składamy bazę Jordana endomorfizmu ϕ.

Dowód twierdzenia o rozkładzie na podprzestrzenie pierwiastkowe.

Implikacja (i) $\Rightarrow$ (ii) wynika stąd, że wymiar każdej z podprzestrzeni pierwiastkowych równy jest krotności algebraicznej odpowiedniej wartości własnej (Wniosek 5.1).

Implikację (ii) $\Rightarrow$ (iii) dowodzi się przez indukcję ze względu na liczbę k podprzestrzeni pierwiastkowych. Dla jednej wartości własnej teza jest jasna, gdyż jeśli $\dim V = \dim V_{[a_1]}$, to $V = V_{[a_i]}$. Jeśli natomiast wartości własnych jest więcej, to z Wniosku 5.1 mamy $V = V_{[a_1]} \oplus W'$, gdzie:

- W' jest ϕ-niezmiennicza,
- $\phi|_{W'}$ nie ma wartości własnej a_1,
- $w_{\phi|_{W'}}(\lambda) = \frac{w_\phi(\lambda)}{(\lambda - a_1)^{\dim V_{[a_1]}}}$.

Stąd podprzestrzenie pierwiastkowe W' względem $\phi|_{W'}$ o wartościach własnych $a_2, \ldots, a_k$ to dokładnie podprzestrzenie $V_{[a_2]}, \ldots, V_{[a_k]}$. Jest ich $k-1$. Z założenia indukcyjnego mamy więc $W' = V_{[a_2]} \oplus \ldots \oplus V_{[a_k]}$, co daje (iii).

Implikacja (iii) $\Rightarrow$ (i) to oczywisty wniosek z postaci blokowo-diagonalnej macierzy w bazie złożonej z baz poszczególnych składników rozkładu ϕ-niezmienniczego (czyli Wniosek 3.1). Punkty (1)-(4) to również proste konsekwencje Twierdzenia 5.2.

Wniosek 5.2: Twierdzenie Jordana wynika z powyższych

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K i niech $\phi \in \mathrm{End}(V)$ będzie triangularyzowalny. Niech $\mathrm{spec}(\phi) = \{a_1, \ldots, a_k\}$ oraz

$$w_\phi(\lambda) = \pm(a_1 - \lambda)^{t_1}(a_2 - \lambda)^{t_2} \cdot \ldots \cdot (a_k - \lambda)^{t_k}.$$

Niech $V^{(a_i)}$ będzie podprzestrzenią pierwiastkową odpowiadającą wartości własnej 0 endomorfizmu $\phi - a_i \,\mathrm{id}$ (czyli podprzestrzenią pierwiastkową odpowiadającą wartości własnej a_i endomorfizmu ϕ). Wówczas:

$$V = V^{(a_1)} \oplus V^{(a_2)} \oplus \ldots \oplus V^{(a_k)},$$

czyli istnieje baza Jordana $\mathcal{B} = \mathcal{B}_{a_1} \cup \mathcal{B}_{a_2} \cup \ldots \cup \mathcal{B}_{a_k}$ przestrzeni V złożona z t_i-elementowych baz Jordana $\mathcal{B}_{a_i}$ endomorfizmów nilpotentnych $\phi - a_i \,\mathrm{id}$ ograniczonych do podprzestrzeni $V^{(a_i)}$.

Pozostałą część wykładu poświęcimy przykładom znajdowania baz Jordana endomorfizmów triangularyzowalnych przestrzeni skończonego wymiaru. Ogólny algorytm jej wyznaczania jest następujący.

1. Znajdź wielomian charakterystyczny $w_\phi(\lambda)$ endomorfizmu ϕ.
2. Weź $a \in \mathrm{spec}(\phi)$ o krotności algebraicznej t; następne kroki powtarzaj dla kolejnych pierwiastków $w_\phi(\lambda)$.
3. Policz $\ker \phi - a \,\mathrm{id} \subseteq \ker(\phi - a \,\mathrm{id})^2 \subseteq \ker(\phi - a \,\mathrm{id})^3 \subseteq \ldots$ (rozwiązując układy równań o macierzach postaci $(A - aI)^k$, gdzie A jest macierzą ϕ). Gdy tylko uzyskasz $\dim \ker(\phi - a \,\mathrm{id})^r = t$, połóż $V_{[a]} = (\phi - a \,\mathrm{id})^r$. Szukaj bazy Jordana endomorfizmu ϕ ograniczonego do $V_{[a]}$.
4. Znajdź t-elementową bazę Jordana $\phi - a \,\mathrm{id}$ ograniczonego do $V_{[a]}$. Konstruuj kolejne klatki Jordana poczynając od największych wymiaru r. Każda klatka to jeden wiersz w odpowiadającym podziałowi t diagramowi Younga. Dokładniej:
   - Weź element $v \in \ker(\phi - a \,\mathrm{id})^r \setminus \ker(\phi - a \,\mathrm{id})^{r-1}$ i wektory
   $$v, \quad (\phi - a \,\mathrm{id})(v), \quad (\phi - a \,\mathrm{id})^2(v), \quad \ldots, \quad (\phi - a \,\mathrm{id})^{r-1}(v)$$
   stanowią bazę Jordana klatki Jordana wymiaru r .
   - Dopóki suma wymiarów klatek, które już skonstruowałeś jest mniejsza od t, to znajdź największe takie $r' \leqslant r$ i $v' \in \ker(\phi - a \,\mathrm{id})^{r'}$, które nie jest zawarte w sumie przestrzeni rozpiętych na już skonstruowanym fragmencie bazy, dodaj nową klatkę, która będzie miała bazę
   $$v', \quad (\phi - a \,\mathrm{id})(v'), \quad (\phi - a \,\mathrm{id})^2(v'), \quad \ldots, \quad (\phi - a \,\mathrm{id})^{r'-1}(v').$$

Zilustrujmy wyznaczanie bazy Jordana kilkoma bardzo prostymi przykładami.

Przykład. Niech $\phi \in \mathrm{End}(\mathbb{R}^4)$ zadany będzie w bazie standardowej macierzą

$$A = M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Oczywiście wielomianem charakterystycznym tego endomorfizmu jest $(1-\lambda)^4$. Skoro rozkłada się on nad $\mathbb{R}$ na iloczyn czynników liniowych, to zgodnie z twierdzeniem Jordana istnieje baza, w której ϕ ma macierz w postaci Jordana. Szukamy najpierw postaci Jordana: badamy potęgi macierzy $A - I = M(\phi - \mathrm{id})^{st}_{st}$. Mamy:

$$A - I = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A zatem $r(A) = 3$ i stąd $q_1 = r(A-I)^0 - r(A-I) = 4 - 3 = 1$. A zatem ϕ ma w bazie Jordana $\mathcal{J}$ macierz o jednej klatce rozmiaru $\geqslant 1$ odpowiadającą wartości własnej 1. Skoro suma rozmiarów klatek odpowiadających tej wartości własnej równa jest krotności algebraicznej wartości własnej 1 (a ich liczba równa jest wymiarowi przestrzeni własnej odpowiadającej wartości własnej 1, czyli jej krotności geometrycznej), czyli 4, to owa jedna klatka ma rozmiar 4×4. W rezultacie szukamy bazy $\mathcal{J} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, w której nasz endomorfizm ma macierz o jednej klatce 4×4:

$$M(\phi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

:
Idąc drogą sugerowaną przez algorytm, przy endomorfizmie $\phi - \mathrm{id}$ poszczególne wektory bazowe przechodzą na siebie w następujący sposób:

$$\alpha_4 \mapsto \alpha_3 \mapsto \alpha_2 \mapsto \alpha_1 \mapsto 0.$$

Zatem α_4 należy do $\ker(\phi-\mathrm{id})^4 \setminus \ker(\phi-\mathrm{id})^3$. Macierzą $(\phi-\mathrm{id})^4$ jest macierz A^4, szukamy więc wektorów takich, że $A^4 v = 0$. Oczywiście każdy wektor z $\mathbb{R}^4$ ma tą własność, czyli

$$\ker(\phi - \mathrm{id})^4 = \mathbb{R}^4.$$

Wyznaczmy $\ker(\phi - \mathrm{id})^3$. Podprzestrzeń ta składa się z wektorów w takich, że $A^3 w = 0$. A zatem jeśli $w_1, \ldots, w_4$ to współrzędne w w bazie standardowej, mamy:

$$A^3 w = \begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

W rezultacie

$$\ker(\phi - \mathrm{id})^3 = \mathrm{lin}((1,0,0,0),(0,1,0,0),(0,0,1,0)).$$

Nietrudno więc wybrać wektor α_4 z $\ker(\phi-\mathrm{id})^4 \setminus \ker(\phi-\mathrm{id})^3$. Jest to dowolny wektor postaci $(0,0,0,a)$, gdzie $0 \neq a \in K$. Każdy wybór a doprowadzi do uzyskania innej bazy Jordana V endomorfizmu ϕ, w której przekształcenie to będzie miało jednakże taką samą macierz. Na przykład bierzemy $\alpha_4 = (0,0,0,1)$. Wówczas wyznaczamy wektory $\alpha_3, \alpha_2, \alpha_1$ stosując kolejne iteracje $(\phi - \mathrm{id})$, czyli:

$$\alpha_3 = (\phi - \mathrm{id})(\alpha_4) = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

$$\alpha_2 = (\phi - \mathrm{id})(\alpha_3) = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 0 \end{bmatrix}.$$

$$\alpha_1 = (\phi - \mathrm{id})(\alpha_2) = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

A zatem przykładowa baza Jordana $\mathcal{J}$ endomorfizmu ϕ to:

$$((6,0,0,0),(5,3,0,0),(1,2,1,0),(0,0,0,1)).$$

Zauważmy, że dla $C = M(\mathrm{id})^{st}_{\mathcal{J}}$ mamy $M(\phi)^{\mathcal{J}}_{\mathcal{J}} = C^{-1}M(\phi)^{st}_{st}C$, czyli

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 5 & 1 & 0 \\ 0 & 3 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 6 & 5 & 1 & 0 \\ 0 & 3 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

* * *

Bazę $\mathcal{J}$ można wyznaczyć nieco łatwiej w przypadku, gdy mamy tylko jedną klatkę Jordana odpowiadającą danej wartości własnej (podprzestrzeń pierwiastkowa jest wtedy cykliczna).

Mając macierz przekształcenia ϕ w bazie Jordana $\mathcal{J} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ odczytujemy z niej, że

$$\begin{aligned} \phi(\alpha_1) &= \alpha_1, \\ \phi(\alpha_2) &= \alpha_1 + \alpha_2, \\ \phi(\alpha_3) &= \alpha_2 + \alpha_3, \\ \phi(\alpha_4) &= \alpha_3 + \alpha_4. \end{aligned}$$

Wyznaczymy najpierw wektor własny α_1, a później napiszemy na podstawie warunków wyżej układ równań, który pozwoli wyznaczyć wektor α_2 i kolejne. Wiemy, że każdy wektor własny x o współrzędnych w bazie standardowej (x_1, x_2, x_3, x_4) spełnia warunek $(\phi - \mathrm{id})(x) = 0$, czyli

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

czyli $(x_1, x_2, x_3, x_4) \in \mathrm{lin}((1,0,0,0))$. Możemy zatem wybrać $\alpha_1 = (6,0,0,0)$ i przejść do wyznaczenia α_2 o współrzędnych (y_1, y_2, y_3, y_4) z warunku $\phi(\alpha_2) = \alpha_1 + \alpha_2$, czyli $(\phi - \mathrm{id})(\alpha_2) = \alpha_1$. Dostajemy układ:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Przykładowym rozwiązaniem jest wektor $\alpha_2 = (5,3,0,0)$. Analogicznie wyznaczamy α_3 i α_4 rozwiązując kolejno układy:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 0 \end{bmatrix},$$

co prowadzi do uzyskania np. wektora $\alpha_3 = (1,2,1,0)$ i wreszcie, rozwiązujemy czwarty układ:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix},$$

dostajemya przykładowe rozwiązanie $\alpha_4 = (0,0,0,1)$. Metoda ta ma jednak wady gdy danej wartości własnej odpowiada więcej klatek.

Przykład. Rozważmy macierz

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \in \mathrm{M}_4(\mathbb{R}).$$

Znaleźć macierz J w postaci Jordana podobną do macierzy X oraz taką macierz odwracalną $C \in M_4(\mathbb{R})$, że $J = C^{-1}XC$.

Wielomian charakterystyczny macierzy X ma postać:

$$w_X(\lambda) = \begin{vmatrix} 1-\lambda & 0 & 0 & 0 \\ 0 & 1-\lambda & -1 & 1 \\ -1 & 0 & 1-\lambda & 0 \\ 0 & 0 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)^3(2-\lambda).$$

Endomorfizm $\phi \in \mathrm{End}(\mathbb{R}^4)$ dany w bazie standardowej macierzą X ma zatem wartości własne $1, 2$. Mamy

$$r(X-I) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = 3.$$

Macierz X można zatem sprowadzić do macierzy w postaci Jordana o jednej klatce rozmiaru 3×3 odpowiadającej wartości własnej 1 oraz o jednej klatce rozmiaru 1×1 odpowiadającej wartości własnej 2.

Endomorfizm ϕ ma wartości własne $1, 2$ i zgodnie z rachunkami wyżej wiemy, że ϕ ma w pewnej bazie $\mathcal{J} = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ macierz:

$$M(\phi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Widzimy, że $(0, 1, 0, 0)$ jest wektorem własnym ϕ o wartości własnej 1 i możemy przyjąć $\beta_1 = (0, 1, 0, 0)$. Aby wyznaczyć β_2 rozwiązujemy układ:

$$(X-I) \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

Mamy zatem $\beta_2 = (0, 0, -1, 0)$. Dalej rozwiązujemy równanie $(X-I)\beta_3 = \beta_2$, uzyskując na przykład $\beta_3 = (1, 0, 0, 0)$. Następnie znajdujemy wektor własny o wartości własnej 2, na przykład $\beta_4 = (0, 1, 0, 1)$.
Przykład. Niech $\phi \in \mathrm{End}(\mathbb{R}^4)$ zadany będzie w bazie standardowej macierzą

$$A = M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Oczywiście wielomianem charakterystycznym tego endomorfizmu jest $(1-\lambda)^4$. Skoro rozkłada się on nad $\mathbb{R}$ na iloczyn czynników liniowych, to zgodnie z twierdzeniem Jordana istnieje baza, w której ϕ ma macierz w postaci Jordana. Wyznaczmy ją przez badanie potęg macierzy $A - I = M(\phi - \mathrm{id})^{st}_{st}$. Mamy:

$$A - I = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A zatem $r(A) = 3$ i stąd $q_1 = r(A-I)^0 - r(A-I) = 4-3 = 1$. A zatem ϕ ma w bazie Jordana $\mathcal{J}$ macierz o jednej klatce rozmiaru $\geqslant 1$ odpowiadającą wartości własnej 1. Skoro suma rozmiarów klatek odpowiadających tej wartości własnej równa jest krotności algebraicznej wartości własnej 1 (a ich liczba równa

jest wymiarowi przestrzeni własnej odpowiadającej wartości własnej 1, czyli jej krotności geometrycznej), czyli 4, to owa jedna klatka ma rozmiar 4×4. W rezultacie:

$$M(\phi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Mamy zatem macierz przekształcenia ϕ w bazie Jordana $\mathcal{J} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ i odczytujemy z niej, że

$$\begin{aligned} \phi(\alpha_1) &= \alpha_1, \\ \phi(\alpha_2) &= \alpha_1 + \alpha_2, \\ \phi(\alpha_3) &= \alpha_2 + \alpha_3, \\ \phi(\alpha_4) &= \alpha_3 + \alpha_4. \end{aligned}$$

Wyznaczymy najpierw wektor własny α_1, a później napiszemy na podstawie warunków wyżej układ równań, który pozwoli wyznaczyć wektor α_2 i kolejne. Wiemy, że każdy wektor własny x o współrzędnych w bazie standardowej (x_1, x_2, x_3, x_4) spełnia warunek $(\phi - \mathrm{id})(x) = 0$, czyli

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

czyli $(x_1, x_2, x_3, x_4) \in \mathrm{lin}((1, 0, 0, 0))$. Możemy zatem wybrać $\alpha_1 = (6, 0, 0, 0)$ i przejść do wyznaczenia α_2 o współrzędnych (y_1, y_2, y_3, y_4) z warunku $\phi(\alpha_2) = \alpha_1 + \alpha_2$, czyli $(\phi - \mathrm{id})(\alpha_2) = \alpha_1$. Dostajemy układ:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Przykładowym rozwiązaniem jest wektor $\alpha_2 = (5, 3, 0, 0)$. Analogicznie wyznaczamy α_3 i α_4 rozwiązując kolejno układy:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 0 \\ 0 \end{bmatrix},$$

co prowadzi do uzyskania np. wektora $\alpha_3 = (1, 2, 1, 0)$ i wreszcie, rozwiązujemy czwarty układ:

$$\begin{bmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix},$$

dostajemya przykładowe rozwiązanie $\alpha_4 = (0, 0, 0, 1)$. W prostych sytuacjach przedstawiony algorytm pozwala wyznaczać bazę Jordana (tzn. w sytuacjach, gdy wprawdzie wartości własnych może być wiele, ale każdej odpowiada jedna klatka). Metoda ta ma jednak wady gdy danej wartości własnej odpowiada więcej klatek. Zacznijmy jednak od omówienia sytuacji, gdy jest więcej wartości własnych.

Przykład. Rozważmy macierz

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \in \mathrm{M}_4(\mathbb{R}).$$

Znaleźć macierz J w postaci Jordana podobną do macierzy X oraz taką macierz odwracalną $C \in M_4(\mathbb{R})$, że $J = C^{-1}XC$.

Wielomian charakterystyczny macierzy X ma postać:

$$w_X(\lambda) = \begin{vmatrix} 1-\lambda & 0 & 0 & 0 \\ 0 & 1-\lambda & -1 & 1 \\ -1 & 0 & 1-\lambda & 0 \\ 0 & 0 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)^3(2-\lambda).$$

Endomorfizm $\phi \in \mathrm{End}(\mathbb{R}^4)$ dany w bazie standardowej macierzą X ma zatem wartości własne $1, 2$. Mamy

$$r(X-I) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = 3.$$

Macierz X można zatem sprowadzić do macierzy w postaci Jordana o jednej klatce rozmiaru 3×3 odpowiadającej wartości własnej 1 oraz o jednej klatce rozmiaru 1×1 odpowiadającej wartości własnej 2.

Endomorfizm ϕ ma wartości własne $1, 2$ i zgodnie z rachunkami wyżej wiemy, że ϕ ma w pewnej bazie $\mathcal{J} = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ macierz:

$$M(\phi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Widzimy, że $(0, 1, 0, 0)$ jest wektorem własnym ϕ o wartości własnej 1 i możemy przyjąć $\beta_1 = (0, 1, 0, 0)$. Aby wyznaczyć β_2 rozwiązujemy układ:

$$(X-I) \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

Mamy zatem $\beta_2 = (0, 0, -1, 0)$. Dalej rozwiązujemy równanie $(X-I)\beta_3 = \beta_2$, uzyskując na przykład $\beta_3 = (1, 0, 0, 0)$. Następnie znajdujemy wektor własny o wartości własnej 2, na przykład $\beta_4 = (0, 1, 0, 1)$.

* * *

Przejdźmy do przykładu, gdy zaprezentowana uproszczona metoda może nie zadziałać i potrzebne jest zastosowanie ogólnego algorytmu.

Przykład. Weźmy endomorfizm $\psi : \mathbb{R}^3 \to \mathbb{R}^3$ dany w bazie standardowej macierzą

$$A = M(\psi)^{st}_{st} = \begin{bmatrix} 4 & -1 & 0 \\ 4 & 0 & 0 \\ 8 & -4 & 2 \end{bmatrix}.$$

Mam nadzieję, że uwierzycie mi Państwo, że $w_\psi(\lambda) = (2-\lambda)^3$ oraz, że $r(A-2I) = 1$, a zatem macierz ψ w bazie Jordana ma 2 bloki rozmiaru $\geqslant 1$, czyli w pewnej bazie $\mathcal{J} = (\alpha_1, \alpha_2, \alpha_3)$ mamy:

$$M(\psi)^{\mathcal{J}}_{\mathcal{J}} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Jak teraz wyznaczycie Państwo α_1? Sprawdzamy, że

$$V_{(2)} = \mathrm{lin}((1, 2, 0), (0, 0, 1)).$$

Czy można przyjąć za α_1 dowolny wektor własny, choćby $(1, 2, 0)$? Wówczas koniecznie $\psi(\alpha_2) = \alpha_1 + 2\alpha_2$, czyli

$$(\psi - 2\,\mathrm{id})(\alpha_2) = \alpha_1.$$

Jeśli $\alpha_2 = (x_1, x_2, x_3)$, to mamy:

$$\begin{bmatrix} 2 & -1 & 0 \\ 4 & -2 & 0 \\ 8 & -4 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Nietrudno sprawdzić, że układ ten nie ma rozwiązań! A zatem wybór wektora własnego $(1,2,0)$ był nieprawidłowy. Wektor α_1 powinien być innym elementem $V_{(2)}$. Jakim? Wyznaczmy najpierw α_2, bowiem $(\phi - 2\,\mathrm{id})(\alpha_2) = \alpha_1$. Skoro tak, to

$$(\phi - 2\,\mathrm{id})^2(\alpha_2) = (\phi - 2\,\mathrm{id})(\alpha_1) = 0.$$

A zatem $\alpha_2 \in \ker(\psi - 2\,\mathrm{id})^2 \setminus \ker(\psi - 2\,\mathrm{id})$. Jednak $\ker(\psi - 2\,\mathrm{id})^2 = \mathbb{R}^3$, bo

$$M((\psi - 2\,\mathrm{id})^2)^{st}_{st} = \begin{bmatrix} 2 & -1 & 0 \\ 4 & -2 & 0 \\ 8 & -4 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 & 0 \\ 4 & -2 & 0 \\ 8 & -4 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A zatem α_2 może być wybrany jako dowolny niezerowy wektor nie-własny. Na przykład $\alpha_2 = (0,1,0)$. Wówczas mamy: $\alpha_1 = (\psi - 2\,\mathrm{id})(\alpha_2)$, czyli $\alpha_1 = (-1,-2,-4)$. Teraz możemy DOBRAĆ do wektora α_1 wektor własny α_3. Wybór α_3 zależy więc od wyboru wektora nie-własnego α_2, a nie tylko od liniowej niezależności z α_1. Czy widzicie Państwo jak wyglądałaby sytuacja, gdyby było to przekształcenie $\mathbb{R}^4 \to \mathbb{R}^4$ i jego postać Jordana miałaby dwie klatki 2×2 odpowiadające tej samej wartości własnej?

Rozumiemy zatem opis endomorfizmu, który można zapisać w bazie Jordana w postaci macierzy o dokładnie jednej klatce Jordana odpowiadającej wartości własnej a. Wiemy jak w takiej sytuacji „dobierać wektory". Spróbujmy teraz przyjrzeć się kolejnej sytuacji, gdy endomorfizm $\phi \in \mathrm{End}(K^{11})$ ma dokładnie jedną wartość własną a, co więcej $w_\phi(\lambda) = -(a-\lambda)^{11}$ oraz w bazie Jordana $\mathcal{B} = \{\beta_1, \ldots, \beta_{11}\}$ macierz ϕ ma postać Jordana o trzech klatkach rozmiarów $5, 4, 2$ odpowiadających wartości własnej a.

$$M(\phi)^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} a & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a \end{bmatrix}.$$

Widzimy zatem, że tym razem wektory bazy Jordana endomorfizmu ϕ układają się w trzy łańcuchy, odpowiadające wynikom działań endomorfizmu $\phi - a \cdot \mathrm{id}$:

$$\beta_5 \mapsto \beta_4 \mapsto \beta_3 \mapsto \beta_2 \mapsto \beta_1 \mapsto 0, \quad \beta_9 \mapsto \beta_8 \mapsto \beta_7 \mapsto \beta_6 \mapsto 0, \quad \beta_{11} \mapsto \beta_{10} \mapsto 0.$$

Gdyby naszym zadaniem było znalezienie bazy Jordana, moglibyśmy to zrobić poprzez staranne wybieranie wektorów z baz jąder odpowiednio wysokich iteracji endomorfizmu $\phi - a\,\mathrm{id}$. W rozważanym przypadku wybralibyśmy najpierw dowolny wektor β_5 z różnicy zbiorów $\ker(\phi - a\,\mathrm{id})^5 \setminus \ker(\phi - a\,\mathrm{id})^4$. Następnie wyznaczylibyśmy wektory $\beta_4, \beta_3, \beta_2, \beta_1$ jak wyżej, rozwiązując warunek $(\phi - a\,\mathrm{id})(\beta_i) = \beta_{i-1}$.

Skąd wzięlibyśmy β_9? Otóż jest to wektor z $(\phi - a\,\mathrm{id})^4 \setminus \ker(\phi - a\,\mathrm{id})^3$ i to liniowo niezależny z już znalezionym wektorem β_3, który jest w tym samym zbiorze. Gdy taki wektor wybierzemy, wyznaczamy β_8, β_7 i β_6.

Wreszcie, wektor β_{11} wybieramy z $\ker(\phi - a\,\mathrm{id})^2 \setminus \ker(\phi - a\,\mathrm{id})$ tak, by układ $\{\beta_2, \beta_7, \beta_{10}\}$ był liniowo niezależny (β_2 dobieramy więc do β_7 i β_{10}), bo te wszystkie wektory są w tej różnicy. W ten sposób wyznaczymy β_{11}. Wydać, że rachunki są w takim przypadku niezwykle żmudne.

5.1 Uzupełnienie. Trudniejsze zadania o postaci i bazie Jordana

Przedstawione na wykładzie przykłady wyznaczania baz Jordana obejmują bardzo podstawowe sytuacje, ilustrując w zasadzie sposób odczytywania macierzy zapisanej w bazie Jordana. Szereg krótszych i dłuższych zadań stawia nam wyzwanie zebrania całej wiedzy dotyczącej endomorfizmów i zastosowania jej w sposób nie zawsze oczywisty. W tym uzupełnieniu prezentujemy przykłady takich zadań, pochodzące zarówno z kolokwiów na potoku zwykłym, jak i gwiazdkowym.

Zadanie 5.1. *Pewna macierz $A \in M_{7\times 7}(\mathbb{C})$ spełnia równanie $(A-I)^3 = 0$. Wiemy też, że $r(A-I)^2 = 2$. Wykaż, że jedyną wartością własną macierzy A jest 1. Opisz postać Jordana macierzy A.*

Rozwiązanie. Niech $Av = \lambda v$, dla pewnego niezerowego wektora v oraz $\lambda \in \mathbb{C}$. Mamy zatem $(A-I)v = (\lambda-1)v$, czyli $(A-I)^3v = (\lambda-1)^3v$. Ale $(A-I)^3v = 0v = 0$, a zatem skoro $v \neq 0$ mamy $(\lambda-1)^3 = 0$, czyli $\lambda = 1$. A zatem A można sprowadzić do postaci Jordana, w której wszystkie klatki odpowiadają wartości własnej 1. Wyznaczamy teraz stałe $q_k = r(A-I)^{k-1} - r(A-I)^k$ opisujące liczby klatek rozmiaru $\geqslant k$.

Skoro $(A-I)^3 = 0$, to $r(A-I)^3 = 0$. A zatem $q_3 = 2-0 = 2$. Oczywiście $q_4 = r(A-I)^3 - r(A-I)^4 = 0-0 = 0$. Zatem w postaci Jordana J macierzy A są dwie klatki rozmiaru nie mniejszego niż 3 i zero klatek rozmiaru nie mniejszego niż 4. A zatem J ma dwie klatki rozmiaru 3×3, i żadnej większej. Skoro A (a więc i J) jest rozmiarów 7×7, to suma rozmiarów klatek Jordana macierzy J to 7. Postać Jordana macierzy A ma zatem jedną klatkę 1×1 oraz dwie klatki 3×3, odpowiadające wartości własnej 1. ■

Zadanie 5.2. *Niech $A \in M_{n\times n}(\mathbb{C})$. Pokaż, że jeśli $A^3 = A$, to $r(A) = \operatorname{tr} A^2$.*

Rozwiązanie. Niech J będzie macierzą w postaci Jordana podobną do A. Wiemy, że $r(A) = r(J)$ oraz $\operatorname{tr} A^2 = \operatorname{tr} J^2$. Macierz J jest górnotrójkątna i na jej przekątnej stoją wartości własne macierzy A. Podobnie macierz J^2 jest górnotrójkątna i na jej przekątnej stoją kwadraty wartości własnych macierzy A, czyli wartości własne macierzy A^2. Twierdzimy, że wartości własne macierzy A wynosić mogą jedynie $-1, 0, 1$. Istotnie, jeśli $Av = \lambda v$, dla pewnego wektora niezerowego v, to $A^3v = \lambda^3 v$. A zatem równość $A^3 = A$ pociąga za sobą równość $A^3v = Av$, czyli $(\lambda^3 - \lambda)v = 0$. Skoro $v \neq 0$, to $\lambda^3 - \lambda = 0$, a więc $\lambda \in \{-1, 0, 1\}$. A zatem elementy na przekątnej macierzy J to jedynie $-1, 0$ lub 1. Stąd elementy na przekątnej J^2 mogą wynosić jedynie $0, 1$. Reszta tezy wynika z faktu, że A jest diagonalizowalna. Jak to pokazać?

Skoro $A^3 = A$, to też $J^3 = J$ (też było na wykładzie). Potęgowanie macierzy Jordana polega zaś na potęgowaniu poszczególnych jej klatek. To oznacza, że wszystkie klatki w J są rozmiarów 1×1, bo tylko wtedy trzecia potęga klatki może być równa niej samej (nad $\mathbb{C}$). Zatem liczba niezerowych elementów na przekątnej macierzy J^2 jest jednocześnie rzędem tej macierzy. Skoro te niezerowe elementy muszą wynosić 1, to liczba ta jest też również śladem J^2. A zatem $r(J^2) = \operatorname{tr}(J^2)$, czyli $r(A^2) = \operatorname{tr}(A^2)$. Skoro J i J^2 są diagonalne, to jest jasne, że $r(A) = r(J) = r(J^2)$. Czyli $r(A) = \operatorname{tr}(A^2)$. ■

Zadanie 5.3. *Macierz $A \in M_{n\times n}(\mathbb{C})$ spełnia równanie $A^2 = I$. Pokaż jest to macierz diagonalizowalna. Jakie wartości własne może mieć ta macierz?*

Rozwiązanie. Każdą kwadratową macierz zespoloną można sprowadzić do postaci Jordana. Niech J będzie macierzą w postaci Jordana podobną do A. Wówczas istnieje macierz odwracalna P taka, że $P^{-1}AP = J$. A zatem $J^2 = P^{-1}A^2P = P^{-1}P = I$. Oznacza to, że każda z klatek Jordana macierzy J po podniesieniu do kwadratu jest macierzą identyczności. Załóżmy, że w J istnieje klatka rozmiaru co najmniej 2×2 odpowiadająca wartości własnej a. Wówczas:

$$\begin{bmatrix} a & 1 & 0 & \dots & 0 \\ 0 & a & 1 & \dots & 0 \\ 0 & 0 & a & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

A zatem dostajemy warunki: $a^2 = 1$ oraz $2a = 0$, co oznacza, że powyższe równanie macierzowe jest sprzeczne. Wszystkie klatki Jordana J są zatem rozmiaru 1×1, a zatem A jest diagonalizowalna. Oczywiście jeśli $a \in \mathbb{C}$ jest wartością własną A, to a^2 jest wartością własną I, a zatem $a^2 = 1$. Równanie to ma w liczbach zespolonych jedynie pierwiastki -1 oraz 1. To jedyne możliwe wartości własne macierzy A. ■

Zadanie 5.4. *Rozważmy endomorfizm ϕ przestrzeni V wymiaru 14 nad ciałem algebraicznie domkniętym, przy czym $\dim\ker\phi = 9$ oraz $\operatorname{im}\phi \subseteq \ker\phi$. Znajdź możliwe postaci Jordana endomorfizmu ψ takiego, że $\phi = \psi^3$.*

Rozwiązanie. Warunek $\operatorname{im}\phi \subseteq \ker\phi$ oznacza, że $\phi^2 = 0$. Stąd wszystkie wartości własne ϕ są równe 0. W istocie wielomian charakterystyczny ϕ musi być[1] postaci $\lambda^{14} = 0$. Mamy również $\psi^6 = 0$, a zatem również ψ ma wielomian charakterystyczny $\lambda^{14} = 0$. Mamy zatem

$$\dim\ker\psi^3 = 9, \quad \dim\ker\psi^6 = 14.$$

A zatem jeśli B jest macierzą ψ, to

$$r(B^3) = 5, \quad r(B^6) = 0.$$

Oczywiście $r(B^0) = 14$, więc mamy:

$$q_1 + q_2 + q_3 = r(B^0) - r(B) + r(B) - r(B^2) + r(B^2) - r(B^3) = 9, \quad q_4 + q_5 + q_6 = r(B^3) - r(B^6) = 5.$$

Mamy jednak $q_1 \leqslant q_2 \leqslant q_3 \leqslant q_4 \leqslant q_5 \leqslant g_6$. Nasza macierz może mieć co najwyżej trzy klatki rozmiaru co najmniej 3. A zatem możliwe trójki (q_4, q_5, q_6) to $(3,2,0)$, $(3,1,1)$, $(2,2,1)$. Uwzględniając pierwszy warunek widzimy, że możliwe konfiguracje klatek Jordana macierzy B są następujące:

- dwie klatki rozmiaru 5×5 oraz jedna klatka rozmiaru 4×4,
- jedna klatka rozmiaru 6×6 oraz dwie klatki rozmiaru 4×4,
- po jednej klatce rozmiaru 6×6, 5×5 oraz 3×3,
- po jednej klatce rozmiaru 6×6, 5×5, 2×2 oraz 1×1,
- po jednej klatce rozmiaru 6×6, 5×5 oraz trzy klatki rozmiaru 1×1.

■

Zadanie 5.5. *Znajdź postać Jordana endomorfizmu $\phi \in \operatorname{End}(\mathbb{C}^{10})$ spełniającego:*

$$\begin{cases} \operatorname{tr}\phi = 20 \\ \dim\ker(\phi - 2\,\mathrm{id})^2 = 5 \\ \dim\ker(\phi - 2\,\mathrm{id})^4 = 8 \\ \det(\phi - 3\,\mathrm{id}) = 0. \end{cases}$$

Rozwiązanie. Jeśli A jest macierzą endomorfizmu ϕ, mamy $r(A-2I)^2 = 5$ oraz $r(A-2I)^4 = 2$. Zatem

$$q_1 + q_2 = r(I) - r(A-2I) + r(A-2I) + r(A-2I)^2 = 5,$$

skąd możliwe wartości (q_1, q_2) to $(5,0)$, $(4,1)$ lub $(3,2)$. Mamy też:

$$q_3 + q_4 = r(A-2I)^2 - r(A-2I)^3 + r(A-2I)^3 - r(A-2I)^4 = 3.$$

Oczywiście $q_1 \leqslant q_2 \leqslant q_3 \leqslant q_4$. A zatem musi być $q_3 = 2$ i wtedy $q_4 = 1$. W rezultacie mamy dokładnie jedną klatkę rozmiaru 3 i co najmniej jedną klatkę rozmiaru 4 — czyli dokładnie jedną, bo więcej się nie zmieści. Jedyna konfiguracja, która tu pasuje to oczywiście $q_1 = 3$ oraz $q_2 = 2$. Stąd mamy jedną klatkę o rozmiarze jeden i wiemy, że wielomian charakterystyczny A dzieli się co najmniej przez $(\lambda-2)^8$.

Endomorfizm ten ma jednak także wartość własną 3. A zatem może mieć jeszcze jedną wartość własną i ma, bowiem ich suma razem z krotnościami równa jest śladowi, czyli liczbie 20. Mamy zatem $20 = 8\cdot 2 + 3 + x$, czyli $x = 1$ jest ostatnią szukaną wartością własną. A zatem postać Jordana endomorfizmu ϕ ma 5 klatek:

- trzy klatki odpowiadające wartości własnej 2 rozmiarów 4×4, 3×3 oraz 1×1,
- po jednej klatce rozmiaru 1×1 odpowiadającej wartości własnej 1 oraz 3.

■

Ostatnie zadanie dotyczyć będzie wyznaczania bazy Jordana oraz potęg macierzy niediagonalizowalnej.

[1] To jest prawdziwe dla endomorfizmu przestrzeni nad dowolnym ciałem, o czym wspominamy w dodatku.

Zadanie 5.6. *Rozważmy endomorfizm ϕ przestrzeni $\mathbb{C}^4$ dany w bazie standardowej macierzą:*

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

Wyznacz macierz ϕ w postaci Jordana. Wyznacz A^{100}.

Rozwiązanie. Macierz A jest blokowo górnotrójkątna i oczywiście jej wielomian charakterystyczny to $(\lambda^2+1)^2 = (x-i)^2(x+i)^2$. A zatem ϕ ma wartości własne i oraz $-i$. Mamy też:

$$A - iI = \begin{bmatrix} -i & 1 & 1 & 0 \\ -1 & -i & 0 & 1 \\ 0 & 0 & -i & 1 \\ 0 & 0 & -1 & -i \end{bmatrix}, \quad (A-iI)^2 = \begin{bmatrix} -2 & -2i & -2i & 2 \\ 2i & -2 & -2 & -2i \\ 0 & 0 & -2 & -2i \\ 0 & 0 & 2i & -2 \end{bmatrix},$$

co oznacza, że $r(A-iI) = 3$, $r(A-iI) = 2$. A zatem mamy jedną klatkę Jordana rozmiaru 2 odpowiadającą wartości własnej $-i$. Również wartości własnej i odpowiada klatka rozmiaru 2.

Druga część wymaga użycia postaci Jordana a także wyznaczenia bazy Jordana. Pokażmy jak to zrobić, a potem jak można było tego uniknąć. Jak skorzystać z postaci Jordana J macierzy A? Zapiszemy $A = C^{-1}JC$, gdzie J jest postacią Jordana A, zaś C^{-1} jest macierzą mającą w kolumnie wektory bazy Jordana. Wtedy $A^{100} = C^{-1}J^{100}C$. Czasami podnoszenie macierzy w postaci Jordana do wysokiej potęgi jest znośne. Są zresztą ogólne (dość istotne) wzory, które pozwalają to robić w dowolnej sytuacji.

Trzeba zacząć od wyznaczenia bazy Jordana. W skrócie, mamy:

$$\ker(\phi - i\,\mathrm{id})^2 = \mathrm{lin}((1,i,0,0),(0,0,1,i)), \quad (\phi - i\,\mathrm{id})(0,0,1,i) = (1,i,0,0),$$
$$\ker(\phi + i\,\mathrm{id})^2 = \mathrm{lin}((1,-i,0,0),(0,0,1,-i)), \quad (\phi + i\,\mathrm{id})(0,0,1,-i) = (1,-i,0,0).$$

A zatem po krótkich rachunkach:

$$C^{-1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ i & 0 & -i & 0 \\ 0 & 1 & 0 & 1 \\ 0 & i & 0 & -i \end{bmatrix}, \quad C = \frac{1}{2}\begin{bmatrix} 1 & -i & 0 & 0 \\ 0 & 0 & 1 & -i \\ 1 & i & 0 & 0 \\ 0 & 0 & 1 & i \end{bmatrix}$$

Teraz podnosimy do potęgi 100 klatki Jordana J_1, J_2 macierzy Jordana J endomorfizmu ϕ, aby dalej skorzystać z zależności:

$$J^{100} = \begin{bmatrix} J_1^{100} & 0 \\ 0 & J_2^{100} \end{bmatrix}.$$

Nietrudno widzieć, że:

$$J_1^n = \begin{bmatrix} i & 1 \\ 0 & i \end{bmatrix}^n = \begin{bmatrix} i^n & i^{n-1}n \\ 0 & i^n \end{bmatrix}, \quad J_2^n = \begin{bmatrix} i & 1 \\ 0 & i \end{bmatrix}^n = \begin{bmatrix} (-i)^n & (-i)^{n-1}n \\ 0 & (-i)^n \end{bmatrix},$$

czyli:

$$A^{100} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ i & 0 & -i & 0 \\ 0 & 1 & 0 & 1 \\ 0 & i & 0 & -i \end{bmatrix} \cdot \begin{bmatrix} 1 & -100i & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 100i \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \frac{1}{2}\begin{bmatrix} 1 & -i & 0 & 0 \\ 0 & 0 & 1 & -i \\ 1 & i & 0 & 0 \\ 0 & 0 & 1 & i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -100 \\ 0 & 1 & 100 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Czy można inaczej? Zauważmy, że kładąc: $C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ mamy $A = \begin{bmatrix} C & I \\ 0 & C \end{bmatrix}$ oraz:

$$A^n = \begin{bmatrix} C^n & nC^{n-1} \\ 0 & C^n \end{bmatrix}$$

Macierz C to tymczasem macierz obrotu o $3\pi/2$, czyli jej setna potęga to macierz obrotu o kąt 150π, czyli I. A nawet jeśli nie widzieliśmy tego obrotu, to widać, że $C^2 = -I$. Mamy więc $C^{99} = -C$, czyli:

$$A^{100} = \begin{bmatrix} 1 & 0 & 0 & -100 \\ 0 & 1 & 100 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

■

5.2 Trivia. Problem Collatza

Przy tak poważnym twierdzeniu jak twierdzenie Jordana nie wypada opowiadać o matematycznych zabawkach, a warto przytoczyć prawdziwy i otwarty od dawna problem badawczy. Będzie on dotyczył następującej prostej funkcji na zbiorze liczb naturalnych.

Definicja 5.2

Definiujemy następującą FUNKCJĘ COLLATZA $f : \mathbb{N} \to \mathbb{N}$:

$$f(n) = \begin{cases} 3n+1, & \text{gdy n jest nieparzyste,} \\ \frac{n}{2}, & \text{gdy n jest parzyste.} \end{cases}$$

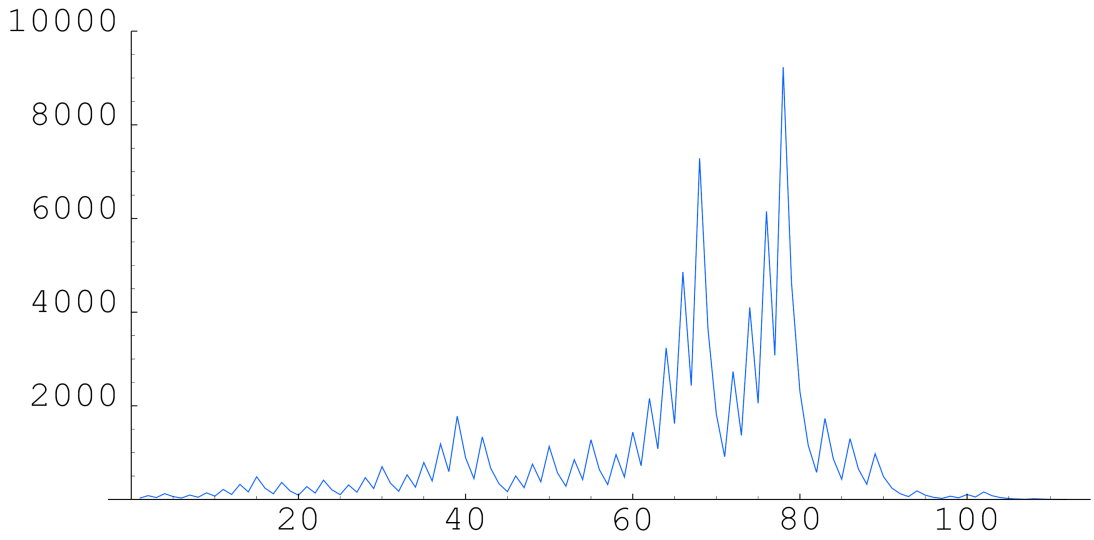
PROBLEM COLLATZA (1937). *Rozstrzygnąć czy dla każdego $k \in \mathbb{N}$ w ciągu $f^n(k)$ znajduje się liczba* 1.

Zagadnienie to było rozważane przez wielu słynnych matematyków, między innymi Ulama i Kakutaniego. Paul Erdős wypowiedział o nim słynne zdanie: „mathematics is not yet ready for such problems".

Przyjrzyjmy się ciągom $(n, f(n), f^2(n), \ldots)$ dla kilku liczb całkowitych:

$$(n, f(n), f^2(n), \ldots) = \begin{cases} n=5: & (5, 16, 8, 4, 2, 1, 4, 2, 1, \ldots) \\ n=6: & (6, 3, 10, 5, 16, 8, 4, 2, 1, 4, 2, 1, \ldots) \\ n=7: & (7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, \ldots) \\ n=8: & (8, 4, 2, 1, 4, 2, 1, \ldots) \\ n=9: & (9, 28, 14, 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, \ldots) \\ n=10: & (10, 5, 16, 8, 4, 2, 1, 4, 2, 1, \ldots) \end{cases}.$$

Jak widzimy, ciągi te w końcu „schodzą" do 1, a potem powtarza się cykl $(1, 4, 2)$. Można by spodziewać się, że ciągi te nie będą osiągały zawrotnych wartości w porównaniu do liczby początkowej.



Przebieg $f^k(27)$. Źródło: Wikipedia.

Tymczasem patrząc na $n = 27$ stwierdzamy, że $f^{36}(n) = 1186$, zaś $f^{77}(n) = 9232$. Okazuje się to jednak być największa wartość tego ciągu, a już $f^{111}(n) = 1$. Zachowanie $f^k(n)$ nie jest łatwo przewidywalne.

Gdy problem jest bardzo trudny matematycy próbują go przetłumaczyć na inny[2] problem. Równoważnych sformułowań wielki problemów takich jak hipoteza Riemanna są dziesiątki, o ile nie setki. Problem Collatza może nie jest milenijny (choć Erdős oferował 500 dolarów za jego rozwiązanie), ale ma ładne przetłumaczenie na problem algebroliniowy, więc kto wie – może Państwo go rozwiążą?

[2]Zgodnie ze sloganem: jeśli chcesz zobaczyć co jest na szczycie góry wejdź na sąsiednią i popatrz przez lornetkę.

Definicja 5.3

Niech $f : \mathbb{N} \to \mathbb{N}$ będzie dowolną funkcją. Dla każdego $n \in \mathbb{N}$ określamy macierz[a] $A_f(n) \in M_n(\mathbb{C})$ o wyrazach a_{ij} postaci:

$$a_{ij} = \begin{cases} 1, & \text{jeśli } i = f(j), \text{ dla pewnych } i, j \in \{1, \ldots, n\}, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Dla (przyspieszonej) funkcji Collatza macierze $A_f(n)$ oznaczamy jako A_n.

[a]Innymi słowy $A_f(n)$ jest macierzą sąsiedztwa grafu skierowanego $\Gamma_f(n)$ o wierzchołkach $1, \ldots, n$ oraz o krawędziach $i \to j$ występujących wtedy i tylko wtedy, gdy $i = f(j)$.

„Przyspieszona" funkcja Collatza opisuje ten sam proces, ale nieco skrócony:

$$f(n) = \begin{cases} \frac{3n+1}{2}, & \text{gdy n jest nieparzyste,} \\ \frac{n}{2}, & \text{gdy n jest parzyste.} \end{cases}$$

Przyjrzyjmy się jak wygląda macierz A_8.

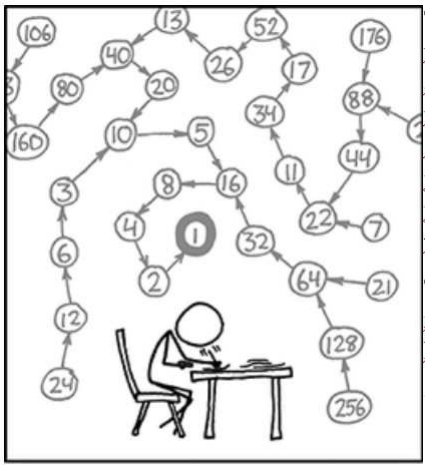
$$A_8 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Zauważmy, że dla ustalonego $m \geqslant 2$ macierz A_m nie zawiera całej informacji o obrazach $f(i)$, dla $1 \leqslant i \leqslant m$. Na przykład dla $m = 5$ mamy $f(5) = 8$, a więc piąty wiersz macierzy A_5 jest zerowy. A zatem na przykład nie wiemy czym są $f^i(5)$, bo $3 \to 5 \to 8 \to 4 \to 2 \to 1$ oraz $8 > 5$. Tymczasem $f^i(4)$ mogą być całkowicie odczytane z A_5.

Przeformułowanie problemu Collatza na problem algebry liniowej można znaleźć w pracy studenckiej „*The Jordan Canonical Form for a Class of Zero–One Matrices*" autorstwa Cardona i Tuckfielda [3]. W szczególności w rozwiązaniu problemu bardzo przydałaby się znajomość wielomianu charakterystycznego $w_A(x)$ macierzy A_n. Hipoteza studentów mówi, że jest on równy

$$(-1)^n \cdot x^{n-2} \cdot (x^2 - 1).$$

Może Państwo to pokażą?



[3]https://repository.upenn.edu/cgi/viewcontent.cgi?article=1314&context=oid_papers

Rozdział 6

Twierdzenie o rozkładzie prymarnym Twierdzenie Jordana-Chevalleya

Przypomnijmy, że jeśli $V = V_1 \oplus \ldots \oplus V_k$ oraz jeśli $\phi \in \mathrm{End}(V)$ jest taki, że $V_1, \ldots, V_k$ są podprzestrzeniami ϕ-niezmienniczymi w V, wówczas ϕ ma w pewnej bazie $\mathcal{A}$ macierz blokowo-diagonalną postaci

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & A_k \end{bmatrix}, \quad \text{gdzie } A_i \in M_{\dim(V_i)}(K).$$

Niestety uzyskiwanie rozkładów przestrzeni liniowych na sumy proste podprzestrzeni niezmienniczych bywa bardzo trudne. Zauważmy, że kryterium diagonalizowalności czy triangularyzowalności macierzy dotyczy dokładnie sytuacji, gdy V jest sumą prostą odpowiednio podprzestrzeni własnych lub pierwiastkowych. Niestety nie jest prawdą nad każdym ciałem, że jeśli $W \subseteq V$ jest podprzestrzenią niezmienniczą względem $\phi \in \mathrm{End}(V)$, to istnieje podprzestrzeń ϕ-niezmiennicza W' taka, ze $W \oplus W' = V$. Wybór odpowiednich podprzestrzeni, dla których takie dopełnienia da się uzyskać jest treścią głębokich twierdzeń.

Aby zilustrować problem, rozważmy endomorfizm ϕ zadany macierzą w postaci Jordana

$$M(\phi)^{st}_{st} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Niech V będzie podprzestrzenią ϕ-niezmienniczą i niech $(x_1, x_2, x_3) \in V$. Wówczas

$$\phi(x_1, x_2, x_3) = (x_2, x_3, 0) \in V, \quad \text{stąd} \quad \phi(x_2, x_3, 0) = (x_3, 0, 0) \in V.$$

A zatem wszystkie podprzestrzenie ϕ-niezmiennicze zawierające wektory o drugiej lub trzeciej współrzędnej niezerowej zawierają wektor $(1, 0, 0)$. Ale poza nimi jest tylko jedna niezerowa podprzestrzeń ϕ-niezmiennicza $\mathrm{lin}((1, 0, 0))$. A zatem każda niezerowa podprzestrzeń ϕ-niezmiennicza zawiera $(1, 0, 0)$. A zatem nie istnieją niezerowe podprzestrzenie ϕ-niezmiennicze takie, że V, W takie, że $\mathbb{R}^3 = V \oplus W$.

Istnieje też bardziej wyrafinowany powód braku rozkładu, o którym mówi tytułowe twierdzenie. Otóż wielomianu charakterystycznego ϕ równego $-\lambda^3$ nie da się rozłożyć na dwa „względnie pierwsze" wielomiany stopnia równego co najmniej 1. A właśnie istnienie takiego rozkładu gwarantuje istnienie wspomnianego rozkładu ϕ-niezmienniczego, jak mówić będzie twierdzenie o rozkładzie prymarnym.

Twierdzenie owo, któremu poświęcimy pierwszą część wykładu, opowiada o pięknym i zaskakującym związku pomiędzy rozkładami na sumy proste oraz rozkładalnością wielomianu charakterystycznego. Prawdziwy „cud" tego twierdzenia polega na tym, że odpowiednia rozkładalność wielomianu charakterystycznego gwarantuje rozkłady ϕ-niezmiennicze. Przyjrzymy się tej sprawie, przywołując wiedzę z materiałów uzupełniających z pierwszego semestru. Interesować nas będzie badanie jądra wielomianu $p(\phi)$ od endomorfizmu ϕ, dla rozkładalnych wielomianów $p \in K[x]$.

Definicja 6.1: Wielomian (nie)rozkładalny

Niech K bedzie ciałem.

- wielomian $f \in K[\lambda]$ nazywamy ROZKŁADALNYM, jeśli istnieją wielomiany $p, q \in K[\lambda]$ takie, że $0 < \deg p, \deg q < \deg f$ oraz
$$f = p \cdot q.$$
Niezerowy wielomian, który nie jest rozkładalny nazywamy wielomianem NIEROZKŁADALNYM.
- NAJWIĘKSZYM WSPÓLNYM DZIELNIKIEM wielomianów $p_1, \ldots, p_n$ należących do $K[\lambda]$ nazywamy wielomian najwyższego możliwego stopnia dzielący wszystkie te wielomiany. Jeśli
$$\mathrm{NWD}(p_1, \ldots, p_n) = 1,$$
wówczas mówimy, że wielomiany te są WZGLĘDNIE PIERWSZE.

W naszych rozważaniach będziemy też korzystali z następującego prostego faktu, dowodzonego analogicznie jak dla liczb całkowitych. W języku algebry mówi ono, że pierścień $K[\lambda]$ jest tzw. dziedziną ideałów głównych. Zachęcam Czytelnika do przeprowadzenia samodzielnego uzasadnienia, korzystającego z przedstawionej niżej wskazówki.

Obserwacja 6.1: Lemat Bezout

Niech K bedzie ciałem. Dla dowolnych $p_1, \ldots, p_n \in K[\lambda]$ oraz dowolnego ich NWD równego g istnieją wielomiany $q_1, \ldots, q_n \in K[\lambda]$ takie, że
$$q_1 p_1 + \ldots + q_n p_n = g.$$

Idea dowodu jest następująca. Rozważamy zbiór I wielomianów postaci:
$$I = \{q_1 p_1 + \ldots + q_n p_n \,|\, q_1, \ldots, q_n \in K[\lambda]\}$$
i wybieramy niezerowy element $d \in I$ najmniejszego możliwego stopnia. Można wykazać, że
$$d = \mathrm{NWD}(p_1, \ldots, p_n).$$

Warto odnotować, że istotne jest założenie o tym, że wielomiany mają współczynniki w ciele. W pierścieniu $\mathbb{Z}[x]$ największy wspólny dzielnik 2 oraz x to 1, ale nie istnieją wielomiany $f, g \in \mathbb{Z}[x]$, że $1 = 2f(x) + xg(x)$.

Twierdzenie 6.1: Twierdzenie o rozkładzie prymarnym

Niech $p \in K[\lambda]$ będzie wielomianem o rozkładzie
$$p = p_1 \cdot \ldots \cdot p_k,$$
gdzie $p_1, \ldots, p_k$ są czynnikami **parami względnie pierwszymi** stopnia większego od 0. Niech $\phi \in \mathrm{End}(V)$. Wówczas ma miejsce rozkład
$$\ker p(\phi) = \ker p_1(\phi) \oplus \ldots \oplus \ker p_k(\phi) = \bigoplus_{i=1}^{k} \ker p_i(\phi)$$
na sumę prostą ϕ-niezmienniczych podprzestrzeni w V. Co więcej:

(a) Naturalne rzutowanie $\pi_i : \ker p(\phi) \to \ker p_i(\phi)$ jest wielomianem od ϕ.

(b) Jeśli $U \subseteq \ker p(\phi)$ jest ϕ-niezmiennicza, to mamy:
$$U = (\ker p_1(\phi) \cap U) \oplus \ldots \oplus (\ker p_k(\phi) \cap U).$$

Dwie kluczowe intuicje dotyczące powyższego fundamentalnego wyniku, gdy jądro odpowiedniego wielomianu od endomorfizmu jest całą przestrzenią.

- Jeśli V jest skończenie wymiarową przestrzenią nad ciałem K, to dla każdego $\phi \in \mathrm{End}(V)$ mamy $w_\phi(\phi) = 0$ (twierdzenie Cayleya-Hamiltona), czyli $\ker p(\phi) = V$ dla $p = w_\phi(\lambda)$. Podobnie rozumować można dla wielomianu minimalnego. Na przykład dla endomorfizmu diagonalizowalnego ϕ o spektrum złożonym z elementów $a_1, \ldots, a_k$ ciała K mamy $m_\phi(\lambda) = (a_1 - \lambda)(a_2 - \lambda) \cdot \ldots \cdot (a_k - \lambda)$ i rzeczywiście

  $$V = \ker(m_\phi(\phi)) = \ker(\phi - a_1 \,\mathrm{id}) \oplus \ker(\phi - a_2 \,\mathrm{id}) \oplus \ldots \oplus \ker(\phi - a_k \,\mathrm{id}) = V_{(a_1)} \oplus V_{(a_2)} \oplus \ldots \oplus V_{(a_k)}.$$

- Nawet jeśli $\dim V = \infty$, ale mimo to $\phi \in \mathrm{End}(V)$ spełnia równanie wielomianowe, np. dowolny rzut $\phi \in \mathrm{End}(V)$ spełnia

  $$(\phi^2 - \phi)(\alpha) = 0,$$

  dla $\alpha \in V$, to twierdzenie działa: mamy $V = \ker(\phi) \oplus \ker(\phi - \mathrm{id})$.

Zobaczmy przykłady rozkładów, o których mówi nasze twierdzenie.

- Rozważmy $\phi \in \mathrm{End}(\mathbb{R}^4)$ dany macierzą:

  $$M(\phi)^{st}_{st} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

  Łatwo widzieć, że $w_\phi(\lambda) = \lambda^4 + 1$. A zatem $w_\phi(\lambda)$ nie ma pierwiastków w $\mathbb{R}$. Skoro $w_\phi(\lambda) = p_1 p_2$, gdzie $p_1 = \lambda^2 - 2\lambda + 2, p_2 = \lambda^2 + 2\lambda + 2$, to

  $$\mathbb{R}^4 = \ker(\phi^4 + \mathrm{id}) = \ker(\phi^2 - 2\phi + 2\,\mathrm{id}) \oplus \ker(\phi^2 + 2\phi + 2\,\mathrm{id}),$$

  zatem mimo, że ϕ nie jest triangularyzowalny, istnieje baza $\mathcal{A}$ przestrzeni $\mathbb{R}^4$, w której ϕ ma macierz blokowo-diagonalną:

  $$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \quad \text{dla pewnych } A, B \in M_2(\mathbb{R}).$$

- Rozważmy $\phi \in \mathrm{End}(\mathbb{R}^4)$ dany macierzą:

  $$M(\phi)^{st}_{st} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

  Oczywiście $w_\phi(\lambda) = (1 - \lambda)^4$, ale ϕ nie jest diagonalizowalny, bo $\dim V_{(1)} = \dim \ker(\phi - \mathrm{id}) = 2$. Mamy też $\mathbb{R}^4 = \ker(\phi - \mathrm{id})^3$. Choć $\ker(\phi - \mathrm{id})^3$ rozkłada się na sumę podprzestrzeni cyklicznych (dwóch), to rozkład ten nie jest sumą prostą jąder endomorfizmów $(\phi - \mathrm{id})^2$ oraz $(\phi - \mathrm{id})$.

Dowód twierdzenia o rozkładzie prymarnym

Przechodzimy do właściwego dowodu. Pokażemy najpierw, że dla każdego i mamy:

$$\ker p_i(\phi) \cap \sum_{j \neq i} \ker p_j(\phi) = \{0\}. \quad (*)$$

Niech $v_1 + \ldots + v_k = 0$, dla pewnych $v_i \in \ker p_i(\phi)$. Wystarczy wykazać, że $v_i = 0$, dla $1 \leqslant i \leqslant k$. Niech $\widehat{p_i} = p/p_i \in K[\lambda]$, dla $1 \leqslant i \leqslant k$. Dla $i \neq j$ mamy $\widehat{p_i}(\phi)(v_j) = 0$, bo $\widehat{p_i}(\phi)$ ma czynnik $p_j(\phi)$. Skoro wielomiany p_i oraz $\widehat{p_i}$ są względnie pierwsze, to na mocy Lematu Bezout istnieją $a_i, b_i \in K[\lambda]$, że $a_i p_i + b_i \widehat{p_i} = 1$, czyli

$$a_i(\phi) p_i(\phi) + b_i(\phi) \widehat{p_i}(\phi) = \mathrm{id}\,.$$

Zatem podstawiając pod tę równość v_i dostajemy $(*)$, bowiem:

$$v_i = a_i(\phi) \underbrace{p_i(\phi)(v_i)}_{0} + b_i(\phi) \widehat{p_i}(\phi)(v_i) = b_i(\phi) \widehat{p_i}(\phi)(-\sum_{j \neq i} v_j) = -\sum_{j \neq i} b_i(\phi) \underbrace{\widehat{p_i}(\phi)(v_j)}_{0} = 0.$$

Pokażemy teraz, że:

$$\ker p(\phi) = \ker p_1(\phi) + \ldots + \ker p_k(\phi).$$

Inkluzja $\supseteq$ jest oczywista, bowiem jeśli wektor przechodzi na zero przy dzielnku $p(\phi)$, to przechodzi na zero i przy $p(\phi)$. Dowodzimy $\subseteq$. Skoro $p_1, \ldots, p_k$ są parami względnie pierwsze, to również $\widehat{p_1}, \ldots, \widehat{p_k}$ są parami względnie pierwsze. Zatem istnieją wielomiany $q_1, \ldots, q_k \in K[\lambda]$, że $q_1\widehat{p_1} + \ldots + q_k\widehat{p_k} = 1$, czyli:

$$q_1(\phi)\widehat{p_1}(\phi) + \ldots + q_k(\phi)\widehat{p_k}(\phi) = \mathrm{id}\,.$$

Niech $v \in \ker p(\phi)$. Z równości wyżej mamy $v = \mathrm{id}(v) = \sum_{i=1}^{k} q_i(\phi)\widehat{p_i}(\phi)(v)$. Wystarczy pokazać, że $q_i(\phi)\widehat{p_i}(\phi)(v) \in \ker p_i(\phi)$, co wynika z:

$$p_i(\phi)\left(q_i(\phi)\widehat{p_i}(\phi)(v)\right) = q_i p(\phi)(v) = q_i(\phi)0 = 0.$$

Pozostaje uzasadnienie punktów (a), (b). Widzimy jednak, że rzutowanie $\pi_i : \ker p(\phi) \to \ker p_i(\phi)$ to po prostu wielomian $q_i(\phi)\widehat{p_i}(\phi)$. Dla dowodu (b) bierzemy $u \in U$ i piszemy $u = u_1 + \ldots + u_k$, gdzie $u_i \in \ker p_i(\phi)$. Zatem $u_i = \pi_i(u) = q_i(\phi)\widehat{p_i}(\phi)(u) \in U$, bowiem U jest ϕ-niezmiennicza. Zatem $u_i \in U$.

* * *

Czytelnik z pewnością dostrzega różnicę jakościową pomiędzy dowodem przedstawionym wyżej, a dowodami z poprzednich wykładów. Różnica jest fundamentalna, ponieważ już w sposób całkowicie jawny korzystamy z tego, że na przestrzeni liniowej V nie działa tylko endomorfizm ϕ, ale cała jego algebra $E(\phi)$ złożona z wielomianów od endomorfizmu ϕ. W istocie przestrzeń V ma strukturę tzw. $K[\lambda]$-modułu, którą można określać dla dowolnego $\phi \in \mathrm{End}(V)$. Jest to struktura przypominająca pozornie strukturę przestrzeni liniowej, gdzie skalary są już niekoniecznie z ciała, ale z pierścienia. Głębia polega nie tylko na wymianie współczynników, ale na to w jaki sposób działają one na V. W tym przypadku „mnożenie" elementu $v \in V$ przez „skalar" $f(\lambda) \in K[\lambda]$ jest w istocie wykonaniem na wektorze v endomorfizmu $f(\phi)$. O tego typu strukturach uczyć się Państwo będą na bardziej zaawansowanych kursach algebraicznych.

Przejdziemy teraz do drugiego zagadnienia blisko związanego z twierdzeniem Jordana i powyższym rezultatem. Z twierdzenia Jordana wynika, że endomorfizm triangularyzowalny można rozłożyć na sumę endomorfizmu nilpotentnego i diagonalizowalnego. Przykładowo, dla macierzy w postaci Jordana tego endomorfizmu można wydzielić „część diagonalną" i „część ściśle górnotrójkątną".

- Czy rozkład ten jest jednoznaczny? Nie, gdyż choćby $A = \begin{bmatrix}1 & 2\\ 0 & 3\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix} + \begin{bmatrix}0 & 2\\ 0 & 0\end{bmatrix}$, przy czym składniki po prawej nie są przemienne.
- W praktyce przydatny jest nie tylko rozkład $A = D + N$, gdzie D to macierz diagonalna i N to macierz nilpotentna, ale też fakt, że $DN = ND$, bo wtedy $A^n = (D+N)^n = \sum_{k=0}^{n} \binom{n}{k} D^k N^{n-k}$.

Jak się okazuje, rozkład endomorfizmu triangularyzowalnego na składową nilpotentną i diagonalną jest jednoznaczny, jeśli zażądamy, by składowe te były przemienne. Jest to tzw. rozkład Jordana-Chevalleya.

Twierdzenie 6.2: Addytywny rozkład Jordana-Chevalleya

Niech V będzie skończenie wymiarową przestrzenią liniową nad K oraz niech $\phi \in \mathrm{End}(V)$ będzie triangularyzowalny. Wówczas istnieje jednoznaczny rozkład

$$\phi = \phi_S + \phi_N, \qquad (*)$$

- gdzie ϕ_S jest diagonalizowalny,
- ϕ_N jest nilpotentny,
- $\phi_S \circ \phi_N = \phi_N \circ \phi_S$.

Co więcej ϕ_S, ϕ_N należą do algebry endomorfizmu ϕ (są wielomianami od ϕ).

Zauważmy, że jeśli zastąpimy diagonalizowalność byciem automorfizmem, to stosowny wynik już znamy. Pojawił się w jednym z dowodów, a któremu przypisuje się w literaturze nazwisko autora.

Obserwacja 6.2: Lemat Fittinga

Niech V będzie skończenie wymiarową przestrzenią liniową nad K oraz niech $\phi \in \mathrm{End}(V)$. Wówczas istnieje jednoznacznie wyznaczony rozkład ϕ-niezmienniczy $V = V_n \oplus V_o$ taki, że:

- $\phi|_{V_n}$ jest nilpotentny,
- $\phi|_{V_o}$ jest odwracalny.

Jeśli $V_n, V_0 \neq 0$, to $\phi = \phi_n + \phi_0$, gdzie ϕ jest nilpotentny, zaś ϕ_o – automorfizm.

Uzasadnijmy jedynie jednoznaczność rozkładu wyżej. Niech $V = A \oplus B$, przy czym $\phi|_A$ jest nilpotentny, zaś $\phi|_B$ – automorfizm. Wówczas $A \subseteq \ker(\phi^k)$, dla **pewnego** k oraz $B \subseteq \mathrm{im}(\phi^k)$, dla **każdego** k. Stąd $A \subseteq V_n$ oraz $B \subseteq V_o$. Aby sumy wymiarów A, B dawały wymiar V, czyli sumę wymiarów V_n oraz V_o musi być $A = V_n$ oraz $B = V_o$.

Dowód (Addytywny rozkład Jordana-Chevalleya)

Dowód rozkładu Jordana-Chevalleya dzieli się na dwie części – istnienie rozkładu i jego jednoznaczność. Zacznijmy od istnienia. Korzystając z twierdzenia o rozkładzie na podprzestrzenie pierwiastkowe piszemy $V = V_{[\lambda_1]} \oplus \ldots \oplus V_{[\lambda_k]}$, gdzie $\lambda_1, \ldots, \lambda_k$ – wartości własne endomorfizmu triangularyzowalnego ϕ. Niech $\pi_i : V \to V_{[a_i]}$ będą naturalnymi rzutowaniami i niech

$$\phi_S := \lambda_1\pi_1 + \ldots + \lambda_k\pi_k \in \mathrm{End}(V).$$

Zauważmy, że endomorfizm ϕ_S ograniczony do $V_{[\lambda_i]}$ to $\lambda_i \,\mathrm{id}$, czyli ϕ_S jest diagonalizowalny. Zgodnie punktem (a) twierdzenia o rozkładzie prymarnym endomorfizmy π_i są wielomianami od ϕ, więc także ϕ_S jest wielomianem od ϕ. Stąd także ϕ_N określony równością $\phi - \phi_S$ jest wielomianem od ϕ. Oczywiście ϕ_S, ϕ_N jako wielomiany od ϕ są przemienne. Mamy też:

$$\phi_N = \phi - \phi_S = \phi \circ \underbrace{(\pi_1 + \ldots + \pi_k)}_{\mathrm{id}} - (\lambda_1\pi_1 + \ldots + \lambda_k\pi_k) = \sum_i (\phi - \lambda_i \,\mathrm{id}) \circ \pi_i,$$

czyli $\phi_N = \phi - \lambda_i \,\mathrm{id}$ na $V_{[\lambda_i]}$, co oznacza, że ϕ_N jest nilpotentny.

Trudniejszy jest nieco dowód jednoznaczności. Wymaga on zastosowania dwóch faktów, z których jeden jest bardzo prosty, a drugi zaskakująco skomplikowany:

- Jeśli $A, B \in M_n(K)$ są diagonalizowalne nad K i przemienne, to $A+B$ również jest diagonalizowalna nad K (wspólnie z A, B).
- Jeśli $A, B \in M_n(K)$ są nilpotentne i przemienne, to $A + B$ też jest nilpotentna.

W przypadku macierzy nilpotentnych uzasadnienie jest bardzo latwe. Niech $A^r = 0$ oraz $A^s = 0$. Weźmy $N = 2\max(r, s)$. Wówczas z wzoru dwumianowego $(A + B)^N = \sum_{k=0}^{N} \binom{N}{k} A^k B^{N-k}$ wynika, że jeden z czynników A^k lub B^{N-k} jest zawsze zerowy, więc $A+B$ jest nilpotentna. Jeśli natomiast chodzi o diagonalizowalność, dowód wymaga zastosowania twierdzenia o wspólnej diagonalizowalności endomorfizmów przemiennych, omówionego w uzupełnieniu do tego wykładu. Odsyłam do uzasadnienia, które zostało tam przedstawione.

Dowodzimy jednoznaczność rozkładu Jordana-Chevalleya. Załóżmy, że dla macierzy $A \in M_n(K)$ mamy $A = A_S + A_N = S + N$, gdzie macierze A_S, S są diagonalizowalne nad K, zaś A_N, N – nilpotentne oraz $A_N A_S = A_S A_N, SN = NS$. Załóżmy też, że A_S, A_N są wielomianami od A. Zatem

$$NA = N(S + N) = NS + N^2 = SN + N^2 = (S + N)N = AN.$$

Analogicznie pokazujemy, że macierz S jest przemienna z A. Co więcej, S, N są też przemienne z każdym wielomianem od A, w tym z A_S, A_N. Wiemy zatem (dwa fakty wyżej), że $A_S - S$ jest diagonalizowalna, zaś $N - A_N$ jest nilpotentna. Ale $A_S - S = N - A_N$, zatem to musi być macierz zerowa (dlaczego?). Zatem $A_S = S$ oraz $A_N = N$. Dowód jest zakończony.

Z dowodu wynika algorytm poszukiwania rozkładu Jordana-Chevalleya macierzy triangularyzowalnej A.

- Rozłóż $w_A(\lambda) = \prod_{i=1}^{k} (a_i - \lambda)^{t_i}$ i połóż $p_i = (a_i - \lambda)^{t_i}$.
- Znajdź wielomiany $q_i \in K[\lambda]$, takie że $\sum q_i \widehat{p_i} = 1$. Jak wiemy $\pi_i = q_i \widehat{p_i}$.
- Zdefiniuj $A_S = \sum a_i q_i(A) \widehat{p_i}(A)$ oraz $A_0 = A - A_S$.

Przykład[1]. Weźmy

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in M_4(\mathbb{C}).$$

Mamy $w_A(\lambda) = (\lambda^2 + 1)^2$. Stąd $a_1 = -i$, $a_2 = i$ oraz $p_1 = \widehat{p_2} = (\lambda + i)^2$ oraz $p_2 = \widehat{p_1} = (\lambda - i)^2$. Szukamy q_1, q_2 takich, że

$$q_1 \widehat{p_1} + q_2 \widehat{p_2} = q_1(\lambda - i)^2 + q_2(\lambda)(\lambda + i)^2 = 1.$$

Na mocy algorytmu Euklidesa (i odwrotnego) uzyskujemy:

$$q_1(\lambda) = \frac{1}{4} i\lambda - \frac{1}{2}, \quad q_2(\lambda) = -\frac{1}{4} i\lambda - \frac{1}{2}.$$

Stąd:

$$-iq_1 \widehat{p_1} + iq_2 \widehat{p_2} = -i(\lambda - i)^2 \left(\frac{1}{4} i\lambda - \frac{1}{2} \right) + i(\lambda + i)^2 \left(-\frac{1}{4} i\lambda - \frac{1}{2} \right) = \frac{1}{2} \lambda^3 + \frac{3}{2} \lambda.$$

Stąd $A_s = \frac{1}{2} A^3 + \frac{3}{2} A$ oraz

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

Zobaczyliśmy zatem przykłady rozkładów na endomorfizm nilpotentny i automorfizm oraz na endomorfizm nilpotentny i diagonalizowalny. Czy są inne? Można sformułować ogólniejszą wersję rozkładu Jordana-Chevalleya w oparciu o pojęcie tzw. endomorfizmu półprostego, czyli takiego, że dla każdej ϕ-niezmienniczej podprzestrzeni U w V istnieje podprzestrzeń ϕ-niezmiennicza W taka, że $V = U \oplus W$. Oczywiście endomorfizm diagonalizowalny jest półprosty, bo obcięcie do podprzestrzeni niezmienniczej jest również diagonalizowalne. Endomorfizm nilpotentny stopnia większego niż 1 nie jest półprosty.

Okazuje się, że prawdziwe jest twierdzenie Jordana-Chevalleya, w którym zamiast diagonalizowalności może być wpisana półprostota endomorfizmu. Stąd składowa diagonalizowalna endomorfizmu w twierdzeniu wyżej miała indeks „S". Dowód jest jednak trudniejszy. Również rozstrzyganie czy endomorfizm jest półposty nie jest zagadnieniem banalnym i w zasadzie prowadzone jest już w innym, ogólniejszym języku teorii modułów. Również zresztą twierdzenie o rozkładzie prymarnym jest częścią teorii modułów i na Algebrze II poznają Państwo jego zdecydowanie ogólniejszą wersję, unifikującą różne działy algebry.

Istnieje również „multyplikatywna wersja" rozkładu Jordana-Chevalleya, której dowód zostawiamy jako ćwiczenie. Jest ona przydatna w szukaniu rozkładów macierzy.

Twierdzenie 6.3: Multyplikatywny rozkład Jordana-Chevalleya

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K oraz niech $\phi \in \mathrm{End}(V)$ będzie izomorfizmem triangularyzowalnym. Wówczas istnieje dokładnie jedno przedstawienie

$$\phi = \phi_S \circ \phi_U = \phi_U \circ \phi_S,$$

gdzie ϕ_S jest diagonalizowalny, zaś ϕ_U jest unipotentny, czyli z definicji: $\phi_U - \mathrm{id}$ jest nilpotentny.

[1] Za `https://mat.polsl.pl/sjpam/zeszyty/z1/znps_mat_stos_01_2011_str_007-028.pdf`.

6.1 Uzupełnienie. Rodziny endomorfizmów przemiennych Wspólna diagonalizowalność i triangularyzowalność

Niech $D_n(K)$ będzie podzbiorem $M_n(K)$ złożonym z macierzy diagonalizowalnych nad K. Wiemy, że zbiór ten jest sumą rozłączną pewnej liczby klas abstrakcji relacji podobieństwa na zbiorze $M_n(K)$. Nie należą tu nigdy (dla $n > 1$) wszystkie klasy, ale pokazaliśmy na wykładzie, że zbiór ten jest stosunkowo złożony. Okazuje się, że suma PRZEMIENNYCH endomorfizmów diagonalizowalnych z $\mathrm{End}(V)$ jest diagonalizowalna.

Definicja 6.2: Endomorfizmy wspólnie diagonalizowalne

Niech V będzie przestrzenią skończonego wymiaru nad ciałem K. Powiemy, że rodzina endomorfizmów $\{\phi_t \in \mathrm{End}(V), t \in T\}$ jest WSPÓLNIE DIAGONALIZOWALNA nad K, jeśli istnieje baza $\mathcal{A}$ przestrzeni V, w której dla każdego $t \in T$ macierz $M(\phi_t)^{\mathcal{A}}_{\mathcal{A}}$ jest diagonalna (czyli $\mathcal{A}$ to baza złożona z wektorów własnych tych ϕ_t).

Naszym celem jest dowód zaskakującego twierdzenia.

Twierdzenie 6.4

Niech V będzie przestrzenią skończonego wymiaru nad ciałem K oraz niech $\{\phi_t \in \mathrm{End}(V), t \in T\}$ będzie rodziną endomorfizmów **diagonalizowalnych**. Następujące warunki sa równoważne:

(1) Dowolne dwa elementy rodziny $\{\phi_t \in \mathrm{End}(V), t \in T\}$ są przemienne.

(2) Rodzina $\{\phi_t \in \mathrm{End}(V), t \in T\}$ jest **wspólnie diagonalizowalna**.

Typowym przykładem przemiennych rodzin są elementy algebry endomorfizmu, która jest przemienna. Innymi słowy dla dowolnych wielomianów $w, v \in K[\lambda]$ oraz endomorfizmu ϕ mamy równość:

$$w(\phi) \circ v(\phi) = v(\phi) \circ w(\phi) = vw(\phi).$$

Oczywiście nie zawsze te endomorfizmy przemienne muszą być diagonalizowalne (to zależy od ϕ). Ciekawym przykładem rodziny przemiennych macierzy diagonalizowalnych są tak zwane macierze cykliczne.

Definicja 6.3: Macierze cykliczne rozmiaru n

Dla dowolnych $a_0, \ldots, a_{n-1} \in K$ macierz rozmiaru n o wyrazach z K postaci

$$C(a_0, \ldots, a_{n-1}) = \begin{bmatrix} a_0 & a_1 & a_2 & \ldots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \ldots & a_{n-3} & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \ldots & a_{n-4} & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_2 & a_2 & a_4 & \ldots & a_0 & a_1 \\ a_1 & a_2 & a_3 & \ldots & a_{n-1} & a_0 \end{bmatrix}.$$

nazywamy macierzą cykliczną rozmiaru n nad K. Zbiór takich macierzy oznaczamy jako $Cykl_n(K)$.

Zachęcam do sprawdzenia co robi macierz cykliczna z wektorem o współrzędnych będących kolejnymi pierwiastkami stopnia n z 1. Jak się okazuje istnieje taka macierz $P \in M_n(K)$, że dla dowolnych $a_0, \ldots, a_{n-1} \in K$ mamy $C(a_0, \ldots, a_{n-1}) = w(P)$, dla pewnego wielomianu $w \in K[\lambda]$. Wnioskujemy stąd, że $Cykl_n(K)$ to podprzestrzeń liniowa w $M_n(K)$, że iloczyn macierzy z $Cykl_n(K)$ jest w $Cykl_n(K)$ oraz, że mnożenie macierzy w $Cykl_n(K)$ jest przemienne. Jak to pokazać?

Rzeczywiście, rozważmy endomorfizm $\phi \in \mathrm{End}(K^n)$ zadany na bazie standardowej wzorem $\phi(\epsilon_i) = \epsilon_{n-1}$, przy czym $\phi(\epsilon_1) = \epsilon_n$. Niech P będzie macierzą tego endomorfizmu w bazie standardowej:

$$P = M(\phi)^{st}_{st} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Zachęcam Czytelnika do sprawdzenia jak zachowują się endomorfizmy ϕ^i i do wywnioskowania, że dla $f(x) = a_0 + a_1 x + \ldots + a_{n-1}x^{n-1} \in K[x]$ mamy $f(P) = C(a_0, \ldots, a_{n-1})$. Okazuje się, że macierze cykliczne są wspólnie diagonalizowalne nad $\mathbb{C}$, a ich wartościami własnymi są pierwiastki stopnia n z 1.

Dowód Twierdzenia 6.3 zaczniemy od następującej obserwacji.

Obserwacja 6.3

Jeśli endomorfizm $\phi : V \to V$ jest diagonalizowalny i W jest podprzestrzenią ϕ-niezmienniczą przestrzeni V, to endomorfizm $\phi|_W : W \to W$ jest również diagonalizowalny.

Dowód. Niech $\lambda_1, \ldots, \lambda_r$ będą parami różnymi wartościami własnymi ϕ. Wynika stąd, że V jest sumą prostą podprzestrzeni własnych $V_{(\lambda_i)}$, zatem dla każdego $w \in W$ możemy zapisać

$$w = v_1 + \ldots + v_r, \text{ gdzie } v_i \in V_{(\lambda_i)}. \qquad (\dagger)$$

Pokazujemy, że $v_i \in W$, co będzie oznaczało, że $W = W_1 \oplus \ldots \oplus W_r$, gdzie $W_i = V_{(\lambda_i)} \cap W$. Oczywiście dla $i \neq j$ mamy $W_i \cap W_j = \{0\}$.

Weźmy najmniejsze r takie, że dla pewnego w w rozkładzie $(\dagger)$ pewne v_i nie należy do W. Mamy

$$\phi(w) - \lambda_1 w = (\lambda_2 - \lambda_1)v_2 + \ldots + (\lambda_k - \lambda_1)v_k \in W.$$

Z minimalności r mamy $(\lambda_i - \lambda_1)v_i \in W$, dla $1 < i < k$. Ale $\lambda_i - \lambda_1 \neq 0$, czyli $v_2, \ldots, v_r \in W$. Na mocy $(\dagger)$ także $v_1 \in W$. Dostaliśmy zatem sprzeczność z wyborem w. Stąd W jest sumą prostą W_i, czyli $\phi|_W$ jest diagonalizowalny. □

Przechodzimy do dowodu implikacji (1) ⇒ (2). Zaczniemy od sytuacji, gdy $T = \{1, 2, \ldots, r\}$ jest zbiorem skończonym. Stosujemy indukcję po r. Dla $r = 1$ teza jest jasna. Załóżmy więc, że $r \geqslant 2$.

Skoro ϕ_r jest diagonalizowalny, to V jest sumą prostą podprzestrzeni własnych $V_{(\lambda)}$ względem ϕ_r, gdzie λ to wartość własna ϕ_r.Dla $i \neq r$ mamy $\phi_i \circ \phi_r = \phi_r \circ \phi_i$, zatem jeśli $v \in V_{(\lambda)}$, to

$$\phi_r(\phi_i(v)) = \phi_i(\phi_r(v)) = \phi_i(\lambda v) = \lambda \cdot \phi_i(v).$$

Zatem $V_{(\lambda)}$ jest ϕ_i-niezmiennicza ($v \in V_{(\lambda)} \Rightarrow \phi_i(v) \in V_{(\lambda)}$). Oczywiście endomorfizmy $\phi_1|_{V_{(\lambda)}}, \ldots, \phi_{r-1}|_{V_{(\lambda)}}$ są przemienne. Zatem z powyższej Obserwacji pomocniczej każde z tych przekształceń jest diagonalizowalne. Z założenia indukcyjnego endomorfizmy te są wspólnie diagonalizowalne przez pewną bazę $\mathcal{A}_\lambda$ przestrzeni $V_{(\lambda)}$ (czyli $\phi_r|_{V_{(\lambda)}}$ też się w niej diagonalizuje). Skoro $V = V_{(\lambda_1)} \oplus \ldots \oplus V_{(\lambda_r)}$, gdzie λ_i to wartości własne ϕ_i, to biorąc bazę $(\mathcal{A}_{(\lambda_1)}, \ldots, \mathcal{A}_{(\lambda_r)})$ przestrzeni V dostajemy wspólną bazę wektorów własnych dla wszystkich ϕ_i, dla $i = 1, \ldots, r$. To kończy krok indukcyjny i dowód dla zbiorów skończonych.

Przypadek ogólny jest teraz oczywisty. niech $U = \mathrm{lin}(\{\phi_t \in \mathrm{End}(V), t \in T\}) \subseteq \mathrm{End}(V)$. Skoro $\mathrm{End}(V)$ jest skończenie wymiarowa, to U też. A zatem wspólna diagonalizowalność rodziny U zależy jedynie od tego, czy wspólnie diagonalizowalne są jej elementy bazowe, których jest skończenie wiele.

Dowód implikacji (2) ⇒ (1). Załóżmy, że dla $A, B \in M_n(K)$ istnieje macierz odwracalna P, że

$$P^{-1}D_1P = A, \quad P^{-1}D_2P = B,$$

gdzie D_1, D_2 są diagonalne. Jest jasne, że $D_1D_2 = D_2D_1$, zatem

$$(\underbrace{P^{-1}D_1P}_{A})(\underbrace{P^{-1}D_2P}_{B}) = P^{-1}D_1D_2P = P^{-1}D_2D_1P = (\underbrace{P^{-1}D_2P}_{B})(\underbrace{P^{-1}D_1P}_{A}).$$

Uwaga. Bardzo ważny w algebrze jest również problem wspólnej triangularyzowalności. Można pokazać, że endomorfizmy triangularyzowalne są wspólnie triangularyzowalne, gdy są przemienne, ale odwrotna implikacja nie zachodzi! Wspólna triangularyzowalność zachodzi w wielu innych sytuacjach i prowadzi do ważnych twierdzeń (patrz H. Radjavi, P. Rosenthal, *Simultaneous Triangularization*, Springer, 2014).

Obserwacja 6.4

Niech $A_1, A_2, \ldots, A_k$ będą przemiennymi macierzami rozmiaru $n \times n$ nad ciałem algebraicznie domkniętym F. Wówczas $A_1, A_2, \ldots, A_k$ są jednocześnie triangularyzowalne, tzn. istnieje macierz odwracalna $C \in M_n(F)$ taka, że $C^{-1}AC$ jest górnotrójkątna, dla $i = 1, 2, \ldots, k$

Powyższy fakt jest dość trudnym ćwiczeniem. Aby jednak przekonać Czytelnika, że zagadnienie to może być przydatne, przywołamy istotne zadanie związane z tak zwanymi równaniami macierzowym Sylvestera, istotnymi w wielu zastosowaniach. Wykorzystamy w jego rozwiązaniu powyższy wynik.

Zadanie 6.1. *Niech F będzie ciałem algebraicznie domkniętym oraz niech $A \in M_n(F)$, $B \in M_m(F)$. Rozważmy endomorfizmy T_A, T_B przestrzeni liniowej $V = M_{n\times m}(F)$ dane wzorami:*

$$T_A(X) = AX, \quad T_B(X) = XB, \qquad \text{dla każdych } X \in V.$$

- *Pokaż, że T_A, T_B są jednocześnie triangularyzowalne.*
- *Pokaż, że jeśli A, B nie mają wspólnej wartości własnej, to równanie $AX - XB = C$ ma dokładnie jedno rozwiązanie $X \in M_{n\times m}(F)$.*

Zanim przedstawimy dowód powiedzmy nieco o równaniu macierzowym $AX - XB = C$. Do równań macierzowych można podchodzić w języku endomorfizmów i wartości własnych. Jednoznaczność rozwiązania prostego równania postaci $AX = Y$ lub $XB = Y$ wymaga, aby A (odpowiednio B) było macierzą odwracalną, a więc by jej wartości własne nie były zerowe. Odpowiada to intuicji ze świata równań: równanie $ax = y$ ma rozwiązanie tylko, gdy $a \neq 0$ (nad ciałem). Analogia ta przenosi się na równania $AX - XB = Y$. Gdybyśmy mieli do rozwiązania równanie $ax - xb = y$, wówczas jest ono jednoznaczne gdy $a - b \neq 0$. Podobnie, gdyby macierze A, B były diagonalne, wówczas równanie to ma jednoznaczne rozwiązanie $X = [x_{ij}]$ postaci $x_{ij} = y_{ij}/(\lambda_i - \mu_j)$, zakładając, że $\lambda_i - \mu_j \neq 0$, dla wszystkich $i, j \neq 0$, gdzie $Y = [y_{ij}]$, zaś na diagonalach A, B są odpowiednio λ_i oraz μ_j. Wszystko to można przetłumaczyć też tak: równanie $AX - XB$ mają jednoznaczne rozwiązanie, gdy A, B nie mają wspólnej wartości własnej. Cud polega na tym, że A oraz B wcale nie muszą być diagonalizowalne.

Rozwiązanie.[2] Najpierw udowodnimy, że T_A, T_B są jednocześnie triangularyzowalne. Z Obserwacji 6.5 wynika, że wystarczy pokazać, iż są one przemienne (jesteśmy nad ciałem algebraicznie domkniętym). Dla każdego $X \in M_{n\times m}(F)$ mamy jednak

$$T_A(T_B(X)) = T_A(XB) = AXB = T_B(AX) = T_B(T_A(X)),$$

bo mnożenie macierzy jest łączne, a to kończy dowód jednoczesnej triangularyzowalności.

Niech $T \in \mathrm{End}(V)$ będzie dane wzorem $T = T_A - T_B$, zaś $\mathcal{A} = \{\alpha_1, ..., \alpha_{nm}\}$ będzie taką bazą przestrzeni V, w której zarówno macierz T_A, jak i macierz T_B mają postać górnotrójkątną (z poprzedniego akapitu wiemy, że takowa istnieje). Niech

$$M(T_A)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} a_1 & * & \ldots & * \\ 0 & a_2 & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nm} \end{bmatrix}, \quad M(T_B)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} b_1 & * & \ldots & * \\ 0 & b_2 & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & b_{nm} \end{bmatrix},$$

gdzie $a_i, b_i \in F$, zaś $*$ oznaczają dowolne elementy F. Wówczas

$$M(T)^{\mathcal{A}}_{\mathcal{A}} = M(T_A - T_B)^{\mathcal{A}}_{\mathcal{A}} = M(T_A)^{\mathcal{A}}_{\mathcal{A}} - M(T_B)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} a_1 - b_1 & * & \ldots & * \\ 0 & a_2 - b_2 & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nm} - b_{nm} \end{bmatrix}.$$

[2] Autorem tego ładnego rozwiązania jest Mieszko Grodzicki.

Pokażemy, że dla $i = 1, 2, ..., nm$ mamy $a_i - b_i \neq 0$. Załóżmy w tym celu nie wprost, że $a_i = b_i$ dla pewnego $1 \leqslant i \leqslant nm$. Oznaczmy $c = a_i = b_i$. Zauważmy, że

$$w_{T_A}(\lambda) = (a_1 - \lambda)(a_2 - \lambda)...(a_{nm} - \lambda),$$
$$w_{T_B}(\lambda) = (b_1 - \lambda)(b_2 - \lambda)...(b_{nm} - \lambda),$$

zatem c jest wartością własną zarówno T_A, jak i T_B. Istnieją więc niezerowe macierze $K, L \in M_{n\times m}(F)$ takie, że

$$T_A(K) = cK, \quad T_B(L) = cL,$$
$$AK = cK, \quad LB = cL.$$

W szczególności, jeśli przez k, l oznaczymy odpowiednio dowolną niezerową kolumnę K i dowolny niezerowy wiersz L (istnieją, skoro K, L są niezerowe), to

$$Ak = ck, \quad lB = cl.$$

Z pierwszej równości wynika, że c jest wartością własną macierzy A. Z drugiej dostajemy

$$B^T l^T = (lB)^T = (cl)^T = c \cdot l^T,$$

czyli c jest wartością własną macierzy B^T. Ale to oznacza, że jest też wartością własną macierzy B, ponieważ $w_B(\lambda) = \det(B - \lambda I) = \det(B - \lambda I)^T = \det(B^T - \lambda I) = w_{B^T}(\lambda)$. To jest jednak wbrew założeniu, że A i B nie mają wspólnej wartości własnej. Otrzymana sprzeczność kończy dowód tego, że $a_i - b_i \neq 0$ dla każdego $i = 1, 2, ..., nm$.

Wiemy więc, że macierz $M(T)_{\mathcal{A}}^{\mathcal{A}}$ jest górnotrójkątna, a na przekątnej ma wszystkie wyrazy różne od 0. To oznacza, że $\mathrm{r}\left(M(T)_{\mathcal{A}}^{\mathcal{A}}\right) = nm$, czyli macierz ta jest odwracalna, a więc T jest izomorfizmem przestrzeni V. Wobec tego dla dowolnego $C \in V$ istnieje dokładnie jedno takie $X \in V$, że $T(X) = C$, czyli dokładnie jedno rozwiązanie równania $AX - XB = C$. ■

Równanie Sylvestera stosować można choćby do badania podobieństwa macierzy blokowych. Rozważmy macierze blokowe rozmiaru $2n \times 2n$ o blokach $n \times n$ postaci:

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}, \quad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

Kiedy takie dwie macierze blokowe są podobne? Zauważmy, że dla dowolnej macierzy $X \in M_n(K)$ mamy:

$$\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}.$$

A zatem dwie macierze blokowe wyżej będą podobne. jeśli znajdziemy X spełniające:

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} \cdot \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} A & C \\ 0 & B \end{bmatrix}.$$

Wymnażając te macierze dostajemy cztery równości, z których tylko jedna nie jest automatycznie spełniona: $AX + C = XB$, czyli $AX - XB = -C$. A zatem zadanie wyżej daje nam wniosek.

Wniosek 6.1

Jeśli $A, B \in M_n(K)$, gdzie K jest ciałem algebraicznie domkniętym oraz jeśli A, B nie mają wspólnych wartości własnych, to dla każdej macierzy $C \in M_n(K)$ poniższe macierze są podobne

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}, \quad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}.$$

Można udowodnić, że powyższy warunek jest konieczny i wystarczający, nawet po opuszczeniu założenia, że ciało K jest algebraicznie domknięte (W. E. Roth, *The equations $AX - YB = C$ and $AX - XB = C$ in matrices*, Proc. Amer. Math. Soc. 3 (1952) 392–396).

6.2 Dodatek. Rozkład na podprzestrzenie cykliczne. Postać kanoniczna Frobeniusa

Twierdzenie Jordana mówi, że endomorfizm triangularyzowalny $\phi \in \mathrm{End}(V)$ wymusza rozkład przestrzeni skończenie wymiarowej V na sumę prostą podprzestrzeni cyklicznych. Co więcej — endomorfizm ten obcięty do podprzestrzeni cyklicznej ma postać Jordana. Pokażemy, że istotne jest tu wymaganie istnienia bazy Jordana. Sam rozkład na podprzestrzenie cykliczne jest bardziej elementarny (i może posłużyć do dowodu tw. Jordana). Poniższy tekst jest przygotowany na podstawie wykładu dr. Andrzeja Strojnowskiego[3]. W kluczowym kroku 3 skorzystamy z twierdzenia o rozkładzie prymarnym.

