# Algebra liniowa z elementami analizy danych

Jacek Tabor

10 czerwca 2016

# Spis treści

# Rozdział 0

# Wstęp

## 0.1 Zasady zaliczenia przedmiotu

Egzamin pisemny będzie zawierał zadania zarówno z ćwiczeń, jak i z wykładu.

Ocena końcowa z przedmiotu Algebra Liniowa, będzie średnią z (średniej z ćwiczeń) i (oceny z egzaminu).

## 0.2 Historia

MUHAMMED IBN MUSA ALCHWARIZMI: „Hisab al-djabr wal-mukabala” (około 830 n.e.)

Algebra: „al-djabr” (początkowo termin lekarski pochodzący od nastawiania kości): odtwarzanie, polegała na likwidowaniu w równaniu wyrazów ujemnych przez dodawanie do obu stron równania wyrazów przeciwnych do danych ujemnych. Od zniekształconej formy nazwiska Alchwarizmi powstało słowo algorytm.

- powstanie matematyki: Grecja, aksjomat, pojęcie dowodu
- „Oświecenie Islamu” ok. IX-XII wiek, powstanie algebry, Alchwarizmi (rozwiązywanie równań kwadratowych)
- XVI wiek: włochy, „rozwiązywanie” wielomianów (Niccolò Fontana Tartaglia, Lodovico Ferrari, Gerolamo Cardano, Vieta)
- wyznaczniki: Leibniz 1693
- reguła Cramera rozwiazywania równań liniowych: 1750
- następnie Gauss: metoda eliminacji Gaussa (zapis, znanan była wcześniej) i metoda najmniejszych kwadratów (bardzo ważne w zastosowaniach!)

## 0.3 Zastosowania

Spośród działów matematyki, chyba najwieksze są zastosowania algebry liniowej:

- przetwarzanie obrazu: filtry (Gaussowski)
- kompresja obrazu: jpg
- przetwarzanie dźwięku: transformata Fouriera
- kompresja dźwięku: mp3

- geometria: czcionki i krzywe (funkcje gięte), gry (obroty)
- interpolacja i ekstrapolacja
- algorytm google pagerank
- sztuczna inteligencja i nauczanie maszynowe (SVM, ...)
- analiza danych

KONIEC WYKŁADU 0

# Rozdział 1

# Równania różnicowe i liczby zespolone

**Streszczenie.** W rozdziale tym zajmiemy się badaniem liniowych równań różnicowych (stopnia drugiego). Pokażemy, że ich badanie prowadzi w sposób naturalny do liczb zespolonych. Drugim wygodnym narzędziem okazuje się pojęcie granicy i pochodnej.

## 1.1 Ciąg Fibonacciego

Chciałbym Państwu przedstawić czym zajmuje się algebra liniowa na podstawie ciągu Fibonacciego. Ciąg Fibonacciego [1170-1250] (pojawił sie już w matematyce hinduskiej u Pringala 200 p.n.e.):

*Jak się rozmnażają króliki: w pewnym miesiącu mamy pewną ilość par młodych i starych królików. W następnym miesiącu każda para królików rodzi nową, natomiast młoda para staje się starą, a każda stara umiera.*

*W miesiącu 1 dostaliśmy jedną starą parę królików. Chcemy prześledzić ile będziemy mieli królików w kolejnych miesiącach.*

| miesiąc | młode | stare | w sumie |
|---|---|---|---|
| 1 | 0 | 1 | 1 |
| 2 | 1 | 0 | 1 |
| 3 | 1 | 1 | 2 |
| 4 | 2 | 1 | 3 |
| 5 | 3 | 2 | 5 |
| 6 | 5 | 3 | 8 |

Zacznijmy od ogólnego wzoru:

$$\begin{cases} stare_{n+1} = mlode_n, \\ mlode_{n+1} = stare_n + mlode_n. \end{cases} \tag{1.1}$$

Nas interesuje ilość królików $x_n = mlode_n + stare_n$. Oczywiście $x_1 = 1, x_2 = 1$. Wyrugujmy na razie zmienną $stare_n = x_n - mlode_n$ z równania (1.1):

$$\begin{cases} x_{n+1} - mlode_{n+1} = mlode_n, \\ mlode_{n+1} = x_n. \end{cases}$$

Czyli

$$x_{n+2} = mlode_{n+2} + mlode_{n+1} = x_{n+1} + x_n.$$

To równanie nazywa sie obecnie *równaniem Fibonacciego* (do tego dochodzą warunki początkowe $x_1 = x_2 = 1$).