Twierdzenie 6.5

Niech ϕ będzie endomorfizmem skończenie wymiarowej przestrzeni liniowej V. Jeśli wielomian charakterystyczny $w_\phi(\lambda)$ rozkłada się na czynniki liniowe, to V jest sumą prostą przestrzeni cyklicznych.

Dowód. Rozumowanie dzieli się na trzy kroki, w zależności od postaci wielomianu minimalnego $m_\phi(\lambda)$.

- Krok 1, gdy $m_\phi(\lambda) = \lambda^t$. Wówczas dla każdego $v \in V$ mamy $\phi^t(v) = 0$ oraz istnieje wektor $\alpha \in V$, że $\phi^{t-1}(\alpha) \neq 0$. Niech B będzie maksymalną (ze względu na inkluzję) podprzestrzenią niezmienniczą V rozłączną z podprzestrzenią cykliczną V_α. Wykażemy, że $V = V_\alpha \oplus B$.

  Niech $W = V_\alpha \oplus B$. Załóżmy nie wprost, że istnieje niezerowy wektor $\delta_1 \in V \setminus W$. Rozważmy maksymalne i takie, że $\delta_2 = \phi^i(\delta_1) \notin W$, tzn. $\delta_2 \notin W$ oraz $\phi(\delta_2) \in W$. Wówczas istnieją takie $a_0, a_1, a_2, \ldots, a_{t-1}$, że

  $$\phi(\delta_2) = \overline{\alpha} + \beta, \qquad \text{gdzie} \qquad \overline{\alpha} = \sum_{i=0}^{t-1} a_i \phi^i(\alpha) \in V_\alpha \text{ oraz } \beta \in B.$$

  Stąd:

  $$0 = \phi^t(\delta_2) = \phi^{t-1}(\overline{\alpha}) + \phi^{t-1}(\beta).$$

  Skoro B jest ϕ-niezmiennicza, to $\phi^{t-1}(\beta) \in B$. Skoro zaś $B \cap V_\alpha = \{0\}$, to równość wyżej oznacza $0 = \phi^{t-1}(\overline{\alpha})$ oraz $0 = \phi^{t-1}(\beta)$. Biorąc pierwszą z tych równości mamy:

  $$0 = \phi^{t-1}(\overline{\alpha}) = \sum_{i=0}^{t-1} a_i \phi^{t+i-1}(\alpha) = a_0 \phi^{t-1}(\alpha),$$

  czyli $\overline{\alpha} = \phi(\alpha_1)$, gdzie

  $$\alpha_1 = \sum_{i=1}^{t-1} a_i \phi^{i-1}(\alpha) \in V_\alpha.$$

  Niech $\delta_3 = \delta_2 - \alpha_1$. Wówczas

  $$\phi(\delta_3) = \phi(\delta_2) - \phi(\alpha_1) = \overline{\alpha} + \beta - \overline{\alpha} \in B. \qquad (\dagger)$$

  Skoro $\delta_2 \notin W$ oraz $\alpha_1 \in W$, to $\delta_3 \notin W$. W szczególności przestrzeń

  $$\overline{B} = B + \mathrm{lin}\{\delta_3\} \supsetneq B$$

  jest również ϕ-niezmiennicza (patrz $(\dagger)$). Do uzyskania sprzeczności z wyborem B jako maksymalnej przestrzeni ϕ-niezmienniczej przecinającej się zerowo z V_α należy pokazać, $V_\alpha \cap \overline{B} = \{0\}$.

  Niech więc $\alpha_3 \in V_\alpha \cap \overline{B}$. Wówczas $\alpha_3 = \beta_1 + r\delta_3$, dla pewnego $\beta_1 \in B$ oraz $r \in K$. Stąd jednak $r\gamma_3 = \alpha_3 - \beta_1 \in W$. Jednak $\gamma_3 \notin W$, więc $r = 0$ oraz $\alpha_3 = \beta_1 \in V_\alpha \cap B = \{0\}$.

  Uzyskana sprzeczność oznacza, że $V = V_\alpha \oplus B$. Zatem rozumowanie w tym punkcie to indukcja ze względu na wymiar V. Ograniczamy ϕ do B i korzystamy z założenia indukcyjnego lub po prostu w B w analogiczny sposób dobieramy podprzestrzenie cykliczne, które są składnikami prostymi, aż do wyczerpania wymiaru. Uzyskaliśmy więc Krok 1. W szczególności udowodniliśmy rozkład na sumą prostą podprzestrzeni cyklicznych dla endomorfizmu nilpotentnego.

[3] https://www.mimuw.edu.pl/~stroa/Gal_Dodatki/Twierdzenie%20o%20przestrzeniach%20cyklicznych.pdf

- Krok 2, gdy wielomian minimalny $m_\phi(\lambda)$ jest postaci $(\lambda-a)^t$, dla pewnego $a \in K$. Wówczas należy rozpatrzeć endomorfizm $\psi = \phi - a\,\mathrm{id}$. Wówczas $m_\psi(\lambda) = \lambda^t$, więc V jest sumą prostą podprzestrzeni cyklicznych względem ψ. Skoro jednak $\phi = \psi + a\,\mathrm{id}$, więc każda podprzestrzeń cykliczna względem ψ jest cykliczna względem ψ. Istotnie, jeśli X jest przestrzenią ψ-cykliczną rozpiętą na wektorze α, zaś Y jest przestrzenią ϕ-cykliczną rozpiętą na α, to czyli

  $$\phi^i(\alpha) = (\psi + a\,\mathrm{id})^i(\alpha) \in X, \quad \psi^i(\alpha) = (\phi - a\,\mathrm{id})^i(\alpha) \in Y$$

  (wykonujemy wzór dwumianowy). Układ rozpinający Y zawarty jest w X i odwrotnie, co oznacza, że $X = Y$. Otrzymaliśmy zatem rozkład V na sumę prostą podprzestrzeni cyklicznych.

- Krok 3, czyli przypadek ogólny, jest indukcją względem liczby nierozkładalnych dzielników wielomianu minimalnego.

  – Jeśli pojedynczy dzielnik nierozkładalny wielomianu minimalnego ma postać $(\lambda-a)^t$, to dowód wynika z poprzedniego kroku.

  – Niech $m_\phi(\lambda) = (a-\lambda)^m \cdot g(x)$ będzie rozkładu wielomianu minimalnego na iloczyn wielomianów względnie pierwszych. Zgodnie Wnioskiem 5.1 mamy

  $$V = \ker(\phi - a\,\mathrm{id})^t \oplus \ker g(\phi).$$

  Wielomian minimalny przestrzeni $\ker(\phi - a\,\mathrm{id})^t$ jest dzielnikiem $(a-\lambda)^m$, wiec przestrzeń ta rozpada się na sumę prostą podprzestrzeni cyklicznych na mocy kroku 2. Rozkład drugiego czynnika wynika z założenia indukcyjnego.

□

Czytelnik może być zaskoczony, że dowodzimy powyższy rezultat jedynie dla endomorfizmów triangularyzowalnych. Celem było pokazanie o ile bardziej podstawowe jest spojrzenie na rozkład na podprzestrzenie cykliczne niż na samo twierdzenie Jordana, które zakłada, że podprzestrzenie te mają szczególną postać. Można w powyższym języku, przy pewnych dodatkowych narzędziach, przeprowadzić dowód ogólny. Zrobimy to jednak w nieco inny sposób, wracając do macierzy wielomianowych, poruszonych już wcześniej. W ten sposób wszystkie wątki połączymy w jeden.

Naszym celem będzie naszkicowanie konstrukcji formy kanonicznej Frobeniusa dla macierzy nad dowolnym ciałem (rational form). Wykorzystamy technologię macierzy wielomianowych z Dodatku 3.2.

Uwaga odnośnie notacji. Macierz blokowo diagonalną o klatkach $A_1, \ldots, A_r$ zapisujemy od tego miejsca jako jako

$$\mathrm{diag}(A_1, \ldots, A_r).$$

Dla macierzy wielomianowej $A(\lambda) \in M_n(K[\lambda])$ o postaci równoważnej mającej formę kanoniczną Smitha (Twierdzenie 3.10) o wielomianach $p_1(\lambda), \ldots, p_n(\lambda) \in K[\lambda]$ na przekątnej będziemy pisać pokrótce:

$$A(\lambda) \sim \mathrm{diag}(p_1(\lambda), \ldots, p_n(\lambda)).$$

Większość podawanych dalej wyników to proste ćwiczenia, wynikające z twierdzeń uzyskanych w Dodatku 3.2. Pamiętajmy, że relacja $\sim$ to relacja mówiąca, że macierze wielomianowe są równoważne, przez odpowiednie dopuszczalne operacje na wierszach i kolumnach macierzy wielomianowej.

Obserwacja 6.5: Ćwiczenie

Dla dowolnego wielomianu $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$ mamy

$$\lambda I - C(f(\lambda)) \sim \mathrm{diag}(\underbrace{1, \ldots, 1}_{n-1}, f(\lambda))$$

Wniosek 6.2

Dla dowolnej macierzy $A \in M_n(K)$ takiej, że $\lambda I - A \sim \operatorname{diag}(\underbrace{1, \ldots, 1}_{n-r}, f_1(\lambda), \ldots, f_r(\lambda))$ oraz macierzy blokowo diagonalnej $C = \operatorname{diag}(C(f_1(\lambda)), \ldots, C(f_r(\lambda)))$ mamy

$$\lambda I - C \sim \operatorname{diag}(\underbrace{1, \ldots, 1}_{n-r}, f_1(\lambda), \ldots, f_r(\lambda)).$$

Na mocy Wniosku 3.4 macierze A oraz C są podobne.

Dowód. Zauważmy, że $\lambda I - C = \operatorname{diag}(\lambda I - C(f_1(\lambda)), \ldots, \lambda I - C(f_r(\lambda)))$. Z poprzedniego ćwiczenia sprowadzamy każdy blok $\lambda I - C(f_1(\lambda))$ do postaci Smitha, czyli:

$$\lambda I - C \sim \operatorname{diag}(\operatorname{diag}(1, \ldots, 1, f_1(\lambda)), \ldots, \operatorname{diag}(1, \ldots, 1, f_r(\lambda))).$$

To oznacza, że porządkując inaczej diagonalę po prawej mamy $\lambda I - C \sim \operatorname{diag}(\underbrace{1, \ldots, 1}_{n-r}, f_1(\lambda), \ldots, f_r(\lambda))$.

Zatem macierze wielomianowe $\lambda I - C$ oraz $\lambda I - A$ są równoważne. □

Wniosek 6.3

Każda macierz $A \in M_n(K)$ jest podobna do macierzy postaci

$$\operatorname{diag}(C(f_1(\lambda)), \ldots, C(f_r(\lambda))),$$

gdzie $f_1(\lambda), \ldots, f_r(\lambda)$ są czynnikami niezmienniczymi $\lambda I - A$.

Uzyskaliśmy zatem rozkład przestrzeni skończenie wymiarowej na sumę prostą podprzestrzeni cyklicznych. Biorąc bowiem $\phi \in \operatorname{End}(K^n)$ dany wzorem $M(\phi)^{st}_{st} = A$ widzimy, że w pewnej bazie ϕ ma macierz blokowo diagonalną, gdzie po ograniczeniu się do odpowiedniego bloku mamy podprzestrzeń cykliczną (taka jest postać macierzy $C(f_i)$ o wielomianie minimalnym f_i).

Twierdzenie 6.6

Niech

$$f(\lambda) = p_1(\lambda)^{e_1} \cdot \ldots \cdot p_r(\lambda)^{e_r} \in K[\lambda]$$

będzie (jednoznacznym) rozkładem $f(\lambda)$ na czynniki nierozkładalne, przy czym $p_1, \ldots, p_r$ są unormowane i parami różne. Wówczas podobne nad K są macierze:

- $C(f(\lambda))$,
- $C = \operatorname{diag}(C(p_1(\lambda)^{e_1}), \ldots, C(p_r(\lambda)^{e_r}))$.

Dowód. Zauważmy, że $\lambda I - C = \operatorname{diag}(\lambda I - C(p_1(\lambda)^{e_1}), \ldots, \lambda I - C(p_1(\lambda)^{e_1}))$. Z ćwiczenia wyżej sprowadzamy każdy blok $\lambda I - C(f_1(\lambda))$ do postaci Smitha, czyli:

$$\lambda I - C \sim \operatorname{diag}(\operatorname{diag}(1, \ldots, 1, p_1(\lambda)^{e_1}), \ldots, \operatorname{diag}(1, \ldots, 1, p_r(\lambda)^{e_r})).$$

Liczymy dzielniki wyznacznikowe $\lambda I - C$. Ostatni dzielnik, powiedzmy d_n to oczywiście $w_C(\lambda)$, czyli $f(\lambda)$. Dzielnik wyznacznikowy $d_{n-1}(\lambda I - C)$ to NWD względnie pierwszych wielomianów $f(\lambda)/p_i(\lambda)^{e_i}$, czyli 1. A zatem

$$\lambda I - C \sim \operatorname{diag}(\underbrace{1, \ldots, 1}_{n-1}, f(\lambda)).$$

Ale również $\lambda I - C(f(\lambda)) \sim \operatorname{diag}(\underbrace{1, \ldots, 1}_{n-1}, f(\lambda))$, co kończy dowód. □

Definicja 6.4: Dzielnik elementarny macierzy/endomorfizmu

Niech $A \in M_n(K)$ oraz niech czynniki niezmiennicze macierzy wielomianowej $\lambda I - A \in M_n(K[\lambda])$ będą postaci:

$$\begin{aligned} c_1(\lambda) &= p_1(\lambda)^{e_{11}} \cdot \ldots \cdot p_t(\lambda)^{e_{1t}} \\ &\vdots \\ c_r(\lambda) &= p_1(\lambda)^{e_{r1}} \cdot \ldots \cdot p_t(\lambda)^{e_{rt}}, \end{aligned}$$

gdzie $p_s(\lambda)$ są parami różne, unormowane i nierozkładalne nad K, zaś $e_{ij} \in \mathbb{N}$. Wielomiany $p_s(\lambda)^{e_{ij}}$, gdzie $e_{ij} \neq 0$, nazywamy DZIELNIKAMI ELEMENTARNYMI macierzy A.

Wniosek 6.4

Macierze $A, B \in M_n(K)$ są podobne nad ciałem K wtedy i tylko wtedy, gdy mają te same dzielniki elementarne.

Wniosek 6.5: Wniosek – twierdzenie o rozkładzie na podprzestrzenie cykliczne

Jeśli $d_1(\lambda), \ldots, d_s(\lambda) \in K[\lambda]$ są dzielnikami elementarnymi macierzy $A \in M_n(K)$, to A jest podobna nad K do macierzy $\operatorname{diag}(C(d_1(\lambda)), \ldots, C(d_r(\lambda))) \in M_n(K)$.

Uwaga. Postać dzielników elementarnych zależy od ciała K. Przykład: jeśli $A \in M_{12}(\mathbb{Q})$ ma czynniki niezmiennicze:

$$\begin{aligned} c_1(\lambda) &= (\lambda^2+1)(\lambda-1), \\ c_2(\lambda) &= (\lambda^2+1)^2(\lambda^2-2)^2(\lambda-1) \end{aligned}$$

to dzielnikami elementarnymi A są:

$$\lambda^2+1, \quad \lambda-1, \quad (\lambda^2+1)^2, \quad (\lambda^2-2)^2, \quad \lambda-1.$$

A zatem A jest podobna nad $\mathbb{Q}$ do

$$\operatorname{diag}\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, 1, \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & -4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}, 1 \right).$$

Natomiast $A \in M_{12}(\mathbb{R})$ ma czynniki niezmiennicze:

$$\begin{aligned} c_1(\lambda) &= (\lambda^2+1)(\lambda-1), \\ c_2(\lambda) &= (\lambda^2+1)^2(\lambda-\sqrt{2})^2(\lambda+\sqrt{2})^2(\lambda-1) \end{aligned}$$

to dzielnikami elementarnymi A są:

$$\lambda^2+1, \quad \lambda-1, \quad (\lambda^2+1)^2, \quad (\lambda-\sqrt{2})^2, \quad (\lambda+\sqrt{2})^2, \quad \lambda-1.$$

A zatem A jest podobna nad $\mathbb{R}$ do

$$\operatorname{diag}\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, 1, \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & 2\sqrt{2} \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & -2\sqrt{2} \end{bmatrix}, 1 \right).$$

Wreszcie macierz $A \in M_{12}(\mathbb{C})$ ma czynniki niezmiennicze:

$$c_1(\lambda) = (\lambda - i)(\lambda + i)(\lambda - 1),$$
$$c_2(\lambda) = (\lambda - i)^2(\lambda + i)^2(\lambda - \sqrt{2})^2(\lambda + \sqrt{2})^2(\lambda - 1)$$

to dzielnikami elementarnymi A są:

$$\lambda - i, \quad \lambda + i, \quad \lambda - 1, \quad (\lambda - i)^2, \quad (\lambda + i)^2, \quad (\lambda - \sqrt{2})^2, \quad (\lambda + \sqrt{2})^2, \quad \lambda - 1.$$

A zatem A jest podobna nad $\mathbb{C}$ do

$$\operatorname{diag}\left(-i, i, 1, \begin{bmatrix} 0 & 1 \\ 1 & 2i \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & -2i \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & 2\sqrt{2} \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 1 & -2\sqrt{2} \end{bmatrix}, 1\right).$$

Teraz przejdziemy do formy wymiernej (Frobeniusa), której szczególnym przypadkiem jest postać kanoniczna Jordana.

Definicja 6.5: Macierz hipertowarzysząca wielomianu

Niech $p(\lambda) \in K[\lambda]$ będzie unormowany stopnia m i niech $N = (n_{ij}) \in M_{m\times m}(K)$ będzie taka, że $n_{m1} = 1$, zaś pozostałe wyrazy są zerowe.

Przez MACIERZ HIPERTOWARZYSZĄCĄ WIELOMIANU $p(\lambda)^r$ rozumiemy macierz blokową $H(p(\lambda)^r) \in M_{mr}(K)$ postaci:

$$H(p(\lambda)^r) = \begin{bmatrix} C(p(\lambda))^T & N & & \dots & & 0 \\ 0 & C(p(\lambda))^T & N & \dots & 0 & 0 \\ 0 & 0 & C(p(\lambda))^T & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & C(p(\lambda))^T & N \\ 0 & 0 & 0 & \dots & 0 & C(p(\lambda))^T \end{bmatrix}.$$

Przykład 1, dla $w_1(\lambda) = (\lambda^2 + 1)^3$.

$$H(w_1(\lambda)) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

Przykład 2, dla $w_2(\lambda) = (\lambda^3 + 5\lambda^2 - 6)^2$:

$$H(w_2(\lambda)) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 6 & 0 & -5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 6 & 0 & -5 \end{bmatrix}$$

Przykład 3, dla $w_3(\lambda) = (\lambda - 3)^4$:

$$H(w_3(\lambda)) = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Czyli macierz hipertowarzysząca $(\lambda - a)^m$ to klatka Jordana.

Obserwacja 6.6: Ćwiczenie

Jeśli $mr \neq 1$, to ponad diagonalą macierzy $H(p(\lambda)^r) \in M_{mr}(K)$ są same jedynki.

Obserwacja 6.7

Niech $p(\lambda) \in K[\lambda]$ będzie wielomianem unormowanym stopnia m. Wówczas dla każdego $r \geqslant 1$ macierze $C = C(p(\lambda)^r)$ oraz $H = H(p(\lambda)^r)$ są podobne nad K.

Dowód. Oczywiście $\det(\lambda I - H) = \det(\lambda I - C) = p(\lambda)^r$. Zatem dzielnik wyznacznikowy d_{rm} macierzy wielomianowych $\lambda I - H$ oraz $\lambda I - C$ jest identyczny. Zauważmy, że jeśli z macierzy H usuniemy pierwszy wiersz i ostatnią kolumnę, to na mocy ćwiczenia wyżej dostajemy macierz dolnotrójkątną mającą na przekątnej same jedynki. Zatem po usunięciu z $\lambda I - H$ pierwszego i ostatniego wiersza dostajemy macierz dolnotrójkątną mającą na przekątnej same -1, więc wyznacznik tego minora wynosi ± 1. Stąd

$$d_{rm-1}(\lambda I - H) = 1.$$

Zatem także $c_{rm-1}(\lambda I - H) = 1$. Zatem z twierdzenia o postaci Smitha mamy:

$$\lambda I - H \sim \operatorname{diag}(1, \ldots, 1, p(\lambda)^r).$$

Wiemy jednak, że $\lambda I - C \sim \operatorname{diag}(1, \ldots, 1, p(\lambda)^r)$, więc H i C są podobne. □

Wniosek 6.6: Twierdzenie o formie kanonicznej wymiernej (Frobeniusa)

Jeśli $d_1(\lambda), \ldots, d_r(\lambda)$ są dzielnikami elementarnymi $A \in M_n(K)$, to A jest podobna do macierzy:

$$\operatorname{diag}(H(d_1(\lambda)), \ldots, H(d_r(\lambda))).$$

Jeśli K jest ciałem algebraicznie domkniętym, to wszystkie elementarne dzielniki A są postaci $(\lambda - a)^r$, gdzie a to wartość własna A. Co więcej $H((\lambda - a)^r)$ jest klatką Jordana rozmiaru r.

Nie wspominamy już o szukaniu macierzy transformacji macierzy do jej formy wymiernej, choć takie algorytmy są oczywiście znane i opisane. Wspomnijmy jednak o choćby jednym prostym wniosku.

Definicja 6.6: Uogólniona klatka Jordana

Uogólnioną klatką Jordana nad $\mathbb{R}$ nazywamy macierz o blokach 2×2 postaci:

$$\begin{bmatrix} A & I & 0 & \ldots & 0 & 0 \\ 0 & A & I & \ldots & 0 & 0 \\ 0 & 0 & A & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & A & I \\ 0 & 0 & 0 & \ldots & 0 & A \end{bmatrix}, \quad \text{gdzie } I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \text{ dla pewnych } a, b \in \mathbb{R}.$$

Twierdzenie 6.7: Rzeczywista postać Jordana

Każda macierz $A \in M_n(\mathbb{R})$ jest podobna do pewnej macierzy blokowo-diagolnalnej B o klatkach Jordana – zwykłych lub ogólnionych.

Rozdział 7

Przestrzenie afiniczne w przestrzeniach liniowych

Tym wykładem rozpoczynamy nowy etap poznawania geometrii z punktu widzenia algebry, związany z tzw. przestrzeniami afinicznymi. Wśród wielu alternatywnych podejść do tego tematu realizujemy najpierw najbardziej elementarny, polegający na wzbogaceniu struktury przestrzeni liniowych dodatkową operacją. Przypisywać ona będzie każdej parze wektorów (v, w) nowy wektor: $w - v$, który oznaczać będziemy jako $\overrightarrow{vw}$. Wydaje się to na pierwszy rzut oka banalne, a jednak doprowadzi do nowego spojrzenia na geometrię choćby przestrzeni K^n, gdzie K jest ciałem. Struktura afiniczna pozwoli na mówienie o punktach, prostych, czy trójkątach, a także na interpretację znanych pojęć odległości czy równoległości. Zobaczymy też, że ów nowy typ struktury na przestrzeni liniowej pozwala badać nie tylko jej podprzestrzenie liniowe, ale także podzbiory będące zbiorami rozwiązań układów niejednorodnych, traktując je jako tzw. podprzestrzenie afiniczne. Gdy oswoimy się z nowym językiem na gruncie przestrzeni liniowych, poznamy również aksjomatykę przestrzeni afinicznych, obejmującą nie tylko podzbiory przestrzeni liniowych z dodatkową strukturą.[1]

Definicja 7.1: Wektor łączący

Niech V będzie przestrzenią liniową nad ciałem K. Wartość funkcji $\omega : V \times V \to V$ przypisującej parze (α, β) wektor

$$\omega(\alpha, \beta) = \beta - \alpha,$$

nazywamy WEKTOREM ŁĄCZĄCYM α z β, albo krócej WEKTOREM OD α DO β i oznaczamy jako $\overrightarrow{\alpha\beta}$.

Zauważmy, że zbiór „wektorów łączących" w przestrzeni V pokrywa się po prostu ze zbiorem elementów V. O co więc tu chodzi? Po co nam taka dodatkowa kopia V? Pierwsza motywacja jest następująca: rozróżnienie to jest wygodne dla opisu zbiorów rozwiązań układów równań niejednorodnych.

Z pierwszego semestru pamiętamy, że jeśli U jest układem niejednorodnym równań liniowych, a U' odpowiadającym mu układem jednorodnym, to zbiór rozwiązań układu U jest postaci

$$H = v + W = \{v + w \,|\, w \in W\},$$

gdzie v jest pewnym wektorem z V spełniającym U oraz W jest przestrzenią rozwiązań układu jednorodnego U'. Pamiętajmy, że zbiór H nie jest podprzestrzenią liniową, o ile $v \neq 0$. Używając języka wprowadzonego wyżej możemy powiedzieć, że zbiór rozwiązań układu U to

$$\{v + \overrightarrow{vw} \,|\, w \in H\}.$$

Po co nam taki opis?

[1] Pod względem koncepcyjnym jest to stosunkowo skomplikowany temat, a powodów należałoby zapewne szukać w niezwykle skomplikowanej historii geometrii. Choć sam język mówiący o przestrzeniach afinicznych jest bardzo elegancki, to ustępuje jak się wydaje pod względem znaczenia językowi, który poznacie Państwo na Topologii, mówiąc o przestrzeniach metrycznych. Niemal każdy podręcznik algebry liniowej podchodzi do tematu przestrzeni afinicznych nieco inaczej. Zasadniczo opieramy się – jak wcześniej – na podejściu ze skryptu wydziałowego „*Wykłady z algebry liniowej II*"

Przykład. Rozwiązaniem równania

$$x_1 + x_2 + x_3 = 3$$

w $\mathbb{R}^3$ jest np. zbiór

$$(3,0,0) + \mathrm{lin}((0,1,-1),(-1,0,1)).$$

Jest to zbiór wektorów łączących wektor $(3,0,0)$ oraz dowolny wektor z $\mathrm{lin}((0,1,-1),(-1,0,1))$. Dlaczego „na przykład"? Nie jest to bowiem jedyny opis. Rozwiązania układu U można także opisać jako zbiór

$$(1,1,1) + \mathrm{lin}((0,1,-1),(-1,0,1)).$$

Po prostu wybieramy inne rozwiązanie układu U. Oczywiście dowolne dwa rozwiązania układu U wyznaczają pewien wektor łączący należący do $\mathrm{lin}((0,1,-1),(-1,0,1))$. Prowadzi to do następującej definicji.

Definicja 7.2: Warstwa podprzestrzeni

Niech W będzie podprzestrzenią przestrzeni liniowej V nad ciałem K i niech $\alpha \in V$. Zbiór

$$\alpha + W = \{\alpha + \gamma, \gamma \in W\}$$

nazywamy warstwą podprzestrzeni W w przestrzeni V.

Obserwacja 7.1

Niech W będzie podprzestrzenią przestrzeni liniowej V i niech $\alpha, \beta \in V$. Wówczas:

(i) $\alpha + W = \beta + W \Leftrightarrow \alpha - \beta \in W$,

(ii) $(\alpha + W) \cap (\beta + W) \neq \emptyset \Leftrightarrow \alpha + W = \beta + W$.

Dowód. Dowodzimy (i). Jeśli $\alpha + W = \beta + W$, to $\alpha \in \beta + W$, więc $\alpha = \beta + \gamma$, dla pewnego $\gamma \in W$. Zatem $\alpha - \beta \in W$. Na odwrót: jeśli $\alpha - \beta = \gamma \in W$, to $\alpha = \beta + \gamma$. Wówczas dla każdego $\gamma' \in W$ mamy

$$\alpha + \gamma' = \beta + \gamma + \gamma' \in \beta + W,$$

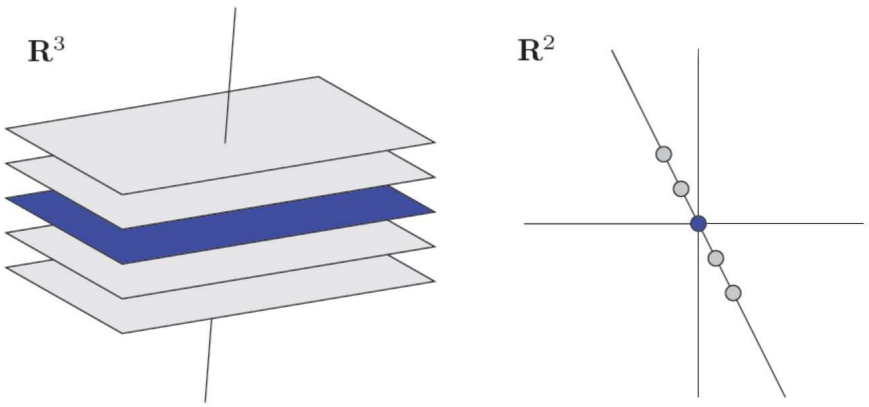
bo $\gamma + \gamma' \in W$. Stąd $\alpha + W \subseteq \beta + W$. Analogicznie dowodzimy $\beta + W \subseteq \alpha + W$.

Dowodzimy (ii). Niech $(\alpha + W) \cap (\beta + W) \neq \emptyset$. Zatem istnieje $\delta \in (\alpha + W) \cap (\beta + W)$. Stąd

$$\alpha + \gamma_1 = \delta = \beta + \gamma_2,$$

dla pewnych $\gamma_1, \gamma_2 \in W$. Wówczas $\alpha - \beta = \gamma_2 - \gamma_1 \in W$, więc na mocy (i) mamy $\alpha + W = \beta + W$. □

Wniosek z punktu (i) jest taki, że dwie warstwy $\alpha + W$, $\beta + W$ podprzestrzeni W są równe wtedy i tylko wtedy, gdy $\overrightarrow{\alpha\beta}$ należy do W.



Warstwy podprzestrzeni wymiaru 2 w $\mathbb{R}^3$ interpretować będziemy w postaci równoległych płaszczyzn. Warstwy podprzestrzeni wymiaru 0 w $\mathbb{R}^2$ interpretować będziemy jako punkty. W pokazanym przypadku punkty te należą akurat do wybranej prostej.

Obserwacja 7.2

Podzbiór $H \subseteq K^n$ jest warstwą podprzestrzeni $W \subseteq K^n$ wtedy i tylko wtedy,[a] gdy H jest zbiorem rozwiązań niesprzecznego układu równań liniowych o n niewiadomych i współczynnikach w K.

[a] To był jeden z punktów twierdzenia Kroneckera-Capellego.

Co to wszystko ma wspólnego z dodatkową strukturą wektorów łączących? Wprawdzie warstwa nie jest sama w sobie podprzestrzenią liniową, ale jest zamknięta na operację ω, którą wprowadziliśmy w V.

Definicja 7.3: Przestrzeń afiniczna, przestrzeń styczna, punkty

Niech V będzie przestrzenią liniową nad ciałem K.

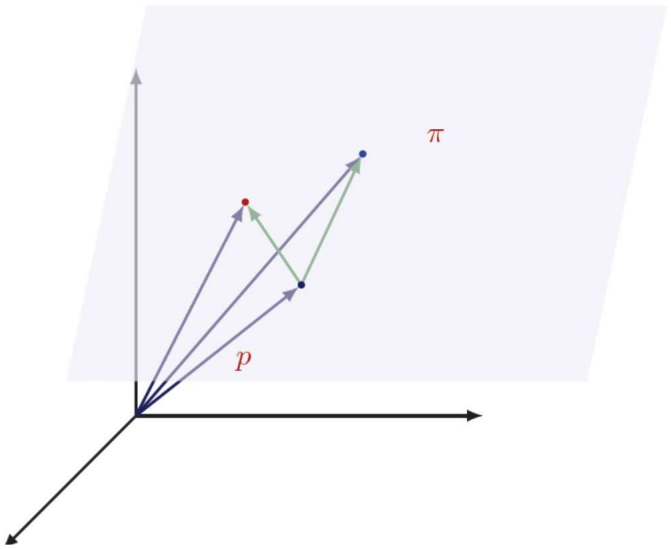
- Mówimy, że podzbiór $H \subseteq V$ jest PRZESTRZENIĄ AFINICZNĄ NAD K, jeśli H jest warstwą pewnej podprzestrzeni w V.
- Elementy przestrzeni afinicznej H nazywamy PUNKTAMI.
- Podprzestrzeń liniową W, której warstwą jest H nazywamy PRZESTRZENIĄ STYCZNĄ lub PRZESTRZENIĄ WEKTORÓW SWOBODNYCH do przestrzeni afinicznej H i oznaczamy $T(H)$.
- Mówimy, że przestrzenie afiniczne $H_1, H_2 \subseteq V$ są RÓWNOLEGŁE, jeśli $T(H_1) = T(H_2)$.

PRZYKŁADY.

- Przestrzeń liniowa V ma strukturę przestrzeni afinicznej – jest to warstwa podprzestrzeni $W = V$ postaci $0 + V$. Dla każdego $q \in V$ zbiór $\{q\}$ jest przestrzenią afiniczną jako warstwa $q + 0$.
- Zbiór $H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 - x_2 = 4, x_2 - 2x_3 = 5\}$ jest przestrzenią afiniczną. Mamy

$$T(H) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 - x_2 = 0, x_2 - 2x_3 = 0\} = \mathrm{lin}((2, 2, 1)),$$

  a więc $H = (11, 7, 1) + \mathrm{lin}((2, 2, 1))$.



Komentarz. Wybierając dowolny punkt p na płaszczyźnie π (tzn. odpowiedni wektor w $\mathbb{R}^3$) możemy traktować π jako zbiór wektorów postaci: $p + T(\pi)$. Zbiór $T(\pi)$ składa się z wektorów swobodnych, ale można je intepretować jako wektory łączące p ze wszystkimi punktami π. Kluczowe jest to, że niezależnie od wyboru punktu p na π mamy równość $\pi = p + T(\pi)$.

Przyjmijmy umowę notacyjną. Elementy przestrzeni afinicznej W zawartej w przestrzeni liniowej V, czyli punkty, oznaczamy małymi literami naszego alfabetu, a więc p, q, r, s itd. Wektory swobodne będziemy oznaczać literami greckimi: α, β, γ itd.

Jako, że zarówno punkty jak i wektory są w istocie elementami V, to w przypadku działania w przestrzeni współrzędnych $V = K^n$ będziemy zarówno współrzędne punktów, jak i wektorów zapisywać jednakowo w okrągłych nawiasach. Stosunkowo delikatnie będziemy natomiast mówić o wykonywaniu operacji mnożenia przez skalar i dodawania wektorów, która jest w przestrzeni V. Wiąże się to z tym, że jeśli punkty p, q należą do przestrzeni afinicznej W, to ani $2p$, ani $p + q$ nie muszą do niej należeć, choć same napisy mają sens, bo działamy w przestrzeni liniowej. Musimy określić kiedy stosowanie tych operacji jest przydatne. Motywacja pochodzi z pierwszego wykładu o układach równań.

Obserwacja 7.3

Rozważmy układ m równań liniowych na zbiorze $X = \{x_1, \ldots, x_n\}$ o współczynnikach w ciele K postaci:

$$\begin{cases} a_{11}x_1 + \ a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \qquad\qquad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases}$$

Załóżmy, że powyższy układ nie jest jednorodny. Wówczas jeśli wektory $v_0, \ldots, v_s$ są rozwiązaniami tego układu, to dla dowolnych $t_0, \ldots, t_s \in K$ kombinacja liniowa $t_0v_0 + \ldots + t_sv_s$ jest rozwiązaniem tego układu wtedy i tylko wtedy, gdy $t_0 + \ldots + t_s = 1$.

Dowód. Zapiszmy nasz układ w postaci macierzowej

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Wówczas to, że $v_0, \ldots, v_s$ są rozwiązaniami powyższego układu zapisać możemy w postaci:

$$Av_0 = b, \quad \ldots, \quad Av_s = b.$$

Zatem dla dowolnych $t_0, \ldots, t_s \in K$ mamy:

$$A(t_0v_0 + \ldots + t_sv_s) = t_0Av_0 + \ldots + t_sAv_s = t_0b + \ldots + t_sb.$$

Wektor $t_0v_0 + \ldots + t_sv_s$ jest zatem rozwiązaniem powyższego układu równań wtedy i tylko wtedy, gdy $(t_0 + \ldots + t_s)b = b$. Skoro $b \neq 0$, to zachodzi równość $t_0 + \ldots + t_s = 1$. □

Dla przykładu: wiedząc, że punkty $(1, 1)$, $(2, 3)$ są rozwiązaniami pewnego układu równań liniowych

$$a_{11}x_1 + a_{12}x_2 = b_1, \quad a_{21}x_1 + a_{22}x_2 = b_2$$

widzimy, że:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

A zatem każdy z punktów postaci $t(1, 1) + (1 - t)(2, 3)$ również jest rozwiązaniem tego układu równań.

Definicja 7.4: Kombinacja afiniczna, wagi

Niech V będzie przestrzenią liniową nad ciałem K. Dla każdych punktów $p_0, p_1, \ldots, p_n \in V$ i każdych $a_0, a_1, \ldots, a_n \in K$ spełniających

$$a_0 + a_1 + \ldots + a_n = 1$$

sumę:

$$a_0p_0 + a_1p_1 + \ldots + a_np_n \in V$$

nazywamy KOMBINACJĄ AFINICZNĄ PUNKTÓW $p_0, \ldots, p_n$ Z WAGAMI $a_0, \ldots, a_n$.

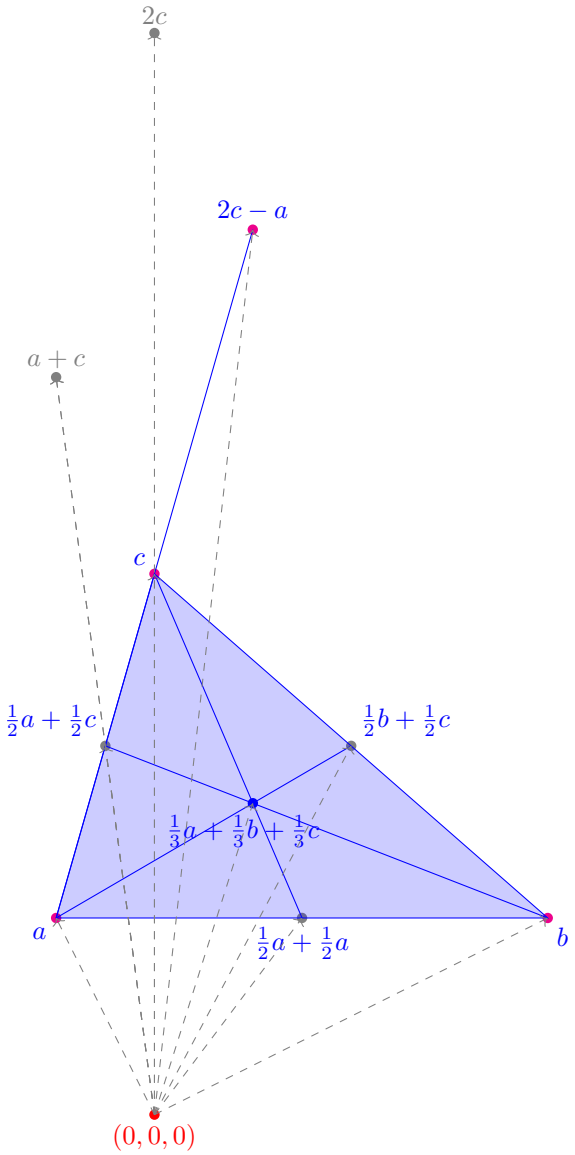
Kombinacje afiniczne są specjalnym typem kombinacji liniowych, w których suma współczynników to 1.

Przykłady:

- Dla $V = \mathbb{R}^3$ i $p_0 = (1,2,1), p_1 = (1,-1,-1), p_2 = (0,1,3)$ mamy
$$2p_0 + 3p_1 - 4p_2 = 2(1,2,1) + 3(1,-1,-1) - 4(0,1,3) = (5,-3,-13),$$
więc $p = (5,-3,-13)$ jest kombinacją afiniczną punktów p_0, p_1, p_2 z wagami $2, 3, -4$.
- Dla dowolnych punktów $x, y \in V$ oraz dowolnego $\lambda \in K$ punkt $\lambda x + (1-\lambda)y$ jest kombinacją afiniczną punktów x, y z wagami $\lambda, 1-\lambda$. W interpretacji geometrycznej punkty te należą do prostej zawierającej punkty x i y. Gdy $K = \mathbb{R}$ oraz $\lambda \in [0,1]$, zbiór punktów postaci $\{\lambda x + (1-\lambda)y\}$ interpretować będziemy jako odcinek w przestrzeni afinicznej o końcach w x i y.

- Dla dowolnych punktów $p, q, r \in V$ punkt $\frac{1}{3}p + \frac{1}{3}q + \frac{1}{3}r$ jest kombinacją afiniczną punktów p, q, r. W interpretacji nad $\mathbb{R}$ jest to środek ciężkości trójkąta o wierzchołkach w punktach p, q, r.

Spójrzmy na rysunek. Weźmy punkty a, b, c w przestrzeni afinicznej $\mathbb{R}^3$ (wierzchołki szarego trójkąta). Na niebiesko zaznaczone są pewne ich kombinacje afiniczne. Punkty $2c$ oraz $a + c$ nie są kombinacjami afinicznymi a, b, c, o ile $a, c \neq (0,0,0)$. Niebieskie odcinki oznaczają kombinacje afiniczne o dodatnich wagach (tzw. kombinacje wypukłe, mające sens nad $\mathbb{R}$) łączących je punktów. To co jest kluczowe to fakt, że z punktu widzenia tej konfiguracji nie ważne gdzie leży punkt $p_0 = (0,0,0)$, a kombinacje afiniczne (tylko one, bo liniowe – nie) punktów a, b, c są jednoznacznie wyznaczone.



Będziemy wielokrotnie korzystać z następującej obserwacji. Wyraża ona wagi $a_1, \ldots, a_k$ kombinacji afinicznej $p = a_0p_0 + a_1p_1 + \ldots + a_kp_k$ jako współczynniki w przedstawieniu wektora $\overrightarrow{p_0p}$ w postaci kombinacji liniowej wektorów $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$.

Obserwacja 7.4

Dla każdych $p_0, p_1, \ldots, p_k, p \in V$ oraz $a_0, \ldots, a_k \in K$ spełniających $a_0 + \ldots + a_k = 1$ zachodzi:

$$p = a_0p_0 + a_1p_1 + \ldots + a_kp_k \iff \overrightarrow{p_0p} = a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k}.$$

Dowód. Jeśli $p = a_0p_0 + \ldots + a_kp_k$, to

$$\begin{aligned}\overrightarrow{p_0p} &= p - p_0 = \\ &= a_0p_0 + \ldots + a_kp_k - p_0 = \\ &= a_0p_0 + \ldots + a_kp_k - (\underbrace{a_0 + \ldots + a_k}_{1})p_0 = \\ &= a_0(p_0 - p_0) + \ldots + a_k(p_k - p_0) = \\ &= a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k}.\end{aligned}$$

Na odwrót, jeśli $\overrightarrow{p_0p} = a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k}$, to:

$$\begin{aligned}p &= p_0 + \overrightarrow{p_0p} \\ &= (\underbrace{a_0 + \ldots + a_k}_{1})p_0 + a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k} = \\ &= a_0p_0 + a_1(p_0 + \overrightarrow{p_0p_1}) + \ldots + a_k(p_k + \overrightarrow{p_0p_k}) = \\ &= a_0p_0 + a_1p_1 + \ldots + a_kp_k.\end{aligned}$$

□

Zauważmy, że wybór punktu p_0 w powyższej uwadze jest zupełnie arbitralny. Można zamiast niego wziąć dowolny z punktów $p_1, \ldots, p_k$, odpowiednio modyfikując sformułowanie.

Przykład. Dla układu punktów w $\mathbb{R}^3$

$$p_0 = (1, 2, 1), p_1 = (1, -1, -1), p_2 = (0, 1, 3), p = (5, -3, -13)$$

mamy $p = 2p_0 + 3p_1 - 4p_2$, a więc równoważnie $\overrightarrow{p_0p} = 3\overrightarrow{p_0p_1} - 4\overrightarrow{p_0p_2}$, a także $\overrightarrow{p_1p} = 2\overrightarrow{p_1p_0} - 4\overrightarrow{p_1p_2}$ oraz $\overrightarrow{p_2p} = 2\overrightarrow{p_2p_0} + 3\overrightarrow{p_2p_1}$.

Omówimy teraz podstawowe własności kombinacji afinicznych.

Obserwacja 7.5

Niech $q_0, \ldots, q_r$ będzie układem punktów, z których każdy jest kombinacją afiniczną punktów $p_0, \ldots, p_k$. Wówczas każda kombinacja afiniczna punktów $q_0, \ldots, q_r$ jest też kombinacją afiniczną punktów $p_0, \ldots, p_k$.

Dowód. Dla każdego $i = 1, \ldots, r$ mamy $q_i = a_{i0}p_0 + \ldots + a_{ik}p_k$, dla pewnych wag $a_{i0}, \ldots, a_{ik}$. Stąd dla każdego układu wag $a_0, \ldots, a_r$ otrzymujemy:

$$\begin{aligned}a_0q_0 + \ldots + a_rq_r &= a_0(a_{00}p_0 + \ldots + a_{0k}p_k) + \ldots + a_r(a_{r0}p_0 + \ldots + a_{rk}p_k) = \\ &= (a_0a_{00} + \ldots + a_ra_{r0})p_0 + \ldots + (a_0a_{0k} + \ldots + a_ra_{rk})p_k\end{aligned}$$

Wystarczy pokazać, że suma współczynników występujących przy p_i w ostatnim wierszu ciągu równości powyżej wynosi 1. Mamy jednak:

$$\begin{aligned}&(a_0a_{00} + \ldots + a_ra_{r0}) + \ldots + (a_0a_{0k} + \ldots + a_ra_{rk}) = \\ &a_0(\underbrace{a_{00} + \ldots + a_{0k}}_{1}) + \ldots + a_r(\underbrace{a_{r0} + \ldots + a_{rk}}_{1}) = \\ &a_0 + \ldots + a_r = 1.\end{aligned}$$

□

Definicja 7.5: Podzbiór zamknięty ze względu na kombinacje afiniczne

Niech V będzie przestrzenią liniową nad K. Mówimy, że podzbiór $H \subseteq V$ JEST ZAMKNIĘTY ZE WZGLĘDU NA KOMBINACJE AFINICZNE, jeśli dla każdych punktów $p_0 \ldots, p_k \in H$ i każdych wag $a_0, \ldots, a_k$ zachodzi $a_0p_0 + \ldots + a_kp_k \in H$.

Poniższe twierdzenie wiąże pojęcia wprowadzone na dzisiejszym wykładzie, stanowiąc jednocześnie alternatywną definicję dla przestrzeni afinicznych (jako podzbiorów przestrzeni liniowych).

Obserwacja 7.6

Niech H będzie niepustym podzbiorem przestrzeni liniowej V nad K. Następujące warunki są równoważne.

(i) H jest zamknięty ze względu na kombinacje afiniczne,

(ii) H jest warstwą podprzestrzeni przestrzeni V,

(iii) H jest przestrzenią afiniczną.

Dowód. Oczywiście równoważność (ii) i (iii) wynika z definicji przestrzeni afinicznej.

Dowodzimy (i) $\Rightarrow$ (ii). Załóżmy, że H jest zamknięty ze względu na kombinacje afiniczne. Wybieramy $p_0 \in H$. Niech $W = \{\overrightarrow{p_0p} \,|\, p \in H\}$. Wykażemy, że W jest podprzestrzenią przestrzeni V. Weźmy dowolne $\alpha_1, \alpha_2 \in W$, czyli pewne $\overrightarrow{p_0p_1}$, $\overrightarrow{p_0p_2}$, dla pewnych $p_1, p_2 \in H$. Weźmy też dowolne $a_1, a_2 \in K$. Niech $a_0 = 1 - a_1 - a_2$. Zbiór H jest zamknięty ze względu na kombinacje afiniczne, więc

$$p = a_0p_0 + a_1p_1 + a_2p_2 \in H.$$

Stąd wektor $\overrightarrow{p_0p}$ należy do W. Ale $\overrightarrow{p_0p} = a_1\overrightarrow{p_0p_1} + a_2\overrightarrow{p_0p_2} = a_1\alpha_1 + a_2\alpha_2$. Stąd $a_1\alpha_1 + a_2\alpha_2 \in W$, co wobec dowolności $a_1, a_2, \alpha_1, \alpha_2$ dowodzi, że W jest podprzestrzenią przestrzeni V. Ponadto $H = p_0 + W$ na mocy definicji W. Zatem H jest warstwą podprzestrzeni W w V.

Dowodzimy (ii) $\Rightarrow$ (i). Załóżmy, że $H = q + W$, dla pewnego $q \in V$ oraz pewnej podprzestrzeni W przestrzeni V. Niech $p_0, \ldots, p_k \in H$ oraz $a_0, \ldots, a_k \in K$, przy czym $a_0 + \ldots + a_k = 1$. Wówczas $p_i = q + \alpha_i$, dla pewnych $\alpha_i \in W$, gdzie $i = 0, \ldots, k$. Zatem

$$\sum_{i=0}^{k} a_ip_i = \sum_{i=0}^{k} a_i(q + \alpha_i) = \sum_{i=0}^{k} a_iq + \sum_{i=0}^{k} a_i\alpha_i = q + \gamma,$$

gdzie $\gamma = a_0\alpha_0 + \ldots + a_k\alpha_k$ jest kombinacją liniową wektorów przestrzeni W, więc należy do W. Stąd $a_0p_0 + \ldots + a_kp_k = q + \gamma \in q + W = H$. Zatem H jest zamknięty ze względu na kombinacje afiniczne. □

Definicja 7.6: Podprzestrzeń przestrzeni afinicznej

Niech H_1, H_2 będą przestrzeniami afinicznymi w przestrzeni liniowej V. Jeśli $H_1 \subseteq H_2$, to mówimy, że H_1 jest PODPRZESTRZENIĄ PRZESTRZENI AFINICZNEJ H_2.

Obserwacja 7.7

Niech $p_0, \ldots, p_k$ będą punktami przestrzeni afinicznej H. Wówczas zbiór wszystkich kombinacji afinicznych punktów $p_0, \ldots, p_k$ jest podprzestrzenią przestrzeni afinicznej H.

Definicja 7.7: Podprzestrzeń rozpięta na układzie punktów

Niech $p_0, \ldots, p_k$ będą punktami przestrzeni afinicznej H. Wówczas zbiór wszystkich kombinacji afinicznych punktów $p_0, \ldots, p_k$ nazywamy PODPRZESTRZENIĄ AFINICZNĄ ROZPIĘTĄ NA $p_0, \ldots, p_k$ (lub PODPRZESTRZENIĄ GENEROWANĄ przez $p_0, \ldots, p_k$) i oznaczamy $\mathrm{af}(p_0, \ldots, p_k)$.

Podobnie jak przestrzeń liniowa może być rozpięta przez dowolny układ wektorów, podobnie i rozważać można podprzestrzenie afiniczne rozpięte przez dowolne układy punktów.

Definicja 7.8

Niech $X \subseteq V$. Przez $\mathrm{af}(X)$ rozumiemy zbiór kombinacji afinicznych elementów zbioru X.

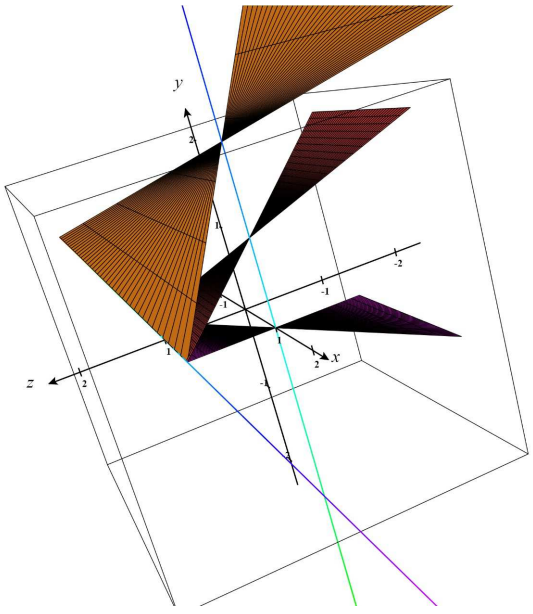
Przykłady:

- Dla dowolnej przestrzeni liniowej V mamy $\mathrm{af}(V) = V$.
- Dla dwóch różnych punktów p, q zbiór $\mathrm{af}(p, q)$ jest prostą.
- Rozważmy proste l_1, l_2 zadane w $\mathbb{R}^3$ układami postaci: $l_1 : x_1 = 0, x_2 = 0, l_2 : x_1 = 1, x_3 = 0$. Te dwie proste nie mają punktu wspólnego – są to tzw. proste skośne. Każdy punkt na prostej l_1 ma postać $(0, 0, s)$, gdzie $s \in \mathbb{R}$, a każdy punkt na l_2 ma postać $(1, t, 0)$, gdzie $t \in \mathbb{R}$. A zatem do prostych tych należą na przykład punkty $(0, 0, 0), (0, 0, 1), (1, 0, 0)$ oraz $(1, 1, 0)$. Dowolny punktu $(x, y, z) \in \mathbb{R}^3$ jest ich kombinacją afiniczną:

  $$(x, y, z) = (x - y)(1, 0, 0) + y(1, 1, 0) + z(0, 0, 1) + (1 - x - z)(0, 0, 0).$$

  A zatem $\mathrm{af}(l_1 \cup l_2) = \mathbb{R}^3$.

Od razu chciałbym przestrzec Czytelnika przed różnymi uproszczeniami. Przestrzeń afiniczna rozpięta na układzie punktów nie jest tym samym co zbiór prostych łączących poszczególne punkty. Proszę rozważyć przykład, który kilka razy pojawiał się na kolokwiach. Dla podzbiorów H_1, H_2 przestrzeni afinicznej $\mathbb{R}^3$ definiujemy zbiór $H_3 = \{tp + (1 - t)q \,|\, p \in H_1, q \in H_2, t \in \mathbb{R}\}$. Wówczas dla $H_1 = (0, 0, 1) + \mathrm{lin}((1, 0, 0))$ oraz $H_2 = (1, 0, 0) + \mathrm{lin}((0, 1, 0))$ zbiór H_3 nie jest równy $\mathbb{R}^3$, a nawet nie jest podprzestrzenią afiniczną.



Idea jest taka: wybierzmy punkt na prostej H_2 i połączmy go prostymi ze wszystkimi punktami z prostej skośnej H_1 (dla ustalonych $p \in H_1, q \in H_2$ zbiór $\{tp + (1 - t)q \,|\, t \in \mathbb{R}\}$ jest prostą). Dostajemy *prawie płaszczyznę* (bez jednej prostej). Można w ten sposób uzyskać *prawie wszystkie* *prawie płaszczyzny* zawierające H_1 poza jedną – równoległą do H_2. Łącznie dostajemy $\mathbb{R}^3$ bez dwóch *prawie płaszczyzn*.

Obserwacja 7.8

Niech $p_0, \ldots, p_k$ będą punktami przestrzeni afinicznej H. Następujące warunki są równoważne:

(1) $H = \mathrm{af}(p_0, \ldots, p_k)$,

(2) $T(H) = \mathrm{lin}(\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k})$.

Dowód. Dowodzimy (1) $\Rightarrow$ (2). Skoro H jest przestrzenią afiniczną, to zgodnie z dowodem faktu charakteryzującego przestrzenie afiniczne jako podzbiory zamknięte na kombinacje afiniczne wiemy, że dla każdego $q \in H$ mamy

$$T(H) = \{\overrightarrow{qp} \,|\, p \in H\}.$$

Jeśli $H = \mathrm{af}(p_0, \ldots, p_k)$, to biorąc $q = p_0$ otrzymujemy

$$\begin{aligned} T(H) &= \{\overrightarrow{p_0p} \,|\, p \in H\} = \\ &= \{\overrightarrow{p_0p} \,|\, \text{ dla wszystkich } p = a_0p_0 + \ldots + a_kp_k, \text{ gdzie } a_0 + \ldots + a_k = 1\} = \\ &= \{a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k} \,|\, a_1, \ldots, a_k \in K\} = \\ &= \mathrm{lin}(\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}). \end{aligned}$$

Na odwrót: jeśli $T(H) = \mathrm{lin}(\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k})$, to już wiemy, że dla każdego $p \in H$ mamy

$$\overrightarrow{p_0p} = a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k},$$

dla pewnych $a_1, \ldots, a_k \in K$. Chcemy, by $p \in \mathrm{af}(p_0, \ldots, p_k)$. Przyjmując $a_0 = 1 - a_1 - \ldots - a_k$ otrzymujemy

$$p = p_0 + \overrightarrow{p_0p} = (a_0 + \ldots + a_k)p_0 + a_1(p_1 - p_0) + \ldots + a_k(p_k - p_0).$$

Po uproszczeniu dostajemy $p = a_0p_0 + \ldots + a_kp_k$. Wobec dowolności p otrzymujemy stąd (1). □

Przykład. W przestrzeni $V = \mathbb{R}^3$ niech $H = \mathrm{af}((1,0,2),(2,1,3),(4,1,1))$. Wówczas przestrzeń styczna $T(H)$ to

$$\mathrm{lin}((1,1,1),(3,1,-1)) = \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, x_1 - 2x_2 + x_3 = 0\}.$$

Stąd sama podprzestrzeń H opisana jest jako zbiór rozwiązań układu złożonego z pojedynczego równania $x_1 - 2x_2 + x_3 = 3$.

Definicja 7.9: Układ afinicznie zależny/niezależny

Niech $p_0, \ldots, p_k$ będzie układem punktów przestrzeni afinicznej H nad ciałem K.

- Mówimy, że układ $p_0, \ldots, p_k$ jest AFINICZNIE ZALEŻNY (albo, że jest W POŁOŻENIU SZCZEGÓLNYM), jeśli jeden z punktów tego układu jest kombinacją afiniczną pozostałych.
- Mówimy, że układ $p_0, \ldots, p_k$ jest AFINICZNIE NIEZALEŻNY (albo, że jest W POŁOŻENIU OGÓLNYM), jeśli nie jest on w położeniu szczególnym.

Przykład. W przestrzeni $H = \mathbb{R}^3$ układ $((3,7,4),(1,9,7),(5,5,1))$ jest afinicznie zależny, bo mamy

$$(1,9,7) = 2(3,7,4) - 1(5,5,1).$$

Układ $(3,1,1),(1,2,1)$ jest natomiast afinicznie niezależny w $\mathbb{R}^3$.

Afiniczna (nie)zależność układu punktów nie zależy od kolejności punktów układu. Każdy podukład układu afinicznie niezależnego jest afinicznie niezależny. Łatwo widzieć, że istnieje zależność pomiędzy afiniczną niezależnością układu punktów i liniową niezależnością odpowiadającego mu układu wektorów. Układ $k+1$ punktów jest afinicznie niezależny wtedy i tylko wtedy, gdy układ k wektorów o początkach w jednym z tych punktów łączących go z pozostałymi punktami układu jest liniowo niezależny.

Obserwacja 7.9

Niech $p_0, \ldots, p_k$ będą punktami przestrzeni afinicznej H nad K. Równoważne są warunki:

(1) układ $p_0, \ldots, p_k$ jest afinicznie niezależny,

(2) układ wektorów $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$ jest liniowo niezależny.

Dowód. Załóżmy, że (2) nie jest prawdą i układ $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$ jest liniowo zależny. Wówczas któryś z wektorów tego układu jest liniową kombinacją pozostałych. Ewentualnie przenumerowując $p_1, \ldots, p_k$ możemy zakładać, że

$$\overrightarrow{p_0p_k} = a_1\overrightarrow{p_0p_1} + \ldots + a_{k-1}\overrightarrow{p_0p_{k-1}},$$

dla pewnych $a_1, \ldots, a_{k-1} \in K$. Niech $a_0 = 1 - a_1 - \ldots - a_{k-1}$. Wówczas

$$\begin{aligned} a_0p_0 + a_1p_1 + \ldots + a_{k-1}p_{k-1} &= a_0(p_0 + \overrightarrow{p_0p_0}) + a_1(p_0 + \overrightarrow{p_0p_1}) + \ldots + a_{k-1}(p_0 + \overrightarrow{p_0p_{k-1}}) = \\ &= (\underbrace{a_0 + a_1 + \ldots + a_{k-1}}_{=1})p_0 + \underbrace{a_1\overrightarrow{p_0p_1} + \ldots + a_{k-1}\overrightarrow{p_0p_{k-1}}}_{=\overrightarrow{p_0p_k}} = \\ &= p_0 + \overrightarrow{p_0p_k} = p_k. \end{aligned}$$

A zatem p_k jest kombinacją afiniczną punktów $p_0, \ldots, p_{k-1}$. A zatem (1) $\Rightarrow$ (2).

Na odwrót: przypuśćmy, że układ $p_0, \ldots, p_k$ jest afinicznie zależny. Rozważamy dwa przypadki.

- Punkt p_0 jest kombinacją afiniczną punktów $p_1, \ldots, p_k$. Niech $p_0 = a_1p_1 + \ldots + a_kp_k$, przy czym $a_1 + \ldots + a_k = 1$. Mamy więc

  $$0 = \overrightarrow{p_0p_0} = \overrightarrow{p_0(a_1p_1 + \ldots + a_kp_k)} = a_1\overrightarrow{p_0p_1} + \ldots + a_k\overrightarrow{p_0p_k}.$$

  Zatem układ $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$ jest liniowo zależny, i mamy sprzeczność z założeniem, że zachodzi (2).

- Punkt p_0 nie jest kombinacją afiniczną[2] punktów $p_1, \ldots, p_k$. Wówczas skoro układ $p_0, \ldots, p_k$ jest afinicznie zależny, to jeden z punktów $p_1, \ldots, p_k$ jest kombinacją afiniczną pozostałych. Po ewentualnym przenumerowaniu $p_1, \ldots, p_k$ możemy założyć, że p_k jest kombinacją afiniczną $p_0, \ldots, p_{k-1}$. W szczególności $p_k = a_0p_0 + \ldots + a_{k-1}p_{k-1}$, gdzie $a_0 + \ldots + a_{k-1} = 1$. A zatem:

  $$\overrightarrow{p_0p_k} = a_0\overrightarrow{p_0p_0} + a_1\overrightarrow{p_0p_1} + \ldots + a_{k-1}\overrightarrow{p_0p_{k-1}} = a_1\overrightarrow{p_0p_1} + \ldots + a_{k-1}\overrightarrow{p_0p_{k-1}},$$

  czyli układ $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$ jest liniowo zależny, co przeczy (2). A zatem (2) $\Rightarrow$ (1).

□

Definicja 7.10: Baza punktowa

Mówimy, że układ punktów $p_0, \ldots, p_k$ punktów przestrzeni afinicznej H jest BAZĄ PUNKTOWĄ przestrzeni H, jeśli spełnia następujące dwa warunki:

- układ $p_0, \ldots, p_k$ jest afinicznie niezależny,
- $H = \mathrm{af}(p_0, \ldots, p_k)$, czyli układ $p_0, \ldots, p_k$ ROZPINA H.

Obserwacja 7.10

Niech $p_0, \ldots, p_k$ będą punktami przestrzeni afinicznej H. Układ $p_0, \ldots, p_k$ jest baza punktową przestrzeni H wtedy i tylko wtedy, gdy układ wektorów $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$ jest bazą przestrzeni $T(H)$. W szczególności każde dwie skończone bazy punktowe przestrzeni afinicznej są równoliczne.

Powyższy wniosek uzasadnia następującą definicję.

[2] Np. układ punktów $p_0 = (1, 2), p_1 = (2, 2), p_2 = (3, 3), p_3 = (4, 4)$ jest afinicznie zależny w $\mathbb{R}^2$, ale $p_0 \notin \mathrm{af}(p_1, p_2, p_3)$.

Definicja 7.11: Wymiar przestrzeni afinicznej; proste, płaszczyzny

WYMIAREM przestrzeni afinicznej H nazywamy wymiar jej przestrzeni stycznej $T(H)$. Wymiar przestrzeni H oznaczamy $\dim H$. Przestrzenie afiniczne wymiaru 1 nazywamy PROSTYMI, przestrzenie wymiaru 2 – PŁASZCZYZNAMI.

- Układ $0, \epsilon_1, \ldots, \epsilon_n$, wektorów przestrzeni liniowej K^n, gdzie $\epsilon_1, \ldots, \epsilon_n$ są wektorami bazy standardowej, tworzy bazę punktową przestrzeni afinicznej K^n. Przestrzeń ta ma wymiar równy n.
- Niech $H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 - x_2 + 3x_3 = 6\}$. Wówczas $T(H)$ opisać można jako zbiór

  $$\{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 - x_2 + 3x_3 = 0\} = \mathrm{lin}((1,1,0),(0,3,1)).$$

  A zatem $\dim H = 2$. Weźmy teraz $p_0 = (1,1,2) \in H$ i niech $p_1 = p_0 + (1,1,0) = (2,2,2)$ oraz $p_2 = p_0 + (0,3,1) = (1,4,3)$. Wówczas układ p_0, p_1, p_2 jest bazą punktową przestrzeni H.
- Dla $M = \mathrm{af}((1,2,3),(5,4,1),(-3,0,5),(2,1,4)) \subseteq \mathbb{R}^3$ znajdujemy opis przestrzeni $T(M)$ jako

  $$\mathrm{lin}((4,2,-2),(-4,-2,2),(1,-1,1)) = \mathrm{lin}((4,2,-2),(1,-1,1)).$$

  Zatem $\dim M = 2$. Układ trzech punktów: $(1,2,3)$, $(1,2,3) + (4,2,-2)$, $(1,2,3) + (1,-1,1)$ jest więc bazą punktową przestrzeni M.

Następujące własności baz punktowych są konsekwencjami odpowiednich własności baz przestrzeni liniowych oraz powyższego wniosku.

Obserwacja 7.11

Każda przestrzeń afiniczna ma bazę punktową. Jeśli $\dim H = k$, to każda baza punktowa przestrzeni H ma $k+1$ punktów. Układ $p_0, \ldots, p_k$ punktów w przestrzeni afinicznej H nad K jest bazą punktową przestrzeni H wtedy i tylko wtedy, dla każdego $p \in H$ istnieje dokładnie jeden układ wag $a_0, \ldots, a_k \in K$ taki, że $p = a_0 p_0 + \ldots + a_k p_k$.

Definicja 7.12: Współrzędne barycentryczne

Niech H będzie przestrzenią afiniczną nad K i niech $p_0, \ldots, p_k$ będzie bazą punktową przestrzeni H. WSPÓŁRZĘDNYMI (BARYCENTRYCZNYMI) PUNKTU $p \in H$ W BAZIE PUNKTOWEJ $p_0, \ldots, p_k$ nazywamy układ wag $a_0, \ldots, a_k \in K$ taki, że $p = a_0 p_0 + \ldots + a_k p_k$.

Definicja 7.13: Układ bazowy i współrzędne w nim

Niech H będzie przestrzenią afiniczną nad K.

- Jeśli p_0 jest punktem przestrzeni H oraz $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ jest bazą przestrzeni $T(H)$, to układ $p_0; \alpha_1, \ldots, \alpha_n$ nazywamy UKŁADEM BAZOWYM dla przestrzeni H.
- Dla punktu p przestrzeni H układ $a_1, \ldots, a_n$ elementów ciała K taki, że $p = p_0 + a_1\alpha_1 + \ldots + a_n\alpha_n$ nazywamy WSPÓŁRZĘDNYMI PUNKTU p W UKŁADZIE BAZOWYM $p_0; \mathcal{A}$.

Odwzorowanie $K^n \to H$ opisane wzorem:

$$(s_1, \ldots, s_n) \mapsto p_0 + s_1\alpha_1 + s_2\alpha_2 + \ldots + s_n\alpha_n,$$

nazywamy PARAMETRYZACJĄ przestrzeni H.

Pojęcie parametryzacji stosuje się także w analizie np. przy opisie krzywych. Funkcja $t \mapsto (\cos t, \sin t)$ jest parametryzacją okręgu na płaszczyźnie, ale oczywiście nie jest to parametryzacja w sensie algebry liniowej. W przeciwieństwie do podanej parametryzacji okręgu, parametryzacje przestrzeni afinicznej opisane w powyższej definicji są zawsze bijekcjami. Oczywiście jasny jest także następujący wniosek.

Obserwacja 7.12

Układ $p_0, \ldots, p_n$ jest bazą punktową przestrzeni H wtedy i tylko wtedy, gdy układ $p_0; \overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_n}$ jest układem bazowym przestrzeni H.

Przykład. Niech $H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, 2x_1 - x_2 + 3x_3 = 5\}$. Wówczas $(1, 0, 1) \in H$ oraz

$$T(H) = \{(x_1, x_2, x_3) \,|\, 2x_2 - x_2 + 3x_3 = 0\} = \mathrm{lin}((1, 2, 0), (0, 3, 1)).$$

Stąd $(1, 0, 1); (1, 2, 0), (0, 3, 1)$ jest układem bazowym w H, zaś przyporządkowanie

$$\mathbb{R}^2 \ni (s_1, s_2) \mapsto (1, 0, 1) + s_1(1, 2, 0) + s_2(0, 3, 1) = (s_1 + 1, 2s_1 + 3s_2, s_2 + 1) \in H$$

jest parametryzacją przestrzeni H.

Zadanie 7.1. *Niech $\pi \subset \mathbb{R}^4$ będzie płaszczyzną opisaną układem:*

$$\begin{cases} 2x_1 + x_2 - x_3 &= 2 \\ 4x_1 + x_4 &= 5. \end{cases}$$

(a) Wyznacz bazę punktową płaszczyzny π.

(b) Podaj przykład takiej płaszczyzny π' przechodzącej przez punkt $(0, 1, 0, 1)$, że $\mathrm{af}(\pi \cup \pi') = \mathbb{R}^4$, poprzez wskazanie parametryzacji płaszczyzny π'.

Rozwiązanie. Zauważmy, że punkt $(x_1, x_2, x_3, x_4) = (1, 0, 0, 1)$ jest rozwiązaniem układu opisującego π, a zatem należy do π. Przestrzeń styczna do π opisana jest układem jednorodnym:

$$\begin{cases} 2x_1 + x_2 - x_3 &= 0 \\ 4x_1 + x_4 &= 0. \end{cases}$$

Łatwo sprawdzamy, że

$$T(\pi) = \mathrm{lin}\left((0, 1, 1, 0), (-1, 2, 0, 4)\right).$$

A zatem punkty $(1, 0, 0, 1) + (0, 1, 1, 0) = (1, 1, 1, 1)$ oraz $(1, 0, 0, 1) + (-1, 2, 0, 4) = (0, 2, 0, 5)$ należą do π, zaś układ

$$(1, 0, 0, 1), (1, 1, 1, 1), (0, 2, 0, 5)$$

jest afinicznie niezależny i rozpina π, czyli jest poszukiwaną bazą punktową.

Przechodzimy do (b). Punkt $(0, 1, 0, 1)$ oczywiście nie należy do π. Szukamy punktu p_4 takiego, by układ

$$p_0 = (1, 0, 0, 1), \quad p_1 = (1, 1, 1, 1), \quad p_2 = (0, 2, 0, 5), \quad p_3 = (0, 1, 0, 1), \quad p_4$$

był bazą punktową $\mathbb{R}^4$. Wtedy rozwiązaniem będzie np. płaszczyzna $\pi' = \mathrm{af}(p_0, p_3, p_4)$. W istocie problem sprowadza się teraz do dopełnienia układu:

$$\overrightarrow{p_0p_1} = (0, 1, 1, 0), \quad \overrightarrow{p_0p_2} = (-1, 2, 0, 4), \quad \overrightarrow{p_0p_3} = (-1, 1, 0, 0)$$

do bazy przestrzeni liniowej $\mathbb{R}^4$ dowolnym wektorem $\overrightarrow{p_0p_4}$. Łatwo widać, że można to zrobić na przykład za pomocą wektora $(0, 0, 1, 0)$. A zatem układ bazowy znalezionej przez nas płaszczyzny to

$$(1, 0, 0, 1); (-1, 1, 0, 0), (0, 0, 1, 0)$$

i możemy napisać jej parametryzację:

$$(s, t) \mapsto (1, 0, 0, 1) + s(-1, 1, 0, 0) + t(0, 0, 1, 0), \quad s, t \in \mathbb{R}.$$

■

Powyższe definicje afinicznej niezależności układu punktów, bazy punktowej i układu bazowego poprowadzą nas, podobnie jak w przypadku przestrzeni liniowych, do pojęcia przekształcenia afinicznego. Wcześniej jednak warto odnotować jeszcze szereg pojęć geometrycznych, które można określić już na początku pracy z przestrzeniami afinicznymi. Najważniejszym jest równoległość przestrzeni afinicznych, o którym mówiliśmy ostatnio. W zadaniach dotyczących przestrzeni afinicznych stosujemy specyficzną dla geometrii elementarnej nomenklaturę. Mówimy chociażby o tym, że

- PUNKT p LEŻY NA prostej/płaszczyźnie, co oznacza, że należy do tej prostej/płaszczyzny,
- prosta, płaszczyzna lub przestrzeń afiniczna PRZECHODZI PRZEZ DANY PUNKT, co znaczy, że ten punkt do niej należy,
- proste, płaszczyzny lub przestrzenie afiniczne PRZECINAJĄ SIĘ mając na myśli to, że odpowiednie przestrzenie afiniczne mają punkt wspólny (lub nie!),
- punkt p LEŻY POMIĘDZY PUNKTAMI q, r, jeśli $p = tq + (1-t)r$, gdzie $0 < t < 1$ (tutaj nad $\mathbb{R}$),
- punkty p, q, r są WSPÓŁLINIOWE, jeśli leżą na jednej prostej,
- podprzestrzenie F_1, F_2 SĄ SKOŚNE, to znaczy: $F_1 \cap F_2 = \emptyset$ oraz $T(F_1) \cap T(F_2) = \{0\}$.

Widzimy wyraźnie, że język geometrii afinicznej korzysta obficie ze znanego nam ze szkoły nazewnictwa obecnego w geometrii elementarnej płaszczyzny czy przestrzeni. Do pełni „geometrycznej operatywności" brakuje nam rzecz jasna kątów oraz odległości, a także izometrii. Te obiekty wprowadzimy rozważając strukturę przestrzeni euklidesowej w przestrzeni afinicznej. Wcześniej poznamy podstawy algebraicznego opisu przekształceń afinicznych[3] – bardzo atrakcyjnych również z geometrycznego punktu widzenia.

Na koniec powiedzmy nieco o aksjomatycznej definicji przestrzeni afinicznej.

Definicja 7.14: Aksjomatyczna definicja przestrzeni afinicznej

Niech W będzie przestrzenią liniową nad ciałem K. PRZESTRZENIĄ AFINICZNĄ nad ciałem K o PRZESTRZENI WEKTORÓW SWOBODNYCH W nazywamy niepusty zbiór H wraz z odwzorowaniem:

$$H \times W \longrightarrow H, \quad (p, \alpha) \mapsto p \boxplus \alpha,$$

zwanym DODAWANIEM WEKTORA DO PUNKTU, przy czym spełnione są następujące warunki, zwane aksjomatami przestrzeni afinicznej.

(1) Dla każdych $\alpha, \beta \in W$ i każdego $p \in H$ mamy:

$$p \boxplus 0 = p, \quad p \boxplus (\alpha + \beta) = (p \boxplus \alpha) \boxplus \beta.$$

(2) Dla każdych $p, q \in H$ istnieje dokładnie jeden $\alpha \in W$ taki, że $p \boxplus \alpha = q$.

Elementy przestrzeni H nazywamy PUNKTAMI przestrzeni H.

Przykłady:

- Przestrzenie afiniczne rozpatrywane dotąd jako warstwy podprzestrzeni liniowych w przestrzeniach liniowych są przestrzeniami afinicznymi w sensie powyższej definicji. Jeśli $H \subseteq V$ jest warstwą podprzestrzeni W przestrzeni liniowej V, to dla każdego $p \in H$ i każdego $\alpha \in W$ mamy $p + \alpha \in H$, więc rolę odwzorowania $\boxplus : H \times W \longrightarrow H$ gra tu po prostu dodawanie wektorów zapożyczone z V.
- Jeśli H_1, H_2 są przestrzeniami afinicznymi nad K, to $H_1 \times H_2 = \{(p_1, p_2) \,|\, p_1 \in H_1, p_2 \in H_2\}$ jest przestrzenią afiniczną nad K. Jeśli W_1, W_2 są przestrzeniami wektorów swobodnych przestrzeni H_1, H_2, odpowiednio, to $W_1 \times W_2$ jest przestrzenią wektorów swobodnych przestrzeni $H_1 \times H_2$.
- Niech H będzie przestrzenią afiniczną nad K i niech $X \neq \emptyset$. Wówczas $P(X, H) = \{$ funkcje $X \to H\}$ jest przestrzenią afiniczną nad K. Jeśli W jest przestrzenią wektorów swobodnych przestrzeni H, to $F(X, W)$ jest przestrzenią wektorów swobodnych przestrzeni $F(X, H)$.
- Struktura $(P, \mathrm{lin}(1,0), \boxplus)$ rozważana w problemie pierwszym jest przestrzenią afiniczną.

Czasem wygodniejszy w użytku jest język przestrzeni afinicznej bez kontekstu przestrzeni liniowej, w której jest ona zanurzona. Nie jest to, jak się okazuje, ogólniejsza teoria. W jednym z dodatków uzasadnimy, że dla każdej abstrakcyjnie zdefiniowanej przestrzeni afinicznej istnieje izomorficzna z nią przestrzeń afiniczna będąca warstwą pewnej przestrzeni liniowej.

[3] Książeczka dr. Jerzego Bednarczuka zawierająca podejście geometryczne do przekształceń afinicznych mówi nawet o „Uroku przekształceń afinicznych" (Biblioteczka Matematyczna, tom 36).

7.1 Uzupełnienie. Wymiar sumy podprzestrzeni afinicznych

Wzajemne położenie przestrzeni afinicznych jest bardziej skomplikowane niż w przypadku przestrzeni liniowych, co odczytujemy choćby przez obecność nieprzecinających się prostych w przestrzeni trójwymiarowej. W trzech wymiarach nie pojawiają się poważniejsze komplikacje dotyczące położenia przestrzeni afinicznych. Jednak już dokonanie klasyfikacji wzajemnego położenia płaszczyzn w przestrzeni czterowymiarowej nie jest trywialnym zadaniem i prowadzi do pięciu konfiguracji. W zrozumieniu tych zależności pomóc nam może analog formuły Grassmanna w przestrzeniach afinicznych.

Twierdzenie 7.1: Formuła Grassmanna w przestrzeniach afinicznych

Niech H_1 oraz H_2 będą przestrzeniami afinicznymi w przestrzeni liniowej V. Wówczas:

- jeśli $H_1 \cap H_2 \neq \emptyset$, to

$$\dim(H_1 + H_2) = \dim(\mathrm{af}(H_1, H_2)) = \dim H_1 + \dim H_2 - \dim(H_1 \cap H_2),$$

- jeśli $H_1 \cap H_2 = \emptyset$, to

$$\dim(H_1 + H_2) = \dim(\mathrm{af}(H_1, H_2)) = \dim H_1 + \dim H_2 - \dim(T(H_1) \cap T(H_2)) + 1.$$

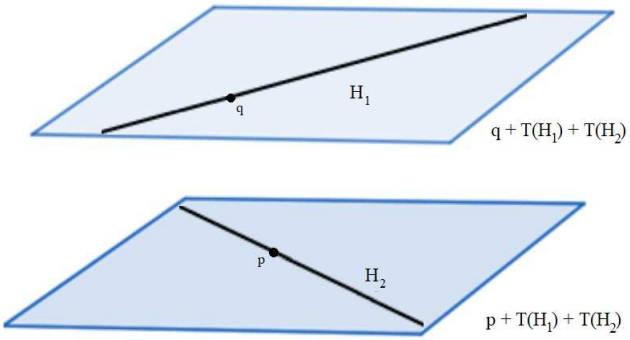
Uzasadnienie rozpoczniemy od następującego lematu.

Obserwacja 7.13

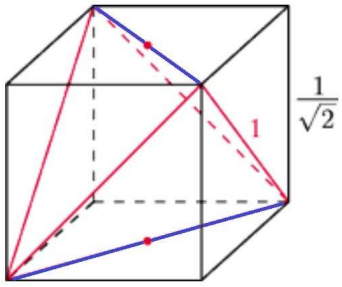
Jeśli dane są przestrzenie afiniczne $H_1 = p_1 + T(H_1)$ oraz $H_2 = p_2 + T(H_2)$, to

$$H_1 \cap H_2 \neq \emptyset \iff \overrightarrow{p_1 p_2} \in T(H_1) + T(H_2).$$

Rozważmy poniższą ilustrację. Pokazuje ona typową sytuację: dwie proste skośne H_1, H_2 w przestrzeni $\mathbb{R}^3$. Zwłaszcza w (przyszłym) kontekście liczenia odległości pomiędzy prostymi, bardzo wygodne jest zawarcie owych prostych w równoległych płaszczyznach, których przestrzenią styczną jest $T(H_1) + T(H_2)$



Oto przykład zastosowania tego spojrzenia dla szkolnego rachunku[4] polegającego na wyznaczeniu odległości pomiędzy środkami przeciwległych krawędzi czworościanu foremnego o krawędzi długości 1.



[4] Więcej: https://www.deltami.edu.pl/temat/matematyka/geometria/stereometria/2014/01/31/2014-02-k25.pdf.

Dowód. Jeśli istnieje punkt $r = p_1 + \alpha_1 = p_2 + \alpha_2$, gdzie $\alpha_1 \in T(H_1)$ oraz $\alpha_2 \in T(H_2)$, to

$$\overrightarrow{p_1p_2} = \overrightarrow{p_1r} + \overrightarrow{rp_2} = \alpha_1 - \alpha_2 \in T(H_1) + T(H_2).$$

Z drugiej strony, jeśli $\overrightarrow{p_1p_2} \in T(H_1) + T(H_2)$, to $\overrightarrow{p_1p_2} = \alpha_1 + \alpha_2$, gdzie $\alpha_1 \in T(H_1)$, $\alpha_2 \in T(H_2)$, co daje nam

$$p_2 = p_1 + \overrightarrow{p_1p_2} = p_1 + \alpha_1 + \alpha_2,$$

a stąd

$$p_1 + \alpha_1 = p_2 - \alpha_2 \in (p_1 + T(H_1)) \cap (p_2 \cap T(H_2)).$$

□

Obserwacja 7.14

Jeśli dane są przestrzenie afiniczne $H_1 = p_1 + T(H_1)$ oraz $H_2 = p_2 + T(H_2)$, to

$$r \in H_1 \cap H_2 \quad \Longrightarrow \quad H_1 \cap H_2 = r + (T(H_1) \cap T(H_2)).$$

Dowód. Na mocy Obserwacji 7.1 wiemy, że $p_1 + T(H_1) = r + T(H_1)$ oraz $p_2 + T(H_2) = r + T(H_2)$. Stąd

$$H_1 \cap H_2 = (r + T(H_1)) + (r + T(H_2)) = r + (T(H_1) \cap T(H_2)).$$

□

Dla dowodu formuły Grassmanna niezbędne jest określenie sumy przestrzeni afinicznych.

Definicja 7.15: Suma podprzestrzeni afinicznych

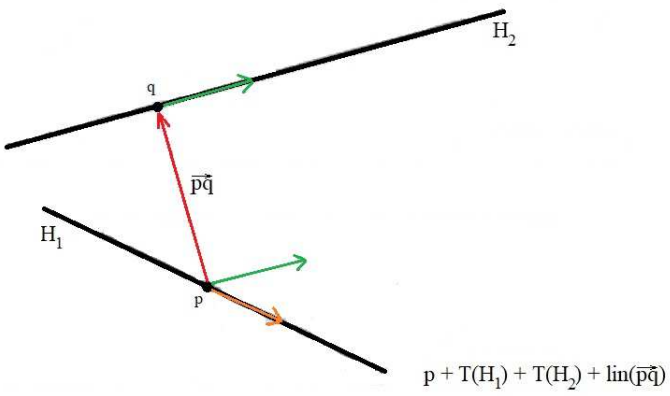
Jeśli H_1, H_2 są przestrzeniami afinicznymi w przestrzeni liniowej V, wówczas przez $H_1 + H_2$ definiujemy najmniejszą podprzestrzeń afiniczną zawierającą H_1 oraz H_2, czyli $\mathrm{af}(H_1, H_2)$.

Opis sumy podprzestrzeni jest bardzo prosty.

Obserwacja 7.15

Jeśli dane są przestrzenie afiniczne $H_1 = p_1 + T(H_1)$ oraz $H_2 = p_2 + T(H_2)$, to

$$H_1 + H_2 = p_1 + (T(H_1) + T(H_2) + \mathrm{lin}(\overrightarrow{p_1p_2})).$$



Dowód. Załóżmy, że L jest przestrzenią afiniczną zawierającą H_1 oraz H_2. Skoro p_1, p_2 należą do L, to istnieje podprzestrzeń W taka, że $L = p_1 + W = p_2 + W$ oraz $\mathrm{lin}(\overrightarrow{p_1p_2}) \subseteq W$. Jednak $H_1 \subseteq L$ implikuje $T(H_1) \subseteq W$ i podobnie $H_2 \subseteq L$ implikuje $T(H_2) \subseteq W$. A zatem

$$T(H_1) + T(H_2) + \mathrm{lin}(\overrightarrow{p_1p_2}) \subseteq L$$

oraz $p_1 + (T(H_1) + T(H_2) + \mathrm{lin}(\overrightarrow{p_1p_2})) \subseteq L$. A zatem $p_1 + (T(H_1) + T(H_2) + \mathrm{lin}(\overrightarrow{p_1p_2}))$ zawarta jest w $\mathrm{af}(H_1, H_2)$. Z drugiej strony $\mathrm{af}(H_1, H_2) \subseteq p_1 + (T(H_1) + T(H_2) + \mathrm{lin}(\overrightarrow{p_1p_2}))$, co daje tezę. □

Jesteśmy gotowi na dowód formuły Grassmanna. Załóżmy najpierw, że $H_1 \cap H_2 \neq \emptyset$. Zgodnie z Obserwacją 7.13 mamy

$$T(H_1) + T(H_2) + \operatorname{lin}(\overrightarrow{p_1 p_2}) = T(H_1) + T(H_2),$$

a zatem zgodnie z Obserwacją 7.15 mamy:

$$H_1 + H_2 = p_1 + T(H_1) + T(H_2).$$

Zatem:

$$\begin{aligned}\dim(H_1 + H_2) &= \dim(T(H_1) + T(H_2)) \\ &= \dim T(H_1) + \dim T(H_2) - \dim(T(H_1) \cap T(H_2)) \\ &= \dim H_1 + \dim H_2 - \dim(H_1 \cap H_2),\end{aligned}$$

bo zgodnie z Obserwacją 7.14 mamy $\dim(H_1 \cap H_2) = \dim(T(H_1) \cap T(H_2))$.

Dowodzimy tezę w przypadku, gdy $H_1 \cap H_2 = \emptyset$. Wówczas zgodnie z Obserwacją 7.13 mamy

$$(T(H_1) + T(H_2)) \cap \operatorname{lin}(\overrightarrow{p_1 p_2}) = 0,$$

a zatem zgodnie z Obserwacją 7.15 mamy

$$\begin{aligned}\dim(H_1 + H_2) &= \dim(T(H_1) + T(H_2) + \operatorname{lin}(\overrightarrow{p_1 p_2})) \\ &= \dim(T(H_1) + T(H_2)) + \dim(\operatorname{lin}(\overrightarrow{p_1 p_2})) \\ &= \dim T(H_1) + \dim T(H_2) - \dim(T(H_1) \cap T(H_2)) + 1 \\ &= \dim H_1 + \dim H_2 - \dim(T(H_1) \cap T(H_2)) + 1\end{aligned}$$

* * *

Zauważmy, że w drugiej części rozumowania nie można wymienić $\dim(T(H_1) \cap T(H_2))$ na $\dim(H_1 \cap H_2)$. Zauważmy, że biorąc w $\mathbb{R}^4$:

$$\begin{aligned}H_1 &= (0,0,0,0) + \operatorname{lin}((1,0,0,0),(0,1,0,0)) \\ H_2 &= (0,0,0,1) + \operatorname{lin}((0,1,0,0),(0,0,1,0))\end{aligned}$$

widzimy, że $H_1 \cap H_2 = \emptyset$, ale płaszczyzny te nie są równoległe. Mamy przy tym

$$\dim(H_1 + H_2) = 2 + 2 - \dim(T(H_1) \cap T(H_2)) + 1 = 5 - 1 = 4.$$

Ogólne możliwe konfiguracje par płaszczyzn w przestrzeni wymiaru 4 opisuje poniższy wniosek. Zachęcam do wskazania przykładów odpowiednich konfiguracji.

Wniosek 7.1

Jeśli dane są w przestrzeni K^4 przestrzenie afiniczne $H_1 = p_1 + T(H_1)$ oraz $H_2 = p_2 + T(H_2)$, przy czym $\dim H_1 = \dim H_2 = 2$, to:

1. Jeśli $H_1 \cap H_2 \neq \emptyset$, to $\dim(H_1 + H_2) = 2 + 2 - \dim(H_1 \cap H_2) = 4 - \dim(H_1 \cap H_2)$, przy czym
   - jeśli $\dim(H_1 \cap H_2) = 0$, to $H_1 \cap H_2$ jest punktem,
   - jeśli $\dim(H_1 \cap H_2) = 1$, to $H_1 \cap H_2$ jest prostą,
   - jeśli $\dim(H_1 \cap H_2) = 2$, to $H_1 = H_2$.
2. Jeśli $H_1 \cap H_2 = \emptyset$, to $\dim(H_1 + H_2) = 2+2-\dim(T(H_1) \cap T(H_2))+1 = 5-\dim(T(H_1) \cap T(H_2))$, przy czym
   - jeśli $\dim(T(H_1) \cap T(H_2)) = 1$, są to płaszczyzny skośne,
   - jeśli $\dim(T(H_1) \cap T(H_2)) = 2$, są to płaszczyzny równoległe.

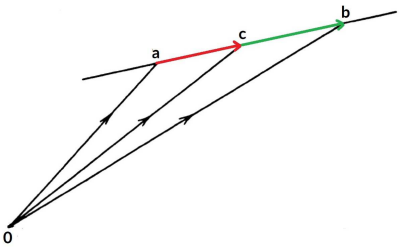
7.2 Dodatek. Współliniowość, współrzędne barycentryczne oraz intuicje i twierdzenia geometrii elementarnej

Przy pomocy rachunku wektorowego w przestrzeni afinicznej dowodzić można klasyczne twierdzenia geometrii[5]. Przyjrzymy się przede wszystkim pojęciu współliniowości w rzeczywistej przestrzeni afinicznej. Mówimy o trójkach punktów, które są albo w położeniu ogólnym, co najczęściej kwitujemy mówiąc, że tworzą one trójkąt niezdegenerowany, albo są współliniowe — są w położeniu szczególnym (w tym równe).

Obserwacja 7.16

Jeśli c jest dowolnym punktem na prostej wyznaczonej przez dwa różne punkty a, b w przestrzeni afinicznej nad $\mathbb{R}$ (czyli $\mathrm{af}(a,b)$), który jest od nich różny, to istnieje $\lambda \in \mathbb{R}$ takie, że $c = (1-\lambda)a+\lambda b$, przy czym zachodzi równość

$$\overrightarrow{cb} = \frac{\lambda}{1-\lambda} \cdot \overrightarrow{ac}.$$



Dowód. Skoro c jest kombinacją afiniczną punktów a, b, to niech $c = (1-\lambda)a+\lambda b$. Zatem $c-a = \lambda(b-a)$, co w konsekwencji daje $\overrightarrow{ac} = \lambda\overrightarrow{ab}$. Z drugiej strony $b-c = (1-\lambda)(b-a)$, czyli $\overrightarrow{cb} = (1-\lambda)\overrightarrow{ab}$. □

Spójrzmy na powyższą równość w inny sposób.

Obserwacja 7.17

Jeśli punkty a, b, c przestrzeni afinicznej nad $\mathbb{R}$ są współliniowe, to istnieją liczby $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$, nie wszystkie równe 0, spełniające warunki:

$$\lambda_1 + \lambda_2 + \lambda_3 = 0 \quad \text{oraz} \quad \lambda_1 a + \lambda_2 b + \lambda_3 c = 0.$$

Dowód. Jeśli $a = b = c$, to teza jest oczywista. Załóżmy, że a, b są różne. Wówczas zgodnie z poprzednią obserwacją możemy zapisać $c = (1-\lambda)a + \lambda b$, dla pewnego λ. Mamy zatem:

$$(1-\lambda)a + \lambda b + (-1)c = 0,$$

czyli biorąc $\lambda_1 = 1-\lambda$, $\lambda_2 = \lambda$ oraz $\lambda_3 = -1$ dostajemy tezę. □

Kluczowe jest odwrócenie powyższego faktu.

Obserwacja 7.18

Jeśli dane są punkty a, b, c oraz liczby rzeczywiste $\lambda_1, \lambda_2, \lambda_3$, nie wszystkie równe 0, dla których $\lambda_1 + \lambda_2 + \lambda_3 = 0$ oraz $\lambda_1 a + \lambda_2 b + \lambda_3 c = 0$, wówczas punkty a, b, c są współliniowe.

Dowód. Ewentualnie zmieniając nazwy punktów możemy założyć, że $\lambda_1 \neq 0$. Możemy zapisać:

$$a = -\frac{\lambda_2}{\lambda_1} b - \frac{\lambda_3}{\lambda_1} c.$$

Skoro $\lambda_1 + \lambda_2 + \lambda_3 = 0$, to $-\frac{\lambda_2}{\lambda_1} - \frac{\lambda_3}{\lambda_1} = 1$. A zatem kładąc $\lambda = -\frac{\lambda_3}{\lambda_1}$, to $1-\lambda = -\frac{\lambda_3}{\lambda_1}$ uzyskujemy

$$a = (1-\lambda)b + \lambda c.$$

Jeśli b, c nie są różne, to oczywiście a, b, c są współliniowe. Jeśli natomiast b, c są różne, to $a \in \mathrm{af}(b,c)$. □

[5]Tekst na podstawie książki D. Pedoe, Geometry, a comprehensive course, Dover Publications 1988.