Metody rozwiązywania: najprostsza – zgadnij wzór i sprawdź. Tu nie za bardzo widać jak to można by było łatwo zrobić.

Typowe: próba analizy danego zjawiska [asymptotyka: zachowanie dla dużych czasów]: zobaczmy, że ilorazy kolejnych elementów są „zbieżne" ($1/1 = 1.0$, $2/1 = 2.0$, $3/2 = 1.5$, $5/3 = 1.66(6)$, $8/5 = 1.6$, $13/8 = 1.625$, $\dots$), wygląda jakby zmierzało $\approx 1.618$ (przypomina się złoty podział!!!). To sugeruje, że rozsądnym jest szukanie rozwiązań w klasie ciągów geometrycznych (przypominam, ciąg geometryczny ma stałe ilorazy kolejnych wyrazów). Zapomnijmy więc na razie o warunku początkowym i poszukajmy rozwiązań postaci $az^n$. Podstawiamy $x_n = az^n$, i dostajemy równanie

$$az^2 = az + a,$$

czyli $a$ może być dowolne, a $z$ musi spełnić równanie

$$z^2 - z - 1 = 0.$$

W konsekwencji

$$z_1 = \frac{1+\sqrt{5}}{2},\ z_2 = \frac{1-\sqrt{5}}{2}.$$

Mamy więc dwa rozwiązania $a_1 z_1^n$, $a_2 z_2^n$. Niestety widać, że żadne z nich nie spełnia warunków początkowych $x_0 = 0, x_1 = 1$ (bo wtedy byłoby $a_1 = a_1 z_1^0 = 0$ i analogicznie $a_2$ byłoby równe zero). Ratuje nas następująca obserwacja:

**Obserwacja 1.1.** *Równanie Fibonacciego jest* równaniem liniowym*, to znaczy, jeżeli* $(v_n)$ *i* $(w_n)$ *spełniają to równanie, to ich* kombinacja liniowa $(av_n + bw_n)$ *też spełnia dla dowolnych* $a, b \in \mathbb{R}$*.*

*Dowód.* Niech $x_n = av_n + bw_n$. Widzimy, że

$$x_{n+2} = av_{n+2} + bw_{n+2} = a(v_{n+1} + v_n) + b(w_{n+1} + w_n)$$

$$= (av_{n+1} + bw_{n+1}) + (av_n + bw_y) = x_{n+1} + x_n.$$

□

Skoro tak, to znaczy, że możemy szukać rozwiązania równania Fibonacciego w postaci

$$x_n = a_1 z^n + a_2 z_2^n.$$

Zobaczymy, czy uda się wyliczyć $a_1$ i $a_2$. Podstawiamy do warunków początkowych:

$$x_1 = a_1 z_1 + a_2 z_2 = 1,$$
$$x_2 = a_1 z_1^2 + a_2 z_2^2 = 1,$$

i po prostych przekształceniach (sprawdzić) dostajemy, że $a_1 = 1/\sqrt{5}$ i całe rozwiązanie dane jest wzorem

$$x_n = \frac{1}{\sqrt{5}}(\frac{1+\sqrt{5}}{2})^n - \frac{1}{\sqrt{5}}(\frac{1-\sqrt{5}}{2})^n.$$

**Zadanie 1.1.** *Proszę sprawdzić bezpośrednio, że to jest rozwiązanie.*

**Zadanie 1.2.** *Proszę rozwiązać równania:*

- $x_{n+1} - 3x_n = 1$, $x_0 = 0$*;*
- $x_{n+2} - 5x_{n+1} + 6x_n = 0$, $x_0 = 0$, $x_1 = 1$*;*
- $x_{n+1} - 2x_n = 2^n$, $x_0 = 0$*;*
- $x_{n+2} - 2x_{n+1} + x_n = 0$, $x_0 = 0$, $x_1 = 1$ *(wsk. proszę zrobić podstawienie* $v_n = x_{n+1} - x_n$*, i najpierw wyliczyć wzór na* $v_n$*);*