Wniosek 7.2

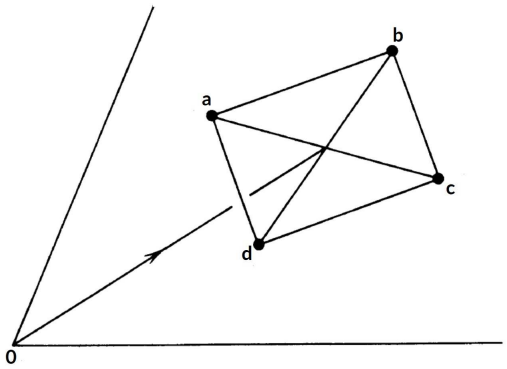
Jeśli a, b, c są niewspółliniowymi punktami oraz jeśli istnieją liczby rzeczywiste λ_1, λ_2 oraz λ_3 takie, że $\lambda_1 + \lambda_2 + \lambda_3 = 0$ oraz $\lambda_1 a + \lambda_2 b + \lambda_3 c = 0$, to $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

Zobaczmy kilka zastosowań.

Obserwacja 7.19

Załóżmy, że a, b, c, d są parami różne. Następujące warunki są równoważne:

- czworokąt a, b, c, d jest równoległobokiem, tzn. $\overrightarrow{ab} = \overrightarrow{dc}$ (równoważnie $\overrightarrow{ad} = \overrightarrow{bc}$),
- $\mathrm{af}(a, c) \cap \mathrm{af}(b, d) = \frac{1}{2}a + \frac{1}{2}c = \frac{1}{2}b + \frac{1}{2}d$, tzn. przekątne czworokąta $abcd$ przecinają się w połowie.



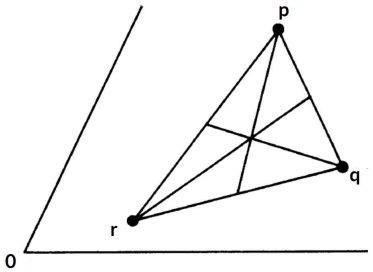
Dowód. Mamy równoważności $\overrightarrow{ab} = \overrightarrow{dc} \iff b - a = c - d \iff b + d = c + a \iff \frac{1}{2}b + \frac{1}{2}d = \frac{1}{2}c + \frac{1}{2}a$. □

Obserwacja 7.20

W trójkącie o wierzchołkach p, q, r środkowe przecinają się w jednym punkcie

$$\frac{1}{3}p + \frac{1}{3}q + \frac{1}{3}r,$$

który dzieli je w stosunku $1 : 2$.



Dowód. Kładąc $\lambda = \frac{2}{3}$ we wzorze $z = (1 - \lambda)x + \lambda y$ widzimy, że punkt, który dzieli wektor $\overrightarrow{xy}$ w stosunku $2 : 1$ to punkt $z = \frac{1}{3}x + \frac{2}{3}y$. Środkiem odcinka qr jest $p' = \frac{1}{2}q + \frac{1}{2}r$ i punkt, który dzieli wektor $\overrightarrow{pp'}$ w stosunku $1 : 2$ to

$$\frac{2}{3}\left(\frac{q + r}{2}\right) + \frac{p}{3} = \frac{p + q + r}{3}.$$

Symetria tego rozumowania daje zatem tezę. □

Przejdziemy teraz do znanego faktu konkursowego, zapewne niepokazywanego w szkole.

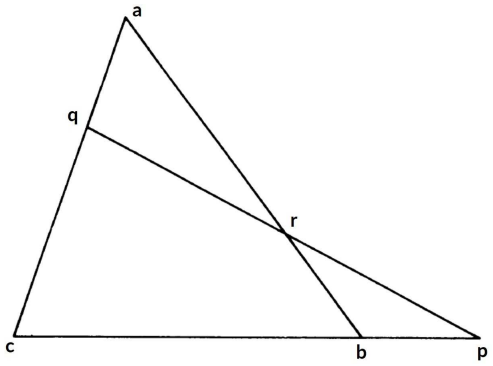
Twierdzenie 7.2: Menelaos

Pewna prosta przecina proste af(b,c), af(c,a), af(a,b) (zawierające boki trójkąta abc) w punktach p,q,r. Wówczas jeśli dla pewnych $x_1,x_2,x_3,y_1,y_2,y_3 \in \mathbb{R}$ spełniających $x_1+y_1=1, x_2+y_2=1$ oraz $x_3+y_3=1$ mamy:

$$p = x_1 b + y_1 c, \quad q = x_2 c + y_2 a, \quad r = x_3 a + y_3 b,$$

wówczas $x_1x_2x_3 = -y_1y_2y_3$, czyli równoważnie[a]:

$$\frac{\overrightarrow{bp}}{\overrightarrow{pc}} \cdot \frac{\overrightarrow{cq}}{\overrightarrow{qa}} \cdot \frac{\overrightarrow{ar}}{\overrightarrow{rb}} = -1.$$



[a]Porównujemy stosunki podziałów wektorów, czyli pisząc $\lambda = \frac{\vec{a}}{\vec{b}}$, mamy na myśli $\vec{a} = \lambda \cdot \vec{b}$,

Dowód. Skorzystamy z Obserwacji 7.17. Skoro p,q,r są współliniowe, to istnieją liczby rzeczywiste $\lambda_1,\lambda_2,\lambda_3$, nie wszystkie równe 0, takie że $\lambda_1+\lambda_2+\lambda_3=0$ oraz $\lambda_1 p + \lambda_2 q + \lambda_3 r = 0$. Stąd

$$\lambda_1(x_1b+y_1c) + \lambda_2(x_2c+y_2a) + \lambda_3(x_3a+y_3b) = 0,$$

czyli równoważnie

$$(\lambda_1x_1+\lambda_3y_3)b + (\lambda_2x_2+\lambda_1y_1)c + (\lambda_3x_3+\lambda_2y_2)a = 0.$$

Zauważmy, że suma współczynników powyższej kombinacji to

$$(\lambda_1x_1+\lambda_3y_3)+(\lambda_2x_2+\lambda_1y_1)+(\lambda_3x_3+\lambda_2y_2) = \lambda_1(x_1+y_1)+\lambda_2(x_2+y_2)+\lambda_3(x_3+y_3) = \lambda_1+\lambda_2+\lambda_3 = 0.$$

Biorąc pod uwagę, że punkty a,b,c nie są współliniowe jako wierzchołki trójkąta (układ afinicznie niezależny), na mocy Wniosku 8.1 mamy zatem

$$\lambda_1x_1+\lambda_3y_3 = \lambda_2x_2+\lambda_1y_1 = \lambda_3x_3+\lambda_2y_2 = 0.$$

Stąd

$$(\lambda_1x_1)(\lambda_2x_2)(\lambda_3x_3) = (-\lambda_3y_3)(-\lambda_1y_1)(-\lambda_2y_2).$$

Pozostaje rozważyć kilka przypadków.

- Jeśli żadna z liczb $\lambda_1,\lambda_2,\lambda_3$ nie jest równa 0, to od razu dostajemy tezę $x_1x_2x_3 = -y_1y_2y_3$.
- Skoro nie wszystkie z liczb $\lambda_1,\lambda_2,\lambda_3$ mogą być równe 0, to nie więcej niż jedna jest równa 0. Załóżmy, że $\lambda_1=0$. Wówczas $\lambda_2+\lambda_3=0$, to $\lambda_2=-\lambda_3\neq 0$, i z równości $\lambda_2q+\lambda_3r=0$, więc $q=r$. Zakładamy jednak, że p,q,r są różne. Inaczej bowiem jego teza jest oczywista.

□

Zachęcam Czytelnika do zastanowienia się nad możliwymi konfiguracjami punktów p,q,r. Niekoniecznie leżą one na bokach trójkąta a,b,c. Proszę pomyśleć też o sformułowaniu twierdzenia odwrotnego i spróbować je udowodnić.

Współrzędne barycentryczne punktu w bazie punktowej przestrzeni afinicznej pozwalają wprowadzić swego rodzaju wewnętrzny układ współrzędnych na przestrzeni afinicznej. Przyjrzymy się temu w przypadku rozważanej już sytuacji, gdy dany jest układ trzech punktów w położeniu ogólnym — trójkąt abc, którego wierzchołki określają nowy „układ współrzędnych". Przypomnijmy fakt z wykładu w tymże języku.

Obserwacja 7.21

Jeśli punkty a, b, c rzeczywistej przestrzeni afinicznej są niewspółliniowe (są afinicznie niezależne), to każdy punkt p płaszczyzny $\mathrm{af}(a, b, c)$ można zapisać w postaci

$$p = \lambda_1 a + \lambda_2 b + \lambda_3 c, \qquad \text{gdzie } \lambda_1 + \lambda_2 + \lambda_3 = 1.$$

Dla danego punktu p liczby $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ są wyznaczone jednoznacznie.

Pomyślmy o naszych współrzędnych jak o trójkach postaci $(\lambda_1 : \lambda_2 : \lambda_3)$. Wówczas mamy

$$a = (1 : 0 : 0), \quad b = (0 : 1 : 0), \quad c = (0 : 0 : 1).$$

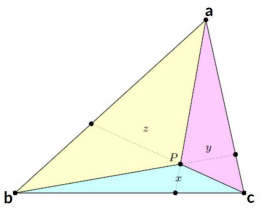
Czy wiemy coś jeszcze? Owszem, środek odcinka ab to $(\frac{1}{2} : \frac{1}{2} : 1)$, zaś środek odcinka bc to $(0 : \frac{1}{2} : \frac{1}{2})$. Jeśli Czytelnikowi wydaje się, że to mało, zachęcam do sprawdzenia[6], że również inne punkty szczególne w trójkącie wyrażają się elegancko w tych współrzędnych. Jeśli $|a|, |b|, |c|$ to długości boków trójkąta abc, zaś α, β, γ to miary kątów wewnętrznych w tym trójkącie (kąt α naprzeciw a itd.) to:

- współrzędne środka ciężkości trójkąta abc to $\left(\frac{1}{3} : \frac{1}{3} : \frac{1}{3}\right)$,
- współrzędne środka okręgu wpisanego to $\left(\frac{|a|}{|a|+|b|+|c|} : \frac{|b|}{|a|+|b|+|c|} : \frac{|c|}{|a|+|b|+|c|}\right)$,
- współrzędne przecięcia wysokości (ortocentrum) to $\left(\frac{\operatorname{tg}\alpha}{\operatorname{tg}\alpha+\operatorname{tg}\beta+\operatorname{tg}\gamma} : \frac{\operatorname{tg}\alpha}{\operatorname{tg}\beta+\operatorname{tg}\beta+\operatorname{tg}\gamma} : \frac{\operatorname{tg}\gamma}{\operatorname{tg}\alpha+\operatorname{tg}\beta+\operatorname{tg}\gamma}\right)$.

Te współrzędne nie wyglądają na pierwszy rzut oka zachęcająco, więc być może warto powiedzieć kilka słów o ich genezie związanej z masami, ciężarami i wyporami, pochodzącej od Ferdinanda Möbiusa. Cytujemy tu świetny artykuł prof. Marka Kordosa[7]. Pomyślmy jakie ciężary należy umieścić w wierzchołkach trójkąta, aby jego środek ciężkości znalazł się we wskazanym punkcie w jego wnętrzu? To może dziwne pytanie, ale jeśli postawimy analogiczne pytanie dla odcinka o końcach w punktach (różnych) a, b i zadamy punkt x, na którym znajdować się ma równowaga po przyłożeniu odpowiednio do a, b mas m_a oraz m_b, to zachodzić musi znane z fizyki prawo (Archimedesa) dźwigni $m_a \cdot \overrightarrow{ax} = m_b \cdot \overrightarrow{xb}$, czyli $\frac{m_a}{m_b} = \frac{\overrightarrow{xb}}{\overrightarrow{ax}}$.

Można tu pytać jak interpretować ujemne współczynniki? Czy chodzi o masy ujemne? Już Archimedes wiedział, że efektywna siła ciężkości może działać zarówno w dół, jak i do góry. Ta druga sytuacja ma miejsce na przykład przy unoszeniu się balonu, albo podczas unoszenia się na powierzchni wody statku. Owe „ujemne ciężary" można nazwać *wyporem*. Tu dochodzimy do związku naszych zagadnień z wypukłością. Otóż okazuje się, że punkt o wszystkich nieujemnych współrzędnych barycentrycznych leżeć musi wewnątrz lub na obwodzie trójkąta abc, a więc jest elementem powłoki wypukłej $\mathrm{conv}(a, b, c)$.

Chcąc jeszcze bardziej namieszać Czytelnikowi w głowie mogę dodać (choć jeszcze nie wiemy co to jest pole), że współrzędne barycentryczne nazywano kiedyś współrzędnymi arealnymi. Wiąże się to z faktem, że punkt p o współrzędnych barycentrycznych $(\lambda_1 : \lambda_2 : \lambda_3)$ wyznacza trójkąty pcb, pba, pac o polach (skierowanych, cokolwiek to znaczy), których stosunek do pola trójkąta abc to odpowiednio $\lambda_1, \lambda_2, \lambda_3$.



Spróbujmy przekonać Czytelnika, że współrzędne barycentryczne mogą mu się do czegoś przydać.

[6]... lub przeczytania, np. pod adresem `https://web.evanchen.cc/handouts/bary/bary-full.pdf`.
[7]`https://www.deltami.edu.pl/temat/matematyka/geometria/planimetria/2012/02/29/wypory.pdf`.

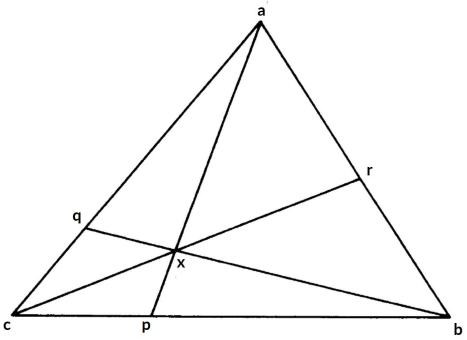
Twierdzenie 7.3: Ceva

Dany jest trójkąt abc i punkt $x \in \mathrm{af}(a,b,c)$. Proste $\mathrm{af}(a,x), \mathrm{af}(b,x), \mathrm{af}(c,x)$ przecinają boki bc, ca, ab trójkąta abc w punktach p, q, r. Niech $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$ będą takimi liczbami spełniającymi $x_1 + y_1 = 1, x_2 + y_2 = 1$ oraz $x_3 + y_3 = 1$, że:

$$p = x_1 b + y_1 c, \quad q = x_2 c + y_2 a, \quad r = x_3 a + y_3 b.$$

Wówczas $x_1 x_2 x_3 = y_1 y_2 y_3$, czyli równowaznie:

$$\frac{\overrightarrow{bp}}{\overrightarrow{pc}} \cdot \frac{\overrightarrow{cq}}{\overrightarrow{qa}} \cdot \frac{\overrightarrow{ar}}{\overrightarrow{rb}} = 1.$$



Dowód. Zapiszmy $x = \lambda_1 a + \lambda_2 b + \lambda_3 c$, gdzie $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Możemy również napisać:

$$x = \lambda_1 a + (\lambda_2 + \lambda_3) y,$$

gdzie

$$y = (\lambda_2 b + \lambda_3 c)/(\lambda_2 + \lambda_3).$$

Na mocy pierwszego z równań wyżej punkty x, a, y są współliniowe, zaś na mocy drugiego — punkty y, b, c są współliniowe. Zatem punkt y leży na ap oraz leży na bc. Zatem mamy $y = p = x_1 b + y_1 c$. Stąd:

$$\frac{x_1}{y_1} = \frac{\lambda_1}{\lambda_2}.$$

Stąd natychmiast wynika teza. □

Warto pokazać jeszcze pewien alternatywny dowód, pokazujący więcej narzędzi stosowanych w rozwiązywaniu zadań olimpijskich[8]. Zachęcam Czytelnika do przeprowadzenia dowodu poniższej obserwacji.

Obserwacja 7.22

Rozważmy płaszczyznę $\mathrm{af}(a,b,c)$, gdzie a, b, c są afinicznie niezależne. Wówczas dla każdego układu liczb u, v, w, gdzie $(u, v, w) \neq (0, 0, 0)$ zbiór punktów o współrzędnych barycentrycznych $(\lambda_1, \lambda_2, \lambda_3)$ takich, że

$$u\lambda_1 + v\lambda_2 + w\lambda_3 = 0 \qquad (*)$$

jest prostą.

Aby Czytelnik poczuł ideę, zauważmy, że równanie prostej zawierającej punkty a, b to po prostu $\lambda_3 = 0$, bowiem równanie $(*)$ muszą spełniać punkty $a = (1:0:0)$ oraz $b = (0:1:0)$. Wracając do twierdzenia Cevy i jego dowodu — zachęcam Czytelnika do pokazania, że jeśli $p = (0, d, 1-d)$, $q = (1-e, 0, e)$ oraz $r = (f, 1-f, 0)$, wówczas proste $\mathrm{af}(a,p)$, $\mathrm{af}(b,q)$, $\mathrm{af}(c,r)$ mają równania odpowiednio

$$\lambda_3 = \frac{1-d}{d}\lambda_2, \quad \lambda_1 = \frac{1-e}{e}\lambda_3, \quad \lambda_2 = \frac{1-f}{f}\lambda_1.$$

Kiedy te proste mają punkt wspólny (o pewnych współrzędnych barycentrycznych)?

[8] Nie sposób nie odesłać tu do Evana Chena: `https://web.evanchen.cc/handouts/bary/bary-full.pdf`.

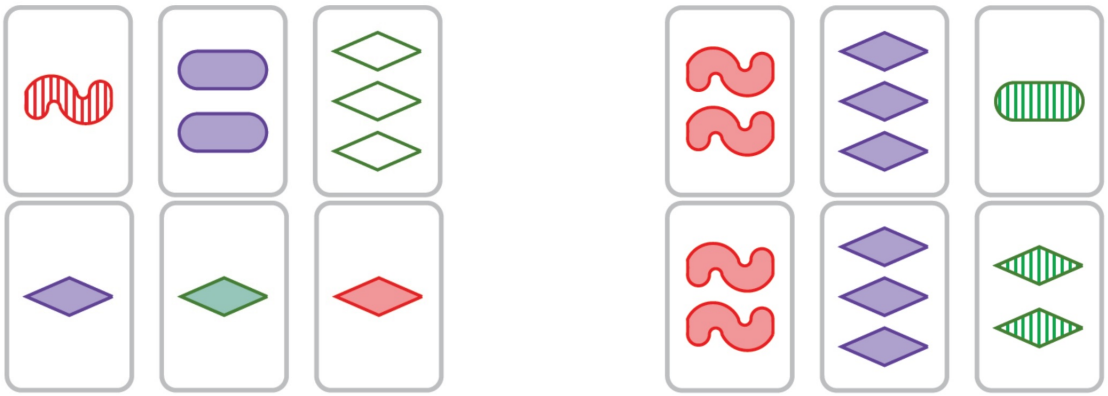
7.3 Trivia. SET, czyli przestrzeń afiniczna nad $\mathbb{Z}_3$

SET jest szybką i dość wciągającą grą karcianą polegającą na rozpoznawaniu prawidłowości, bardzo popularną wśród studentów i entuzjastów matematyki[9]. Zadaniem jest identyfikowanie trzech kart z obiektami, które spełniają określone zasady – taki zbiór to właśnie Set.

Zasady[10] są następujące. Dana jest talia 81 kart. Każda karta jest określona przez cztery cechy, z których każda ma jedną z trzech możliwych wartości zgodnie z poniższą tabelą (sprawę $\mathbb{Z}_3$ wyjaśnimy dalej).

cecha/element $\mathbb{Z}_3$	0	1	2
liczba	raz	dwa	trzy
wypełnienie	puste	częściowe	pełne
kolor	czerwony	zielony	fioletowy
kształt	romb	elipsa	fala

Na początku rozgrywki rozkładamy na stole 12 kart (3 rzędy po 4 kolumny). Celem jest zebranie **Setów**, czyli 3 kart na których każda pojedyncza cecha jest różna lub taka sama. Kiedy któryś z graczy go zauważy mówi głośno „Set", gra się zatrzymuje i sprawdzamy czy jest on poprawny. Każdy Set jest wart 3 punkty, jednak kiedy wytypujemy zły Set tracimy jedną kartę. Następnie karty uzupełniane są do 12.



Po lewej – dwa Sety, po prawej – dwa nie-Sety.

Proszę zauważyć, że jeśli mamy dowolne dwie karty, to istnieje dokładnie jedna karta w zestawie, która tworzy z nimi Set. To się kojarzy z własnością:

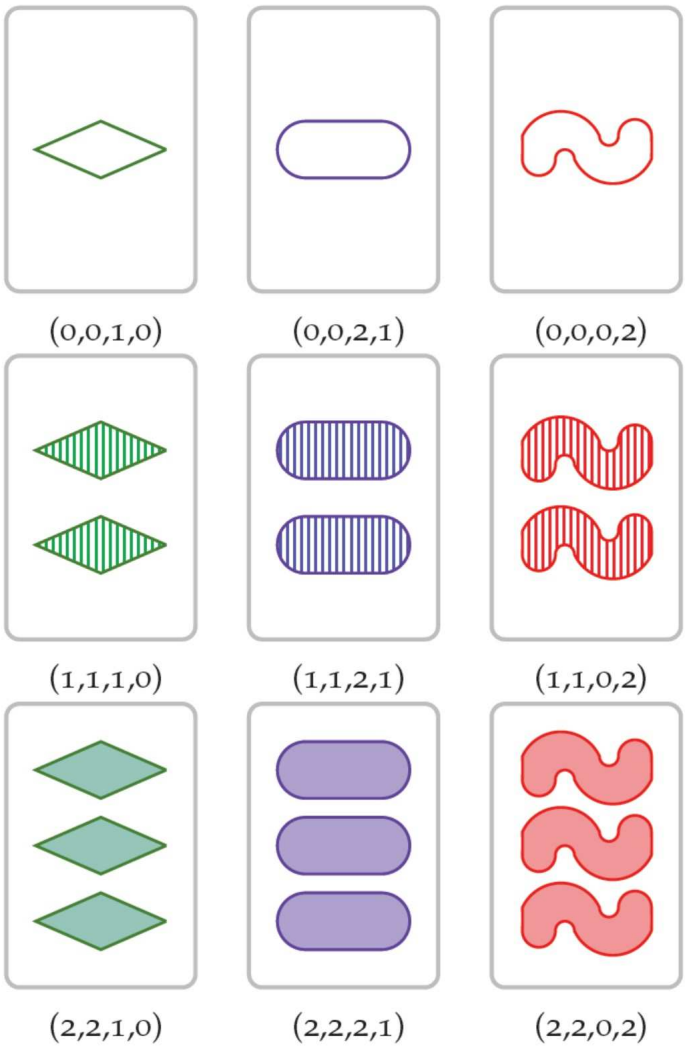
Przez dowolne dwa punkty można przeprowadzić prostą.

Pomysł jest taki, że rozważamy przestrzeń $\mathbb{Z}_3^4$ traktowaną jako przestrzeń afiniczna i w tej przestrzeni karty są punktami, a Sety – prostymi. Jak to działa? Mamy po prostu punkty o współrzędnych (x_1, x_2, x_3, x_4), które oznaczają kolejno: liczbę, wypełnienie, kolor, kształt, a liczby $0, 1, 2$ z ciała $\mathbb{Z}_3$ przypisujemy kolejno tak jak w tabeli wyżej. A zatem np. elementowi $(0, 1, 0, 2)$ przypisujemy kartę:

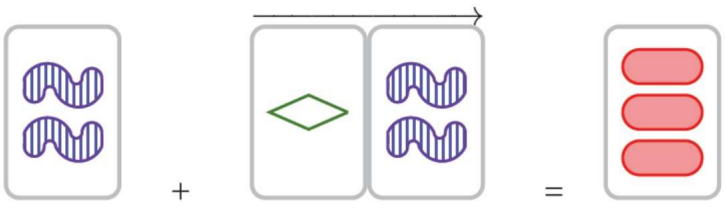
[9] Podstawą merytoryczną jest artykuł *Algebra From Geometry in the Card Game SET* autorstwa Timothy E. Goldberga, The College Mathematics Journal , Vol. 47, No. 4, pp. 265-273 oraz materiały prof. J. Wiśniewskiego

[10] Zasady przytoczone za stroną `https://boardtime.pl/2012/11/set-recenzja.html`. W Seta można zagrać także online: `https://smart-games.org/en/set/submit_set`.

Wyobraźmy sobie teraz np. zbiór rozwiązań układu równań liniowych $x_1 - x_2 = 0, x_3 - x_4 = 1$. Jest to warstwa dwuwymiarowej przestrzeni liniowej nad $\mathbb{Z}_3$, a dokładniej zbiór 9 kart. Poniżej znajdują się te karty i odpowiadające im punkty. Wyraźnie widzimy, że jest to płaszczyzna afiniczna w przestrzeni $\mathbb{Z}_3^4$ złożona z 12 prostych (proszę wskazać).



Zauważmy też, że mamy operacje w tej przestrzeni, na przykład dodawanie do punktów wektorów:



$$(1,1,2,2) \quad + \quad \overrightarrow{(0,0,1,0)(1,1,2,2)} \quad = \quad (2,2,0,1)$$

Nie chcę przywoływać rozważań teoretycznych[11] dotyczących gry Set, ale mam nadzieję, że to spojrzenie uprzyjemni Państwu myślenie o przestrzeniach afinicznych także w kontekście matematyki dyskretnej. Przede wszystkim natomiast polecam Set — jakże popularny wśród matematyków.

[11] No dobrze, przywołam jedno: ile można wyłożyć kart, by nie było wśród nich żadnego Seta? Odpowiedź w artykule: `http://homepages.warwick.ac.uk/staff/D.Maclagan/papers/set.pdf`

7.4 Coda. Geometria wewnętrzna i zewnętrzna

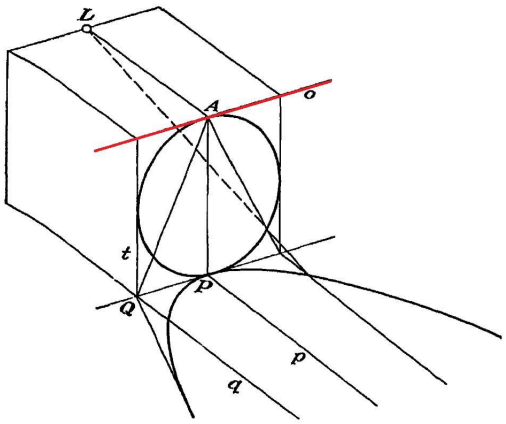
Pojęcie przestrzeni afinicznej jest na tyle istotnym i delikatnym punktem naszego programu, że warto poświęcić mu nieco mniej matematyczny fragment naszych rozważań, niebędący jednak tylko dodatkiem historycznym, ale tekstem bliższym wykładowi. Koncepcja przestrzeni afinicznej i jej związków z otaczającym nas światem jest punktem zwrotnym kursu. Zacznijmy od idei prostej, podkreślonej pięknie przez jednego z polskich autorów, Jacka Komorowskiego, w świetnej (i bardzo polecanej) książce *Od liczb zespolonych do tensorów, spinorów, algebr Liego i kwadryk*, gdzie pojawia się następujący argument[12].

> Nie wydaje się naturalne przyjmowanie przestrzeni wektorowej jako ośrodka, w którym będzie uprawiana geometria. Taki wybór, aczkolwiek wygodny ze względów rachunkowych, jest niefortunny ze światopoglądowego punktu widzenia; przecież w otaczającym nas świecie, którego modelem ma być geometria, brak jest np. wyróżnionego punktu, „pępka świata", jakim w przestrzeni wektorowej jest zero.

Warto napisać kilka zdań o geometrycznych źródłach geometrii afinicznej, będących w zasadzie tłumaczeniem (bez większego komentarza) fragmentu rozważań jednego z głównych autorytetów XX-wiecznej geometrii – H. S. M. Coxetera. Pochodzą one z książki *The Real and Projective Plane*.

Zwykła geometria, którą poznawaliśmy w szkole, zajmowała się okręgami, kątami, prostymi równoległymi, trójkątami podobnymi itd., zwana jest geometrią euklidesową[13], ponieważ została ona po raz pierwszy usystematyzowana przez *Elementy* Euklidesa, żyjącego ok. 300 r. p.n.e. Do dziś, przy odpowiednim języku i redakcji, jest to pozycja zawierająca wzorcowe podejście ilustrujące stosowanie metody dedukcyjnej[14]. W XIX wieku pojawiła się tendencja do wyciągania z geometrii euklidesowej pewnych fundamentalnych idei, zwłaszcza takich, które nie wymagają mierzenia odległości lub kątów, i do wykorzystywania tych idei do budowania ogólniejszych spojrzeń na geometrię.

Skąd się bierze nazwa geometrii afinicznej? Weźmy dwie figury znajdujące się na dwóch różnych płaszczyznach. Powiemy, że jedną figurę można uzyskać z drugiej przez rzut równoległy, jeśli odpowiednie punkty znajdujące się na tych figurach można połączyć za pomocą równoległych prostych. To jest ta sytuacja, która ma miejsce gdy słońce rzuca cień na ziemię, np. okrągła moneta rzuca eliptyczny cień. Jeśli dwie płaszczyzny są równoległe, dwie figury będą przystające; w przeciwnym przypadku będą miały wprawdzie nieco inne kształty, ale linie proste pozostaną proste, styczne do krzywych pozostaną stycznymi, proste równoległe pozostaną równoległe, odcinki przecięte na pół pozostaną takie po rzutowaniu itd.



Słowo „afiniczne" (*affinis* – z łac. podobieństwo, powinowactwo) zostało użyte po raz pierwszy przez Eulera w drugim tomie *Introductio in Analysis Infinitorum* z 1748 roku, w kontekście stycznych do krzywej. Niedługo później zostało podjęte przez Möbiusa i Monge'a, który dali początek ogólniejszej teorii – geometrii rzutowej. Stawia ona mniej restrykcyjne wymagania, ograniczając się do tych własności, które nie zmieniają się przy rzucie środkowym. Tu pojawiają się dodatkowe wyzwania, bo nie każda prosta przechodzi na prostą, tak jak czerwona prosta na rysunku powyżej.

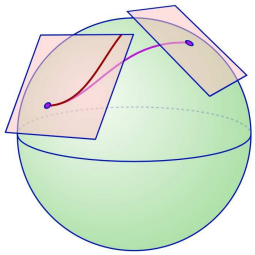
---

[12] Argument ten stosowany również przez inne nauki, na przykład fizyczne, o czym wspomina znakomity podręcznik prof. A. Herdegena z Uniwersytetu Jagiellońskiego: `http://eigenspace.pl/herdegen_algebra.pdf`.

[13] Za kilka wykładów poznamy tzw. przestrzenie euklidesowe, która mają być bezpośrednim obiektem realizującym geometrię euklidesową w kontekście przestrzeni liniowych.

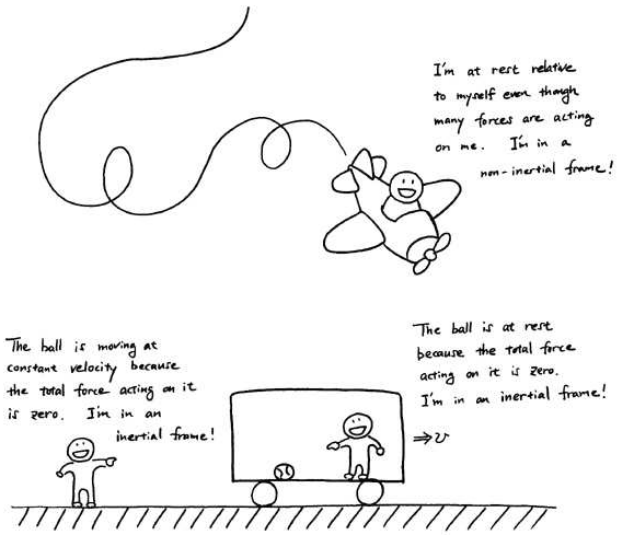
[14] Polecam gorąco kolorowe opracowanie on-line `https://www.c82.net/euclid/`.

Na wykładzie zdefiniowaliśmy przestrzenie afiniczne będące warstwami w przestrzeniach liniowych. Jest to podejście bliskie naszej geometrycznej intuicji i dogodne w licznych zastosowaniach – nie jest ono jednak wystarczające. Jeden z powodów jest następujący: w przyszłości będziecie Państwo badać obiekty nieliniowe i ich własności przypisując im „lokalnie" pewne obiekty liniowe i afiniczne.



Nie zawsze wygodne będzie zanurzanie owych nieliniowych obiektów w przestrzeń liniową z jednym ustalonym „układem współrzędnych" i rozpatrywanie w ramach tej przestrzeni złożonych problemów analitycznych. Stąd potrzeba, by móc w jakiś alternatywny sposób określić przestrzeń afiniczną. Przywołamy dwa podstawowe podejścia Jedno z nich oparte jest o wprowadzoną już abstrakcyjną definicję aksjomatyczną przestrzeni afinicznej, a drugie oparte jest o ideę „obiektów niezmienniczych". Jest to stosowny moment, by na podstawie naszych dotychczasowych doświadczeń opowiedzieć Państwu o różnych spojrzeniach na zagadnienia geometryczne. Będzie to również zapowiedź tego, że na kolejnych latach studiów matematycznych poznamy problemy geometryczne „z wyższego punktu widzenia".

W geometrii afinicznej badamy obiekty odnoszące się silnie do „rzeczywistości". Bardziej niż w przypadku przestrzeni liniowych, strukturze afinicznej przypisujemy miano „geometrii świata" – co prowadzi naturalnie do wielu trudnych pytań. Rzecz jasna aksjomatyka przestrzeni liniowych czerpie z intuicji fizycznych, ale pojawiają się pewne ograniczenia. Związane są one ze sformułowaną przez Galileusza tzw. *zasadą względności* mówiącą poglądowo, że prawa fizyki w dwóch inercjalnych układach odniesienia są takie same. Choć w różnych układach odniesienia prawa fizyki mają być te same, ich „postrzeganie" i „opisywanie" stawia wielkie matematyczne wyzwania. Opis zjawisk fizycznych chętnie czerpał z geometrii analitycznej Kartezjusza i Fermata, i możliwości przedstawiania obiektów w układach współrzędnych, a następnie pracowicie tworzył sposoby przechodzenia z jednego układu odniesienia do drugiego. W tym kluczu możemy między innymi interpretować przestrzenie liniowe i ich endomorfizmy – przechodząc z macierzy przekształcenia w jednej bazie do innej „zmieniamy układ odniesienia". Zawsze jednak istniało napięcie związane z pytaniem: czy istnieje „uniwersalny układ inercjalny", albo z jego nieprecyzyjną, ale obrazową wersją: czy ktokolwiek we Wszechświecie jest w stanie stwierdzić czy się porusza, czy nie?

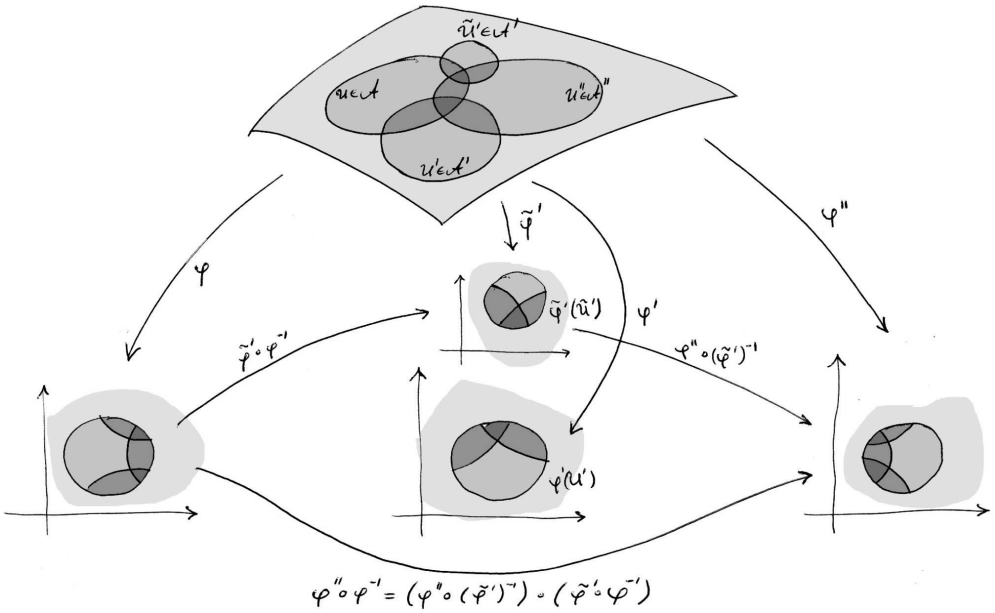


Źródło: `https://www1.phys.vt.edu/~takeuchi/relativity/notes/section02.html`. Polecam inne świetne rysunki na tej stronie.

Zarówno podejście aksjomatyczne do geometrii afinicznej, jak i podejście oparte na traktowaniu przestrzeni afinicznych jako warstwy ma pewne zauważalne już wady. W ujęciu aksjomatycznym wydajemy się mieć za dużą swobodę, mogąc choćby tak jak w problemie pierwszym przypisać strukturę afiniczną niemal dowolnemu obiektowi. W ujęciu „warstwowym" – zasadniczo nie tracimy nic ze swobody, jaką daje nam podejście aksjomatyczne[15], ale trafiamy do drugiego ekstremum: musimy traktować „liniowo" obiekty, które jako takie nie mają natury liniowej. Tak było w przypadku paraboli opisanej w ostatniej części wykładu. Parabola jest jednak niedoskonałym przykładem, bo choć nie jest warstwą, to jednak jej „naturę" odczytujemy stąd, że jest opisana w afinicznym układzie współrzędnych na płaszczyźnie. Gdybyśmy nie wiedzieli, że P składa się z punktów postaci (x, x^2), słabiej byśmy oponowali przeciwko myśleniu o $(P, \mathrm{lin}(1,0), \boxplus)$ jako o prostej afinicznej.

Problem ten rozważa się także w ogólnej sytuacji: mamy pewien zbiór X i nie mamy w nim struktury liniowej. Czasem nie chcemy myśleć o naszym obiekcie jako zanurzonym w przestrzeni liniowej z układem współrzędnych, nawet w ten sposób, w jaki parabola P z problemu pierwszego była związana z $\mathbb{R}^2$. Naprawdę nie traktujemy X jako części większej całości tylko jako obiekt sam w sobie. Co można powiedzieć o jego geometrii? To wbrew pozorom niezwykle ważne pytanie, zwłaszcza gdy ze zbiorem X związane są pewne funkcje, przez których badanie możemy wypowiadać się o naturze geometrycznej tego zbioru. Pięknie o tym pisze prof. Toruńczyk w swoich notatkach (`https://duch.mimuw.edu.pl/~torunczy/GAL/011-12/Wyk/VIII-AFI.pdf`), mówiąc o intuicji pojęcia, zwanego ROZMAITOŚCIĄ.

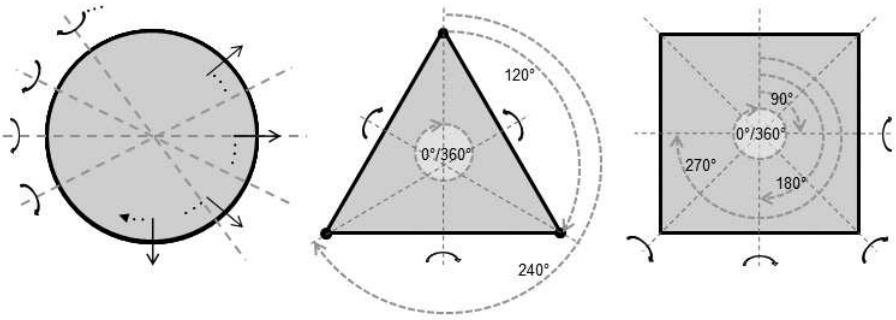
> „Przypuśćmy, że grono osób bada niedostępny obiekt $\mathbb{A}$ (np. powierzchnię niewidocznej strony Księżyca), mając do dyspozycji pewien zbiór jego map, rozdzielonych między badających. Jeśli ktoś zaproponuje nazwanie zbioru $X \subseteq \mathbb{A}$ elipsą, gdy na jego mapie zbiór ten jest odzwierciedlony jako elipsa, to natychmiast pojawia się pytanie, czy na innych mapach też jest on tak odzwierciedlony. Oznaczmy przez S_i przekształcenie, które punktowi zbioru $\mathbb{A}$ przyporządkowuje jego obraz na i-tej mapie. Wówczas $S_j S_i^{-1}$ przyporządkowuje punktom i-tej mapy odpowiadające im punkty mapy j-tej. Możliwość uzgodnienia przez badających, jakie zbiory $\mathbb{A}$ nazwać elipsami, zależy od tego, czy wszystkie PRZEKSZTAŁCENIA ZMIANY MAP przeprowadzają elipsy na elipsy. Ogólniej, badający zdołają uzgodnić te pojęcia dotyczące obiektu $\mathbb{A}$, które są niezmiennicze względem wszystkich tych przekształceń. (Zostawiamy tu pewne niedopowiedzenie.) [...] Materiał ten jest ważny i dlatego, że stwarza pewien zakres pojęć umożliwiających badanie przestrzeni, w której żyjemy – co było zadaniem geometrii od jej początków. Zadanie to o tyle sobie ułatwiamy, że istnienie odpowiednio zgodnych map przyjmujemy, podczas gdy ważnym osiągnięciem geometrii klasycznej było odkrycie ich istnienia i wskazanie sposobów konstrukcji."



Rysunek ilustrujący rozważanie powyżej. Źródło: `http://www.few.vu.nl/~vdvorst/notes-2012.pdf`.

[15] Każdej przestrzeni afinicznej zdefiniowanej przez aksjomaty można w sposób jednoznaczny przypisać izomorficzną z nią – czyli identyczną pod względem abstrakcyjnej struktury afinicznej – warstwę pewnej podprzestrzeni. Służy temu konstrukcja tzw. przestrzeni uniwersalnej.

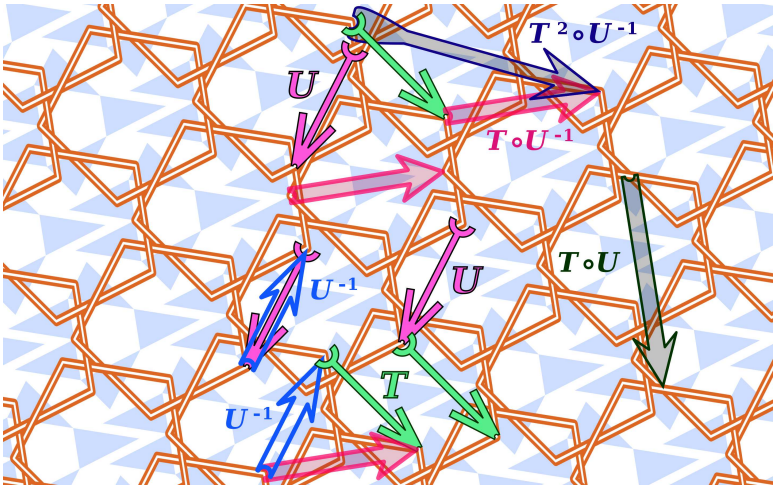
Zaczęliśmy rozważania tego wykładu od nadania paraboli w sposób sztuczny struktury afinicznej przez przypisanie jej jednej tylko „mapy”. To skrajny przykład zaprzeczający idei powyższych rozważań. „Poprawne” nadanie paraboli P zestawu funkcji $S_i : P \to \mathbb{R}^2$, które jednoznacznie **wyróżniają ją** w zbiorze innych (pewnych) podzbiorów $\mathbb{R}^2$, opisanych innymi zestawami map (i opisanie sposobów na porównywanie „zestawów map”) – to jest właściwy problem. Unikam tu jak mogę precyzyjnego języka – kiedyś na analizie i geometrii różniczkowej poznacie Państwo szczegóły. Powiem tylko, że podejście to pozwala na jeszcze inną definicję przestrzeni afinicznej wymiaru n nad ciałem K – mianowicie jako zbioru $\mathbb{A}$ z wyróżnioną rodziną bijekcji $M_{\mathbb{A}}$ idących z $\mathbb{A}$ do K^n takich, że dla $f_1, f_2 \in M_{\mathbb{A}}$ odwzorowanie $f_1 f_2^{-1}$ należy do przestrzeni przekształceń afinicznych $A(K^n, K^n)$. Podejście to, opisane w notatkach prof. Toruńczyka, pozwala widzieć rozmaite obiekty geometryczne jako „niezmiennicze” przy działaniu pewnych „grup przekształceń”. Sposób ten promował F. Klein w tzw. programie erlangeńskim (1872 r.). Nie sposób wejść w szczegóły, ale chciałbym przynajmniej powiedzieć co rozumieć przez „obiekty niezmiennicze”.



Izometrie własne okręgu, trójkąta równobocznego i kwadratu. Źródło: Michael J Richardson.

W matematyce możemy badać obiekty poprzez ich zbiory przekształceń, na przykład zbiory symetrii w siebie. Zbiory te wyglądają inaczej dla różnych obiektów geometrycznych. Interesujące jest jednak podejście, w którym to nie obiekt definiuje swoje przekształcenia, ale to zbiór przekształceń definiuje obiekt! Weźmy zbiór X bez żadnej „odziedziczonej” struktury. Do badania X będziemy mieli do dyspozycji zestaw przekształceń $\mathcal{F}$ zbioru X w siebie. Umówimy się, że jeśli dla pewnego podzbioru $O \subseteq X$ mamy $f(O) = O$, dla każdego $f \in \mathcal{F}$, to zbiór O nazywamy *$\mathcal{F}$-okręgiem*. Uwaga: nic nie wiemy ani o zbiorze O, ani o X (może to np. zbiór Cantora), a nawet nic wstępnie nie zakładamy o $\mathcal{F}$. Przy tak minimalistycznym podejściu wydaje się, że zbieżność nazewnictwa *$\mathcal{F}$-okręgu* z „prawdziwymi okręgami” (czyli czym?) jest zupełnie przypadkowa i bezzasadna. Czy rzeczywiście?

Umówmy się zatem, że jeśli w zbiorze X istnieje inny podzbiór: P taki, że tylko dwa przekształcenia z $\mathcal{F}$ (a nie wszystkie) zachowują P, to P nazywamy *$\mathcal{F}$-prostokątem*. Nonsens? Pojęcia $\mathcal{F}$-okręgu i $\mathcal{F}$-prostokąta pozbawione są być może na pozór geometrycznego sensu – ale można pytać co nas dokładnie interesowało? Myślenie geometryczne polegające na rozróżnianiu obiektów ze względu na pewne niezmienniki jest niezwykle ciekawe i prowadzi do głębokich pytań, zupełnie innych jakościowo od pytań geometrii szkolnej. Proces formułowania problemów bywa nie mniej fascynujący, co samo poszukiwanie na nie rozwiązania. Ot choćby: jak wysłowić różnicę między preclem, a pączkiem, czyli między torusem i sferą (dwuwymiarowymi)? Przecież „widzimy” tę różnicę („dziura”), ale jak ująć tę różnicę w pojęcia matematyczne i sformułować odpowiedni wynik? Pytanie to dotyczy niezwykle ważnych problemów.



Rozdział 8

Przekształcenia afiniczne

W tym rozdziale omawiamy pojęcie przekształcenia afinicznego, czyli z geometrycznego punktu widzenia: takiego przekształcenia pomiędzy przestrzeniami afinicznymi, które zachowuje kombinacje afiniczne. Przekształcenie takie przeprowadza choćby trójki punktów współliniowych na trójki punktów współliniowych. Z algebraicznego punktu widzenia można szukać innej motywacji: mamy w pewnych przestrzeniach liniowych V, W zawarte przestrzenie afiniczne H i M szukamy przekształceń $f : H \to M$, które zachowują operację wektora łączącego zdefiniowaną na V oraz W. Oznacza to, że z przekształceniem f stowarzyszone jest przekształcenie liniowe $\phi : T(H) \to T(M)$ takie, że dla każdych $p_1 \in H$ oraz $p_2 \in M$ mamy

$$\phi(\overrightarrow{p_1p_2}) = \overrightarrow{f(p_1)f(p_2)}.$$

Wykażemy, że są to podejścia równoważne. Zacznijmy od podejścia algebraicznego.

Obserwacja 8.1

Niech H, M będą przestrzeniami afinicznymi nad ciałem K i niech $\phi : T(H) \to T(M)$ będzie przekształceniem liniowym. Dla dowolnej funkcji $f : H \to M$ następujące warunki są równoważne:

(1) istnieją punkty $p_0 \in H$, $q_0 \in M$ takie, że dla każdego punktu $p \in H$ zachodzi

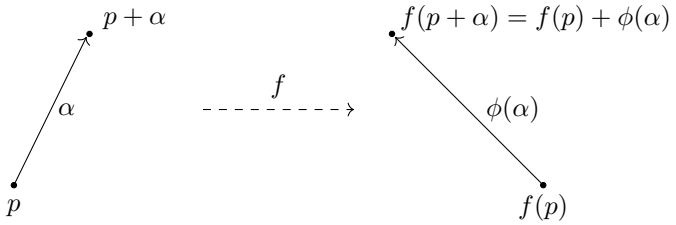
$$f(p) = q_0 + \phi(\overrightarrow{p_0p}),$$

(2) dla każdego $p \in H$ i każdego $\alpha \in T(H)$:

$$f(p + \alpha) = f(p) + \phi(\alpha),$$

(3) dla każdych $p, p' \in H$ mamy:

$$\phi\left(\overrightarrow{pp'}\right) = \overrightarrow{f(p)f(p')}.$$



Powyższa obserwacja daje nam dwa warunki charakteryzujące ów związek pomiędzy poszukiwanymi (i jeszcze niezdefiniowanymi) przekształceniami przestrzeni afinicznych, a odpowiadającymi im przekształceniami liniowymi zachowującymi wektory łączące (to dokładnie wyraża warunek (3)).

Punkt (1) podaje wprost „przepis" na znalezienie każdego takiego przekształcenia $f : H \to M$ mając dane przekształcenie $\phi : T(H) \to T(M)$. Warunek ten mówi, że po ustaleniu na co ma przechodzić jeden wybrany punkt z przestrzeni H możemy, wykorzystując ϕ, określić jednoznacznie przekształcenie f.

Dowód. Dowodzimy (1) $\Rightarrow$ (2). Z (1) wynika, że $f(p_0) = q_0$. Co więcej, dla każdego $p \in H$ oraz $\alpha \in T(H)$ mamy

$$p + \alpha = p_0 + \overrightarrow{p_0p} + \alpha.$$

Stąd ponownie na mocy (1) oraz z liniowości ϕ:

$$f(p+\alpha) = f(p_0+\overrightarrow{p_0p}+\alpha) \overset{(1)}{=} q_0+\phi(\overrightarrow{p_0(p_0 + \overrightarrow{p_0p} + \alpha)}) = q_0+\phi(\overrightarrow{p_0p}+\alpha) = q_0+\phi(\overrightarrow{p_0p})+\phi(\alpha) = f(p)+\phi(\alpha).$$

Dowodzimy (2) $\Rightarrow$ (3). Weźmy dowolne punkty $p, p' \in H$ i niech $\alpha = \overrightarrow{pp'}$. Wówczas $p' = p + \alpha$, więc:

$$f(p') = f(p + \alpha) \overset{(2)}{=} f(p) + \phi(\alpha) = f(p) + \phi(\overrightarrow{pp'}).$$

Stąd $\phi(\overrightarrow{pp'}) = f(p') - f(p) = \overrightarrow{f(p)f(p')}$.

Dowodzimy (3) $\Rightarrow$ (1). Weźmy dowolny punkt $p_0 \in H$ i przyjmijmy $q_0 = f(p_0)$. Wówczas dla każdego $p \in H$ mamy

$$\phi(\overrightarrow{p_0p}) \overset{(3)}{=} \overrightarrow{f(p_0)f(p)} = \overrightarrow{q_0f(p)},$$

więc

$$f(p) - q_0 = \phi(\overrightarrow{p_0p}) \Rightarrow f(p) = q_0 + \phi(\overrightarrow{p_0p}).$$

□

Drugie, bardzo naturalne z geometrycznego punktu widzenia podejście każe szukać przekształceń, które zachowują kombinacje afiniczne punktów (czyli także ich współrzędne barycentryczne w układzie bazowym). Poniższe twierdzenie mówi, że przekształcenia te uzyskiwane są jedynie w sposób opisany wyżej.

Twierdzenie 8.1

Niech H, M będą przestrzeniami afinicznymi nad ciałem K. Wówczas dla funkcji $f : H \to M$ następujące warunki są równoważne:

(1) dla każdych punktów $p_0, \ldots, p_k \in H$ oraz układu wag $a_0, \ldots, a_k \in K$ mamy

$$f(a_0p_0 + \ldots + a_kp_k) = a_0f(p_0) + \ldots + a_kf(p_k),$$

(2) istnieje przekształcenie liniowe $\phi : T(H) \to T(M)$ oraz punkty $p_0 \in H$ i $q_0 \in M$ takie, że dla każdego $p \in H$ mamy

$$f(p) = q_0 + \phi(\overrightarrow{p_0p}).$$

Dowód. Wykażemy (1) $\Rightarrow$ (2). Wybierzmy $p_0 \in H$ i określmy przekształcenie $\phi : T(H) \to T(M)$ wzorem

$$\phi(\alpha) = \overrightarrow{f(p_0)f(p_0 + \alpha)} = f(p_0 + \alpha) - f(p_0).$$

Wykażemy, że ϕ jest liniowe. Niech $\alpha, \beta \in T(H)$. Wówczas $p_0 + \alpha + \beta$ jest kombinacją afiniczną punktów $p_0, p_0 + \alpha, p_0 + \beta$ z wagami $-1, 1, 1$ postaci:

$$p_0 + \alpha + \beta = -p_0 + (p_0 + \alpha) + (p_0 + \beta).$$

Z warunku (1) dostajemy więc, że:

$$f(p_0 + \alpha + \beta) = -f(p_0) + f(p_0 + \alpha) + f(p_0 + \beta).$$

A zatem z definicji ϕ mamy:

$$\begin{aligned}
\phi(\alpha + \beta) &= \overrightarrow{f(p_0)f(p_0 + \alpha + \beta)} = \\
&= f(p_0 + \alpha + \beta) - f(p_0) = \\
&= -f(p_0) + f(p_0 + \alpha) + f(p_0 + \beta) - f(p_0) = \\
&= f(p_0 + \alpha) - f(p_0) + f(p_0 + \beta) - f(p_0) = \\
&= \overrightarrow{f(p_0)f(p_0 + \alpha)} + \overrightarrow{f(p_0)f(p_0 + \beta)} = \\
&= \phi(\alpha) + \phi(\beta).
\end{aligned}$$

Pozostaje pokazać, że $\phi(a\alpha) = a\phi(\alpha)$, dla każdego $\alpha \in T(H)$ oraz dla każdego $a \in K$. Jednak również wektor $p_0 + a\alpha$ możemy przedstawić jako kombinację afiniczną postaci

$$p_0 + a\alpha = a(p_0 + \alpha) + (1 - a)p_0,$$

otrzymując stąd równość:

$$f(p_0 + a\alpha) = af(p_0 + \alpha) + (1 - a)f(p_0).$$

W rezultacie opierając się ponownie na definicji ϕ mamy:

$$\begin{aligned}
\phi(a\alpha) &= \overrightarrow{f(p_0)f(p_0 + a\alpha)} = \\
&= f(p_0 + a\alpha) - f(p_0) = \\
&= af(p_0 + \alpha) + (1 - a)f(p_0) - f(p_0) = \\
&= a(f(p_0 + \alpha) - f(p_0)) = \\
&= a\phi(\alpha).
\end{aligned}$$

Pokazaliśmy zatem, że zdefiniowane przez nas ϕ jest przekształceniem liniowym. Trzeba jeszcze sprawdzić, że spełnia ono wymagania postawione w (2). Z definicji ϕ mamy jednak: $f(p_0 + \alpha) = f(p_0) + \phi(\alpha)$, dla każdego $\alpha \in T(H)$. Przyjmując więc $q_0 = f(p_0)$ dostajemy:

$$f(p) = f(p_0 + \overrightarrow{p_0p}) = f(p_0) + \phi(\overrightarrow{p_0p}) = q_0 + \phi(\overrightarrow{p_0p}).$$

Dowodzimy (2) $\Rightarrow$ (1). Przypuśćmy więc, że istnieje przekształcenie liniowe $\phi : T(H) \to T(M)$ oraz punkty $s_0 \in H, t_0 \in M$ takie, że dla każdego $p \in H$ zachodzi $f(p) = t_0 + \phi(\overrightarrow{s_0p})$ (w szczególności $f(s_0) = t_0$). Wówczas dla każdych punktów $p_0, \ldots, p_k \in H$ i wag $a_0, \ldots, a_k \in K$ mamy:

$$a_0p_0 + \ldots + a_kp_k = s_0 - (a_0 + \ldots + a_k)s_0 + a_0p_0 + \ldots + a_kp_k = s_0 + a_0\overrightarrow{s_0p_0} + \ldots + a_k\overrightarrow{s_0p_k},$$

Zatem

$$\begin{aligned}
f(a_0p_0 + \ldots + a_kp_k) &= f(s_0 + a_0\overrightarrow{s_0p_0} + \ldots + a_k\overrightarrow{s_0p_k}) = \\
&\overset{\text{Obs. 9.1 (2)}}{=} t_0 + \phi(a_0\overrightarrow{s_0p_0} + \ldots + a_k\overrightarrow{s_0p_k}) = \\
&= t_0 + a_0\phi(\overrightarrow{s_0p_0}) + \ldots + a_k\phi(\overrightarrow{s_0p_k}) = \\
&\overset{\text{Obs 9.1 (3)}}{=} t_0 + a_0\overrightarrow{f(s_0)f(p_0)} + \ldots + a_k\overrightarrow{f(s_0)f(p_k)} = \\
&= t_0 + a_0\overrightarrow{t_0f(p_0)} + \ldots + a_k\overrightarrow{t_0f(p_k)} = \\
&= a_0(t_0 + \overrightarrow{t_0f(p_0)}) + \ldots + a_k(t_0 + \overrightarrow{t_0f(p_k)}) = \\
&= a_0f(p_0) + \ldots + a_kf(p_k).
\end{aligned}$$

□

Powyższe twierdzenie motywuje sformułowanie definicji przekształcenia afinicznego.

Definicja 8.1: Przekształcenie afiniczne

Niech H, M będą przestrzeniami afinicznymi nad ciałem K. Mówimy, że funkcja $f : H \to M$ jest PRZEKSZTAŁCENIEM AFINICZNYM, jeśli f spełnia jeden z równoważnych warunków:

(1) dla każdych punktów $p_0, \ldots, p_k \in H$ oraz układu wag $a_0, \ldots, a_k \in K$ zachodzi równość $f(a_0p_0 + \ldots + a_kp_k) = a_0f(p_0) + \ldots + a_kf(p_k)$,

(2) istnieje przekształcenie liniowe $\phi : T(H) \to T(M)$ oraz punkty $p_0 \in H$ i $q_0 \in M$ takie, że dla każdego $p \in H$ mamy $f(p) = q_0 + \phi(\overrightarrow{p_0p})$.

Przekształcenie liniowe $\phi : T(H) \to T(M)$ z warunku (2) nazywamy POCHODNĄ (albo PRZEKSZTAŁCENIEM WEKTORÓW SWOBODNYCH) przekształcenia afinicznego f i oznaczamy f'.

Z pierwszego udowodnionego dziś twierdzenia wynika, że ϕ jest przekształceniem wyznaczonym jednoznacznie przez f, niezależnie od wyboru p_0 i $q_0 = f(p_0)$. Rozważmy kilka przykładów.

- Rozważmy przyporządkowanie $f : \mathbb{R}^2 \to \mathbb{R}^3$ zadane wzorem $f((x_1, x_2)) = (x_1 + 2, x_2 + 1, x_2)$. Jest to przekształcenie afiniczne, gdyż mamy przekształcenie liniowe $\phi : \mathbb{R}^2 \to \mathbb{R}^3$ dane wzorem $\phi((x_1, x_2)) = (x_1, x_2, x_2)$ oraz punkty $p_0 = (0, 0), q_0 = (2, 1, 0)$ takie, że
$$f((x_1, x_2)) = (2, 1, 0) + \phi(x_1 - 0, x_2 - 0) = (2, 1, 0) + \phi((x_1, x_2)).$$

- Niech
$$H = \{(x_1, x_2, x_3) \in K^3 \,|\, x_1 + x_2 + x_3 = 1\}, \quad M = \{(x_1, x_2) \in K^2 \,|\, x_1 - x_2 = 2\}.$$
Funkcja $f : H \to M$ dana wzorem $f((x_1, x_2, x_3)) = (x_1, -x_2 - x_3 - 1)$ jest przekształceniem afinicznym. Mamy
$$T(H) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 + x_2 + x_3 = 0\}, \quad T(M) = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, x_1 - x_2 = 0\}.$$
Mamy też $\phi(T(H)) = T(M)$, gdzie $\phi : \mathbb{R}^3 \to \mathbb{R}^2$ dane jest wzorem $\phi((x_1, x_2, x_3)) = (x_1, -x_2 - x_3)$. Możemy więc określić $f' = \phi|_{T(H)} : T(H) \to T(W)$ oraz $p_0 = (0, -1, 2)$, $q_0 = (0, -2)$.

- Niech V, Z będą przestrzeniami liniowymi nad K. Każde przekształcenie liniowe $\phi : V \to Z$ jest przekształceniem afinicznym, przy czym $\phi' = \phi$. Każde przekształcenie afiniczne $f : V \to Z$ jest postaci $f(\alpha) = f(0 + \alpha) = f(0) + f'(\alpha)$.

- Każda parametryzacja $K^n \to H$ przestrzeni afinicznej H jest przekształceniem afinicznym.

- Złożenie przekształceń afinicznych $g \circ f$ jest afiniczne, gdyż w sposób oczywisty zachowuje kombinacje afiniczne. Co więcej, nietrudno widzieć, że $(g \circ f)' = g' \circ f'$.

Wniosek 8.1

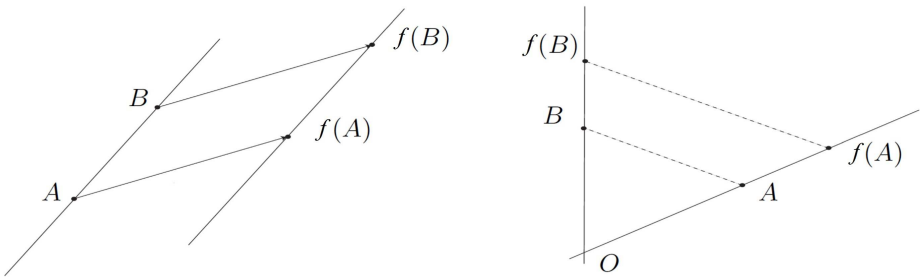
Każde przekształcenie afiniczne $f : K^n \to K^m$ jest zadane wzorem:
$$f((x_1, \ldots, x_n)) = (a_{11}x_1 + \ldots + a_{1n}x_n + b_1, \quad \ldots \quad , a_{m1}x_1 + \ldots + a_{mn}x_n + b_m),$$
gdzie $(b_1, \ldots, b_m) = f((0, \ldots, 0))$ oraz f' to przekształcenie liniowe o macierzy $M(f')^{st}_{st} = A = [a_{ij}]$.

Omówmy kilka ważnych klas przekształceń afinicznych.

- Dla dowolnych przestrzeni afinicznych H, M nad K i każdego $q \in M$ PRZEKSZTAŁCENIE STAŁE $f : H \to M$, o wartości $f(p) = q$, dla każdego $p \in H$, jest afiniczne. Przy tym $f' = 0$.

- Niech H będzie przestrzenią afiniczną nad K i niech $\alpha \in T(H)$. Przekształcenie $f : H \to H$ spełniające $f(p) = p + \alpha$, dla każdego $p \in H$ jest afiniczne. Nazywamy je PRZESUNIĘCIEM (TRANSLACJĄ) O WEKTOR α i oznaczamy τ_α. Zatem $(\tau_\alpha)' = \mathrm{id}$.

- Niech H będzie przestrzenią afiniczną nad K, $p_0 \in H$ i $a \in K$. Przekształcenie afiniczne $f : H \to H$ takie, że $f(p_0) = p_0$ oraz $f' = a \cdot \mathrm{id}$ nazywamy JEDNOKŁADNOŚCIĄ O ŚRODKU p_0 I SKALI a.



Ilustracja przesunięcia (po lewej) i homotetii o środku O (po prawej).

- Niech H będzie przestrzenią afiniczną nad K oraz $T(H) = W_1 \oplus W_2$. Dla $p_1, p_2 \in H$ określamy podprzestrzenie afiniczne $H_1 = p_1 + W_1$, $H_2 = p_2 + W_2$. Przekształcenie afiniczne $f : H \to H$ takie, że f' jest RZUTEM na W_1 wzdłuż W_2 oraz $f(p_1) = p_1$ nazywamy RZUTEM na H_1 wzdłuż H_2. Przekształcenie afiniczne $g : H \to H$ takie, że $g(p_1) = p_1$ oraz g' jest SYMETRIĄ względem W_1 wzdłuż W_2 nazywamy SYMETRIĄ względem H_1 wzdłuż H_2.

Obserwacja 8.2: Jednoznaczność przekształcenia afinicznego na bazie punktowej

Niech H, M będą przestrzeniami afinicznymi nad ciałem K.

- Jeśli $p_0, \ldots, p_n$ jest bazą punktową przestrzeni H oraz $q_0, \ldots, q_n$ jest dowolnym układem punktów przestrzeni M, to istnieje dokładnie jedno przekształcenie afiniczne $f : H \to M$ takie, że $f(p_i) = q_i$, dla $i = 0, \ldots, n$. Jest ono zadane przez:

$$f(a_0p_0 + \ldots + a_np_n) = a_0q_0 + \ldots + a_nq_n, \quad \text{dla dowolnych wag } a_0, \ldots, a_n \in K.$$

- Jeśli $p_0; \alpha_1, \ldots, \alpha_n$ jest układem bazowym przestrzeni H oraz $q_0 \in M$, a $\beta_1, \ldots, \beta_n$ jest dowolnym układem wektorów w $T(M)$, to istnieje dokładnie jedno przekształcenie afiniczne $f : H \to M$ takie, że $f(p_0) = q_0$ oraz $f'(\alpha_i) = \beta_i$, dla $i = 1, \ldots, n$. Jest ono zadane wzorem

$$f(p_0 + a_1\alpha_1 + \ldots + a_n\alpha_n) = q_0 + a_1\beta_1 + \ldots + a_n\beta_n, \quad \text{dla dowolnych skalarów } a_1, \ldots, a_n \in K.$$

W geometrii elementarnej powyższa obserwacja daje ważne wnioski:

- dla każdej trójki punktów niewspółliniowych na płaszczyźnie H istnieje izomorfizm afiniczny, który przeprowadza te punkty na wierzchołki trójkąta równobocznego,
- dla każdej czwórki punktów będących wierzchołkami równoległoboku na płaszczyźnie H istnieje izomorfizm afiniczny, który przeprowadza te punkty na wierzchołki kwadratu,
- dla każdej czwórki punktów niewspółpłaszczyznowych w przestrzeni trójwymiarowej H istnieje izomorfizm afiniczny, który przeprowadza te punkty na wierzchołki czworościanu foremnego.

Prosty przykład zastosowania. Jeśli chcemy udowodnić, że w dowolnym trójkącie środkowe przecinają się w jednym punkcie i dzielą w stosunku $2 : 1$, licząc od wierzchołka, to wystarczy wiedzieć, że trójkąt równoboczny ma żądane własności. Dowolny inny trójkąt jest jego obrazem przy pewnym przekształceniu afinicznym, które zachowuje środki boków, a więc także środkowe, ich współpękowość oraz stosunek podziału. Podobnie argumentować można np. dowód twierdzenia Cevy czy Menelaosa.

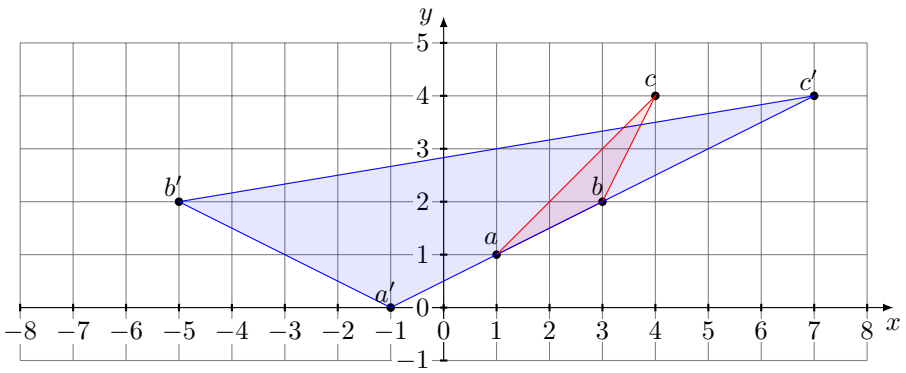
Przykład. Znajdź (jedyne) przekształcenie afiniczne $f : \mathbb{R}^2 \to \mathbb{R}^2$, które przeprowadza punkty $a = (1, 1), b = (3, 2), c = (4, 4)$ odpowiednio na punty $a' = (-1, 0)$, $b' = (-5, 2)$ oraz $c' = (7, 4)$.

Rozważmy układy bazowe $a; \overrightarrow{ab}, \overrightarrow{ac}$, czyli $(1, 1); (2, 1), (3, 3)$ oraz $a'; \overrightarrow{a'b'}, \overrightarrow{a'c'}$, czyli $(-1, 0); (-4, 2), (8, 4)$. Stąd f' jest przeprowadzają $(2, 1)$ w $(-4, 2)$ oraz $(3, 3)$ w $(8, 4)$. Zatem (jak w pierwszym semestrze):

$$f'((x_1, x_2)) = \left(-\frac{20}{3}x_1 + \frac{28}{3}x_2, \frac{2}{3}x_1 + \frac{2}{3}x_2\right).$$

Biorąc dowolne $p = (x_1, x_2)$ oraz $p_0 = (1, 1)$, $q_0 = f(p_0) = (-1, 0)$ mamy:

$$f((x_1, x_2)) = q_0 + f(\overrightarrow{p_0p}) = (-1, 0) + f'(\overrightarrow{(1, 1), (x_1, x_2)}) = (-1, 0) + f'((x_1 - 1, x_2 - 1))$$
$$= (-1, 0) + \left(-\frac{20}{3}(x_1 - 1) + \frac{28}{3}(x_2 - 1), \frac{2}{3}(x_1 - 1) + \frac{2}{3}(x_2 - 1)\right) = \left(-\frac{20}{3}x_1 + \frac{28}{3}x_2 - \frac{11}{3}, \frac{2}{3}x_1 + \frac{2}{3}x_2 - \frac{4}{3}\right).$$



Opisane przekształcenie przeprowadza trójkąt abc na trójkąt $a'b'c'$. Dokładniej, punkt o współrzędnych barycentrynych (x, y, z) w bazie punktowej a, b, c przechodzi na punkt o współrzędnych barycentrycznych (x, y, z) w bazie punktowej a', b', c'.

Definicja 8.2: Izomorfizm afiniczny

Niech H, M będą przestrzeniami afinicznymi nad ciałem K. Mówimy, że przekształcenie afiniczne $f : H \to M$ jest IZOMORFIZMEM, jeśli f jest różnowartościowe i „na". Mówimy, że przestrzenie H, M są IZOMORFICZNE, jeśli istnieje izomorfizm $f : H \to M$.

Poniższa charakteryzacja izomorfizmów przestrzeni afinicznych jest bezpośrednią konsekwencją odpowiedniej charakteryzacji izomorfizmów przestrzeni liniowych i związku między nimi.

Obserwacja 8.3

Niech H, M będą przestrzeniami afinicznymi nad K i niech $f : H \to M$ będzie przekształceniem afinicznym. Następujące warunki są równoważne:

- f jest izomorfizmem,
- f przeprowadza pewną (każdą) bazę punktową H na bazę punktową M,
- istnieje przekształcenie afiniczne $g : M \to H$ takie, że $g \circ f = \mathrm{id}_H$ oraz $f \circ g = \mathrm{id}_M$,
- $f' : T(H) \to T(M)$ jest izomorfizmem przestrzeni liniowych.

Przesunięcia równoległe, jednokładności o skalach różnych od 0, symetrie – są przykładami izomorfizmów afinicznych. Skończenie wymiarowe przestrzenie liniowe nad K są izomorficzne wtedy i tylko wtedy, gdy mają ten sam wymiar. Stąd wynika wniosek.

Wniosek 8.2

Niech H, M będą skończenie wymiarowymi przestrzeniami afinicznymi nad K. Przestrzenie H, M są izomorficzne wtedy i tylko wtedy, gdy $\dim T(H) = \dim T(M)$. W szczególności każda n-wymiarowa przestrzeń afiniczna nad K jest izomorficzna z przestrzenią afiniczną K^n.

Przyjrzymy się związkom pomiędzy przekształceniami afinicznymi i macierzami ich pochodnych.

Obserwacja 8.4

Niech H, M będą przestrzeniami afinicznymi nad K, niech $p_0; \mathcal{A}$ będzie układem bazowym przestrzeni H oraz niech $q_0; \mathcal{B}$ będzie układem bazowym przestrzeni M, przy czym $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$, $\mathcal{B} = (\beta_1, \ldots, \beta_m)$. Niech $f : H \to M$ będzie przekształceniem afinicznym, przy czym:

- $f(p_0) = q_0 + w_1\beta_1 + \ldots + w_m\beta_m$, dla pewnych $w_1, \ldots, w_n$,
- $M(f')^{\mathcal{B}}_{\mathcal{A}} = A$.

Wówczas dla każdego $p \in H$: jeśli p ma w układzie bazowym $p_0; \mathcal{A}$ współrzędne $a_1, \ldots, a_n$ oraz $f(p)$ ma w układzie bazowym $q_0; \mathcal{B}$ współrzędne $b_1, \ldots, b_m$, to:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = A \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}. \qquad (*)$$

W szczególności jeśli $H = M$ oraz $f = \mathrm{id}$, dostajemy następujący związek między współrzędnymi $a_1, \ldots, a_n$, $b_1, \ldots, b_n$ punktu $p \in H$ odpowiednio w układach bazowych $p_0; \mathcal{A}$, $q_0; \mathcal{B}$:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = C \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, \quad \text{gdzie} \quad C = M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}} \text{ oraz } p_0 = q_0 + w_1\beta_1 + \ldots + w_n\beta_n.$$

Dowód. Skoro $A = M(f')^{\mathcal{B}}_{\mathcal{A}}$, to dla $d_1, \ldots, d_m$ spełniających

$$\begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix} = A \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

mamy $f'(a_1\alpha_1 + \ldots + a_n\alpha_n) = d_1\beta_1 + \ldots + d_m\beta_m$. Skoro $p = p_0 + a_1\alpha_1 + \ldots + a_n\alpha_n$, to:

$$\begin{aligned} f(p) &= f(p_0) + f'(a_1\alpha_1 + \ldots + a_n\alpha_n) = \\ &= q_0 + w_1\beta_1 + \ldots + w_m\beta_m + d_1\beta_1 + \ldots + d_m\beta_m = \\ &= q_0 + (w_1 + d_1)\beta_1 + \ldots + (w_m + d_m)\beta_m \end{aligned}$$

Wobec równości $f(p) = q_0 + b_1\beta_1 + \ldots + b_m\beta_m$ mamy $\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = A \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}$. □

Uwaga. Stosuje się niekiedy następującą wygodną notację blokową zapisującą skrótowo formułę $(*)$:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_m \\ 1 \end{bmatrix} = \left[\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & w_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} & w_m \\ \hline 0 & \ldots & 0 & 1 \end{array}\right] \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ 1 \end{bmatrix}.$$

Wprowadzimy element nieobowiązkowy: macierz przekształcenia afinicznego w układach bazowych[1].

Definicja 8.3

Niech H, M będą przestrzeniami afinicznymi nad K, niech $p_0; \mathcal{A}$ będzie układem bazowym przestrzeni H oraz niech $q_0; \mathcal{B}$ będzie układem bazowym przestrzeni M, przy czym $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$, $\mathcal{B} = (\beta_1, \ldots, \beta_m)$. MACIERZĄ PRZEKSZTAŁCENIA AFINICZNEGO $f : H \to M$ W UKŁADACH BAZOWYCH $p_0; \mathcal{A}$, $q_0; \mathcal{B}$ nazywamy macierz blokową

$$M(f)^{q_0;\mathcal{B}}_{p_0;\mathcal{A}} = \left[\begin{array}{c|c} A & w \\ \hline 0 & 1 \end{array}\right],$$

gdzie $A = M(f')^{\mathcal{B}}_{\mathcal{A}}$ spełnia $(*)$ oraz w jest wektorem współrzędnych $f(p_0)$ w układzie bazowym $q_0; \mathcal{B}$

Wróćmy do rozważanego już wcześniej przekształcenia afinicznego przeprowadzającego trójkąty o wierzchołkach $a = (1,1), b = (3,2), c = (4,4)$ trójkąta abc odpowiednio na wierzchołki trójkąta $a'b'c'$ postaci $a' = (-1,0), b' = (-5,2), c' = (7,4)$ Popatrzymy na f w dwóch sytuacjach.

- W układach bazowych: $\mathcal{X} = \{a; \overrightarrow{ab}, \overrightarrow{ac}\}$ oraz $\mathcal{Y} = \{a'; \overrightarrow{a'b'}, \overrightarrow{a'c'}\}$ mamy $f(a) = a' + 0 \cdot \overrightarrow{a'b'} + 0 \cdot \overrightarrow{a'c'}$, czyli współrzędne $f(a)$ w układzie bazowym $\mathcal{Y}$ wynoszą $w_1 = 0, w_2 = 0$. A zatem

$$M(f)^{\mathcal{Y}}_{\mathcal{X}} = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array}\right].$$

- W standardowych układach bazowych $st = \{(0,0), (1,0), (0,1)\}$ mamy $f(a) = a' = (-1,0)$. Mamy też $f'((1,0)) = (-\frac{20}{3}, \frac{2}{3})$ oraz $f'((0,1)) = (\frac{28}{3}, \frac{2}{3})$. A zatem

$$M(f)^{st}_{st} = \left[\begin{array}{cc|c} -\frac{20}{3} & \frac{28}{3} & -1 \\ \frac{2}{3} & \frac{2}{3} & 0 \\ \hline 0 & 0 & 1 \end{array}\right].$$

[1] Pojęcie to nie jest standardowym elementem kursu, ale pojawia się u wielu autorów i prowadzi do nieco łatwiejszych rachunków, gdy składamy przekształcenia afiniczne. Wtedy bowiem dla przekształceń afinicznych $f : H_1 \to H_2$ oraz $g : H_2 \to H_3$ oraz układów bazowych $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ odpowiednio w H_1, H_2, H_3 mamy

$$M(f \circ g)^{\mathcal{Z}}_{\mathcal{X}} = M(g)^{\mathcal{Z}}_{\mathcal{Y}} \cdot M(f)^{\mathcal{Y}}_{\mathcal{X}}.$$

Warto również wspomnieć, że zbiór przekształceń afinicznych $A(H, M)$ z H do M jest izomorficzny z przestrzenią $M_{\dim H+1 \times \dim M}(K)$ wymiaru $(\dim H + 1) \dim M$.

Przykład. W przestrzeni afinicznej $\mathbb{R}^3$ rozważamy podprzestrzenie afiniczne

$$L = \mathrm{af}((1,1,1),(3,5,4)), \quad H : x_1 + 2x_2 - 2x_3 = 4.$$

Niech $f : \mathbb{R}^3 \to \mathbb{R}^3$ będzie rzutem na H wzdłuż L oraz niech $g : \mathbb{R}^3 \to \mathbb{R}^3$ będzie jednokładnością o środku $(5,5,5)$ i skali 3. Wyznaczyć wzór przekształcenia f.

Rozwiązanie 1. Mamy $T(L) = \mathrm{lin}(2,4,3)$, zaś $T(H)$ jest opisana równaniem $x_1 + 2x_2 - 2x_3 = 0$, czyli np. $T(H) = \mathrm{lin}((0,1,1),(2,0,1))$. Mamy zatem bazę $\mathcal{A}$ przestrzeni $\mathbb{R}^3$, której pierwszy element rozpina $T(L)$, a kolejne dwa rozpinają $T(H)$: $\mathcal{A} = ((2,4,3),(0,1,1),(2,0,1))$. Pochodna f' jest rzutem $\mathbb{R}^3$ na $T(H)$ wzdłuż $T(L)$, czyli

$$M(f')^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Wyznaczamy $M(f')^{st}_{st}$ ze wzoru $M(f')^{st}_{st} = M(\mathrm{id})^{st}_{\mathcal{A}} \cdot M(f')^{\mathcal{A}}_{\mathcal{A}} \cdot M(\mathrm{id})^{\mathcal{A}}_{st}$. Mamy:

$$M(\mathrm{id})^{st}_{\mathcal{A}} = \begin{bmatrix} 2 & 0 & 2 \\ 4 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \quad \text{oraz } M(\mathrm{id})^{\mathcal{A}}_{st} = (M(\mathrm{id})^{st}_{\mathcal{A}})^{-1} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & -\frac{1}{2} \\ -1 & -1 & 2 \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Zatem:

$$M(f')^{st}_{st} = \begin{bmatrix} 2 & 0 & 2 \\ 4 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & -\frac{1}{2} \\ -1 & -1 & 2 \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & -\frac{1}{2} \\ -1 & -1 & 2 \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -1 & 1 \\ -1 & -1 & 2 \\ -\frac{3}{4} & -\frac{3}{2} & \frac{5}{2} \end{bmatrix}$$

Oczywiście f przeprowadza każdy punkt z H na siebie, a więc mamy na przykład $f((4,0,0)) = (4,0,0)$. Mamy więc, dla dowolnego punktu $(x_1,x_2,x_3) \in \mathbb{R}^3$:

$$\begin{aligned} f(x_1,x_2,x_3) &= (4,0,0) + f'((x_1,x_2,x_3) - (4,0,0)) = (4,0,0) + f'(x_1-4,x_2,x_3) = \\ &= (4,0,0) + \left(\frac{1}{2}(x_1-4) - x_2 + x_3, -(x_1-4) - x_2 + 2x_3, -\frac{3}{4}(x_1-4) - \frac{3}{2}x_2 + \frac{5}{2}x_3\right) = \\ &= \left(\frac{1}{2}x_1 - x_2 + x_3 + 2, -x_1 - x_2 + 2x_3 + 4, -\frac{3}{4}x_1 - \frac{3}{2}x_2 + \frac{5}{2}x_3 + 3\right) \end{aligned}$$

Rozwiązanie 2 (mniej rachunkowe). Obrazem punktu (x_1,x_2,x_3) przy rzucie f jest taki punkt

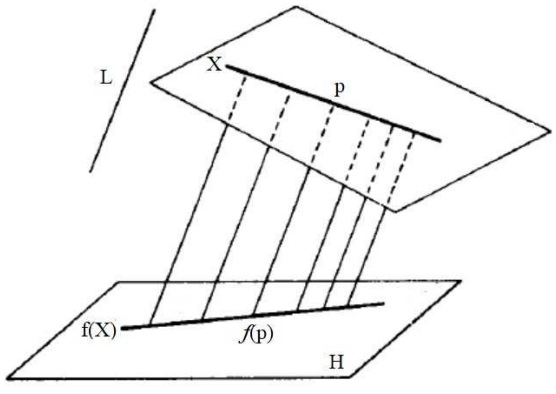
$$f((x_1,x_2,x_3)) = (x_1,x_2,x_3) + t(2,4,3),$$

który należy do H. A zatem wyznaczamy t za pomocą x_1,x_2,x_3 rozwiązując równanie

$$(x_1+2t) + 2(x_2+4t) - 2(x_3+3t) = 4,$$

czyli równoważnie $4t = 4 - x_1 - 2x_2 + 2x_3$. Zatem $f((x_1,x_2,x_3)) = (x_1,x_2,x_3) + (1 - \frac{x_1}{4} - \frac{x_2}{2} + \frac{x_3}{2})(2,4,3)$.

Morał jest taki: pomocne jest rozumienie geometrii przekształcenia. Wyjaśnienie drugiego sposobu jest widoczne gdy patrzymy na geometrię rzutu, nazywanego czasem w geometrii rzutem równoległym. Rzut w kierunku pewnej prostej L w przestrzeni $\mathbb{R}^3$ na płaszczyznę H polega na tym, że wybierając punkt $(x_1,x_2,x_3) \in \mathbb{R}^3$ poruszamy się po prostej równoległej do L przechodzącej przez ten punkt. Zapisanie tej prostej w postaci parametrycznej ułatwia sprawdzanie które z jej punktów trafiają na H.



Przykład. Niech $s : \mathbb{R}^3 \to \mathbb{R}^3$ będzie symetrią względem płaszczyzny $H : x_1 + x_2 - x_3 = 2$ wzdłuż prostej $L = (0,1,0) + \mathrm{lin}((2,1,2))$. Znajdziemy $s((1,0,1))$ oraz wyznaczymy wzór przekształcenia s.

Niech $p = (1,0,1)$. Dla $r \in \mathbb{R}^3$ mamy $s(p) = s(r + \overrightarrow{rp}) = s(r) + s'(\overrightarrow{rp})$. Płaszczyzna H opisana jest układem bazowym

$$\underbrace{(1,1,0)}_{\in H}; \underbrace{(1,0,1),(0,1,1)}_{\in T(H)}.$$

Zatem obraz punktu $r = (1,1,0) \in H$ w tej symetrii to $s(r) = (1,1,0)$. Rozważmy wektor $\overrightarrow{rp}$ równy $(0,-1,1)$. Znajdźmy współrzędne tego wektora w bazie $(1,0,1), (0,1,1), (2,1,2)$ przestrzeni $\mathbb{R}^3$ (pierwsze dwa wektory rozpinają $T(H)$, a ostatni $T(L)$). Mamy:

$$(0,-1,1) = 4(1,0,1) + 1(0,1,1) - 2(2,1,2).$$

A zatem z definicji symetrii liniowej s':

$$\begin{aligned} s'((0,1,-1)) &= 4s'((1,0,1)) + 1s'((0,1,1)) - 2s'((2,1,2)) = \\ &= 4(1,0,1) + 1(0,1,1) + 2(2,1,2) = (8,3,9). \end{aligned}$$

Zatem

$$s(p) = s(r) + s'(\overrightarrow{rp}) = (1,1,0) + (8,3,9) = (9,4,9).$$

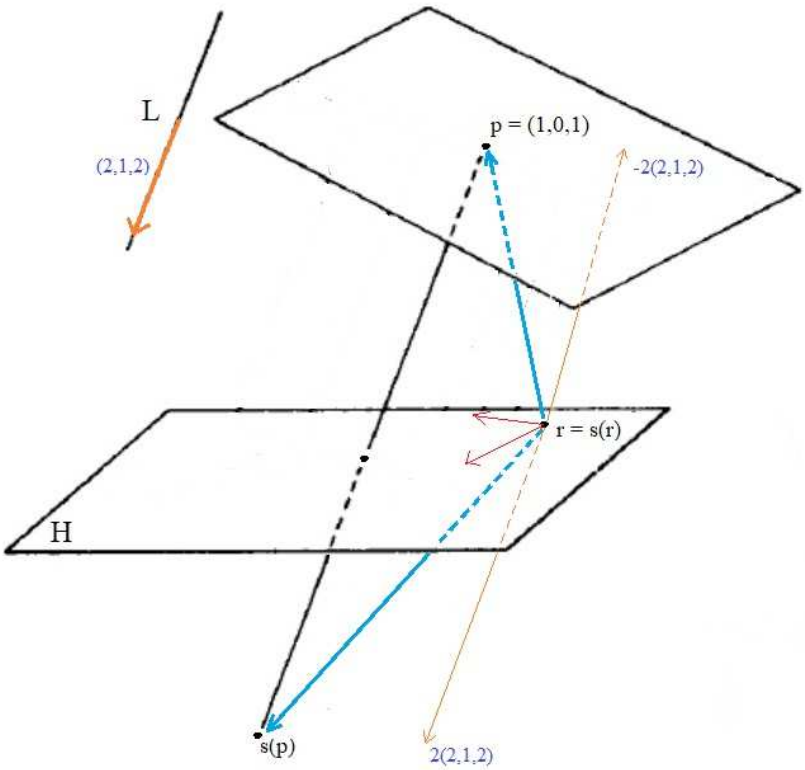
Przejdźmy do wyznaczenia wzoru na s. Układem bazowym płaszczyzny H jest $(1,1,0); (1,0,1), (0,1,1)$. Zatem obraz punktu $r = (1,1,0) \in H$ w tej symetrii równy jest $s(r) = (1,1,0)$. Dla $x \in \mathbb{R}^3$ mamy zatem $s(x) = s(r) + s'(\overrightarrow{rx})$, czyli w bazie standardowej:

$$s\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + M(s')^{st}_{st} \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right).$$

Oczywiście biorąc bazę $\mathcal{A} = ((1,0,1),(0,1,1),(2,1,2))$ przestrzeni $\mathbb{R}^3$ (pierwsze dwa wektory rozpinają $T(H)$, a ostatni $T(L)$) wyznaczamy macierz

$$M(s')^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

i stąd potem (znanymi metodami) wyznaczamy $M(s')^{st}_{st}$.



Rozdział 9

Zbiory i kombinacje wypukłe

Omówimy podstawy teorii zbiorów wypukłych w przestrzeni afinicznej $\mathbb{R}^n$.

Definicja 9.1: Podzbiór wypukły

Niepusty podzbiór X przestrzeni afinicznej $\mathbb{R}^n$ nazywamy WYPUKŁYM jeśli dla dowolnych $x, y \in X$ oraz $\lambda \in [0,1]$ mamy: $(1-\lambda)x + \lambda y \in X$. Zbiór pusty również uznajemy za wypukły.

Oczywistymi przykładami zbiorów wypukłych są: cała przestrzeń afiniczna $\mathbb{R}^n$ oraz każdy punkt tej przestrzeni. Podstawowymi przykładami sa natomiast sympleksy i połprzestrzenie.

Definicja 9.2

SYMPLEKSEM n-wymiarowym rozpiętym na afinicznie niezależnym układzie punktów $p_0, \ldots, p_n$ nazywamy zbiór $S(p_0, \ldots, p_n)$ określony warunkami:

$$S(p_0, \ldots, p_n) = \{a_0 p_0 + \ldots + a_n p_n \in \mathbb{R}^n \,|\, a_0 + \ldots + a_n = 1, a_i \geqslant 0\}.$$

Sympleks jednowymiarowy nazywamy ODCINKIEM, dwuwymiarowy – TRÓJKĄTEM, a trójwymiarowy – CZWOROŚCIANEM. Punkty $p_0, \ldots, p_n$ nazywamy też WIERZCHOŁKAMI sympleksu.

Sympleks jest oczywiście zbiorem wypukłym. Jeśli $p = a_0 p_0 + \ldots + a_n p_n$ oraz $q = a'_0 p_0 + \ldots + a'_n p_n$, gdzie $a_i, a'_i \geqslant 0$ oraz $a_0 + \ldots + a_n = a'_0 + \ldots + a'_n = 1$. to

$$t(a_0 p_0 + \ldots + a_n p_n) + (1-t)(a'_0 p_0 + \ldots + a'_n p_n) = (ta_0 + (1-t)a'_0)p_1 + \ldots + (ta_n + (1-t)a'_n)p_n.$$

Oczywiście mamy $ta_i + (1-t)a'_i \geqslant 0$. Natomiast suma owych nieujemnych wag jest równa 1, tzn.

$$(ta_0 + (1-t)a'_0) + \ldots + (ta_n + (1-t)a'_n) = t(a_0 + \ldots + a_n) + (1-t)(a'_0 + \ldots + a'_n) = t + 1 - t = 1.$$

Definicja 9.3: Półprzestrzeń w $\mathbb{R}^n$

PÓŁPRZESTRZENIĄ w $\mathbb{R}^n$ nazywamy zbiór H rozwiązań nietrywialnej nierówności liniowej postaci:

$$H = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \,|\, a_1 x_1 + a_2 x_2 + \ldots + a_n x_n \leqslant b, \text{ gdzie } (a_1, a_2, \ldots, a_n) \neq 0\}.$$

Półprzestrzeń jest zbiorem wypukłym. Istotnie, niech $p = (p_1, \ldots, p_n)$ oraz $q = (q_1, q_2, \ldots, q_n) \in H$ oraz niech $t \in [0,1]$ Należy wykazać, że $tp + (1-t)q \in H$. Mamy $a_1 p_1 + a_2 p_2 + \ldots + a_n p_n \leqslant b$ oraz $a_1 q_1 + a_2 q_2 + \ldots + a_n q_n \leqslant b$. Stąd

$$a_1(tp_1) + a_2(tp_2) + \ldots + a_n(tp_n) \leqslant tb, \quad a_1(1-t)q_1 + a_2(1-t)q_2 + \ldots + a_n(1-t)q_n \leqslant (1-t)b.$$

W rezultacie:

$$\sum_{i=1}^{n} a_i(tp_i + (1-t)q_i) \leqslant b, \quad \text{oraz} \quad tp + (1-t)q \in H.$$

Bez trudu sprawdzamy (zostawiam to jako ćwiczenie), że zbiorem wypukłym jest również (standardowa euklidesowa) KULA $K_{p,r} \subseteq \mathbb{R}^n$ o środku w punkcie $p = (p_1, \ldots, p_n)$ i promieniu $r \geqslant 0$, czyli zbiór postaci

$$K_{p,r} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid (x_1 - p_1)^2 + \ldots + (x_n - p_n)^2 \leqslant r^2\}.$$

Obserwacja 9.1

Część wspólna zbiorów wypukłych jest zbiorem wypukłym.

Dowód. Niech $A = \bigcap_i A_i$ będzie przecięciem rodziny zbiorów wypukłych A_i. Weźmy $p, q \in A$. Wtedy dla każdego i mamy $p \in A_i$ oraz $q \in A_i$, co z wypukłości A_i oznacza, że dla każdego $t \in [0,1]$ mamy $tp + (1-t)q \in A_i$. W szczególności $tp + (1-t)q \in A$. □

Definicja 9.4: Kombinacja wypukła, otoczka wypukła

Niech $x_0, \ldots, x_k$ będą punktami w przestrzeni afinicznej $\mathbb{R}^n$ oraz $a_0, \ldots, a_k$ będzie układem wag, czyli $a_0 + \ldots + a_k = 1$. Kombinację afiniczną $a_0x_0 + \ldots + a_kx_k$ nazywamy KOMBINACJĄ WYPUKŁĄ, jeśli $a_0, a_1, \ldots, a_k \geqslant 0$. Niech $X \subset \mathbb{R}^n$. Zbiór wszystkich kombinacji wypukłych punktów z X nazywamy UWYPUKLENIEM lub OTOCZKĄ WYPUKŁĄ zbioru X, oznaczaną jako $\mathrm{conv}(X)$.

Zgodnie z powyższą definicją $S(p_0, \ldots, p_n) = \mathrm{conv}(\{p_0, \ldots p_n\})$. Wykażemy następujący fakt.

Obserwacja 9.2

Zbiór $\mathrm{conv}(X)$ jest najmniejszym (względem inkluzji) zbiorem wypukłym zawierającym X.

Dowód. Wypukłość zbioru $\mathrm{conv}(X)$ uzasadnia się identycznie jak wypukłość sympleksu. Pozostaje więc uzasadnić minimalność. Niech C będzie zbiorem wypukłym zawierającym X. Wykażemy przez indukcję względem długości $k+1$ zapisu kombinacji wypukłej, że każdy punkt $p \in \mathrm{conv}(X)$ należy do C. Niech $p = a_0p_0 + \ldots + a_kp_k \in \mathrm{conv}(X)$, gdzie $p_i \in X$, $a_0 + \ldots + a_k = 1$ oraz $a_i \geqslant 0$.

Jeśli $k = 0$, to $p = p_0 \in X \subseteq C$. Przechodzimy do kroku indukcyjnego. Załóżmy, że $k > 0$ oraz, że każda kombinacja wypukła $\leqslant k$ punktów z X należy do C. Niech $p = a_0p_0 + \ldots + a_kp_k$. Jeśli dowolna z wag a_i tej kombinacji równa jest 1, to ponownie opisuje ona punkt ze zbioru X. Załóżmy więc, że mamy $a_k < 1$. Dla $i = 0, 1, 2, \ldots, k-1$ niech

$$a'_i = \frac{a_i}{1 - a_k}.$$

Wówczas $a'_i \geqslant 0$ oraz

$$a'_0 + \ldots + a'_{k-1} = \frac{a_0}{1-a_k} + \ldots + \frac{a_{k-1}}{1-a_k} = \frac{a_0 + \ldots + a_{k-1}}{1-a_k} = 1.$$

Zatem na mocy założenia indukcyjnego punkt $p' = a'_0p_1 + \ldots + a'_{k-1}p_{k-1}$ należy do C. Co więcej

$$p = a_0p_0 + \ldots + a_kp_k = (1-a_k)(a'_0p_0 + \ldots + a'_{k-1}p_{k-1}) + a_kp_k = (1-a_k)p' + a_kp_k,$$

czyli z definicji zbioru wypukłego C skoro $p' \in C$ oraz $p_k \in X \subseteq C$, to $p \in C$. □

Wniosek 9.1

Następujące warunki są równoważne dla podzbioru $X \subseteq \mathbb{R}^n$:

- X jest wypukły,
- X jest zamknięty na kombinacje wypukłe, tzn. $\mathrm{conv}(X) = X$.

Pięknym i ważnym aspektem teorii zbiorów wypukłych jest ich związek z dodawaniem zbiorów (rozumianych w sensie przestrzeni liniowych). Operacja ta ma w kontekście tej teorii nazwę sumy Minkowskiego.

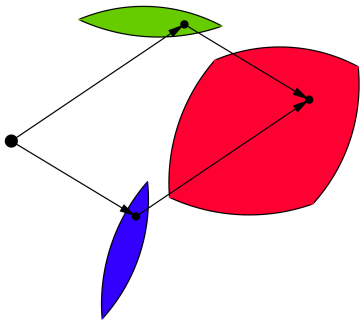
Definicja 9.5: Suma i iloczyn Minkowskiego

Dla podzbiorów A, B przestrzeni liniowej $\mathbb{R}^n$ SUMĄ MINKOWSKIEGO tych podzbiorów nazywamy:

$$A + B = \{a + b \,|\, a \in A, b \in B\}.$$

Dla podzbioru $A \subseteq \mathbb{R}^n$ przez ILOCZYN ZBIORU A PRZEZ SKALAR $t \in \mathbb{R}$ określamy zbiór

$$tA = \{ta \,|\, a \in A\}.$$



Suma Minkowskiego zbioru niebieskiego i zielonego to zbiór czerwony.

Zachęcam Czytelnika do uzasadnienia prostych obserwacji:

- zbiór $\{0\}$ jest elementem neutralnym dodawania Minkowskiego,
- suma Minkowskiego jest działaniem łącznym i przemiennym,
- mnożenie przez skalar jest rozdzielne względem dodawania,
- suma i mnożenie przez skalar w sensie Minkowskiego zachowują inkluzje, tzn. jeśli $A_1 \subseteq B_1$ oraz $A_2 \subseteq B_2$, to $A_1 + A_2 \subseteq B_1 + B_2$ oraz $tA_1 \subseteq tB_1$.

Obserwacja 9.3

Jeśli A, B są zbiorami wypukłymi, to $A + B$ oraz tA są wypukłe, dla każdego $t \in \mathbb{R}$. Co więcej, dla dowolnych podzbiorów $A, B \subseteq \mathbb{R}^n$ mamy:

$$\operatorname{conv}(A) + \operatorname{conv}(B) = \operatorname{conv}(A + B).$$

Dowód. Oczywiście tA jest obrazem zbioru A przy jednokładności o środku 0 i skali t, więc jest to zbiór wypukły. Również suma Minkowskiego zachowuje wypukłość. Jeśli zbiory A_1, A_2 są wypukłe to niech $p, q \in A_1 + A_2$. Wówczas $p = p_1 + p_2$ oraz $q = q_1 + q_2$, dla pewnych $p_i, q_i \in A_i$. Ponieważ A_i jest zbiorem wypukłym, to $S(p_i, q_i) \subseteq A_i$. Zatem:

$$S(p, q) \subseteq \{(1-t)p_1 + tq_1 \,|\, t \in [0,1]\} + \{(1-t)p_2 + tq_2 \,|\, t \in [0,1]\} \subset S(p_1, q_1) + S(p_2, q_2) \subseteq A_1 + A_2.$$

□

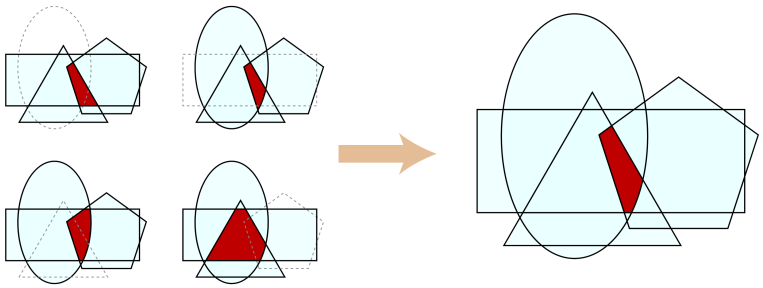
Wprowadzone wyżej operacje mają liczne proste własności analityczne i topologiczne. Czytelnika zainteresowanego tymi własnościami odsyłam do podręcznika prof. M. Moszyńskiej *Geometria zbiorów wypukłych*. Sumy Minkowskiego mają istotne znaczenie w zastosowaniach, m.in. w tomografii czy robotyce.

* * *

Przejdziemy teraz do słynnego twierdzenia z tzw. geometrii kombinatorycznej, mającego wiele intepretacji zarówno w geometrii elementarnej, jak i zastosowaniach.

Twierdzenie 9.1: Helly, 1913

Dana jest skończona rodzina $\mathcal{I}$ zbiorów wypukłych w $\mathbb{R}^n$, przy czym $|\mathcal{I}| > n$. Jeśli każde $n+1$ z tych zbiorów ma niepuste przecięcie, to przecięcie wszystkich elementów rodziny $\mathcal{I}$ jest niepuste.



Twierdzenie to ukazuje swoistą sztywność zbiorów wypukłych. Jako ilustrację proponujemy kilka zadań elementarnych.

Zadanie 9.1. *Na płaszczyźnie leży n punktów $\{a_1, a_2, \dots a_n\}$, gdzie $n \geqslant 3$. Dla każdych trzech z nich istnieje koło o promieniu 1, który je zawiera. Udowodnij, że wszystkie te punkty zawarte są w pewnym kole o promieniu 1.*

Rozwiązanie. Niech c_1, c_2, c_3 będą kołami o promieniach równych 1 i środkach w punktach a_1, a_2, a_3. Wówczas zgodnie z założeniem istnieje koło c o promieniu 1 i środku a_0, że $a_1, a_2, a_3 \in c$. Równoważnie: $c_1 \cap c_2 \cap c_3 \neq \emptyset$. Z twierdzenia Helly'ego zastosowanego do rodziny kół $\{c_1, c_2, \dots, c_n\}$ o promieniach 1 i środkach w $\{a_1, a_2, \dots a_n\}$ mamy: $c_1 \cap c_2 \cap \dots \cap c_n \neq 0$. Niech x należy do przecięcia tej rodziny kół. Zgodnie z definicją $\{a_1, a_2, \dots a_n\}$ są odległe od x o co najwyżej 1. Istnieje zatem koło o środku w x i promieniu 1, zawierający je wszystkie. ■

Dowód jest oczywiście prawdziwy gdy zamiast kół o promieniach równych 1 rozważymy koła o promieniach równych pewnemu ustalonemu $r > 0$. Zachęcam do samodzielnego zastosowania poprzedniego ćwiczenia w poniższym zadaniu.

Zadanie 9.2. *Niech $z_1, z_2, \dots, z_n$ będą punktami płaszczyzny takimi, że dowolne dwa są odległe od siebie nie więcej niż o 1. Wtedy wszystkie te punkty można pokryć kołem o promieniu $\frac{1}{\sqrt{3}}$.*

Zadanie 9.3. *Na płaszczyźnie położona jest dowolna skończona rodzina wielokątów $\mathcal{F}$ (niekoniecznie wypukłych) o tej własności, że każde dwa jej elementy mają punkt wspólny. Pokaż, że dla dowolnego punktu P na płaszczyźnie istnieje okrąg o środku w P mający punkt wspólny z każdym wielokątem z $\mathcal{F}$.*

Rozwiązanie. Wybieramy P i ustalamy dowolną półprostą l przechodzącą przez ten punkt. Rozważamy przekształcenie $f : \mathbb{R}^2 \to l$ zadane w następujący sposób: każdy okrąg O_r o środku w P i promieniu $r \geqslant 0$ przeprowadzamy na punkt $O_r \cap l$. Nietrudno widzieć[1], że każdy z wielokątów należących do $\mathcal{F}$ przejdzie na pewien odcinek domknięty na l. Z założenia $f(W_1) \cap f(W_2) \neq \emptyset$, gdzie $W_1, W_2 \in \mathcal{F}$. Z twierdzenia Helly'ego przecięcie wszystkich odcinków postaci $f(W)$, $W \in \mathcal{F}$ jest niepuste. Wiadomo już zatem jaki okrąg wybrać, prawda?

Zadanie 9.4. *Na płaszczyźnie leży n prostokątów o bokach równoległych do osi OX i OY: $P_1, P_2, \dots, P_n$ o tej własności, że dowolne dwa mają punkt wspólny. Pokaż, ze wszystkie prostokąty mają punkt wspólny.*

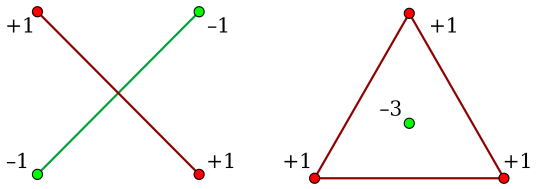
Rozwiązanie. Zrzutujmy te prostokąty na osie układu współrzędnych. Obraz każdego prostokąta po zrzutowaniu na dowolną z osi jest odcinkiem. Niech f, g będą tymi rzutowaniami. Wówczas dla dowolnych dwóch prostokątów P_1, P_2 rozważanej rodziny mamy: $f(P_1) \cap f(P_2) \neq \emptyset$. Podobnie dla g. Stosując na osi OX twierdzenie Helly'ego do rodziny $f(P_i)$ dostajemy, że istnieje $x \in OX$, że $x \in \bigcap f(P_i)$. Podobnie na osi OY istnieje punkt $y \in OY$, który zawiera się w przecięciu $g(P_i)$. Zatem punkt (x, y) należy do przecięcia wszystkich tych prostokątów. ■

Widzimy zatem, że można układać różne sympatyczne zadania związane z twierdzeniem Helly'ego. Zanim udowodnimy to twierdzenie potrzebujemy ważnego lematu z teorii zbiorów wypukłych.

[1] Czy przez dowolny punkt P można przeprowadzić pewną prostą przecinającą każdy wielokąt z $\mathcal{F}$?

Obserwacja 9.4: Radon

Niech $S = \{x_0, x_1, \ldots, x_r\}$ będzie zbiorem punktów przestrzeni $\mathbb{R}^n$, gdzie $r \geqslant n+1$. Wówczas można S podzielić na dwa niepuste i rozłączne podzbiory S_1, S_2, spełniające: $\operatorname{conv}(S_1) \cap \operatorname{conv}(S_2) \neq \emptyset$.



Dowód. Rozważmy wektory $y_i = (x_i, 1) \in \mathbb{R}^{n+1}$. Mamy $r + 1 \geqslant n + 2$ wektorów w $n + 1$ wymiarowej przestrzeni liniowej, więc istnieją $\lambda_0, \ldots, \lambda_r$, nie wszystkie równe zero, że

$$\lambda_0 y_0 + \ldots + \lambda_r y_r = 0.$$

Niech $S_1 = \{i \,|\, \lambda_i \geqslant 0\}$ oraz niech $S_2 = \{i \,|\, \lambda_i < 0\}$. Skoro ostatnia współrzędna y_i równa jest 1, mamy:

$$\sum_{i \in S_1} \lambda_i = -\sum_{j \in S_2} \lambda_j = \lambda \neq 0.$$

W szczególności punkt

$$\sum_{i \in S_1} \frac{\lambda_i}{\lambda} x_i = \sum_{j \in S_2} \frac{-\lambda_j}{\lambda} x_j$$

należy zarówno do $\operatorname{conv}(S_1)$ jak i do $\operatorname{conv}(S_2)$, co kończy dowód. □

Można udowodnić także odwrotną wersję tego lematu: jeśli S jest dowolnym skończonym podzbiorem $\mathbb{R}^n$, afinicznie niezależnym, wówczas dla każdego podziału S na dwa niepuste i rozłączne podzbiory S_1, S_2 mamy $\operatorname{conv} S_1 \cap \operatorname{conv} S_2 \neq \emptyset$.

Dowód twierdzenia Helly'ego. Niech $\mathcal{F} = \{B_1, B_2, \ldots B_r\}$. Dowód to indukcja ze względu na $|\mathcal{F}| = r$. Dla $r = n + 1$ teza jest oczywista. Niech $r \geqslant n + 2$ i załóżmy, że teza jest spełniana dla mniejszych r. Rozważmy zbiory

$$\widehat{B}_i = \bigcap_{j \neq i} B_i.$$

Zgodnie z założeniem indukcyjnym wszystkie te zbiory są niepuste, więc istnieją takie punkty $x_1, x_2, \ldots, x_r$, że $x_i \in \widehat{B}_i$. Skoro $r \geqslant n + 2$, to możemy zaaplikować do tego zbioru punktów – nazwijmy go S – lemat Radona i stwierdzić bez straty ogólności, że istnieje $S_1 = \{x_1, x_2, \ldots x_k\}$, że $\operatorname{conv} S_1 \cap \operatorname{conv}(S \setminus S_1) \neq \emptyset$. Niech więc:

$$x \in \operatorname{conv}(x_1, x_2, \ldots, x_k) \cap \operatorname{conv}(x_{k+1}, x_{k+2}, \ldots, x_r).$$

Wówczas twierdzimy, że x jest w przecięciu wszystkich B_i (to da tezę). Istotnie, weźmy $x_i \in \{x_1, x_2, \ldots x_k\}$. Wiemy, że dla $j > k$ mamy $x_j \in \widehat{B}_i \subseteq B_i$. Skoro B_i jest wypukły, to

$$\operatorname{conv}(x_{k+1}, x_{k+2}, \ldots, x_r) \subseteq B_i.$$

Ale również x należy do $\operatorname{conv}(x_{k+1}, x_{k+2}, \ldots, x_r)$, więc $x \in B_i$, dla $1 \leqslant i \leqslant k$. Analogicznie pokazujemy, że $x \in B_j$, gdzie $k + 1 \leqslant j \leqslant r$.

Łatwo pokazać, że założeń tw. Helly'ego nie da się w żadnym przypadku istotnie poprawić. Nie można oczywiście opuścić założenia wypukłości (prosty kontrprzykład). Nie jest też możliwe obniżenie założenia, że przecinać się (w $\mathbb{R}^n$) musi każda podrodzina $n + 1$ zbiorów. Inaczej bowiem boki dowolnego trójkąta stanowią kontrprzykład do tego twierdzenia.

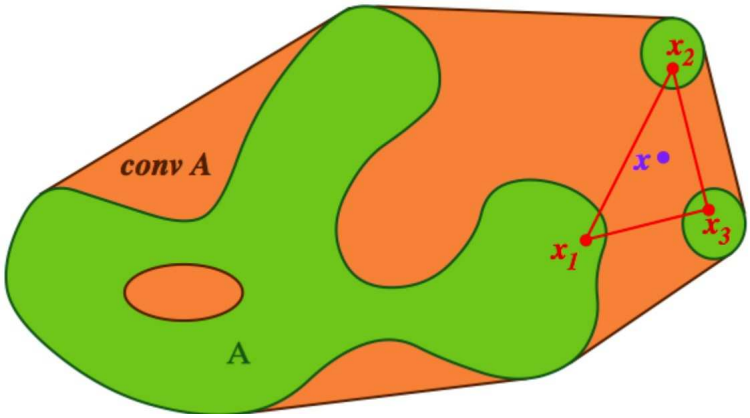
Podstawowym rezultatem, ważnym dla teorii zbiorów wypukłych, jest również następujący wynik.

Twierdzenie 9.2: Caratheodory

Jeśli X jest niepustym podzbiorem przestrzeni afinicznej $\mathbb{R}^n$ oraz $p \in \text{conv}(X)$, wówczas p jest kombinacją wypukłą

$$p = a_0x_0 + \ldots + a_nx_n,$$

dla pewnych $x_0, \ldots, x_n \in X$ oraz nieujemnych wag $a_0, \ldots, a_n$, których suma wynosi 1.



Dowód. Skoro $p \in \text{conv}(X)$, istnieje skończona kombinacja wypukła $k+1$ punktów ze zbioru X, która daje p, czyli

$$p = a_0x_0 + \ldots + a_kx_k,$$

gdzie $a_i \geqslant 0$ oraz $a_0 + \ldots + a_k = 0$. Jeśli $k+1 \leqslant n+1$, to nie ma czego dowodzić, Przypuśćmy, że $k+1 > n+1$. Wykażemy, że istnieje podzbiór S zbioru X o liczbie elementów mniejszej niż $k+1$, że $p \in \text{conv}(S)$. To będzie oznaczało, że możemy redukować $k+1$ aż otrzymamy $n+1$.

Rozważmy $n+1$ wektorów łączących $\overrightarrow{x_0x_1}, \ldots, \overrightarrow{x_0x_{n+1}}$. Wektory te są są liniowo zależne w $\mathbb{R}^n$, więc istnieją takie $b_1, \ldots, b_{n+1}$, nie wszystkie równe 0, że

$$b_1\overrightarrow{x_0x_1} + \ldots + b_{n+1}\overrightarrow{x_0x_{n+1}} = 0.$$

Niech $b_0 = -b_1 - \ldots - b_{n+1}$ oraz niech $b_j = 0$, dla $j > n+1$. Wówczas mamy $\sum_i b_ip_i = 0$ oraz $\sum_i b_i = 0$. Stąd dla każdego $\lambda \in K$ mamy:

$$p = \sum_i a_ix_i = \sum_i a_ix_i - \lambda\sum_i b_ix_i, \quad \text{zatem} \quad p = \sum_i (a_i - \lambda b_i)x_i.$$

Skoro powyższe zachodzi dla każdego λ, to wybieramy λ takie, by $a_i - \lambda b_i \geqslant 0$, ale tak, by co najmniej jedna z tych nierówności była równością. W ten sposób uzyskujemy przedstawienie p jako kombinacji wypukłej mniej niż $k+1$ punktów ze zbioru $\text{conv}(X)$, co kończy dowód. □

Twierdzenie Helly'ego było pierwszym rezultatem tzw. **dyskretnej geometrii kombinatorycznej**. Typowe zagadnienia tej dziedziny związane są między innymi z kolorowaniem obiektów geometrycznych, z najlepszymi pokryciami, upakowaniami lub wypełnieniami, z podziałami i rozkładami figur, z ich symetriami. Choć znakomita większość problemów do niej należących da się wysłowić w sposób całkowicie elementarny, ich dowody nie należą zawsze do łatwych. Oto kilka zadań z tego typu geometrii[2].

Zadanie 9.5. *Na płaszczyźnie dane jest pięć punktów. Żadne trzy z nich nie są współliniowe. Wówczas można z nich wybrać cztery będące wierzchołkami czworokąta wypukłego.*

Zadanie 9.6. *Na płaszczyźnie dane jest sześć punktów. Żadne trzy z nich nie są współliniowe. Wówczas można utworzyć dwa rozłączne trójkąty o wierzchołkach w tych sześciu punktach.*

Zadanie 9.7. *Na płaszczyźnie dane jest siedem punktów. Wówczas istnieją takie punkty A, B (niekoniecznie spośród danych siedmiu), że jeśli pewien zbiór wypukły X zawiera przynajmniej cztery z wybranych siedmiu punktów, to X zawiera też A lub B.*

[2]Zachęcam też do zajrzenia do poniższych notatek `https://pi.math.cornell.edu/~eranevo/homepage/ConvNote.pdf`. Czytelnika zainteresowanego porządnym wykładem z licznymi wątkami współczesnymi odsyłam do notatek T. Tkocza: `https://www.math.cmu.edu/~ttkocz/teaching/1920/conv-discr-geom-notes.pdf`.

9.1 Trivia. Spektrum macierzy stochastycznej

Celem naszych rozważań będzie rozwiązanie zadania dotyczącego ważnych dla zastosowań, zwłaszcza dla teorii probabilistycznych, macierzy stochastycznych.

Definicja 9.6: Macierz stochastyczna

Macierz $P \in M_n(\mathbb{C})$ nazwiemy STOCHASTYCZNĄ, jeśli wszystkie jej wyrazy są nieujemnymi liczbami rzeczywistymi oraz gdy suma wyrazów w każdym jej wierszu wynosi 1.

Zadanie 9.8. *Niech M_n będzie podzbiorem $\mathbb{C}$ złożonym ze wszystkich wartości własnych macierzy stochastycznych w $M_n(\mathbb{C})$. Przez wielokąt wypukły o wierzchołkach $p_1, \ldots, p_r \in \mathbb{C}$ rozumieć będziemy podzbiór $\mathbb{C}$ złożony ze wszystkich elementów postaci $a_1p_1 + \ldots + a_rp_r$, gdzie $a_i \in \mathbb{R}$, $a_i \geqslant 0$. dla $i = 1, \ldots, r$, oraz $a_1 + \ldots + a_r = 1$.*

(a) Niech $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$ będzie wektorem własnym macierzy stochastycznej P o wartości własnej λ. Wykaż, że jeśli $Q \subset \mathbb{C}$ jest wielokątem wypukłym o wierzchołkach we współrzędnych wektora z, to zachodzi zawieranie $\lambda \cdot Q = \{\lambda \cdot q \,|\, q \in Q\} \subseteq Q$.

(b) Pokaż, że M_n zawiera 1, jest symetryczny względem prostej $\mathrm{Re}(z)$ oraz, że $M_n \subset \{z : |z| \leqslant 1\}$. Wykaż ponadto, że jeśli $\lambda \in M_n$ i $|\lambda| = 1$, to λ jest pierwiastkiem z 1.

(c) Wykaż, że jeśli $\lambda_1, \lambda_2, \lambda_3$ są zespolonymi wartościami własnymi macierzy P rozmiaru 3×3 o nieujemnych wyrazach rzeczywistych, to $3(\lambda_1^2 + \lambda_2^2 + \lambda_3^2) \geqslant (\lambda_1 + \lambda_2 + \lambda_3)^2$. Wywnioskuj stąd, że zbiór $M_3 \setminus \mathbb{R}$ jest zawarty w trójkącie równobocznym o wierzchołkach w pierwiastkach stopnia 3 z 1.

Rozwiązanie (a). Niech $z = (z_1, \ldots, z_n)$ będzie wektorem własnym macierzy stochastycznej $P = (p_{ij})$ o wartości własnej λ, czyli $Pz = \lambda z$. Przechodząc w tej równości do współrzędnej $i = 1, \ldots, n$ mamy

$$\lambda z_i = p_{i1}z_1 + p_{i2}z_2 + \ldots + p_{in}z_n. \qquad (*)$$

Rozważając wielokąt wypukły Q o wierzchołkach $z_1, \ldots, z_n$ (nie wszystkie one są zerami) oraz dowolny należący doń punkt $q = q_1z_1 + \ldots + q_nz_n$, gdzie $q_i \geqslant 0$ oraz $q_1 + \ldots + q_n = 1$ mamy, na mocy $(*)$:

$$\begin{aligned}\lambda q &= \lambda(q_1z_1 + \ldots + q_nz_n) = \\ &= q_1(p_{11}z_1 + p_{12}z_2 + \ldots + p_{1n}z_n) + \ldots + q_n(p_{n1}z_1 + p_{i2}z_2 + \ldots + p_{nn}z_n) = \\ &= (q_1p_{11} + q_np_{n1})z_1 + \ldots + (q_np_{n1} + \ldots + q_np_{nn})z_n.\end{aligned}$$

Oczywiście $q_1p_{1i} + \ldots + q_np_{ni} \geqslant 0$, zaś suma tych współczynników to $q_1 + \ldots + q_n = 1$. Zatem $\lambda q \in Q$.

Uwaga. Wykazaliśmy, że jeśli $\lambda \in M_n$, to istnieje wielokąt wypukły Q taki, że $\lambda Q \subseteq Q$. Zachodzi też fakt odwrotny. Istotnie, niech $\lambda Q \subset Q$, dla pewnego wielokąta wypukłego o wierzchołkach $z_1, \ldots, z_q$, $q \leqslant n$. Wtedy λz_j leży w Q, a zatem istnieją takie $p_{ij} \geqslant 0$ oraz $p_{i1} + \ldots + p_{iq} = 1$, że $\lambda z_i = p_{i1}z_1 + \ldots + p_{iq}z_q$. Niech P będzie macierzą, która w pierwszych q wierszach i kolumnach ma wyrazy p_{ij}, a poza nimi zera. Wówczas macierz P jest stochastyczna i ma wektor własny $(z_1, \ldots, z_q, \underbrace{0, \ldots, 0}_{n-q})$ o wartości własnej λ.

Rozwiązanie (b). Oczywiście dla dowolnej macierzy stochastycznej rozmiaru n wektor złożony z samych jedynek jest wektorem własnym o wartości własnej 1. Zbiór M_n jest sumą mnogościową zbiorów pierwiastków zespolonych wielomianów o współczynnikach rzeczywistych, a zatem z każdym pierwiastkiem nierzeczywistym z takiego wielomianu należącym do M_n, do zbioru M_n wchodzi także sprzężony z nim pierwiastek $\overline{z}$. Wreszcie, korzystając[3] z (a) bierzemy wielokąt Q odpowiadający wartości własnej λ i wybieramy z niego takie z, że $|z|$ jest maksymalny. Wówczas $\lambda z \in \lambda Q \subseteq Q$, zatem $|\lambda z| \leqslant |z|$, czyli $|\lambda| \leqslant 1$.

[3]Można także argumentować bez punktu (a) rozważając na $\mathbb{C}^n$ (nad $\mathbb{R}$) tzw. normę maksimum $\|\cdot\|_\infty$ zadaną wzorem: $\|(z_1, \ldots, z_n)\| = \max(|z_i|)$. Wówczas rozważając na $M_n(\mathbb{C})$ tzw. normę wierszową, tzn. $\|P\|_\infty = \max_i \sum_{j=1}^n |p_{ij}|$ oraz pokazując, że $\|\lambda v\|_\infty = \|Pv\|_\infty \leqslant \|P\|_\infty \cdot \|v\|_\infty$, dostajemy $|\lambda| \leqslant 1$. To znana sztuczka z teorii zbiorów wypukłych.

Pozostało wykazanie, że jeśli $\lambda \in M_n$ oraz $|\lambda| = 1$, to λ jest pierwiastkiem z 1. Weźmy znowu wielokąt Q odpowiadający λ i taki punkt $0 \neq z \in Q$, że $|z|$ jest maksymalny. Jeśli $z_1, \ldots, z_n$ to wierzchołki Q, to

$$|z| = |a_1 z_1 + \ldots + a_n z_n| \leqslant a_1|z_1| + \ldots + a_n|z_n| \leqslant (a_1 + \ldots + a_n) \max_i |z_i| = \max_i |z_i| \leqslant |z|.$$

Równość w powyższej nierówności zachodzi tylko gdy wszystkie z_i są proporcjonalne. A zatem z jest jednym z wierzchołków. Zatem $\lambda z, \lambda^2 z, \ldots$, które również należą do Q (bo $\lambda Q \subseteq Q$), mają również moduł $|z|$ (bo $|\lambda| = 1$), więc powtarzając powyższy argument – są wierzchołkami Q. Skoro jednak Q ma skończenie wiele wierzchołków, to istnieją s, t, że $\lambda^{s+t} z = \lambda^s z$. Skoro $|\lambda| \neq 0$, to $z \neq 0$, czyli $\lambda^t = 1$.

Rozwiązanie (c). Niech $P = (p_{ij})$ będzie macierzą o wyrazach nieujemnych i wartościach własnych $\lambda_1, \lambda_2, \lambda_3$. Oczywiście $\mathrm{tr}(P) = \lambda_1 + \lambda_2 + \lambda_3$. Macierz P^2 ma wartości własne $\lambda_1^2, \lambda_2^2, \lambda_3^2$, a zatem z tego, że $p_{ij} \geqslant 0$ oraz z nierówności Schwarza $\|(p_{11}, p_{22}, p_{33})\|^2 \cdot \|(1,1,1)\|^2 \geqslant \langle (p_{11}, p_{22}, p_{33}), (1,1,1) \rangle_{st}^2$ mamy:

$$\begin{aligned}\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \mathrm{tr}(P^2) = \\ = p_{11}p_{11} + p_{12}p_{21} + p_{13}p_{31} + p_{21}p_{12} + p_{22}p_{22} + p_{23}p_{32} + p_{31}p_{13} + p_{32}p_{23} + p_{33}p_{33} \geqslant \\ \geqslant p_{11}^2 + p_{22}^2 + p_{33}^2 \geqslant \frac{1}{3}(p_{11} + p_{22} + p_{33})^2 = \frac{1}{3}(tr(P))^2 = \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)^2.\end{aligned}$$

Teraz pokazujemy tezę dotyczącą zbioru M_3. Zauważmy, że jeśli macierz P rozważana wyżej jest stochastyczna, to jej wielomian charakterystyczny ma pierwiastki $1, \lambda, \overline{\lambda}$, dla pewnego $\lambda \in \mathbb{C}$. Jak wyżej, $\mathrm{tr}(P) = \lambda + \overline{\lambda} + 1$. Skoro P ma wyrazy nieujemne to jej ślad też jest nieujemny, co oznacza, że dla $\lambda \in M_3$ mamy $\mathrm{Re}(\lambda) \geqslant -\frac{1}{2}$. Oczywiście $\mathrm{Re}(z) = -\frac{1}{2}$ to prosta zawierająca jeden z boków rozważanego w tezie trójkąta. Proste zawierające pozostałe boki opisane są równaniami: $1 - \mathrm{Re}(z) \pm \sqrt{3}\,\mathrm{Im}(z) = 0$. Zauważmy, że nierówność wykazana wyżej przybiera dla macierzy P (o wartościach własnych $1, \lambda, \overline{\lambda}$) postać:

$$3(1 + 2\,\mathrm{Re}(\lambda)^2) \geqslant (1 + 2\,\mathrm{Re}(\lambda))^2,$$

czyli

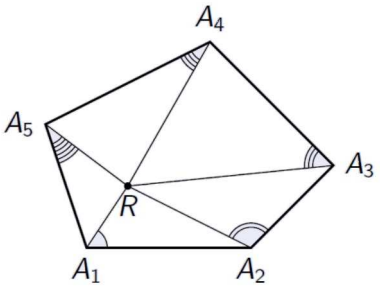
$$(1 - \mathrm{Re}(\lambda))^2 \geqslant 3\,\mathrm{Im}(\lambda)^2.$$

A zatem:

$$(1 - \mathrm{Re}(\lambda) - \sqrt{3}\,\mathrm{Im}(\lambda))(1 - \mathrm{Re}(\lambda) + \sqrt{3}\,\mathrm{Im}(\lambda)) \geqslant 0.$$

Biorąc pod uwagę, że M_3 zawiera się w dysku $|z| \leqslant 1$, dostajemy tezę.

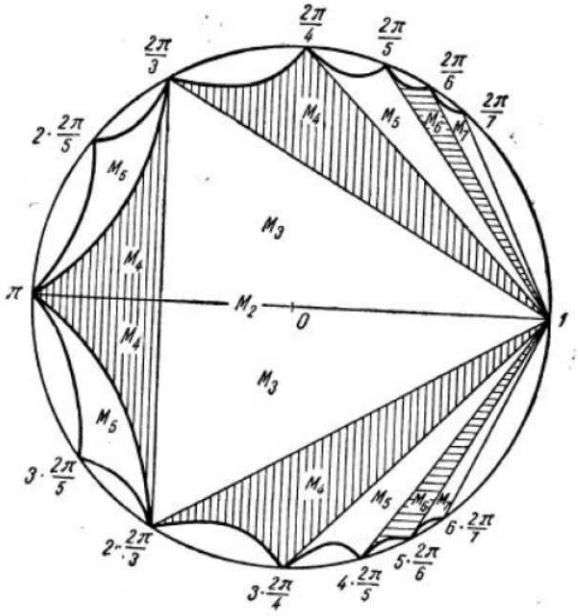
KOMENTARZ. Problem określenia położenia spektrum macierzy stochastycznych postawił w 1938 roku wielki radziecki matematyk Andriej Kołmogorow. W roku akademickim 1944-45 podczas seminarium dotyczącego teorii procesów Markova pytanie Kołmogorowa trafiło do grupy studentów. Niedługo później dwójka dwudziestolatków: K. Dmitriev oraz E. Dynkin znalazła niezwykły pomysł ataku tego problemu. Dmitriev został przyjęty na Uniwersytet w wieku 14 lat i był powszechnie uznawany za geniusza. Dynkin przyszedł na uczelnię jako szesnastolatek i okazał się nie mniej wybitnym. Obydwaj doktoryzowali się u Kołmogorowa. Tej dwójce udało się dokonanie geometrycznego przeformułowania problemu, czego fragment widzieliśmy w zadaniu. Kolejne pomysły to między innymi uogólnienie twierdzenia o kącie Brocarda, ograniczające, dla n-kąta wypukłego, miarę najmniejszego z kątów zaznaczonych niżej, po wszystkich R (piękne i proste rozwiązanie w prezentacji niżej; dla $n = 3$ było to zadanie na IMO 1991):



Swoje wyniki publikowali z Kołmogorowem w latach 1945-46. W roku 1951 problem doczekał się całkowitego rozwiązania przez studenta Dynkina: F. Karpielewicza. To właśnie nazwiskiem Dynkina opatrzone są grafy, dla których forma Titsa (komentarz do zadania 3) jest dodatnio określona. Warto zajrzeć do prezentacji Ádáma Besenyei z Eötvös Loránd University pt. *Stochastic matrices and geometry: a gem from Dmitriev and Dynkin*, poświęconej temu problemowi: `https://abesenyei.web.elte.hu/publications/dynkin.pdf`.

ТЕОРЕМА VII. *Для* $n \leqslant 5$ M_n *будет объединением циклических* k*-угольников* $(k \leqslant n)$.

В заключение приведем чертеж (фиг. 11), схематически изображающий фигуры M_2, M_3, M_4, M_5 и те части фигур M_n для $n > 5$, которые нам известны.



Z oryginalnej pracy Dmitrieva i Dynkina: opis zbiorów M_2, M_3, M_4, M_5. Jest on do zrozumienia! Źródło: `https://abesenyei.web.elte.hu/publications/dynkin.pdf`.

Rozdział 10

Wielościany, stożki i dualność

Na poprzednim wykładzie wprowadziliśmy podstawowy język związany z opisem zbiorów wypukłych. Celem tego wykładu jest przyjrzenie się pewnym szczególnym klasom zbiorów wypukłych, mającym duże znaczenie geometryczne i ważne miejsce w zastosowaniach. Na poprzednim wykładzie podawaliśmy półprzestrzeń jako przykład zbioru wypukłego. Powiedzieliśmy też, że przecięcie zbiorów wypukłych jest wypukłe. Wychodząc od tych faktów formułujemy definicję bryły wielościennej i wielościanu wypukłego.

Definicja 10.1: Wypukła bryła wielościenna i wielościan wypukły

Rozważmy układ m nierówności o n zmiennych $x_1, \ldots, x_n$ i współczynnikach rzeczywistych postaci:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leqslant b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leqslant b_2 \\ \quad\vdots \qquad\quad \vdots \qquad\qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leqslant b_m \end{cases} \qquad (*)$$

Zbiór rozwiązań układu $(*)$ nazywamy (WYPUKŁĄ) BRYŁĄ WIELOŚCIENNĄ i oznaczamy $\Delta = \Delta(*)$. Jeśli dodatkowo zbiór ten jest ograniczony, czyli nie zawiera półprostej $\{x + ty \,|\, t \geqslant 0\}$, dla żadnego $y \neq 0$ to Δ nazywamy WIELOŚCIANEM WYPUKŁYM. Definiujemy też $\dim(\Delta) = \dim \operatorname{lin}(\Delta)$.

Przy rozważaniu przestrzeni rzeczywistych użyteczna będzie dla nas następująca konwencja. Powiemy, że dla wektorów $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ mamy $v \leqslant w$, jeśli $v_i \leqslant w_i$, dla każdego $1 \leqslant i \leqslant n$. Układ nierówności $(*)$ możemy zatem oznaczyć krótko jako $Ax \leqslant b$, gdzie $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ oraz $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$.

Podstawowe przykłady brył wielościennych i wielościanów znamy rzecz jasna z geometrii elementarnej.

- Dowolna podprzestrzeń afiniczna w $\mathbb{R}^n$ jest bryłą wielościenną (opisujący ją układ m równań ma ten sam zbiór rozwiązań, co odpowiedni układ $2m$ nierówności).
- Półprzestrzeń w $\mathbb{R}^1$ nazywamy PÓŁPROSTĄ. Wielościanami w $\mathbb{R}^1$ są: prosta $\mathbb{R}^1$, półproste i wszystkie odcinki domknięte. Przedział $[0, 1)$ nie jest wielościanem.
- W przestrzeni $\mathbb{R}^2$ zbiór wektorów (x_1, x_2) spełniający układ $-x_1 \leqslant 0, -x_2 \leqslant 0$ odpowiada pierwszej ćwiartce – wektorom o obydwu współrzędnych nieujemnych.
- Wielościan wypukły w dwuwymiarowej przestrzeni afinicznej nazywamy WIELOKĄTEM WYPUKŁYM.
- Kostką n-wymiarową nazywamy n-wymiarowe uogólnienie sześcianu, czyli zbiór:

$$C_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid -1 \leqslant x_i \leqslant 1\} = \operatorname{conv}(\{(\delta_1, \ldots, \delta_n), \text{ gdzie } \delta_i \in \{-1, 1\}\})$$

- Warto jednocześnie zdefiniować uogólnienie trójwymiarowego ośmiościanu foremnego:

$$C_n^{\triangle} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \,|\, \delta_1 x_1 + \ldots + \delta_n x_n \leqslant 1, \text{ gdzie } \delta_i \in \{-1, 1\}\} = \operatorname{conv}(\{\epsilon_1, -\epsilon_1, \ldots, -\epsilon_n, \epsilon_n\}).$$

Oznaczenie w ostatnim przykładzie związane jest z wzajemną dualnością (polarnością) sześcianu i ośmiościanu. Wyraża się ona w następującym warunku (przyjmujemy izomorfizm $\mathbb{R}^n$ i $(\mathbb{R}^n)^*$ przez $\mathrm{st} \mapsto \mathrm{st}^*$):

$$C_n^{\triangle} = \{\lambda \in (\mathbb{R}^n)^* : \lambda(x) \leqslant 1, \text{ dla każdego } x \in C_n\}, \quad C_n = \{\lambda \in (\mathbb{R}^n)^* : \lambda(x) \leqslant 1, \text{ dla każdego } x \in C_n^{\triangle}\}.$$

Dualności wielościanów poświęcimy dodatek do tego wykładu. Odnotujmy natomiast w tym miejscu, że szereg ważnych dla zastosowań problemów polega na szukaniu ekstremów funkcji liniowych wielu zmiennym na pewnym wielościanie. Ich waga polega na tym, że nie wymagają one stosowania rachunku różniczkowego. Zagadnieniami tymi zajmuje się dział zwany programowaniem liniowym, wykładany na naszym Wydziale jako Optymalizacja. Zainteresowanego Czytelnika odsyłam choćby do skryptu dr. Andrzeja Strojnowskiego `https://mst.mimuw.edu.pl/wyklady/op1/wyklad.pdf`. Naszym zaś celem będzie klasyczny wynik – opis wielościanów wypukłych jako otoczek wypukłych skończonych układów punktów.

Definicja bryły wielościennej nie zależy od przestrzeni afinicznej, w której ją rozpatrujemy. Prostym ćwiczeniem dla Czytelnika jest następujący fakt: jeżeli $W \neq \emptyset$ jest podzbiorem w przestrzeniach afinicznych $H_1 \subseteq H_2$, to W jest bryłą wielościenną w H_1 wtedy i tylko wtedy, gdy jest bryłą wielościenną w H_2.

Definicja 10.2: Hiperpłaszczyzna podpierająca, brzeg bryły wielościennej

Niech $W \subseteq \mathbb{R}^n$ będzie zbiorem wypukłym. Powiemy, że $n-1$-wymiarowa podprzestrzeń afiniczna $H \subset \mathbb{R}^n$ jest HIPERPŁASZCZYZNĄ PODPIERAJĄCĄ zbiór W w punkcie $p = (p_1, \ldots, p_n) \in W$, jeżeli H jest opisana równaniem $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$, gdzie $(a_1, \ldots, a_n) \in \mathbb{R}^n$, $b \in \mathbb{R}$ spełniają:

- zbiór W „leży po jednej stronie" H, czyli dla każdego $x = (x_1, \ldots, x_n) \in W$ zachodzi nierówność $a_1x_1 + a_2x_2 + \ldots + a_nx_n \leqslant b$,
- punkt p należy do H, czyli $a_1p_1 + a_2p_2 + \ldots + a_np_n = b$.

Niech W będzie bryłą wielościenną. Zbiór punktów $p \in W$, dla których istnieje płaszczyzna podpierająca W w punkcie p nazwiemy BRZEGIEM W.

Z uwagi na to, że nie chcemy wprowadzać elementów analitycznych, definiujemy brzeg jedynie dla bryły wielościennej, co pozwala na użycie płaszczyzny podpierającej, zamiast pojęć metrycznych. Można udowodnić, że dla brzegu ∂W zdefiniowanego w sposób analityczny[1] zachodzi twierdzenie: jeśli W jest zbiorem wypukłym i domkniętym oraz $p \in \partial W$, to istnieje hiperpłaszczyzna podpierająca W w punkcie p.

Definicja 10.3: Ściana, krawędź, wierzchołek zbioru wypukłego

ŚCIANĄ zbioru wypukłego W nazwiemy zbiór $W \cap H$, gdzie H jest hiperpłaszczyzną podpierającą W. WYMIAREM ŚCIANY $W \cap H$ nazywamy liczbę $\dim \mathrm{af}(W \cap H)$. Ścianę wymiaru 0 nazywamy WIERZCHOŁKIEM zbioru W, zaś ścianę wymiaru 1 nazywamy KRAWĘDZIĄ.

Pojęcie wymiaru ściany jest oczywiście dobrze określone (np. punkt wewnętrzny krawędzi nie jest wierzchołkiem). Dla opisu ścian brył wielościennych przydatne jest pojęcie półprzestrzeni dopełniającej.

Definicja 10.4: Półprzestrzeń dopełniająca

Niech $H = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_1x_1 + \ldots + a_nx_n \leqslant b\}$ będzie półprzestrzenią. PÓŁPRZESTRZENIĄ DOPEŁNIAJĄCĄ H nazywamy półprzestrzeń $H^- = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_1x_1 + \ldots + a_nx_n \geqslant b\}$.

Odnotujmy następujące łatwe w dowodzie obserwacje. Zostawiamy je jako ćwiczenie.

- Jeśli H jest półprzestrzenią w $\mathbb{R}^n$, to $\mathbb{R}^n = H \cup H^-$, natomiast $H \cap H^-$ jest brzegiem zarówno H, jak i H^-. Zbiór ten jest, zgodnie z definicją wyżej, jedyną ścianą półprzestrzeni H.
- Niech $W = \bigcap_{i=1}^t H_i$ będzie bryłą wielościenną, gdzie H_i są półprzestrzeniami. Wówczas $W \cap H^-$ jest ścianą W lub zbiorem pustym (patrz np. $H_1 \cap H_2^-$, gdzie $H_1 : x_1 \in \mathbb{R} : x_1 \geqslant 2, H_2 : x_1 \in \mathbb{R} : x_1 \geqslant 1$). Co więcej, jeśli dla pewnego s zbiór $S = W \cap \bigcap_{i=1}^s H_i$ jest niepusty, to jest on ścianą W.

[1] Tzn. $\partial W = \{p \in \mathbb{R}^n : \forall_{\epsilon>0} \exists_{q_1,q_2}\ q_1 \in K(p,\epsilon) \cap W, q_2 \in K(p,\epsilon) \setminus W\}$, gdzie $K(p,\epsilon)$ jest kulą o środku p i promieniu ϵ.

Obserwacja 10.1: Opis ścian brył wielościennych

Niech $p = (p_1, \ldots, p_n)$ będzie punktem bryły wielościennej $W \subset \mathbb{R}^n$ będącej przecięciem t półprzestrzeni postaci $H_i = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_{i1}x_1 + \ldots + a_{in}x_n \leqslant b_i\}$, dla $1 \leqslant i \leqslant t$, przy czym zakładamy, że punkt p należy do brzegu dokładnie s pierwszych półprzestrzeni, tzn.

$$a_{i1}p_1 + \ldots + a_{in}p_n = b_i, \text{ dla każdego } 1 \leqslant i \leqslant s,$$
$$a_{j1}p_1 + \ldots + a_{jn}p_n < b_j, \text{ dla każdego } s+1 \leqslant j \leqslant t.$$

Rozważmy macierz $A_p = [a_{ij}] \in M_{t\times n}(\mathbb{R})$. Wówczas zbiór

$$S = W \cap \bigcap_{i\leqslant s} H_i^-$$

jest ścianą wymiaru $n - r(A_p)$ zawierającą p.

Dowód. Jeśli przez H określimy zbiór rozwiązań układu s równań postaci $a_{i1}x_1 + \ldots + a_{in}x_n = b_i$, to na mocy tw. Kroneckera-Capellego zbiór H jest przestrzenią afiniczną wymiaru $n - r(A_p)$. Mamy też $H = \bigcap_{i\leqslant s} H_i \cap \bigcap_{i\leqslant s} H_i^-$. Oczywiście $S \subseteq H$, więc $\dim S \leqslant n - r(A_p)$.

Niech $m = n - r(A_p)$ i niech układ $\alpha_1, \ldots, \alpha_m$ wektorów w $\mathbb{R}^n$ stanowi bazę przestrzeni liniowej $T(H)$. Uzupełnijmy ten układ do bazy $\alpha_1, \ldots, \alpha_n$ całej przestrzeni $\mathbb{R}^n$. Wówczas w układzie bazowym $p; \alpha_1, \ldots, \alpha_n$ punkt należy do H wtedy i tylko wtedy, gdy ma ostatnie $n - m$ współrzędnych zerowych. Skoro dla $j > s$ punkt p należy do każdej z półprzestrzeni H_j i nie leży na jej brzegu – mamy bowiem $a_{j1}p_1 + \ldots + a_{jn}p_n < b_j$ – to dla każdego $i \leqslant m$ istnieje dostatecznie małe r_i, że $p + r_i\alpha_i$ również należy do przecięcia wszystkich H_j, dla $t \geqslant j > s$. Jednocześnie $p + r_i\alpha_i$ jest w H, ponieważ jego ostatnie $n - m$ współrzędnych w układzie bazowym $p; \alpha_1, \ldots, \alpha_n$ to zera. A zatem zbiór S zawiera układ $m+1$ afinicznie niezależnych punktów $\{p, p + r_1\alpha_1, \ldots, p + r_m\alpha_m\}$, czyli $\dim S \geqslant m$. W rezultacie $\dim S = m$. □

Uzyskany wynik pozwala sprawdzać kiedy punkt przestrzeni afinicznej jest wierzchołkiem wielościanu. W tym sensie możemy uzyskać prosty algorytm szukania wierzchołków czy krawędzi wielościanu – istotny dla rozwiązywania praktycznych problemów programowania liniowego. Naszym celem będzie natomiast dowód twierdzenia strukturalnego, wymagającego zastosowania tzw. teorii dualności.

Twierdzenie 10.1: Twierdzenie strukturalne – Minkowski, Steinitz, Weyl

Następujące warunki są równoważne dla podzbioru $\Delta \subseteq \mathbb{R}^n$.

(1) podzbiór Δ jest wielościanem wypukłym,

(2) $\Delta = \operatorname{conv}(p_0, \ldots, p_r)$, dla pewnego układu punktów $p_1, \ldots, p_r$ w $\mathbb{R}^n$.

Definicja 10.5: Stożek

Dla dowolnego podzbioru $X \subseteq \mathbb{R}^n$ definiujemy STOŻEK WYPUKŁY $\operatorname{cone}(X)$ rozpięty (generowany) przez X jako zbiór nieujemnych kombinacji liniowych układów punktów z X:

$$\operatorname{cone}(X) = \{a_1p_1 + \ldots + a_rp_r \mid p_1, \ldots, p_r \in X, a_1, \ldots, a_r \geqslant 0\}.$$

Definiujemy też WYMIAR STOŻKA jako $\dim \operatorname{lin}(\operatorname{cone}(X))$. Jeśli X jest zbiorem skończonym to stożek $\operatorname{cone}(X)$ nazywamy SKOŃCZENIE GENEROWANYM.

Jak widzimy, definicja wykorzystuje wektor zerowy w $\mathbb{R}^n$. Alternatywnie można definiować stożek $C \subseteq \mathbb{R}^n$ tak, aby dla każdego niezerowego $x \in C$ zachodziło $\{tx \in C \mid t \geqslant 0\}$. Innymi słowy, stożek jest sumą półprostych. Jest jasne, że stożek jest zbiorem wypukłym. Wykażemy, że w istocie stożek skończenie generowany jest bryłą wielościenną, a nawet przecięciem skończenie wielu półprzestrzeni przechodzacych przez 0. Teoria dowolnych stożków wymaga bardziej zaawansowanych narzędzi.

Twierdzenie 10.2: Twierdzenie Fouriera-Motzkina

Obraz bryły wielościennej Δ przy rzutowaniu $p : \mathbb{R}^n \to \mathbb{R}^{n-1}$, danym wzorem $p(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1})$, jest bryłą wielościenną.

Dowód. Załóżmy, że bryła wielościenna Δ zadana jest układem m nierówności postaci

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \leqslant b_i. \qquad (\dagger)$$

Popatrzmy na współczynniki w n-tej kolumnie macierzy $A = [a_{ij}]$ i rozważmy trzy zbiory postaci

$$P = \{i : a_{in} > 0\}, \quad N = \{j : a_{jn} < 0\}, \quad Z = \{k : a_{kn} = 0\}.$$

Dla indeksów i, j należących do $P \cup N$ przyjmujemy, po podzieleniu odpowiednich nierówności przez a_{in}:

$$x_n \leqslant a'_{i1}x_1 + \ldots + a'_{i,n-1}x_{n-1} + b'_i, \text{ dla } i \in P$$
$$a'_{j_1}x_1 + \ldots + a'_{j,n-1}x_{n-1} + b'_j \leqslant x_n, \text{ dla } j \in N.$$

W konsekwencji, warunek $(x_1, \ldots, x_{n-1}) \in p(\Delta)$ można zapisać za pomocą układu nierówności[2]

$$a'_{j_1}x_1 + \ldots + a'_{j,n-1}x_{n-1} + b'_j \leqslant x_n \leqslant a_{i1'}x_1 + \ldots + a'_{i,n-1}x_{n-1} + b'_i,$$

gdzie $i \in P$ oraz $j \in N$. Oznacza to, że zbiór $p(\Delta)$ jest zadany za pomocą $|P| \cdot |N|$ nierówności postaci:

$$(a'_{j1} - a'_{i1})x_1 + \ldots + (a'_{j,n-1} - a'_{i,n-1})x_{n-1} \leqslant (b'_i - b'_j)$$

oraz nierówności typu $(\dagger)$, dla których $a_{in} = 0$ (które nie zawierają x_n), mających indeksy spoza zbioru $P \cup N$. Stąd $p(\Delta)$ jest bryłą wielościenną. □

Udowodnione twierdzenie jest podstawą tzw. algorytmu Fouriera-Motzkina rozwiązywania układów nierówności liniowych. Możemy rozważyć układ m nierówności liniowych od n zmiennych $x_1, \ldots, x_n$, przy czym zmienna x_n ma być wyeliminowana. Wyjściowy układ nierówności ma rozwiązanie $(s_1, \ldots, s_{n-1}, s_n)$, które jest punktem bryły wielościennej Δ wtedy i tylko wtedy, gdy po redukcji punkt $(s_1, \ldots, s_{n-1})$ jest rozwiązaniem układu opisującego $p(\Delta)$. W ten sposób eliminujemy kolejne zmienne. Na końcu stwierdzamy, że nierówności typu $a \leqslant x_1$ oraz $x_1 \leqslant b$ mają rozwiązanie wtedy i tylko wtedy, gdy $a \leqslant b$.

Rozważmy układ nierówności od zmiennych x_1, x_2, x_3 postaci:

$$\begin{cases} -5x_1 + 4x_2 + 2x_3 & \leqslant 12 \\ -4x_1 + 3x_2 + 3x_3 & \leqslant 9 \\ 5x_1 - 2x_2 - x_3 & \leqslant -7 \\ 2x_1 + 6x_2 - 3x_3 & \leqslant 12 \end{cases}$$

Aby wyeliminować zmienną x_3 zapisujemy nierówności w równoważnej postaci:

$$\begin{cases} x_3 & \leqslant \frac{12+5x_1-4x_2}{2} \\ x_3 & \leqslant \frac{9+4x_1-3x_2}{3} \\ x_3 & \geqslant 7 + 5x_1 - 2x_2 \\ x_3 & \geqslant \frac{-12+2x_1+6x_2}{3}. \end{cases}$$

Mamy zatem dwie nierówności $\leqslant$ oraz dwie $\geqslant$. Zatem rozważany układ ma rozwiązanie jeśli prawa strona każdej nierówności $\leqslant$ jest równa co najmniej prawej stronie każdej nierówności $\geqslant$. To prowadzi do układu:

$$\begin{cases} 7 + x_1 - 2x_2 & \leqslant \frac{12+5x_1-4x_2}{2} \\ 7 + 5x_1 - 2x_2 & \leqslant \frac{9+4x_1-3x_2}{3} \\ \frac{-12+2x_1+6x_2}{3} & \leqslant \frac{12+5x_1-4x_2}{2} \\ \frac{-12+2x_1+6x_2}{3} & \leqslant \frac{9+4x_1-3x_2}{3}. \end{cases}$$

Czytelnika zainteresowanego większą liczbą przykładów i związkami algorytmu Fouriera z klasyczną eliminacją Gaussa, odsyłam do świetnego elementarnego tekstu Nielsa Lauritzena w „Undergraduate Convexity": `https://www.worldscientific.com/doi/suppl/10.1142/8527/suppl_file/8527_chap01.pdf`.

Zastosujmy powyższe twierdzenie dla stożków.

[2] Jeśli $P = \emptyset$ lub $N = \emptyset$, to $p(\Delta)$ opisany jest nierównościami $(\dagger)$, dla $i \in Z$. Jeśli również Z jest zbiorem pustym, to $p(\Delta)$ jest przestrzenią afiniczną wymiaru $n - 1$. Np. dla półprzestrzeni $\Delta = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \leqslant 0\}$ zbiór $p(\Delta)$ jest prostą.

Obserwacja 10.2

W przestrzeni afinicznej $\mathbb{R}^{n+1}$ złożonej z punktów postaci $(x_0, \ldots, x_n)$ rozważmy przestrzeń afiniczną H zadaną równaniem $x_0 = 1$. Dla dowolnego $X \subseteq H$ zachodzi

$$\operatorname{conv}(X) = \operatorname{cone}(X) \cap H.$$

Dowód. Inkluzja $\operatorname{conv}(X) \subseteq \operatorname{cone}(X) \cap H$ jest jasna, bo $\operatorname{cone}(X)$ oraz H są wypukłe. Załóżmy, że dla $i = 1, \ldots, r$ mamy punkty $p_i = (1, b_{i1}, \ldots, b_{in}) \in \mathbb{R}^{n+1}$. Wówczas wektor $a_1p_1 + \ldots + a_rp_r$ ma na zerowej współrzędnej sumę $a_1 + \ldots + a_r$. Stąd kombinacjom liniowym punktów p_i w $\mathbb{R}^{n+1}$ odpowiadają kombinacje afiniczne p_i w H, a nieujemnym kombinacjom w $\mathbb{R}^{n+1}$ – kombinacje wypukłe w H. □

Twierdzenie 10.3: Skończenie generowane stożki są bryłami wielościennymi

Dla dowolnego układu r wektorów $v_i = (b_{1i}, \ldots, b_{ni})$ w przestrzeni $\mathbb{R}^n$, gdzie $1 \leqslant i \leqslant r$, stożek $\sigma = \operatorname{cone}(v_1, \ldots, v_r)$ jest bryłą wielościenną. Co więcej, stożek ten jest opisany układem skończenie wielu nierówności typu

$$a_{i1}x_1 + \ldots + a_{in}x_n \leqslant 0.$$

Dowód. Naszym celem będzie wykazanie, że stożek σ jest obrazem pewnej bryły wielościennej przy rzutowaniu $p : \mathbb{R}^{n+r} \to \mathbb{R}^n$ (a w istocie przy złożeniu r rzutowań).

Niech $B \in M_{n\times r}(\mathbb{R})$ będzie macierzą, której i-ta kolumna zawiera współrzędne v_i w bazie standardowej, czyli b_{ki}. W przestrzeni $\mathbb{R}^{n+r}$ ze współrzędnymi $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_r)$ rozpatrzmy bryłę wielościenną Δ zadaną układem nierówności, który można zapisać skrótowo w postaci

$$(x, y) \in \Delta \iff x = By \text{ oraz } y \geqslant 0$$

czyli po zapisaniu we współrzędnych $(x_1, \ldots, x_n, y_1, \ldots, y_r)$ chodzi o układ nierówności

$$\begin{cases} x_i - b_{i1}y_1 - b_{i2}y_2 - \ldots - b_{ir}y_r & \leqslant 0, \text{ dla } 1 \leqslant i \leqslant n, \\ b_{j1}y_1 + b_{j2}y_2 + \ldots + b_{jr}y_r - x_j & \leqslant 0, \text{ dla } 1 \leqslant j \leqslant n, \\ -y_k & \leqslant 0, \text{ dla } 1 \leqslant k \leqslant r. \end{cases}$$

Wówczas rzutowanie $p : \mathbb{R}^{n+r} \to \mathbb{R}^n$ określone wzorem $p(x, y) = x$ oczywiście przeprowadza Δ na σ, czyli

$$p(\Delta) = \operatorname{cone}(v_1, \ldots, v_r) = \sigma.$$

Z drugiej strony z twierdzenia Fouriera-Motzkina wiemy, że σ jest bryłą wielościenną, czyli spełnia pewien układ nierówności postaci

$$a_{i1}x_1 + \ldots + a_ix_n \leqslant b.$$

Pozostaje zatem wykazać, ze $b_i = 0$. Wiemy jednak, że $0 \in \operatorname{cone}(v_1, \ldots, v_r)$, więc wszystkie b_i są nieujemne. Z drugiej strony, gdyby $b_i > 0$, dla pewnego i, wówczas mielibyśmy wektor $v = (x_1, \ldots, x_n) \in \sigma$, że

$$0 < a_{i1}x_1 + \ldots + a_{in}x_n \leqslant b_i.$$

Skoro jednak dla każdego $c \in \mathbb{R}$ mamy $cv \in \sigma$, to dla $c \gg 0$ dostalibyśmy sprzeczność. □

Dowód twierdzenia wyżej pozwala udowodnić jedną z implikacji w twierdzeniu strukturalnym.

Wniosek 10.1

Dla układu punktów $X = \{p_0, \ldots, p_r\}$ w przestrzeni $\mathbb{R}^n$ zbiór $\operatorname{conv}(X)$ jest wielościanem wypukłym.

Dowód. Rozważmy układ $Y \subset \mathbb{R}^{n+1}$ złożony z punktów postaci $y_i = (1, p_i)$, gdzie $1 \leqslant i \leqslant r$. Wówczas $\operatorname{cone}(Y)$ jest, jako stożek skończenie generowany, bryłą wielościenną. Stąd $\operatorname{conv}(Y) = \operatorname{cone}(Y) \cap H$ jest wielościanem, gdzie $H : \{(1, x), x \in \mathbb{R}^n\}$. Po zrzutowaniu $p : \mathbb{R}^{n+1} \to \mathbb{R}^n$ postaci $p((x_0, \ldots, x_n)) \mapsto (x_1, \ldots, x_n)$ mamy zatem $p(\operatorname{conv}(Y)) = \operatorname{conv}(X)$. Z twierdzenia Fouriera-Motzkina wynika więc, że $\operatorname{conv}(X)$ jest bryłą wielościenną. Jest jasne, że $\operatorname{conv}(X)$ jest zbiorem ograniczonym, co kończy dowód. □

Uwaga. Warto uzasadnić samodzielnie wniosek opisujący dokładnie nierówności spełniane przez dowolny sympleks n-wymiarowy (mający oczywiście $n+1$ ścian wymiaru $n-1$).

Wniosek 10.2

Niech $p_0, \ldots, p_n$ będzie afinicznie niezależnym układem punktów w $\mathbb{R}^n$, takim że

$$\det \begin{bmatrix} p_0 & 1 \\ p_1 & 1 \\ \vdots & \vdots \\ p_n & 1 \end{bmatrix} > 0.$$

Wówczas $W = \operatorname{conv}(p_0, p_1, \ldots, p_n)$ jest wielościanem złożonym z punktów $x = (x_1, \ldots, x_n)$ spełniających układ $n+1$ nierówności:

$$\det \begin{bmatrix} x & 1 \\ p_1 & 1 \\ \vdots & \vdots \\ p_n & 1 \end{bmatrix} \geqslant 0, \quad \det \begin{bmatrix} x & 1 \\ p_1 & 1 \\ \vdots & \vdots \\ p_n & 1 \end{bmatrix} \geqslant 0, \quad \det \begin{bmatrix} p_0 & 1 \\ x & 1 \\ \vdots & \vdots \\ p_n & 1 \end{bmatrix} \geqslant 0, \quad \det \begin{bmatrix} p_0 & 1 \\ p_1 & 1 \\ \vdots & \vdots \\ x & 1 \end{bmatrix} \geqslant 0.$$

Co więcej, dla każdego j punkt p_j spełnia wszystkie nierówności wyżej, za wyjątkiem j-tej, czyli jest wierzchołkiem wielościanu.

Jeśli Państwo myślą, że nic ciekawego nie wynika z takiego opisu, warto przypomnieć sobie dodatek 5.4 ze skryptu z GAL I. Opisywał on – choć wtedy nie można było tego powiedzieć – jak wyglada wielościan rozpięty na wszystkich macierzach permutacyjnych rozmiaru n. Okazuje się, że jest to zbiór macierzy podwójnie stochastycznych, czyli takich, gdzie wiersze i kolumny mają sumy wyrazów równe 1, co udowodnił w 1946 roku Birkhoff (i do czego Państwa zachęcam). Fakt ten ma wiele zastosowań.

Aby uzasadnić, że każdy wielościan wypukły jest uwypukleniem skończonego zbioru punktów, będziemy potrzebowali uzasadnić wcześniej fakt mówiący, że bryła wielościenna pewnej szczególnej postaci jest stożkiem. W tym celu wprowadzimy podstawowe narzędzia teorii dualności.

Wcześniej jednak wzmocnijmy nieco geometryczne intuicje stojące za wykazywanymi rezultatami, tak by wprowadzenie do naszych rozumowań funkcjonałów wydało się naturalne. Intuicje te związane są bezpośrednio związane z faktem, że każdy funkcjonał odpowiada geometrycznie pewnej podprzestrzeni kowymiaru 1.

Oto *podstawowy rezultat teorii nierówności liniowych.*

Twierdzenie 10.4: Lemat Farkasa

Niech $v_1, \ldots, v_m, b$ będą wektorami w przestrzeni $\mathbb{R}^n$. Wówczas zachodzi alternatywa wyłączająca:

(1) albo $b \in \operatorname{cone}(v_1, \ldots, v_m)$,

(2) albo istnieje funkcjonał $\lambda \in (\mathbb{R}^n)^*$, taki że $\lambda(b) > 0$ oraz $\lambda(v_1) \leqslant 0, \ldots, \lambda(v_m) \leqslant 0$.

Punkt (2) można sformułować geometrycznie, mianowicie: istnieje podprzestrzeń afiniczna H wymiaru $n-1$ (hiperpowierzchnia), która „rozdziela" stożek $\operatorname{cone}(v_1, \ldots, v_m)$ oraz prostą $\operatorname{lin}(b)$. Co to znaczy? Jeśli H opisana jest równaniem $a_1x_1+a_2x_2+\ldots+a_nx_n = 0$, to dla wektorów $(s_1, s_2, \ldots, s_n) \in \operatorname{cone}(v_1, \ldots, v_m)$ mamy $a_1x_1 + \ldots + a_nx_n \leqslant 0$, zaś dla wektora $b = (b_1, \ldots, b_n)$ mamy $a_1b_1 + a_2b_2 + \ldots + a_nb_n > 0$.

Prostą ilustrację daje choćby układ wektorów $v_1 = (1, 0)$, $v_2 = (1, 1)$ oraz $b = (0, 1)$. Nie ma wątpliwości, że $b \notin \operatorname{cone}(v_1, v_2)$. Lemat powyżej mówi, że istnieje prosta rozdzielająca $\operatorname{cone}(\beta)$ od $\operatorname{cone}(v_1, v_2)$. Wystarczy wziąć prostą $y = \operatorname{lin}(-1, 1)$. Odpowiada jej funkcjonał $\lambda \in (\mathbb{R}^2)^*$ dany wzorem $\lambda(x_1, x_2) = -x_1 + x_2$. Oczywiście $\lambda(v_1) = -1, \lambda(v_2) = 0$, zaś $\lambda(b) = 1$.

Aby udowodnić powyższy lemat wykorzystamy jeszcze jedną ważną konwencję. Niech $v_i = (a_{i1}, \ldots, a_{in})$ i niech $A = [a_{ij}] \in M_{n\times m}(\mathbb{R})$. Punkt pierwszy lematu mówi, że istnieje $x \in \mathbb{R}^m$, $x \geqslant 0$ (nieujemne współrzędne), że

$$Ax = b.$$

Drugi punkt można natomiast przepisać w postaci: istnieje $y = (a_1, \ldots, a_n) \in \mathbb{R}^n$, że

$$A^T y \leqslant 0 \quad \text{oraz} \quad b^T y > 0.$$

Oczywiście y ma być taki, że $\lambda(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n$. Widać po przemnożeniu macierzy, że dostajemy warunki równoważne opisanym w lemacie. Można powyższe warunki interpretować także w języku funkcjonałów. Rozważmy przekształcenie liniowe $\phi : \mathbb{R}^m \to \mathbb{R}^n$ dane macierzą A. Macierz A^T jest macierzą przekształcenia sprzężonego $\phi^* : (\mathbb{R}^n)^* \to (\mathbb{R}^m)^*$. Warunek $A^T y \leqslant 0$ oznacza, dokładnie tyle, że obraz każdego z v_i jest przy λ niedodatni, dla $1 \leqslant i \leqslant m$. Zauważmy jeszcze, że używając języka macierzowego można przeformułować Obserwację 10.3 w następujący sposób: dla dowolnego stożka C w przestrzeni $\mathbb{R}^n$ generowanego przez r wektorów istnieje macierz $A \in M_{m\times n}(\mathbb{R})$, że

$$C = \{v \in \mathbb{R}^n \mid Av \leqslant 0\}.$$

Dowód Lematu Farkasa. Jeśli zachodzi pierwszy warunek, to

$$b^T y = (Ax)^T y = (x^T A^T)y = x^T(A^T y) \leqslant 0,$$

jeśli $y^T A \leqslant 0$ oraz $x \geqslant 0$. Ostatnia nierówność przeczy warunkowi $y^T b > 0$. Trudniej wykazać, że drugi warunek zachodzi, gdy nie zachodzi pierwszy.

Niech C będzie stożkiem generowanym przez kolumny macierzy A. Na mocy Obserwacji 10.3 istnieje macierz Y, taka że $C = \{v \in \mathbb{R}^n \mid Yv \leqslant 0.\}$. Jeśli pierwszy warunek nie zachodzi, to $b \notin C$. To oznacza, że $Yb \not\leqslant 0$, czyli istnieje wektor wierszowy y w macierzy Y, że $y^T b > 0$. Jeśli v_i jest wektorem kolumnowym w A, to z definicji C mamy jednak $y^T v_i \leqslant 0$. Stąd $y^T A \leqslant 0$ i dowód jest zakończony.

Definicja 10.6: Stożek dualny

Dla dowolnego zbioru $X \subseteq \mathbb{R}^n$ definiujemy tzw. stożek dualny do $\text{cone}(X)$ postaci:

$$\text{cone}^*(X) = \{\lambda \in (\mathbb{R}^n)^* \mid \forall_{v\in X}\, \lambda(v) \geqslant 0\}.$$

Podobnie dla $Y \subseteq (\mathbb{R}^n)^*$ definiujemy

$$\text{cone}^*(Y) = \{v \in \mathbb{R}^n \mid \forall_{\lambda\in Y}\, \lambda(v) \geqslant 0\}.$$

Czasami definiujemy stożek dualny do stożka $\text{cone}(X)$ jako podzbiór $\mathbb{R}^n$ złożony z takich wektorów $(a_1, \ldots, a_n)$, że dla każdego wektora $(x_1, \ldots, x_n) \in \text{cone}(X)$ mamy $a_1x_1 + \ldots + a_nx_n \geqslant 0$.

Zachęcam Czytelnika do prostego dowodu następujących wartości stożków.

Obserwacja 10.3: Własności stożków dualnych

Niech X będzie podzbiorem $\mathbb{R}^n$.

(i) Jeśli $X \subseteq Y \subseteq \mathbb{R}^n$, to $\text{cone}^*(X) \supseteq \text{cone}^*(Y)$.

(ii) $\text{cone}^*(\text{conv}(X)) = \text{cone}^*(X)$.

(iii) $\text{cone}^*(\text{cone}(X)) = \text{cone}^*(X)$.

(iv) $\text{cone}^*(\text{cone}^*(X)) = \text{cone}(X)$.

Przechodzimy do naszkicowania twierdzenie o dualności. Dla uproszczenia notacji przyjmijmy, że jeśli $\sigma = \text{cone}(X)$ w przestrzeni V lub V^*, to $\text{cone}^*(\sigma)$ oznaczamy przez σ^*. Wówczas kluczowy czwarty punkt powyższej obserwacji sprowadza się do stwierdzenia, że $(\sigma^*)^* = \sigma$.

Twierdzenie 10.5: Dualność stożków – Farkas, Minkowski, Weyl

Następujące warunki są równoważne dla podzbiorów $\sigma \subseteq \mathbb{R}^n$ oraz $\sigma^* \subseteq (\mathbb{R}^n)^*$.

(a) σ jest skończenie generowany, czyli $\sigma = \mathrm{cone}(v_1, \ldots, v_r)$,

(a') σ^* jest skończenie generowany,

(b) σ jest bryłą wielościenną, czyli jest rozwiązaniem układu $Ax \leqslant 0$.

(b)' σ^* jest bryłą wielościenną.

Dowód. Implikacje (a) $\Rightarrow$ (b) oraz (a') $\Rightarrow$ (b') są już nam znane. Jeśli σ spełnia warunek (b), to wiemy już, że jest opisana układem r nierówności $a_{i1}x_1 + \ldots + a_{in}x_n \leqslant 0$. Biorąc więc funkcjonały $\lambda_i : \mathbb{R}^n \to \mathbb{R}$ dane wzorami $\lambda_i(x_1, \ldots, x_n) = a_{i1}x_1 + \ldots + a_{in}x_n$ bez trudu stwierdzamy, że $\sigma^* = \mathrm{cone}(-\lambda_1, \ldots, -\lambda_n)$, co daje (a'). Równoważności (a) oraz (a') wynikają natomiast wprost z uwag wyżej. □

Jesteśmy gotowi na wykazanie finalnej implikacji w twierdzeniu strukturalnym.

Jeśli mamy wielościan Δ zadany przez układ nierówności postaci $a_{i1}x_1 + \ldots + a_{in}x_n \leqslant b_i$, to bierzemy stożek $\sigma \in \mathbb{R}^{n+1}$ taki, aby $p(\sigma \cap H) = \Delta$, czyli zadany nierównościami $-b_ix_0 + a_{i1}x_1 + \ldots + a_{in}x_n \leqslant 0$. Na mocy twierdzenia o dualności stożków stwierdzamy, że σ jest stożkiem rozpiętym na pewnym układzie wektorów $v_1, \ldots, v_r$, gdzie $v_i = (c_{i0}, \ldots, c_{in})$. Rozpatrujemy trzy przypadki, w zależności od znaku c_{i0}.

Jeśli dla każdego i mamy $c_{i0} > 0$, to kładąc $p_i = \frac{1}{c_{i0}} v_i$ widzimy, że $p(\sigma) = \Delta = \mathrm{conv}(p_1, \ldots, p_r)$. Jeśli σ zawiera niezerowy wektor $v = (0, c_1, \ldots, c_n)$, to $\Delta \subset \Delta + tv$, dla $t > 0$, co jest sprzeczne z tym, że Δ jest ograniczony. Jeśli wreszcie $v = (c_0, \ldots, c_n) \in \sigma$, gdzie $c_0 < 0$, to łącząc punkty wielościanu $\Delta \subset H = \{x_0 = 1\}$ z punktem odpowiadającym wektorowi v dostajemy wektor z zerową współrzędną i wracamy do poprzedniego przypadku. Twierdzenie strukturalne jest udowodnione.

Zachęcam Czytelnika do docenienia uzyskanego twierdzenia strukturalnego poprzez zauważenie, że poniższe obserwacje są zwykle trywialne przy jednym opisie wielościanu, a zdecydowanie nie – przy drugim.

- Przecięcie bryły wielościennej z podprzestrzenią afiniczną jest bryłą wielościenną.
- Przecięcie bryły wielościennej z bryłą wielościenną jest bryłą wielościenną.
- Suma Minkowskiego dwóch brył wielościennych jest bryłą wielościenną.
- Każdy wielościan wypukły jest sumą Minkowskiego sympleksów.
- Każdy wielościan wypukły jest obrazem sympleksu przy pewnym rzucie afinicznym.

Można udowodnić ogólniejsze twierdzenie, związane z operacją sumy Minkowskiego.

Twierdzenie 10.6

Podzbiór W przestrzeni afinicznej $\mathbb{R}^n$ jest bryłą wielościenną wtedy i tylko wtedy, gdy istnieje przestawienie:

$$W = \mathrm{conv}(V) + \mathrm{cone}(Y),$$

gdzie V oraz Y są zbiorami skończonymi. Innymi słowy każda bryła wielościenna jest sumą wielościanu wypukłego i stożka skończenie generowanego.

* * *

Teoria wielościanów związana jest nie tylko z optymalizacją liniową czy teorią zbiorów wypukłych, ale także z osiągnięciami tzw. rzeczywistej geometrii algebraicznej i jej aktywnie uprawianymi dziedzinami. Czytelnika zainteresowanego tekstem popularyzującym jeden z możliwych kierunków dalszego pogłębienia tematu odsyłam do artykułu Davida A. Coxa: *Mirror Symmetry and Polar Duality of Polytopes*, pod adres: `https://www.mdpi.com/2073-8994/7/3/1633`. Zainteresowanych natomiast monografią prezentującą dogłębnie algebraiczną teorię wielościanów i ich kombinatorykę, odsyłam do klasycznego już podręcznika Zieglera *Lectures on Polytopes*; GTM 152, Springer, 1995.

Rozdział 11

Formy dwuliniowe i iloczyny skalarne

Klasyfikacja wszystkich endomorfizmów (nad ciałem algebraicznie domkniętym) w języku postaci Jordana mimo niewątpliwego uroku ma też swoje wady – wiemy wprawdzie coś o geometrii każdego endomorfizmu z osobna, ale nie rozumiemy zbyt dobrze „struktury algebraicznej" stojącej za tym opisem geometrycznym. Mówiąc prościej: mając dwa endomorfizmy ϕ, ψ przestrzeni V, na przykład nad ciałem $\mathbb{C}$, możemy wyznaczyć bazy (i postaci) Jordana $\mathcal{J}_1$, $\mathcal{J}_2$ każdego z tych endomorfizmów. Uzyskamy informację geometryczną o każdym z nich. Nie wiemy natomiast czym jest z geometrycznego punktu widzenia $\phi \circ \psi$. Jego macierzą nie musi być przecież iloczyn macierzy w postaci Jordana $M(\psi)^{\mathcal{J}_2}_{\mathcal{J}_2} \cdot M(\phi)^{\mathcal{J}_1}_{\mathcal{J}_1}$, który nie musi mieć nawet postaci Jordana. Nie znamy relacji baz Jordana dla złożeń innych niż wielomian od endomorfizmu. Takie podejście niestety nie zadziała. Określimy jednak nowy rodzaj baz, wymagający rozważania dodatkowej struktury na przestrzeni liniowej, który umożliwi to w istotnych sytuacjach. Struktura ta pozwoli między innymi na mówienie o prostopadłości wektorów, i to nie tylko w kontekście geometrycznym.

Definicja 11.1: Forma dwuliniowa

Niech V będzie przestrzenią liniową nad ciałem K. Funkcję:

$$h : V \times V \to K$$

nazywamy FORMĄ DWULINIOWĄ (albo FUNKCJONAŁEM DWULINIOWYM) na przestrzeni V, jeśli dla każdych $\alpha, \beta, \gamma, \delta \in V$ i $a, b, c, d \in K$ zachodzi:

(1) $h(a \cdot \alpha + b \cdot \beta, \gamma) = a \cdot h(\alpha, \gamma) + b \cdot h(\beta, \gamma)$ liniowość względem pierwszej zmiennej,
(2) $h(\alpha, c \cdot \gamma + d \cdot \delta) = c \cdot h(\alpha, \gamma) + d \cdot h(\alpha, \delta)$ liniowość względem drugiej zmiennej,

Zobaczmy kilka przykładów.

- Dla $V = \mathbb{R}^2$ funkcja $h : V \times V \to \mathbb{R}$ dana wzorem:

  $$h((x_1, x_2), (y_1, y_2)) = x_1 y_1 - x_2 y_2$$

  jest formą dwuliniową. Mamy bowiem:

  $$\begin{aligned} h(a \cdot (x_1, x_2) + b \cdot (x_1', x_2'), (y_1, y_2)) &= h((ax_1 + bx_1', ax_2 + bx_2'), (y_1, y_2)) \\ &= (ax_1 + bx_1')y_1 - (ax_2 + bx_2')y_2 = a(x_1 y_1 - x_2 y_2) + b(x_1' y_1 - x_2' y_2) \\ &= a \cdot h((x_1, x_2), (y_1, y_2)) + b \cdot h((x_1', x_2'), (y_1, y_2)). \end{aligned}$$

- Dla $V = K^n$ oraz $A = [a_{ij}] \in M_n(K)$ dwuliniowa jest forma $h_A : V \times V \to K$ zadana wzorem

  $$h_A((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{1 \leqslant i,j \leqslant n} a_{ij} x_i y_j \quad (\dagger)$$

  Wynika to z rozdzielności mnożenia względem dodawania dla macierzy $V_1, V_2 \in M_{1 \times n}(K)$ oraz $W_1, W_2 \in M_{n \times 1}(K)$:

  $$(V_1 + V_2)AW = V_1 AW + V_2 AW \quad \text{oraz} \quad VA(W_1 + W_2) = VAW_1 + VAW_2.$$

- Dla dowolnej pary wektorów $(a_1, a_2), (b_1, b_2) \in K^2$ określamy:

$$h((a_1, a_2), (b_1, b_2)) = \det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}.$$

  Rzeczywiście, dla dowolnych $a_1, a_2, b_1, b_2, a_1', a_2', b_1', b_2' \in K$ mamy:

$$\det \begin{bmatrix} a_1 + a_1' & b_1 \\ a_2 + a_2' & b_2 \end{bmatrix} = \det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} + \det \begin{bmatrix} a_1' & b_1 \\ a_2' & b_2 \end{bmatrix},$$
$$\det \begin{bmatrix} a_1 & b_1 + b_1' \\ a_2 & b_2 + b_2' \end{bmatrix} = \det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} + \det \begin{bmatrix} a_1 & b_1' \\ a_2 & b_2' \end{bmatrix}.$$

- Dla $V = M_n(K)$ oraz dowolnych $A, B \in V$ określamy:

$$h((A, B)) = \mathrm{tr}(AB^T).$$

- Dla $V = F_c[0, 1]$ – funkcji „całkowalnych" z $[0, 1]$ do $\mathbb{R}$, określamy:

$$h(f, g) = \int_0^1 f(x)g(x)dx.$$

- Dla przestrzeni $V = (P(X), \triangle, \emptyset)$ nad $\mathbb{Z}_2$ oraz $A, B \in V$ określamy:

$$h(A, B) = |A \cap B| \pmod 2.$$

Odnotujmy kilka własności dowolnej formy dwuliniowej h na przestrzeni V nad ciałem K.

- $h(a \cdot \alpha, b \cdot \beta) = ab \cdot h(\alpha, \beta)$, dla dowolnych $a, b \in K$ oraz $\alpha, \beta \in V$,
- $h(\alpha, 0) = h(\alpha, \alpha - \alpha) = h(\alpha, \alpha) - h(\alpha, \alpha) = 0$, dla dowolnego $\alpha \in V$,
- $h(\alpha + \beta, \alpha + \beta) = h(\alpha, \alpha) + h(\alpha, \beta) + h(\beta, \alpha) + h(\beta, \beta)$, dla dowolnych $\alpha, \beta \in V$.

Obserwacja 11.1

Każda forma dwuliniowa h na przestrzeni K^n zadana jest wzorem:

$$h((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sum_{1 \leqslant i,j \leqslant n} a_{ij} x_i y_j,$$

dla pewnych $a_{ij} \in K$.

Dowód. Niech h będzie formą dwuliniową na K^n oraz niech $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ będzie bazą standardową. Dla każdych $i, j = 1, 2, \ldots, n$ niech też $a_{ij} = h(\epsilon_i, \epsilon_j) \in K$. Wówczas mamy:

$$\begin{aligned} &h((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = h(x_1\epsilon_1 + x_2\epsilon_2 + \ldots + x_n\epsilon_n, y_1\epsilon_1 + y_2\epsilon_2 + \ldots + y_n\epsilon_n) = \\ &\overset{(1)}{=} x_1 \cdot h(\epsilon_1, y_1\epsilon_1 + y_2\epsilon_2 + \ldots + y_n\epsilon_n) + \quad \ldots \quad + x_n \cdot h(\epsilon_n, y_1\epsilon_1 + y_2\epsilon_2 + \ldots + y_n\epsilon_n) = \\ &\overset{(2)}{=} x_1y_1 \cdot h(\epsilon_1, \epsilon_1) + \ldots + x_1y_n \cdot h(\epsilon_1, \epsilon_n) + \quad \ldots \quad + x_ny_1 \cdot h(\epsilon_n, \epsilon_1) + \ldots + x_ny_n \cdot h(\epsilon_n, \epsilon_n) = \\ &= a_{11}x_1y_1 + a_{12}x_1y_2 + \ldots + a_{1n}x_1y_n + \quad \ldots \quad + a_{n1}x_ny_1 + a_{n2}x_ny_2 + \ldots + a_{nn}x_ny_n. \end{aligned}$$

Wynika stąd, że h zadana jest wzorem (†), czyli jest dwuliniowa zgodnie z wcześniejszym rachunkiem. □

W kontekście formy dwuliniowych szczególnie istotne będzie rozważanie przypadku nad ciałem liczb rzeczywistych. Dla ilustracji rozmaitych zagadnień potrzebujemy odróżniać skalary dodatnie od ujemnych. Równolegle przyglądać się będziemy strukturom dwuliniowym definiowanym nad innymi ciałami.

Definicja 11.2: Iloczyn skalarny

Niech V będzie przestrzenią liniową nad ciałem $\mathbb{R}$. Formę dwuliniową

$$\langle\, , \rangle : V \times V \to \mathbb{R}$$

nazywamy ILOCZYNEM SKALARNYM NA PRZESTRZENI V, jeśli

- $\langle\, , \rangle$ jest SYMETRYCZNA, to znaczy dla dowolnych $\alpha, \beta \in V$ mamy $\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$,
- $\langle\, , \rangle$ jest DODATNIO OKREŚLONA, to znaczy dla niezerowego $\alpha \in V$ mamy $\langle \alpha, \alpha \rangle > 0$.

Możliwych iloczynów skalarnych na danej przestrzeni jest oczywiście wiele. Zobaczmy kilka przykładów na przestrzeni $V = \mathbb{R}^n$. Zacznijmy od fundamentalnego.

- Definiujemy $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ wzorem:

  $$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n$$

  zadaje iloczyn skalarny na przestrzeni $\mathbb{R}^n$, który nazywać będziemy STANDARDOWYM ILOCZYNEM SKALARNYM i oznaczamy jako $\langle\, , \rangle_{st}$. Dzięki (standardowemu) iloczynowi skalarnemu odczytamy w nowym języku wiele poznanych wcześniej konfiguracji. Jest to również podstawowy obiekt w elementarnej geometrii analitycznej, choć niestety już nie geometrii uczonej w szkole.

  Rozważmy układ równań o współczynnikach rzeczywistych postaci:

  $$\begin{cases} a_{11}x_1 + \ a_{12}x_2 + \ldots + \ a_{1n}x_n = b_1 \\ \qquad\qquad\qquad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases}$$

  Przyjmijmy, że $v = (x_1, \ldots, x_n)$ oraz $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{in})$ są elementami $\mathbb{R}^n$, dla $1 \leqslant i \leqslant m$. Wówczas powyższy układ równań liniowych zapisać możemy w postaci układu warunków:

  $$\langle \alpha_1, v \rangle_{st} = b_1, \quad \ldots \quad , \langle \alpha_n, v \rangle_{st} = b_n.$$

  Zauważmy, że jeśli $\epsilon_1, \ldots, \epsilon_n$ jest bazą standardową przestrzeni $\mathbb{R}^n$, to mamy:

  $$\langle \epsilon_i, \epsilon_j \rangle = \begin{cases} 1, & \text{dla } i = j, \\ 0, & \text{dla } i \neq j \end{cases}.$$

- Funkcja $\langle\, , \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ dana wzorem

  $$\langle (x_1, x_2, x_3), (y_1, y_2, y_3) \rangle = x_1y_1 + 2x_1y_2 + 2x_2y_1 + 9x_2y_2 + 4x_3y_3$$

  jest dwuliniową formą symetryczną, zgodnie z obserwacją wyżej. Warunek dodatniej określoności jest spełniony, bo wyrażenie $\langle (x_1, x_2, x_3), (x_1, x_2, x_3) \rangle$ zapisuje się jako:

  $$x_1^2 + 2x_1x_2 + 2x_2x_1 + 9x_2^2 + 4x_3^2 = (x_1 + 2x_2)^2 + 5x_2^2 + 4x_3^2.$$

- Funkcja $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ dana wzorem

  $$f((x_1, x_2, x_3), (y_1, y_2, y_3)) = x_1y_1 + 4x_1y_2 + 4x_2y_1 + 5x_2y_2$$

  jest dwuliniowa symetryczna, ale nie jest dodatnio określona, bowiem

  $$f((1, -1, 0), (1, -1, 0)) = 1 \cdot 1 + 4 \cdot 1 \cdot (-1) + 4 \cdot (-1) \cdot 1 + 5 \cdot (-1) \cdot (-1) = -2,$$

  więc nie jest ona iloczynem skalarnym na $\mathbb{R}^3$.

Iloczyny skalarne pełnią fundamentalną rolę w badaniu ważnych klas przestrzeni nieskończonego wymiaru. Ograniczmy się w tym miejscu do podania dwóch przykładów ze skryptu.

- W przestrzeni $\mathbb{R}^\infty$ rozpatrujemy podprzestrzeń l^2 złożoną ze wszystkich ciągów (x_i) takich, że

$$\sum_{i=1}^{\infty} x_i^2 < \infty.$$

  Wówczas (i to trzeba udowodnić) funkcja $\langle\, , \rangle : l^2 \times l^2 \to \mathbb{R}$ dana wzorem

$$\langle (x_i), (y_i) \rangle = \sum_{i=1}^{\infty} x_i y_i$$

  jest iloczynem skalarnym na l^2.

- Niech $C[a,b]$ oznacza przestrzeń wszystkich funkcji ciągłych określonych na przedziale $[a,b]$ o wartościach w $\mathbb{R}$. Funkcja $\langle\, , \rangle : C[a,b] \times C[a,b] \to \mathbb{R}$ dana wzorem

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

  jest iloczynem skalarnym na przestrzeni $C[a,b]$.

Widzimy, że w obydwu przypadkach wprowadzenie iloczynu skalarnego wymaga nietrywialnych definicji i faktów analitycznych (i nie tylko). Stąd tez rozważania na temat nieskończenie wymiarowych przestrzeni liniowych (m.in.) z iloczynem skalarnym stanowią przedmiot semestralnego wykładu na wyższych latach.

Definicja 11.3: Przestrzeń euklidesowa (liniowa)

Parę $(V, \langle\, , \rangle)$ gdzie V jest skończenie wymiarową przestrzenią liniową nad $\mathbb{R}$, zaś $\langle\, , \rangle$ jest iloczynem skalarnym na V nazywamy PRZESTRZENIĄ EUKLIDESOWĄ LINIOWĄ.

Przejdźmy do zasadniczych pojęć geometrycznych.

Definicja 11.4: Długość/norma wektora

Niech $\langle\, , \rangle$ będzie iloczynem skalarnym na przestrzeni V. DŁUGOŚCIĄ (albo NORMĄ) wektora $\alpha \in V$ nazywamy liczbę $\|a\| = \sqrt{\langle \alpha, \alpha \rangle}$.

Norma zależy od wyboru iloczynu skalarnego. Wektor $(3,4)$ ma:

- normę $\sqrt{3^2+4^2} = 5$, jeśli $\langle\, , \rangle$ to standardowy iloczyn skalarny,
- normę $\sqrt{3^2+(2\cdot 4)^2} = \sqrt{73}$, jeśli $\langle (x_1,x_2),(y_1,y_2) \rangle = x_1y_1 + 2x_2y_2$.

Nowo zdefiniowana funkcja „zachowuje" się jak znana ze szkoły długość, co wyrażają następujące fakty.

Obserwacja 11.2: Podstawowe własności normy

Niech $\langle\, , \rangle$ będzie iloczynem skalarnym na przestrzeni V. Wówczas dla każdych $\alpha, \beta \in V$ oraz $a \in \mathbb{R}$ zachodzi:

- $\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2 + 2\langle \alpha, \beta \rangle$,
- $\|a \cdot \alpha\| = |a| \cdot \|\alpha\|$,
- $\|\alpha\| = 0 \iff \alpha = 0$.

Twierdzenie 11.1: Uogólnienia nierówności szkolnych

Niech $\langle\,,\rangle$ będzie iloczynem skalarnym na przestrzeni V. Wówczas dla każdych $\alpha, \beta \in V$ zachodzi:

(1) nierówność Schwarza: $|\langle \alpha, \beta \rangle| \leqslant \|\alpha\| \cdot \|\beta\|$,

(2) nierówność trójkąta: $\|\alpha + \beta\| \leqslant \|\alpha\| + \|\beta\|$,

(3) twierdzenie Pitagorasa: $\langle \alpha, \beta \rangle = 0 \implies \|\alpha\|^2 + \|\beta\|^2 = \|\alpha + \beta\|^2$.

Dowód. Dowód (1). Jeśli $\alpha = 0$, to obie strony nierówności są zerami. W przypadku, gdy $\alpha \neq 0$ rozważamy funkcję $f : \mathbb{R} \to \mathbb{R}$ zadaną wzorem $f(t) = \|t\alpha + \beta\|^2$. Zatem

$$\begin{aligned} f(t) &= \langle t\alpha + \beta, t\alpha + \beta \rangle = \\ &= t\langle \alpha, t\alpha + \beta \rangle + \langle \beta, t\alpha + \beta \rangle = \\ &= t^2\langle \alpha, \alpha \rangle + t\langle \alpha, \beta \rangle + t\langle \beta, \alpha \rangle + \langle \beta, \beta \rangle = \\ &= t^2\|\alpha\|^2 + 2t\langle \alpha, \beta \rangle + \|\beta\|^2. \end{aligned}$$

Na mocy dodatniej określoności iloczynu skalarnego dla każdego $t \in \mathbb{R}$ mamy $f(t) \geqslant 0$, a zatem wyróżnik trójmianu kwadratowego

$$t^2\|\alpha\|^2 + 2\langle \alpha, \beta \rangle t + \|\beta\|^2$$

musi być niedodatni, tj. $\Delta \leqslant 0$, gdzie

$$\Delta = (2\langle \alpha, \beta \rangle)^2 - 4\|\alpha\|^2\|\beta\|^2 = 4\langle \alpha, \beta \rangle^2 - 4\|\alpha\|^2\|\beta\|^2.$$

Stąd wynika (1). Do dowodu (2) i (3) wykorzystujemy przydatną w rachunkach tożsamość, wypisaną wcześniej w Obserwacji 10.2 $\|\alpha + \beta\|^2 = \|\alpha\|^2 + 2\langle \alpha, \beta \rangle + \|\beta\|^2$, którą wyprowadza się jak powyżej. Wówczas na mocy (1) mamy:

$$\|\alpha + \beta\|^2 \leqslant \|\alpha\|^2 + 2\|\alpha\| \cdot \|\beta\| + \|\beta\|^2 = (\|\alpha\| + \|\beta\|)^2.$$

□

Gorąco zachęcam Czytelnika do przeanalizowania kiedy w nierównościach wyżej zachodzą równości. Wymaga to chwili namysłu i jest istotnym ćwiczeniem na analizę dowodu wyżej.

Z nierówności Schwarza wynika natychmiast, że dla każdej pary niezerowych wektorów $\alpha, \beta \in V$ mamy:

$$-1 \leqslant \frac{\langle \alpha, \beta \rangle}{\|\alpha\| \cdot \|\beta\|} \leqslant 1.$$

Dla każdej liczby rzeczywistej $r \in [-1, 1]$ istnieje dokładnie jedna liczba $\theta \in [0, \pi]$, że $\cos\theta = r$.