## 1.2 Liczby zespolone: $\Delta < 0$

**Problem 1.1.** Spróbujmy rozwiązać równanie:

$$x_{n+2} - 2x_{n+1} + 2x_n = 0,\ x_0 = 0,\ x_1 = 1.$$

Dostajemy równanie

$$z^2 - 2z + 2 = 0.$$

Chcemy rozwiązać – ale problem, bo $\Delta = -4 < 0$! Nie przejmujemy się tym, przyjmujemy oznaczenie $i = \sqrt{-1}$, i rozwiązujemy jakby nigdy nic:

$$z_1 = \tfrac{-b+\sqrt{\Delta}}{2a} = \tfrac{2+2i}{2} = 1 + i,$$
$$z_2 = \tfrac{-b-\sqrt{\Delta}}{2a} = \tfrac{2-2i}{2} = 1 - i.$$

Teraz dobieramy $a_1, a_2$ tak by spełnione były warunki początkowe:

$$a_1 z_1^0 + a_2 z_2^0 = 0,\ a_1 z_1^1 + a_2 z_2^1 = 1,$$

czyli $a_2 = -a_1$ i $a_1(1+i) - a_1(1-i) = 1$, co oznacza, że

$$a_1 = \frac{1}{2i} \text{ oraz } a_2 = -\frac{1}{2i}.$$

W konsekwencji ostateczne rozwiązanie jest dane wzorem

$$x_n = a_1 z_1^n + a_2 z_2^n = \frac{1}{2i}(1+i)^n - \frac{1}{2i}(1-i)^n.$$

Niby wszystko gra, ale musimy a) nauczyć się operować na tych liczbach (no bo co to jest tak naprawdę $\frac{1}{2i}$), b) liczyć potęgi (potrzebujemy jawny wzór na $(1+i)^n$). I to będzie właśnie celem niniejszej sekcji.

**Historia 1.1.** *liczby zespolone - wzory Cardano - niezbędne do rozwiązywania wielomianów od stopnia 3*

Zacznijmy dokładniejszy opis od liczb zespolonych. Nieformalnie możemy napisać, że $i = \sqrt{-1}$. Inaczej mówiąc, $i$ to taka liczba, że $i^2 = -1$.

Najprostsze działania na liczbach zespolonych:

- dodawanie $(a + bi) + (c + di) = (a + c) + (b + d)i$
- mnożenie $(a + bi) \cdot (c + di) = ac + adi + bci + bdi^2 = (ac - bd) + (ad + bc)i$
- dzielenie – analogicznie jak pozbywanie się niewymierności w mianowniku (bo w końcu $i$ to jakby pierwiastek z $-1$), dla przykładu:

$$\frac{1+2i}{1-i} = \frac{1+2i}{1-i} \cdot \frac{1+i}{1+i} = \frac{(1+2i)(1+i)}{1-i^2} = \frac{1+2i+i+2i^2}{2} = \frac{-1+3i}{2} = \frac{-1}{2} + \frac{3}{2}i.$$

## 1.3 Układ kartezjański i biegunowy na płaszczyźnie

Aby policzyć potęgę liczby zespolonej potrzebna nam będzie jej interpretacja geometryczna. W szczególności przypomnimy sobie podstawowe informacje na temat płaszczyzny.

UKŁAD KARTEZJAŃSKI: Kartezjusz 1637, jedna oś (czyli prosta). Dwie osie „La Géométrie” (dodane w tłumaczeniu na łacinę 1649 Frans van Schooten).

UKŁAD BIEGUNOWY. połowa XVII wieku: Grégoire de Saint-Vincent and Bonaventura Cavalieri. Newton, (napisane 1671, opublikowane 1736), Isaac Newton używał chyba 10 układów krzywoliniowych, między innymi biegunowego.

rysunek: czołg (dwie wersje w obie strony)

KONWERSJA MIĘDZY UKŁADAMI

Zespolone używają (potrzebują) obu podejść.