Definicja 11.5: Kąt między wektorami

Niech $\langle\,,\rangle$ będzie iloczynem skalarnym na przestrzeni V. Niech α, β będą niezerowymi wektorami przestrzeni V. Liczbę $\theta \in [0, \pi]$ taką, że

$$\frac{\langle \alpha, \beta \rangle}{\|\alpha\| \cdot \|\beta\|} = \cos\theta$$

nazywamy (niezorientowanym) KĄTEM MIĘDZY WEKTORAMI α i β.

Przykład. Wyznacz kosinus kąta θ pomiędzy wektorami $\alpha = (2, 1, -4)$ oraz $\beta = (1, -1, 3)$ w przestrzeni $(\mathbb{R}^3, \langle,\rangle)$, gdzie $\langle,\rangle$ jest iloczynem skalarnym postaci $h((x_1, x_2, x_3), (y_1, y_2, y_3)) = x_1y_1 + 2x_2y_2 + 3x_3y_3$.

ROZWIĄZANIE. Szukany kosinus równy jest $-\frac{2}{\sqrt{5}}$. Wyznaczamy go z zależności $\|\alpha\| \cdot \|\beta\| \cdot \cos\theta = \langle \alpha, \beta \rangle$. Mamy $\|\alpha\| = \sqrt{1 \cdot 2^2 + 2 \cdot 1^2 + 3 \cdot (-4)^2} = \sqrt{54}$ oraz $\|\beta\| = \sqrt{1 \cdot 1^2 + 2 \cdot (-1)^2 + 3 \cdot 3^2} = \sqrt{30}$, a także $\langle \alpha, \beta \rangle = 2 \cdot 1 + 2 \cdot 1 \cdot (-1) + 3 \cdot (-4) \cdot 3 = -36$. Stąd

$$\cos\theta = \frac{-36}{\sqrt{30} \cdot \sqrt{54}} = -\frac{2}{\sqrt{5}}.$$

Badanie form dwuliniowych na przestrzeniach liniowych jest ciekawe, ponieważ wśród układów liniowo niezależnych szczególne miejsce mają układy wektorów, które nazwiemy prostopadłymi lub ortogonalnymi. Bazy złożone z takich układów są kluczem do całej dalszej teorii. Samo pojęcie prostopadłości jest centralnym motywem całej matematyki, który odkrywać będziecie Państwo w rozmaitych kontekstach.

Definicja 11.6: Wektory prostopadłe, dopełienie ortogonalne podzbioru

Niech h będzie formą dwuliniową na przestrzeni liniowej V.

1. Mówimy, że wektory $\alpha, \beta \in V$ są PROSTOPADŁE, ozn. $\alpha \perp \beta$, jeśli
$$h(\alpha, \beta) = 0.$$
2. Dla podzbioru $X \subseteq V$ definiujemy jego DOPEŁNIENIE ORTOGONALNE
$$X^{\perp} = \{\alpha \in V : \alpha \perp \beta, \text{ dla każdego } \beta \in X\}.$$

Powyższa definicja obejmuje oczywiście także iloczyny skalarne, określone na przestrzeniach rzeczywistych. Różnym formom dwuliniowym na tej samej przestrzeni liniowej V odpowiadają różne pojęcia prostopadłości. Wektory bazy standardowej
$$(1,0,0), (0,1,0)$$
są prostopadłe przy standardowym iloczynie skalarnym na przestrzeni $\mathbb{R}^3$, ale $\langle (1,0,0), (0,1,0) \rangle = 2$, jeśli iloczyn skalarnym $\langle \, , \rangle$ jest określony na tej samej przestrzeni $\mathbb{R}^3$ wzorem
$$\langle (x_1,x_2,x_3), (y_1,y_2,y_3) \rangle = x_1y_1 + 2x_1y_2 + 2x_2y_1 + 9x_2y_2 + 4x_3y_3.$$

Obserwacja 11.3

Dla każdego podzbioru $X \subseteq V$ zbiór $X^{\perp}$ jest podprzestrzenią V.

Dowód. Dla dowolnych $a, b \in \mathbb{R}$, $\alpha, \beta \in X^{\perp}$ oraz dla dowolnego $x \in X$ mamy:
$$\langle x, a\alpha + b\beta \rangle = a \langle x, \alpha \rangle + b \langle x, \beta \rangle = 0.$$
□

Odnotujmy też, że wyznaczanie podprzestrzeni $X^{\perp}$ ma wiele wspólnego z rozwiązywaniem układów równań liniowych. Jeśli na przykład $X = \{(1,1,2), (1,2,0)\}$ i szukamy w $\mathbb{R}^3$ zbioru $X^{\perp}$ przy standardowym iloczynie skalarnym, to każdy wektor (x_1, x_2, x_3) należący do tego zbioru spełniać musi układ równań:
$$\begin{cases} \langle (x_1,x_2,x_3), (1,1,2) \rangle &= 0 \\ \langle (x_1,x_2,x_3), (1,2,0) \rangle &= 0 \end{cases} \iff \begin{cases} x_1 + x_2 + 2x_3 &= 0 \\ x_2 + 2x_2 &= 0 \end{cases}.$$

Definicja 11.7: Układ prostopadły, baza ortogonalna

Niech h będzie formą dwuliniową na przestrzeni liniowej V.

1. Układ wektorów $X \subseteq V$ nazwiemy PROSTOPADŁYM (albo ORTOGONALNYM), jeśli $\alpha \perp \beta$, dla każdych $\alpha, \beta \in X$.
2. Układ prostopadły będący bazą przestrzeni V nazywamy BAZĄ PROSTOPADŁĄ (albo ortogonalną) przestrzeni V względem h.

Zobaczmy kilka przykładów.

- Jeśli $\langle\,,\rangle_{st}$ jest standardowym iloczynem skalarnym na przestrzeni $\mathbb{R}^n$, to baza standardowa tej przestrzeni jest bazą prostopadłą.
- Układ $(1,1,2),(2,2,-2),(1,-1,0)$ jest bazą prostopadłą przestrzeni $\mathbb{R}^3$ ze standardowym iloczynem skalarnym.
- Jeśli $\langle\,,\rangle$ jest iloczynem skalarnym na przestrzeni V oraz $W \subseteq V$ jest podprzestrzenią, to ograniczenie $\langle\,,\rangle|_W : W \times W \to \mathbb{R}$ dane wzorem $\langle w_1, w_2\rangle|_W = \langle w_1, w_2\rangle$ jest iloczynem skalarnym na W.

  Niech $V = \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, x_1+2x_2-2x_3 = 0\}$. Na przestrzeni V rozważyć możemy ograniczenie standardowego iloczynu skalarnego z przestrzeni $\mathbb{R}^3$. Wówczas układ $(0,1,1),(4,-1,1)$ jest bazą prostopadłą przestrzeni $(V, \langle\,,\rangle|_V)$.
- Funkcja $\langle (x_1,x_2,x_3),(y_1,y_2,y_3)\rangle = x_1y_1 + 2x_1y_2 + 2x_2y_1 + 9x_2y_2 + 4x_3y_3$ jest, jak już wiemy, iloczynem skalarnym na $\mathbb{R}^3$. Układ $(1,0,0)$, $(2,-1,0)$, $(0,0,1)$ jest bazą prostopadłą przestrzeni $\mathbb{R}^3$ z tym iloczynem skalarnym.
- Warunek dodatniej określoności iloczynu skalarnego wymusza, aby w przestrzeni z iloczynem skalarnym niezerowy wektor nie był prostopadły do siebie samego. Dla dowolnej formy dwuliniowej jest to jednak możliwe. Na przykład: dla formy $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ zadanej wzorem

  $$h((x_1,x_2),(y_1,y_2)) = x_1y_2 + x_2y_1$$

  mamy $(1,0) \perp (1,0)$ oraz $(0,1) \perp (0,1)$. To rodzi następujące pytanie: czy w przestrzeni $(\mathbb{R}^2, h)$ istnieje baza ortogonalna? Okazuje się, że tak — wystarczy obrać układ $((1,1),(1,-1))$.

Układy i bazy prostopadłe można teoretycznie rozważać zarówno w przestrzeniach skończonego, jak i nieskończonego wymiaru, ale tylko pierwsza z możliwości daje istotne dla nas wyniki. W przestrzeniach nieskończonego wymiaru problem istnienia bazy ortogonalnej nie jest taki prosty, jak się wydaje[1].

Rozważmy podprzestrzeń $X \subseteq \mathbb{R}^\infty$ złożoną z ciągów ograniczonych $(a_1, a_2, \ldots)$ z iloczynem skalarnym $\langle\,,\rangle$ danym wzorem:

$$\langle (a_n),(b_n)\rangle = \sum_{n=1}^{\infty} \frac{a_n b_n}{n^2}.$$

Niech U będzie podprzestrzenią X złożoną z ciągów o skończonej liczbie niezerowych wyrazów. Wówczas ciągi $(\epsilon_i)_n$ dane wzorami

$$(\epsilon_i)_n = \begin{cases} n, & n = i \\ 0, & n \neq i \end{cases}$$

tworzą bazę ortogonalną U. Jednak $U^\perp = \{0\}$ (sprawdź!) To pokazuje, że układ $\{\epsilon_i\}_{i\in\mathbb{N}}$ jest maksymalnym układem ortogonalnym w X, ale nie jest bazą X. Okazuje się, że X takiej bazy nie ma!

W dalszym ciągu zajmujemy się jedynie przestrzeniami euklidesowymi Do rozważań dotyczących przestrzeni z dowolną formą dwuliniową wrócimy za kilka wykładów. Zacznijmy od obserwacji wskazującej na to, że układy ortogonalne niezerowych wektorów w przestrzeni euklidesowej są liniowo niezależne.

Obserwacja 11.4

Niech $X \subseteq V$ będzie prostopadłym układem niezerowych wektorów przestrzeni V z iloczynem skalarnym $\langle\,,\rangle$. Wówczas układ X jest liniowo niezależny.

Dowód. Weźmy dowolny skończony podukład X, np. $\{\alpha_1,\ldots,\alpha_n\}$. Wystarczy pokazać, że jest on liniowo niezależny. Przypuśćmy, że $\sum_{i=1}^n a_i\alpha_i = 0$, dla pewnych $a_1,\ldots,a_n$. Wówczas dla każdego j:

$$0 = \langle 0, \alpha_j\rangle = \langle \sum_{i=1}^n a_i\alpha_i, \alpha_j\rangle = \sum_{i=1}^n a_i \langle \alpha_i, \alpha_j\rangle = a_j \langle \alpha_j, \alpha_j\rangle,$$

bo $\langle \alpha_i, \alpha_j\rangle = 0$, dla $i \neq j$. Wobec $\langle \alpha_j, \alpha_j\rangle > 0$ dostajemy $a_j = 0$, dla każdego $j = 1,\ldots,n$. □

[1] W analizie funkcjonalnej wprowadza się pojęcie bazy ortogonalnej, ale nie jest to *zwykle* baza w sensie algebry liniowej.

Obserwacja 11.5

Każda przestrzeń euklidesowa ma bazę prostopadłą.

Dowód. Stosujemy indukcję ze względu na wymiar przestrzeni liniowej. Dla przestrzeni wymiaru 1 twierdzenie jest oczywiste, bo każdy jej niezerowy wektor stanowi bazę prostopadłą. Załóżmy, że wykazaliśmy twierdzenie dla przestrzeni wymiaru mniejszego od n. Niech $(V, \langle\,,\rangle)$ będzie n wymiarową przestrzenią euklidesową. Wybierzmy dowolny niezerowy wektor $\alpha \in V$. Pokażemy najpierw, że zachodzi rozkład:

$$V = \operatorname{lin}(\alpha) \oplus \operatorname{lin}(\alpha)^{\perp}.$$

W tym celu trzeba uzasadnić kolejno równości $\operatorname{lin}(\alpha) \cap \operatorname{lin}(\alpha)^{\perp} = \{0\}$ oraz $\operatorname{lin}(\alpha) + \operatorname{lin}(\alpha)^{\perp} = V$.

Załóżmy zatem, że jakiś niezerowy wektor v należy zarówno do $\operatorname{lin}(\alpha)$ jak i do podprzestrzeni $\operatorname{lin}(\alpha)^{\perp}$ ortogonalnej do $\operatorname{lin}(\alpha)$. Zatem $v = a\alpha$, dla pewnego $0 \neq a \in \mathbb{R}$. Skoro wektor $a\alpha$ należy także do $\operatorname{lin}(\alpha)^{\perp}$ to $\langle a\alpha, \alpha\rangle = 0$. Wobec tego, że $\langle a\alpha, \alpha\rangle = a\langle \alpha, \alpha\rangle$ oraz, że $\langle \alpha, \alpha\rangle > 0$, dla każdego $0 \neq \alpha$, dostajemy sprzeczność z wyborem a, bo iloczyn dwóch niezerowych z założenia wartości $a\langle \alpha, \alpha\rangle$ daje zero. Pokazaliśmy zatem, że żaden niezerowy wektor nie może należeć do $\operatorname{lin}(\alpha) \cap \operatorname{lin}(\alpha)^{\perp}$.

Aby pokazać równość $\operatorname{lin}(\alpha) + \operatorname{lin}(\alpha)^{\perp} = V$ musimy rozłożyć dowolny wektor $\beta \in V$ na sumę wektorów, z których pierwszy należy do $\operatorname{lin}(\alpha)$, a drugi do $\operatorname{lin}(\alpha)^{\perp}$. Oto ten rozkład:

$$\beta = \frac{\langle \beta, \alpha\rangle}{\langle \alpha, \alpha\rangle}\alpha + \left(\beta - \frac{\langle \beta, \alpha\rangle}{\langle \alpha, \alpha\rangle}\alpha\right).$$

Pokażemy, że drugi składnik powyższej sumy jest prostopadły do α. Mamy:

$$\langle \beta - \frac{\langle \beta, \alpha\rangle}{\langle \alpha, \alpha\rangle}\alpha, \alpha\rangle = \langle \beta, \alpha\rangle - \frac{\langle \beta, \alpha\rangle}{\langle \alpha, \alpha\rangle}\langle \alpha, \alpha\rangle = \langle \beta, \alpha\rangle - \langle \beta, \alpha\rangle = 0.$$

Pokazaliśmy, że $V = \operatorname{lin}(\alpha) \oplus \operatorname{lin}(\alpha)^{\perp}$. Z twierdzenia o wymiarze sumy prostej $\dim \operatorname{lin}(\alpha)^{\perp} = n - 1$. Przyjmijmy $W = \operatorname{lin}(\alpha)^{\perp}$ i rozważmy przekształcenie $\langle\,,\rangle|_W : W \times W \to \mathbb{R}$ zadane wzorem $\langle w, w'\rangle|_W = \langle w, w'\rangle$, dla dowolnych $w, w' \in W$. Jest to obcięcie przekształcenia $\langle\,,\rangle$ do zbioru $W \times W$, będące oczywiście iloczynem skalarnym. A zatem z założenia indukcyjnego $(W, \langle\,, f\rangle|_W)$ ma bazę prostopadłą, powiedzmy $\beta_1, \ldots, \beta_{n-1}$. Zgodnie z definicją $\langle\,,\rangle|_W$ jest to również układ ortogonalny w V. Co więcej każdy z jego elementów jest prostopadły do α. Zatem układ ortogonalny $\alpha, \beta_1, \ldots, \beta_{n-1}$ jest liniowo niezależny i tworzy bazę ortogonalną V. □

Definicja 11.8: Układ ortonormalny

Mówimy, że układ $\alpha_1, \ldots, \alpha_n$ wektorów przestrzeni euklidesowej $(V, \langle\,,\rangle)$ jest PROSTOPADŁY UNORMOWANY (albo ORTONORMALNY), jeśli $\alpha_1, \ldots, \alpha_n$ jest układem prostopadłym oraz $\|\alpha_i\| = 1$, dla każdego $i = 1, \ldots, n$. Układ prostopadły unormowany będący bazą przestrzeni V nazywamy BAZĄ ORTONORMALNĄ przestrzeni $(V, \langle\,,\rangle)$.

Przykłady.

- Baza standardowa w $\mathbb{R}^n$ jest bazą ortonormalną $(\mathbb{R}^n, \langle\,,\rangle_{st})$.
- Poniższy układ jest bazą ortonormalną przestrzeni $(\mathbb{R}^3, \langle\,,\rangle_{st})$:

$$\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \quad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \quad \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right).$$

Bazy ortonormalne będą miały znaczenie przy omawianiu izometrii. Oczywiście każda przestrzeń euklidesowa $(V, \langle\,,\rangle)$ ma bazę ortonormalną. Istotnie, jeśli $\alpha_1, \ldots, \alpha_n$ jest bazą prostopadłą V, to układ $\frac{1}{\|\alpha_1\|}\alpha_1, \ldots, \frac{1}{\|\alpha_n\|}\alpha_n$ jest w sposób oczywisty ortonormalny, bo $\|a\alpha\| = |a| \cdot \|\alpha\|$, dla $a \in \mathbb{R}$, $\alpha \in V$.

11.1 Uzupełnienie. Ważne typy form dwuliniowych

W tym materiale wspomnimy krótko pewne podstawowe typy form dwuliniowych oraz motywacje dla ich wprowadzania. Najbardziej znanymi formami są symetryczne, ale są naturalne powody rozważania także innych form. Zanim do nich przejdziemy warto wspomnieć pewien ogólny przykład.

Definicja 11.9: Ewaluacja

Niech V będzie przestrzenią liniową nad ciałem K. Funkcję $\epsilon : V \times V^* \to K$ daną wzorem:

$$V \times V^* \ni (v, f) \longrightarrow \epsilon(v, f) = f(v) \in K$$

nazywamy EWALUACJĄ.

Przykład ten omówimy bardziej szczegółowo w dodatku, ale warto wspomnieć w tym miejscu, że mamy następujące naturalne przyporządkowania.

- Dla ustalonego $f \in V^*$ mamy przekształcenie liniowe $V \ni v \longrightarrow \epsilon(v, f) = f(v) \in K$,
- Dla ustalonego $v \in V$ mamy przekształcenie liniowe $V^* \ni f \longrightarrow \epsilon(v, f) \in K$.

Warto zastanowić się nad relacjami przyporządkowań napisanych wyżej. Jeśli zbiór wszystkich form dwuliniowych na V oznaczymy przez $\mathrm{Bil}(V)$, to jest jasne, że jest to przestrzeń liniowa. Czy Czytelnik byłby skłonny przyznać, że powyższe przyporządkowania zadają izomorfizm przestrzeni liniowych? A czy umieliby Państwo uzasadnić izomorfizm $\mathrm{Bil}(V \times V^*) \simeq \mathrm{End}(V)$? A jaka jest relacja $\mathrm{Bil}(V)$ oraz $\mathrm{Bil}(V \times V^*)$? Wreszcie, skoro $\mathrm{Bil}(V)$ jest przestrzenią liniową, to jaka jest jej relacja z podprzestrzenią form dwuliniowych symetrycznych?

Definicja 11.10: Forma dwuliniowa antysymetryczna

Formę dwuliniową h na przestrzeni liniowej V nazywamy ANTYSYMETRYCZNĄ (albo skośnie symetryczną), jeśli dla każdych $\alpha, \beta \in V$ mamy $h(\alpha, \beta) = -h(\beta, \alpha)$.

Formy antysymetryczne mają duże znaczenie chociażby w związku z pojęciem wyznacznika, bowiem choćby forma

$$h((x_1, x_2), (y_1, y_2)) = x_1 y_2 - x_2 y_1$$

odpowiada wyznacznikowi jako formie antysymetrycznej na przestrzeni wierszy lub kolumn macierzy 2×2. Nietrudno widzieć, że jeśli $\mathrm{char}(K) \neq 2$ to mamy jednoznaczny rozkład:

$$h(x, y) = \underbrace{\frac{1}{2}(h(x, y) + h(y, x))}_{forma\ symetryczna} + \underbrace{\frac{1}{2}(h(x, y) - h(y, x))}_{forma\ antysymetryczna}.$$

Stąd przestrzeń form dwuliniowych jest sumą prostą przestrzeni form symetrycznych i antysymetrycznych.

Definicja 11.11: Forma alternująca i forma refleksywna

Niech h będzie formą dwuliniową na przestrzeni liniowej V. Mówimy, że forma h jest

- REFLEKSYWNA, gdy $h(\alpha, \beta) = 0$ implikuje $h(\beta, \alpha) = 0$,
- ALTERNUJĄCA, gdy $h(\alpha, \alpha) = 0$, dla każdego $\alpha \in V$.

Nietrudno wykazać, że każda forma symetryczna jest refleksywna. Również każda antysymetryczna forma jest refleksywna. To są jasne obserwacje. Natomiast nieco mniej oczywiste jest wykazanie, że każda przestrzeń alternująca jest refleksywna. W dalszych rozdziałach wspomnimy o pewnych klasyfikacjach macierzy związanych z wymienionymi wyżej formami.

11.2 Dodatek. Iloczyn skalarny i funkcjonały

Na wykładzie wspomnieliśmy o związkach pomiędzy iloczynem skalarnym na przestrzeni liniowej oraz o przestrzeni sprzężonej. Rozwińmy nieco ten wątek.

Obserwacja 11.6: Twierdzenie Riesza o reprezentacji

Niech $(V, \langle , \rangle)$ będzie liniową przestrzenią euklidesową wymiaru n. Każdemu wektorowi $v \in V$ przyporządkowujemy funkcjonał $f_v \in V^*$ określony wzorem:

$$f_v(u) = \langle u, v \rangle, \text{ gdzie } u \in V.$$

Przyporządkowanie $\Phi : V \to V^*$ zadane wzorem $\Phi(v) = f_v$ jest izomorfizmem przestrzeni liniowych. Jeśli $\mathcal{V} = (v_1, \ldots, v_n)$ jest bazą ortonormalną w V, to układ $(f_{v_1}, \ldots, f_{v_n})$ jest bazą dualną do $\mathcal{V}$.

Dowód. Przekształcenie f_v jest oczywiście liniowe dla każdego $v \in V$, bo $f_v(au) = \langle au, v \rangle = a \langle u, v \rangle = a \cdot f_v(u)$ oraz $f_v(u_1 + u_2) = \langle u_1 + u_2, v \rangle = \langle u_1, v \rangle + \langle u_2, v \rangle = f_v(u_1) + f_v(u_2)$. A zatem rzeczywiście $f_v \in V^*$. Także przyporządkowanie Φ jest liniowe, ponieważ $f_{av}(u) = \langle u, av \rangle = a \langle u, v \rangle = a \cdot f_v(u)$, czyli funkcjonały $\Phi(av)$ oraz $a \cdot \Phi(v)$ są identyczne na V. Także $f_{v_1+v_2}(u) = \langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle = f_{v_1}(u) + f_{v_2}(u)$, czyli funkcjonały $\Phi(v_1+v_2)$ oraz $\Phi(v_1)+\Phi(v_2)$ są równe. Skoro V oraz V^* są tego samego wymiaru to do wykazania, że Φ jest izomorfizmem wystarczy pokazać, że Φ to monomorfizm. Jeśli mamy $\Phi(v) = 0$, to f_v jest tożsamościowo równe 0. W szczególności $f_v(v) = \langle v, v \rangle = \|v\|^2 = 0$. Zatem $v = 0$. Czyli Φ jest izomorfizmem. Ostatnie stwierdzenie dotyczące baz dualnych jest oczywiste (z definicji). □

Niech V będzie przestrzenią liniową skończonego wymiaru oraz V^* niech będzie przestrzenią dualną do V. Wiemy, że przestrzenie V oraz V^* są tego samego wymiaru, a więc istnieje między nimi izomorfizm. Żaden izomorfizm nie jest jednak wyróżniony lub „naturalny". Dlaczego? Można wprawdzie wybrać bazę w V oraz bazę w V^* i zażądać przekształcenia liniowego $i : V \to V^*$, które przeprowadza jedną bazę w drugą. Jest jednak pewien problem. Załóżmy, ze mamy dwie skończenie wymiarowe przestrzenie liniowe V_1 oraz V_2 oraz przekształcenie liniowe $\lambda : V_1 \to V_2$. Wiemy, że indukuje ono przekształcenie liniowe $\lambda^* : V_2^* \to V_1^*$. Chciałoby się uważać, że skoro V i V^* są izomorficzne, to istnieje jakiś naturalny sposób przechodzenia z endomorfizmów V do endomorfizmów V^*. Okazuje się, że nie musi istnieć izomorfizm $i : V \to V^*$ taki, że dla każdego odwracalnego λ spełniona jest równość:

$$\begin{array}{ccc} V & \xrightarrow{i} & V^* \\ {\scriptstyle\lambda}\downarrow & & \uparrow{\scriptstyle\lambda^*} \\ V & \xrightarrow[i]{} & V^* \end{array}$$

Oczywiście jeśli wybierzemy pewien izomorfizm $V \to V^*$ w dolnym wierszu tego diagramu, to można określić i górny, znając λ. Ale nie dobierzemy JEDNEGO izomorfizmu i do WSZYSTKICH λ. Pokazanie tego to stosunkowo trudne ćwiczenie z niesymetrycznych form dwuliniowych[2]. Inaczej mają się sprawy, gdy V jest przestrzenią euklidesową, czyli jest parą $(V, \langle , \rangle)$. Wówczas, jak można pokazać, każdemu endomorfizmowi $\lambda : V \to V$ odpowiada jednoznacznie endomorfizm sprzężony $\lambda' : V \to V$ taki, że

$$\langle \lambda(v), w \rangle = \langle v, \lambda'(w) \rangle .$$

Dzięki niemu mamy naturalną odpowiedniość między $End(V)$ i $End(V^*)$:

$$\begin{array}{ccc} V & \xrightarrow{i} & V^* \\ \uparrow{\scriptstyle\lambda'} & & {\scriptstyle\lambda^*}\uparrow \\ V & \xrightarrow[i]{} & V^* \end{array}$$

Rzeczywiście, dzięki izomorfizmowi Φ możemy o każdym funkcjonale z V^* myśleć jako o pewnym f_v (dla pewnego wektora v). A zatem: bierzemy taki funkcjonał f_v, wykonujemy na nim ϕ^* i dostajemy znowu funkcjonał z V^*, czyli pewien funkcjonał $f_{v'}$ (bo tak nakazuje Φ). Co mają ze sobą wspólnego v oraz v'? Musimy mieć $\phi^*(f_v) = f_{v'}$, czyli dla każdego $u \in V$ mamy $\phi^*(f_v)(u) = f_v(\phi(u)) = f_{v'}(u)$. A zatem z definicji funkcjonałów f_v oraz $f_{v'}$ mamy $\langle \phi(u), v \rangle = \langle u, v' \rangle$.

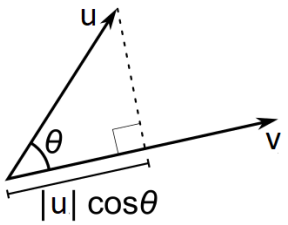
[2] https://www.ams.org/journals/tran/1945-058-00/S0002-9947-1945-0013131-6/S0002-9947-1945-0013131-6.pdf, str. 4.

11.3 Trivia. Dlaczego cosinus?

Powiedzieliśmy, że liczbę $\theta \in [0, \pi]$ taką, że

$$\frac{\langle u, v \rangle}{\|u\| \cdot \|v\|} = \cos\theta$$

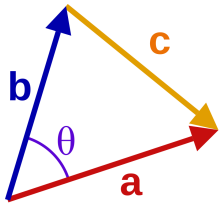
nazywamy (niezorientowanym) kątem między wektorami u i v. Czy taka definicja ma coś wspólnego z geometrią i tym, co znamy jako cosinus? Na poziomie intuicji dać można następującą odpowiedź: jeśli chcielibyśmy pomnożyć dwa wektory poprzez mnożenie ich długości, wypadałoby aby wektory te miały ten sam zwrot. Rozważmy następujący rysunek



Zauważmy, że z definicji kąta wynika równość $\langle u, v \rangle = \|u\| \cdot \|v\| \cdot \cos\theta$, a liczba $\|u\| \cdot |\cos\theta|$ równa jest długości rzutu prostopadłego wektora u na podprzestrzeń $\mathrm{lin}(v)$. O prostopadłości powiemy więcej na następnym wykładzie (a o rzutach — za dwa wykłady). Jednak już teraz możemy powiedzieć, że jeśli u_v jest długością rzutu wektora u na $\mathrm{lin}(v)$ oraz v_u jest długością rzutu v na $\mathrm{lin}(u)$, to zachodzi równość:

$$|\langle u, v \rangle| = u_v \cdot \|v\| = v_u \cdot \|u\|.$$

Bardzo nas ta równość nie dziwi, bowiem $\sqrt{\|u\|^2 - u_v^2} \cdot \|v\| = \sqrt{\|v\|^2 - v_u^2} \cdot \|u\|$ równe jest polu równoległoboku rozpiętego przez wektory u, v (wysokość razy podstawa). Z geometrycznego punktu widzenia równość wyżej może być traktowana jako definicja iloczynu skalarnego, o ile wiemy czym jest kosinus (lub pole). Rozważmy kolejny rysunek



Zauważmy, że posługując się wprowadzoną definicją kąta możemy wykazać twierdzenie cosinusów. Istotnie, dla trójkąta wyżej, wyznaczonego przez kierunki a, b oraz $c = a - b$ mamy:

$$\|c\|^2 = \langle c, c \rangle = \langle a - b, a - b \rangle = \|a\|^2 - 2\langle a, b \rangle + \|b\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\| \cdot \|b\| \cdot \cos\theta.$$

Jaki ma to związek ze współrzędnymi i definicjami iloczynu skalarnego za pomocą różnych form dwuliniowych? Zobaczmy to zagadnienie w $\mathbb{R}^3$ ze standardowym iloczynem skalarnym. Jeśli $a = (x_1, y_1, z_1)$ oraz $b = (x_2, y_2, z_2)$, to $a - b = (x_1 - x_2, y_1 - y_2, z_1 - z_2)$. Stąd (przyjmując standardową normę) mamy:

$$\|c\| = \|a - b\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Jeżeli wystartujemy od twierdzenia cosinusów zapisanego w postaci następujących równości $2\langle a, b \rangle = 2\|a\| \cdot \|b\| \cdot \cos\theta = \|a\|^2 + \|b\|^2 - \|c\|^2$, to otrzymamy:

$$\begin{aligned} 2\langle a, b \rangle = 2\|a\| \cdot \|b\| \cdot \cos\theta &= (x_1^2 + y_1^2 + z_1^2) + (x_2^2 + y_2^2 + z_2^2) - ((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2) = \\ &= 2(x_1x_2 + y_1y_2 + z_1z_2). \end{aligned}$$

A zatem $\langle a, b \rangle = x_1x_2 + y_1y_2 + z_1z_2$. Widzimy więc ponownie — można określać kąt (i za chwilę również pole) za pomocą iloczynu skalarnego i to jest podejście ważne z punktu widzenia naszego kursu. Ze szkolnego punktu widzenia można określić iloczyn skalarny za pomocą podstawowych narzędzi geometrii elementarnej, dowodząc wcześniej twierdzenie cosinusów — odsyłam do tekstu dra Krycha wspominanego w zasadniczym tekście wykładu. Wiadomo jest jednak, że szkolne definicje kąta czy funkcji trygonometrycznej mają w sobie pewne przemilczenia (np. czym dokładnie jest *miara łukowa* czy *pole*). Stąd nasz „kierunek definicji" jest w pewnym sensie odwrotny do „szkolnego". W naszym ujęciu startujemy od abstrakcji (aksjomaty) i idziemy do intuicyjnie znanych obiektów: najpierw wektory i przestrzenie liniowe, potem punkty i przestrzenie afiniczne: proste, płaszczyzny itd., dalej abstrakcyjny iloczyn skalarny, a z nim: kąty, odległości, pole, a jeszcze dalej: odcinki, trójkąty, okręgi, parabole, hiperbole itd.

11.4 Trivia. Zastosowania nierówności Cauchy-ego Schwarza

Nierówność Cauchy'ego Schwarza ma liczne zastosowania algebraiczne, w tym konkursowe. Wykorzystuje się zwłaszcza wersję dla pary wektorów w przestrzeni $\mathbb{R}^n$ ze standadowym iloczynem skalarnym.

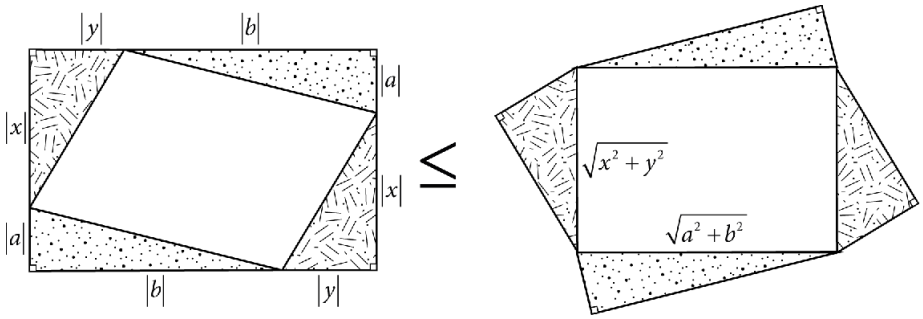
Obserwacja 11.7

Dla dowolnych liczb rzeczywistych $a_1, \ldots, a_n$ oraz $b_1, \ldots, b_n$ mamy:

$$(a_1b_1 + \ldots + a_nb_n)^2 \leqslant (a_1^2 + \ldots + a_n^2)(b_1^2 + \ldots + b_n^2),$$

przy czym równość zachodzi wtedy i tylko wtedy, gdy istnieje $\lambda \in \mathbb{R}$ taka, że $a_i = \lambda b_i$, dla wszystkich i lub $b_i = \lambda a_i$, dla wszystkich i.

Poniższy rysunek ilustruje możliwą geometryczną interpretację nierówności powyżej na płaszczyźnie (po spierwiarstkowaniu obydwu stron).



Źródło. Proofs without words, R. Nielsen.

Bezpośrednim zastosowaniem powyższej nierówności jest dla $a_i = \sqrt{x_i}$ oraz $b_i = 1/\sqrt{x_i}$, gdzie $x_i \geqslant 0$, nierówność między średnią arytmetyczną i harmoniczną (a raczej jej równoważna forma).

$$(x_1 + \ldots + x_n)\left(\frac{1}{x_1} + \ldots + \frac{1}{x_n}\right) \geqslant n^2.$$

Biorąc $b_i = 1$ uzyskujemy nierówność między średnią kwadratową i arytmetyczną

$$\sqrt{\frac{a_1^2 + \ldots + a_n^2}{n}} \geqslant \frac{a_1 + \ldots + a_n}{n}.$$

Jeszcze innym zastosowaniem jest nierówność Minkowskiego (to w zasadzie nierówność trójkąta):

$$\sqrt{x_1^2 + \ldots + x_n^2} + \sqrt{y_1^2 + \ldots + y_n^2} \geqslant \sqrt{(x_1 + y_1)^2 + \ldots + (x_n + y_n)^2}.$$

Oto przykład z I etapu LXXIV OM. Pokazać, że dla dowolnych liczb dodatnich a, b, c mamy:

$$\sqrt{ab^2(b+c)} + \sqrt{bc^2(c+a)} + \sqrt{ca^2(a+b)} \leqslant \frac{1}{2}\sqrt{(a+b+c)^4 - (a^2+b^2+c^2)^2}.$$

Przyjmijmy $x_1 = \sqrt{ab}, x_2 = \sqrt{bc}, x_3 = \sqrt{ca}$ oraz

$$y_1 = \sqrt{b(b+c)}, \quad y_2 = \sqrt{c(c+a)}, \quad y_3 = \sqrt{a(a+b)}.$$

Wówczas teza wynika natychmiast z nierówności Cauchy'ego-Schwarza, bowiem:

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= ab + bc + ca \\ y_1^2 + y_2^2 + y_3^2 &= a^2 + b^2 + c^2 + ab + bc + ca \\ x_1y_1 + x_2y_2 + x_3y_3 &= \sqrt{ab^2(b+c)} + \sqrt{bc^2(c+a)} + \sqrt{ca^2(a+b)} \end{aligned}$$

Z drugiej strony korzystając ze wzorów skróconego mnożenia (wzór na różnicę kwadratów) mamy:

$$(a+b+c)^4 - (a^2+b^2+c^2)^2 = 4(a^2+b^2+c^2+ab+bc+ca)(ab+bc+ca).$$

Zachęcam Czytelnika do poszukania ciekawych wniosków z nierówności Cauchy'ego Schwarza dla niestandardowych iloczynów skalarnych na $\mathbb{R}^n$ oraz dla innych przestrzeni euklidesowych.

Rozdział 12

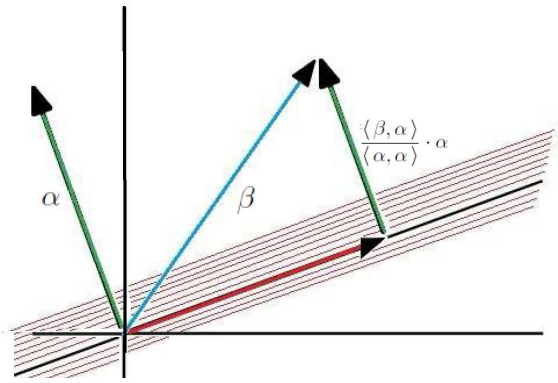
Rzut prostopadły i ortogonalizacja Przestrzenie euklidesowe afiniczne

Na poprzednim wykładzie wprowadziliśmy pojęcie prostopadłości pary wektorów, a także pojęcie dopełnienia ortogonalnego. W kontekście przestrzeni euklidesowych pojęcia te mają szczególnie przejrzysty opis. Wykazaliśmy, że każda taka przestrzeń posiada bazę prostopadłą. Dziś zaczniemy od wywnioskowania, że dla dowolnej podprzestrzeni W przestrzeni euklidesowej V ma miejsce rozkład $V = W \oplus W^{\perp}$. Wychodzimy od następującej obserwacji, poczynionej na poprzednim wykładzie.

Obserwacja 12.1

Jeśli α jest niezerowym wektorem w przestrzeni liniowej V z iloczynem skalarnym $\langle\, , \rangle$ oraz $\beta \in V$, to mamy rozkład, w którym drugi składnik należy do $\mathrm{lin}(\alpha)^{\perp}$

$$\beta = \frac{\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle} \cdot \alpha + \left(\beta - \frac{\langle \beta, \alpha \rangle}{\langle \alpha, \alpha \rangle} \cdot \alpha \right),$$



Rozpocznijmy od banalnej obserwacji.

Obserwacja 12.2

Jeśli $\alpha_1, \ldots, \alpha_n$ jest bazą podprzestrzeni $W \subseteq V$, to następujące warunki są równoważne dla $v \in V$:

(i) $v \in W^{\perp}$,

(ii) $\langle v, \alpha_i \rangle = 0$, dla każdego i.

Dowód. Implikacja $(i) \Rightarrow (ii)$ jest jasna. W drugą stronę: chcemy, by $v \in V$ był prostopadły do każdego wektora w z W, czyli wektora postaci $w = a_1\alpha_1 + \ldots + a_n\alpha_n$. Zatem $\langle v, w \rangle = \sum_i a_i \langle v, \alpha_i \rangle = 0$. □

Obserwacja 12.3

Niech $(V, \langle\,,\rangle)$ będzie n wymiarową przestrzenią euklidesową i niech W będzie podprzestrzenią przestrzeni V. Wówczas:

(a) $V = W \oplus W^{\perp}$,

(b) jeśli $\dim(W) = k$, to $\dim(W^{\perp}) = n - k$,

(c) $(W^{\perp})^{\perp} = W$.

Dowód. Dowodzimy (a). Podobnie jak w dowodzie twierdzenia o istnieniu bazy prostopadłej musimy wykazać, że $W \cap W^{\perp} = \{0\}$ oraz, że $W + W^{\perp} = V$. Pierwszy punkt jest oczywisty. Jeśli $w \in W \cap W^{\perp}$, to $\langle w, w \rangle = 0$, czyli $w = 0$. Niech $\alpha_1, \ldots, \alpha_k$ będzie bazą prostopadłą W, czyli $\langle \alpha_i, \alpha_j \rangle = 0$, dla $i \neq j$ (już wiemy, że taka baza istnieje). Dla każdego wektora $\alpha \in V$ definiujemy:

$$\alpha' = \sum_{i=1}^{k} \frac{\langle \alpha, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \alpha_i.$$

Wówczas oczywiście $\alpha' \in W$. Co więcej, okazuje się, że $\alpha - \alpha' \in W^{\perp}$. Sprawdzimy to pokazując, że $\alpha - \alpha'$ jest prostopadły do każdego wektora bazowego w W. Istotnie, dla każdego $j = 1, \ldots, k$ mamy:

$$\langle \alpha - \alpha', \alpha_j \rangle = \langle \alpha - \sum_{i=1}^{k} \frac{\langle \alpha, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \alpha_i, \alpha_j \rangle = \langle \alpha, \alpha_j \rangle - \sum_{i=1}^{k} \frac{\langle \alpha, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \langle \alpha_i, \alpha_j \rangle =$$
$$= \langle \alpha, \alpha_j \rangle - \frac{\langle \alpha, \alpha_j \rangle}{\langle \alpha_j, \alpha_j \rangle} \langle \alpha_j, \alpha_j \rangle = \langle \alpha, \alpha_j \rangle - \langle \alpha, \alpha_j \rangle = 0.$$

A zatem każdy wektor $\alpha \in W$ ma rozkład $\alpha = \alpha' + (\alpha - \alpha') \in W + W^{\perp}$. Stąd wynika punkt (a). Oczywiście (b) jest natychmiastową konsekwencją (a). Jeśli chodzi o (c) zauważmy, że mamy $W \subseteq (W^{\perp})^{\perp}$, bo każdy wektor z W jest prostopadły do każdego wektora z $W^{\perp}$. Zauważmy jednak, że stosując (b) mamy $\dim(W^{\perp})^{\perp} = n - \dim(W^{\perp}) = n - (n - \dim W) = \dim W$. Zatem $W = (W^{\perp})^{\perp}$. □

Twierdzenie to pozwala opisywać zbiory ortogonalne i jego teza jest szczególnie przejrzysta gdy weźmiemy przestrzeń $\mathbb{R}^n$ ze standardowym iloczynem skalarnym i w niej podprzestrzeń W opisaną układem równań jednorodnych o macierzy $A \in M_{m \times n}(K)$. Jeśli przez $\alpha_1, \ldots, \alpha_m$ oznaczymy wiersze macierzy A, to $W^{\perp} = \mathrm{lin}(\alpha_1, \ldots, \alpha_m)$. Oczywiście gdy iloczyn skalarny nie jest standardowy tak nie musi być.

Zobaczmy przykład wyznaczania bazy prostopadłej oparty o pokazany dowód. Niech V będzie przestrzenią $\mathbb{R}^5$ z iloczynem skalarnym $\langle\,,\rangle$ zadanym wzorem:

$$\langle (x_1, x_2, x_3, x_4, x_5), (y_1, y_2, y_3, y_4, y_5) \rangle = 2x_1y_1 + x_1y_4 + x_2y_2 + 2x_3y_3 + x_4y_1 + 2x_4y_4 + x_5y_5$$

Niech $W \subseteq \mathbb{R}^5$ będzie opisana układem równań:

$$\begin{cases} 3x_1 + x_2 + 3x_3 + 2x_4 &= 0 \\ 2x_1 + x_2 + 2x_3 + 3x_4 &= 0 \end{cases}.$$

Nietrudno widzieć, że W jest trójwymiarowa. Weźmiemy dowolny niezerowy wektor α do niej należący, powiedzmy

$$\alpha = (-1, 0, 1, 0, 1)$$

i wyznaczmy przestrzeń $\mathrm{lin}(\alpha)^{\perp}$. Zawiera ona wszystkie wektory $v \in \mathbb{R}^5$ postaci $(x_1, x_2, x_3, x_4, x_5)$ takie, że $\langle \alpha, v \rangle = 0$, czyli (pamiętajmy — $\langle\,,\rangle$ nie jest standardowy iloczyn skalarny):

$$2 \cdot (-1) \cdot x_1 + 1 \cdot (-1) \cdot x_4 + 1 \cdot 0 \cdot x_2 + 2 \cdot 1 \cdot x_3 + 1 \cdot 0 \cdot x_1 + 2 \cdot 0 \cdot x_4 + 1 \cdot 1 \cdot x_5 = 0.$$

A zatem przestrzeń $\mathrm{lin}(\alpha)^{\perp}$ opisana jest równaniem:

$$-2x_1 + 2x_3 - x_4 + x_5 = 0.$$

Wektory β, γ, które mają dopełnić α do bazy ortogonalnej W pochodzą z $\lin(\alpha)^\perp$, oraz oczywiście są elementami W. A zatem β, γ należą do przestrzeni $W' = \lin(\alpha)^\perp \cap W$ opisanej układem:

$$\begin{cases} 3x_1 + x_2 + 3x_3 + 2x_4 &= 0 \\ 2x_1 + x_2 + 2x_3 + 3x_4 &= 0 \\ -2x_1 + 2x_3 - x_4 + x_5 &= 0. \end{cases}.$$

Widzimy zatem, że dowolne dwa wektory prostopadłe z W' dopełnią α do bazy W. Jak je zatem znaleźć? Tak, jakbyśmy startowali od W'. Bierzemy dowolne niezerowe $\beta \in W'$ i szukamy wektora $\gamma \neq 0$ w podprzestrzeni $\lin(\beta)^\perp \cap W'$. Jest jasne, że

$$\lin(\beta)^\perp \cap W' = \lin(\alpha)^\perp \cap \lin(\beta)^\perp \cap W.$$

A zatem biorąc $\beta = (-1, 0, 1, 0, -4)$ widzimy, że $\lin(\beta)^\perp$ opisana jest równaniem $-x_1 + x_3 - x_5 = 0$. Stąd brakujący wektor γ szukanej bazy leży w przestrzeni $\lin(\beta)^\perp \cap W'$ opisanej układem:

$$\begin{cases} 3x_1 + x_2 + 3x_3 + 2x_4 &= 0 \\ 2x_1 + x_2 + 2x_3 + 3x_4 &= 0 \\ -2x_1 + 2x_3 - x_4 + x_5 &= 0 \\ -x_1 + x_3 - x_5 &= 0 \end{cases}.$$

A zatem baza tworzona jest krok po kroku: zaczynamy od $\alpha \in W$, potem bierzemy niezerowy wektor $\beta \in \lin(\alpha)^\perp \cap W$, potem (tak naprawdę) wektor $\gamma \in \lin(\alpha, \beta)^\perp \cap W$. Odrobinę prostszym zadaniem jest prośba o wyznaczenie dowolnej bazy prostopadłej przestrzeni $\mathbb{R}^5$ przy zadanym iloczynie skalarnym. Idea jest wówczas taka sama, przy czym przyjmujemy $W = \mathbb{R}^5$.

Obserwacja 12.4

Niech $\mathcal{A} = \{\alpha_1, \ldots, \alpha_n\}$ będzie bazą prostopadłą przestrzeni euklidesowej $(V, \langle , \rangle)$. Wówczas dla każdego wektora $\alpha = a_1\alpha_1 + \ldots + a_n\alpha_n \in V$ jego współrzędne w bazie $\mathcal{A}$ wynoszą:

$$a_1 = \frac{\langle \alpha, \alpha_1 \rangle}{\langle \alpha_1, \alpha_1 \rangle}, \quad a_2 = \frac{\langle \alpha, \alpha_2 \rangle}{\langle \alpha_2, \alpha_2 \rangle}, \quad \ldots \quad , \quad a_n = \frac{\langle \alpha, \alpha_n \rangle}{\langle \alpha_n, \alpha_n \rangle}.$$

Jeśli $\mathcal{A}$ jest bazą ortonormalną, wówczas mamy

$$a_1 = \langle \alpha, \alpha_1 \rangle, \quad a_2 = \langle \alpha, \alpha_2 \rangle, \quad \ldots \quad , \quad a_n = \langle \alpha, \alpha_n \rangle.$$

Dowód. Dla każdego $i = 1, \ldots, n$ oraz dla każdego $j \neq i$ mamy $\langle \alpha_j, \alpha_i \rangle = 0$, czyli:

$$\langle \alpha, \alpha_i \rangle = \langle a_1\alpha_1 + a_2\alpha_2 + \ldots + a_n\alpha_n, \alpha_i \rangle = a_1 \langle \alpha_1, \alpha_i \rangle + \ldots + a_i \langle \alpha_i, \alpha_i \rangle + \ldots + a_n \langle \alpha_n, \alpha_i \rangle = a_i \langle \alpha_i, \alpha_i \rangle.$$

Stąd $a_i = \frac{\langle \alpha, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle}$. □

Powyższe rozważania prowadzą do następującej definicji.

Definicja 12.1: Rzut i symetria prostopadła

Niech $(V, \langle , \rangle)$ będzie przestrzenią euklidesową i niech $W \subseteq V$ będzie podprzestrzenią.

- Przekształcenie liniowe $\phi : V \to V$ będące rzutem na W wzdłuż $W^\perp$ nazywamy RZUTEM PROSTOPADŁYM na W.
- Przekształcenie liniowe $\psi : V \to V$ będące symetrią względem W wzdłuż $W^\perp$ nazywamy SYMETRIĄ PROSTOPADŁĄ względem W.

Z opisu współrzędnych w bazie prostopadłej wynika natychmiast następująca obserwacja.

Obserwacja 12.5: Rzut prostopadły

Niech $W \subseteq V$ będzie podprzestrzenią przestrzeni euklidesowej $(V, \langle\,,\rangle)$, niech $(\alpha_1, \ldots, \alpha_k)$ będzie bazą prostopadłą przestrzeni W oraz niech $(\alpha_{k+1}, \ldots, \alpha_n)$ niech będzie bazą prostopadłą przestrzeni $W^\perp$. Wówczas dla każdego wektora $\alpha \in V$:

- rzut prostopadły $\phi(\alpha)$ wektora α na W wynosi:

$$\phi(\alpha) = \sum_{i=1}^{k} \frac{\langle\, \alpha, \alpha_i \,\rangle}{\langle\, \alpha_i, \alpha_i \,\rangle} \alpha_i,$$

- symetria prostopadła $\psi(\alpha)$ wektora α względem W wynosi:

$$\psi(\alpha) = \sum_{i=1}^{k} \frac{\langle\, \alpha, \alpha_i \,\rangle}{\langle\, \alpha_i, \alpha_i \,\rangle} \alpha_i - \sum_{j=k+1}^{n} \frac{\langle\, \alpha, \alpha_j \,\rangle}{\langle\, \alpha_j, \alpha_j \,\rangle} \alpha_j.$$

Oczywiście powyższe wzory upraszczają się zdecydowanie w sytuacji, gdy baza $\alpha_1, \ldots, \alpha_n$ jest ortonormalna. Wówczas wzory na rzut i symetrię mają odpowiednio postać:

$$\phi(\alpha) = \sum_{i=1}^{k} \langle\, \alpha, \alpha_i \,\rangle\, \alpha_i, \quad \psi(\alpha) = \sum_{i=1}^{k} \langle\, \alpha, \alpha_i \,\rangle\, \alpha_i - \sum_{j=k+1}^{n} \langle\, \alpha, \alpha_j \,\rangle\, \alpha_j.$$

Bazę prostopadłą przestrzeni euklidesowej $(V, \langle\,,\rangle)$ można otrzymać nie tylko poprzez wyliczanie dopełnień ortogonalnych, ale też przez odpowiednią modyfikację dowolnej bazy zadanej przestrzeni V. Metoda ta nazywa się ortogonalizacją Grama-Schmidta i posiada liczne zastosowania. Geometrycznie polega ona na zastąpieniu bazy $\gamma_1, \ldots, \gamma_n$ bazą $\alpha_1, \ldots, \alpha_n$, w której dla $i = 2, \ldots, n$ wektor α_i jest rzutem prostopadłym wektora γ_i na podprzestrzeń $\operatorname{lin}(\gamma_1, \ldots, \gamma_{i-1})^\perp$.

Twierdzenie 12.1: Ortogonalizacja Grama-Schmidta

Jeśli $(V, \langle\,,\rangle)$ jest przestrzenią euklidesową oraz $\gamma_1, \ldots, \gamma_n$ jest bazą przestrzeni V, to układ $\alpha_1, \ldots, \alpha_n$ zadany indukcyjnie wzorami $\alpha_1 = \gamma_1$ oraz dla $j > 1$:

$$\alpha_j = \gamma_j - \sum_{i=1}^{j-1} \frac{\langle\, \gamma_j, \alpha_i \,\rangle}{\langle\, \alpha_i, \alpha_i \,\rangle} \alpha_i$$

jest bazą prostopadłą przestrzeni V. Ponadto dla każdego $j = 1, \ldots, n$ mamy:

$$\operatorname{lin}(\alpha_1, \ldots, \alpha_j) = \operatorname{lin}(\gamma_1, \ldots, \gamma_j).$$

Uwaga. Bardzo istotne jest założenie $V_s = \operatorname{lin}(\alpha_1, \ldots, \alpha_s) = \operatorname{lin}(\gamma_1, \ldots, \gamma_s)$, które mówi, że w $s+1$-wszym kroku algorytmu, gdy udało nam się już zamienić bazę $\gamma_1, \ldots, \gamma_s$ przestrzeni V_s na bazę prostopadłą $\alpha_1, \ldots, \alpha_s$ – rzutujemy γ_{s+1} na $\operatorname{lin}(\alpha_1, \ldots, \alpha_s)^\perp$ i w ten sposób dostajemy α_{s+1}, uzupełniający $\alpha_1, \ldots, \alpha_s$ do bazy prostopadłej V_{s+1}.

Przykład. Rozważmy $(\mathbb{R}^3, \langle\,,\rangle_{st})$ z bazą $\gamma_1 = (1, -1, 1)$, $\gamma_2 = (1, 0, 1)$, $\gamma_3 = (1, 1, 2)$. Zaprezentujmy algorytm ortogonalizacji.

$$\alpha_1 = (1, -1, 1), \quad \alpha_2 = (1, 0, 1) - \frac{\langle (1,0,0), (1,-1,1) \rangle}{\langle (1,-1,1), (1,-1,1) \rangle} (1, -1, 1) = (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}),$$

$$\alpha_3 = (1, 1, 2) - \frac{\langle (1,1,2), (1,-1,1) \rangle}{\langle (1,-1,1), (1,-1,1) \rangle} (1, -1, 1) - \frac{\langle (1,1,2), (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}) \rangle}{\langle (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}) \rangle} (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}) = (\frac{-1}{2}, 0, \frac{1}{2}).$$

Dostaliśmy bazę prostopadłą $(1, -1, 1), (\frac{1}{3}, \frac{2}{3}, \frac{1}{3}), (\frac{-1}{2}, 0, \frac{1}{2})$.

Dowód. Stosujemy indukcję względem n. Dla $n = 1$ teza jest oczywista. Załóżmy, że wykazaliśmy twierdzenie dla $n-1$ i dowodzimy dla n. Niech $W = \mathrm{lin}(\gamma_1, \ldots, \gamma_{n-1})$. Na mocy założenia indukcyjnego wzory określone w tezie określają bazę prostopadłą $\alpha_1, \ldots, \alpha_{n-1}$ przestrzeni W. Stąd mamy rozkład wektora γ_n na sumę wektorów z W oraz $W^\perp$, której składnikami są

$$\sum_{i=1}^{n-1} \frac{\langle \gamma_n, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \alpha_i \in W \quad \text{oraz} \quad \gamma_n - \sum_{i=1}^{n-1} \frac{\langle \gamma_n, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \alpha_i \in W^\perp.$$

Drugi z tych wektorów jest niezerowy, bo $\gamma_n \notin \mathrm{lin}(\gamma_1, \ldots, \gamma_{n-1}) = \mathrm{lin}(\alpha_1, \ldots, \alpha_{n-1})$ (tu wykorzystujemy ponownie założenie indukcyjne, tym razem wobec do drugiej części tezy). A zatem układ:

$$\alpha_1, \ldots, \alpha_{n-1}, \gamma_n - \sum_{i=1}^{n-1} \frac{\langle \gamma_n, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} \alpha_i$$

składa się z n niezerowych wektorów. Jest to więc baza prostopadła V. □

Wniosek 12.1

Niech $(V, \langle\,,\rangle)$ będzie przestrzenią euklidesową.[a] Wówczas każdy układ prostopadły (odp. ortonormalny) dopełnić można do bazy prostopadłej (odp. ortonormalnej) przestrzeni V.

[a] A co się w zasadzie dzieje, gdy próbujemy wykonać proces ortogonalizacji na układzie liniowo zależnym?

Dowód. Niech $\alpha_1, \ldots, \alpha_k$ będzie układem prostopadłym w V. Na mocy twierdzenia Steinitza możemy dopełnić ten układ do bazy V wektorami $\gamma_{k+1}, \ldots, \gamma_n$. Stosujemy proces ortogonalizacji do rozszerzonego przez nas układu $\alpha_1, \ldots, \alpha_k, \gamma_{k+1}, \ldots, \gamma_n$ i uzyskujemy bazę prostopadłą $\alpha_1, \ldots, \alpha_k, \alpha_{k+1}, \ldots, \alpha_n$. Rzeczywiście, w pierwszych k krokach dostaniemy po prostu układ prostopadły $\alpha_1, \ldots, \alpha_k$. □

Odnotujmy poboczny, ale prosty wniosek dotyczący teorii endomorfizmów. Czy Czytelnik zechciałby uzasadnić poniższy rezultat?

Obserwacja 12.6

Niech $(V, \langle\,,\rangle)$ będzie przestrzenią euklidesową oraz niech $\phi \in \mathrm{End}(V)$ będzie triangularyzowalny. Wówczas istnieje baza prostopadła przestrzeni V, w której macierz ϕ jest górnotrójkątna.

Rzut prostopadły ma ważną interpretację dotyczącą szacowania norm, o licznych zastosowaniach.

Obserwacja 12.7

Niech W będzie podprzestrzenią przestrzeni euklidesowej V oraz niech $v \in V$ rozkłada się (jednoznacznie) na sumę $v' + v''$, gdzie $v' \in W$ oraz $v'' \in W^\perp$. Wówczas

$$\|v - v'\| = \min_{w \in W} \|v - w\|.$$

Dowód. Dowód wynika z definicji normy i twierdzenia Pitagorasa. Dla każdego $w \in W$ mamy:

$$\|v - v'\|^2 \leqslant \|v - v'\|^2 + \|v' - w\|^2 = \|(v - v') + (v' - w)\|^2 = \|v - w\|^2.$$

□

Przejdziemy teraz do rozważania na rzeczywistej przestrzeni afinicznej struktury przestrzeni euklidesowej. Uzyskamy w ten sposób „dostęp" do wielu obiektów geometrycznych znanych ze szkoły.

Definicja 12.2: Przestrzeń euklidesowa afiniczna

Parę $(H, \langle , \rangle)$, gdzie H jest skończenie wymiarową przestrzenią afiniczną nad ciałem $\mathbb{R}$, oraz gdzie $(T(H), \langle , \rangle)$ jest przestrzenią euklidesową nazywamy PRZESTRZENIĄ EUKLIDESOWĄ AFINICZNĄ

Każda przestrzeń euklidesowa liniowa jest przestrzenią euklidesową afiniczną. Jeśli M jest podprzestrzenią przestrzeni euklidesowej afinicznej $(H, \langle , \rangle)$, to $(M, \langle , \rangle|_{T(M)})$ jest przestrzenią euklidesową afiniczną. Każda podprzestrzeń afiniczna przestrzeni euklidesowej liniowej jest przestrzenią euklidesową afiniczną.

Definicja 12.3: Odległość

Niech $(H, \langle , \rangle)$ będzie przestrzenią euklidesową afiniczną. ODLEGŁOŚCIĄ punktów $p, q \in H$ nazywamy długość wektora łączącego p z q. Liczbę tę oznaczamy $\rho(p, q)$. Zatem $\rho(p, q) = \|\overrightarrow{pq}\|$.

To, co najczęściej nazywa się w potocznym matematycznym języku odległością euklidesową[1] punktów $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$ w przestrzeni $\mathbb{R}^n$, to norma wektora $\|\overrightarrow{ab}\|$ przy standardowym iloczynie skalarnym dana wzorem:

$$\rho(a, b) = \sqrt{(b_1 - a_1)^2 + \ldots + (b_n - a_n)^2}.$$

Na przykład w $\mathbb{R}^2$ ze standardowym iloczynem skalarnym odległość punktów $(2, 3, 1), (6, 1, 4)$ to

$$\rho((2, 3, 1), (6, 1, 4)) = \sqrt{(6 - 2)^2 + (1 - 3)^2 + (4 - 1)^2} = \sqrt{29}.$$

Obserwacja 12.8

Dla punktów p, q, r przestrzeni euklidesowej afinicznej mamy:

$$(a)\ \rho(p, q) = 0 \Leftrightarrow p = q, \quad (b)\ \rho(p, q) = \rho(q, p), \quad (c)\ \rho(p, q) + \rho(q, r) \geqslant \rho(p, r).$$

Dowód. Dowód (a): $\|\overrightarrow{pq}\| = 0 \iff \overrightarrow{pq} = 0 \iff p = q$. Punkt (b) wynika z formuły: $\|\overrightarrow{pq}| = \| - \overrightarrow{pq}\|$, a punkt (c) z nierówności trójkąta dla przestrzeni euklidesowych liniowych, czyli: $\|\alpha\| + \|\beta\| \geqslant \|\alpha + \beta\|$. Mamy zaś: $\overrightarrow{pr} = \overrightarrow{pq} + \overrightarrow{qr}$, więc $\|\overrightarrow{pq}\| + \|\overrightarrow{qr}\| \geqslant \|\overrightarrow{pr}\|$. □

Na kolejnych latach studiów dowiecie się Państwo, że wiele innych rzeczywistych funkcji (niż zdefiniowana wyżej norma różnicy wektorów zadająca „odległość euklidesową") na zbiorze $X \times X$ spełnia warunki (a)-(c) podane wyżej. Funkcje te nazywane są ogólnie METRYKAMI. Wiele z nich można określić w $\mathbb{R}^n$. Przykładem jest „metryka miejska" – dla punktów $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n)$ w $\mathbb{R}^n$ określamy:

$$\mu(a, b) = |a_1 - b_1| + \ldots + |a_n - b_n|.$$

Można liczyć „odległości" obiektów niekoniecznie przypominających geometryczne, na przykład definiując

- METRYKĘ RZĘDU: dla $A, B \in M_n(K)$ określamy $d(A, B) = r(A - B)$,
- ODLEGŁOŚĆ HAMMINGA dwóch słów długości n nad alfabetem Σ: liczba pozycji, na których słowa się różnią. Na przykład odległość między ZAGRABIĆ i ZATRĄBIŁ wynosi 3,
- ODLEGŁOŚĆ GEODEZYJNĄ wierzchołków grafu: długość najkrótszej ścieżki (geodezyjnej) między wierzchołkami grafu (jeśli dwa wierzchołki nie są połączone to za ich odległość przyjmuje się ∞.).

Kluczowym narzędziem w przestrzeniach euklidesowych liniowych są bazy prostopadłe i bazy ortonormalne. Podobnie jest w przestrzeniach afinicznych. Chcemy bowiem (chociażby) wyznaczać odległości nie tylko pomiędzy punktami, ale też pomiędzy podzbiorami, a zwłaszcza podprzestrzeniami afinicznymi – na przykład odległość punktu od prostej, odległość dwóch prostych równoległych itd.

[1] Niestandardowe odległości euklidesowe wyrażają się „jakościowo" identycznymi wzorami. Dla nas umiejętność liczenia niestandardowej odległości w $\mathbb{R}^n$ jest istotna, natomiast z szerszego punktu widzenia stosuje się jednolite nazewnictwo.

Definicja 12.4: Prostopadły układ bazowy

Mówimy, że układ $p_0; \alpha_1, \ldots, \alpha_n$ jest UKŁADEM BAZOWYM PROSTOPADŁYM (odpowiednio: UKŁADEM BAZOWYM ORTONORMALNYM) przestrzeni euklidesowej afinicznej $(H, \langle\,,\rangle)$, jeśli p_0 jest punktem przestrzeni H oraz $\alpha_1, \ldots, \alpha_n$ jest bazą prostopadłą (odpowiednio: bazą ortonormalną) przestrzeni euklidesowej liniowej $(T(H), \langle\,,\rangle)$.

Z istnienia baz prostopadłych i baz ortonormalnych przestrzeni euklidesowych liniowych wynika oczywiście istnienie układów bazowych prostopadłych i układów bazowych ortonormalnych przestrzeni euklidesowych afinicznych.

Przykład. Niech

$$H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, 2x_1 + x_2 - 3x_3 = 4\}$$

i niech $\langle\,,\rangle_{st}$ będzie standardowym iloczynem skalarnym w $\mathbb{R}^3$. Wówczas $(1,2,0); (1,1,1), (4,-5,1)$ jest prostopadłym układem bazowym przestrzeni afinicznej $(H, \langle\,,\rangle\,|_H)$, a układ $(1,2,0); \frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{42}}(4,-5,1)$ jest ortonormalnym układem bazowym tej przestrzeni.

Definicja 12.5: Rzut i symetria prostopadła

Niech $(H, \langle\,,\rangle)$ będzie przestrzenią euklidesową afiniczną i niech $M \subseteq H$ będzie podprzestrzenią przestrzeni H.

- RZUTEM PROSTOPADŁYM na M nazywamy przekształcenie afiniczne $f : H \to H$ będące rzutem na M wzdłuż $q + T(M)^\perp$, dla pewnego $q \in H$.
- SYMETRIĄ PROSTOPADŁĄ względem M nazywamy przekształcenie afiniczne $g : H \to H$ będące symetrią względem M wzdłuż $q + T(M)^\perp$, dla pewnego $q \in H$.

Zgodnie z definicją afinicznych rzutów i symetrii równoległych wybór punktu $q \in H$ jest tu dowolny; wynik jest ten sam dla wszystkich q. Zauważmy też, że jeśli $f : H \to H$ jest afinicznym rzutem prostopadłym na M, to jego pochodna $f' : T(H) \to T(H)$ jest rzutem prostopadłym na $T(M)$, podobnie dla symetrii.

Obserwacja 12.9

Niech $M \subseteq H$ będzie podprzestrzenią przestrzeni euklidesowej afinicznej $(H, \langle\,,\rangle)$ Niech p_0 będzie punktem przestrzeni M, niech $\alpha_1, \ldots, \alpha_k$ będzie bazą prostopadłą przestrzeni $T(M)$ i niech $\alpha_{k+1}, \ldots, \alpha_n$ będzie bazą przestrzeni $T(M)^\perp$. Wówczas dla każdego wektora $\alpha \in T(H)$ rzut prostopadły punktu $p_0 + \alpha$ na M wynosi:

$$p_0 + \frac{\langle\,\alpha, \alpha_1\,\rangle}{\langle\,\alpha_1, \alpha_1\,\rangle}\alpha_1 + \ldots + \frac{\langle\,\alpha, \alpha_k\,\rangle}{\langle\,\alpha_k, \alpha_k\,\rangle}\alpha_k,$$

a obraz punktu $p_0 + \alpha$ w symetrii prostopadłej względem M wynosi:

$$p_0 + \left(\frac{\langle\,\alpha, \alpha_1\,\rangle}{\langle\,\alpha_1, \alpha_1\,\rangle}\alpha_1 + \ldots + \frac{\langle\,\alpha, \alpha_k\,\rangle}{\langle\,\alpha_k, \alpha_k\,\rangle}\alpha_k\right) - \left(\frac{\langle\,\alpha, \alpha_{k+1}\,\rangle}{\langle\,\alpha_{k+1}, \alpha_{k+1}\,\rangle}\alpha_{k+1} + \ldots + \frac{\langle\,\alpha, \alpha_n\,\rangle}{\langle\,\alpha_n, \alpha_n\,\rangle}\alpha_n\right).$$

Jeśli powyższy układ bazowy jest ortonormalny, to rzut prostopadły punktu $p_0 + \alpha$ na M wynosi:

$$p_0 + \langle\,\alpha, \alpha_1\,\rangle\,\alpha_1 + \ldots + \langle\,\alpha, \alpha_k\,\rangle\,\alpha_k,$$

zaś obraz punktu $p_0 + \alpha$ w symetrii prostopadłej względem M wynosi:

$$p_0 + (\langle\,\alpha, \alpha_1\,\rangle\,\alpha_1 + \ldots + \langle\,\alpha, \alpha_k\,\rangle\,\alpha_k) - (\langle\,\alpha, \alpha_{k+1}\,\rangle\,\alpha_{k+1} + \ldots + \langle\,\alpha, \alpha_n\,\rangle\,\alpha_n).$$

Definicja 12.6: Odległość punktu od przestrzeni afinicznej

Niech $(H, \langle , \rangle)$ będzie przestrzenią euklidesową afiniczną. Odległością punktu $p \in H$ do podprzestrzeni $M \subseteq H$ nazywamy odległość punktu p od jego rzutu prostopadłego na M. Odległość tę oznaczamy $\rho(p, M)$.

Obserwacja 12.10

Jeśli M jest podprzestrzenią przestrzeni euklidesowej afinicznej H i p jest punktem przestrzeni H, to dla każdego punktu $q \in M$ zachodzi $\rho(p, q) \geqslant \rho(p, M)$.

Dowód. Niech p_0 będzie rzutem prostopadłym punktu p na M. Mamy zatem $\rho(p, M) = \rho(p, p_0)$. Przy tym $\overrightarrow{pp_0} \perp \overrightarrow{p_0q}$, dla każdego $q \in M$. Zatem z twierdzenia Pitagorasa $\|\overrightarrow{pq}\|^2 = \|\overrightarrow{pp_0}\|^2 + \|\overrightarrow{p_0q}\|^2$, a stąd

$$\rho(p, q) = \|\overrightarrow{pq}\| \geqslant \|\overrightarrow{pp_0}\| = \rho(p, p_0).$$

□

Przykład. W przestrzeni afinicznej $(\mathbb{R}^3, \langle , \rangle_{st})$ rozpatrzmy: prostą $M = (2, 3, 1) + \mathrm{lin}((1, 2, -1))$ i punkt $p = (5, 4, 7) \notin M$. Mamy $p = (2, 3, 1) + (3, 1, 6)$, a więc rzut prostopadły p na M to

$$(2, 3, 1) + \frac{\langle (3, 1, 6), (1, 2, -1) \rangle}{\langle (1, 2, -1), (1, 2, -1) \rangle}(1, 2, -1) = (2, 3, 1) - \frac{1}{6}(1, 2, -1) = \left(\frac{11}{6}, \frac{16}{6}, \frac{7}{6}\right).$$

A zatem odległość punktu p od prostej M równa jest

$$\rho\left((5, 4, 7), \left(\frac{11}{6}, \frac{16}{6}, \frac{7}{6}\right)\right) = \|\frac{1}{6}(19, 8, 35)\| = \frac{\sqrt{1650}}{6} = \sqrt{\frac{275}{6}}$$

Udowodnijmy teraz wynikający z powyższych obserwacji wzór na odległość punktu od podprzestrzeni wymiaru o 1 mniejszego niż cała przestrzeń afiniczna. Uogólnia on znany ze szkoły wzór na odległość punktu od prostej na płaszczyźnie.

Obserwacja 12.11

W przestrzeni euklidesowej afinicznej $H = \mathbb{R}^n$ ze standardowym iloczynem skalarnym niech M będzie podprzestrzenią opisaną równaniem $a_1x_1 + \ldots + a_nx_n = b$, gdzie $(a_1, \ldots, a_n) \neq (0, \ldots, 0)$. Wówczas odległość punktu $p = (y_1, \ldots, y_n)$ od M wynosi:

$$\rho(p, M) = \frac{|a_1y_1 + \ldots + a_ny_n - b|}{\sqrt{a_1^2 + \ldots + a_n^2}}.$$

Wzór ten uogólnia się bez zmian na dowolną przestrzeń euklidesową afiniczną $(H, \langle , \rangle)$ i jej podprzestrzeń M wymiaru $\dim H - 1$ pod warunkiem, że

$$\begin{aligned} M &= \{p_0 + x_1\alpha_1 + \ldots + x_n\alpha_n \,|\, a_1x_1 + \ldots + a_nx_n = b\}, \\ p &= \; p_0 + y_1\alpha_1 + \ldots + y_n\alpha_n \end{aligned}$$

dla pewnego ortonormalnego układu bazowego $p_0, \alpha_1, \ldots, \alpha_n$ w H.

Dowód. Niech q będzie rzutem prostopadłym punktu p na M. Wówczas, jak wiemy z wcześniejszych wykładów $\rho(p, M) = \|p - q\|$. Jak wiadomo $p - q \in T(M)^\perp$. Przestrzeń $T(M)$ jest opisana równaniem $a_1x_1 + \ldots + a_nx_n = 0$, a zatem wektor z $\mathbb{R}^n = T(H)$ należy do $T(M)$ wtedy i tylko wtedy gdy jest on prostopadły do wektora $(a_1, \ldots, a_n)$ (bo mamy standardowy iloczyn skalarny). Stąd

$$T(M) = \mathrm{lin}((a_1, \ldots, a_n))^\perp \quad \text{oraz} \quad T(M)^\perp = (\mathrm{lin}((a_1, \ldots, a_n))^\perp)^\perp = \mathrm{lin}((a_1, \ldots, a_n)).$$

Niech $q = (z_1, \ldots, z_n)$. Skoro $p - q \in T(M)^\perp$, to istnieje takie $t \in \mathbb{R}$, że $p - q = t(a_1, \ldots, a_n) = (y_1 - z_1, \ldots, y_n - z_n)$. A zatem

$$\rho(p, M) = \|p - q\| = |t|\sqrt{a_1^2 + \ldots + a_n^2}. \qquad (\spadesuit)$$

Wyliczmy teraz wartość t:

$$\begin{aligned} t(a_1^2 + \ldots + a_n^2) &= \langle\, t(a_1, \ldots, a_n), (a_1, \ldots, a_n)\,\rangle = \langle (y_1 - z_1, \ldots, y_n - z_n), (a_1, \ldots, a_n)\,\rangle = \\ &= a_1y_1 + \ldots + a_ny_n - a_1z_1 - \ldots - a_nz_n = a_1y_1 + \ldots + a_ny_n - b, \end{aligned}$$

przy czym równość $a_1z_1 + \ldots + a_nz_n = b$ wynika stąd, że $q \in M$. A więc

$$t = \frac{a_1y_1 + \ldots + a_ny_n - b}{a_1^2 + \ldots + a_n^2}.$$

Wstawiamy wyliczone $|t|$ do $(\spadesuit)$ dostajemy

$$\rho(p, M) = \left|\frac{a_1y_1 + \ldots + a_ny_n - b}{a_1^2 + \ldots + a_n^2}\right| \cdot \sqrt{a_1^2 + \ldots + a_n^2} = \frac{|a_1y_1 + \ldots + a_ny_n - b|}{\sqrt{a_1^2 + \ldots + a_n^2}}.$$

□

Definicja 12.7: Odległość podzbiorów przestrzeni afinicznej

Niech X, Y będą niepustymi podzbiorami przestrzeni afinicznej euklidesowej $(H, \langle\,,\rangle)$. Odległością podzbiorów X, Y nazywamy liczbę $\rho(X, Y) = \inf\limits_{x \in X, y \in Y} \|x - y\|$.

Ta definicja jest bardzo ogólna i dopuszcza sytuacje, gdy rozłączne podzbiory mają odległość zero. Wystarczy na prostej $\mathbb{R}$ ze standardową odległością zadaną przez wartość bezwzględną przyjąć $X = \{0\}$ oraz $Y = (0, 1)$. W przypadku, gdy X, Y są przestrzeniami afinicznymi taka sytuacja nie zachodzi.

Obserwacja 12.12

Załóżmy, że $A, B \neq \emptyset$ są podprzestrzeniami afinicznej przestrzeni euklidesowej X. Jeżeli $A \cap B = \emptyset$, to istnieje prosta $L \subseteq X$ prostopadła zarówno do A jak i do B, spełniająca

$$L \cap A \neq \emptyset, \quad L \cap B \neq \emptyset.$$

Dla dowolnych punktów $p \in L \cap A$ oraz $q \in L \cap B$ mamy $\rho(A, B) = \|\overrightarrow{pq}\|$.

Dowód. Twierdzimy, że skoro $A \cap B = \emptyset$, to $T(A) + T(B) \neq T(X)$. Inaczej bowiem biorąc dowolne punkty $a \in A$ oraz $b \in B$ mielibyśmy, że $\overrightarrow{ab} \in T(A) + T(B)$, czyli $\overrightarrow{ab} = v_a + v_b$, gdzie $v_a \in T(A)$ oraz $v_b \in T(B)$. Wtedy jednak $a + v_a = b - v_b$. Lewa strona należy do A, a prawa do B, co jest niemożliwe.

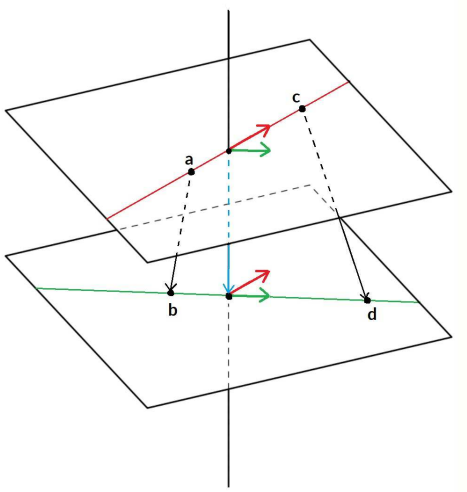
Weźmy $a \in A$ oraz $b \in B$ oraz zrzutujmy wektor $\overrightarrow{ab}$ na $(T(A)+T(B))^\perp$. Twierdzimy, że ten rzut nie zależy od wyboru a, b. Weźmy bowiem $c \in A$ oraz $d \in B$ i również zrzutujmy $\overrightarrow{cd}$ na $(T(A)+T(B))^\perp$ Dokładniej, niech $b - a = v_1 + v_2$ oraz $d - c = v_1' + v_2'$, gdzie $v_1, v_1' \in T(A) + T(B)$ oraz $v_2, v_2' \in (T(A) + T(B))^\perp$. Wówczas:

$$b - a - (d - c) = (b - d) + (c - a) \in T(A) + T(B)$$

Jednak powyższa suma równa jest $v_1 + v_2 - v_1' - v_2'$. Oznacza to, że $v_2 - v_2' \in T(A) + T(B)$, a w konsekwencji: $v_2 - v_2' = 0$, bo przecież $v_2 - v_2' \in (T(A) + T(B))^\perp$. Dostaliśmy więc $v_2 = v_2'$. Ale v_2 to rzut $\overrightarrow{ab}$ na $(T(A) + T(B))^\perp$, zaś v_2' to rzut $\overrightarrow{cd}$ na $(T(A) + T(B))^\perp$, czyli mamy to, co chcieliśmy: rzut nie zależy od wyboru $a \in A, b \in B$.

Wniosek jest taki, że jeśli $A \cap B = \emptyset$, to odległość pomiędzy podprzestrzeniami A oraz B jest długością rzutu wektora łączącego dowolny punkt z A z dowolnym punktem z B na $(T(A) + T(B))^\perp$. Dlaczego? Co z szukaną prostą? Wśród wszystkich możliwych wektorów łączących $\overrightarrow{ab}$, dla $a \in A, b \in B$, któryś jest w $(T(A)+T(B))^\perp$. Punkty, które łączy rozpinają szukaną prostą. Ich odległość to odległość A od B. □

Geometrycznie sprawa jest prosta i można ją sobie zobrazować próbując wyznaczyć odległość dwóch nieprzecinających się prostych A oraz B w przestrzeni trójwymiarowej afinicznej. Proste te mogą być równoległe, czyli $T(A) = T(B)$ lub skośne, czyli $T(A) \cap T(B) = \{0\}$.



Mamy $A = a+T(A)$ oraz $B = b+T(B)$ i te przestrzenie afiniczne możemy zawrzeć w dwóch równoległych przestrzeniach:

$$A' = a + T(A) + T(B), \quad B' = b + T(A) + T(B).$$

W przypadku prostych skośnych podprzestrzenie A' oraz B' to płaszczyzny równoległe zawierające proste A oraz B. Odległość tych płaszczyzn wyznaczamy zgodnie z dowodem twierdzenia: rzutujemy dowolny wektor łączący $a \in A$ oraz $b \in B$ na $(T(A) + T(B))^{\perp}$ i bierzemy normę uzyskanego rzutu.

Przykład. W przestrzeni euklidesowej afinicznej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym znajdź odległość prostych $L_1 = (1, 2, 3) + \text{lin}((1, 0, 0))$ oraz $L_2 = (0, 0, 0) + \text{lin}((0, 2, 4))$.

Rozwiązanie. Niech $p = (1, 2, 3)$ oraz $q = (0, 0, 0)$. Oczywiście $T(L_1) + T(L_2) = \text{lin}((1, 0, 0), (0, 2, 4))$. A zatem $(T(L_1) + T(L_2))^{\perp} = \text{lin}((0, 4, -2))$. Rozkładamy teraz wektor $\overrightarrow{pq}$ na składowe z $T(L_1) + T(L_2)$ oraz $(T(L_1) + T(L_2))^{\perp}$. Po krótkich rachunkach mamy:

$$\overrightarrow{pq} = (-1, -2, -3) = -(1, 0, 0) - \frac{4}{5}(0, 2, 4) - \frac{1}{10}(0, 4, -2).$$

W szczególności wyznaczamy długość rzutu $\overrightarrow{pq}$ na $(T(L_1) + T(L_2))^{\perp}$, czyli długość ostatniego składnika:

$$\rho(L_1, L_2) = \left| -\frac{1}{10}(0, 4, -2) \right| = \frac{\sqrt{5}}{5}.$$

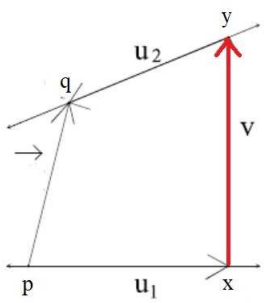
Szukając punktów x, y, dla których ta odległość jest osiągalna ustalmy, że jeśli

$$\overrightarrow{pq} = u_1 + u_2 + v,$$

gdzie $u_1 \in T(L_1), u_2 \in T(L_2), v \in (T(L_1) + T(L_2))^{\perp}$, to punkty te mają postać

$$x = p + u_1 \quad \text{oraz} \quad y = q - u_2$$

tak, by $\overrightarrow{xy} = v$, jak na rysunku poniżej.



12.1 Uzupełnienie. Dlaczego ogólny wzór na odległość działa?

Do Obserwacji 14.4 dodany został komentarz mówiący o odległości punktu od przestrzeni afinicznej wymiaru $n-1$ opisanej równaniem zawierającym współrzędne w dowolnym układzie bazowym.

Obserwacja 12.13

W przestrzeni euklidesowej afinicznej $(H, \langle\,,\rangle)$ wymiaru n niech $p_0, \alpha_1, \ldots, \alpha_n$ będzie ortonormalnym układem bazowym i niech M będzie zbiorem punktów

$$M = \{p_0 + x_1\alpha_1 + \ldots + x_n\alpha_n \,|\, a_1x_1 + \ldots + a_nx_n = b\}, \quad (a_1, \ldots, a_n) \neq (0, \ldots, 0).$$

Wówczas odległość punktu $p = p_0 + y_1\alpha_1 + \ldots + y_n\alpha_n$ od M wynosi:

$$\rho(p, M) = \frac{|a_1y_1 + \ldots + a_ny_n - b|}{\sqrt{a_1^2 + \ldots + a_n^2}}.$$

Zanim przedstawimy dowód warto podać wyjaśnienie. Co to znaczy, że zapisujemy przestrzeń afiniczną za pomocą równania w innym układzie bazowym? Jeśli na przykład weźmiemy w $\mathbb{R}^2$ prostą o równaniu $x_1 - x_2 = 1$, czyli $(1,0) + \mathrm{lin}((1,1))$, to biorąc na przykład układ bazowy $(1,0); (1,1), (1,-1)$ widzimy, że każdy punkt tej prostej zapisuje się w tym układzie bazowym w postaci $(1,0) + y_1(1,1)$, czyli każdy punkt tej prostej ma w nowym układzie bazowym współrzędne $(y_1, 0)$.

Opis podprzestrzeni afinicznych, a także ogólniej — zbiorów algebraicznych w przestrzeni afinicznej we współrzędnych w danym układzie bazowym ma kluczowe znaczenie w dalszych wykładach. Możliwość uzyskiwania „różnych równań" opisujących zbiory spełniające równania liniowe (a także wielomianowe) prowadzi do pytania: jak uzyskać „najprostsze" opisy zbiorów za pomocą równań.

Dowód. Niech q będzie rzutem prostopadłym punktu p na M. Wówczas, jak wiemy z wcześniejszych wykładów $\rho(p, M) = \|p - q\|$. Jak wiadomo $p - q \in T(M)^\perp$. Przestrzeń $T(M)$ jest złożona z wektorów postaci $x_1\alpha_1 + \ldots + x_n\alpha_n$, które spełniają równanie $a_1x_1 + \ldots + a_nx_n = 0$, bo układ $\alpha_1, \ldots, \alpha_n$ jest ortonormalny, a zatem wektor z $T(H)$ należy do $T(M)$ wtedy i tylko wtedy gdy jest on prostopadły do wektora $a_1\alpha_1 + \ldots + a_n\alpha_n$. Stąd

$$T(M)^\perp = \mathrm{lin}(a_1\alpha_1 + \ldots + a_n\alpha_n).$$

Niech $q = p_0 + z_1\alpha_1 + \ldots + z_n\alpha_n$. Skoro $p - q \in T(M)^\perp$, to istnieje takie $t \in \mathbb{R}$, że

$$p - q = ta_1\alpha_1 + \ldots + ta_n\alpha_n = (y_1 - z_1)\alpha_1 + \ldots + (y_n - z_n)\alpha_n.$$

A zatem (znowu korzystamy z ortonormalności układu $\alpha_1, \ldots, \alpha_n$):

$$\rho(p, M) = \|p - q\| = |t|\sqrt{a_1^2 + \ldots + a_n^2}. \qquad (\spadesuit)$$

Wyliczmy teraz wartość t:

$$\begin{aligned} t(a_1^2 + \ldots + a_n^2) &= \langle\, ta_1\alpha_1 + \ldots + ta_n\alpha_n, a_1\alpha_1 + \ldots + a_n\alpha_n \,\rangle = \\ &= \langle (y_1 - z_1)\alpha_1 + \ldots + (y_n - z_n)\alpha_n, a_1\alpha_1 + \ldots + a_n\alpha_n \,\rangle = \\ &= a_1y_1 + \ldots + a_ny_n - a_1z_1 - \ldots - a_nz_n = a_1y_1 + \ldots + a_ny_n - b, \end{aligned}$$

przy czym równość $a_1z_1 + \ldots + a_nz_n = b$ wynika stąd, że $q \in M$. A więc

$$t = \frac{a_1y_1 + \ldots + a_ny_n - b}{a_1^2 + \ldots + a_n^2}.$$

Wstawiamy wyliczone $|t|$ do $(\spadesuit)$ dostajemy

$$\rho(p, M) = \left|\frac{a_1y_1 + \ldots + a_ny_n - b}{a_1^2 + \ldots + a_n^2}\right| \cdot \sqrt{a_1^2 + \ldots + a_n^2} = \frac{|a_1y_1 + \ldots + a_ny_n - b|}{\sqrt{a_1^2 + \ldots + a_n^2}}.$$

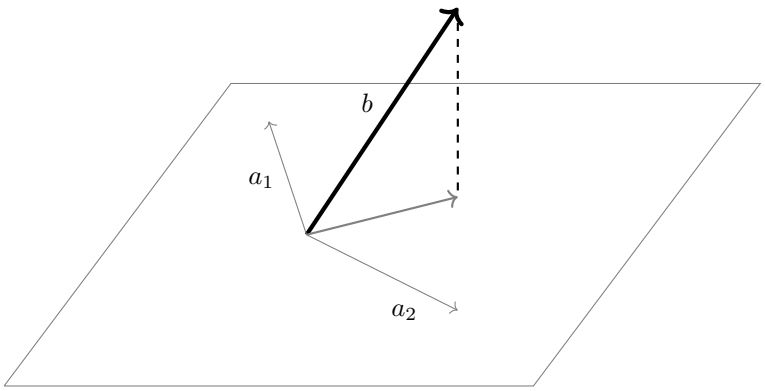
□

12.2 Dodatek. Metoda najmniejszych kwadratów

Zacznijmy od układów równań[2]. Jeśli $A \in M_{m\times n}(R)$ oraz $b \in \mathbb{R}^m$, to układ $Ax = b$ może nie mieć rozwiązania $x \in \mathbb{R}^n$. Jeśli jednak rozważamy strukturę przestrzeni euklidesowej $(\mathbb{R}^n, \langle , \rangle)$, to możemy zastanawiać się czy jest jakiś wektor, który byłby „najbliżej" rozwiązania. Rozważmy przykład:

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \iff x_1 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_2 \cdot \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Nietrudno widzieć, że równanie to nie ma rozwiązań, bo wektor $(4, 5, 6)$ nie jest kombinacją liniową wektorów $a_1 = (1, 1, 0), a_2 = (2, 3, 0)$. Zobaczmy rysunek w przestrzeni $\mathbb{R}^3$



Zaznaczona „płaszczyzna" to przestrzeń rozpięta przez kolumny a_1, a_2 macierzy naszego układu. Każdy wektor na tej płaszczyźnie jest postaci $x_1a_1 + x_2a_2$, dla pewnych $x_1, x_2 \in \mathbb{R}$. To są właśnie wektory Ax. Linią przerywaną oznaczony jest wektor opisujący $Ax - b$. Niestety może on nigdy nie być zerowy. Ale skoro jesteśmy w przestrzeni euklidesowej, to możemy próbować minimalizować wyrażenie $\|Ax - b\|$.

Przypomnijmy Obserwację z wykładu, która mówiła, że jeśli W jest podprzestrzenią przestrzeni euklidesowej V oraz v' jest rzutem prostopadłym wektora $v \in V$ na W, wówczas $\|v - v'\| = \min_{w\in W} \|v - w\|$.

W naszym przypadku podprzestrzeń $W = \{Ax, \text{ gdzie } x \in \mathbb{R}^3\}$ jest przestrzenią kolumnową macierzy A. Zatem wektor $\hat{x}$, który minimalizuje $\|Ax - b\|$ jest rzutem wektora b na podprzestrzeń kolumnową A.

Jak wyznaczyć ten rzut? Popatrzmy na tzw. wektor błędu $b - A\hat{x}$. Wektor ten musi być prostopadły do każdej kolumny $a_1, \ldots, a_n$ macierzy A, co możemy zapisać schematycznie w postaci:

$$\begin{cases} \langle\, a_1^T, b - A\hat{x} \,\rangle & = 0 \\ & \vdots \\ \langle\, a_n^T, b - A\hat{x} \,\rangle & = 0 \end{cases},$$

co w przypadku standardowego iloczynu skalarnego przepisać można do postaci $A^T(b - A\hat{x}) = 0$. Innymi słowy, wektor $\hat{x}$ jest w tym przypadku rozwiązaniem tzw. równania normalnego $A^TA\hat{x} = A^Tb$, powstającego przez przemnożenie wyjściowego równania przez A^T. Zachodzi następujące twierdzenie.

Obserwacja 12.14

Jeśli równanie $Ax = b$ nie ma rozwiązań w przestrzeni euklidesowej $(\mathbb{R}^n, \langle , \rangle_{st})$, to $\|Ax - b\|$ jest minimalna dla $x = \hat{x}$, będącego rozwiązaniem równania $A^TA\hat{x} = A^Tb$. Macierz A^TA jest odwracalna wtedy i tylko wtedy, gdy kolumny A są liniowo niezależne. Wówczas:

$$\hat{x} = (A^TA)^{-1}A^Tb.$$

Zachęcam Czytelnika do dowodu tego faktu. Proszę przy okazji pokazać, że jednorodne układy równań liniowych zadane macierzami A oraz A^TA mają ten sam zbiór rozwiązań.

[2]Ten dodatek napisany jest na podstawie znakomitej książki Gilberta Stranga „Linear algebra and its applications".

Przejdźmy na koniec do naszego przykładu. Rzut prostopadły (standardowy) wektora $b = (4, 5, 6)$ na przestrzeń kolumnową macierzy o kolumnach $(1, 1, 0)^T$ oraz $(2, 3, 0)^T$ to po prostu rzut prostopadły na przestrzeń $\mathrm{lin}((1, 0, 0), (0, 1, 0))$, i jest to oczywiście $(4, 5, 0)$. Potwierdza się to przy wyliczeniach proponowanych przez fakt wyżej:

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

Co to wszystko ma wspólnego z zastosowaniami matematyki? Załóżmy, że wykonaliśmy serię eksperymentów i oczekujemy, że wynik b będzie funkcją liniową argumentu t. Szukamy zatem prostej o równaniu $C + Dt = b$. Przykłady:

- Wyznaczamy odległość satelity od Marsa. W tym przypadku t jest zmienną czasową, a b – dystansem od planety. O ile nie wyłączy się silnika i o ile grawitacja nie będzie bardzo silna, satelita powinien poruszać się z mniej więcej stałą prędkością v, czyli $b = b_0 + vt$.
- W teorii elastyczności badamy rozszerzenie b struny obciążonej określoną masą t. O ile obciążenie nie jest nadmierne, rozważa się liniową zależność $b = C + Dt$.
- Koszt produkcji t książek jest niemal liniową funkcją $b = C + Dt$, gdzie C jest kosztem przygotowania, a D jest kosztem każdej dodatkowej książki

Jak wyznaczyć C i D? Jeśli nie ma błędu, wówczas każde dwa pomiary b powinny określać prostą $b = C + Dt$. Ale jeśli jest błąd, wówczas musimy być gotowi na znajdowanie „optymalnej" prostej przybliżającej te rozwiązania. Wykonujemy te pomiary dostając układ równań:

$$\begin{cases} C + Dt_1 &= b_1 \\ C + Dt_2 &= b_2 \\ & \vdots \\ C + Dt_m &= b_m. \end{cases}.$$

Ten układ zawiera nadmiarowe informacje: mamy m równań i tylko dwie niewiadome C, D. Jeśli wystąpiły błędy, układ ten nie ma rozwiązania. Jeśli wystąpiły, to najlepsze rozwiązanie $\hat{x} = (\hat{C}, \hat{D})^T$, przybliżające błąd (w normie $\|, \|_{st}$) ma minimalizować wyrażenie:

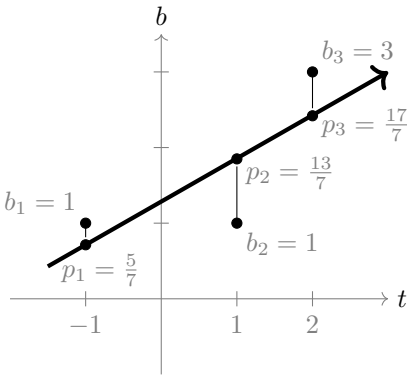
$$\|Ax - b\| = (b_1 - C - Dt_1)^2 + \ldots + (b_m - C - Dt_m)^2.$$

Wektor $p = (p_1, p_2, p_3) = A\hat{x}$ jest „tak blisko" $b = (b_1, b_2, b_3)$, jak to możliwe, tzn. suma kwadratów różnic $p_i - b_i$ jest minimalna.