- $\mathrm{re}(a+bi) = a$ (część rzeczywista), $\mathrm{im}(a+bi) = b$ (część urojona)
- liczbę $(a+bi)$ zaznaczamy na płaszczyźnie jako parę $(a,b)$
- moduł liczby zespolonej $a+bi$ to jej odległość od zera, czyli z Tw. Pitagorasa $|a+bi| = \sqrt{a^2+b^2}$
- Liczba sprzężona: jeśli $z = a+bi$, to $\bar{z} = a - bi$ (symetria względem osi rzeczywistej)

**Zadanie 1.3.** *Proszę sprawdzić, że* $|z_1 \cdot z_2| = |z_1| \cdot z_2|$, $|z_1 + z_2| \leqslant |z_1 + z_2|$.

Postać trygonometryczna:
$$z = |z| \cdot (\cos\phi + i\sin\phi),$$
gdzie $|z|$ to moduł liczby $z$, zaś $\phi$ to argument $z$ (piszemy $\phi = \arg z$), czyli dowolny kąt taki, że $a/|z| = \cos\phi$, $b/|z| = \sin\phi$ (często rozważamy także argument główny $\mathrm{Arg} z$ liczby zespolonej, czyli jedyny argument $\in [-\pi, \pi)$, niektórzy rozważają $[0, 2\pi)$).

Koniec Wykładu 1

Wzory de'Moivre'a:

$$r_1(\cos\phi_1 + i\sin\phi_1) \cdot r_2(\cos\phi_2 + \sin\phi_2) = r_1 r_2(\cos(\phi_1+\phi_2) + i\sin(\phi_1+\phi_2)),$$

$$\frac{r_1(\cos\phi_1 + i\sin\phi_1)}{r_2(\cos\phi_2 + \sin\phi_2)} = \frac{r_1}{r_2}(\cos(\phi_1-\phi_2) + i\sin(\phi_1-\phi_2)).$$

**Zadanie 1.4.** *Proszę sprawdzić powyższe wzory.*

Jako konsekwencję używając powyższego wzoru $n$ razy, dostajemy wzór na potęgowanie liczby zespolonej w postaci trygonometrycznej:

$$z^n = r^n(\cos n\varphi + i\sin n\varphi) \text{ gdzie } r = |z|, \varphi = \arg z.$$

Wracamy do rozwiązania równania różnicowego od którego zaczęliśmy: wystarczy teraz wyliczyć wzór na

$$x_n = \frac{1}{2i}(1+i)^n - \frac{1}{2i}(1-i)^n.$$

Jak łatwo sprawdzić (było na wykładzie), po prostych przeliczeniach dostajemy:

$$x_n = \sqrt{2}^n \sin(n\pi/4).$$

**Dygresja 1.1.** *Grafika/gry – wzory de Moivre'a stosuje się do obrotów na płaszczyźnie (bywa przydatne w grach). Jeżeli mamy punkt* $z = (a,b)$ *na płaszczyźnie, i chcemy go obrócić o kąt* $\alpha$ *to wystarczy przejść do postaci zespolonej* $a + ib$*, i przemnożyć przez* $(\cos\alpha + i\sin\alpha)$*, co da nam łatwo gotowy wynik. Podobny trik używa się także w przestrzeni – ale zamiast liczb zespolonych potrzebne są tak zwane kwaterniony (wymyślił Hamilton, pojawiają się* $i, j, k$ *które spełniają równania* $i^2 = j^2 = k^2 = -1$, $ijk = -1$*, są zaimplementowane na przykład w Direct X). Nie będę teraz o nich więcej mówił, ale jeżeli ktoś jest zainteresowany, odsyłam do wikipedii.*

Liczby zespolone są ważne między innymi z powodu twierdzenia ZASADNICZE TWIERDZENIE ALGEBRY które mówi, że każdy wielomian ma przynajmniej jeden pierwiastek zespolony. W konsekwencji każdy wielomian $W$ stopnia $n$ możemy przedstawić w postaci iloczynu $a(x - x_1)\ldots(x - x_n)$, gdzie $x_i \in \mathbb{C}$.