Dla przykładu: biorąc $b_1 = 1, t_1 = 1$, $b_2 = 1, t_2 = 1, b_3 = 3, t_3 = 2$ widzimy, że punkty (b_i, t_i) nie leżą na jednej prostej. Istnieje jednak prosta $\hat{C} + \hat{D}t = \frac{9}{7} + \frac{4}{7}t$, która „najlepiej" przybliża rozwiązania układu

$$\begin{cases} C - D &= 1 \\ C + D &= 1 \\ C + 2D &= 3 \end{cases}.$$

Rysunek poniżej przedstawia uzyskaną prostą i „punkty pomiarowe".



12.3 Trivia. Czy względna pierwszość to prostopadłość?

Jak wiemy każda dodatnia liczba naturalna n może być jednoznacznie przedstawiona jako iloczyn

$$n = \prod_p p^{v_p(n)},$$

gdzie v_p są nieujemnymi liczbami naturalnymi. Wspominaliśmy o tym choćby w podrozdziale o liczbach p-adycznych. Liczby $v_p(n)$ nazywamy wykładnikami p-adycznymi liczby n. Są takie liczby całkowite k, że $p^k \mid n$ oraz $p^{k+1} \nmid n$. Wprawdzie powyższy iloczyn jest nieskończony, ale jedynie skończenie wiele z jego czynników jest różnych od 1. Przypomnijmy kilka elementarnych własności wykładnika.

Obserwacja 12.15

Niech p będzie liczbą pierwszą, zaś a, b niech będą dodatnimi liczbami całkowitymi.

- $v_p(ab) = v_p(a) + v_p(b)$ oraz $v_p(a/b) = v_p(a) - v_p(b)$,
- $v_p(a^n) = nv_p(a)$,
- $v_p(\mathrm{NWD}(a,b)) = \min\{v_p(a), v_p(b)\}$ oraz $v_p(\mathrm{NWW}(a,b)) = \max\{v_p(a), v_p(b)\}$,
- $v_p(a \pm b) \geqslant \min\{v_p(a), v_p(b)\}$. Gdy $v_p(a) \neq v_p(b)$, wówczas mamy równość.

Co więcej następujące warunki są równoważne.

- $x \mid y$,
- dla każdej liczby pierwszej p zachodzi nierówność $v_p(x) \leqslant v_p(y)$

W szczególności $x = y$ wtedy i tylko wtedy, gdy $v_p(x) = v_p(y)$, dla każdej liczby pierwszej p.

Przykładowe proste zastosowania powyższego faktu.

Zadanie 12.1. *Niech p będzie liczbą pierwszą. Wykaż, że równanie $x^2 - py^2 = 0$ nie ma rozwiązań w liczbach całkowitych.*

Rozwiązanie. Aby zachodziła równość $x^2 = py^2$ konieczne jest, aby $v_p(x^2) = v_p(py^2)$. Jedna strona tej równości jest liczbą parzystą $2v_p(x)$, druga zaś – liczbą nieparzystą $2v_p(y) + 1$, co jest niemożliwe. ■

Zadanie 12.2. *Dla dodatnich liczb całkowitych a, b mamy* $\mathrm{NWD}(a,b) \cdot \mathrm{NWW}(a,b) = a \cdot b$.

Rozwiązanie. Policzmy wartość wyrażenia $v_p(\mathrm{NWD}(a,b) \cdot \mathrm{NWW}(a,b)) - v_p(ab)$. Zgodnie z własnościami podanymi wyżej wyrażenie to jest równe $\min\{v_p(a), v_p(b)\} + \max\{v_p(a), v_p(b)\} - v_p(a) - v_p(b)$. Jest jasne, że uzyskane wyrażenie jest zawsze równe 0. ■

Na potrzeby tego dodatku stosować będziemy notację przyjętą przez Ronalda L. Grahama, Donalda D. Knutha i Orena Patashnika w książce *Matematyka Konkretna*, oznaczając liczby względnie pierwsze symbolem prostopadłości $\perp$, tzn. $n \perp m \iff \mathrm{NWD}(n,m) = 1$. Jaka jest motywacja tego podejścia?

Podstawowe własności wektorów prostopadłych można podsumować w następującym wyliczeniu. Niech v, w będą wektorami w przestrzeni rzeczywistej. Wówczas:

- dla każdego $\overrightarrow{v}$ istnieje $\overrightarrow{w}$, że $\overrightarrow{v} \perp \overrightarrow{w}$,
- dla każdych $\overrightarrow{v}, \overrightarrow{w}$ mamy $\overrightarrow{v} \perp \overrightarrow{w} \Rightarrow \overrightarrow{w} \perp \overrightarrow{v}$,
- dla każdego $\overrightarrow{v} \neq 0$ mamy $\overrightarrow{v} \not\perp \overrightarrow{v}$,
- dla każdych $\overrightarrow{v}, \overrightarrow{w}, \overrightarrow{z}$ mamy:

$$\begin{cases} \overrightarrow{v} \perp \overrightarrow{w} \\ \overrightarrow{v} \perp \overrightarrow{z} \end{cases} \quad \Rightarrow \quad \overrightarrow{v} \perp (\overrightarrow{w} + \overrightarrow{z})?$$

Czy te własności spełnione są przez pary liczb względnie pierwszych? Pierwsze dwie – oczywiście tak. Każda dodatnia liczba całkowita ma choć jedną względnie pierwszą z nią (a nawet nieskończenie wiele). Relacja względnej pierwszości jest symetryczna. A pozostałe dwa?

Warunek trzeci może budzić wątpliwość, bo w ogóle nie rozważamy na razie niezerowych liczb. Ma on jednak sens o tyle, że jedyną dodatnią liczbą całkowitą względnie pierwszą ze sobą jest 1 — ta liczba gra więc rolę zera.

A warunek czwarty? Odpowiedź jest pozytywna, jeśli działanie $+$ jest mnożeniem! Jak to rozumieć?

Rozważamy podzbiór przestrzeni ciągów $\mathbb{R}^\infty$ złożony z ciągów postaci:

$$n \mapsto v(n) = (v_2(n), v_3(n), v_5(n), v_7(n), v_{11}(n), \ldots),$$

Zobaczmy przykłady:

$$80 = 2^4 \cdot 3^0 \cdot 5^1 \cdot 7^0 \cdot \ldots \rightsquigarrow (4, 0, 1, 0, 0, \ldots)$$
$$189 = 2^0 \cdot 3^3 \cdot 5^0 \cdot 7^1 \cdot \ldots \rightsquigarrow (0, 3, 0, 1, 0, \ldots)$$

W tej konfiguracji oczywiście liczba 1 przechodzi na wektor zerowy:

$$1 = 2^0 \cdot 3^0 \cdot 5^0 \cdot 7^0 \cdot \ldots \rightsquigarrow (0, 0, 0, 0, 0, \ldots).$$

Jasna jest również interpretacja liczb pierwszych – są to kolejne „wersory”, np.

$$5 = 2^0 \cdot 3^0 \cdot 5^1 \cdot 7^0 \cdot \ldots \rightsquigarrow (0, 0, 1, 0, 0, \ldots)$$

Zauważmy, że w istocie można rozszerzyć powyższe przypisanie na dowolne dodatnie liczby wymierne, przyjmując $v(p/q) = v(p) - v(q)$, na przykład:

$$80/27 = 2^4 \cdot 3^{-3} \cdot 5^1 \cdot 7^0 \cdot \ldots \rightsquigarrow (4, -3, 1, 0, 0, \ldots)$$
$$189/25 = 2^0 \cdot 3^3 \cdot 5^{-2} \cdot 7^1 \cdot \ldots \rightsquigarrow (0, 3, -2, 1, 0, \ldots)$$
$$3/7 = 2^0 \cdot 3^1 \cdot 5^0 \cdot 7^{-1} \cdot \ldots \rightsquigarrow (0, 1, 0, -1, 0, \ldots)$$

Gdybyśmy mogli tu używać języka teorii grup, powiedzielibyśmy, że mamy izomorfizm grupy multyplikatywnej liczb wymiernych z grupą addytywną wielomianów o współczynnikach całkowitych. Oczywiście nie otrzymujemy podprzestrzeni liniowej (ale otrzymujemy pewien podmoduł...). Zauważmy, że przyjmując w wymienionych wyżej ciągach naturalne działania dodawania wektorów i mnożenia przez skalar widzimy, że

$$v(nm) = v(n) + v(m), \quad v(n^k) = k \cdot v(n).$$

Oto ilustracja dla iloczynu $80 \cdot 189$ oraz dla liczby $3920^3 = 2^{12} \cdot 5^3 \cdot 7^6$:

$$(4, 0, 1, 0, 0, \ldots) + (0, 3, 0, 1, 0, \ldots) = (4, 3, 1, 1, 0, \ldots), \quad 3 \cdot (4, 0, 1, 2, 0, \ldots) = (12, 0, 3, 6, 0, \ldots).$$

We wprowadzonej konwencji rozważać można również kombinacje liniowe wektorów:

$$\overrightarrow{a} = 3, \quad \overrightarrow{b} = 10, \quad \overrightarrow{a} + \overrightarrow{2b} = 3 \cdot 10^2 = 300.$$

Jest wreszcie jasne, że względna pierwszość liczb całkowitych odpowiada prostopadłości, przy przyjeciu „standardowego iloczynu skalarnego” w $\mathbb{R}^\infty$, np.

$$\langle 80, 189 \rangle = 4 \cdot 0 + 0 \cdot 3 + 1 \cdot 0 + 0 \cdot 1 + \ldots = 0.$$

Podobnie zdefiniować można normę oraz kąt, np.

$$\|432\| = \sqrt{\langle 2^4 \cdot 3^3, 2^4 \cdot 3^3 \rangle} = \sqrt{4^2 + 3^2} = 5, \qquad \cos \measuredangle(6, 10) = \frac{\langle 6, 10 \rangle}{\|6\| \cdot \|10\|} = \frac{1}{\sqrt{2} \cdot \sqrt{2}} = \frac{1}{2}.$$

Można więc powiedzieć, że pary liczb względnie pierwszych można traktować w tej konwencji jako liczby prostopadłe, a potęgi tej samej liczby całkowitej – jako liczby równoległe. Zachęcam Czytelnika do przemyślenia dlaczego w ramach tej konwencji „trójkąt” o wierzchołkach 6, 10, 15 jest „równoboczny”.

Czy to wszystko przekłada się na jakiekolwiek zastosowania? Okazuje się, że tak, gdy rozważamy powyższe wektory jako elementy $\mathbb{Z}_2^\infty$ (co wtedy oznacza prostopadłość?), ale to już inna historia...

Rozdział 13

Relacja kongruencji macierzy. Kryteria Sylvestera i Jacobiego

W rozdziale 11 omówione zostały podstawowe przykłady form dwuliniowych, w tym iloczyny skalarne. Wprowadziliśmy też podstawowe intuicje geometryczne, związane z pojęciem prostopadłości. Mamy dostatecznie dużo motywacji, aby z każdą formą dwuliniową związać pewne macierze, tzw. macierze formy dwuliniowej. Ich podstawowa postać obecna była w pierwszych przykładach form na przestrzeni K^n. Macierze tej samej formy w różnych bazach nazwiemy kongruentnymi. W języku wyznaczników związanych z tymi macierzami sformułujemy kryterium wyróżniające iloczyny skalarne spośród symetrycznych funkcji dwuliniowych $h : V \times V \to \mathbb{R}$, nazwane imieniem Jacoba Sylvestera, a także jego uogólnienie.

Definicja 13.1

Niech $h : V \times V \to K$ będzie formą dwuliniową na przestrzeni V i niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ będzie bazą przestrzeni V. MACIERZĄ FORMY h W BAZIE $\mathcal{A}$ nazywamy macierz:

$$G(h; \mathcal{A}) = [h(\alpha_i, \alpha_j)] \in M_n(K)$$

Rozważmy dwa przykłady

- Niech $h : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ dana będzie wzorem $h((x_1, x_2, x_3), (y_1, y_2, y_3)) = x_1y_1 + x_2y_2 - x_3y_3$. Wówczas:
$$G(h; st) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$
- Niech $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ dana będzie wzorem $h((x_1, x_2), (y_1, y_2)) = x_1y_1 - 2x_1y_2 + 3x_2y_1 + 5x_2y_2$. Wówczas jeśli $\mathcal{A} = ((1, 1), (0, 4))$, to
$$G(h; st) = \begin{bmatrix} 1 & -2 \\ 3 & 5 \end{bmatrix}, \quad G(h; \mathcal{A}) = \begin{bmatrix} 7 & 12 \\ 32 & 80 \end{bmatrix}.$$

Sformułujmy jako ćwiczenie oczywistą formułę wynikającą z definicji macierzy formy.

Obserwacja 13.1

Jeśli $h : V \times V \to K$ jest formą dwuliniową, zaś $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ jest bazą przestrzeni V, to dla dowolnych wektorów $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$, oraz $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ przestrzeni V zachodzi:

$$h(\alpha, \beta) = \sum_{1 \leqslant i,j \leqslant n} h(\alpha_i, \alpha_j) \cdot x_i y_i = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \cdot G(h; \mathcal{A}) \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Podstawowy wniosek jest zupełnie elementarny, ale stanowi ważną motywację geometryczną.

Wniosek 13.1

Jeśli $h : V \times V \to K$ jest formą dwuliniową, zaś $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ jest bazą przestrzeni V, to następujące warunki są równoważne:

- baza $(\alpha_1, \ldots, \alpha_n)$ jest ortogonalna,
- macierz $G(h; \mathcal{A})$ jest diagonalna.

Póki co istnienie baz ortogonalnych (oraz ortonormalnych) wykazaliśmy jedynie dla iloczynów skalarnych. Odnotujmy kluczowy wniosek.

Wniosek 13.2

Niech h będzie formą dwuliniową na przestrzeni skończenie wymiarowej V nad ciałem $\mathbb{R}$. Wówczas następujące warunki są równoważne:

(i) dla pewnej bazy $\mathcal{A}$ przestrzeni V mamy $G(h; \mathcal{A}) = I$,

(ii) h jest iloczynem skalarnym na V.

Dowód. Implikacja (ii) $\Rightarrow$ (i) jest oczywista — wystarczy przyjąć, że $\mathcal{A}$ jest bazą ortonormalną (V, h). W drugą stronę — aby pokazać, że h jest iloczynem skalarnym, należy uzasadnić jej symetryczność i dodatnią określoność. Symetria jest oczywista, bo dla dowolnych wektorów $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$, $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ przestrzeni V mamy

$$h(\alpha, \beta) = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \cdot I \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{1 \leqslant i \leqslant n} x_i y_i = \begin{bmatrix} y_1 & \ldots & y_n \end{bmatrix} \cdot I \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = h(\beta, \alpha).$$

Dodatnia określoność również jest jasna. Dla dowolnego niezerowego wektora $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ mamy

$$h(\alpha, \alpha) = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \cdot I \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{1 \leqslant i \leqslant n} x_i^2.$$

Ostatnia suma jest większa od 0, bowiem co najmniej jeden z wyrazów x_i jest niezerowy. □

Intuicja powyższego faktu jest następująca: wszystkie iloczyny skalarne na przestrzeni wymiaru n nad ciałem $\mathbb{R}$ działają „w ten sam sposób" jak iloczyn standardowy, oczywiście w pewnej bazie. Innymi słowy, jeśli $\alpha_1, \ldots, \alpha_n$ jest bazą ortonormalną przestrzeni euklidesowej $(V, \langle , \rangle)$, to dla dowolnych wektorów $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n, \beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ przestrzeni V zachodzi:

$$\langle \alpha, \beta \rangle = x_1 y_1 + \ldots + x_n y_n.$$

Widzimy już jak ważnym wątkiem naszych rozważań będzie badanie macierzy form dwuliniowych w różnych bazach. Odnotujmy podstawową uwagę.

Obserwacja 13.2

Niech $h : V \times V \to K$ będzie formą dwuliniową oraz niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$, $\mathcal{B} = (\beta_1, \ldots, \beta_n)$ będą bazami przestrzeni V. Jeśli $A = G(h; \mathcal{A}), B = G(h; \mathcal{B})$ oraz $C = M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}}$, to

$$B = C^T A C.$$

Dowód. Niech $C = [c_{ij}]$. Opiszmy wyraz z i-tego wiersza i j-tej kolumny macierzy $C^T AC$. Nietrudno widzieć (bo tak jest dla dowolnego iloczynu trzech macierzy), że wyraz ten powstaje przez przemnożenie i-tego wiersza macierzy C^T, macierzy A oraz j-tej kolumny macierzy C. Mnożymy więc w rezultacie współrzędne i-tego oraz j-tego elementu bazy $\mathcal{B}$ zapisanych w bazie $\mathcal{A}$ przez macierz formy h w bazie $\mathcal{A}$. A zatem zgodnie z Obserwacją 13.1 wyraz ten wynosi $h(\beta_i, \beta_j)$. Dokładnie tej samej postaci jest także z definicji wyraz w i-tym wierszu i j-tej kolumnie macierzy $B = G(h; \mathcal{B})$. □

Uwaga ta motywuje wprowadzenie następującego fundamentalnego dla algebry liniowej pojęcia.

Definicja 13.2

Mówimy, że macierze $A, B \in M_n(K)$ są KONGRUENTNE NAD K jeśli istnieje macierz odwracalna $C \in M_n(K)$ taka, że

$$B = C^T AC.$$

Obserwacja 13.3

Macierze $A, B \in M_n(K)$ są kongruentne wtedy i tylko wtedy, gdy są macierzami tej samej formy dwuliniowej (w pewnych bazach).

Dowód. Jeśli istnieje forma dwuliniowa $h : V \times V \to K$ na n wymiarowej przestrzeni liniowej V nad K oraz bazy $\mathcal{A}, \mathcal{B}$ przestrzeni V takie, że $A = G(h, \mathcal{A})$, $B = G(h, \mathcal{B})$, to macierze A, B są kongruentne na mocy uwagi wyżej. Na odwrót, niech $B = C^T AC$, dla pewnej macierzy odwracalnej $C \in M_n(K)$. Niech $A = [a_{ij}]$. Określmy formę dwuliniową $h : K^n \times K^n \to K$ warunkiem $G(h; st) = A$. To znaczy:

$$h((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sum_{1 \leqslant i,j \leqslant n} a_{ij} x_i y_j.$$

Niech $B = (\beta_1, \ldots, \beta_n)$ będzie bazą przestrzeni K^n zadaną przez $M(\mathrm{id})^{st}_{\mathcal{B}} = C$. Wówczas z dowodu poprzedniej uwagi wynika natychmiast, że $B = G(h; \mathcal{B})$. □

Dla (rozważanej wcześniej) formy $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ zadanej wzorem $h((x_1, x_2), (y_1, y_2)) = x_1 y_1 - 2x_1 y_2 + 3x_2 y_1 + 5x_2 y_2$ wyliczyliśmy

$$G(h; st) = \begin{bmatrix} 1 & -2 \\ 3 & 5 \end{bmatrix}, \quad G(h; \mathcal{A}) = \begin{bmatrix} 7 & 12 \\ 32 & 80 \end{bmatrix},$$

więc macierze te są kongruentne nad $\mathbb{R}$. Przestrzeń $(\mathbb{R}^2, h)$ nie jest jednak euklidesowa, bo $G(h; st)$ nie jest nawet symetryczna.

Kongruentność, podobnie jak podobieństwo, jest relacją równoważności w zbiorze $M_n(K)$. Rzeczywiście:

- zwrotność relacji kongruencji wynika z faktu, że $A = I^T AI$,
- symetryczność relacji kongruencji wynika z tego, że jeśli $A = C^T BC$, dla pewnej macierzy odwracalnej C, to $B = (C^T)^{-1} AC^{-1} = (C^{-1})^T AC^{-1}$,
- przechodniość relacji kongruencji wynika z tego, że jeśli $A = X^T BX, B = Y^T CY$, to

$$A = X^T Y^T CYX = (YX)^T C(YX).$$

Nietrudno widzieć, że również po ograniczeniu do macierzy symetrycznych kongruencja jest relacją równoważności. Stwierdzenie kiedy macierze są kongruentne może być bardzo trudnym zadaniem. Jednym z celów całego kursu jest dokonać takiej klasyfikacji dla macierzy symetrycznych nad ciałami liczb rzeczywistych i zespolonych. Sformułujmy teraz kilka warunków koniecznych, aby macierze były kongruentne.

Obserwacja 13.4

Jeśli macierze $A, B \in M_n(K)$ są kongruentne nad K, to $r(A) = r(B)$ oraz $\det(A) \cdot \det(B)$ jest kwadratem w ciele K.

Dowód. Skoro istnieje macierz odwracalna C taka, że $B = C^T AC$, to $r(B) = r(C^T AC) = r(A)$, bo mnożenie przez macierz odwracalną (z dowolnej strony) nie zmienia rzędu. Mamy również:

$$\det(A) \cdot \det(B) = \det(A) \cdot \det(C^T AC) = (\det A \cdot \det C)^2.$$

□

Intuicyjnie, problem badania kongruencji macierzy nad K jest tym trudniejszy, im więcej elementów ciała K nie można utożsamić za pomocą kwadratu. Stąd na przykład klasyfikacja macierzy kongruentnych nad ciałem $\mathbb{Q}$ jest bardzo skomplikowana, w porównaniu do ciał $\mathbb{C}$ oraz $\mathbb{R}$. Już dla macierzy rozmiaru 1×1 klas kongruencji jest nieskończenie wiele, bowiem dla dowolnych dwóch różnych liczb pierwszych p, q liczba pq^{-1} jest kwadratem w $\mathbb{Q}$ wtedy i tylko wtedy, gdy $p = q$.

Za kilka wykładów wykażemy bez trudu, że każda skończenie wymiarowa przestrzeń liniowa z symetryczną formą dwuliniową ma bazę prostopadłą, a więc i macierz w formie diagonalnej. Badanie macierzy kongruentnych sprowadzi się zatem do problemu kongruencji macierzy diagonalnych. Jak się okaże, nad ciałami $\mathbb{R}$ oraz $\mathbb{C}$ liczba klas kongruencji w macierzach rozmiaru n jest skończona.

Póki co postawimy sobie jednak inne cele – bliższe geometrii. Związane one będą, mówiąc ogólnie, z wyznacznikami macierzy formy i pewnych ich podmacierzy. Dziś sformułujemy w języku wyznaczników kryterium dodatniej określoności symetrycznej formy dwuliniowej oraz jego ważne uogólnienie. Na kolejnym wykładzie przyjrzymy się natomiast geometrycznym zastosowaniom wyznacznika macierzy układu wektorów iloczynu skalarnego (nieco ogólniejsze pojęcie) i związanym z nim pojeciem miary n-wymiarowej.

Wniosek 13.3

Jeśli h jest iloczynem skalarnym na przestrzeni liniowej V ciałem $\mathbb{R}$ oraz $\mathcal{A}$ jest dowolną bazą V, to

$$\det G(h; \mathcal{A}) > 0.$$

Dowód. Rzeczywiście, macierz $G(h; \mathcal{A})$ jest kongruentna do macierzy identycznościowej. Innymi słowy, istnieje macierz odwracalna $C \in M_n(\mathbb{R})$, że:

$$G(h; \mathcal{A}) = C^T \cdot I \cdot C.$$

Wynika stąd, że $\det G(h; \mathcal{A}) = (\det C)^2 > 0$, bo C jest macierzą odwracalną. □

Definicja 13.3: Główne minory wiodące macierzy kwadratowej

Dla dowolnej macierzy $A \in M_n(K)$ oraz dla $i = 1, \ldots, n$ niech $A^{(i)} \in M_i(K)$ oznacza macierz powstałą z A przez usunięcie ostatnich $n - i$ wierszy i ostatnich $n - i$ kolumn. Wyznacznik tej macierzy nazywamy WIODĄCYM MINOREM GŁÓWNYM STOPNIA i macierzy A.

Przykład dla macierzy symetrycznej $A \in M_3(\mathbb{R})$ w postaci

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

mamy 3 główne minory wiodące, równe $\det A^{(1)} = 4, \det A^{(2)} = 1, \det A^{(3)} = 3$, będące wyznacznikami macierzy:

$$A^{(1)} = \begin{bmatrix} 4 \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}, \quad A^{(3)} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Twierdzenie 13.1: Kryterium Sylvestera

Niech h będzie symetryczną formą dwuliniową na n-wymiarowej przestrzeni liniowej V nad ciałem $\mathbb{R}$ oraz niech $A = G(h; \mathcal{A})$ będzie macierzą tej formy w bazie $\mathcal{A}$. Następujące warunki są równoważne:

(i) h jest iloczynem skalarnym na przestrzeni V,

(ii) $\det A^{(i)} > 0$, dla $i = 1, \ldots, n$.

Przykład: niech $h : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, gdzie

$$h((x_1, x_2, x_3), (y_1, y_2, y_3)) = (4-r)x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_2y_2 + x_2y_3 + x_3y_1 + x_3y_2 + rx_3y_3.$$

- Dla jakich $r \in \mathbb{R}$ funkcja h jest iloczynem skalarnym na $\mathbb{R}^3$?
- Dla jakich $r \in \mathbb{R}$ funkcja $h|_W$ jest iloczynem skalarnym na przestrzeni $W = \mathrm{lin}((1,1,-1),(0,1,1))$?

Rozwiązanie. Funkcja h jest dwuliniowa i symetryczna, a zatem możemy zastosować kryterium Sylvestera. Bierzemy bazę standardową $(\epsilon_1, \epsilon_2, \epsilon_3)$ przestrzeni $\mathbb{R}^3$ i macierz o wyrazach $h(\epsilon_i, \epsilon_j)$ postaci:

$$\begin{bmatrix} 4-r & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & r \end{bmatrix}.$$

Aby macierz ta była macierzą iloczynu skalarnego potrzeba i wystarcza, aby:

$$4 - r > 0, \quad \begin{vmatrix} 4-r & 1 \\ 1 & 1 \end{vmatrix} > 0,$$

oraz

$$\begin{vmatrix} 4-r & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & r \end{vmatrix} > 0.$$

A zatem dostajemy układ warunków:

$$4 > r, \quad 3 > r, \quad -r^2 + 4r - 3 = -(r-1)(r-3) > 0.$$

Stąd h jest iloczynem skalarnym na $\mathbb{R}^3$ wtedy i tylko wtedy, gdy $r \in (1,3)$.

Odpowiedzmy na drugie pytanie. Bierzemy

$$\alpha_1 = (1,1,-1), \quad \alpha_2 = (0,1,1).$$

Wyznaczamy macierz o wyrazach $h(\alpha_i, \alpha_j)$. Ma ona postać:

$$\begin{bmatrix} 3 & 3-r \\ 3-r & 3+r \end{bmatrix}.$$

A zatem jest ona macierzą iloczynu skalarnego wtedy i tylko wtedy, gdy $3(3+r) - (3-r)^2 > 0$, co daje $-r^2 + 9r > 0$, czyli $r(9-r) > 0$. A zatem w tym przypadku $r \in (0,9)$.

Dowód. Dowodzimy (i) $\Rightarrow$ (ii). Jeśli h jest iloczynem skalarnym na V, to oczywiście A jest macierzą formy h w bazie $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$. Ograniczenie $h|_{W_i}$ do podprzestrzeni $W_i = \mathrm{lin}(\alpha_1, \ldots, \alpha_i)$ jest iloczynem skalarnym na przestrzeni W_i, dla każdego $i = 1, \ldots, n$. A zatem macierze $A^{(i)} = G(\alpha_1, \ldots, \alpha_i)$ mają dodatnie wyznaczniki, na mocy Wniosku 13.3.

Zakładamy, że dla bazy $\mathcal{A}$ macierz $G(h; \mathcal{A})$ ma dodatnie główne minory wiodące. Rozważmy postać blokową powyższej macierzy postaci:

$$G(h; \mathcal{A}) = \begin{bmatrix} G(h; \mathcal{A})^{(n-1)} & \alpha \\ \alpha^T & b \end{bmatrix}, \text{ gdzie } \alpha \in \mathbb{R}^{n-1}, b \in \mathbb{R}.$$

Główne minory wiodące macierzy $G(h; \mathcal{A})^{(n-1)}$ są głównymi minorami wiodącymi macierzy $G(h; \mathcal{A})$. Zatem z założenia indukcyjnego forma $g = h|_{W_{n-1}}$ jest iloczynem skalarnym na podprzestrzeni $W_{n-1} = \mathrm{lin}(\alpha_1, \ldots, \alpha_{n-1})$.

Jeśli $\mathcal{C} = (\beta_1, \ldots \beta_{n-1})$ jest bazą W_{n-1}, to w bazie $\mathcal{B} = (\beta_1, \ldots \beta_{n-1}, \alpha_n)$ przestrzeni V macierz $G(h; \mathcal{B})$ ma postać:

$$G(h; \mathcal{B}) = \begin{bmatrix} G(g; \mathcal{C}) & \gamma \\ \gamma^T & b \end{bmatrix}, \text{ gdzie } \gamma \in \mathbb{R}^{n-1}$$

i skoro $G(h;\mathcal{B}) = C^T G(h;\mathcal{A})C$, dla pewnej macierzy odwracalnej C, to wobec założenia $\det G(h;\mathcal{A}) > 0$ (to jeden z minorów) mamy $\det G(h;\mathcal{B}) > 0$. Przyjmijmy dalej, że baza $\mathcal{C}$ jest ortonormalna.

Wobec tego macierz formy h w bazie $(\beta_1, \ldots, \beta_{n-1}, \alpha_n)$ ma postać

$$G(h;\mathcal{B}) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & a_1 \\ 0 & 1 & 0 & \ldots & 0 & a_2 \\ 0 & 0 & 1 & \ldots & 0 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & a_{n-1} \\ a_1 & a_2 & a_3 & \ldots & a_{n-1} & b \end{bmatrix}.$$

Niech P^T będzie macierzą powstałą przez zamianę ostatniego wiersza w macierzy identycznościowej $I_n \in M_n(\mathbb{R})$ na wiersz

$$\begin{bmatrix} -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} & 1 \end{bmatrix},$$

czyli macierzą postaci:

$$P^T = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} & 1 \end{bmatrix}.$$

Łatwo sprawdzić patrząc jakie operacje elementarne związane są z mnożeniem przez P oraz P^T, że iloczyn $P^T X$ powstaje z X przez odejmowanie od ostatniego wiersza wiersza wiersza i-tego pomnożonego przez $-a_i$, dla $1 \leqslant i \leqslant n-1$ i podobnie macierz XP powstaje z X przez odejmowanie od ostatniej kolumny kolumny j-tej pomnożonej przez $-a_j$, dla $1 \leqslant j \leqslant n-1$.

Wynika stąd, że macierz $P^T G(h;\mathcal{B})P$ ma postać:

$$\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \\ -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & a_1 \\ 0 & 1 & 0 & \ldots & 0 & a_2 \\ 0 & 0 & 1 & \ldots & 0 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & a_{n-1} \\ a_1 & a_2 & a_3 & \ldots & a_{n-1} & b \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ 0 & 0 & 1 & \ldots & 0 & -a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -a_{n-1} \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix},$$

czyli istnieje $d \in \mathbb{R}$, że

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

Skoro zaś uzyskana macierz jest to kolejna macierz formy h (znowu dokonaliśmy zmiany bazy), to musi mieć dodatni wyznacznik (bo największy minor jest dodatni). Zatem $d > 0$ i h jest w sposób oczywisty iloczynem skalarnym. □

W dalszych wykładach skupimy się na geometrycznych aspektach wyznaczników macierzy układów wektorów na przestrzeni z iloczynem skalarnym — tak zwanej macierzy Grama. Poprowadzą nas one w kierunku fundamentalnych i niezwykle ważnych dla zastosowań (choćby analitycznych) pojęć geometrycznych: miary, orientacji i iloczynu wektorowego. Poniżej przestawimy natomiast kilka uwag dotyczących zastosowań wyznaczników do badania kongruencji macierzy nad dowolnym ciałem.

* * *

Wróćmy na koniec do dowolnego kontekstu przestrzeni z formą dwuliniową symetryczną. Wprowadzimy pewien typ form dwuliniowych, zwanych formami nieosobliwymi lub niezdegenerowanymi, który pozwoli nam uogólnić na przestrzenie z dowolnymi symetrycznymi formami dwuliniowymi rozkład ortogonalny $V = W \oplus W^\perp$, uzyskany na poprzednim wykładzie. Formy te są daleko „namiastką” form dodatnio określonych w przypadku dowolnego ciała. Ich teorią zajmiemy się za kilka wykładów.

Obserwacja 13.5

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K z formą dwuliniową symetryczną h. Następujące warunki są równoważne dla dowolnej bazy $\mathcal{A}$ przestrzeni V:

(i) $\det(G(h,\mathcal{A})) \neq 0$,

(ii) dla każdego niezerowego wektora $\alpha \in V$ istnieje wektor $\beta \in V$ taki, że $h(\alpha,\beta) \neq 0$.

Dowód. Załóżmy, że forma h spełnia warunek (i) oraz dla pewnego niezerowego $\alpha \in V$ mamy $h(\alpha,\beta) = 0$, dla wszystkich $\beta \in V$. Wektor α dopełniamy do bazy $\mathcal{A}$ przestrzeni V postaci $(\alpha, \beta_1, \ldots, \beta_s)$. Wówczas $G(h,\mathcal{A})$ ma zerowy pierwszy wiersz i kolumnę, bo $h(\alpha,\alpha) = 0$ oraz $h(\alpha,\beta_i) = h(\beta_i,\alpha) = 0$, dla $i = 1,\ldots,s$. A zatem $\det(G(h,\mathcal{A})) = 0$, sprzeczność.

Na odwrót: załóżmy, że zachodzi (ii), ale dla pewnej bazy $\mathcal{A} = (\alpha_1,\ldots,\alpha_n)$ mamy $\det(G(h,\mathcal{A})) = 0$. Oznacza to, że rząd macierzy $G(h,\mathcal{A})$ nie jest maksymalny. A zatem kolumny tej macierzy są liniowo zależne, czyli istnieje wektor niezerowy $(x_1,\ldots,x_n) \in K^n$ taki, że:

$$x_1 \begin{bmatrix} h(\alpha_1,\alpha_1) \\ \vdots \\ h(\alpha_1,\alpha_n) \end{bmatrix} + x_2 \begin{bmatrix} h(\alpha_2,\alpha_1) \\ \vdots \\ h(\alpha_2,\alpha_n) \end{bmatrix} + \ldots + x_n \begin{bmatrix} h(\alpha_n,\alpha_1) \\ \vdots \\ h(\alpha_n,\alpha_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Korzystając z liniowości formy dwuliniowej względem pierwszej zmiennej mamy zatem:

$$\begin{bmatrix} h(x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n, \alpha_1) \\ \vdots \\ h(x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n, \alpha_n) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Niech $\alpha = x_1\alpha_1 + x_2\alpha_2 + \ldots + x_n\alpha_n$. Zauważmy, że skoro $h(\alpha,\alpha_i) = 0$, dla każdego $i = 1,\ldots,n$, to $h(\alpha,\beta) = 0$, dla każdego $\beta \in V$ (bo każde β jest sumą α_i). Skoro w (V,h) zachodzi (ii), to $\alpha = 0$. Ale α to kombinacja wektorów bazowych przestrzeni V o współczynnikach $x_1,\ldots,x_n$. A zatem $x_1 = \ldots = x_n = 0$, co przeczy wyborowi elementów x_i jako współczynników liniowej zależności kolumn $G(h;\mathcal{A})$. □

Wniosek 13.4

Niech V będzie skończenie wymiarową przestrzenią liniową nad ciałem K z formą dwuliniową symetryczną h i niech W będzie podprzestrzenią przestrzeni V. Następujące warunki są równoważne dla dowolnej bazy $\mathcal{A}$ przestrzeni W:

(1) $V = W \oplus W^{\perp}$,

(2) $\det(G(h|_{W\times W},\mathcal{A})) \neq 0$.

Dowód. Załóżmy, że $V = W \oplus W^{\perp}$. Mamy zatem $W \cap W^{\perp} = \{0\}$. Gdyby W nie spełniała (2), to na mocy Obserwacji 13.5, istniałby wektor $\alpha \in W$ taki, że dla każdego $\beta \in W$ mielibyśmy $h(\alpha,\beta) = 0$. W szczególności $h(\alpha,\alpha) = 0$, czyli $\alpha \in W^{\perp}$. A zatem $\alpha \in W \cap W^{\perp}$, sprzeczność.

Na odwrót: przypuśćmy, że forma $h|_{W\times W}$ spełnia (2). Mamy wykazać, że $V = W \oplus W^{\perp}$, czyli że dla każdego $\alpha \in V$ istnieją jednoznacznie wyznaczone wektory $\alpha' \in W$ oraz $\alpha'' \in W^{\perp}$ spełniające warunek $\alpha = \alpha' + \alpha''$. Niech $\mathcal{A} = (\alpha_1,\ldots,\alpha_k)$ będzie bazą przestrzeni W. Istnienie i jednoznaczność wektora postaci $\alpha' = x_1\alpha_1 + \ldots + x_k\alpha_k$ spełniającego powyższe warunki jest równoważne istnieniu i jednoznaczności współczynników $x_1,\ldots,x_k \in K$ takich, że

$$\alpha - (x_1\alpha_1 + \ldots + x_k\alpha_k) \in W^{\perp}.$$

Sprawdzenie czy jakiś wektor leży w $W^{\perp}$ jest równoważne sprawdzeniu, czy wektor ten jest prostopadły do każdego wektora z bazy W (na mocy dwuliniowości h). A zatem teza (czyli (1)) jest równoważna temu, że zachodzi układ warunków

$$h(\alpha_j, \alpha - (x_1\alpha_1 + \ldots + x_k\alpha_k)) = 0, \text{ dla każdego } j = 1,\ldots,k,$$

czyli (1) równoważne jest, z liniowości h, istnieniu jednoznacznego rozwiązania $(x_1, \ldots, x_k)$ układu równań

$$\begin{cases} x_1 h(\alpha_1, \alpha_1) + x_2 h(\alpha_1, \alpha_2) + \ldots + x_k h(\alpha_1, \alpha_k) & = h(\alpha_1, \alpha) \\ x_1 h(\alpha_2, \alpha_1) + x_2 h(\alpha_2, \alpha_2) + \ldots + x_k h(\alpha_2, \alpha_k) & = h(\alpha_2, \alpha) \\ & \vdots \\ x_1 h(\alpha_k, \alpha_1) + x_2 h(\alpha_k, \alpha_2) + \ldots + x_k h(\alpha_k, \alpha_k) & = h(\alpha_k, \alpha) \end{cases}.$$

Macierzą współczynników tego (być może niejednorodnego) układu jest macierz $G(h|_{W\times W}, \mathcal{A})$. Skoro W spełnia (2), to macierz ta jest odwracalna, a zatem jej rząd wynosi k. Także rząd macierzy całego układu (rozmiarów $k \times k+1$) wynosi k, a zatem na mocy Twierdzenia Kroneckera-Capellego układ ten ma dokładnie jedno rozwiązanie. To kończy dowód istnienia i jednoznaczności wektora α', więc i α''. □

Twierdzenie 13.2: Jacobi

Niech h będzie symetryczną formą dwuliniową na przestrzeni skończenie wymiarowej V nad ciałem K, taką że wiodące minory główne $\Delta_i = \det A^{(i)}$ macierzy symetrycznej $A = G(h, \mathcal{A})$ są niezerowe, dla pewnej bazy $\mathcal{A}$ przestrzeni V. Wówczas istnieje baza $\mathcal{B}$ przestrzeni V taka, że

$$G(h, \mathcal{B}) = \operatorname{diag}\left(\frac{\Delta_1}{\Delta_0}, \ldots, \frac{\Delta_n}{\Delta_{n-1}}\right), \quad \text{gdzie } \Delta_0 = 1.$$

Przykład: dla poniższej macierzy $A \in M_n(\mathbb{Q})$ mamy $\Delta_0 = 1, \Delta_1 = 3, \Delta_2 = 2, \Delta_3 = -7$.

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

Macierz ta jest zatem kongruentna do macierzy diagonalnej (wymiernej) o wyrazach $3, \frac{2}{3}, -\frac{7}{2}$.

Dowód. Rozumowanie jest indukcją ze względu na wymiar n przestrzeni V. Teza jest oczywista dla $n = 1$. Niech $n = \dim V > 1$. Niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$. Niech $W = \operatorname{lin}(\alpha_1, \ldots, \alpha_{n-1})$. Zauważmy, że macierz $A^{(n-1)}$ jest macierzą formy $h|_{W\times W}$ i jej wiodące minory główne, to wyznaczniki macierzy $A^{(1)}, A^{(2)}, \ldots, A^{(n-1)}$. Wyznaczniki te są niezerowe. W szczególności W spełnia warunki Wniosku 13.4, bo $\det A^{(n-1)} \neq 0$, czyli $V = W \oplus W^\perp$, więc zgodnie z założeniem indukcyjnym istnieje taka baza $\beta_1, \ldots, \beta_{n-1}$ przestrzeni W oraz dopełniający ją do bazy V wektor $\gamma_n \in W^\perp$, że w bazie $\mathcal{C} = (\beta_1, \ldots, \beta_{n-1}, \gamma_n)$ przestrzeni V mamy:

$$G(h, \mathcal{C}) = \begin{bmatrix} \Delta_1/\Delta_0 & 0 & \ldots & 0 & 0 \\ 0 & \Delta_2/\Delta_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \Delta_{n-1}/\Delta_{n-2} & 0 \\ 0 & 0 & \ldots & 0 & h(\gamma_n, \gamma_n) \end{bmatrix}.$$

Niech $C = M(\mathrm{id})_{\mathcal{C}}^{\mathcal{A}}$ i $\det C = d$. Mamy $C^T \cdot A \cdot C = G(h, \mathcal{C})$ oraz $\det A = \Delta_n$. Zatem

$$d^2 \cdot \Delta_n = \Delta_{n-1} \cdot h(\gamma_n, \gamma_n).$$

Możemy podmienić wektor γ_n przez $\beta_n = d^{-1} \cdot \gamma_n$ dostając, że istnieje baza $\mathcal{B} = (\beta_1, \ldots, \beta_n)$ przestrzeni V, że:

$$G(h, \mathcal{B}) = \begin{bmatrix} \Delta_1/\Delta_0 & 0 & \ldots & 0 & 0 \\ 0 & \Delta_2/\Delta_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \Delta_{n-1}/\Delta_{n-2} & 0 \\ 0 & 0 & \ldots & 0 & h(\gamma_n, \gamma_n)/d^2 \end{bmatrix}.$$

Oczyiście $\det M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}} = 1$, zatem

$$\Delta_n = \Delta_{n-1} \cdot \frac{h(\gamma_n, \gamma_n)}{d^2},$$

co kończy dowód. □

13.1 Uzupełnienie. Dowód kryterium Sylvestera ze skryptu

Czytelnika pracującego ze skryptem wydziałowym zachęcam do prześledzenia alternatywnego dowodu tam przedstawionego, opartego o pojęcie macierzy Grama, które dopiero poznamy. Poniżej przedstawiam ten dowód w wersji nieco rozszerzonej w języku już poznanych faktów.

Nie dowodzimy implikacji (i) $\Rightarrow$ (ii), ponieważ bez pojęcia kongruencji nie mamy chwilowo dobrego narzędzia do wykazania, że wyznacznik macierzy iloczynu skalarnego w dowolnej bazie jest dodatni. W skrypcie czyni się to za pomocą macierzy Grama, w pewnym sensie powtarzając główne argumenty wypisane wyżej w sposób ogólny — mianowicie pokazujemy, że macierz Grama jest iloczynem macierzy C^TC, dla pewnej macierzy odwracalnej C.

Dowodzimy implikację (ii) $\Rightarrow$ (i) przez indukcję po wymiarze n przestrzeni V. Trzeba jedynie pokazać, że h jest dodatnio określona, tzn. $h(\alpha,\alpha) > 0$ dla dowolnego $0 \neq \alpha \in V$.

Dla $n = 1$ teza jest jasna, bo $V = \mathrm{lin}(\alpha_1)$ oraz $A = [a_{11}]$, przy czym $a_{11} > 0$ (bo to jest $\det A^{(1)}$). A zatem dla każdego niezerowego $a\alpha_1 \in V$

$$h(a\alpha_1, a\alpha_1) = a^2 a_{11} > 0.$$

Zatem h jest iloczynem skalarnym. Załóżmy, że dowodzona implikacja jest prawdziwa dla $n-1$. Dowodzimy tezę dla n. Niech $W = \mathrm{lin}(\alpha_1, \ldots, \alpha_{n-1})$.

Z założenia indukcyjnego $(W, h|_W)$ jest przestrzenią euklidesową. Niech $\beta_1, \ldots, \beta_{n-1}$ będzie bazą prostopadłą W otrzymaną z bazy $\alpha_1, \ldots, \alpha_{n-1}$ metodą Grama-Schmidta. W szczególności $h(\beta_i, \beta_j) = 0$, dla $i \neq j$. Definiujemy wektor β_n postaci:

$$\beta_n = \alpha_n - \sum_{i=1}^{n-1} \frac{h(\alpha_n, \beta_i)}{h(\beta_i, \beta_i)} \beta_i.$$

Wówczas nowo zdefiniowany wektor spełnia $h(\beta_n, \beta_i) = 0$, dla wszystkich $i = 1, \ldots, n-1$ (sprawdzamy to podobnie jak w dowodach z poprzedniego wykładu, stosując dwuliniowość oraz symetryczność h).

Rozpatrzmy teraz macierz $B = [b_{ij}] \in M_n(\mathbb{R})$, gdzie $b_{ij} = h(\beta_i, \beta_j)$, dla $i, j = 1, \ldots, n$, czyli:

$$B = \begin{bmatrix} h(\beta_1,\beta_1) & 0 & \ldots & 0 \\ 0 & h(\beta_2,\beta_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & h(\beta_n,\beta_n) \end{bmatrix}.$$

Teraz udowodnimy coś intuicyjnie jasnego, ale technicznego w dowodzie. W istocie macierz B powstaje z macierzy A przez wykonanie operacji elementarnych typu (1) na wierszach i kolumnach (dodawanie wiersza pomnożonego przez stałą do innego wiersza, analogicznie dla kolumn). Dlaczego to wystarczy?

Otóż operacje elementarne typu (1) nie zmieniają wyznacznika. Więc $\det A = \det B$. Wiemy też, że B jest diagonalna i ma na diagonali wyrazy $h(\beta_1,\beta_1), \ldots, h(\beta_n,\beta_n)$. Ich iloczyn to wyznacznik macierzy B. Wiemy z założenia indukcyjnego, że $h(\beta_1,\beta_1), \ldots, h(\beta_{n-1},\beta_{n-1})$ są dodatnie, bo $h|_W$ jest iloczynem skalarnym i korzystamy z (i) $\Rightarrow$ (ii). Ostatni element diagonali $h(\beta_n,\beta_n)$ też musi być jednak dodatni, bo z (ii) mamy

$$\det A^{(n)} = \det A > 0.$$

To oznacza, że h jest iloczynem skalarnym, bowiem biorąc $0 \neq \alpha = x_1\beta_1 + \ldots + x_n\beta_n$ i korzystając z warunku $h(\beta_i, \beta_j) = 0$, dla $i \neq j$, mamy

$$h(\alpha,\alpha) = \sum_{i=1}^{n} x_i^2 h(\beta_i, \beta_i) > 0,$$

co zakończy dowód.

Dowód wymaga dwóch kroków. Po pierwsze zajmiemy się macierzą $X_{\mathcal{A}}$ mającą w wierszach wektory $\alpha_1, \ldots, \alpha_n$ bazy $\mathcal{A}$, a następnie dopiero popatrzymy na minory macierzy A. Po pierwsze twierdzimy, że macierz $X_{\mathcal{B}}$ mająca w kolejnych wierszach wektory[1] $\beta_1, \ldots, \beta_n$ powstaje z macierzy $X_{\mathcal{A}}$ poprzez operacje typu (1) na wierszach.

Pokażmy przez indukcję ze względu na m, że sprowadzenie pierwszych m wierszy macierzy $X_{\mathcal{A}}$ do pierwszych m wierszy macierzy $X_{\mathcal{B}}$ wymaga stosowania jedynie operacji (1) na pierwszych m wierszach.

Dla $m = 1$ [2] mamy $\beta_1 = \alpha_1$, więc teza jest jasna, bo nie trzeba wykonywać żadnych operacji. Natomiast jeśli wiemy, że jest tak dla $m-1$, to bierzemy macierz $X_{\mathcal{A}}$. Zamieniamy pierwsze $m-1$ wierszy z założenia indukcyjnego na $\beta_1, \ldots, \beta_{m-1}$ operacjami (1) wykonywanymi w obrębie $m-1$ pierwszych wierszy. Dostajemy macierz $X_{\mathcal{A}'}$ której pierwsze $m-1$ wierszy to $\beta_1, \ldots, \beta_{m-1}$, a kolejne to $\alpha_m, \ldots, \alpha_n$.

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{m-1} \\ \alpha_m \\ \vdots \\ \alpha_n \end{bmatrix} \xrightarrow{\text{m-1 operacji}} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{m-1} \\ \alpha_m \\ \vdots \\ \alpha_n \end{bmatrix}$$

Z twierdzenia o ortogonalizacji oraz z definicji wektora β_n mamy

$$\beta_m = \alpha_m - c_1\beta_1 - c_2\beta_2 - \ldots - c_{m-1}\beta_{m-1}, \text{ gdzie } c_i = h(\alpha_m, \beta_i)/h(\beta_i, \beta_i).$$

A zatem od wiersza m-tego $X_{\mathcal{A}'}$ odejmujemy c_1 razy pierwszy wiersz. Od uzyskanego m-tego wiersza odejmujemy c_2 razy drugi wiersz itd, na końcu odejmując c_{m-1} razy wiersz $m-1$ od wiersza m-tego. W rezultacie dostajemy macierz $X_{\mathcal{B}'}$, która w pierwszych m wierszach ma wektory $\beta_1, \ldots, \beta_m$, uzyskane przez stosowanie operacji (1) na pierwszych m wierszach, a dalej (jeśli są) wektory $\alpha_{m+1}, \ldots, \alpha_n$. To kończy dowód indukcyjny, że $X_{\mathcal{B}}$ można uzyskać z $X_{\mathcal{A}}$ przez operacje (1).

Niech A' będzie macierzą formy h w bazie $\beta_1, \ldots, \beta_{m-1}, \alpha_m, \ldots, \alpha_n$ oraz niech B' niech będzie macierzą formy h bazy $\beta_1, \ldots, \beta_m, \alpha_{m+1}, \ldots, \alpha_n$.

Do dowodu, że B jest uzyskiwana z A przez operacje elementarne wierszowe i kolumnowe typu (1) wystarczy pokazać, że B' można uzyskać z A' przez operacje typu (1) zmieniające jedynie wiersze i kolumny o indeksie m. Wynika to z faktu, że macierze A' oraz B' mają identyczne wyrazy poza m-tym wierszem i m-tą kolumną, bo układy $\beta_1, \ldots, \beta_{m-1}, \alpha_m, \ldots, \alpha_n$ oraz $\beta_1, \ldots, \beta_m, \alpha_{m+1}, \ldots, \alpha_n$ różnią się tylko m-tym elementem. Zobaczmy:

$$A' = \begin{bmatrix} & \ldots & h(\beta_1, \beta_{m-1}) & h(\beta_1, \alpha_m) & \ldots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h(\beta_{m-1}, \beta_1) & \ldots & h(\beta_{m-1}, \beta_{m-1}) & h(\beta_{m-1}, \alpha_m) & \vdots & h(\beta_{m-1}, \alpha_n) \\ h(\alpha_m, \beta_1) & \ldots & h(\alpha_m, \beta_{m-1}) & h(\alpha_m, \alpha_m) & \vdots & h(\alpha_m, \alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \ldots & h(\alpha_n, \beta_{m-1}) & h(\alpha_n, \alpha_m) & \ldots & \end{bmatrix}$$

$$B' = \begin{bmatrix} & \ldots & h(\beta_1, \beta_{m-1}) & h(\beta_1, \beta_m) & \ldots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h(\beta_{m-1}, \beta_1) & \ldots & h(\beta_{m-1}, \beta_{m-1}) & h(\beta_{m-1}, \beta_m) & \vdots & h(\beta_{m-1}, \alpha_n) \\ h(\beta_m, \beta_1) & \ldots & h(\beta_m, \beta_{m-1}) & h(\beta_m, \beta_m) & \vdots & h(\beta_m, \alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \ldots & h(\alpha_n, \beta_{m-1}) & h(\alpha_n, \beta_m) & \ldots & \end{bmatrix}$$

[1] Mała delikatność: jeszcze nie wiemy czy układ $\beta_1, \ldots, \beta_n$ jest bazą V, bo nie wiemy czy h to iloczyn skalarny na V.

[2] Intuicja: wykonujemy na kolejnych wierszach $X_{\mathcal{A}}$ te same operacje, co przy ortogonalizacji G-S, choć nie wiemy czy h to iloczyn skalarny. Na i-tym wierszu stosujemy $i-1$ operacji (1), co zamienia α_i na $\beta_i = \alpha_i - c_1\beta_1 - \ldots - c_{i-1}\beta_{i-1}$.

Wiemy, że $\beta_m = \alpha_m - c_1\beta_1 - c_2\beta_2 - \ldots - c_{m-1}\beta_{m-1}$, dla znanych c_i. Macierz $X_{\mathcal{B}'}$ uzyskiwaliśmy z $X_{\mathcal{A}'}$ przez wykonanie $m-1$ operacji odjęcia od m-tego wiersza wielokrotności c_i wiersza i-tego. Aby uzyskać macierz B' z A' trzeba wykonać $2m-2$ operacji typu (1). Pierwsze dwie to: odjęcie od m-tego wiersza c_1 razy wiersz pierwszy, potem od kolumny m-tej trzeba odjąć c_1 razy kolumnę pierwszą. Oto rezultat:

$$\begin{bmatrix} & \ldots & h(\beta_1,\beta_{m-1}) & h(\beta_1,\alpha_m - c_1\beta_1) & \ldots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ h(\beta_{m-1},\beta_1) & \ldots & h(\beta_{m-1},\beta_{m-1}) & h(\beta_{m-1},\alpha_m - c_1\beta_1) & \vdots & h(\beta_{m-1},\alpha_n) \\ h(\alpha_m - c_1\beta_1,\beta_1) & \ldots & h(\alpha_m - c_1\beta_1,\beta_{m-1}) & h(\alpha_m - c_1\beta_1,\alpha_m - c_1\beta_1) & \vdots & h(\alpha_m - c_1\beta_1,\alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \ldots & h(\alpha_n,\beta_{m-1}) & h(\alpha_n,\alpha_m - c_1\beta_1) & \ldots & \end{bmatrix}$$

Następne dwie operacje to: od m-tego wiersza odjąć c_2 razy drugi wiersz i potem od m-tej kolumny odjąć c_2 razy kolumnę drugą:

$$\begin{bmatrix} & \ldots & h(\beta_1,\alpha_m - c_1\beta_1 - c_2\beta_2) & \ldots & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ h(\alpha_m - c_1\beta_1 - c_2\beta_2,\beta_2) & \ldots & h(\alpha_m - c_1\beta_1 - c_2\beta_2,\alpha_m - c_1\beta_1 - c_2\beta_2) & \vdots & h(\alpha_m - c_1\beta_1 - c_2\beta_2,\alpha_n) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ & \ldots & h(\alpha_n,\alpha_m - c_1\beta_1 - c_2\beta_2) & \ldots & \end{bmatrix},$$

i tak dalej. W ten sposób, korzystając z liniowości h względem każdej ze zmiennych dostajemy żądaną postać B'. Zauważmy wszak, że m-ty wiersz macierzy B' to kombinacja liniowa c_1 razy pierwszy wiersz $+\ c_2$ razy drugi wiersz $+ \ldots + c_{m-1}$ razy $m-1$-wszy wiersz. Podobnie dla kolumn. Zatem pokazaliśmy, że B można dostać z A operacjami elementarnymi typu (1), co kończy dowód implikacji (ii) $\Rightarrow$ (i).

* * *

Dowód był stosunkowo zawiły, ale stanowi ilustrację poglądu, że można przeprowadzić go bez pojęcia kongruencji. Jeszcze inny elementarny dowód przytacza dr Krych w notatkach z Analizy II, cytując rozumowanie prof. Musielaka: `https://www.mimuw.edu.pl/~krych/matematyka/AM2skrypt/am2cz02L.pdf`.

Idea jest taka, by patrzeć na tzw. formę kwadratową q (definicję poznamy wkrótce) stowarzyszoną z formą dwuliniową h postaci $q((x_1,\ldots,x_n)) = h((x_1,\ldots,x_n),(x_1,\ldots,x_n))$. Aby stwierdzić czy forma h jest dodatnio określona wystarczy pokazać, że $q(x) > 0$, dla każdego niezerowego $x \in \mathbb{R}^n$. A zatem badamy wyrażenie postaci $q((x_1,\ldots,x_n)) = \sum_{1\leqslant i,j\leqslant n} a_{ij}x_ix_j$, dla pewnych $a_{ij} \in \mathbb{R}$, przy czym $a_{ij} = a_{ij}$, bo forma h jest symetryczna. Myślmy o q jak funkcji kwadratowej f zmiennej x_1. Warunkiem koniecznym i dostatecznym na to, by $f(x_1) > 0$ jest $a_{11} > 0$ oraz to, by wyróżnik tego trójmianu był ujemny. Jak się okazuje, wyróżnik ten zależy jedynie od $n-1$ zmiennych $x_2,\ldots,x_n$, a nie zależy od x_1. To pozwala przeprowadzić rozumowanie indukcyjne. Szczegóły — we wspomnianym wyżej tekście dra Krycha.

Ostatnia metoda ilustruje następującą myśl: czasem łatwiej pomyśleć niż korzystać z gotowych kryteriów. Rozważmy, dla $a \in \mathbb{R}$ macierz $A_n \in M_n(\mathbb{R})$ postaci:

$$A_n = G(h_a, st) = \begin{bmatrix} a & 1 & 0 & \ldots & 0 & 0 \\ 1 & a & 1 & \ldots & 0 & 0 \\ 0 & 1 & a & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a & 1 \\ 0 & 0 & 0 & \ldots & 1 & a \end{bmatrix}.$$

Czy h jest iloczynem skalarnym? Wystarczy sprawdzić dla jakich a dla dowolnego niezerowego wektora $v \in \mathbb{R}^n$ mamy $h_a(v,v) > 0$. Robimy to (niemal) wprost z definicji: niech $v = (v_1,\ldots,v_n)$. Wtedy dla dowolnego a wartość $h_a(v,v) = v^TA_nv$ równa jest:

$$a(v_1^2+\ldots+v_n^2)+2v_1v_2+2v_2v_3+\ldots+2v_{n-1}v_n = (a-2)(v_1^2+\ldots+v_n^2)+v_1^2+(v_1+v_2)^2+\ldots+(v_{n-1}+v_n)^2+v_n^2 > 0.$$

Jest więc jasne, że dla $a > 2$ forma h_a jest iloczynem skalarnym. Można też rozwiązać ten punkt korzystając z kryterium Sylvestera, wyprowadzając z rozwinięcia Laplace'a rekurencję $|A_n| = a|A_{n-1}| - |A_{n-2}|$.

13.2 Dodatek. Grupa klas kwadratów ciała

W części wykładu dotyczącej kongruencji macierzy nad dowolnym ciałem i warunków koniecznych do tego, by macierze były kongruentne, wspomnieliśmy o tym, że „im więcej kwadratów", tym trudniejsza klasyfikacja macierzy z dokładnością do kongruencji. W tym dodatku przedstawimy kilka uwag dotyczących tej kwestii. Po szczegóły odsyłam do podręcznika „Wykłady z algebry dwuliniowej" prof. Kazimierza Szymiczka. Załóżmy, że chcemy rozstrzygnąć następujący problem.

Znaleźć warunki koniecznie i dostateczne na to, by dla nad ciałem K
zachodziła kongruencja macierzy diagonalnych:

$$\operatorname{diag}(a_1,\dots,a_n) \cong \operatorname{diag}(b_1,\dots,b_n).$$

Trywialna uwaga 1. Jeśli $A = \operatorname{diag}(a_1,\dots,a_n)$ oraz $B = \operatorname{diag}(a_{\sigma(1)},\dots,a_{\sigma(n)})$, dla pewnej permutacji $\sigma \in S_n$, to jeśli P jest macierzą permutacji σ, to $B = P^T AP$, czyli: $\operatorname{diag}(a_1,\dots,a_n) \cong \operatorname{diag}(a_{\sigma(1)},\dots,a_{\sigma(n)})$. Inaczej – jeśli $A = G(h,\mathcal{A})$, gdzie $\mathcal{A} = (\alpha_1,\dots,\alpha_n)$, to $B = G(h,\mathcal{B})$, dla bazy $\mathcal{B} = (\alpha_{\sigma(1)},\dots,\alpha_{\sigma(n)})$.

Trywialna uwaga 2. Dla dowolnych $a_1,\dots,a_n,b_1 \neq 0,\dots,b_n \neq 0 \in K$ zachodzi kongruencja $\operatorname{diag}(a_1,\dots,a_n) \cong \operatorname{diag}(b_1^2a_1,\dots,b_n^2a_n)$.

Definicja 13.4: Ciało kwadratowo domknięte

Ciało K nazywamy KWADRATOWO DOMKNIĘTYM, jeśli każdy element K jest kwadratem pewnego elementu z K.

Oczywiście każde ciało algebraicznie domknięte jest kwadratowo domknięte. Kwadratowo domknięte jest również ciało $\mathbb{Z}_2$, a także (co nie jest takie łatwe — dowolne skończone ciało charakterystyki 2). W każdym ciele $\mathbb{Z}_p$, dla $p > 2$ istnieją niekwadraty (o czym dalej). Oczywiście $\mathbb{Q}$ czy $\mathbb{R}$ kwadratowo domknięte nie są. Nie jest więc zupełnie trywialne podanie innych przykładów. Stosunkowo naturalny przykład to tzw. zespole liczby konstruowalne, czyli kwadratowe domknięcie ciała $\mathbb{Q}$. Liczby te są podciałem ciała liczb zespolonych złożonym z liczb $a+bi$, gdzie $|a|, |b|$ są liczbami, które można skonstruować za pomocą cyrkla i linijki dysponując odcinkiem długości 1. Innym, również niezupełnie oczywistym przykładem jest ciało $\mathbb{Z}_{5^\infty}$, czyli suma wszystkich skończonych ciał charakterystyki 5 (czyli w istocie ciał $\mathbb{Z}_{5^n}$, dla $n \geqslant 1$).

Obserwacja 13.6

Jeśli K jest ciałem kwadratowo domkniętym oraz jeśli $A = \operatorname{diag}(a_1,\dots,a_r,0,\dots,0) \in M_n(K)$, gdzie $a_i \neq 0$, to

$$\operatorname{diag}(a_1,\dots,a_r,0,\dots,0) \simeq \operatorname{diag}(\underbrace{1,\dots,1}_{r},0,\dots,0).$$

ROZWIĄZANIE.Jeśli $A = \operatorname{diag}(a_1,\dots,a_r,0,\dots,0) \in M_n(\mathbb{C})$, gdzie $a_i \neq 0$, jest macierzą formy h na V w bazie $(\alpha_1,\dots,\alpha_n)$, to biorąc bazę $\mathcal{B} = (\beta_1,\dots\beta_n)$ postaci:

$$\beta_j = \begin{cases} \frac{1}{\sqrt{a_i}}\alpha_i, & \text{dla } i \leqslant r, \\ \alpha_i, & \text{dla } i > r \end{cases}$$

dostajemy: $G(h,\mathcal{B}) = \operatorname{diag}(1,\dots,1,0,\dots,0)$, czyli:

$$\operatorname{diag}(a_1,\dots,a_r,0,\dots,0) \simeq \operatorname{diag}(\underbrace{1,\dots,1}_{r},0,\dots,0).$$

■

Teorią kongruencji macierzy diagonalnych zajmiemy się na kolejnych wykładach. Celem tego dodatku jest wprowadzenie następującego pojęcia.

Definicja 13.5: Grupa klas kwadratów ciała

Na zbiorze $\dot{K}$ niezerowych elementów ciała K definiujemy relację równoważności: $a \sim b$ wtedy i tylko wtedy, gdy ab^{-1} jest kwadratem w K. Zbiór klas abstrakcji $\dot{K}/\dot{K}^2$ ma strukturę grupy z działaniem

$$[a] \cdot [b] = [ab]$$

nazywanej GRUPĄ KLAS KWADRATÓW CIAŁA K.

Przedstawmy dla informacji Czytelnika kilka przykładów, nie wchodząc w (niekiedy łatwe) uzasadnienia.

- Dla ciał kwadratowo domkniętych mamy $|\dot{K}/\dot{K}^2| = 1$.
- Jeśli $0 \neq a \in \mathbb{R}$, to $a \sim 1$ lub $a \sim -1$, czyli $|\dot{\mathbb{R}}/\dot{\mathbb{R}}^2| = 2$.
- Jeśli $p, q \in \mathbb{Q}$ są liczbami pierwszymi, to $p \sim q \Leftrightarrow p = q$, czyli $|\dot{\mathbb{Q}}/\dot{\mathbb{Q}}^2| = \infty$.
- $(*)$ jeśli $|\dot{K}/\dot{K}^2|$ jest skończona, to ma 2^k elementów, dla pewnego k.
- $(*)$ Dla każdego ciała skończonego K charakterystyki $\neq 2$ mamy $|\dot{K}/\dot{K}^2| = 2$.
- $(*)$ Rozważa się tzw. ciała szeregów formalnych (ilorazy szeregów formalnych) o współczynnikach w ciele K, ozn. $K((x))$. Okazuje się, że

$$|K(\dot{(}x)) / K(\dot{(}x))^2| = 2 \cdot |\dot{K}/\dot{K}^2|.$$

Czytelnik bez trudu sprawdzi, że działanie na klasach kwadratów jest istotnie działaniem grupowym. Uzyskiwane grupy są oczywiście przemienne. W teorii, którą rozważamy kluczowa jest klasyfikacja macierzy kongruentnych nad ciałem K dla $|\dot{K}/\dot{K}^2| = 2$, obejmująca także macierze rzeczywiste. Nie będziemy się tym tematem zajmować w tym miejscu. Wspomnimy natomiast jeszcze kilka definicji z teorii ciał.

Definicja 13.6

Ciało K nazywa się:

- FORMALNIE RZECZYWISTYM, jeśli element $-1 \in K$ nie jest sumą kwadratów z ciała K,
- NIERZECZYWISTYM, jeśli K nie jest formalnie rzeczywiste,
- UPORZĄDKOWANYM, jeśli na K jest określona relacja $>$, zwana porządkiem, o następujących własnościach, dla każdych $a, b, c \in K$:
  - przechodniość: $a > b$ i $b > c \Rightarrow a > c$,
  - trychotomia $a > b$ albo $a = b$ albo $b > a$,
  - zgodność z działaniami w K: $a > b \Rightarrow a + c > b + c$ oraz $a > b$ i $c > 0 \Rightarrow ac > bc$.

Przykłady i uwagi – niektóre jeszcze w tym momencie trudne do uzasadnienia.

- Każde ciało charakterystyki $p \neq 0$ jest nierzeczywiste, bo $-1 = (p-1) \cdot 1^2$.
- Każde ciało kwadratowo domknięte jest nierzeczywiste.
- Jeśli K jest ciałem formalnie rzeczywistym i $|\dot{K}/\dot{K}^2| = 2$, to relacja

$$a > b \Leftrightarrow a - b \in \dot{K}^2$$

  zadaje na K strukturę ciała uporządkowanego. Elementy $a > 0$ ciała K nazywamy DODATNIMI (w porządku $>$), zaś elementy $a < 0$ nazywamy UJEMNYMI.
- $(*)$ Jeśli K jest ciałem formalnie rzeczywistym i $|\dot{K}/\dot{K}^2| = 2$, to dla macierzy symetrycznych nad tym ciałem zachodzi twierdzenie o bezwładności, to znaczy – są one kongruentne do macierzy diagonalnych o wyrazach $\{-1, 0, 1\}$ i wiadomo, które z tych macierzy są kongruentne.
- $(*)$ Jeśli K jest ciałem nierzeczywistym o charakterystyce $\neq 2$ oraz $|\dot{K}/\dot{K}^2| = 2$ (np. ciało skończone charakterystyki $\neq 2$), to macierze symetryczne nad K są kongruentne wtedy i tylko wtedy, gdy ich rzędy oraz tzw. wyróżniki (wyznaczniki modulo kwadrat z ciała K) są równe.

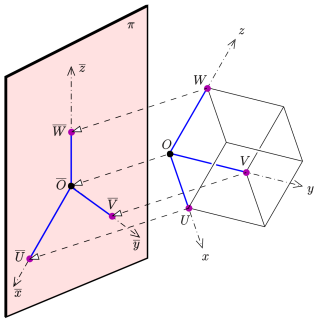
13.3 Trivia. Inne kątomierze, czyli o aksonometrii

Widzieliście Państwo kiedyś kątomierz izometryczny? Czy wiecie, że na trzy rozłożone palce rzutować można sześcian? Jaki jest kąt pomiędzy (1,0) a (0,1) i jaki jest kierunek ich „dwusiecznej”? Spróbujmy powiązać te tematy tak, by uzyskać jeszcze jedno spojrzenie na iloczyny skalarne. Spójrzmy na poniższy rysunek

Twierdzimy, że kierunki wyznaczone przez trzy palce, z których żadne dwa nie są połączone wyznaczają sześcian. Co to znaczy? Jest to poglądowe wysłowienie następującego rezultatu, wspomnianego przez samego Hugona Steinhausa w *Kalejdoskopie Matematycznym*

Twierdzenie 13.3: Pohlke (1853), dowód – Schwarz (1864)

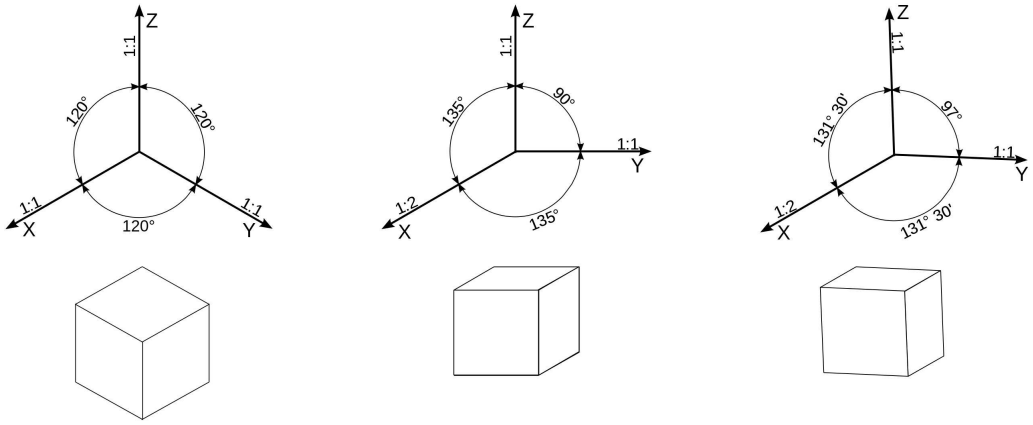
Trzy dowolne odcinki na płaszczyźnie wychodzące z jednego punktu, które nie leżą na jednej prostej można rozważać jako rzut równoległy krawędzi sześcianu.



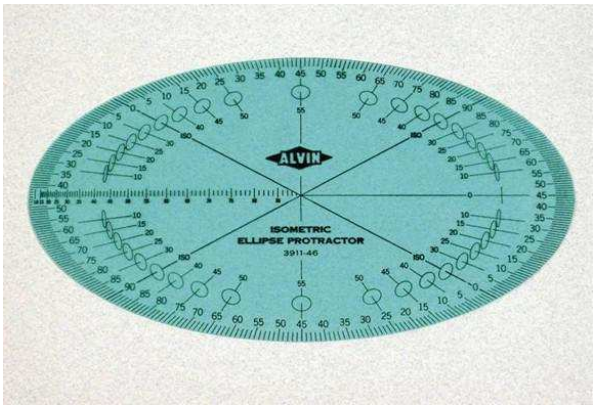
Nietrudny dowód tego rezultatu, wymagający jednak wiedzy, którą dopiero posiądziemy, znaleźć można w tekście `https://www.heldermann-verlag.de/jgg/jgg22/j22h2manf.pdf`. Postawmy sobie jednak inny problem: co to twierdzenie mówi o naszej percepcji kątów? Czy dla pary wektorów nierównoległych znaleźć można zawsze iloczyn skalarny taki, że kąt między nimi będzie prosty? Zagadnienia te mają związki z zastosowaniami matematyki w tzw. aksonometrii — dziedzinie mającej zastosowanie w rysunku technicznym, a dotyczącej odwzorowań przestrzeni na płaszczyznę z wykorzystaniem prostokątnego układu osi. Dąży się przy tym do zachowania prawdziwych wymiarów rzutowanych obiektów, przynajmniej w jednym wybranym kierunku. Trzy podstawowe typy aksonometrii to:

- izometria – wszystkie osie układu prostokątnego w przestrzeni tworzą jednakowy kąt z rzutnią i ich obrazy ulegają jednakowemu skrótowi – na rzutni powstaje obraz trzech osi tworzących pomiędzy sobą kąty po 120°, często na rysunkach izometrycznych pomija się wpływ skrótu;
- dimetria – dwie z osi układu prostokątnego tworzą z rzutnią jednakowe kąty (bywa, że są do niej równoległe), a zatem układ współrzędnych posiada jednakowe skróty na co najmniej dwóch osiach,
- trimetria (anizometria) – każda z osi układu prostokątnego tworzy z rzutnią inny kąt i podlega innemu skrótowi.

Oto przykładowe rzuty sześcianu i kąty pomiędzy ścianami przy tych typach odwzorowań.



Do rzutowania aksonometrycznego używa się specjalnych kątomierzy, podobnych do tego poniżej.



Aby dodać do naszej opowieści nieco matematyki, rozwiążmy prosty problem.

Zadanie. W przestrzeni euklidesowej $(\mathbb{R}^2, \langle\,,\rangle)$ dane są wektory $(1,0),(0,1)$. Jaki może być, w zależności od $\langle\,,\rangle$ kąt między tymi wektorami? Opisać iloczyny skalarne na przestrzeni $(\mathbb{R}^2, \langle\,,\rangle)$, dla których kąt pomiędzy $(1,0)$, a $(1,1)$ wynosi $\frac{\pi}{2}$.

Rozwiązanie. Ogólna postać iloczynu skalarnego wektorów $\langle (x_1,x_2),(y_1,y_2)\rangle$ w $\mathbb{R}^2$ to:

$$a_{11}x_1y_1 + a_{12}x_1y_2 + a_{21}x_2y_1 + a_{22}x_2y_2,$$
$$\text{gdzie } a_{12} = a_{21} = \langle (1,0),(0,1)\rangle, \quad a_{11} > 0, \quad a_{11}a_{22} - a_{12}a_{21} > 0.$$

Na przykład $\langle (1,0),(0,1)\rangle = 0$ zachodzi tylko dla iloczynu skalarnego postaci $ax_1y_1 + bx_2y_2$, dla $a,b > 0$. Natomiast $\langle (1,0),(1,1)\rangle = 0$ zachodzi tylko dla iloczynu skalarnego postaci $ax_1y_1 - ax_1y_2 - ax_2y_1 + bx_2y_2$, dla $b > a > 0$. A zatem $(1,0)$ i $(1,1)$ są prostopadłe dla każdego iloczynu skalarnego postaci:

$$\langle (x_1,x_2),(y_1,y_2)\rangle = x_1y_1 - x_1y_2 - x_2y_1 + bx_2y_2, \quad \text{gdzie } b > 1.$$

Zatem $\langle (1,0),(0,1)\rangle = -1, \|(1,0)\| = \sqrt{\langle (1,0),(1,0)\rangle} = 1$ oraz $\|(0,1)\| = \sqrt{\langle (0,1),(0,1)\rangle} = \sqrt{b}$. W zależności od b, cosinus kąta pomiędzy $(1,0)$, a $(0,1)$ wynosi:

$$\frac{-1}{\sqrt{b}} \in (-1,0).$$

■

Pytania do samodzielnego przemyślenia.

1. Czy kąt pomiędzy $(1,0)$, $(0,1)$ może dla pewnego $\langle\,,\rangle$ wynosić π?
2. Czy niezależnie od $\langle\,,\rangle$ suma kątów w trójkącie to π?
3. Czy trójkąty wyznaczają iloczyny skalarne w $\mathbb{R}^2$?

Rozdział 14

Macierz Grama układu wektorów. Miara, orientacja, iloczyn wektorowy

Przyjrzymy się kolejnemu zastosowaniu macierzy formy dwuliniowej, ograniczając się z jednej strony do iloczynów skalarnych, a rozszerzając z drugiej strony definicję na dowolne układy wektorów. Celem będzie mówienie o n-wymiarowej objętości i związanych z nią konfiguracjach geometrycznych.

Definicja 14.1: Macierz i wyznacznik Grama układu wektorów

Niech $\alpha_1, \ldots, \alpha_k$ będzie układem wektorów przestrzeni euklidesowej $(V, \langle , \rangle)$. Macierzą Grama układu $\alpha_1, \ldots, \alpha_k$ nazywamy macierz $G(\alpha_1, \ldots, \alpha_k) \in M_k(\mathbb{R})$, która w i-tym wierszu i j-tej kolumnie ma wyraz $\langle \alpha_i, \alpha_j \rangle$. Wyznacznikiem Grama układu $\alpha_1, \ldots, \alpha_k$ nazywamy natomiast liczbę

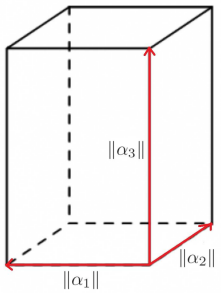
$$W(\alpha_1, \ldots, \alpha_k) = \det G(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}.$$

Rozmiar macierzy Grama nie ma z założenia nic wspólnego z wymiarem przestrzeni V. W n-wymiarowej przestrzeni możemy rozpatrywać układ złożony choćby z dwóch wektorów i wyznaczyć jego macierz Grama rozmiarów 2×2. Oto przykład. W przestrzeni euklidesowej $(\mathbb{R}^3, \langle , \rangle_{st})$ możemy rozważyć układ $\alpha_1 = (1, 0, 1), \alpha_2 = (1, 1, 2)$. Wówczas $\langle \alpha_1, \alpha_1 \rangle = 2$, $\langle \alpha_1, \alpha_2 \rangle = \langle \alpha_2, \alpha_1 \rangle = 3$, $\langle \alpha_2, \alpha_2 \rangle = 6$. Stąd:

$$G(\alpha_1, \alpha_2) = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}.$$

Zobaczmy co wyznacznik Grama może mieć wspólnego z polem czy objętością. Wracając do przykładu wyżej możemy stwierdzić, że $\|(1, 0, 1)\| = \sqrt{2}$ oraz $\|(1, 1, 2)\| = \sqrt{6}$. Mamy też $\langle (1, 0, 1), (1, 1, 2) \rangle = 3$, zatem kosinus kąta pomiędzy rozważanymi wektorami to $3/\sqrt{12} = \sqrt{3}/2$, czyli kąt ten równy jest $30°$. Z geometrii szkolnej wiemy zatem, że pole równoległoboku o bokach długości $\sqrt{2}, \sqrt{6}$ oraz kącie $30°$ pomiędzy nimi wynosi $\sqrt{2} \cdot \sqrt{6} \cdot \sin 30°$, czyli $\sqrt{3}$. Wyznacznik macierzy wyżej jest kwadratem tego pola – wynosi 3. Nie jest to przypadek.

Oto inna charakterystyczna sytuacja, gdy macierz Grama wektorów $\alpha_1, \alpha_2, \alpha_3$ jest diagonalna. Czym są wyrazy na jej przekątnej? Są kwadratami długości wektorów α_i, czyli $\|\alpha_1\|^2, \|\alpha_2\|^2, \|\alpha_3\|^2$. Układ $\alpha_1, \alpha_2, \alpha_3$ jest prostopadły, bo macierz Grama jest diagonalna, a wyznacznik tej macierzy to oczywiście $\|\alpha_1\|^2\|\alpha_2\|^2\|\alpha_3\|^2$, czyli kwadrat objętości prostopadłościanu o krawędziach długości $\|\alpha_1\|, \|\alpha_2\|, \|\alpha_3\|$.



Zacznijmy od następującego faktu.

Wniosek 14.1

Dla układu wektorów $\alpha_1, \ldots, \alpha_k$ przestrzeni euklidesowej $(V, \langle\,,\,\rangle)$ oraz dla bazy ortonormalnej $\beta_1, \ldots, \beta_n$ tej przestrzeni niech A będzie macierzą rozmiaru $n \times k$ mającą w j-tej kolumnie współrzędne wektora α_j w bazie ortonormalnej $\beta_1, \ldots, \beta_n$, to znaczy: $A = (a_{ij})$, gdzie $a_{ij} = \langle\, \alpha_j, \beta_i \,\rangle$, dla $i = 1, \ldots, k$, $j = 1, \ldots, n$. Wówczas:

$$G(\alpha_1, \ldots, \alpha_k) = A^T A.$$

W szczególności dla $k = n$ mamy $W(\alpha_1, \ldots, \alpha_n) = (\det A)^2$.

Dowód. Wyraz w i-tym wierszu oraz j-tej kolumnie $G(\alpha_1, \ldots, \alpha_n)$ jest równy $\langle \alpha_i, \alpha_j \rangle$, czyli to iloczyn i-tego wiersza A^T przez macierz $G(\beta_1, \ldots, \beta_n) = I$ oraz przez j-tą kolumnę macierzy A. Ta ostatnia zawiera współrzędne α_j w bazie $\beta_1, \ldots, \beta_n$. Druga część wynika z tego, że $\det A = \det A^T$, gdy $k = n$, a zatem z twierdzenia Cauchy'ego o wyznacznikach mamy $\det G(\alpha_1, \ldots, \alpha_k) = \det A^T \det A = (\det A)^2$. □

Obserwacja 14.1

Dla dowolnego układu $\alpha_1, \ldots, \alpha_k$ wektorów przestrzeni euklidesowej $(V, \langle\,,\,\rangle)$ zachodzi nierówność $W(\alpha_1, \ldots, \alpha_k) \geqslant 0$. Co więcej następujące warunki są równoważne:

(a) $\alpha_1, \ldots, \alpha_k$ jest liniowo niezależny,

(b) $W(\alpha_1, \ldots, \alpha_k) > 0$.

Dowód. Niech $\mathcal{B} = \{\beta_1, \ldots, \beta_n\}$ będzie bazą ortonormalną przestrzeni euklidesowej $W = \operatorname{lin}(\alpha_1, \ldots, \alpha_k)$ z iloczynem skalarnym $\langle\,,\,\rangle|_W$. Niech A ponownie macierzą rozmiaru $n \times k$ mającą w j-tej kolumnie współrzędne wektora α_j w bazie $\beta_1, \ldots, \beta_n$. Zauważmy, że jeśli układ $\mathcal{A} = \{\alpha_1, \ldots, \alpha_k\}$ jest liniowo niezależny, to jest on bazą W, czyli $k = n$ i macierz A jest odwracalną macierzą kwadratową postaci $M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}$. Skoro $G(\alpha_1, \ldots, \alpha_k) = A^T A$, to w tym przypadku $W(\alpha_1, \ldots, \alpha_k) = \det(A^T A) = \det(A)^2 > 0$.

Jeśli $\alpha_1, \ldots, \alpha_k$ jest układem liniowo zależnym, to macierz A ma rząd mniejszy niż k. Rzeczywiście, warunkiem równoważnym liniowej zależności układu $\alpha_1, \ldots, \alpha_k$ jest to, że jeden z wektorów α_i jest kombinacją liniową wektorów α_j, dla $i \neq j$. Niech to będzie

$$\alpha_i = \sum_{j \neq i} a_j \alpha_j.$$

Wówczas i-ta kolumna macierzy A zawierająca współrzędne α_i w bazie ortonormalnej $\mathcal{B}$ ma postać:

$$\begin{bmatrix} \cdots & \left\langle \sum_{j \neq i} a_j \alpha_j, \beta_1 \right\rangle & \cdots \\ \ddots & \vdots & \ddots \\ \cdots & \left\langle \sum_{j \neq i} a_j \alpha_j, \beta_n \right\rangle & \cdots \end{bmatrix} = \sum_{j \neq i} a_j \begin{bmatrix} \cdots & \langle \alpha_j, \beta_1 \rangle & \cdots \\ \ddots & \vdots & \ddots \\ \cdots & \langle \alpha_j, \beta_n \rangle & \cdots \end{bmatrix}.$$

A zatem wykonując ciąg operacji elementarnych typu (1) polegających na odejmowaniu od i-tej kolumny j-tej przemnożonej przez a_j uzyskujemy wyzerowanie i-tej kolumny. Skoro rząd macierzy nie zmienia się przy operacjach elementarnych, to rząd macierzy A jest mniejszy niż k. Z pierwszego semestru wiemy, że dla dowolnych macierzy X, Y, dla których istnieje iloczyn XY, mamy[1]

$$r(XY) \leqslant \min\{r(X), r(Y)\}.$$

Stąd $r(G(\alpha_1, \ldots, \alpha_k)) = r(A^T A) \leqslant \min\{r(A^T), r(A)\} = r(A) < k$. Ostatecznie $W(\alpha_1, \ldots, \alpha_k) = 0$. W szczególności pokazaliśmy też pierwszą część tezy (także dla macierzy, gdzie A nie jest kwadratowa). □

Zobaczmy jak zmienia się wyznacznik Grama przy elementarnych modyfikacjach układu wektorów. Elementy poniższego dowodu wykorzystać można w dowodzie kryterium Sylvestera (uzup. wykładu 14).

[1] Przypominam — to wynika stąd, że kolumny macierzy XY są kombinacjami liniowymi kolumn macierzy Y.

Obserwacja 14.2

Niech $\alpha_1, \ldots, \alpha_k$ będzie układem wektorów w przestrzeni euklidesowej $(V, \langle , \rangle)$. Wówczas jeśli układ $\beta_1, \ldots, \beta_k$ powstaje z $\alpha_1, \ldots, \alpha_k$ za pomocą dowolnej z trzech poniższych operacji:

(a) przez dodanie do któregoś z wektorów α_i kombinacji liniowej pozostałych wektorów układu $\alpha_1, \ldots, \alpha_k$,

(b) przez zmianę wektora z układu $\alpha_1, \ldots, \alpha_k$ na wektor przeciwny,

(c) przez zamianę kolejności wektorów z układu $\alpha_1, \ldots, \alpha_k$, to

$$W(\alpha_1, \ldots, \alpha_k) = W(\beta_1, \ldots, \beta_k).$$

Jeśli $\beta_1, \ldots, \beta_k$ powstaje z $\alpha_1, \ldots, \alpha_k$ przez przemnożenie jednego z wektorów α_i przez $a \in \mathbb{R}$, to

$$a^2 \cdot W(\alpha_1, \ldots, \alpha_k) = W(\beta_1, \ldots, \beta_k).$$

Wiemy już, że wyznacznik Grama jest liczbą nieujemną (dodatnią, gdy układ jest liniowo niezależny). Chcielibyśmy myśleć o jego pierwiastku kwadratowym jako o liczbie przypisującej układowi wektorów „miarę" ich wzajemnego położenia w przestrzeni euklidesowej. Prowadzi to do definicji równoległościanu.

Definicja 14.2: Równoległościan w przestrzeni afinicznej

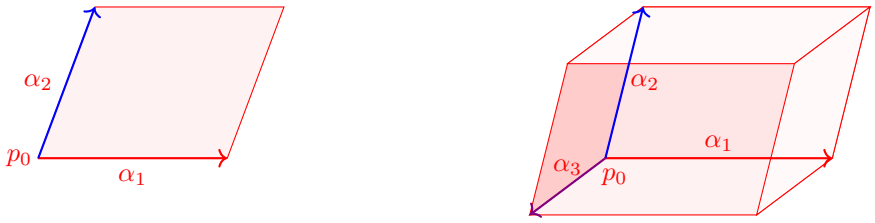
Niech H będzie przestrzenią euklidesową afiniczną. Podzbiór $R(p_0; \alpha_1, \ldots, \alpha_k) \subset H$ postaci

$$\{p_0 + a_1\alpha_1 + \ldots + a_k\alpha_k \,|\, a_1, \ldots, a_k \in [0,1]\}$$

gdzie $p_0 \in H$ oraz $\alpha_1, \ldots, \alpha_k$ są liniowo niezależne w $T(H)$ nazywamy k-WYMIAROWYM RÓWNOLEGŁOŚCIANEM w H ROZPIĘTYM NA WEKTORACH $\alpha_1, \ldots, \alpha_k$ ZACZEPIONYCH W PUNKCIE p_0.

Wprowadzamy szczególne nazewnictwo dla równoległościanów w przestrzeniach niskich wymiarów.

- Równoległościan $R(p_0; \alpha_1)$ nazywamy ODCINKIEM o końcach w punktach $p_0, p_0 + \alpha_1$.
- Równoległościan $R(p_0; \alpha_1, \alpha_2)$ nazywamy RÓWNOLEGŁOBOKIEM.



Równoległobok oraz równoległościan. Źródło: materiały prof. Jarosława Wiśniewskiego

Dla równoległościanu $R(p_0; \alpha_1, \ldots, \alpha_k) \subset H$ zbiór punktów postaci $p_0 + \epsilon_1\alpha_1 + \ldots + \epsilon_k\alpha_k$, gdzie $\epsilon_i \in \{0, 1\}$ nazywamy często zbiorem WIERZCHOŁKÓW równoległościanu, a dla $k = 1$ – końcami odcinka.

Definicja 14.3: Miara n-wymiarowa

Niech $R = R(p_0; \alpha_1, \ldots, \alpha_k)$ będzie k-wymiarowym równoległościanem w afinicznej przestrzeni euklidesowej. Przez $\mu_n(R)$ określać będzie liczbę rzeczywistą zwaną n-WYMIAROWĄ MIARĄ) (albo n-wymiarową objętością) równoległościanu R, przy czym

$$\mu_k(R) = \sqrt{W(\alpha_1, \ldots, \alpha_k)}.$$

Ponadto przyjmujemy $\mu_n(R) = 0$, dla $n > k$ oraz $\mu_l(R) = \infty$, dla $n < k$.

Jednowymiarową miarę nazywamy długością, dwuwymiarową – polem, a trójwymiarową – objętością. Stwierdzenie z początku wykładu podaje szereg intuicji uzasadniających przyjętą przez nas definicję. Objętość nie zależy od kolejności wektorów ani zmiany zwrotu jednego z nich. Zmienia się ona proporcjonalnie do zmiany normy każdego z wektorów rozpinających równoległościan.

Zauważmy, że dla dowolnych wektorów $\alpha_1, \ldots, \alpha_n$ przestrzeni liniowej $\mathbb{R}^n$ ze standardowym iloczynem skalarnym zachodzi $W(\alpha_1, \ldots, \alpha_n) = (\det A)^2$, bo możemy traktować macierz A jako macierz $M(\mathrm{id})^{st}_{\mathcal{A}}$. Stąd $|\det A|$ jest n-wymiarową objętością równoległościanu rozpiętego przez rozważany układ wektorów.

W przypadku, gdy układ $\alpha_1, \ldots, \alpha_n$ wektorów rozpinających równoległościan równy jest wymiarowi n przestrzeni, w której się on znajduje, wówczas miara n-wymiarowa tego równoległościanu równa jest wyznacznikowi macierzy zawierającej w kolumnach współrzędne $\alpha_1, \ldots, \alpha_n$ w dowolnej bazie ortonormalnej. Jest to też pojęcie intuicyjne z punktu widzenia szkolnego wzoru na pole jako „iloczynu podstawy i wysokości", o czym mowa w kolejnym stwierdzeniu.

Twierdzenie 14.1: „Iloczyn pola podstawy i wysokości"

Niech $\alpha_1, \ldots, \alpha_k$ będzie układem wektorów przestrzeni euklidesowej $(V, \langle\,,\rangle)$ i niech γ będzie rzutem prostopadłym α_k na $\mathrm{lin}(\alpha_1, \ldots, \alpha_{k-1})^{\perp}$.

$$W(\alpha_1, \ldots, \alpha_k) = W(\alpha_1, \ldots, \alpha_{k-1}) \cdot \|\gamma\|^2.$$

Dowód. Niech $W = \mathrm{lin}(\alpha_1, \ldots, \alpha_{k-1})$. Wiemy z wcześniejszych wykładów, że

$$V = W \oplus W^{\perp} \quad \text{oraz} \quad \alpha_k = \beta + \gamma,$$

dla pewnego $\beta \in W$. A zatem:

$$W(\alpha_1, \ldots, \alpha_k) = \begin{vmatrix} \langle \alpha_1, \alpha_1 \rangle & \ldots & \langle \alpha_1, \alpha_{k-1} \rangle & \langle \alpha_1, \beta + \gamma \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \ldots & \langle \alpha_2, \alpha_{k-1} \rangle & \langle \alpha_2, \beta + \gamma \rangle \\ \vdots & \ddots & \vdots & \vdots \\ \langle \alpha_{k-1}, \alpha_1 \rangle & \ldots & \langle \alpha_{k-1}, \alpha_{k-1} \rangle & \langle \alpha_{k-1}, \beta + \gamma \rangle \\ \langle \beta + \gamma, \alpha_1 \rangle & \ldots & \langle \beta + \gamma, \alpha_{k-1} \rangle & \langle \beta + \gamma, \beta + \gamma \rangle \end{vmatrix}$$

Skoro $\gamma \in \mathrm{lin}(\alpha_1, \ldots, \alpha_{k-1})^{\perp}$ oraz $\beta \perp \gamma$, to mamy:

$$W(\alpha_1, \ldots, \alpha_k) = \begin{vmatrix} \langle \alpha_1, \alpha_1 \rangle & \ldots & \langle \alpha_1, \alpha_{k-1} \rangle & \langle \alpha_1, \beta \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \ldots & \langle \alpha_2, \alpha_{k-1} \rangle & \langle \alpha_2, \beta \rangle \\ \vdots & \ddots & \vdots & \vdots \\ \langle \alpha_{k-1}, \alpha_1 \rangle & \ldots & \langle \alpha_{k-1}, \alpha_{k-1} \rangle & \langle \alpha_{k-1}, \beta \rangle \\ \langle \beta, \alpha_1 \rangle & \ldots & \langle \beta, \alpha_{k-1} \rangle & \langle \beta, \beta \rangle + \langle \gamma, \gamma \rangle \end{vmatrix}$$

Niech $\beta = a_1\alpha_1 + \ldots + a_{k-1}\alpha_{k-1}$. Wtedy po odjęciu od n tego wiersza i-tego wiersza przemnożonego przez a_i, dla wszystkich $1 \leqslant i \leqslant k-1$, wyznacznik się nie zmienia i dostajemy:

$$W(\alpha_1, \ldots, \alpha_k) = \begin{vmatrix} \langle \alpha_1, \alpha_1 \rangle & \ldots & \langle \alpha_1, \alpha_{k-1} \rangle & \langle \alpha_1, \beta \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \ldots & \langle \alpha_2, \alpha_{k-1} \rangle & \langle \alpha_2, \beta \rangle \\ \vdots & \ddots & \vdots & \vdots \\ \langle \alpha_{k-1}, \alpha_1 \rangle & \ldots & \langle \alpha_{k-1}, \alpha_{k-1} \rangle & \langle \alpha_{k-1}, \beta \rangle \\ \langle 0, \alpha_1 \rangle & \ldots & \langle 0, \alpha_{k-1} \rangle & \langle \gamma, \gamma \rangle \end{vmatrix}.$$

Po odjęciu od n tej kolumny i-tej kolumny przemnożonej przez a_i, dla wszystkich $1 \leqslant i \leqslant k-1$, wyznacznik się nie zmienia i **ostatni wiersz się nie zmienia**, zaś ostatnia kolumna staje się *transpozycją* ostatniego wiersza, co oznacza, że:

$$W(\alpha_1, \ldots, \alpha_k) = \begin{vmatrix} G(\alpha_1, \ldots, \alpha_{k-1}) & 0 \\ 0 & \|\gamma\|^2 \end{vmatrix}.$$

□

Wniosek 14.2: Odległość od podprzestrzeni

Niech $(V, \langle\,,\rangle)$ będzie przestrzenią euklidesową liniową. Dla każdego $v \in V$ oraz dla dowolnej podprzestrzeni $W \subseteq V$ odległość $d(v, W)$ punktu v od W, czyli długość $\|v - p_W(v)\|$, gdzie $p_W(v)$ jest rzutem v na W równa jest

$$d(v, W)^2 = \frac{G(v, \alpha_1, \ldots, \alpha_k)}{G(\alpha_1, \ldots, \alpha_k)},$$

gdzie $\alpha_1, \ldots, \alpha_k$ jest bazą W.

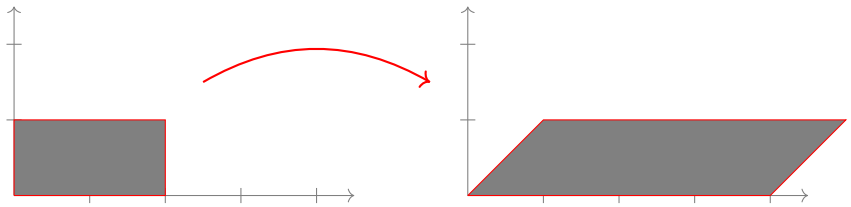
Powiedzmy o jeszcze jednym fundamentalnym dla naszej intuicji fakcie mającym olbrzymie znaczenie w interpretacji analitycznej: objętość równoległościanu zmienia przy endomorfizmie liniowym się tak, jak moduł wyznacznika tego endomorfizmu.

Obserwacja 14.3

Niech $\phi : V \to V$ będzie endomorfizmem przestrzeni euklidesowej $(V, \langle\,,\rangle)$. Dla dowolnej bazy $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ w V zachodzi równość:

$$\sqrt{W(\phi(\alpha_1), \phi(\alpha_2), \ldots, \phi(\alpha_n))} = |\det(\phi)| \cdot \sqrt{W(\alpha_1, \ldots, \alpha_n)}.$$

Przykład. Bierzemy $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ dane wzorem $\phi(x, y) = (2x + y, y)$. Równoległościan rozpięty na wektorach $(2, 0), (0, 1)$ przechodzi na równoległościan rozpięty na wektorach $(4, 0), (1, 1)$.



Wyznacznik rozważanego przekształcenia liniowego to 2.

Dowód. Zakładamy, że $\dim V = n$. Niech $\mathcal{B}$ będzie bazą ortonormalną przestrzeni $(V, \langle\,,\rangle)$. Wiemy, że:

$$G(\alpha_1, \ldots, \alpha_n) = (M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}})^T \cdot M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}},$$
$$G(\phi(\alpha_1), \ldots, \phi(\alpha_n)) = (M(\phi)_{\mathcal{A}}^{\mathcal{B}})^T \cdot M(\phi)_{\mathcal{A}}^{\mathcal{B}}.$$

Mamy też: $M(\phi)_{\mathcal{A}}^{\mathcal{B}} = M(\phi)_{\mathcal{B}}^{\mathcal{B}} \cdot M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}$, czyli $\det(M(\phi)_{\mathcal{A}}^{\mathcal{B}}) = \det(\phi) \cdot \det(M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}})$. Zatem

$$W(\alpha_1, \ldots, \alpha_n) = \det((M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}})^T \cdot \det(M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}) = (\det(M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}))^2$$
$$W(\phi(\alpha_1), \ldots, \phi(\alpha_n)) = (\det(\phi) \cdot \det(M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}))^2.$$

□

Powyższy fakt możemy przepisać w języku przekształceń afinicznych na równoległościanach.

Obserwacja 14.4

Niech $f : V \to V$ będzie przekształceniem afinicznym przestrzeni euklidesowej afinicznej $(V, (V, \langle\,,\rangle))$ w siebie. Dla dowolnego układu bazowego $(p_0; \mathcal{A}) = (p_0; \alpha_1, \ldots, \alpha_n)$ w V zachodzi równość:

$$\mu_n(R(f(p_0); f'(\alpha_1), \ldots, f'(\alpha_n))) = |\det(f')| \cdot \mu_n(R(p_0; \alpha_1, \ldots, \alpha_n)).$$

Przykład. Niech $f : \mathbb{R}^2 \to \mathbb{R}^2$ dane wzorem $f(x, y) = (2x+y-1, y)$. Równoległobok $R((1, 1); (2, 0), (0, 1))$ przechodzi na $R((2, 1); (4, 0), (1, 1))$, a jego pole zmienia się z 2 na 4, zgodnie z modułem wyznacznika pochodnej przekształcenia f.

Definicja 14.4: Miara sympleksu

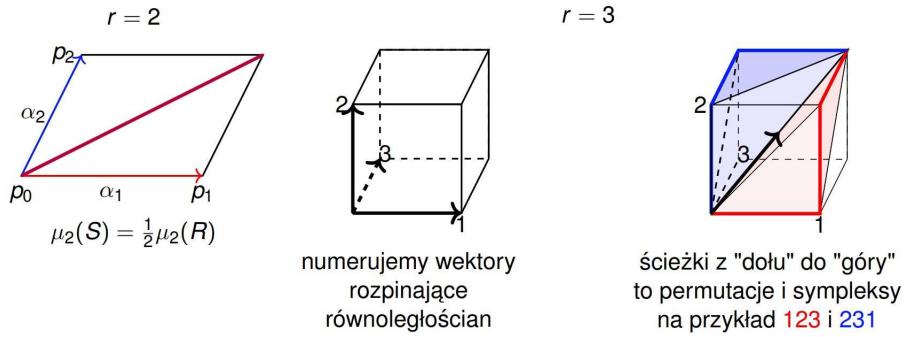
Niech $S = S(p_0, \ldots, p_k)$ będzie k-wymiarowym sympleksem w $(H, \langle , \rangle)$. Przez k-wymiarową miarę (lub k-wymiarową objętość) sympleksu S rozumiemy liczbę:

$$\mu_k(S) = \frac{1}{k!}\sqrt{W(\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k})}.$$

Skąd czynnik $1/k!$ w definicji miary k-wymiarowej sympleksu? Rozumować można na wiele sposobów. W dodatku przedstawimy szkic podejścia wywodzącego się z analizy. Przedstamy podejście oparte o definicję. Mając równoległościan $R(p_0; \alpha_1, \ldots, \alpha_k)$ możemy rozważyć wszystkie sympleksy rozpięte przez takie układy k wierzchołków tego równoległościanu, aby:

- układy te zawierały wierzchołek p_0 oraz przeciwległy, czyli $p_0 + \alpha_1 + \ldots + \alpha_k$,
- aby zbiór wektorów łączących pary wierzchołków tego układu był wybrany ze zbioru $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$

Zgodnie z naszą definicją wszystkie te sympleksy mają tę samą objętość. A jest ich przecież $k!$ Czy Czytelnik wie jak to pokazać? Wybierając taki sympleks trzeba określić w jakiej kolejności poruszamy się po wierzchołkach równoległościanu w kierunkach k wektorów $\overrightarrow{p_0p_1}, \ldots, \overrightarrow{p_0p_k}$. Innymi słowy: poruszamy się od wierzchołka $p_0 + 0 \cdot \overrightarrow{p_0p_1} + \ldots + 0 \cdot \overrightarrow{p_0p_k}$ i w pewnej kolejności zamieniamy współczynniki 0 na współczynniki 1. W ten sposób po k krokach otrzymamy wierzchołek $p_0 + \alpha_1 + \ldots + \alpha_k$. Oto przykłady:



Źródło: materiały prof. Jarosława Wiśniewskiego.

Inne podejście: k-wymiarową kostkę $[0,1]^k$ „triangulować" możemy na $k!$ sympleksów postaci[2]:

$$S_\sigma = \{(x_1, \ldots, x_k) \in \mathbb{R}^d \,|\, 0 \leqslant x_{\sigma(1)} \leqslant x_{\sigma(2)} \leqslant \ldots \leqslant x_{\sigma(k)} \leqslant 1\}, \text{ gdzie } \sigma \in S_k.$$

W dalszej części wykładu przyjrzymy się pojęciom orientacji oraz iloczynu wektorowego, stanowiących podstawowe narzędzia zaawansowanej geometrii analitycznej i różniczkowej. Mają one nie tylko ważne zastosowania w innych naukach, ale też wnoszą ważne intuicje do naszego rozumienia geometrii. Z elementarnego punktu widzenia pozwalają powiedzieć m.in. o tzw. mierze skierowanej (np. pole ze znakiem). Czytelnika zainteresowanego skarbcem ciekawych zagadnień geometrycznych dotyczącej miary odsyłam na stronę prof. Jacka Świątkowskiego (UWr): `www.math.uni.wroc.pl/~swiatkow/Repozytorium/pole.html`.

Definicja 14.5: Bazy zgodnie i przeciwnie zorientowane

Niech V będzie n-wymiarową przestrzenią nad ciałem $\mathbb{R}$. Mówimy, że bazy $\mathcal{A} = \{\alpha_1, \ldots, \alpha_n\}$, $\mathcal{B} = \{\beta_1, \ldots \beta_n\}$ przestrzeni V są:

- ZGODNIE ZORIENTOWANE, jeśli $\det M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}} > 0$,
- PRZECIWNIE ZORIENTOWANE, jeśli $\det M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}} < 0$.

[2]Ogólną teorią takich działań zajmuje się topologiczna teoria tzw. kompleksów symplicjalnych, patrz np. paragraf 5.1 w `https://ww2.ii.uj.edu.pl/~zgliczyn/dydaktyka/2013-14/top-obliczeniowa/topologia.pdf`.

Przykład. Bazy $\mathcal{A} = \{(3,2),(7,4)\}$ oraz $\mathcal{B} = \{(1,2),(1,0)\}$ przestrzeni $\mathbb{R}^2$ są zgodnie zorientowane, bo:

$$\det M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}} = \begin{vmatrix} 1 & 2 \\ 2 & 5 \end{vmatrix} > 0.$$

Relacja zgodnego zorientowania jest relacją równoważności w zbiorze wszystkich baz w przestrzeni liniowej V wymiaru n. Aby się o tym przekonać sprawdzamy trzy własności relacji równoważności.

- **Zwrotność**. Jeśli $\mathcal{A}$ jest bazą V to $M(\mathrm{id})_{\mathcal{A}}^{\mathcal{A}} = I$, czyli bazy $\mathcal{A}$ oraz $\mathcal{A}$ są zgodnie zorientowane.
- **Symetryczność**. Jeśli $\mathcal{A}, \mathcal{B}$ są bazami V to $M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}} = (M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}})^{-1}$, czyli
$$\det M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}} > 0 \Rightarrow \det M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}} > 0.$$
- **Przechodniość**. Jeśli $\mathcal{A}, \mathcal{B}, \mathcal{C}$ są bazami V to $M(\mathrm{id})_{\mathcal{A}}^{\mathcal{C}} = M(\mathrm{id})_{\mathcal{B}}^{\mathcal{C}} \cdot M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}$, czyli
$$\det M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}} > 0, \det M(\mathrm{id})_{\mathcal{B}}^{\mathcal{C}} > 0 \Rightarrow \det M(\mathrm{id})_{\mathcal{A}}^{\mathcal{C}} > 0.$$

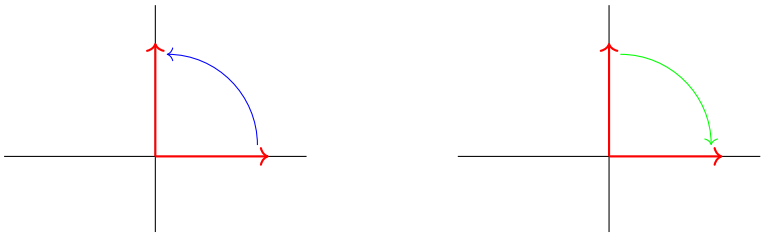
Uwaga: dla każdej bazy $\mathcal{A} = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n)$ przestrzeni V nad ciałem $\mathbb{R}$ oraz bazy $\mathcal{A}' = (\alpha_2, \alpha_1, \alpha_3, \ldots, \alpha_n)$ powstałej przez zamianę kolejności (pierwszych) dwóch wektorów, wówczas:

- bazy $\mathcal{A}$ oraz $\mathcal{A}'$ są przeciwnie zorientowane,
- każda baza przestrzeni liniowej V jest zgodnie zorientowana z $\mathcal{A}$ lub $\mathcal{A}'$.

Definicja 14.6: Orientacja

Rodzinę wszystkich baz zgodnie zorientowanych z pewną bazą przestrzeni V nad ciałem $\mathbb{R}$ nazywamy ORIENTACJĄ przestrzeni V. Mówimy, że przestrzeń V jest ZORIENTOWANA, jeśli wybrana jest jedna z jej (dwóch) orientacji. W przestrzeni zorientowanej baza $\mathcal{A}$ jest DODATNIO (UJEMNIE) ZORIENTOWANA, jeśli jest zorientowana zgodnie (przeciwnie) z wybraną orientacją przestrzeni V.

Zauważmy, że dokonanie wyboru jednej z dwóch dostępnych orientacji przestrzeni liniowej (czy afinicznej) nie jest w żaden sposób uzasadnione matematycznie. Innymi słowy — wybór orientacji prowadzi do dwóch dualnych „geometrii". Mimo wszystko wyborów tych często dokonujemy w oparciu o czynniki pozamatematyczne (np. biorąc pod uwagę ruch Ziemi czy ludzką anatomię). Ustalając orientację na płaszczyźnie mówimy o kierunku zgodnym lub przeciwnym ze wskazówkami zegara ustalamy tak naprawdę kolejność wektorów obranej bazy (np. standardowej). Jeśli płaszczyzna zorientowana jest zgodnie z bazą standardową, wówczas mówimy o orientacji przeciwnej do wskazówek zegara.



Szereg innych pojęć matematycznych można określać w oparciu o orientację, na przykład „leżenie po stronie" prostej (na płaszczyźnie) czy płaszczyzny (w przestrzeni) czy też „leżenie wewnątrz" krzywej zamkniętej. Są to pojęcia delikatne, w zależności od rozważanego obiektu, ale ważne choćby w analize czy topologii.

Po ustaleniu orientacji w przestrzeni euklidesowej możemy mówić o zorientowanej mierze n-wymiarowej. Po prostu mówimy, że objętość równoległościanu rozpiętego na bazie $\alpha_1, \ldots, \alpha_n$ wynosi $\pm\sqrt{W(\alpha_1, \ldots, \alpha_n)}$, przy czym wybór znaku zależy od tego, czy baza $\alpha_1, \ldots, \alpha_n$ jest dodatnio, czy ujemnie zorientowana. Czytelnika zainteresowanego tymi zagadnieniami w kontekście elementarnym zachęcam do lektury pracy „Zorientowane pole wielokątów" Agnieszki Burskiej[3].

Możliwość rozważania przestrzeni o zadanej (przez pewną bazę) orientacji jest istotną pomocą w rozmaitych rozważaniach geometrycznych, między innymi w czekających nas klasyfikacjach izometrii. Pojęcie orientacji nie wymaga, jak widzieliśmy, wprowadzania iloczynu skalarnego (mówimy też o orientowalności innych zbiorów, o czym wspominam w dodatku), ale ich wspólne stosowanie jest bardzo wygodne.

[3] `http://www.math.uni.wroc.pl/~swiatkow/Repozytorium/pole/prace/skan_burska_zorient_pole.pdf`.

Definicja 14.7

Załóżmy, że V jest zorientowaną przestrzenią linową skończonego wymiaru nad $\mathbb{R}$. Powiemy, że automorfizm ϕ przestrzeni V:

- ZACHOWUJE ORIENTACJĘ V, jeśli dla każdej jej (równow. pewnej) bazy dodatnio zorientowanej $(\alpha_1, \ldots, \alpha_n)$ baza $(\phi(\alpha_1), \ldots, \phi(\alpha_n))$ jest również dodatnio zorientowana,
- ZMIENIA ORIENTACJĘ V, jeśli nie zachowuje orientacji.

Łatwo sprawdzić, że automorfizm ϕ zachowuje orientację wtedy i tylko wtedy, gdy $\det \phi > 0$, a zmienia ją tylko wtedy, gdy $\det \phi < 0$. W kontekście izometrii nabiera to istotnego znaczenia, o czym powiemy dalej. Wprowadźmy jeszcze jedno pojęcie, istotne z punktu widzenia rachunkowego i dla wielu zastosowań. Mówi ono o tym, że wybór iloczynu skalarnego i orientacji w n-wymiarowej przestrzeni liniowej V nad $\mathbb{R}$ pozwala podać geometryczną metodę dopełniania liniowo niezależnego układu $n-1$ wektorów z V do dodatnio zorientowanej bazy V. [4]

Definicja 14.8: Iloczyn wektorowy

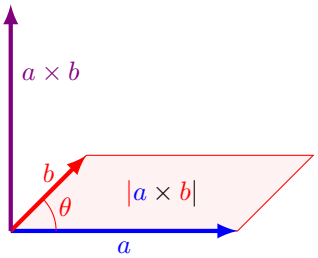
Niech $(V, \langle\,,\rangle)$ będzie n-wymiarową przestrzenią euklidesową liniową, zorientowaną (przez pewną bazę). Niech $\alpha_1, \ldots, \alpha_{n-1}$ będzie układem wektorów przestrzeni V. ILOCZYNEM WEKTOROWYM układu $\alpha_1, \ldots, \alpha_{n-1}$ nazywamy wektor β, oznaczany dalej: $\alpha_1 \times \ldots \times \alpha_{n-1}$, taki, że:

(i) jeśli układ $\alpha_1, \ldots, \alpha_{n-1}$ jest liniowo zależny, to $\beta = 0$,

(ii) jeśli układ $\alpha_1, \ldots, \alpha_{n-1}$ jest liniowo niezależny, to:

- $\beta \in \operatorname{lin}(\alpha_1, \ldots, \alpha_{n-1})^{\perp}$,
- $\|\beta\| = \sqrt{W(\alpha_1, \ldots, \alpha_{n-1})}$,
- baza $\alpha_1, \ldots, \alpha_{n-1}, \beta$ jest dodatnio zorientowana.

Zobaczmy ilustrację iloczynu wektorowego w przestrzeni trójwymiarowej.



Warto odnotować następujący wniosek dotyczący równoległościanu 2-wymiarowego rozpiętego przez układ dwóch wektorów u, v w przestrzeni trójwymiarowej. W tym przypadku długość wektora $u \times v$ to po prostu pierwiastek z wyznacznika macierzy

$$\begin{bmatrix} \langle u, u \rangle & \langle u, v \rangle \\ \langle v, u \rangle & \langle v, v \rangle \end{bmatrix},$$

czyli $\|u\|^2 \cdot \|v\|^2 - \langle u, v \rangle^2$. A zatem dostaliśmy następującą równość:

$$\|u\|^2 \|v\|^2 = \langle u, v \rangle^2 + \|u \times v\|^2.$$

Wzór ten można interpretować jako znaną nam ze szkoły jedynkę trygonometryczną. Korzystając z elementarnego wzoru na pole równoległoboku postaci $\|a\| \cdot \|b\| \cdot \sin \alpha$, gdzie α jest kątem między a oraz b, możemy przyjąć, że dla niezerowych wektorów a oraz b ów sinus jest z definicji liczbą $\|a \times b\| / \|a\| \cdot \|b\|$.

[4] Definicja ta ma również ważne motywacje fizyczne; wiele momentów fizycznych za pomocą iloczynu wektorowego definiuje się moment siły lub moment pędu. Jest ono szczególnie przydatne w geometrii rzeczywistej przestrzeni trójwymiarowej.

Przykład. W przestrzeni liniowej $(\mathbb{R}^3, \langle\,,\,\rangle)$ mamy orientację wyznaczoną przez (przeciwnie zorientowaną do standardowej) bazę $((0,1,0),(1,0,0),(0,0,1))$. Niech $\alpha_1 = (1,0,2)$ oraz $\alpha_2 = (3,2,1)$. Wówczas:

$$\operatorname{lin}((1,0,2),(3,2,1))^{\perp} = \operatorname{lin}((4,-5,-2)), \quad \sqrt{W(\alpha_1,\alpha_2)} = \sqrt{45}.$$

Wektor $(4,-5,-2)$ jest prostopadły do α_1, α_2, ma długość $\sqrt{45}$ i baza $\alpha_1, \alpha_2, (4,-5,-2)$ jest przeciwnie zorientowana z bazą standardową, a więc zgodna z wybraną orientacją. Mamy więc $\alpha_1 \times \alpha_2 = (4,-5,-2)$.

Przykład. W przestrzeni liniowej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym zorientowanej zgodnie z bazą standardową szukamy wektora α takiego, że $(1,-1,0) \times \alpha = (1,1,0)$.

Jest to na przykład wektor $\alpha = (-\frac{1}{2}, -\frac{1}{2}, -\frac{1}{\sqrt{2}})$. Jest on prostopadły do $(1,1,0)$ i nie jest wielokrotnością $(1,-1,0)$. Ma on normę 1, a więc

$$\|(1,-1,0) \times \alpha\| = \|(1,1,0)\|.$$

Wreszcie, układ wektorów $\mathcal{A} = ((1,-1,0),(-\frac{1}{2},-\frac{1}{2},-\frac{1}{\sqrt{2}}),(1,1,0))$ jest bazą $\mathbb{R}^3$ oraz

$$M(\mathrm{id})^{st}_{\mathcal{A}} = \begin{bmatrix} 1 & -\frac{1}{2} & 1 \\ -1 & -\frac{1}{2} & 1 \\ 0 & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Skoro $\det M(\mathrm{id})^{st}_{\mathcal{A}} > 0$, to baza $\mathcal{A}$ jest dodatnio zorientowana i rzeczywiście α spełnia równanie wyżej.

Przykład. W przestrzeni liniowej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym zorientowanej zgodnie z bazą standardową znaleźć równanie płaszczyzny L zawierającej punkt $(0,0,1)$, która przechodzi przez punkty $a = (1,1,3), b = (4,1,1), c = (5,2,0)$.

Zauważmy, że $T(L)$ jest płaszczyzną opisaną przez równanie $a_1x_1 + a_2x_2 + a_3x_3 = 0$ i niezerowy wektor (a_1,a_2,a_3) wyznacza ją z dokładnością do stałej. Jest to wektor prostopadły do $T(L)$. Możemy więc przyjąć $(a_1,a_2,a_3) = \overrightarrow{ab} \times \overrightarrow{ac}$. Gdy wyznaczymy ten wektor pozostanie wyznaczyć równanie postaci $a_1x_1 + a_2x_2 + a_3x_3 = d$ płaszczyzny L przez wstawienie do otrzymanego równania punktu $(0,0,1)$.

Obserwacja 14.5

Niech $(V, \langle\,,\,\rangle)$ będzie trójwymiarową zorientowaną przestrzenią euklidesową liniową oraz niech $\mathcal{B} = (\beta_1, \beta_2, \beta_3)$ będzie jej bazą ortonormalną, dodatnio zorientowaną. Dla dowolnych wektorów $\alpha = a_1\beta_1 + a_2\beta_2 + a_3\beta_3$ oraz $\beta = b_1\beta_1 + b_2\beta_2 + b_3\beta_3$. zachodzi

$$\alpha \times \beta = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \beta_1 - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \beta_2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \beta_3.$$

Dowód. Niech γ będzie wektorem zdefiniowanym w tezie naszej obserwacji (po prawej stronie ostatniej równości). Mamy

$$\langle \alpha, \gamma \rangle = a_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - a_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + a_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \end{vmatrix} = 0.$$

Analogicznie $\langle \beta, \gamma \rangle = 0$. Stąd $\gamma \in \operatorname{lin}(\alpha,\beta)^{\perp}$. Mamy dalej:

$$\begin{aligned} \|\gamma\|^2 &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2 = \\ &= (a_2b_3 - a_3b_2)^2 + (a_1b_3 - a_3b_1)^2 + (a_1b_2 - a_2b_1)^2 = \\ &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 = \\ &= \langle \alpha, \alpha \rangle \cdot \langle \beta, \beta \rangle - \langle \alpha, \beta \rangle^2 = W(\alpha, \beta). \end{aligned}$$

Gdy α, β są liniowo niezależne, to dla $\mathcal{A} = (\alpha, \beta, \gamma)$ oraz $\mathcal{B} = (\beta_1, \beta_2, \beta_3)$:

$$\det M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}} = \begin{vmatrix} a_1 & b_1 & \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \\ a_2 & b_2 & -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \\ a_3 & b_3 & \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ a_2 & b_2 & -\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} \\ a_3 & b_3 & \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \end{vmatrix},$$

co po rozwinięciu względem trzeciej kolumny daje:

$$\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2 > 0,$$

bo inaczej $(a_1, a_2, a_3), (b_1, b_2, b_3)$ będą liniowo zależne. Zatem układ α, β, γ tworzy bazę dodatnio zorientowaną (zorientowaną zgodnie z bazą $\beta_1, \beta_2, \beta_3$). Stąd $\alpha \times \beta$ jest równy γ. □

Iloczyn wektorowy posiada także szereg ciekawych własności algebraicznych, które warto sprawdzić.

Obserwacja 14.6

Niech $u, v, w \in \mathbb{R}^3$ oraz niech $a \in \mathbb{R}$. Wtedy ma miejsce:

- antysymetria iloczynu wektorowego: $u \times v = -v \times u$,
- formuła Lagrange'a:
$$u \times (v \times w) = \langle u, w \rangle \cdot v - \langle u, v \rangle \cdot w,$$
- tożsamość Jacobiego (imitująca łączność):
$$u \times (v \times w) + v \times (w \times u) + w \times (u \times v) = 0.$$

Poniższy rezultat wiąże objętość równoległościanu z iloczynem wektorowym w przestrzeni trójwymiarowej. Wynik ten stanowi jednocześnie popularne kryterium sprawdzania czy proste o określonych kierunkach leżą na jednej płaszczyźnie.

Obserwacja 14.7

Jeśli $\alpha = (a_1, a_2, a_3), \beta = (b_1, b_2, b_3), \gamma = (c_1, c_2, c_3)$ należą do $(\mathbb{R}^3, \langle , \rangle_{st})$, to miara 3-wymiarowa równoległościanu $R(\alpha, \beta, \gamma)$ równa jest

$$|\langle \alpha \times \beta, \gamma \rangle| = \left| \det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \right|.$$

Dowód. Jak już wiemy objętość równoległościanu równa jest iloczynowi pola podstawy oraz wysokości. Podstawą niech będzie równoległościan rozpięty przez wektory α, β. Jego miara 2-wymiarowa to $\|\alpha \times \beta\|$. Szukamy wysokości tego równoległościanu. Wektor $\alpha \times \beta$ jest prostopadły do każdego z wektorów α, β, więc wysokość jest odcinkiem równoległym do $\alpha \times \beta$. A zatem rzutujemy γ na $\mathrm{lin}(\alpha \times \beta)$. Ten rzut to wektor

$$\frac{\langle \alpha \times \beta, \gamma \rangle}{\|\alpha \times \beta\|^2} \cdot (\alpha \times \beta).$$

Jego długość, to oczywiście $\frac{|\langle \alpha \times \beta, \gamma \rangle|}{\|\alpha \times \beta\|}$. Pokazaliśmy pierwszą część tezy.

Wzór wypisany na końcu wynika z rozwinięcia wyznacznika względem trzeciego wiersza. □

14.1 Uzupełnienie. Orientowalność i transformacje ciągłe

Badanie macierzy nie ogranicza się jedynie do algebry liniowej. W dalszych etapach studiów poznacie Państwo także ich „grupową" i „pierścieniową" naturę. Tak się składa, że podejście „grupowe" pojawiło się na naszym wykładzie. Warto więc dodać kilka komentarzy.

Definicja 14.9: Grupa macierzy odwracalnych

Niech K będzie ciałem oraz niech n będzie dodatnią liczbą całkowitą. Przez $GL(n, K)$ oznaczamy podzbiór $M_n(K)$ złożony z macierzy odwracalnych.

Oznaczenie wprowadzone wyżej nie pojawia się na kursie, bowiem macierze $GL(n, K)$ nie tworzą podprzestrzeni. Tworzą one jednak grupę ze względu na mnożenie, co oznacza, że

- złożenie macierzy odwracalnych jest macierzą odwracalną,
- każda macierz z $GL(n, K)$ ma macierz odwrotną,
- macierz $I \in GL(n, K)$ ma tę własność, że $AI = IA = A$, czyli jest to element neutralny mnożenia.

Zauważmy jeszcze, że zbiór $GL(n, K)$ utożsamić można ze zbiorem izomorfizmów przestrzeni K^n.

W kontekście naszych geometrycznych rozważań, a więc w kontekście przestrzeni nad $\mathbb{R}$, warto wyróżnić macierze odwracalne o dodatnim wyznaczniku i macierze o ujemnym wyznaczniku. Pierwsze oznaczamy przez $GL_+(n, \mathbb{R})$, a drugie przez $GL_-(n, \mathbb{R})$. Mamy $GL(n, \mathbb{R}) = GL_+(n, \mathbb{R}) \cup GL_-(n, \mathbb{R})$.

Kluczowe obserwacje to algebraiczne sformułowania obserwacji poczynionych w trakcie wykładu:

- Iloczyn dowolnych macierzy o dodatnim wyznaczniku jest kolejną macierzą o dodatnim wyznaczniku. Również odwrotność macierzy z $GL_+(n, \mathbb{R})$ należy do tego zbioru. Jest w nim również I, co oznacza, że $GL_+(n, \mathbb{R})$ jest tak zwaną podgrupą $GL(n, \mathbb{R})$.
- Iloczyn macierzy z $GL_+(n, \mathbb{R})$ oraz $GL_-(n, \mathbb{R})$ należy do $GL_-(n, \mathbb{R})$.
- Iloczyn macierzy z $GL_-(n, \mathbb{R})$ to macierz z $GL_+(n, \mathbb{R})$.

Dostajemy następujący wniosek. Istnieje funkcja $f : GL(n, \mathbb{R}) \to \{-1, 1\}$, która spełnia warunek:[5]

$$f(AB) = f(A)f(B).$$

Dlaczego to jest takie ważne? Ta dwuelementowa grupa $\{-1, 1\}$ reprezentuje dwie różne orientacje, jakie mogą pojawić się w przestrzeni euklidesowej. Gdyby miał istnieć jakiś inny sposób definicji orientacji, to mimo wszystko oczekiwalibyśmy, że relacja baz wspólnie zorientowanych będzie relacją równoważności oraz, że będzie miała jakiś związek z mnożeniem macierzy. Nawet więcej – chcielibyśmy, aby klasy równoważności relacji wspólnego zorientowania tworzyły grupę.

To nie są argumenty bardzo precyzyjnie, ale można powiedzieć jedno: nie da się wprowadzić nietrywialnej (nie jednoelementowej) orientacji w oparciu o homomorfizm z $GL(n, \mathbb{R})$ do pewnej skończonej grupy innej niż $\{-1, 1\}$. Dowiecie się Państwo tego na Topologii, odwołując się do ważnego pojęcia `drogi`.

Definicja 14.10: Droga nieosobliwa

Niech $A, B \in GL(n, \mathbb{R})$ będą odwracalne. Przyporządkowanie: $f : [0, 1] \to M_n(K)$ nazwiemy DROGĄ NIEOSOBLIWĄ w $M_n(\mathbb{R})$ łączącą A, B, jeśli

- $f(0) = A, f(1) = B$,
- dla $t \in [0, 1]$ macierz $f(t)$ jest odwracalna,
- jeśli $f(t) = [a_{ij}(t)]$, to funkcje $t \mapsto a_{ij}(t)$ są ciągłe.

[5] Funkcje spełniające równość obok nazywamy homomorfizmami grup. Zbiór $\{-1, 1\}$ z definicją mnożenia jest grupą.

Przykład. Funkcja $[0,1] \ni t \mapsto \begin{bmatrix} 1+t & t \\ 0 & 1 \end{bmatrix}$ jest drogą nieosobliwą łączącą $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ oraz $B = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$.

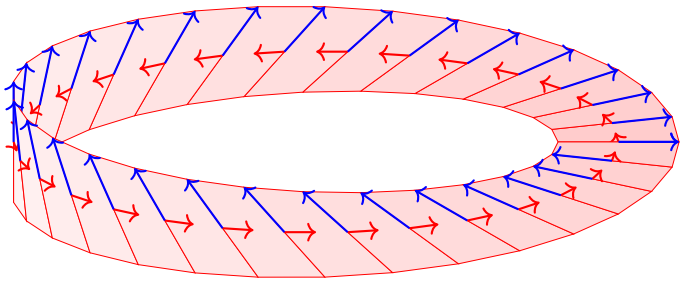
Zachodzi następujące twierdzenie.

Obserwacja 14.8

Baza $(A_1, \ldots, A_n)$ jest zorientowana zgodnie z bazą standardową $(E_1, \ldots, E_n)$ (w $\mathbb{R}^n$) wtedy i tylko wtedy, gdy istnieje nieosobliwa droga pomiędzy macierzami $[A_1, \ldots, A_n]$ oraz $[E_1, \ldots, E_n]$,

A zatem widzimy, że zgodność orientacji dwóch baz ma charakter wybitnie geometryczny. Dowód tego wyniku znaleźć można na przykład w skrypcie prof. Chabera i Pola w punkcie 14.3. Rozumowanie wymaga zastosowania macierzy ortogonalnych, o których dowiemy się za kilka wykładów. Warto odnotować, że w matematyce mówi się również o orientowalności rozmaitych obiektów, nie mających struktury przestrzeni liniowej. Jest to związane z pojęciem przestrzeni stycznej, które stosować można np. dla powierzchni lub rozmaitości gładkich.

W myśl takiego podejścia obiekt jest orientowalny jeśli można na zbiorze przestrzeni stycznych do wszystkich jego punktów wprowadzić orientacje, które będzie można w sposób ciągły deformować. Rozróżnienie kiedy powierzchnia jest orientowalna, a kiedy nie, będzie jednym z tematów, który rozważać będziecie Państwo w dalszych studiach.



Przykład powierzchni nieorientowalnej – wstęgi Möbiusa. Czy Czytelnik widzi jaki jest związek z zacytowanym wynikiem?
Źródło: materiały prof. Jarosława Wiśniewskiego.

Z orientowalnością wiąże się wiele ważnych tematów. Oto prosty przykład: na płaszczyźnie litera L oraz jej symetryczne odwrócenie można traktować jako obiekty różne, bo nie można ich nałożyć na siebie (za pomocą izometrii) bez wyjścia poza płaszczyznę (dlaczego?).

Dla Płaszczaka – mieszkańca Flatlandii[6] – niemogącego wyobrazić sobie ruchu w trzech wymiarach – powyższe obiekty mogą wydawać się być różnych kształtów. Jedną może nazwać lewą L, a drugą – prawą. Jeśli jednak znałby przestrzeń trójwymiarową i mógł w niej umieścić obydwie te figury – byłby w stanie za pomocą obrotu zamienić jedną w drugą. Dla mieszkańca Flatlandii byłby to szok – z jego punktu widzenia figura zmieniłaby kształt.

Podobnie jest z rozróżnianiem lewej ręki od prawej – status lewej i prawej ręki w przestrzeni trójwymiarowej jest identyczny, jak owych liter w przykładzie wyżej. Poprzez odpowiednie przekształcenie w przestrzeni czterowymiarowej możemy nałożyć rękę lewą na prawą. Może to budzić pewne zdziwienie. Jak sobie wyobrazić taki ruch?

We Flatlandii nie jest możliwe przesunięcie dwóch płaskich odwróconych plecami L tak, by dostać literę. Ale jest to możliwe na dwuwymiarowej, nieorientowalnej powierzchni – wstędze Möbiusa (patrz rysunek wyżej). Jest ona niezanurzalna w liniowej przestrzeni dwuwymiarowej (i dowód też Państwo poznacie).

[6]Serdecznie polecam wydaną w Polsce XIX-wieczną klasykę popularyzacji matematyki: *Flatlandia, czyli Kraina Płaszczaków* autorstwa Edwina Abbotta.

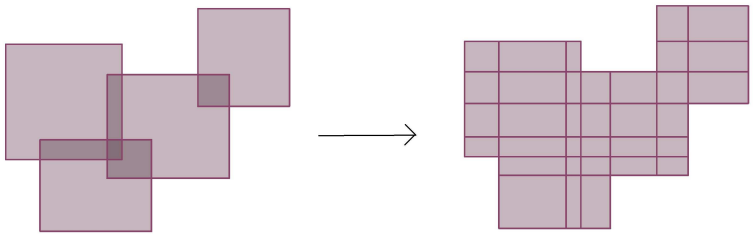
14.2 Dodatek. Miara Jordana i zasada Cavalieriego

Materiał przedstawiony w tym dłuższym niż zwykle dodatku, ma na celu wprowadzenie Państwa (bez szczegółów) w jedną z kluczowych obserwacji analitycznych prowadzących do wyznaczania objętości zbiorów... no właśnie, jakich? Zbiory te nazwiemy w tym dodatku mierzalnymi w sensie[7] Jordana. Chciałbym pokazać jak wprowadzona przez nas definicja miary równoległościanu wiąże się z definicją miary sympleksu, oraz innych obiektów. Wyjaśnienie podane na końcu wykładu zyska w ten sposób szerszy kontekst.

Naszym celem jest określenie miary Jordana. Jest to sposób wyznaczania miary n-wymiarowej μ_n w przestrzeni $\mathbb{R}^n$, którą przypisujemy coraz bardziej skomplikowanym podzbiorom przestrzeni afinicznej, począwszy od równoległościanów. Robimy to w trzech krokach.

- **Krok 1.** Bierzemy przestrzeń $\mathbb{R}^n$ i ustalamy w niej bazę $\mathcal{A}$. Niech $P_1, \ldots, P_s$ będzie zbiorem równoległościanów rozpiętych przez skalarne wielokrotności wektorów z $\mathcal{A}$ o rozłącznych wnętrzach. Wówczas określamy $\mu_n(P_1 \cup \ldots \cup P_s) = \mu_n(P_1) + \ldots + \mu_n(P_s)$.

- **Krok 2.** Rozważamy zbiór równoległościanów $P_1, \ldots, P_s$ rozpiętych przez skalarne wielokrotności wektorów z $\mathcal{A}$, ale nie zakładamy, że ich wnętrza są rozłączne.



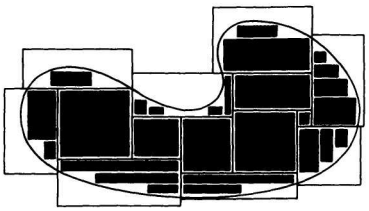
Przykład dekompozycji zbioru prostokątów o bokach równoległych do ustalonej bazy $\mathbb{R}^2$.
Źródło: Wikipedia.

Nietrudno widzieć, że można dokonać ich dekompozycji na sumę prostokątów $Q_1, \ldots, Q_r$ o rozłącznych wnętrzach, jak w poprzednim kroku. Okazuje się, że niezależnie od wybranej dekompozycji wartość $\mu_n(Q_1 \cup \ldots \cup Q_r)$ jest taka sama, co pozwala rozszerzyć definicję funkcję μ_n na zbiory $P_1 \cup \ldots \cup P_s$.

Uzyskaliśmy zatem nową klasę zbiorów, na której mamy miarę μ_n – są to zbiory równoległościanów równoległych do ustalonej bazy $\mathbb{R}^n$. Czas na argument analityczny.

- **Krok 3.** Bierzemy podzbiór **ograniczony** $\Omega \subset \mathbb{R}^n$, to znaczy: Ω jest zawarty w pewnym równoległościanie.

  – Przez **miarę zewnętrzną** zbioru Ω, ozn. $\overline{\mu_n(\Omega)}$, określamy infimum po wszystkich sumach skończonych $\mu_n(P_1)+\ldots+\mu_n(P_s)$, gdzie $P_1, \ldots, P_s$ są równoległościanami (jak w poprzednich krokach) takimi, że $\Omega \subseteq P_1 \cup \ldots \cup P_n$.



Źródło: K. Spindler Abstract Algebra With Applications, Chapmann & Hall.

  – Przez **miarę wewnętrzną** zbioru Ω, ozn. $\underline{\mu_n(\Omega)}$, określamy supremum po wszystkich sumach skończonych $\mu_n(Q_1)+\ldots+\mu_n(Q_r)$, gdzie $\overline{Q_1, \ldots, Q_r}$ są równoległościanami (jak w poprzednich krokach) takimi, że: $\Omega \supseteq Q_1 \cup \ldots \cup Q_r$.

[7] W trakcie studiów poznają Państwo różne klasy zbiorów mierzalnych, zarówno w ramach kursu analizy, jak i teorii miary czy rachunku prawdopodobieństwa (najważniejszym będzie miara Lebesgue'a).

Definicja 14.11: Zbiór mierzalny w sensie Jordana

Mówimy, że zbiór ograniczony $\Omega \subseteq \mathbb{R}^n$ jest **mierzalny w sensie Jordana**, jeśli

$$\overline{\mu_n(\Omega)} = \underline{\mu_n(\Omega)}.$$

Liczbę powyżej nazywamy **miarą Jordana** zbioru Ω, ozn. $\mu_n(\Omega)$.

Mówiąc równoważnie: ograniczony zbiór $\Omega \subseteq \mathbb{R}^n$ jest mierzalny w sensie Jordana jeśli dla każdego $\epsilon > 0$ istnieją równoległościany $P_1, \ldots, P_m$ oraz $R_1, \ldots, R_n$ o rozłącznych wnętrzach rozpięte przez skalarne wielokrotności wektorów z bazy $\mathcal{A}$, że

$$P_1 \cup \ldots \cup P_r \subseteq \Omega \subseteq R_1 \cup \ldots \cup R_s \quad \text{oraz} \quad 0 \leqslant \sum_{j=1}^{s} \mu_n(R_j) - \sum_{i=1}^{r} \mu_n(P_i) < \epsilon.$$

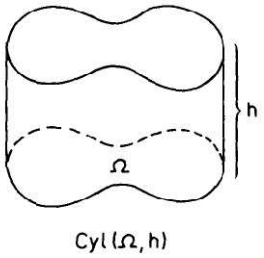
Czytelnika zainteresowanego szczegółami odsyłam do świetnego tekstu dra Michała Krycha: *Elementy teorii całkowania*, `https://www.mimuw.edu.pl/~krych/staszic/skrypt27-calki_D.pdf`.

Wymieńmy kilka własności miary, którą określiliśmy jako rozszerzenie zdefiniowanej na wykładzie miary.

- Suma skończenie wielu mierzalnych podzbiorów $\mathbb{R}^n$ w sensie Jordana jest mierzalna w sensie Jordana, ale suma nieskończenie wielu – już niekoniecznie.
- Można pokazać, że dla zbiorów mierzalnych $\Omega_1 \subseteq \Omega_2$ mamy $\mu_n(\Omega_1) \leqslant \mu_n(\Omega_2)$.

Uogólnienia poznanych przez nas rezultatów, w tym niezwykle istotna zasada Cavalieriego.

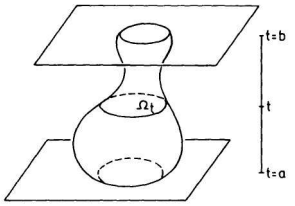
- **Twierdzenie.** Niech $U \subseteq (\mathbb{R}^n, \langle , \rangle_{st})$ będzie podprzestrzenią afiniczną wymiaru $n-1$, czyli tzw. hiperpłaszczyzną, oraz niech $v \in U^{\perp}$, gdzie $\|v\| = 1$. Niech μ_n oraz μ_{n-1} będą miarami Jordana na $\mathbb{R}^n$ oraz na U. Dla dowolnego podzbioru $\Omega \subseteq U$ mierzalnego w sensie Jordana oraz dodatniej liczby $h > 0$ określamy **cylinder nad Ω wysokości** h jako zbiór $Cyl(\Omega, h) := \{x + tv \in \mathbb{R}^n \mid x \in \Omega, 0 \leqslant t \leqslant h\}$.



Źródło: K. Spindler Abstract Algebra With Applications, Chapmann & Hall.

Wówczas $\mu_n(Cyl(\Omega, h)) = h \cdot \mu_{n-1}(\Omega)$.

- **Zasada Cavalieriego (1635)**. Niech Ω będzie mierzalnym (w s. J.) podzbiorem $(\mathbb{R}^n, \langle , \rangle_{st})$, niech U będzie hiperpłaszczyzną i niech $n_0 \in U^{\perp}$ ma długość 1. Dla $t \in \mathbb{R}$ niech $\Omega_t := \Omega \cap (tn_0 + U)$.



Źródło: K. Spindler Abstract Algebra With Applications, Chapmann & Hall.

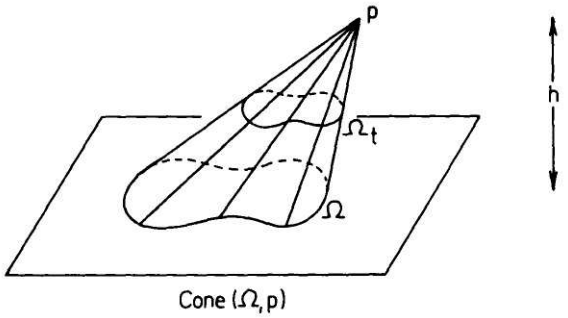
Jeśli Ω_t jest mierzalny (w sensie Jordana) dla $a \leqslant t \leqslant b$ oraz pusty dla pozostałych t, to:

$$\mu_n(\Omega) = \int_a^b \mu_{n-1}(\Omega_t) dt.$$

Czytelnik zechce wpisać w dowolną wyszukiwarkę hasło „zasada Cavalieriego”, a przekona się o jej licznych i pięknych zastosowaniach. Można też obejrzeć wiele przyjemnych filmów, np. `https://youtu.be/0xCe6LGWWjU`. Cytowany wyżej tekst da Państwu ścisły pogląd na te zagadnienia.

- **Wniosek.** Niech U będzie hiperpłaszczyzną w $(\mathbb{R}^n, \langle\,,\rangle_{st})$ i niech Ω będzie mierzalnym (w s. J) podzbiorem U. Dla każdego elementu $p \in V \setminus U$ przez **stożek nad** Ω określamy zbiór:

$$Cone(\Omega, p) := \{(1-t)x + tp \in \mathbb{R}^n \,|\, x \in \Omega, 0 \leqslant t \leqslant 1\},$$



Źródło: K. Spindler Abstract Algebra With Applications, Chapmann & Hall.

czyli jest to suma wszystkich odcinków łączących elementy Ω z p. Jeśli h jest odległością p od U, to

$$\mu_n(Cone(\Omega, p)) = \frac{h}{n}\mu_{n-1}(\Omega).$$

Wykażmy ostatni wniosek. Zgodnie z zasadą Cavalieriego rozważamy zbiór

$$\Omega_t = \Omega \cap (tn_0 + U),$$

gdzie $0 \leqslant t \leqslant h$. Jest to obraz Ω przy jednokładności o środku w p o skali $\frac{h-t}{h}\Omega$. Zgodnie z twierdzeniem o zmianie miary przy przekształceniu afinicznym

$$\mu_{n-1}(\Omega_t) = \left(\frac{h-t}{h}\right)^{n-1} \mu_{n-1}(\Omega)$$

(zobacz macierz tej jednokładności). Zgodnie z zasadą Cavalieriego mamy zatem $\mu_n(Cone(\Omega, p))$ równe jest:

$$\int_0^h \left(\frac{h-t}{h}\right)^{n-1} \mu_{n-1}(\Omega)dt = h \cdot \mu_{n-1}(\Omega) \int_0^1 \tau^{n-1}d\tau = \frac{h}{n} \cdot \mu_{n-1}(\Omega).$$

Obserwacja 14.9: Wniosek – pole, wzór 1.

Niech $p_0, \ldots, p_n \in (\mathbb{R}^n, (V, \langle\,,\rangle)_{st})$ i niech $S = S(p_0, \ldots, p_n)$. Niech też $\alpha_i = \overrightarrow{p_0p_i}$, dla $i = 1, \ldots, n$. Wówczas[a]:

$$\mu_n(S) = \frac{1}{n!}\sqrt{W(\alpha_1, \ldots, \alpha_n)} \overset{*}{=} \frac{1}{n!}|\det \begin{bmatrix} | & | & & | \\ p_0 & p_1 & \ldots & p_n \\ | & | & & | \\ 1 & 1 & & 1 \end{bmatrix}|.$$

przy czym w kolumnach macierzy z prawej strony $(*)$ są współrzędne $p_0, \ldots, p_n$ w pewnym układzie bazowym $p_0, \alpha_1, \ldots, \alpha_n$ przestrzeni $\mathbb{R}^n$.

[a] **Szczególny przypadek.** Dla standardowego układu bazowego w $\mathbb{R}^2$ dostajemy wzór na pole S trójkąta o wierzchołkach w punktach $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ oraz warunek konieczny i wystarczający do współliniowości tych punktów. Dostajemy także równanie prostej przechodzącej przez dwa zadane punkty.

Dowód. Odejmując pierwszą kolumnę od wszystkich pozostałych dostajemy łatwo jedną z równości:

$$\det\begin{bmatrix} | & | & & | \\ p_0 & p_1 & \dots & p_n \\ | & | & & | \\ 1 & 1 & & 1 \end{bmatrix}^2 = \det\begin{bmatrix} | & | & & | \\ p_0 & \overrightarrow{p_0p_1} & \dots & \overrightarrow{p_0p_n} \\ | & | & & | \\ 1 & 0 & & 0 \end{bmatrix}^2 =$$

$$= \det\begin{bmatrix} | & & | \\ \alpha_1 & \dots & \alpha_n \\ | & & | \end{bmatrix}^T \cdot \det\begin{bmatrix} | & & | \\ \alpha_1 & \dots & \alpha_n \\ | & & | \end{bmatrix} =$$

$$= \det\begin{bmatrix} \langle\alpha_1,\alpha_1\rangle & \dots & \langle\alpha_1,\alpha_n\rangle \\ \vdots & \ddots & \vdots \\ \langle\alpha_n,\alpha_1\rangle & \dots & \langle\alpha_n,\alpha_n\rangle \end{bmatrix} = W(\alpha_1,\dots,\alpha_n)^2.$$

Indukcja ze względu na n. Dla $n=1$ mamy z definicji (równoległościanu) $\mu_1(S(p_0,p_1)) = \sqrt{W(\alpha_1)}$.

Rozważmy zatem sympleks rozpięty na afinicznie niezależnym układzie $p_0,\dots,p_{n+1}$. punktów w $\mathbb{R}^n$. Wiemy też (poprzedni fakt o mierze n-wymiarowej stożka), że

$$\mu_{n+1}(S(p_0,\dots,p_{n+1})) = \frac{h}{n+1}\mu_n(S(p_0,\dots,p_n)).$$

A zatem możemy wstawić uzyskaną wyżej formułę na $\mu_n(S(p_0,\dots,p_n))$ i dostajemy dwa przedstawienia szukanej objętości $\mu_{n+1}(S(p_0,\dots,p_{n+1}))$:

$$\frac{h}{(n+1)!}|\det\begin{bmatrix} | & | & & | \\ p_0 & p_1 & \dots & p_n \\ | & | & & | \\ 1 & 1 & & 1 \end{bmatrix}| = \frac{1}{(n+1)!}|\det\begin{bmatrix} | & | & & | & | \\ p_0 & p_1 & \dots & p_n & * \\ | & | & & | & | \\ 0 & 0 & \dots & 0 & h \\ 1 & 1 & & 1 & 1 \end{bmatrix}|.$$

Powyższa równość wynika z rozwinięcia Laplace'a. Twierdzimy, że macierz (której wyznacznik liczymy) po prawej jest macierzą współrzędnych punktów $p_0,\dots,p_n,p_{n+1}$ w pierwotnej bazie punktowej $p_0;\alpha_1,\dots,\alpha_n$ uzupełnionej o wektor $\frac{1}{h}\overrightarrow{p_{n+1}p_0}$. Mamy wtedy

$$p_{n+1} = p_0 + h\frac{1}{h}\overrightarrow{p_{n+1}p_0}$$

zaś punkty $p_0,\dots,p_n$ mają w rozszerzonym układzie identyczne n współrzędnych, jak w pierwotnym (i ostatnie równe 0). □

Widzicie zatem Państwo, że od dość ogólnych abstrakcyjnych rozważań przeszliśmy do bardzo konkretnych zastosowań. Jest ich więcej.

Obserwacja 14.10: Wniosek – pole, wzór 2.

Niech $p_0,\dots,p_n \in (\mathbb{R}^n,(V,\langle\,,\rangle)_{st})$ i niech $S = S(p_0,\dots,p_n)$. Niech też $\alpha_i = \overrightarrow{p_0p_i}$, dla $i=1,\dots,n$ oraz przyjmijmy $d_{ij} = \|p_i - p_j\|$, dla $0 \leqslant i,j \leqslant n$. Wówczas:

$$\mu_n(S)^2 = \frac{1}{2^n\cdot(n!)^2}\cdot|\det\begin{bmatrix} 0 & d_{01}^2 & \dots & d_{0n}^2 & 1 \\ d_{10}^2 & 0 & \dots & d_{1n}^2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{n0}^2 & d_{n1}^2 & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}|^2.$$

Przypadek szczególny. Wzór Herona. Niech $\triangle$ będzie trójkątem o bokach długości a,b,c i niech V będzie polem $\triangle$. Dla $p=(a+b+c)/2$ mamy:

$$V^2 = \frac{1}{2^2\cdot(2!)^2}\cdot|\det\begin{bmatrix} 0 & a^2 & c^2 & 1 \\ a^2 & 0 & b^2 & 1 \\ c^2 & b^2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}| = \frac{|-16p(p-a)(p-b)(p-c)|}{16}.$$

Dowód. Już wiemy, że

$$n! \cdot \mu_n(S) = |\det \begin{bmatrix} | & \dots & | \\ p_0 & \dots & p_n \\ | & \dots & | \\ 1 & \dots & 1 \end{bmatrix}| = |\det \begin{bmatrix} | & \dots & | \\ p_0 & \dots & p_n \\ | & \dots & | \\ 1 & \dots & 1 \end{bmatrix}^T|.$$

Możemy też zmodyfikować te macierze, nie zmieniając wyznaczników, i dostać:

$$n! \cdot \mu_n(S) = |\det \begin{bmatrix} - & p_0^T & - & 1 & 0 \\ \vdots & & \vdots & \vdots & \\ - & p_n^T & - & 1 & 0 \\ - & 0^T & - & 0 & 1 \end{bmatrix}| = |\det \begin{bmatrix} | & \dots & | & | \\ p_0 & \dots & p_n & 0 \\ | & \dots & | & | \\ 0 & \dots & 0 & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}^T|.$$

Mnożymy dwie macierze i na chwilę przypominamy sobie, że p_i to wektory. Zatem $(n! \cdot \mu_n(S))^2$ jest równe:

$$|\det \begin{bmatrix} \langle p_0, p_0\rangle & \dots & \langle p_0, p_n\rangle & 1 \\ \vdots & & \vdots & \vdots \\ \langle p_n, p_0\rangle & \dots & \langle p_n, p_n\rangle & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}| = 2^{-n} \cdot |\det \begin{bmatrix} 2\langle p_0, p_0\rangle & \dots & 2\langle p_0, p_n\rangle & 1 \\ \vdots & & \vdots & \vdots \\ 2\langle p_n, p_0\rangle & \dots & 2\langle p_n, p_n\rangle & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}|.$$

Dostaliśmy tę równość mnożąc pierwsze $n+1$ wierszy przez 2 i jednocześnie dzieląc ostatnią kolumnę przez $\frac{1}{2}$. Teraz upraszczamy ten wyznacznik. Dla każdego k, gdzie $0 \leqslant k \leqslant n$ odejmujemy $\|p_k\|^2$-razy ostatni wiersz od $(k+1)$-wszego wiersza, a jednocześnie odejmujemy $\|p_k\|^2$-razy ostatnią kolumnę od kolumny $(k+1)$-wszej. Aby ogarnąć to, co powstanie wprowadzamy oznaczenia:

$$\begin{aligned} a_{ij} &:= 2\langle p_i, p_j\rangle - \|p_i\|^2 \\ b_{ij} &:= 2\langle p_i, p_j\rangle - \|p_i\|^2 - \|p_j\|^2 = -\|p_i - p_j\|^2 = -d_{ij}^2. \end{aligned}$$

Dostajemy (pamiętając, że $d_{ii} = \|p_i - p_i\| = 0$):

$$\begin{aligned} 2^n \cdot (n!)^2 \mu_n(S)^2 &= |\det \begin{bmatrix} 2\langle p_0, p_0\rangle & \dots & 2\langle p_0, p_n\rangle & 1 \\ \vdots & & \vdots & \vdots \\ 2\langle p_n, p_0\rangle & \dots & 2\langle p_n, p_n\rangle & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}| = \\ &= |\det \begin{bmatrix} a_{00} & \dots & a_{0n} & 1 \\ \vdots & & \vdots & \vdots \\ a_{n0} & \dots & a_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}| \quad = |\det \begin{bmatrix} b_{00} & \dots & b_{0n} & 1 \\ \vdots & & \vdots & \vdots \\ b_{n0} & \dots & b_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}| = \\ &= |\det \begin{bmatrix} d_{00}^2 & \dots & d_{0n}^2 & -1 \\ \vdots & & \vdots & \vdots \\ d_{n0}^2 & \dots & d_{nn}^2 & -1 \\ 1 & \dots & 1 & 0 \end{bmatrix}| = |\det \begin{bmatrix} 0 & \dots & d_{0n}^2 & 1 \\ \vdots & & \vdots & \vdots \\ d_{n0}^2 & \dots & 0 & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}|. \end{aligned}$$

Obserwacja 14.11: Wniosek – pole, wzór 3.

Niech $p_0, \dots, p_n \in (\mathbb{R}^n, (V, \langle\,,\rangle)_{st})$ są równej długości: $\|p_i\| = r$ tzn. jeśli wszystkie wierzchołki p_i leżą na sferze o promieniu r o środku (gdzieś, czyli) w punkcie 0, to dla $S = S(p_0, \dots, p_n)$ mamy:

$$\mu_n(S)^2 = \frac{1}{2^{n+1}(n!)^2 r^2} \cdot |\det \begin{bmatrix} 0 & d_{01}^2 & \dots & d_{0n}^2 \\ d_{10}^2 & 0 & \dots & d_{1n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n0}^2 & d_{n1}^2 & \dots & 0 \end{bmatrix}|.$$

Ilustracja. Niech $\triangle$ będzie trójkątem o bokach długości a, b, c i niech V będzie polem $\triangle$. Jeśli r jest promieniem okręgu opisanego na trójkącie $\triangle$, to $4Vr = abc$.

$$V^2 = \frac{1}{2^{2+1} \cdot (2!)^2 \cdot r^2} \cdot \det \begin{bmatrix} 0 & a^2 & c^2 \\ a^2 & 0 & b^2 \\ c^2 & b^2 & 0 \end{bmatrix} = \frac{1}{2^{2+1} \cdot (2!)^2 \cdot r^2} \cdot 2a^2b^2c^2.$$

Dowód. Wiemy, że $r \cdot n! \cdot \mu_n(S)$ to (także):

$$r \cdot |\det \begin{bmatrix} | & & | \\ p_0 & \cdots & p_n \\ | & & | \\ 1 & & 1 \end{bmatrix}| = |\det \begin{bmatrix} | & & | \\ -p_0 & \cdots & -p_n \\ | & & | \\ r & & r \end{bmatrix}|.$$

A zatem $r \cdot n! \cdot \mu_n(S)$ równe jest:

$$r \cdot |\det \begin{bmatrix} - & p_0^T & - & 1 \\ & \vdots & & \vdots \\ - & p_n^T & - & 1 \end{bmatrix}| = \det \begin{bmatrix} - & p_0^T & - & r \\ & \vdots & & \vdots \\ - & p_n^T & - & r \end{bmatrix}|.$$

$$r \cdot |\det \begin{bmatrix} | & & | \\ p_0 & \cdots & p_n \\ | & & | \\ 1 & & 1 \end{bmatrix}| = |\det \begin{bmatrix} | & & | \\ -p_0 & \cdots & -p_n \\ | & & | \\ r & & r \end{bmatrix}|.$$

A zatem wymnażając:

$$(n!)^2 \cdot \mu_n(S)^2 \cdot r^2 = |\det \begin{bmatrix} r^2 - \|p_0\|^2 & r - \langle p_0, p_1\rangle & \cdots & r^2 - \langle p_0, p_n\rangle \\ \vdots & \vdots & & \vdots \\ r^2 - \langle p_0, p_n\rangle & r^2 - \langle p_1, p_n\rangle & \cdots & r^2 - \|p_n\|^2 \end{bmatrix}.$$

Stąd już wynika teza, bowiem mamy:

$$\begin{aligned} d_{ij}^2 = \|p_i - p_j\|^2 &= \|p_i\|^2 + \|p_j\|^2 - 2\langle p_i, p_j\rangle \\ &= r^2 + r^2 - 2\langle p_i, p_j\rangle \\ &= 2(r^2 - \langle p_i, p_j\rangle). \end{aligned}$$

Obserwacja 14.12: Wniosek - jeszcze jeden (ćwiczenie).

Niech $p_0, \ldots, p_{n+1}$ będą w $\mathbb{R}^n$. Wówczas dla $d_{ij} = \|p_i - p_j\|$ mamy:

$$-(-2)^n \det \underbrace{\begin{bmatrix} \|p_0\|^2 & p_0^T & 1 \\ \vdots & \vdots & \vdots \\ \|p_{n+1}\|^2 & p_{n+1}^T & 1 \end{bmatrix}^2}_{A} = \det \begin{bmatrix} 0 & d_{01}^2 & \ldots & d_{0n}^2 & 1 \\ d_{10}^2 & 0 & \ldots & d_{1n}^2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{n0}^2 & d_{n1}^2 & \ldots & 0 & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{bmatrix}.$$

Co więcej, następujące warunki są równoważne.

(1) Punkty $v_0, \ldots, v_{n+1}$ leżą na hiperpłaszczyźnie lub na sferze w $\mathbb{R}^n$.

(2) Istnieją $a, c \in \mathbb{R}$, $b \in \mathbb{R}^n$ (nie wszystkie równe 0) takie, że $v = (v_0, \ldots, v_{n+1})$ spełniają $a\|v\|^2 + 2\langle b, v\rangle + c = 0$;

(3) $\det A = 0$.

Przykładowe zastosowanie: twierdzenie Ptolemeusza o czworokącie.

Jak Państwo widzicie, także z poziomu algebry liniowej (nie tylko analizy) dowodzić można szereg elementarnych rezultatów, także elementarnych. Można je traktować jako elementy geometrii analitycznej w n-wymiarach. Nie są one zasadniczą częścią wykładu, ponieważ zajmujemy się w nim strukturami i pojęciami przydatnymi nie tylko w geometrii, ale w i całej matematyce.

14.3 Trivia. Zastosowania macierzy Grama w zadaniach

Zadanie 1. *Miasteczko Nieparzystów liczy sobie n mieszkańców. Ich głównym zajęciem jest formowanie rozmaitych klubów, co grozi wielkim politycznym chaosem. Rada Nieparzystowa rozporządziła jednak, aby każdy klub miał nieparzystą liczbę członków. Co więcej każde dwa kluby muszą mieć parzystą liczbę wspólnych członków. Pokazać, że nie jest możliwe stworzenie więcej klubów, niż wynosi ludność Nieparzystowa.*

Liczba potencjalnych klubów, czyli podzbiorów zbioru n-elementowego wynosi 2^n, a więc więcej niż n. Ile wśród nich jest klubów o nieparzystej liczbie członków? Wydaje się, że również możliwe jest, żeby było ich więcej niż n. Gdy $n = 3$ mamy możliwe trzy kluby o jednym członku i jeden klub o trzech członkach, a więc więcej niż 3. Łatwo jednak sprawdzić, że konfiguracja taka nie może wystąpić w przypadku, gdy każde dwa kluby muszą mieć parzystą liczbę wspólnych członków. Jak to pokazać dla każdego n?

Dowód. Ponumerujmy mieszkańców liczbami $1, 2, \ldots, n$, a kluby nazwijmy $C_1, \ldots, C_m$. Definiujemy macierz $A \in M_{n\times m}(\mathbb{R})$, w której kolumnach stoją „listy przynależności" do klubów. Dokładniej $A = (a_{ij})$:

$$a_{ij} = \begin{cases} 1, & \text{jeśli mieszkaniec } i \text{ należy do } C_j \\ 0, & \text{jeśli mieszkaniec } i \text{ nie należy do } C_j. \end{cases}$$

Przyjrzyjmy się teraz iloczynowi A^TA. Jest to macierz rozmiarów $m \times m$, której wyraz w wierszu i-tym oraz w kolumnie j-tej dostajemy ze wzoru: $\sum_{k=1}^{n} a_{ki}a_{kj}$. Co ta liczba oznacza?

Zauważmy, że i-ty wiersz macierzy A^T liczy sobie dokładnie n wyrazów i wyraz a_{ki} koduje przynależność k-tego mieszkańca do klubu C_i. Wyraz ten wynosi 1, jeśli k-ty mieszkaniec należy do klubu C_i oraz 0, jeśli nie należy. Mnożenie i-tego wiersza A^T oraz j-tego wiersza A to po prostu standardowy iloczyn skalarny „list przynależności" do klubów C_i i C_j. Proszę zauważyć, że iloczyn ten ma składniki równe 0 lub 1. Składnik $a_{ki}a_{kj}$ jest równy 1 wtedy i tylko wtedy gdy k-mieszkaniec jest członkiem zarówno klubu C_i jak i klubu C_j. A zatem wyraz macierzy A^TA stojący w i-tym wierszu i j-tej kolumnie będący sumą takich $a_{ki}a_{kj}$ opisuje dokładnie liczbę wspólnych członków klubu i-tego i j-tego! To jest dokładnie macierz Grama układu wektorów opisujących „listy przynależności" do klubów od 1 do m. Zauważmy jeszcze, że można mówić o „prostopadłości" wektorów z listami klubów. Oznacza ona po prostu, że kluby nie mają wspólnych członków.

Co to oznacza dla naszego problemu? Prawo w Nieparzystowie implikuje, że na przekątnej naszej macierzy Grama A^TA stoi nieparzysta liczba członków i-tego klubu (bo i-ty klub ma wszystkich członków wspólnych z i-tym klubem), a wszędzie indziej stoją liczby parzyste. Nietrudno pokazać, że macierz A^TA rozmiaru $m \times m$ jest pełnego rzędu m. Wynika to stąd, że wyznacznik macierzy Grama jest liczbą całkowitą (wszystkie wyrazy są całkowite) zaś wyznacznik macierzy o współczynnikach całkowitych ma resztę z dzielenia przez 2 taką samą jak wyznacznik macierzy powstałej przez podmiany pierwotnych wyrazów przez ich reszty modulo 2. Można to pokazać na wiele sposobów. Natychmiast wynika to na przykład z wzoru permutacyjnego na wyznacznik. Np. dla zbadania parzystości wyznacznika macierzy całkowitoliczbowej:

$$\begin{vmatrix} 102495 & 550428 & 873298 & 660694 \\ 370628 & 909093 & 127450 & 925600 \\ 835044 & 601178 & 624653 & 263392 \\ 663780 & 487252 & 292276 & 593107 \end{vmatrix}$$

wystarczy wziąć każdy ze współczynników modulo 2 i dostajemy macierz I_4, której wyznacznik jest nieparzysty.

W przypadku naszej „klubowej" macierzy Grama. Po wzięciu jej współczynników modulo 2 dostajemy I_m, a więc macierz $m \times m$ o niezerowym wyznaczniku 1. Zatem $r(A^TA) = m$. Na wykładzie przypomnieliśmy nierówność

$$m = r(A^TA) \leqslant r(A) = r(A^T) \leqslant n.$$

Klubów jest zatem nie więcej, niż mieszkańców. □

Podobnych zadań wykorzystujących „wektory przynależności" (lub wektory incydencji) jest znacznie więcej. Zachęcam Czytelnika do zmierzenia się z nimi.

Zadanie 2. Zbiór punktów $\{p_0, p_1, \ldots, p_n\}$ przestrzeni euklidesowej afinicznej $(\mathbb{R}^n, \langle\ ,\ \rangle)$ ma następującą własność: odległość dowolnych dwóch różnych punktów tego zbioru równa jest 1.

a) Niech $\alpha_i = \overrightarrow{p_0p_i}$, dla $i = 1, \ldots, n$. Wyznacz macierz Grama $G(\alpha_1, \ldots, \alpha_n)$.

b) Pokaż, że układ punktów $\{p_0, p_1, \ldots, p_n\}$ jest bazą punktową przestrzeni afinicznej $(\mathbb{R}^n, \langle\ ,\ \rangle)$.

c) Dla każdego $0 < k < n$ określamy $L_k = \mathrm{af}(p_0, \ldots, p_k)$, $S_k = \mathrm{af}(p_{k+1}, \ldots, p_n)$. Wykaż, że $L_k \cap S_k = \emptyset$.

d) Dla $n = 3$ wyznacz odległość prostych $\mathrm{af}(p_0, p_1)$ oraz $\mathrm{af}(p_2, p_3)$.

Rozwiązanie. Zacznijmy od (a). Zauważmy, że $\alpha_i - \alpha_j = \overrightarrow{p_0p_i} - \overrightarrow{p_0p_j} = \overrightarrow{p_jp_i}$. Przypomnijmy, że dla wektorów α, β przestrzeni euklidesowej mamy:

$$2\langle\alpha, \beta\rangle = \|\alpha\|^2 + \|\beta\|^2 - \|\alpha - \beta\|^2 \quad \Rightarrow \quad 2\langle\alpha_i, \alpha_j\rangle = \|\overrightarrow{p_0p_i}\|^2 + \|\overrightarrow{p_0p_i}\|^2 - \|\overrightarrow{p_jp_i}\|^2.$$

Dowodzimy (b). To, że $\{p_0, p_1, \ldots, p_n\}$ jest bazą punktową $\mathbb{R}^n$ jest równoważne temu, że $\alpha_1, \ldots, \alpha_n$ jest bazą $\mathbb{R}^n$. Zgodnie z twierdzeniem z wykładu jest równoważne temu, że macierz Grama $G(\alpha_1, \ldots, \alpha_n)$ jest odwracalna. To z kolei pokazujemy dowodząc, że $r(G(\alpha_1, \ldots, \alpha_n)) = n$. Robimy to przez operacje elementarne. Wypisujemy macierz $2G(\alpha_1, \ldots, \alpha_n)$ i odejmujemy ostatni wiersz od poprzednich. Dostajemy:

$$r\begin{bmatrix} 2 & 1 & \ldots & 1 & 1 \\ 1 & 2 & \ldots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \ldots & 2 & 1 \\ 1 & 1 & \ldots & 1 & 2 \end{bmatrix} = r\begin{bmatrix} 1 & 0 & \ldots & 0 & -1 \\ 0 & 1 & \ldots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -1 \\ 1 & 1 & \ldots & 1 & 2 \end{bmatrix}.$$

Jest jasne, że pierwsze $n-1$ wierszy uzyskanej macierzy jest liniowo niezależnych, a ostatni nie jest kombinacją liniową wcześniejszych (patrz pierwsze $n-1$ współrzędnych). Stąd mamy tezę.

Dowodzimy (c). Załóżmy przeciwnie, że dla pewnych układów wag $a_0, \ldots, a_k$ oraz $b_{k+1}, \ldots, b_n$ mamy

$$a_0p_0 + a_1p_1 + \ldots + a_kp_k = b_{k+1}p_{k+1} + \ldots + b_np_n.$$

Mamy $a_0 + a_1 + \ldots + a_k - b_{k+1} - \ldots - b_n = 1 - 1 = 0$ i wyrażenie wyżej można równoważnie przekształcić:

$$\begin{aligned} &a_0p_0 + a_1p_1 + \ldots + a_kp_k - b_{k+1}p_{k+1} - \ldots - b_np_n + (a_0 + a_1 + \ldots + a_k - b_{k+1} - \ldots - b_n)p_0 = \\ &a_1\overrightarrow{p_0p_1} + a_2\overrightarrow{p_0p_2} + \ldots + a_k\overrightarrow{p_0p_k} - b_{k+1}\overrightarrow{p_0p_{k+1}} - \ldots - b_n\overrightarrow{p_0p_n} = \\ &a_1\alpha_1 + a_2\alpha_2 + \ldots + a_k\alpha_k - b_{k+1}\alpha_{k+1} - \ldots - b_n\alpha_n = 0 \end{aligned}$$

Jednak jedna z liczb a_i, b_j jest niezerowa, a układ $\alpha_1, \ldots, \alpha_n$ jest liniowo niezależny. Sprzeczność.

Przechodzimy do ostatniej części. Niech $K = \mathrm{af}(p_0, p_1)$ oraz $L = \mathrm{af}(p_2, p_3)$. Mamy

$$T(K) = \alpha_1, \quad T(L) = \overrightarrow{p_2p_3} = \overrightarrow{p_0p_3} - \overrightarrow{p_0p_2} = \alpha_3 - \alpha_2.$$

Odległość d prostych K oraz L równa jest długości rzutu wektora łączącego dowolne dwa jej punkty na przestrzeń prostopadłą do $\mathrm{lin}(T(K), T(L))$. Ten wektor łączący to może być np. $\overrightarrow{p_0p_3} = \alpha_3$, zaś owa przestrzeń, na którą rzutujemy to $\mathrm{lin}(\alpha_1, \alpha_3 - \alpha_2)^\perp$. Zgodnie z twierdzeniem z wykładu (typu „objętość = podstawa x wysokość") mamy:

$$d = \sqrt{\frac{W(\alpha_1, \alpha_3 - \alpha_2, \alpha_3)}{W(\alpha_1, \alpha_3 - \alpha_2)}}.$$

Bez trudu wyznaczamy

$$G(\alpha_1, \alpha_3 - \alpha_2, \alpha_3) = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix},$$

skąd szukana odległość to $\frac{\sqrt{2}}{2}$.

Uwaga 1. Idea geometryczna: realizując wierzchołki czworościanu foremnego jako wierzchołki sześcianu o krawędzi $\frac{\sqrt{2}}{2}$ (krawędzie czworościanu to przekątne ścian bocznych sześcianu) dostajemy uzyskaną odpowiedź, bo szukana odległość prostych równa jest odległości przeciwległych ścian sześcianu.

Uwaga 2. Zachęcam Czytelnika do pokazania, że $\rho(L_k, S_k)$ równa jest, w zależności od n:

$$\sqrt{\frac{W(\alpha_1, \ldots, \alpha_k, \alpha_n - \alpha_{k+1}, \ldots, \alpha_n - \alpha_{n-1}, \alpha_n)}{W(\alpha_1, \ldots, \alpha_k, \alpha_n - \alpha_{k+1}, \ldots, \alpha_n - \alpha_{n-1})}} = \sqrt{\frac{n+1}{2(k+1)(n-k)}}.$$

Powyższe zadanie pochodzące z jednego z kolokwiów dotyczy przykładu zbioru, który nazywamy czasem równobocznym. Oto definicja pochodząca ze znanych 33 Miniatur J. Matouska.

Definicja 14.12: Zbiór równoboczny

Podzbiór równoboczny w przestrzeni afinicznej $\mathbb{R}^n$ to zbiór punktów $p_1, \ldots, p_k$ takich, że wszystkie pary punktów p_i, p_j są równoodległe.

Na podstawie naszych rozważań nietrudno zobaczyć, że zbiór równoodległy w $\mathbb{R}^n$ ma nie więcej niż $n+1$ punktów — tak jak sympleks d-wymiarowy oraz zbiór rozważany w zadaniu wyżej.

Problem znajdowania zbiorów równoodległych komplikuje się, gdy przejdziemy do innej metryki[8], na przykład do metryki rzeka, zwanej w matematyce metryką l_1: dla punktów $x = (x_1, \ldots, x_n)$ oraz $y = (y_1, \ldots, y_n)$ w przestrzeni $\mathbb{R}^n$ określamy:

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n|.$$

Przypomnijmy, że w metryce tej poruszamy się tylko wzdłuż prostych równoległych do osi. Dla tej metryki możemy łatwo wskazać zbiór równoodległy złożony z $2n$ elementów $\{e_1, -e_1, e_2, -e_2, \ldots, e_n, -e_n\}$ gdzie e_i jest wektorem standardowym. Co ciekawe istnieje problem otwarty Kusnera z 1983 roku mówiący, że zbiór równoodległy nie może mieć więcej elementów. Aż do 2001 nie było jasno, że ograniczenie górne ma nie więcej niż $2^n - 1$ elementów. W następnych latach pojawiło się wiele ulepszeń tego wyniku, ale ostatecznego rozstrzygnięcia hipotezy wciąż nie ma.

Na koniec warto wspomnieć, że z macierzy Grama można odczytywać interesujące fakty dotyczące układów odległości w przestrzeni euklidesowej. Oto przykład.

Obserwacja 14.13

Nie istnieją cztery punkty na płaszczyźnie, kórych odległości są wszystkie liczbami nieparzystymi.

Dowód. Gdyby taki układ punktów istniał, to bez straty ogólności moglibyśmy założyć, że jeden z nich jest 0, a pozostałe to a, b, c (zakładamy, że jesteśmy w przestrzeni liniowej $\mathbb{R}^2$). Zgodnie z założeniami obserwacji liczby $\|a\|, \|b\|, \|c\|, \|a-b\|, \|b-c\|, \|c-a\|$ są nieparzyste.

Kluczowe jest, że gdy m jest liczbą nieparzystą, wtedy m^2 daje resztę 1 z dzielenia przez 8. Z twierdzenia kosinusów mamy zatem

$$2\langle a, b\rangle = \|a\|^2 + \|b\|^2 - \|a-b\|^2$$

co oznacza, że liczby $2\langle a, b\rangle$, $2\langle a, c\rangle$ oraz $2\langle b, c\rangle$ też dają resztę 1 z dzielenia przez 8. W szczególności macierz Grama $G(a, b, c)$ ma tę własność, że oglądana modulo 8 ma postać:

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

Wyznacznik tej macierzy wynosi 4, a więc rząd macierzy $G(a, b, c)$ wynosi 3 (nie przystaje 0 modulo 8). Z drugiej strony mówimy o macierzy Grama układu 3 wektorów na płaszczyźnie — układu liniowo zależnego. A więc jej wyznacznik wynosi 0, zgodnie z faktem z wykładu. Uzyskaliśmy sprzeczność. □

[8] Znowu odsyłam do 33 Miniatur J. Matouska: `https://kam.mff.cuni.cz/~matousek/stml-53-matousek-1.pdf`

Zadanie 3. W zorientowanej przestrzeni euklidesowej afinicznej $(\mathbb{R}^3, \langle\ ,\ \rangle_{st})$ dany jest układ ośmiu punktów $p_1, \ldots, p_8$ stanowiących zbiór wierzchołków równoległościanu $R = R(p_1; \overrightarrow{p_1p_2}, \overrightarrow{p_1p_3}, \overrightarrow{p_1p_5})$, przy czym spełnione są warunki $\overrightarrow{p_1p_2} + \overrightarrow{p_1p_3} = \overrightarrow{p_1p_4}$ oraz $\overrightarrow{p_5p_6} + \overrightarrow{p_5p_7} = \overrightarrow{p_5p_8}$.

(a) Czy z równości $\|\overrightarrow{p_1p_2}\| = \|\overrightarrow{p_1p_3}\| = \|\overrightarrow{p_1p_5}\|$ wynikają równości $\|\overrightarrow{p_8p_7}\| = \|\overrightarrow{p_8p_6}\| = \|\overrightarrow{p_8p_4}\|$?

(b) Czy z równości $\overrightarrow{p_1p_5} = \overrightarrow{p_1p_2} \times \overrightarrow{p_1p_3}$ wynika równość $\overrightarrow{p_8p_4} = \overrightarrow{p_8p_7} \times \overrightarrow{p_8p_6}$?

(c) Załóżmy, że R jest sześcianem, oraz że dla pewnej płaszczyzny $P \subset \mathbb{R}^3$ odległość punktu p_i od P równa jest i, dla $1 \leqslant i \leqslant 8$. Wyznacz długość krawędzi sześcianu R.

Uwaga. Wierzchołki $R = R(p_1; \alpha_1, \alpha_2, \alpha_3)$ to punkty postaci $p_1 + \epsilon_1\alpha_1 + \epsilon_2\alpha_2 + \epsilon_3\alpha_3$, gdzie $\epsilon_i \in \{0, 1\}$.

Rozwiązanie. Z oznaczmy $\alpha = \overrightarrow{p_1p_2}$, $\beta = \overrightarrow{p_1p_3}$ oraz $\gamma = \overrightarrow{p_1p_5}$. Z warunku $\overrightarrow{p_1p_2} + \overrightarrow{p_1p_3} = \overrightarrow{p_1p_4}$ wynika, że $p_4 = p_1 + \alpha + \beta$ należy do $\mathrm{af}(p_1, p_2, p_3)$, natomiast z warunku $\overrightarrow{p_5p_6} + \overrightarrow{p_5p_7} = \overrightarrow{p_5p_8}$ wynika, że $p_8 = p_5 + \overrightarrow{p_5p_6} + \overrightarrow{p_5p_7}$. Wiedząc dodatkowo, że $p_5 = p_1 + \overrightarrow{p_1p_5}$ dostajemy $p_8 = p_1 + \alpha + \beta + \gamma$. A zatem punkty p_6, p_7 są w pewnej kolejności postaci $p_1+\alpha+\gamma$ oraz $p_1+\beta+\gamma$. Będzie to miało znaczenie w (b).

W punkcie (a) odpowiedź jest twierdząca. Z rozważań wyżej wiemy, że $\overrightarrow{p_8p_7} = p_7 - p_8$, czyli jest to wektor $-\beta$ lub $-\alpha$. Podobnie $\overrightarrow{p_8p_6}$ równy jest $-\beta$ lub $-\alpha$. Wreszcie, $\overrightarrow{p_8p_4} = p_4 - p_8 = -\gamma$. Skoro $\|\alpha\| = \|\beta\| = \|\gamma\|$, to $\|\overrightarrow{p_8p_7}\| = \|\overrightarrow{p_8p_6}\| = \|\overrightarrow{p_8p_4}\|$.

Punkt (b) nie jest prawdą. Wiemy, że baza α, β, γ jest zorientowana dodatnio, bowiem $\gamma = \alpha \times \beta$. To oznacza, że również bazy $-\alpha, -\beta, \gamma$ oraz $-\beta, -\alpha, -\gamma$ są zorientowane dodatnio. Mamy $-\gamma = \overrightarrow{p_8p_4}$. W zależności od wyboru p_6, p_7 jako $p_1 + \alpha + \gamma$ oraz $p_1 + \beta + \gamma$ mamy

$$\overrightarrow{p_8p_7} \times \overrightarrow{p_8p_6} = -\beta \times -\alpha \quad \text{lub} \quad -\alpha \times -\beta.$$

Tylko jeden z tych iloczynów jest równy $-\gamma$, co kończy argument.

Odpowiedź w punkcie (c) to $\sqrt{21}$. Niech $T(P)^\perp = \mathrm{lin}(\delta)$. Załóżmy też bez straty ogólności, że $\|\delta\| = 1$. Zauważmy, że zgodnie z założeniem rzuty α, β, γ na $\mathrm{lin}(\delta)$ to wektory mające długości odpowiednio $1, 2, 4$. Rzuty te mają jednocześnie postać (odpowiednio):

$$\frac{\langle\, \alpha, \delta \,\rangle}{\|\delta\|^2}\delta, \quad \frac{\langle\, \beta, \delta \,\rangle}{\|\delta\|^2}\delta, \quad \frac{\langle\, \gamma, \delta \,\rangle}{\|\delta\|^2}\delta$$

Liczby $\langle\, \alpha, \delta \,\rangle, \langle\, \beta, \delta \,\rangle, \langle\, \gamma, \delta \,\rangle$ są zatem równe co do modułu odpowiednio $1, 2, 4$.

Skoro R jest sześcianem, to $\|\alpha\| = \|\beta\| = \|\gamma\|$ oraz układ α, β, γ jest prostopadły. To oznacza, że:

$$G(\alpha, \beta, \gamma, \delta) = \begin{bmatrix} \langle\, \alpha, \alpha \,\rangle & \langle\, \alpha, \beta \,\rangle & \langle\, \alpha, \gamma \,\rangle & \langle\, \alpha, \delta \,\rangle \\ \langle\, \beta, \alpha \,\rangle & \langle\, \beta, \beta \,\rangle & \langle\, \beta, \gamma \,\rangle & \langle\, \beta, \delta \,\rangle \\ \langle\, \gamma, \alpha \,\rangle & \langle\, \gamma, \beta \,\rangle & \langle\, \gamma, \gamma \,\rangle & \langle\, \gamma, \delta \,\rangle \\ \langle\, \delta, \alpha \,\rangle & \langle\, \delta, \beta \,\rangle & \langle\, \delta, \gamma \,\rangle & \langle\, \delta, \delta \,\rangle \end{bmatrix} = \begin{bmatrix} \|\alpha\|^2 & 0 & 0 & \pm 1 \\ 0 & \|\alpha\|^2 & 0 & \pm 2 \\ 0 & 0 & \|\alpha\|^2 & \pm 4 \\ \pm 1 & \pm 2 & \pm 4 & 1 \end{bmatrix},$$

przy czym macierz wyżej jest symetryczna. Układ $\alpha, \beta, \gamma, \delta$ jest liniowo zależny, czyli $W(\alpha, \beta, \gamma, \delta) = 0$. Wyznacznik ten jest natomiast równy $\|\alpha\|^6 - 21\|\alpha\|^4$, niezależnie od wyboru znaków $\langle\, \alpha, \delta \,\rangle, \langle\, \beta, \delta \,\rangle, \langle\, \gamma, \delta \,\rangle$. Wynika to z faktu, że przy operacji zamiany jednego z wektorów α, β, γ na przeciwny (co zmieniłoby tylko znak iloczynu skalarnego z δ) wyznacznik Grama się nie zmienia. Skoro $\|\alpha\| > 0$, to $\|\alpha\| = \sqrt{21}$.

Autorem punktu (c) jest Łukasz Bożyk, patrz: `https://www.deltami.edu.pl/delta/archiwum/2018/12/2018/12/04/0-DELTA-2018-12.pdf`, str 3. Elementarne rozwiązanie (jedno z wielu) — na str. 5.

Rozdział 15

Izometrie liniowe i afiniczne. Macierze ortogonalne

Celem tego wykładu jest powiedzieć o przekształceniach liniowych przestrzeni euklidesowych mających równoważne sobie własności: zachowywania iloczynu skalarnego oraz zachowywania długości wektorów. Podobnie jak prostopadłość (czy ogólniej mówiąc – struktura euklidesowa) wywiera wpływ na własności układów wektorów, np. gwarantując ich liniową niezależność, tak i w przypadku przekształceń liniowych zachowywanie struktury euklidesowej ma konsekwencje dla struktury samego przekształcenia.

Definicja 15.1

Niech $(V_1, \langle\,,\,\rangle_1), (V_2, \langle\,,\,\rangle_2)$ będą przestrzeniami euklidesowymi liniowymi oraz niech $\phi : V_1 \to V_2$ będzie przekształceniem liniowym.

(a) Mówimy, że ϕ ZACHOWUJE ILOCZYN SKALARNY, jeśli dla każdych $\alpha, \beta \in V_1$ zachodzi:

$$\langle\, \alpha, \beta \,\rangle_1 = \langle\, \phi(\alpha), \phi(\beta) \,\rangle_2 .$$

(b) Mówimy, że ϕ ZACHOWUJE DŁUGOŚĆ WEKTORÓW jeśli dla każdego $\alpha \in V_1$ zachodzi

$$\|\alpha\|_1 = \|\phi(\alpha)\|_2,$$

gdzie $\|\alpha\|_1 = \sqrt{\langle\, \alpha, \alpha \,\rangle_1}$, $\|\gamma\|_2 = \sqrt{\langle\, \gamma, \gamma \,\rangle_2}$, dla każdych $\alpha \in V_1, \gamma \in V_2$.

Zacznijmy od podstawowej obserwacji.

Obserwacja 15.1

Niech $(V_1, \langle\,,\,\rangle_1), (V_2, \langle\,,\,\rangle_2)$ będą przestrzeniami euklidesowymi liniowymi oraz niech $\phi \in L(V_1, V_2)$. Wówczas ϕ zachowuje iloczyn skalarny wtedy i tylko wtedy, gdy ϕ zachowuje długość wektorów.

Dowód. Jeśli ϕ zachowuje iloczyn skalarny, to dla $\alpha \in V_1$ mamy:

$$\|\phi(\alpha)\|_2 = \sqrt{\langle\, \phi(\alpha), \phi(\alpha) \,\rangle_2} = \sqrt{\langle\, \alpha, \alpha \,\rangle_1} = \|\alpha\|_1.$$

Na odwrót: załóżmy, że ϕ zachowuje długość wektorów. Wówczas $\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2 + 2\,\langle\, \alpha, \beta \,\rangle$, a więc

$$2\,\langle\, \alpha, \beta \,\rangle = \|\alpha + \beta\|^2 - \|\alpha\|^2 - \|\beta\|^2,$$

dla dowolnych wektorów α, β. Zatem dla każdych $\alpha, \beta \in V_1$ mamy:

$$\begin{aligned} 2\,\langle\, \phi(\alpha), \phi(\beta) \,\rangle_2 &= \|\phi(\alpha) + \phi(\beta)\|_2^2 - \|\phi(\alpha)\|_2^2 - \|\phi(\beta)\|_2^2 = \\ &= \|\alpha + \beta\|_1^2 - \|\alpha\|_1^2 - \|\beta\|_1^2 = 2\,\langle\, \alpha, \beta \,\rangle_1 . \end{aligned}$$

□

Obserwacja 15.2

Jeśli przekształcenie liniowe ϕ zachowuje iloczyn skalarny, to ϕ jest monomorfizmem.

Dowód. Dla $\alpha \in \ker(\phi)$ mamy $\phi(\alpha) = 0$, czyli $\|\phi(\alpha)\|_2 = 0$, skąd $\|\alpha\|_1 = 0$, więc $\alpha = 0$. □

Definicja 15.2: Izometria przestrzeni euklidesowej (liniowej)

Niech $(V_1, \langle\,,\rangle_1), (V_2, \langle\,,\rangle_2)$ będą przestrzeniami euklidesowymi liniowymi. Mówimy, że przekształcenie liniowe $\phi : V_1 \to V_2$ jest IZOMETRIĄ przestrzeni $(V_1, \langle\,,\rangle_1)$ na przestrzeń $(V_2, \langle\,,\rangle_2)$ jeśli:

(a) ϕ jest izomorfizmem przestrzeni liniowej V_1 na przestrzeń liniową V_2,

(b) ϕ zachowuje iloczyn skalarny.

Izometrię przestrzeni $(V, \langle\,,\rangle)$ na $(V, \langle\,,\rangle)$ nazywamy IZOMETRIĄ PRZESTRZENI EUKLIDESOWEJ $(V, \langle\,,\rangle)$.

Oto dwa podstawowe przykłady izometrii przestrzeni euklidesowych.

1. **Symetrie prostopadłe**. Niech W będzie podprzestrzenią przestrzeni euklidesowej liniowej $(V, \langle\,,\rangle)$ i niech $\psi : V \to V$ będzie symetrią prostopadłą względem W. Wówczas ψ jest izomorfizmem oraz dla dowolnego $v = v_1 + v_2$, gdzie $v_1 \in W, v_2 \in W^\perp$, mamy $\psi(v) = v_1 - v_2$, czyli skoro $v_1 \perp v_2$, to:
$$\|\psi(v)\|^2 = \langle\, v_1 - v_2, v_1 - v_2 \,\rangle = \|v_1\|^2 + \|v_2\|^2 - 2\,\langle\, v_1, v_2 \,\rangle = \|v_1\|^2 + \|v_2\|^2 = \|v_1 + v_2\|^2 = \|v\|^2.$$
Zauważmy, że zgodnie z przyjętą definicją rzut prostopadły nie jest (zwykle) izometrią.

2. **Obroty**. Niech $(V, \langle\,,\rangle)$ będzie dwuwymiarową przestrzenią euklidesową liniową. Załóżmy, że V jest zorientowana i niech $\mathcal{A}$ będzie bazą ortonormalną tej przestrzeni, zorientowaną zgodnie z zadaną orientacją. Wówczas przekształcenie $\phi : V \to V$ zadane warunkiem
$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$
nazywamy OBROTEM O KĄT θ.

Definicja 15.3

Niech $(V, \langle\,,\rangle)$ będzie przestrzenią euklidesową liniową i niech $W \subseteq V$ będzie jej 2-wymiarową podprzestrzenią zorientowaną. Przekształcenie liniowe $\phi : V \to V$ takie, że $\phi|_W$ jest obrotem o kat θ w przestrzeni W, a $\phi|_{W^\perp}$ jest identycznością na przestrzeni $W^\perp$ nazywamy OBROTEM O KAT θ WOKÓŁ PODPRZESTRZENI $W^\perp$. Jeśli $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ jest taką bazą przestrzeni V, że $\mathcal{A}' = (\alpha_1, \alpha_2)$ jest bazą ortonormalną przestrzeni W zorientowaną zgodnie z orientacją W, natomiast $(\alpha_3, \ldots, \alpha_n)$ jest dowolną bazą przestrzeni $W^\perp$, to $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest w postaci blokowej
$$\begin{bmatrix} O_\theta & 0 \\ 0 & I_{n-2} \end{bmatrix}, \quad \text{gdzie} \quad O_\theta = M(\phi|_W)^{\mathcal{A}'}_{\mathcal{A}'} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Pojęcie obrotu wymaga dobrego zrozumienia pojęcia orientacji przestrzeni euklidesowej, ponieważ wyznaczenie właściwego kąta obrotu i podprzestrzeni wokół której obracamy możliwe jest jedynie wówczas, gdy blok O_θ macierzy endomorfizmu ϕ przestrzeni euklidesowych uzyskany jest w dodatnio zorientowanej bazie podprzestrzeni W. Inaczej macierz endomorfizmu ϕ ma właściwą postać blokową, ale możemy ją źle zinterpretować. Oto przykłady, które powinny dać jakąś intuicję.

- Niech $(V, \langle\,,\rangle)$ będzie przestrzenią $\mathbb{R}^2$ ze standardowym iloczynem skalarnym i orientacją wyznaczoną przez bazę standardową. Bierzemy $\phi : V \to V$ zadane macierzą
$$M(\phi)^{st}_{st} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$
Wówczas ϕ jest obrotem o kąt $-\frac{\pi}{3}$.

- $V = \mathbb{R}^2$ ze standardowym iloczynem skalarnym i orientacją wyznaczoną przez bazę $\mathcal{A} = ((0,1),(1,0))$ (to nie jest baza standardowa!). Bierzemy ϕ zadane tym samym wzorem, co wyżej. Wówczas ϕ jest obrotem o kąt $\frac{\pi}{3}$. Dlaczego? Wprawdzie dalej w bazie standardowej przekształcenie ϕ ma macierz, jak wyżej, ale teraz baza standardowa nie jest zorientowana zgodnie z bazą $\mathcal{A}$ wyznaczającą orientację. Mamy jednak
$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = \begin{bmatrix} \frac{1}{2} & \frac{-\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Czas powiedzieć kilka słów o macierzach izometrii. Jak się okazuje, szczególnie istotne jest rozważanie macierzy w bazach ortonormalnych.

Twierdzenie 15.1

Niech $(V_1, \langle\,,\rangle_1), (V_2, \langle\,,\rangle_2)$ będą przestrzeniami euklidesowymi liniowymi. Następujące warunki są równoważne dla $\phi \in L(V_1, V_2)$:

(1) ϕ jest izometrią przestrzeni euklidesowej $(V_1, \langle\,,\rangle_1)$ na $(V_2, \langle\,,\rangle_2)$,

(2) ϕ przeprowadza każdą bazę ortonormalną przestrzeni $(V_1, \langle\,,\rangle_1)$ na bazę ortonormalną $(V_2, \langle\,,\rangle_2)$,

(3) ϕ przeprowadza pewną bazę ortonormalną przestrzeni $(V_1, \langle\,,\rangle_1)$ na bazę ortonormalną $(V_2, \langle\,,\rangle_2)$.

Co więcej, jeśli $A = M(\phi)^{\mathcal{B}}_{\mathcal{A}} \in M_n(\mathbb{R})$ jest macierzą izometrii przestrzeni euklidesowych w bazach ortonormalnych, wówczas kolumny A są prostopadłe w $(\mathbb{R}^n, \langle\,,\rangle_{st})$. Innymi słowy $A^T A = I$.

Dowód. W pierwszej części dowodu pokażemy tylko implikację (3) $\Rightarrow$ (1), bo (2) $\Rightarrow$ (3) jest trywialna, a implikacja (1) $\Rightarrow$ (2) wynika stad, że ϕ jako izometria zachowuje długość i prostopadłość.

Przypuśćmy zatem, że dla pewnej bazy ortonormalnej $\alpha_1, \ldots, \alpha_n$ przestrzeni euklidesowej $(V_1, \langle\,,\rangle)_1)$ układ wektorów $\phi(\alpha_1), \ldots, \phi(\alpha_n)$ jest bazą ortonormalną przestrzeni $(V_2, \langle\,,\rangle)_2)$. Wykażemy, że ϕ zachowuje iloczyn skalarny. Dla dowolnych $\alpha = x_1\alpha_1 + \ldots + x_n\alpha_n$ oraz $\beta = y_1\alpha_1 + \ldots + y_n\alpha_n$ w przestrzeni V_1 mamy
$$\phi(\alpha) = x_1\phi(\alpha_1) + \ldots + x_n\phi(\alpha_n) \quad \text{oraz} \quad \phi(\beta) = y_1\phi(\alpha_1) + \ldots + y_n\phi(\alpha_n),$$
a zatem z ortonormalności obydwu wyróżnionych baz mamy warunek $\langle \alpha_i, \alpha_i \rangle_1 = \langle \phi(\alpha_j), \phi(\alpha_j) \rangle_2 = 1$, dla każdych i, j, czyli[1]:
$$\begin{aligned} \langle \alpha, \beta \rangle_1 &= \langle x_1\alpha_1 + \ldots + x_n\alpha_n, y_1\alpha_1 + \ldots + y_n\alpha_n \rangle_1 = \\ &= x_1y_1 \langle \alpha_1, \alpha_1 \rangle_1 + \ldots + x_ny_n \langle \alpha_n, \alpha_n \rangle_1 = \\ &= x_1y_1 \langle \phi(\alpha_1), \phi(\alpha_1) \rangle_2 + \ldots + x_ny_n \langle \phi(\alpha_n), \phi(\alpha_n) \rangle_2 = \\ &= \langle x_1\phi(\alpha_1) + \ldots + x_n\phi(\alpha_n), y_1\phi(\alpha_1) + \ldots + y_n\phi(\alpha_n) \rangle_2 = \\ &= \langle \phi(\alpha), \phi(\beta) \rangle_2 . \end{aligned}$$

Przechodzimy do drugiej części dowodu. Niech $\phi : (V_1, \langle\,,\rangle)_1) \to (V_2, \langle\,,\rangle_2)$ będzie dowolną izometrią. Niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n), \mathcal{B} = (\beta_1, \ldots, \beta_n)$ będą bazami ortonormalnymi odpowiednio $(V_1, \langle\,,\rangle)_1), (V_2, \langle\,,\rangle)_2)$. Rozważamy macierz $A = M(\phi)^{\mathcal{B}}_{\mathcal{A}} \in M_n(\mathbb{R})$ i chcemy pokazać, że $A^T A = I$.

Niech $i \neq j$ oraz niech $\phi(\alpha_i) = x_1\beta_1 + \ldots + x_n\beta_n$ i $\phi(\alpha_j) = y_1\beta_1 + \ldots + y_n\beta_n$. Aby pokazać tezę wystarczy pokazać, że:
$$\begin{aligned} x_1y_1 + \ldots + x_ny_n &= 0, \\ x_1^2 + \ldots + x_n^2 = y_1^2 + \ldots + y_n^2 &= 1. \end{aligned}$$

Skoro jednak ϕ jest izometrią, to z ortonormalności układu α_i wynika ortonormalność układu $\phi(\alpha_i)$, to znaczy:
$$\langle \alpha_i, \alpha_j \rangle_1 = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \Rightarrow \langle \phi(\alpha_i), \phi(\alpha_j) \rangle_2 = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}.$$

[1] W przejściu między trzecią a czwartą linią dodajemy elementy typu $x_iy_j \langle \phi(\alpha_i), \phi(\alpha_j) \rangle_2 = 0$, dla $i \neq j$, co można zrobić korzystając z ortonormalności układu $\phi(\alpha_i)$.

Z drugiej strony $\langle \beta_i, \beta_j \rangle = 0$, dla $i \neq j$ oraz $\langle \beta_i, \beta_i \rangle = 1$, czyli:

$$\begin{aligned} \langle\, \phi(\alpha_i), \phi(\alpha_j)\,\rangle_2 &= \langle\, x_1\beta_1 + \ldots + x_n\beta_n, y_1\beta_1 + \ldots + y_n\beta_n \,\rangle_2 = \\ &= x_1y_1 \langle\, \beta_1, \beta_1 \,\rangle_2 + \ldots + x_ny_n \langle\, \beta_n, \beta_n \,\rangle_2 = \\ &= x_1y_1 + \ldots + x_ny_n. \end{aligned}$$

co natychmiast daje (♠) i tezę twierdzenia. □

Macierze opisane w twierdzeniu wyżej mają fundamentalne znaczenie.

Definicja 15.4: Macierz ortogonalna

Macierz $A \in M_n(\mathbb{R})$ nazywamy ORTOGONALNĄ, jeśli $A^TA = I$. Inaczej mówiąc, kolumny macierzy ortogonalnej są bazą ortonormalną $\mathbb{R}^n$ ze standardowym iloczynem skalarnym.

Macierze ortogonalne są to więc odwracalne macierze kwadratowe, których odwrotność równa jest ich transpozycji. Wśród wielu ważnych przykładów tych macierzy wyróżnić można macierze permutacyjne (w tym I), a także macierze obrotu i symetrii prostopadłej (w bazach ortonormalnych). Powyższe rozważania pozwalają sformułować następujący wniosek.

Obserwacja 15.3

Niech $(V_1, \langle\,,\,\rangle_1), (V_2, \langle\,,\,\rangle_2)$ będą przestrzeniami euklidesowymi Przekształcenie liniowe $\phi : V_1 \to V_2$ jest izometrią przestrzeni euklidesowej $(V_1, \langle\,,\,\rangle_1)$ na przestrzeń euklidesową $(V_2, \langle\,,\,\rangle_2)$ wtedy i tylko wtedy, gdy dla pewnych (równoważnie: dla każdych) baz ortonormalnych $\mathcal{A}_1$ przestrzeni $(V_1, \langle\,,\,\rangle_1)$ oraz $\mathcal{A}_2$ przestrzeni $(V_2, \langle\,,\,\rangle_2)$ macierz $M(\phi)^{\mathcal{A}_2}_{\mathcal{A}_1}$ jest ortogonalna.

Wniosek ten daje wygodnie kryterium sprawdzania czy przekształcenie liniowe przestrzeni euklidesowych zadane wzorem jest izometrią, czy nie. Dla przykładu, weźmy endomorfizm przestrzeni $\mathbb{R}^2$ ze standardowym iloczynem skalarnym dany wzorem $\phi((x_1, x_2)) = (x_1, -x_1 + x_2)$. Baza standardowa jest bazą ortonormalną tej przestrzeni euklidesowej, a zatem do sprawdzenia czy ϕ jest izometrią wystarczy przekonać się czy $M(\phi)^{st}_{st}$ jest ortogonalna (wtedy będzie też pewność, że w dowolnych innych bazach ortonormalnych przestrzeni $\mathbb{R}^2$ przekształcenie to ma macierz ortogonalną). Mamy natomiast:

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}^T \cdot \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

W rezultacie ϕ nie jest izometrią.

Warto sformułować kolejny wniosek dotyczący zmiany baz.

Obserwacja 15.4

Jeśli $\mathcal{A}$ oraz $\mathcal{B}$ są bazami ortonormalnymi $(\mathbb{R}^n, \langle\,,\,\rangle)$, to

$$(M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}})^{-1} = (M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}})^T.$$

Zatem dla izometrii ϕ przestrzeni $(\mathbb{R}^n, \langle\,,\,\rangle)$:

$$M(\phi)^{\mathcal{A}}_{\mathcal{A}} = (M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}})^T \cdot M(\phi)^{\mathcal{B}}_{\mathcal{B}} \cdot M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}}.$$

Na koniec pozostawiam Państwu do wykazania dwa istotne fakty. Pierwszy jest bardzo prosty, bo wymaga jedynie analizy tego jak może wyglądać macierz ortogonalna rozmiaru 2.

Obserwacja 15.5

Każda izometria przestrzeni euklidesowej liniowej wymiaru 2 jest albo obrotem o kąt θ albo symetrią prostopadłą względem $\mathrm{lin}((\sin\theta, 1-\cos\theta))$, więc w pewnej bazie ortonormalnej ma jedną z poniższych macierzy:

$$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}, \quad \begin{bmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{bmatrix}$$

Drugi fakt wymaga nieco więcej wysiłku, w tym analizy możliwych wartości własnych izometrii.

Obserwacja 15.6

Każda izometria przestrzeni euklidesowej liniowej wymiaru 3 jest albo obrotem, albo obrotem złożonym z symetrią prostopadłą, albo symetrią płaszczyznową. Przekształcenia te mają w pewnej bazie ortonormalnej macierze postaci, odpowiednio:

$$\begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}, \quad \begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & -1\end{bmatrix}, \quad \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{bmatrix}.$$

Ostatnie twierdzenie można pokazać także korzystając z twierdzenia Cartana o rozkładzie izometrii przestrzeni trójwymiarowej na nie więcej niż trzy symetrie (kolejny wykład). Odnotujmy proste uwagi dotyczące wyznacznika izometrii.

Obserwacja 15.7

Jeśli ϕ jest izometrią przestrzeni euklidesowej liniowej $(V, \langle\,,\,\rangle)$, wówczas $\det\phi = \pm 1$.
Jeśli izometria ϕ ma rzeczywiste wartości własne, to są one równe ± 1.

Dowód. Niech $\mathcal{A}$ będzie bazą ortonormalną V. Wtedy $M(\phi)^{\mathcal{A}}_{\mathcal{A}}$ jest macierzą ortogonalną, więc

$$(M(\phi)^{\mathcal{A}}_{\mathcal{A}})^T \cdot M(\phi)^{\mathcal{A}}_{\mathcal{A}} = I.$$

Wyznacznik macierzy równy jest wyznacznikowi jej transpozycji, a zatem:

$$\det((M(\phi)^{\mathcal{A}}_{\mathcal{A}})^T \cdot M(\phi)^{\mathcal{A}}_{\mathcal{A}}) = (\det M(\phi)^{\mathcal{A}}_{\mathcal{A}})^2 = (\det\phi)^2 = 1.$$

Skoro $\det\phi \in \mathbb{R}$, to $\det\phi = \pm 1$. Jeśli natomiast dla $\alpha \neq 0$ mamy $\phi(\alpha) = \lambda\cdot\alpha$, to skoro ϕ jest izometrią, mamy

$$\|\alpha\| = \|\phi(\alpha)\| = \|\lambda\cdot\alpha\| = |\lambda|\cdot\|\alpha\|.$$

Stąd $|\lambda| = 1$. □

Zauważmy, że z wyżej wypisanych twierdzeń wynika, że w $\mathbb{R}^2$ oraz $\mathbb{R}^3$ jedynie obroty zachowują orientację. Jak się okaże, zbiory obrotów w $\mathbb{R}^2$ oraz $\mathbb{R}^3$ tworzą ważne obiekty, zwane specjalnymi grupami ortogonalnymi. Na sam koniec warto odnotować: czasem dzięki wyznacznikowi jesteśmy w stanie szybko stwierdzić czy dane przekształcenie liniowe jest danym typem izometrii, czy nie. Na przykład izometria o wyznaczniku równym -1 nie może być obrotem, ani złożeniem dwóch odbić. Jeszcze inna uwaga: jeśli np. wiemy, że izometria ϕ jest symetrią, to jak znaleźć podprzestrzeń, względem której jest ta symetria? Oczywiście jest to podprzestrzeń własna ϕ odpowiadająca wartości własnej 1.

Omówiwszy izometrie przestrzeni liniowej i związane z nimi macierze ortogonalne, przechodzimy do wątków geometrycznych i pojęcia izometrii przestrzeni afinicznej. Zagadnienie to omawiamy bardzo pobieżnie, odsyłając jednak w dodatkach do niezliczonych wątków geometrycznych, zarówno elementarnych, jak i bardzo klasycznych i nieoczywistych, choćby dotyczących „ruchu sztywnego” w $\mathbb{R}^3$.

Definicja 15.5: Izometria przestrzeni euklidesowej afinicznej

Mówimy, że przekształcenie afiniczne $f : H_1 \to H_2$ jest IZOMETRIĄ przestrzeni euklidesowej afinicznej $(H_1, \langle\,,\rangle_1)$ na przestrzeń euklidesową afiniczną $(H_2, \langle\,,\rangle_2)$ jeśli:

- f jest izomorfizmem przestrzeni afinicznej H_1 na przestrzeń afiniczną H_2,
- $f' : T(H_1) \to T(H_2)$ jest izometrią $(T(H_1), \langle\,,\rangle_1)$ na $(T(H_2), \langle\,,\rangle_2)$.

Dla każdej przestrzeni euklidesowej afinicznej $(H, \langle\,,\rangle)$ izometrię przestrzeni $(H, \langle\,,\rangle)$ na $(H, \langle\,,\rangle)$ nazywamy IZOMETRIĄ PRZESTRZENI EUKLIDESOWEJ AFINICZNEJ $(H, \langle\,,\rangle)$.

Przykłady.

- PRZESUNIĘCIE. Niech $(H, \langle\,,\rangle)$ będzie przestrzenią euklidesową afiniczną oraz niech $\alpha \in T(H)$. Przekształcenie afiniczne $f_\alpha : H \to H$. zadane wzorem $f_\alpha(p) = p + \alpha$, jest izometrią przestrzeni afinicznej $(H, \langle\,,\rangle)$. Pochodną przesunięcia jest identyczność.
- SYMETRIA PROSTOPADŁA. Dla podprzestrzeni M przestrzeni euklidesowej afinicznej $(H, \langle\,,\rangle)$ symetria prostopadła względem M jest izometrią H.
- OBRÓT WOKÓŁ PUNKTU. Niech $(H, \langle\,,\rangle)$ będzie 2-wymiarową przestrzenią euklidesową afiniczną i załóżmy, że w $T(H)$ wybrana jest orientacja. Dla dowolnego punktu $p \in H$ przekształcenie afiniczne $f : H \to H$ nazywamy OBROTEM WOKÓŁ PUNKTU p o kąt θ, gdy:
  - $f(p) = p$,
  - $f' : T(H) \to T(H)$ jest obrotem o kąt θ

  Dla $\dim M \geqslant 2$ przekształcenie afiniczne $f : M \to M$ jest obrotem o kąt θ wokół podprzestrzeni W wymiaru $\dim(M) - 2$, jeśli:
  - dla każdego $p \in W$ mamy $f(p) = p$,
  - $f'|_{T(W)^\perp}$ jest obrotem o kąt θ (na $T(W)^\perp$ musi być zadana orientacja).

Warto wspomnieć też dwa przykłady przekształceń, które nie są izometriami:

- homotetia o skali różnej od ± 1 nie jest izometrią,
- rzut prostopadły różny od id nie jest izometrią.

Uzyskujemy więc charakteryzację izomerii przestrzeni euklidesowych afinicznych.

Obserwacja 15.8

Jeśli $(H_1, \langle\,,\rangle_1)$, $(H_2, \langle\,,\rangle_2)$ są przestrzeniami euklidesowymi afinicznymi oraz $\mathcal{A}_i$ jest bazą ortonormalną przestrzeni $(T(H_i), \langle\,,\rangle_i)$, dla $i = 1, 2$, to przekształcenie afiniczne $f : H_1 \to H_2$ jest izometrią wtedy i tylko wtedy, gdy macierz $M(f')_{\mathcal{A}_1}^{\mathcal{A}_2}$ jest ortogonalna.

Z punktu widzenia geometrii afinicznej kluczową własnością izometrii jest jej odniesienie do pojęcia odległości, wyrażone w poniższej definicji.

Definicja 15.6

Niech $(H_1, \langle\,,\rangle_1)$, $(H_2, \langle\,,\rangle_2)$ będą przestrzeniami euklidesowymi afinicznymi i niech ρ_i oznacza odległość w przestrzeni $(H_i, \langle\,,\rangle_i)$, dla $i = 1, 2$. Powiemy, że funkcja $f : H_1 \to H_2$ ZACHOWUJE ODLEGŁOŚĆ PUNKTÓW, jeśli dla każdych $p, q \in H_1$ zachodzi:

$$\rho_1(p, q) = \rho_2(f(p), f(q)).$$

Obserwacja 15.9

Niech $(H, \langle\,,\rangle)$ będzie przestrzenią euklidesową afiniczną i niech ρ będzie odległością w $(H, \langle\,,\rangle)$. Dla dowolnego przekształcenia afinicznego $f : H \to H$ następujące warunki są równoważne:

(i) f jest izometrią przestrzeni H,

(ii) f zachowuje odległość punktów.

Dowód. Zacznijmy od (i) $\Rightarrow$ (ii). Zgodnie z (i) przekształcenie $f' : T(H) \to T(H)$ zachowuje iloczyn skalarny, więc f' zachowuje normę wektorów. Stąd dla każdych punktów $p, q \in H$ mamy:

$$\rho(f(p), f(q)) = \|\overrightarrow{f(p)f(q)}\| = \|f'(\overrightarrow{pq})\| = \|\overrightarrow{pq}\| = \rho(p, q),$$

czyli f zachowuje odległość punktów. Również dowód (ii) $\Rightarrow$ (i) jest natychmiastowy. Przypuśćmy, że f jest przekształceniem afinicznym. Wówczas dla każdego punktu $p \in H$ oraz każdego wektora $\alpha \in T(H)$ mamy $f'(\alpha) = \overrightarrow{f(p)f(p+\alpha)}$, więc:

$$\|f'(\alpha)\| = \|\overrightarrow{f(p)f(p+\alpha)}\| = \rho(f(p), f(p+\alpha)) \overset{(ii)}{=} \rho(p, p+\alpha) = \|\overrightarrow{p\,(p+\alpha)}\| = \|\alpha\|.$$

Zatem f' zachowuje długość wektorów, czyli f jest izometrią przestrzeni $(H, \langle\,,\rangle)$. □

Naszym celem jest pokazanie, że dowolna funkcja z przestrzeni afinicznej $(H, \langle\,,\rangle)$ w $(H, \langle\,,\rangle)$ – nie tylko przekształcenie afiniczne, zachowuje odległość na H wtedy i tylko wtedy, gdy jest jej izometrią. Jak jednak badać dowolne funkcje zachowujące odległość? Trzeba pokazać, że funkcje takie muszą być przekształceniami afinicznymi.

Twierdzenie 15.2: Twierdzenie Mazura-Ulama

Niech $(H, \langle\,,\rangle)$ będzie przestrzenią euklidesową afiniczną i niech ρ będzie odległością w $(H, \langle\,,\rangle)$. Jeśli funkcja $f : H \to H$ zachowuje odległość ρ, to f jest przekształceniem afinicznym.

Dowód jest długi, ale idea jest następująca. Weźmiemy dowolną funkcję f zachowującą odległość punktów i zobaczymy jak zachowuje się ona na pewnym układzie ortonormalnym w H. Potem weźmiemy izometrię g, która na tym samym układzie ortonormalnym w H zachowuje się jak f i pokażemy, że f i g muszą być tym samym przekształceniem. Ten sposób dowodzenia jest bardzo charakterystyczny.

Dowód. **Krok 1.** Wykażemy, że dla każdej bazy punktowej $p_0, \ldots, p_n$ przestrzeni H położenie punktu w przestrzeni H jest jednoznacznie wyznaczone przez odległości od punktów $p_0, \ldots, p_n$. Innymi słowy, jeśli dla pewnych $x, y \in H$ mamy $\rho(x, p_i) = \rho(y, p_i)$, dla $i = 0, \ldots, n$, to koniecznie $x = y$.

Oznaczmy $\overrightarrow{p_0x} = \beta$, $\overrightarrow{p_0y} = \gamma$ oraz $\overrightarrow{p_0p_i} = \alpha_i$, dla $i = 1, \ldots, n$. Wówczas:

$$\begin{aligned}
\overrightarrow{xp_i} &= \overrightarrow{xp_0} + \overrightarrow{p_0p_i} = -\beta + \alpha_i = \alpha_i - \beta \\
\overrightarrow{yp_i} &= \overrightarrow{yp_0} + \overrightarrow{p_0p_i} = -\gamma + \alpha_i = \alpha_i - \gamma \\
\overrightarrow{xy} &= \overrightarrow{xp_0} + \overrightarrow{p_0y} = -\beta + \gamma = \gamma - \beta.
\end{aligned}$$

Stąd dla każdego $i = 1, \ldots, n$ mamy ciąg równoważności:

$$\begin{aligned}
\rho(x, p_i) = \rho(y, p_i) &\Leftrightarrow \|\overrightarrow{xp_i}\|^2 = \|\overrightarrow{yp_i}\|^2 \Leftrightarrow \\
&\Leftrightarrow \|\alpha_i - \beta\|^2 = \|\alpha_i - \gamma\|^2 \Leftrightarrow \\
&\Leftrightarrow \|\alpha_i\|^2 - 2\langle\,\alpha_i, \beta\,\rangle + \|\beta\|^2 \\
&= \|\alpha_i\|^2 - 2\langle\,\alpha_i, \gamma\,\rangle + \|\gamma\|^2 \Leftrightarrow \\
&\Leftrightarrow \|\alpha_i\|^2 - 2\langle\,\alpha_i, \beta\,\rangle + \rho(x, p_0)^2 \\
&= \|\alpha_i\|^2 - 2\langle\,\alpha_i, \gamma\,\rangle + \rho(y, p_0)^2 \Leftrightarrow \\
&\Leftrightarrow 2\langle\,\alpha_i, \beta\,\rangle = 2\langle\,\alpha_i, \gamma\,\rangle \Leftrightarrow \\
&\Leftrightarrow \langle\,\alpha_i, \beta - \gamma\,\rangle = 0
\end{aligned}$$

Układ $\alpha_1, \ldots, \alpha_n$ jest bazą przestrzeni $T(H)$, więc równości $\langle \alpha_i, \beta - \gamma \rangle = 0$, dla $i = 1, \ldots, n$, oznaczają, że $\beta - \gamma$ jest prostopadły do $T(H)$, czyli $\beta = \gamma$. Stąd $x = y$.

Krok 2. Niech $f : H \to H$ będzie funkcją zachowującą odległość punktów. Niech $p_0; \alpha_1, \ldots, \alpha_n$ będzie ortonormalnym układem bazowym przestrzeni H i niech $p_i = p_0 + \alpha_i$, dla $i = 1, \ldots, n$. Niech $q_i = f(p_i)$, dla $i = 0, \ldots, n$ oraz $\beta_i = \overrightarrow{q_0 q_i}$, dla $i = 1, \ldots, n$. Wykażemy, że $q_0; \beta_1, \ldots, \beta_n$ też jest ortonormalnym układem bazowym w H.

Po pierwsze zauważmy, że funkcja f przeprowadza wektory o normie 1 na wektory o normie 1.

$$\|\beta_i\| = \|\overrightarrow{q_0 q_i}\| = \rho(q_0, q_i) = \rho(f(p_0), f(p_i)) = \|\overrightarrow{p_0 p_i}\| = \|\alpha_i\| = 1.$$

Pozostaje pokazać, że f przeprowadza prostopadłe wektory na prostopadłe. Skorzystamy w tym celu z następującej obserwacji prawdziwej w każdej przestrzeni euklidesowej afinicznej: wśród wszystkich trójkątów równoramiennych o ramionach długości 1 tylko trójkąt prostokątny ma trzeci bok długości $\sqrt{2}$. Formalnie: dla dowolnych $p \in H$ oraz $\gamma, \delta \in T(H)$ spełniających $\|\gamma\| = \|\delta\| = 1$:

$$\gamma \perp \delta \Leftrightarrow \rho(p + \gamma, p + \delta) = \sqrt{2}.$$

Dowód $\rho(p + \gamma, p + \delta)^2 = \|\gamma - \delta\|^2 = \|\gamma\|^2 - 2\langle \gamma, \delta \rangle + \|\delta\|^2 = 2 - 2\langle \gamma, \delta \rangle$, więc $\langle \gamma, \delta \rangle = 0$ wtedy i tylko wtedy, gdy $\rho(p + \gamma, p + \delta)^2 = 2$.

Zauważmy, że dla każdego $i \neq j$ mamy $\rho(q_0 + \beta_i, q_0 + \beta_j) = \rho(q_i, q_j) = \rho(p_i, p_j) = \rho(p_0 + \alpha_i, p_0 + \alpha_j) = \sqrt{2}$, więc $\beta_i \perp \beta_j$, co kończy dowód ortonormalności układu $\beta_1, \ldots, \beta_n$.

Krok 3. Niech $p_0, \ldots, p_n$ oraz $q_0, \ldots, q_n$ będą bazami punktowymi z Kroku 2. Niech $g : H \to H$ będzie przekształceniem afinicznym zadanym warunkiem $g(p_i) = q_i$, dla $i = 0, \ldots, n$. Wykażemy, że $f = g$, czyli dla każdego $x \in H$ mamy $f(x) = g(x)$.

Przekształcenie g jest izometrią przestrzeni $(H, \langle , \rangle)$, bo g' przeprowadza bazę ortonormalną $\alpha_1, \ldots, \alpha_n$ na bazę ortonormalną $\beta_1, \ldots, \beta_n$. Dla każdego punktu $x \in H$ zachodzi

$$\rho(f(x), q_i) = \rho(f(x), f(p_i)) = \rho(x, p_i) = \rho(g(x), q_i),$$

dla $i = 0, \ldots, n$. Pokazaliśmy, że $q_0; \beta_1, \ldots, \beta_n$ jest układem bazowym przestrzeni H, więc $q_0, \ldots, q_n$ jest bazą punktową przestrzeni H. Zatem z powyższego ciągu równości i tezy Kroku 1 wynika, że $f(x) = g(x)$, dla każdego $x \in H$. Stąd f jest przekształceniem afinicznym, co kończy dowód. □

Wniosek 15.1

Niech $(H, \langle , \rangle)$ będzie przestrzenią euklidesową afiniczną i niech ρ będzie odległością w tej przestrzeni. Dla dowolnej funkcji $f : H \to H$ następujące warunki są równoważne:

(1) f jest izometrią przestrzeni $(H, \langle , \rangle)$,

(2) f zachowuje odległość punktów.

Warto dodać komentarz, że Twierdzenie Mazura-Ulama prawdziwe jest dla dowolnych przestrzeni unormowanych V i W nad $\mathbb{R}$, a dokładniej mówi, że dowolne przekształcenie $f : V \to W$, które jest surjekcją i zachowuje normę, jest afiniczne. Przestrzeń unormowana $(V, \|\cdot\|)$, to zgodnie definicją Banacha z 1922 roku[2], (który postawił ten problem) przestrzeń liniowa nad $\mathbb{R}$ (lub $\mathbb{C}$) z funkcją $\|\cdot\| : V \to \mathbb{R}$ spełniającą:

(a) dla dowolnych $x \in V$ mamy $\|x\| \geqslant 0$ oraz $\|x\| = 0$ wtedy i tylko wtedy, gdy $x = 0$,

(b) dla dowolnych $x \in V$ oraz $\lambda \in \mathbb{R}$ mamy $\|\lambda x\| = |\lambda| \cdot x$,

(c) dla dowolnych $x, y \in V$ mamy $\|x + y\| \leqslant \|x\| + \|y\|$.

Norma jest pojęciem ogólniejszym od tego poznanego w teorii przestrzeni euklidesowych. Są funkcje $\|\cdot\|$ spełniające warunki powyzej, dla których nie istnieje iloczyn skalarny $\langle , \rangle$, że $\|x\| = \langle x, x \rangle$ (przykłady pojawiły się na ćwiczeniach, np. funkcje $\|(x_1, x_2)\| = |x_1| + |x_2|$ lub $\|(x_1, x_2)\| = \max(|x_1|, |x_2|)$ na $\mathbb{R}^2$).

[2]Patrz słynny artykuł „Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales" w Fundamenta Mathematicae 3 (1922), 133-181, `http://matwbn.icm.edu.pl/ksiazki/fm/fm3/fm3120.pdf`, gdzie na stronie 135 jest jedna z pierwszych definicji przestrzeni liniowej i pierwsza definicja przestrzeni unormowanej.

15.1 Uzupełnienie. Rozpoznawanie i zliczanie izometrii

W zadaniach dotyczących izometrii istotne znaczenie mają dwa aspekty: pierwszym jest samo rozpoznawanie izometrii, a drugim — ich opis. Zobaczmy pewne przykłady, zaczynając od przykładu ze skryptu.

Zadanie 15.1. *Rozważmy przestrzenie liniowe $V_1 = V_2 = \mathbb{R}^2$, na których rozważamy iloczyny skalarne:*

$$\langle (x_1, x_2), (y_1, y_2) \rangle_1 = 10x_1y_1 + 14x_1y_2 + 14x_2y_1 + 20x_2y_2, \quad \langle (x_1, x_2), (y_1, y_2) \rangle_2 = x_1y_1 + x_2y_2$$

Pokaż, że $\phi : (V_1, \langle\,,\rangle_1) \to (V_1, \langle\,,\rangle_2)$ dane wzorem $\phi((x_1, x_2)) = (x_1 + 2x_2, 3x_1 + 4x_2)$ jest izometrią.

Możemy podejść do rozwiązania na kilka sposobów:

- sprawdzić, czy ϕ zachowuje normę,
- sprawdzić, czy ϕ przeprowadza bazę ortonormalną na ortonormalną,
- sprawdzić, czy macierz ϕ w bazach ortonormalnych jest ortonormalna.

Z uwagi na to, że każda z metod stanowi zarazem przypomnienie wiadomości, wykonajmy wszystkie trzy. Być może docenimy, że w tym przypadku pierwsza działa dość szybko.

- **Sposób 1**. (skrypt) — bierzemy wektor $\alpha = (x_1, x_2)$ oraz jego obraz $\phi(\alpha) = (x_1 + 2x_2, 3x_1 + 4x_2)$ i porównujemy ich normy w odpowiedniej przestrzeni euklidesowej:

  $$\|\alpha\|_1^2 = 10x_1^2 + 28x_1x_2 + 20x_2^2,$$
  $$\|\phi(\alpha)\|_2 = (x_1 + 2x_2)^2 + (3x_1 + 4x_2)^2.$$

  Po uproszczeniu widzimy, że wyrażenia te są równe. Ten sposób uderza nas jako krótki i naturalny.

- **Sposób 2**. Aby wyznaczyć jakąś bazę ortonormalną $(V_1, \langle\,,\rangle_1)$ weźmy wektor $(1, 0) \in V_1$. Nie ma on normy 1, ale wektor $\alpha_1 = \frac{1}{\sqrt{10}}(1, 0)$ ma. Zbiór wektorów prostopadłych do α_1 wyznaczamy z warunku:

  $$\begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 10 & 14 \\ 14 & 20 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 10x_1 + 14x_2 = 0.$$

  A zatem $\mathrm{lin}(\alpha_1)^\perp = \mathrm{lin}((14, -10))$. Mamy $\|(14, -10)\|_1 = \sqrt{40} = 2\sqrt{10}$, a zatem biorąc wektor $\alpha_2 = \frac{1}{\sqrt{10}}(7, -5)$ dostajemy bazę ortonormalną α_1, α_2 przestrzeni $(V_1, \langle\,,\rangle_1)$. Mamy też:

  $$\phi(\alpha_1) = \frac{1}{\sqrt{10}}(1, 3), \quad \phi(\alpha_2) = \frac{1}{\sqrt{10}}(-3, 1).$$

  Z łatwością widzimy, że w standardowym iloczynie skalarnym $\langle\,,\rangle_2$ normy tych obrazów wynoszą 1 i są to wektory prostopadłe. A zatem rzeczywiście pewna baza ortonormalna $(V_1, \langle\,,\rangle_1)$ przechodzi na bazę ortonormalną $(V_2, \langle\,,\rangle_2)$.

- **Sposób 3**. Skoro już w poprzednim sposobie wyznaczyliśmy bazy ortonormalne, to oznaczając bazy $\mathcal{A} = (\alpha_1, \alpha_2)$ oraz $\mathcal{B} = ((1, 0), (0, 1))$ mamy

  $$M(\phi)_{\mathcal{A}}^{\mathcal{B}} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}.$$

  Oczywiście macierz ta jest ortogonalna. A można jeszcze prościej: wziąć $\mathcal{C} = (\phi(\alpha_1), \phi(\alpha_2))$ i mamy:

  $$M(\phi)_{\mathcal{A}}^{\mathcal{C}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

  Niech ten ostatni przykład nie zdziwi Czytelnika — nie oznacza on oczywiście, że każda izometria ma wyznacznik 1. Dlaczego?

Zobaczyliśmy zatem jak fakty dowiedzione na wykładzie działają na opisanym przykładzie. Oczywiście wiedząc, że przekształcenie jest izometrią – a zwłaszcza wiedząc jaka to izometria (np. obrót wokół danej podprzestrzeni, symetria prostopadła względem podprzestrzeni) możemy wyznaczać macierze, wzory tych przekształceń wykorzystując choćby wzór z Obserwacji 17.4. Zobaczmy teraz nieco trudniejszy przykład, związany z izometriami afinicznymi.

Zadanie 15.2. *W przestrzeni euklidesowej afinicznej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym dana jest prosta $L = (0,1,0) + \lin((1,2,1))$ oraz punkt $p_0 = (-3,-1,1)$.*

(a) Dla jakich $s \in \mathbb{R}$ istnieje izometria $f : \mathbb{R}^3 \to \mathbb{R}^3$ spełniająca $f(L) = L$ oraz $f(p_0) = (s,1,-s)$?

(b) Ile jest izometrii $g : \mathbb{R}^3 \to \mathbb{R}^3$ takich, że $g((0,0,1)) = (0,0,1)$ oraz dla każdego punktu $p \in L$ mamy $g(p) = p$? Dla każdej takiej izometrii obliczyć $g((0,0,0))$.

Rozwiązanie. Zacznijmy od (a). Izometria to przekształcenie zachowujące odległość. Zauważmy, że skoro prosta L przechodzi przy f na siebie, to odległość punktów $p_0 = (-3,-1,1)$ oraz $(s,1,-s)$ od prostej L musi być jednakowa. Odległość punktu od prostej L to norma rzutu ortogonalnego wektora łączącego ten punkt z dowolnym punktem L na $T(L)^\perp$. Wektor łączący $(-3,-1,1)$ oraz punkt $(0,1,0) \in L$ to $(3,2,-1)$. Przestrzeń $T(L)^\perp$ opisana jest równaniem $x_1 + 2x_2 + x_3 = 0$. Baza ortogonalna tej przestrzeni to $(1,0,-1)$ oraz $(1,-1,1)$. A zatem rzut ortogonalny $(3,2,-1)$ na $T(L)^\perp$ to :

$$\frac{\langle (3,2,-1),(1,0,-1)\rangle}{\langle (1,0,-1),(1,0,-1)\rangle}\cdot(1,0,-1) + \frac{\langle (3,2,-1),(1,-1,1)\rangle}{\langle (1,-1,1),(1,-1,1)\rangle}\cdot(1,-1,1) = (2,0,-2).$$

Wektor łączący $(s,1,-s)$ z $(0,1,0)$ to $(s,0,-s) \in T(L)^\perp$. A zatem aby f było izometrią potrzeba, aby $\|(2,0,-2)\| = \|(s,0,-s)\|$, czyli $s = \pm 2$.

Pokażmy, że dla $s = \pm 2$ rozważane izometrie istnieją. Rozważmy wektor łączący punkt $(0,1,0)$ z punktem $f(p_0) = (s,1,-s)$. Jest to wektor $(s,0,-s) \in T(L)^\perp$, bo przestrzeń $T(L)^\perp$ opisana jest równaniem $x_1 + 2x_2 + x_3 = 0$. Skoro $f(L) = L$, to $(0,1,0) = f(q_0)$, dla pewnego $q_0 = (0,1,0) + t(1,2,1) \in L$. Skoro wektor łączący $f(q_0)$ oraz $f(p_0)$ jest prostopadły do $T(L)$, to wobec faktu, że f' jest izometrią (zachowującą iloczyn skalarny), również wektor łączący q_0 oraz $p_0 = (-3,-1,1)$ musi być prostopadły do L. Mamy $\overrightarrow{q_0p_0} = (t, 2t+1, t) - (-3,-1,1) = (t+3, 2t+2, t-1)$. Ten wektor należy do $T(L)^\perp$, jeśli spełnia równanie $x_1 + 2x_2 + x_3 = 0$, czyli $t + 3 + 4t + 4 + t - 1 = 0$, co daje $t = -1$. Stąd

$$q_0 = (-1,-1,-1), \quad f((-1,-1,-1)) = (0,1,0).$$

A zatem na prostej L przekształcenie f jest przesunięciem o wektor $(1,2,1)$. Dla parametru $s = -2$ mamy $f((-3,-1,1)) = (-2,1,2)$, więc za szukaną izometrię f na $\mathbb{R}^3$ można wziąć przesunięcie o wektor $(1,2,1)$. Dla $s = 2$ niech f będzie przesunięciem o wektor $(1,2,1)$ złożonym z odbiciem względem prostej L wzdłuż $T(L)^\perp \ni \lin(1,0,-1)$. Wtedy f to izometria, $f(L) = L$ oraz $f((-3,-1,1)) = (2,0,-2)$, bo przy tym odbiciu punkt $(-2,1,2) = (0,1,0) + (-2,0,2)$ przechodzi na $(0,1,0) - (-2,0,2) = (2,1,-2)$.

Przechodzimy do punktu (b). Skoro dla każdego $p \in L$ mamy mieć $g(p) = p$, to w szczególności $g((0,1,0)) = (0,1,0)$ oraz $g(-1,-1,-1) = (-1,-1,-1)$. Mamy też oczywiście $g((0,0,1)) = (0,0,1)$. W rezultacie $g'((0,1,-1)) = (0,1,-1)$ oraz $g'((1,2,1)) = (1,2,1)$, czyli g' jest identycznością na podprzestrzeni $\lin((0,1,-1),(1,2,1))$. Przestrzeń prostopadła do tej dwuwymiarowej przestrzeni stałej (dla g') jest jednowymiarowa i złożona z wektorów (x_1,x_2,x_3) spełniających:

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

czyli jest to przestrzeń $\lin((-3,1,1))$. Ta przestrzeń jest niezmiennicza przy izometrii g', bo izometria zachowuje prostopadłość. Zatem wektor $(-3,1,1)$ przechodzi na $t(-3,1,1)$, dla pewnego t. Jednak jego norma nie zmienia się przy izometrii, więc $t = \pm 1$. Obrazem $(-3,1,1)$ przy g' musi być więc $(-3,1,1)$ lub $(3,-1,-1)$. Mamy zatem dwie izometrie g spełniające warunki zadania. Rozważmy je osobno.

- Jeśli $g'((-3,1,1)) = (-3,1,1)$, to $g' = \mathrm{id}$ i skoro $g((0,0,1)) = (0,0,1)$, to dostajemy $g = \mathrm{id}$. Stąd $g((0,0,0)) = (0,0,0)$.
- Jeśli $g'((-3,1,1)) = (3,-1,-1)$, to g' jest symetrią prostopadłą względem płaszczyzny T postaci $(0,0,1) + \lin((0,1,-1),(1,2,1))$. Płaszczyzna ta jest opisana równaniem $-3x_1 + x_2 + x_3 = 1$. Rzut prostopadły $(0,0,0)$ na tą płaszczyznę to pewien punkt $(0,0,0) + t(-3,1,1)$, który spełnia równanie wyżej. Stąd $t = \frac{1}{11}$. A zatem obraz $(0,0,0)$ w symetrii prostopadłej względem płaszczyzny T to

$$(0,0,0) + \frac{2}{11}(-3,1,1) = \left(-\frac{6}{11}, \frac{2}{11}, \frac{2}{11}\right) = g((0,0,0)).$$

■

Zadanie 15.3. *Dla jakich $s,t \in \mathbb{R}$ istnieje izometria ϕ przestrzeni $\mathbb{R}^3$ ze standardowym iloczynem skalarnym taka, że $\phi(0,4,s) = (4,t,0)$ oraz $\phi(1,-4,1) = (4,1,1)$?*

Rozwiązanie. Skoro ϕ jest izometrią to musi zachowywac standardowy iloczyn skalarny i normę wyznaczoną przez ten iloczyn. A zatem

$$\|(0,4,s)\| = \|(4,t,0)\| \quad \text{oraz} \quad \|(1,-4,1)\| = \|(4,1,1)\|,$$

a także

$$\langle (0,4,s),(1,-4,1)\rangle = \langle (4,t,0),(4,1,1)\rangle.$$

Pierwszy warunek oznacza, że $s^2 = t^2$, drugi zaś, że $-16+s = 16+t$. A zatem $s \neq t$ i dostajemy $s = -t = 16$. Oczywiście istnieje izometria $(\mathbb{R}^3, \langle\,,\rangle_{st})$ spełniająca powyższe warunki. Aby wypisać jej wzór wystarczy na przykład zażądać, by wektor $(0,4,16)\times(1,-4,1)$ przechodził przy ϕ na $(4,-16,0)\times(4,1,1)$. ■

* * *

Zobaczmy wreszcie przykład ze zliczaniem izometrii związany z pojęciem orientacji.

Zadanie 15.4. *Niech układ $(v_1,...,v_n)$ będzie bazą zorientowanej przestrzeni euklidesowej $(V,\langle,\rangle)$. Ile jest zachowujących orientację izometrii $\phi : V \to V$ takich, że*

$$\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$$

dla wszystkich $i = 1,...,n$?

Rozwiązanie. Niech $(v_1',\ldots,v_n')$ będzie bazą ornonormalną $(V,\langle,\rangle)$ uzyskaną poprzez ortogonalizację Grama-Schmidta. Wówczas

$$\mathrm{lin}(v_1,\ldots,v_i) = \mathrm{lin}(v_1',\ldots,v_i'),$$

dla każdego $i = 1,\ldots,n$. Stąd $\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$ jest równoważny z następującym warunkiem $\phi(\mathrm{lin}(v_1',...,v_i')) = \mathrm{lin}(v_1',...,v_i')$ i możemy od początku założyć, że baza $(v_1',\ldots,v_n')$ jest ortogonalna.

Policzymy najpierw wszystkie izometrie takie, że $\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla każdego $i = 1,2,\ldots,n$, przy czym $(v_1,\ldots,v_n)$ jest bazą ortonormalną. Wykażemy przez indukcję po n, że izometrii tych jest 2^n.

Dla $n = 1$ mamy $\phi(v_1) = v_1$ lub $\phi(v_1) = -v_1$, bo ϕ zachowuje normę wektora. Przechodzimy do kroku indukcyjnego. Weźmy dowolną izometrię $(V,\langle,\rangle)$ spełniającą warunki $\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla każdego $i = 1,2,\ldots,n$, i obetnijmy ją do podprzestrzeni $W = \mathrm{lin}(v_1,...,v_{n-1})$. Zgodnie z założeniem $\phi(\mathrm{lin}(v_1,...,v_{n-1})) = \mathrm{lin}(v_1,...,v_{n-1})$ obcięcie $\phi|_W$ jest „na". Jego jądro jest zerowe, bo jądro ϕ jest zerowe. Oczywiscie $\phi|_W$ zachowuje długość, więc to izometria. Co więcej, $\phi|_W(v_1,\ldots,v_i) = (v_1,\ldots,v_i)$, dla $i = 1,\ldots,n-1$. A zatem z założenia indukcyjnego wiemy, że $\phi|_W$ można wybrać na 2^{n-1} sposobów. Zauważmy teraz, że $\phi(v_n) \notin W$, bo $v_n \in W^\perp$, skoro baza $(v_1,\ldots,v_n)$ jest ortonormalna. A zatem $\phi(v_n) = v_n$ lub $\phi(v_n) = -v_n$. A zatem mamy 2^n izometrii $(V,\langle,\rangle)$ spełniających warunek $\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla $i = 1,\ldots,n$. Krok indukcyjny jest zakończony.

Każda izometria przestrzeni zorientowanej $(V,\langle,\rangle)$ zachowuje orientację lub jej nie zachowuje. Co więcej, każdej izometrii ϕ_1 spełniającej warunek $\phi_1(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla $i = 1,\ldots,n$ zachowującej orientację $(V,\langle,\rangle)$ odpowiada dokładnie jedna izometria ϕ_2, która spełnia warunek $\phi_2(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla $i = 1,\ldots,n$ i nie zachowuje orientacji. Wystarczy żądać, by $\phi_1(v_i) = \phi_2(v_i)$, dla $i = 1,\ldots,n-1$ oraz $\phi_2(v_n) = -\phi_1(v_n)$. A zatem liczba izometrii $(V,\langle,\rangle)$ spełniających warunek $\phi(\mathrm{lin}(v_1,...,v_i)) = \mathrm{lin}(v_1,...,v_i)$, dla $i = 1,\ldots,n$ wynosi 2^{n-1}. ■

Rozdział 16

Składanie izometrii

Złożenie izometrii jest izometrią. Ta oczywista obserwacja prowadzi do pięknych wniosków geometrycznych. Na tym wykładzie powiemy o rozkładaniu izometrii liniowych na symetrie prostopadłe, a dokładniej – na tak zwane odbicia. Na koniec (i w dodatku) przyjrzymy się podstawowym przykładom i faktom dotyczącym składania izometrii w przestrzeniach afinicznych. Zacznijmy od kilku przykładów.

W przestrzeni euklidesowej liniowej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym, zorientowanej zgodnie z bazą standardową, rozważmy endomorfizm ϕ dany w bazie standardowej macierzą:

$$M(\phi)^{st}_{\text{st}} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Czytelnik bez trudu widzi, że jest to macierz izometrii. W istocie: jest to macierz przekształcenia liniowego przestrzeni euklidesowej w bazach ortonormalnych — i jest ona ortogonalna. Co więcej, mamy:

$$M(\phi)^{st}_{\text{st}} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

co oznacza, że ϕ jest w istocie złożeniem **przemiennych ze sobą** izometrii:

- symetrii prostopadłej względem $\text{lin}((1,0,0),(0,1,0))$,
- obrotu wokół $\text{lin}((0,0,1))$ o kąt θ.

Rozważmy izometrię $\phi \in \text{End}(\mathbb{R}^3, \langle\,,\rangle_{st})$ daną macierzą:

$$M(\phi)^{st}_{st} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}.$$

Znowu bez trudu widzimy, że macierz jest ortogonalna. Rozważmy analogiczny do powyższego rozkład:

$$M(\phi)^{st}_{st} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}}_{B}.$$

O ile widzimy, że macierz A po lewej to macierz symetrii prostopadłej względem $\text{lin}((0,1,0),(0,0,1))$, to może nie być jasne, że macierz B po prawej jest w istocie macierzą obrotu. Aby to stwierdzić, trzeba sprawdzić, że jedną z wartości własnych macierzy B jest 1, a odpowiadający tej wartości własnej wektor własny to $(1,1,-1)$. Innych rzeczywistych wartości własnych ta macierz nie ma. Okazuje się, że biorąc $\text{lin}((1,1,-1))^{\perp}$ i ograniczając do tej podprzestrzeni izometrię zadaną przez macierz B dostajemy... obrót o kąt $\pm\frac{4\pi}{3}$, w zależności od przyjętej orientacji. Podsumowując: ϕ jest w istocie złożeniem (od prawej, zgodnie z kolejnością składania):

- obrotu wokół $\text{lin}((1,1,-1))$ o kąt $\pm\frac{4\pi}{3}$,
- symetrii prostopadłej względem $\text{lin}((0,1,0),(0,0,1))$.

Aby dać Czytelnikowi jaśniejszy pogląd na to co my tu tak naprawdę robimy, i w jaki sposób formułujemy wnioski o powyższej macierzy, odnotujmy nastepującą prostą obserwację.

Obserwacja 16.1

Niech $\phi : V \to V$ będzie izometrią przestrzeni euklidesowej liniowej $(V, \langle , \rangle)$. Jeśli $W \subseteq V$ jest podprzestrzenią ϕ-niezmienniczą, tzn. $\phi(W) \subseteq W$, to $W^\perp$ jest również podprzestrzenią ϕ-niezmienniczą. W szczególności $\phi|_W$ oraz $\phi|_{W^\perp}$ są izometriami i istnieje baza ortonormalna V taka, że macierz ϕ ma w niej postać blokowo-diagonalną:

$$\begin{bmatrix} M(\phi|_W) & 0 \\ 0 & M(\phi|_{W^\perp}) \end{bmatrix} = \begin{bmatrix} M(\phi|_W) & 0 \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} I & 0 \\ 0 & M(\phi|_{W^\perp}) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & M(\phi|_{W^\perp}) \end{bmatrix} \cdot \begin{bmatrix} M(\phi|_W) & 0 \\ 0 & I \end{bmatrix}.$$

Dowód. Niech $\beta \in W^\perp$. Chcemy pokazać, że $\phi(\beta) \in W^\perp$. Wystarczy pokazać, że dla każdego $\alpha \in W$ mamy $\langle \phi(\beta), \alpha \rangle = 0$. Skoro ϕ jest izomorfizmem, to $\phi|_W : W \to W$ jest mono, czyli jest izomorfizmem (z uwagi na skończony wymiar). Czyli $\phi^{-1}(W) = W$. Skoro jednak ϕ jest izometrią, to

$$\langle \phi(\beta), \alpha \rangle = \langle \beta, \phi^{-1}(\alpha) \rangle = 0.$$

Druga część jest teraz jasna: biorąc bazę ortonormalną $\mathcal{A} = (\alpha_1, \ldots, \alpha_k)$ przestrzeni W oraz bazę ortonormalną $\mathcal{B} = (\alpha_{k+1}, \ldots, \alpha_n)$ widzimy, że macierz ϕ ma w bazie $\alpha_1, \ldots, \alpha_n$ żądaną postać. □

Gdy teraz Czytelnik spojrzy na przykład macierzy

$$M(\psi)^{st}_{st} = \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$$

rozważanej wyżej, to wiedząc, że ma ona jedynie jedną wartość własną 1 o jednowymiarowej podprzestrzeni własnej $W = \mathrm{lin}((1, 1, -1))$, będzie jasne, że $W^\perp$ jest dwumiarową podprzestrzenią ψ-niezmienniczą. A to znaczy, że $\psi|_{W^\perp}$ jest izometrią przestrzeni dwuwymiarowej, która nie ma wartości własnych. To musi być zatem obrót. Kąt ujawni się po wypisaniu macierzy $\psi|_{W^\perp}$ w bazie ortonormalnej $W^\perp$.

Naszym celem będzie teraz zwrócenie się w bardziej geometryczną stronę i opowiedzenie o rozkładzie dowolnej izometrii liniowej na odbicia. Czym one są?

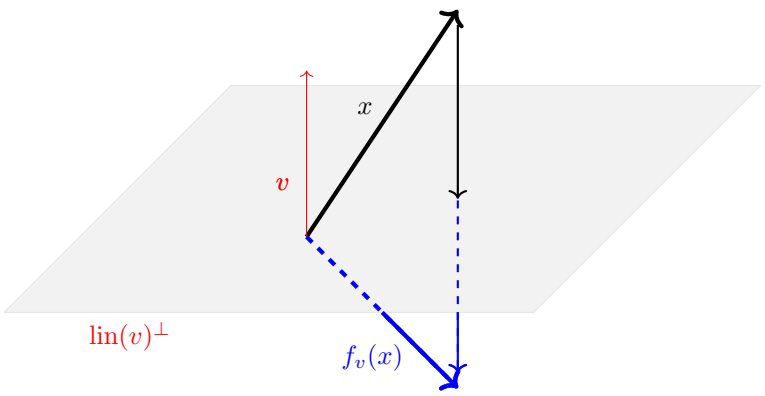
Definicja 16.1: Odbicie

Załóżmy, że $(V, \langle , \rangle)$ jest przestrzenią euklidesową liniową. Dla dowolnego wektora $0 \neq v \in V$ symetrię prostopadłą V względem podprzestrzeni $\mathrm{lin}(v)^\perp$ określoną wzorem

$$f_v(x) = x - 2\frac{\langle x, v \rangle}{\langle v, v \rangle} v$$

nazywamy ODBICIEM V względem $\mathrm{lin}(v)^\perp$.

Oto ilustracja odbicia.

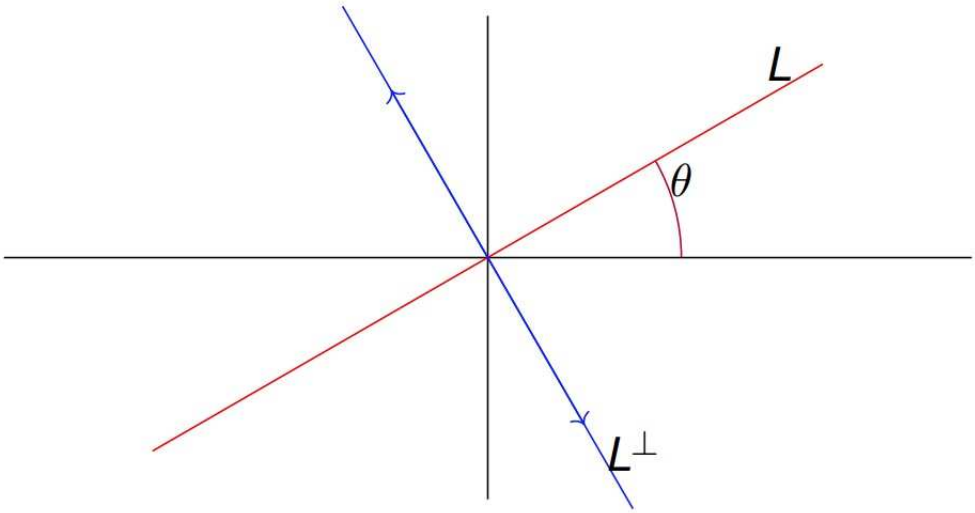


Warto przedstawić ilustrację odbicia na płaszczyźnie. Rozważmy kierunek nachylony do osi OX (kierunku $(1,0)$) pod kątem θ, czyli wektor $(\cos\theta, \sin\theta)$. Zauważmy, że:

$$\begin{bmatrix}\cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta\end{bmatrix} \cdot \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} = \begin{bmatrix}\cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta 0\cos^2\theta\end{bmatrix} \cdot \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} = \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}$$

$$\begin{bmatrix}\cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta\end{bmatrix} \cdot \begin{bmatrix}-\sin\theta \\ \cos\theta\end{bmatrix} = \begin{bmatrix}\cos^2\theta - \sin^2\theta & 2\cos\theta\sin\theta \\ 2\cos\theta\sin\theta & \sin^2\theta 0\cos^2\theta\end{bmatrix} \cdot \begin{bmatrix}-\sin\theta \\ \cos\theta\end{bmatrix} = -\begin{bmatrix}-\sin\theta \\ \cos\theta\end{bmatrix}$$

Widzimy zatem, że macierz kwadratowa po lewej definiuje odbicie prostopadłe do prostej L nachylonej pod kątem θ do osi OX (zdefiniowaliśmy je na bazie ortonormalnej).

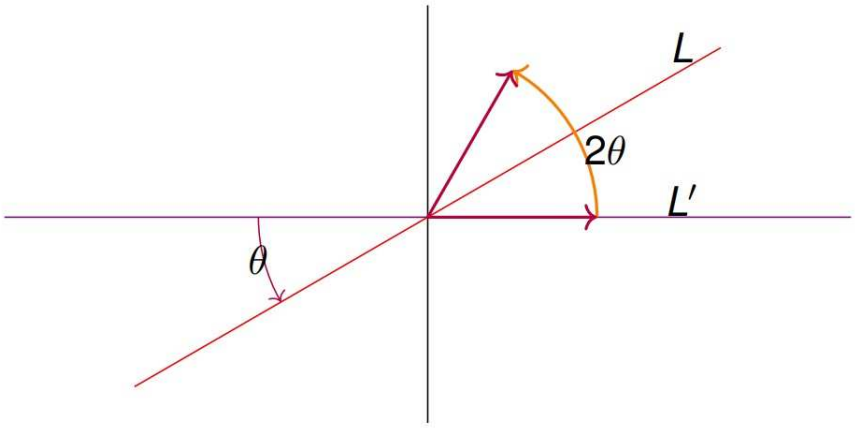


Źródło: materialy prof. Jarosława Wiśniewskiego

Po co nam teraz ta prosta pozioma $\text{lin}(1,0)$? Otóż kluczowe jest to, że znamy macierz odbicia względem niej. Gdy wykonamy to odbicie, a potem odbicie względem $\text{lin}((\cos\theta, \sin\theta))$, dostaniemy macierz

$$\begin{bmatrix}\cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta\end{bmatrix} \cdot \begin{bmatrix}1 & 0 \\ 0 & -1\end{bmatrix} = \begin{bmatrix}\cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta\end{bmatrix}.$$

A zatem złożenie dwóch odbić względem prostych pomiędzy którymi jest kąt θ, jest obrotem o 2θ w kierunku zależnym od kolejności złożenia odbić.



Źródło: materialy prof. Jarosława Wiśniewskiego

Widzimy zatem, że na płaszczyźnie każdą izometrię można przedstawić jako złożenie nie więcej niż 2 odbić. Okazuje się, że twierdzenie to przenosi się na izometrie liniowe wyższych wymiarów.

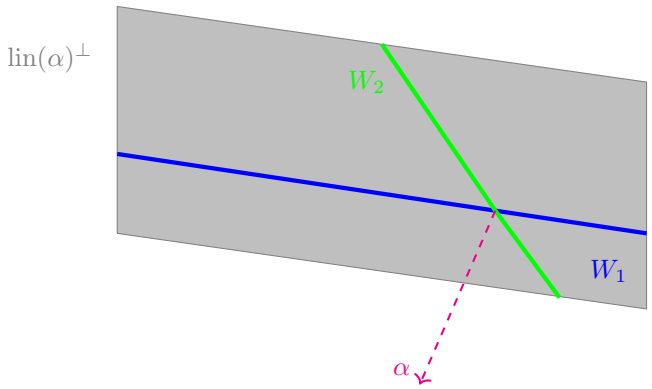
Twierdzenie 16.1: Cartan

Niech $(V, \langle , \rangle)$ będzie n wymiarową przestrzenią euklidesową liniową. Dla każdej izometrii ϕ przestrzeni $(V, \langle , \rangle)$ istnieje liczba $k \leqslant n$ taka, że ϕ jest złożeniem k (być może różnych) odbić przestrzeni V (tzn. symetrii prostopadłych wzgl. podprzestrzeni wymiaru $n-1$).

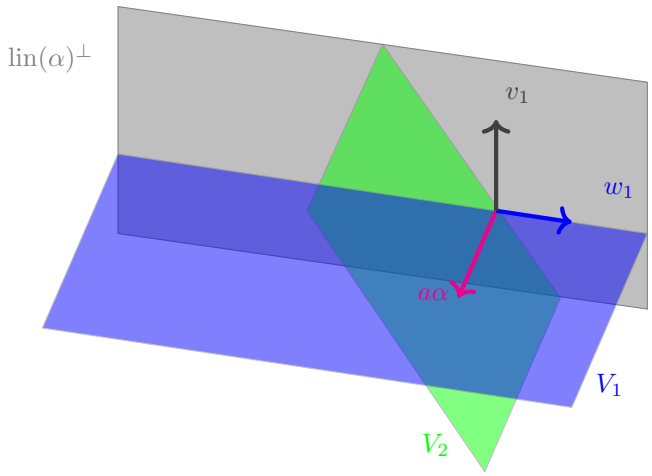
Dowód. Stosujemy indukcję po n. Dla $n = 0$ nie ma czego dowodzić. Dla $n = 1$ twierdzenie jest oczywiste, bo są tylko dwie izometrie 1-wymiarowej przestrzeni euklidesowej liniowej: id oraz $-\mathrm{id}$. Dla pierwszej mamy $k = 0$, dla drugiej zaś $k = 1$. Wyżej ilustrowaliśmy tezę dla $n = 2$.

Załóżmy, że udowodniliśmy twierdzenie dla przestrzeni euklidesowych wymiaru mniejszego od $n > 1$. Niech $\phi : V \to V$ będzie izometrią n wymiarowej przestrzeni euklidesowej liniowej $(V, \langle , \rangle)$. Wybierzmy dowolny wektor $0 \neq \alpha \in V$ i rozpatrzmy $\mathrm{lin}(\alpha)$. Możliwe są dwa przypadki: $\phi(\alpha) = \alpha$ oraz $\phi(\alpha) \neq \alpha$.

- Przypadek 1, gdy $\phi(\alpha) = \alpha$. Wówczas zgodnie z Obserwacją 16.1 $\mathrm{lin}(\alpha)$ oraz $\mathrm{lin}(\alpha)^\perp$ są ϕ-niezmiennicze, bo α to wektor własny. Przekształcenie $\phi|_{\mathrm{lin}(\alpha)^\perp}$ jest zatem izometrią przestrzeni euklidesowej $\mathrm{lin}(\alpha)^\perp$ wymiaru $n-1$ z iloczynem skalarnym będącym obcięciem $\langle , \rangle$ do $\mathrm{lin}(\alpha)^\perp$. Na mocy założenia indukcyjnego mamy dwa podprzypadki:
  - Gdy $\phi|_{\mathrm{lin}(\alpha)^\perp} = \mathrm{id}_{\mathrm{lin}(\alpha)^\perp}$ (złożenie 0 symetrii), wtedy $\phi(\alpha) = \alpha \Rightarrow \phi = \mathrm{id}_V$.
  - Gdy $\phi|_{\mathrm{lin}(\alpha)^\perp} = \phi_k \circ \ldots \circ \phi_1$, gdzie $k \leqslant n-1$ oraz $\phi_i : \mathrm{lin}(\alpha)^\perp \to \mathrm{lin}(\alpha)^\perp$ jest odbiciem $\mathrm{lin}(\alpha)^\perp$ względem podprzestrzeni $W_i \subseteq \mathrm{lin}(\alpha)^\perp$. Zobaczmy poglądowy rysunek sytuacji w przypadku $V = \mathbb{R}^3$. Wówczas $\mathrm{lin}(\alpha)^\perp$ to płaszczyzna i przyjmijmy, że nasza izometria obcięta do $\mathrm{lin}(\alpha)^\perp$ to złożenie dwóch odbić:



Chcemy „podnieść ten rozkład" do rozkładu endomorfizmu ϕ na całej przestrzeni V. Inaczej, chcemy pokazać, że $\phi = \psi_k \circ \ldots \circ \psi_1$, gdzie ψ_i są pewnymi odbiciami, które po obcięciu do $\mathrm{lin}(\alpha)^\perp$ są przekształceniami ϕ_i. Określamy $V_i = \mathrm{lin}(\alpha) + W_i$, dla $i = 1, \ldots, k$. Skoro $W_i \subset \mathrm{lin}(\alpha)^\perp$, to $V_i = \mathrm{lin}(\alpha) \oplus W_i$.



Zauważmy, że jeśli ϕ_i jest symetrią prostopadłą względem W_i w przestrzeni $\lin(\alpha)^\perp$, to przekształcenie liniowe („podniesienie") $\psi_i : V \to V$ określone, dla $w \in \lin(\alpha)^\perp$, wzorem

$$\psi_i(a\alpha + w) = a\alpha + \phi_i(w)$$

jest symetrią V względem $n-1$ wymiarowej podprzestrzeni V_i. Rzeczywiście, możemy zapisać $w = a\alpha + w = a\alpha + w_i + v_i$, gdzie $w_i \in W_i, v_i \in (W_i^\perp \cap \lin(\alpha)^\perp)$. W ten sposób:

$$\psi_i(a\alpha + w) = a\alpha + \phi_i(w) = a\alpha + \phi_i(w_i + v_i) = a\alpha + w_i - v_i,$$

czyli ψ jest odbiciem V względem V_i. Pozostaje pokazać, że mamy $\phi = \psi_k \circ \ldots \circ \psi_1$. Rzeczywiście, dwa endomorfizmy V określone identycznie na dwóch składnikach sumy prostej $V = W \oplus W'$ muszą być identycznie, bo pokrywają się na pewnej bazie całej V (złożonej z baz każdego ze składników).

Obetnijmy więc złożenie $\psi_k \circ \ldots \circ \psi_1$ do $\lin(\alpha)$. Dostaniemy identyczność tak, jak dla ϕ, bo $\phi(\alpha) = \alpha$. Jeśli obetniemy złożenie do $\lin(\alpha)^\perp$ to mamy $\phi_k \circ \ldots \circ \phi_1$, podobnie jak dla ϕ. Stąd $\phi = \psi_k \circ \ldots \circ \psi_1$, bowiem przekształcenia te są identyczne na całym V.

- Przypadek 2, gdy $\phi(\alpha) \neq \alpha$. W rozumowaniu skorzystamy z następującego lematu.

Obserwacja 16.2

Załóżmy, że $(V, \langle\,,\,\rangle)$ jest przestrzenią euklidesową liniową. Jeżeli $u, v \in V$ spełniają $\|u\| = \|v\|$, to

$$u + v \perp u - v$$

oraz istnieje taka izometria f, będąca identycznością lub odbiciem, że $f(u) = v$ oraz $f(v) = u$.

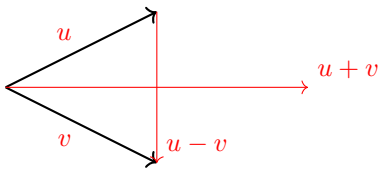
Pokażmy najpierw jak z lematu wynika teza. Mamy $\|\phi(\alpha)\| = \|\alpha\|$. Zgodnie z obserwacją wyżej istnieje odbicie f, które przeprowadza $\phi(\alpha)$ na α. Stąd przekształcenie $f \circ \phi : V \to V$ spełnia warunek

$$(f \circ \phi)(\alpha) = \alpha.$$

Na mocy Przypadku 1 mamy więc albo $f \circ \phi = \mathrm{id}_V$, albo $f \circ \phi = \psi_k \circ \ldots \circ \psi_1$, gdzie $k \leqslant n-1$ oraz $\psi_k : V \to V$ są odbiciami. Zatem (korzystając z $f^2 = \mathrm{id}_V$) mamy albo $\phi = f$, albo $\phi = f \circ \psi_k \circ \ldots \circ \psi_1$, co jest złożeniem nie więcej niż n odbić. Dowód jest zakończony.

□

Pokazujemy lemat.



Oczywiście jeśli $u = v$ teza jest oczywista, bo $u - u = 0$ jest prostopadły do każdego wektora i możemy wziąć $f = \mathrm{id}$. Załóżmy więc, ze $u \neq v$. Mamy: $\langle u + v, u - v \rangle = \|u\|^2 - \|v\|^2 = 0$. Skoro:

$$u = \underbrace{\frac{1}{2}(u+v)}_{\in \lin(u-v)^\perp} + \underbrace{\frac{1}{2}(u-v)}_{\in \lin(u-v)}.$$

to odbicie f_{u-v} względem $\lin(u-v)^\perp$ spełnia warunki zadania:

$$u = \frac{1}{2}(u-v) + \frac{1}{2}(u+v) \xrightarrow{f_{u-v}} \frac{1}{2}(u+v) - \frac{1}{2}(u-v) = \frac{1}{2}v + \frac{1}{2}v = v.$$

Twierdzenie jest udowodnione. Warto odnotować ważną obserwację dotyczącą izomerii i orientacji, wynikającą z twierdzenia Cartana.

Wniosek 16.1

Każdą izometrię ϕ zorientowanej przestrzeni euklidesowej liniowej można przedstawić w postaci złożenia parzystej albo nieparzystej liczby odbić. Jeśli $\det\phi = 1$, to izometria ϕ jest złożeniem parzystej liczby odbić. Gdy $\det\phi = -1$, to izometria ϕ jest złożeniem nieparzystej liczby odbić.

Powyższy wniosek motywuje następującą definicję.

Definicja 16.2: Izometria parzysta/nieparzysta

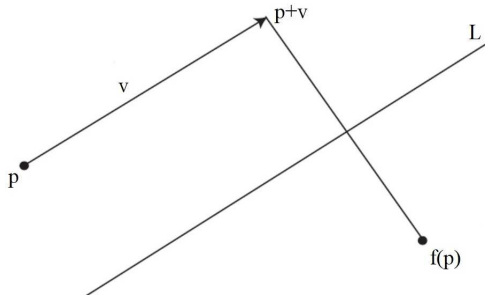
Izometrie przestrzeni euklidesowej liniowej o wyznaczniku równym 1 nazywamy IZOMETRIAMI PARZYSTYMI, a izometrie o wyznaczniku -1 nazywamy IZOMETRIAMI NIEPARZYSTYMI.

Oczywiście izometrie parzyste zachowują orientację i są złożeniem parzystej liczby odbić, a nieparzyte – nie zachowują orientacji, będąc złożeniem nieparzystej liczby odbić.

* * *

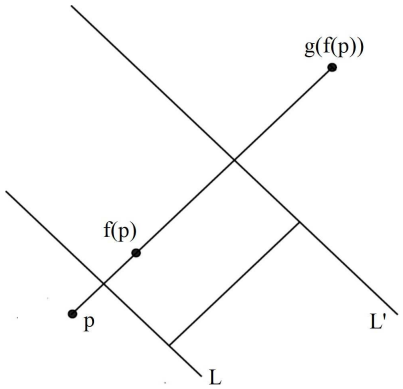
Na sam koniec powiedzmy o kilku przykładach złożeń przekształceń afinicznych. Szczególnie przydatne może być omówienie kilku przykładów, gdzie składamy znane nam obroty i symetrie z przesunięciami. Jest chyba jasne, że **każda izometria afiniczna $\mathbb{R}^n$ jest złożeniem przesunięcia i izometrii liniowej**. Oczywiście jest tak dlatego, że $\mathbb{R}^n$ zawiera wektor zerowy.

- SYMETRIA Z POŚLIZGIEM. Rozważmy dwuwymiarową przestrzeń afiniczną H i złożenie przekształcenia $f_v : H \to H$ będącego przesunięciem o wektor v oraz $s : H \to H$ będącego symetrią względem prostej L (nie musi być ona nawet prostopadła. Nietrudno widzieć, że $f = f_v \circ s = s \circ f_v$ działa jak na rysunku poniżej. Przekształcenie to nazywamy symetrią z poślizgiem.



- ZŁOŻENIE DWÓCH ODBIĆ AFINICZNYCH. Powiedzieliśmy, że złożenie odbić (liniowych) względem prostych przecinających się pod kątem α jest obrotem o kąt 2α. Jeśli proste te są podprzestrzeniami afinicznymi przecinającymi się w punkcie p, to złożenie odbić afinicznych względem nich będzie obrotem wokół punktu p. Zachęcam Czytelnika do formalnego dowodu tego faktu.

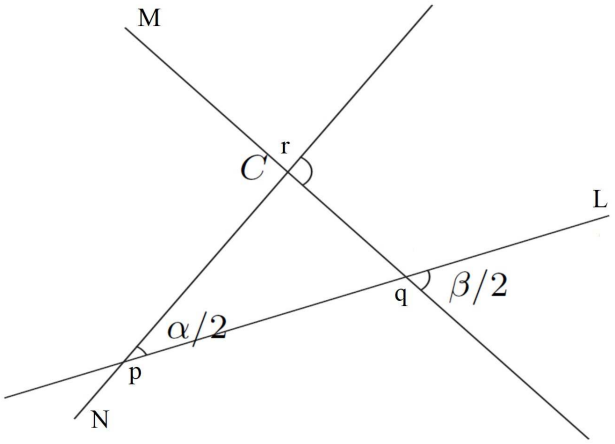
  A jeśli proste, względem których odbijamy się nie przecinają? Wówczas w wyniku złożenia otrzymujemy przesunięcie. O jaki wektor? Co się stanie, gdy zamienimy kolejność przestrzeni, względem których odbijamy? Zachęcam zarówno do geometrycznego, jak i algebraicznego dowodu.



- Złożenie obrotów. Na zorientowanej płaszczyźnie niech f będzie obrotem o kąt α wokół p oraz g niech będzie obrotem o kąt β wokół q. Wiemy z wcześniejszych rozważań, że obrot o kąt 2θ można dostać poprzez złożenie dwóch odbić względem prostych przecinających się pod kątem θ. W szczególności weźmy prostą L przechodzącą przez punkty p oraz q. Niech prosta N przechodzi przez p i tworzy z L kąt $-\alpha/2$ (zorientowany), zaś prosta M niech przechodzi przez q i tworzy z L kąt zorientowany $\beta/2$. Niech s_X będzie odbiciem względem dowolnej prostej X. Wówczas

$$f = s_N \circ s_L, \quad g = s_L \circ s_M, \quad \Longrightarrow f \circ g = s_N \circ s_L \circ s_L \circ s_M = s_N \circ s_M,$$

bowiem złożenie tej samej symetrii ze sobą to identyczność. Zachodzą dwie możliwości. Albo proste M oraz N są równoległe, co oznacza, że $\alpha + \beta = 180°$, skąd prosty wniosek, że złożenie obrotów $f \circ g$ jest przesunięciem. Jeśli natomiast $\alpha + \beta \neq 180°$, to niech $r = N \cap M$. Stąd nasze złożenie $f \circ g$ to złożenie odbić względem prostych M oraz N. Proste te przecinają się w punkcie r, więc złożenie jest obrotem o kąt dwa razy większy niż kąt pomiędzy prostymi M oraz N, czyli jest to obrór o kąt $\alpha + \beta$.



Fakt ten ma liczne zastosowania w geometrii elementarnej. Niektóre można znaleźć w następującym artykule prof. Piotra Grzeszczuka, w Delcie: `http://www.deltami.edu.pl/temat/matematyka/geometria/planimetria/2011/02/16/Obroty_w_zadaniach_geometrycznyc/`. a także innych źródłach, choćby w książeczce „*Wokół obrotów*" dra Waldemara Pompe czy też w klasycznym zbiorze z geometrii Wiktora Prasołowa. Zastosowania te mają urok głównie w języku elementarnym.

16.1 Uzupełnienie. Grupy izometrii i obrotów

Możliwość przedstawienia dowolnej izometrii przestrzeni euklidesowej liniowej jako złożenia pewnej liczby symetrii to jedynie wierzchołek góry lodowej olbrzymiej teorii związanej z tak zwanymi liniowymi i afinicznymi grupami odbić, badanymi już w XIX wieku, w latach 20. XX wieku zastosowanymi w teorii grup Liego przez Weyla i Cartana, później zaś usamodzielnionymi przez prace Coxetera. Tamte wyniki mają wciąż wielkie znaczenie. Elementarnie formułowalne geometryczne wyniki algebry liniowej, choćby klasyfikacja izometrii przestrzeni trójwymiarowej, nie są tylko ubieraniem matematyki „szkolnej" w język algebraiczny. Zaczniemy od definicji, którą zajmować się będziecie Państwo na Algebrze I.

Definicja 16.3: Grupa

Grupą nazywamy trójkę$(G, \circ, e)$, gdzie G jest zbiorem niepustym, $\circ : G \times G \to G$ jest działaniem 2-argumentowym, $e \in G$, i w której spełnione są aksjomaty:

- łączność $\circ$; dla każdych $a, b, c \in G$ mamy $(a \circ b) \circ c = a \circ (b \circ c)$,
- element e spełnia $a \circ e = a = e \circ a$, dla każdego $a \in G$,
- każdy element ma odwrotność, tzn. dla każdego $a \in G$ istnieje $b \in G$, że $a \circ b = b \circ a = e$.

Znamy już kilka grup, choć nie używaliśmy dotąd tego języka.

- Jeśli $(K, 0, 1, +, \cdot)$ jest ciałem, to trójki $(K, 0, +)$ oraz $(K \setminus \{0\}, 1, \cdot)$ są grupami, zwanymi odpowiednio: grupą addytywną oraz grupą multiplikatywną ciała K.
- Jeśli V jest przestrzenią liniową to zbiór izomorfizmów liniowych przestrzeni V w siebie, ozn. $Aut(V)$ jest grupą (ze względu na składanie przekształceń oraz id_V).
- Podzbiór $GL_n(K)$ złożony z macierzy odwracalnych rozmiaru n nad ciałem K z działaniem mnożenia macierzy i elementem neutralnym I_n jest grupą, zwaną pełną grupą liniową nad K.
- Zbiór S_n macierzy permutacyjnych rozmiaru n, więc w istocie: zbiór permutacji zbioru n-elementowego wraz z operacją składania permutacji, jest grupą.

Grupy mają olbrzymie znacznie w matematyce. Z punktu widzenia algebry liniowej również grają one poważną rolę, ale prowadzą do wielu skomplikowanych zagadnień, na które nie ma czasu w toku wykładu. Są im poświęcone nierzadko osobne przedmioty. Być może nie będzie zbyt dużym uproszczeniem, jeśli przyjmiemy, że bardzo ważną ideą teorii grup jest to, że można próbować wyróżnić pewien „reprezentatywny" podzbiór elementów grupy i próbować przedstawiać wszystkie inne za pomocą ciągów napisów złożonych z tych elementów oraz łączących je działań grupowych[1]. Dla przykładu:

- w grupie $(\mathbb{Z}, +, 0)$ każdy element jest sumą jedynek lub elementów przeciwnych do jedynki,
- w grupie liczb całkowitych podzielnych przez 3 (z działaniem dodawania) każdy element jest sumą pewnej liczby elementów 3 lub -3,
- w grupie S_n każdy element jest złożeniem permutacji zamieniających sąsiednie wyrazy,
- w grupie $GL_n(K)$ każdy element jest iloczynem macierzy operacji elementarnych.

Obserwacja 16.3

Niech $(V, \langle , \rangle)$ będzie przestrzenią euklidesową. Zbiór $O(V) \subseteq End(V)$ izometrii przestrzeni V na siebie jest grupą z działaniem składania przekształceń oraz elementem neutralnym równym id_V.

Dowód. Zacznijmy od zobaczenia, że złożenie dwóch izometrii H jest izometrią. Jeśli $f, g : H \to H$ są izometriami, to dla każdego $v \in V$ mamy $\|(g \circ f)(v)\| = \|g(f(v))\| = \|f(v)\| = \|v\|$, czyli $g \circ f$ zachowuje normę na H i jest izometrią. Przekształcenie id_V jest izometrią, bo nie zmienia normy. Jest to element neutralny działania składania izometrii w $O(V)$. Wreszcie, każda izometria f ma odwrotną. Rzeczywiście, f jest izomorfizmem, więc istnieje przekształcenie odwrotne f^{-1}. Skoro f zachowuje normę, to f^{-1} oczywiście też to robi. Jest więc izometrią. □

[1] Takie zbiory elementów nazywamy zbiorami generatorów. Grupa G może mieć różne zbiory generatorów, np. całe G.

Definicja 16.4

W zbiorze $Gl_n(K)$ wyróżniamy podzbiory:

$$O_n(K) = \{A \in Gl_n(K) \mid A^T A = I_n\},$$
$$SO_n(K) = \{A \in O_n(K) \mid \det(A) = 1\}.$$

Zbiory $O_n(K)$ oraz $SO_n(K)$ wraz działaniami mnożenia macierzy oraz elementem neutralnym I_n tworzą (odpowiednio) tzw. GRUPĘ ORTOGONALNĄ oraz SPECJALNĄ GRUPĘ ORTOGONALNĄ.

Grupy $SO_2(\mathbb{R})$ oraz $SO_3(\mathbb{R})$ złożone są z macierzy ortogonalnych o wyznaczniku 1. Muszą to być macierze obrotów, a więc w istocie te grupy to grupy obrotów. Grupa $SO_4(\mathbb{R})$ nie jest już złożona z samych obrotów, ale je zawiera. Dlaczego? Naszym celem będzie zrozumienie, że grupa $O_n(\mathbb{R})$ jest „w istocie" (kiedyś poznamy termin: izomorfizmu grup) grupą izometrii przestrzeni $\mathbb{R}^n$, niezależnie od przyjętego na niej iloczynu skalarnego.

Dowiedliśmy, że każda symetria przestrzeni euklidesowej jest złożeniem odbić. W szczególności oznacza to, że każda macierz ortogonalna jest iloczynem macierzy odbić. Co to za macierze, nie ma w tym momencie aż tak wielkiego znaczenia (ale będzie miało znaczenie np., gdy poznacie Państwo odbicie jako transformację Householdera na matematyce obliczeniowej), ale nawet sama świadomość, że każda macierz z $O(V)$ jest iloczynem macierzy, które w kwadracie są identycznościami daje do myślenia. Interesujące pytanie jest następujące: bierzemy grupę $O_n(\mathbb{R})$ i rozważamy następujący problem, postawiony w 1934 przez wielkiego geometrę Coxetera[2].

> Rozważmy skończoną liczbę ODBIĆ $s_1, \ldots, s_r$ przestrzeni euklidesowej liniowej $(\mathbb{R}^n, \langle , \rangle)$ i zacznijmy je ze sobą składać, a dalej składać złożenia itd. Innymi słowy rozważmy GRUPĘ GENEROWANĄ przez symetrie $s_1, \ldots, s_r$, ozn. $W(s_1, \ldots, s_r)$, nazywaną GRUPĄ SYMETRII. Uwaga: grupa symetrii nie składa się z samych symetrii, ale ze złożeń symetrii. $W(s_1, \ldots, s_r)$ to „najmniejsza" grupa zawierająca wszystkie symetrie $s_1, \ldots, s_r$. Kiedy ta grupa jest skończona?

Pytamy tu w istocie o tzw. skończone podgrupy $O_n(\mathbb{R})$. To, że takie podgrupy istnieją jest oczywiste, nawet jeśli ograniczymy się do badania izometrii płaszczyzny. Okazuje się bowiem, że skończone podgrupy $O_2(\mathbb{R})$ to grupy symetrii wielokątów foremnych. Kilka przykładów.

- Niech $\theta = 2\pi/m$ oraz niech:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

  Wówczas elementy grupy H_m (tzw. grupa dihedralna) generowanej przez symetrie T oraz $R_\theta \cdot T$ to izometrie zachowujące m-kąt foremny na płaszczyźnie (o środku w środku układu współrzędnych).

- Niech $\Delta^3 = S(e_1, e_2, e_3, e_4)$ będzie sympleksem trójwymiarowym w przestrzeni $\mathbb{R}^4$, rozpiętym przez punkty $e_1 = (1,0,0,0)$, $e_2 = (0,1,0,0)$, $e_3 = (0,0,1,0)$, $e_4 = (0,0,0,1)$. Każdej parze wierzchołków odpowiada symetria s_{ij}, która przeprowadza $i-$ty wierzchołek na j-ty, a pozostałych nie rusza. Każda izometria zachowująca czworościan Δ^3 jest złożeniem elementarnych symetrii. **Grupa symetrii** czworościanu, to po prostu okazuje się być grupa permutacji S_4 zbioru 4-elementowego.

- **Twierdzenie**: każda skończona grupa **generowana przez obroty**[3] w przestrzeni euklidesowej $\mathbb{R}^3$ (w języku macierzy: skończona podgrupa w $SO(3)$) *jest* albo grupą obrotów lub **grupą symetrii** wielokąta foremnego, albo jedną z trzech **grup symetrii** zachowujących wielościany platońskie (!).

Oczywiście Coxeter nie postawił tylko pytań, ale w serii artykułów w latach 1934-1935 dał początek powstaniu wielkiej geometrycznej teorii, dziś obejmującej swoim zakresem algebrę, geometrię algebraiczną, teorię reprezentacji i wiele innych dziedzin. Rozwiązanie problemu Coxetera wymagało wprowadzenia nowych obiektów i zrozumienia na nowo wielu obiektów uważanych za znane. Była to piękna realizacja XIX-wiecznego programu Felixa Kleina.

[2] Harold Scott MacDonald Coxeter (1907-2003), „Król geometrii" — jeden z najbardziej znanych geometrów XX wieku.

[3] Prosty tekst, gdzie (z obrazkami) opowiedziane jest jakie są te grupy odbić: *Symmetry Groups of Platonic Solids*, David Newcomb: `https://dnewcomb.com/pdfs/Platonic%20Solids.pdf`

16.2 Dodatek. Punkty stałe i klasyfikowanie izometrii

Badanie wielu klas przekształceń związane jest z poszukiwaniem punktu stałego. Mówi się nawet o teorii punktu stałego. Jest to jeden z bardzo często powracających motywów w matematyce.

Definicja 16.5: Punkt stały izometrii

Niech $(H, \langle , \rangle)$ będzie przestrzenią afiniczną oraz $f : H \to H$. Powiemy, że $x \in H$ jest PUNKTEM STAŁYM funkcji f jeśli $f(x) = x$.

Obserwacja 16.4

Niech $(H, \langle , \rangle)$ będzie przestrzenią euklidesową afiniczną. Dla dowolnego przekształcenia afinicznego $f : H \to H$ następujące warunki są równoważne:

(1) istnieje baza punktowa H złożona z punktów stałych H,

(2) f jest przekształceniem identycznościowym.

Polecam każdemu próbę policzenia permutacji zbioru n-elementowego, które nie ruszają dokładnie jednego elementu[4]. Szereg twierdzeń o punkcie stałym pochodzących z rozmaitych działów matematyki nosi nazwiska znamienitych uczonych: Poincarego, Borela, Atiyaha, Lefschetza, Kakutaniego, Brouwera, Kleene, choć wydaje się, że jednym najważniejszych rezultatów tego typu jest twierdzenie Stefana Banacha. Wielu zresztą polskich matematyków, zwłaszcza związanych z topologią lub analizą funkcjonalną, dowodziło znakomite rezultaty dotyczące punktów stałych, jak choćby Schauder, Ryll-Nardzewski, Knaster, Kuratowski, Mazurkiewicz, Tarski, Borsuk, Sieklucki, a także prof. A. Białynicki-Birula.

Klasyfikacje różnych typów przekształceń prowadzi się niekiedy w oparciu o analizę punktów stałych[5]. Nie starczy nam na tym wykładzie miejsca na przeprowadzenie klasyfikacji izometrii przestrzeni afinicznych niskich wymiarów, a opiera się ona właśnie o rozgraniczenie izometrii, które nie mają punktów stałych (np. przesunięcia), które mają dokładnie jeden punkt stały (np. obrót na płaszczyźnie) lub wyżej wymiarową podprzestrzeń afiniczną punktów stałych (np. symetria).

Ograniczę się do kilku uwag wraz z podaniem źródeł, gdzie można doczytać szczegóły i niektóre dowody. Rozważania te można prowadzić zarówno algebraicznie (w języku wektorów i wartości własnych), jak i czysto geometrycznie (zwłaszcza w $\mathbb{R}^2$).

Twierdzenie 16.2

Każda izometria skończenie wymiarowej przestrzeni afinicznej $(H, \langle , \rangle)$ może być przedstawiona jednoznacznie w postaci złożenia $t \circ k$, gdzie t jest przesunięciem, zaś k jest izometrią mającą punkt stały. Innymi słowy każda izometria f przestrzeni afinicznej ma punkt stały albo prostą niezmienniczą, na której f jest przesunięciem.[a]

[a] Wykład dr. Strojnowskiego, tw. 19.4: `https://www.mimuw.edu.pl/~stroa/GAL2wyk14.pdf`.

W szczególności jeśli $H = \mathbb{R}^n$, to powyższa uwaga mówi, że każda izometria jest złożeniem przesunięcia i izometrii zachowującej wektor zerowy. Możemy zatem alternatywnie do tw. Cartana odwołać się do Twierdzenia 17.5. Przedstawię teraz rezultaty klasyfikujące[6] wszystkie izometrie przestrzeni afinicznych $\mathbb{R}$, $\mathbb{R}^2$ oraz $\mathbb{R}^3$ (nie są to fakty obowiązkowe, ale na ćwiczeniach będą przykłady).

[4] Trivia: liczba permutacji zbioru n elementowego bez punktu stałego „ma taką samą asymptotykę" jak funkcja $n!/e$.

[5] Kto by chciał dowiedzieć się więcej o punktach stałych, izometriach i ich zastosowaniach w ładnych zadaniach, polecam tekst dra Krzysztofa Żyjewskiego (UAM) *O składaniu symetrii i obrotów. Punkty stałe izometrii*, pod adresem: `http://wmii.uwm.edu.pl/f/images/PTM/SZM-2016/o_skladaniu_symetrii_i_obrotow._punkty_stale_izometrii_-_krzysztof_zyjewski.pdf`. Tekst zawiera również ładne zastosowania elementarne.

[6] Opis ten odwołuje się do stron 42-43 w `https://www.mimuw.edu.pl/~stroa/GAL2wyk14.pdf`.

Obserwacja 16.5

Każda izometria prostej rzeczywistej $\mathbb{R}$ jest postaci $f(x) = x + c$ lub $f(x) = -x + c$, dla pewnego $c \in \mathbb{R}$. Są to zatem: przesunięcie lub symetria środkowa.

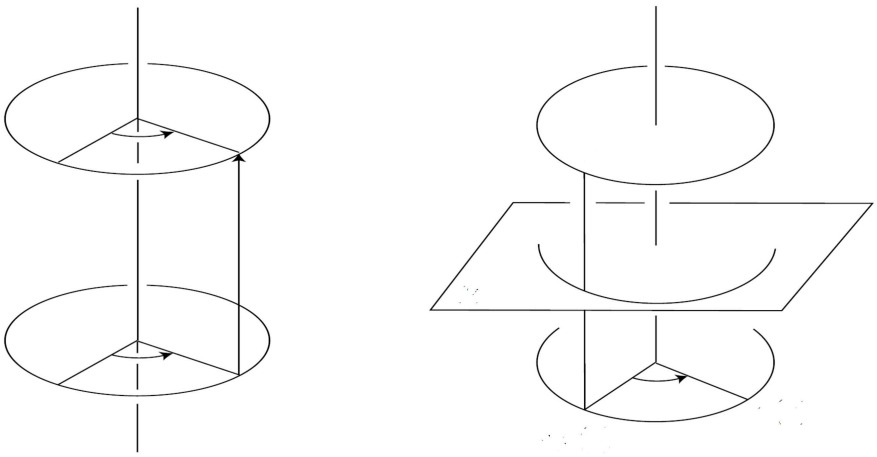
Obserwacja 16.6

Każda izometria płaszczyzny afinicznej $\mathbb{R}^2$ jest jednym z czterech przekształceń: przesunięciem, obrotem, symetrią lub tzw. symetrią z poślizgiem, czyli złożeniem niezerowego przesunięcia o wektor α oraz symetrii względem przestrzeni $\mathrm{lin}(\alpha)$.

Dowód algebraiczny tego faktu wynika łatwo z twierdzenia klasyfikującego izometrie liniowe płaszczyzny jako obroty lub symetrie prostopadłe. Na mocy wcześniejszej uwagi każda izometria afiniczna jest złożeniem przesunięcia i jednego z tych przekształceń. Identyczność ma trzy afinicznie niezależne punkty stałe, symetria osiowa – dwa. Obrót ma jeden punkt stały, a przesunięcie i symetria z poślizgiem ich nie mają.

Twierdzenie 16.3

Każda izometria trójwymiarowej przestrzeni afinicznej jest identycznością, obrotem, symetrią płaszczyznową, obrotem z odbiciem (te izometrie mają punkty stałe) lub przesunięciem, obrotem z poślizgiem (ruch śrubowy) bądź symetrią płaszczyznową z poślizgiem (brak punktów stałych).



Obrót z poślizgiem w kierunku prostopadłym (ruch śrubowy), symetria płaszczyznowa z poślizgiem. Źródło: M. Audin. *Geometry.*

Fakt ten wywodzi się z twierdzenia Cartana o izometriach. Kto by chciał zobaczyć elementarny dowód, zachęcam do lektury pięknego tekstu prof. Kordosa: Rzut butem, czyli twierdzenie Chaslesa: `http://www.deltami.edu.pl/temat/matematyka/geometria/stereometria/2015/10/27/1511delta-chasles.pdf`. Opowiada on o wyjaśnieniu następującego fenomenu.

> Zdjąłem z nogi but i cisnąłem nim byle jak, po czym on upadł byle gdzie i jakoś tam leży. Istnieje ruch po linii śrubowej, za pomocą którego można kulturalnie przenieść ten but z obecnego położenia na moją nogę.

Sens tego twierdzenia jest następujący: obrót z poślizgiem jest jedyną izometrią przestrzeni trójwymiarowej, której pochodna ma dodatni wyznacznik. Takie izometrie nazywa się RUCHAMI SZTYWNYMI. Istotnie, rzut butem nie zmienia przecież jego orientacji — trajektoria lotu w jedną stronę może nie być przekształceniem afinicznym, ale można ten but przywrócić na właściwe miejsce za pomocą ruchu po linii śrubowej (oczywiście może się zdarzyć, że obrót lub przesunięcie są trywialne).