## 1.4 A co z $\Delta = 0$?

Przykład $x_{n+2} - 4x_{n+1} + 4x_n = 0$ z warunkami $x_0 = 0, x_1 = 1$.

jedno rozwiązanie równania charakterystycznego $z^2 - 4z + 4 = 0$ dane wzorem $z_1 = 2$, czyli mamy tylko rozwiązanie postaci $x_n = a2^n$. Nie jesteśmy w stanie tak dobrać $a$, by spełnione były warunki początkowe.

Perturbacja: lekko zaburzamy równanie, aby mieć dwa pierwiastki, a następnie z zaburzeniem zmierzamy do zera.

dwa: $(2+\varepsilon)^k$ i $(2-\varepsilon)^k$, jakie równanie – bardzo bliskie a mianowicie $z^2 - 4z + (4-\varepsilon^2)$, czyli rozpatrujemy równanie

$$x_{n+2} - 4x_{n+1} + (4-\varepsilon^2)x_n = 0, \tag{1.2}$$

z warunkiem początkowym $x_0 = 0, x_1 = 1$. Rozwiązanie ogólne (1.2) dane jest

$$x_n = a_1(2+\varepsilon)^n + a_2(2-\varepsilon)^n.$$

Uwzględniamy warunki początkowe

$$x_0 = a_1 + a_2 = 0,\ x_1 = a_1(2+\varepsilon) + a_2(2-\varepsilon) = 1,$$

dostając $a_2 = -a_1$ oraz $a_1 = 1/(2\varepsilon)$. Konkludując dostajemy

$$x_n = \frac{(2+\varepsilon)^n - (2-\varepsilon)^n}{2\varepsilon}.$$

Przechodzimy do granicy z $\varepsilon \to 0$! Jak łatwo zauważyć to jest dokładnie pochodna funkcji $f(x) = x^n$ w punkcie $x = 2$, czyli $f'(x) = nx^{n-1}$, czyli $f'(2) = n2^{n-1}$. W konsekwencji rozwiązanie naszego równania dane jest wzorem

$$x_n = n2^{n-1}.$$

Przypominam *pochodna* $f'$ funkcji $f$ to granica przy $\varepsilon \to 0$ wyrażenia

$$f'(x) = \frac{f(x+\varepsilon) - f(x-\varepsilon)}{2\varepsilon} \text{ lub } \frac{f(x+\varepsilon) - f(x)}{\varepsilon}.$$

Interpretacja geometryczna – współczynnik kierunkowy stycznej do krzywej.

Jeżeli ktoś nie miał pochodnej, to przyda się nam *wzór Newtona* (dwumian Newtona):

$$(a+b)^n = a^n + \tbinom{n}{1}a^{n-1}b + \ldots + \tbinom{n}{n-1}ab^{n-1} + b^n.$$

Wygodnie współczynniki wyliczyć z Trójkąta Pascala:

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ 1 & & 3 & & 3 & & 1 \end{array}$$

(na bokach trójkata kładziemy jedynki, a każdy współczynnik jest sumą dwóch nad nim).

DLA ĆWICZEŃ: Wzory de'Moivre'a są ważne, gdyż przy pomocy dwumianu Newtona pozwalają łatwo wyliczać wzory na $\cos(n\alpha)$ i $\sin(n\alpha)$ za pomocą potęg $\cos\alpha$ i $\sin\alpha$.

**Przykład 1.1.** *Mamy:*

$$\cos 3\alpha + i\sin 3\alpha = (\cos\alpha + i\sin\alpha)^3 = \cos^3\alpha + 3\cos^2\alpha(i\sin\alpha) + 3\cos\alpha(i\sin\alpha)^2 + (i\sin\alpha)^3$$

$$= \cos^3\alpha - 3\cos\alpha\sin^2\alpha + i(3\cos^2\alpha\sin\alpha - \sin^3\alpha),$$

*czyli* $\cos 3\alpha = \cos^3\alpha - 3\cos\alpha\sin^2\alpha,\ \sin 3\alpha = 3\cos^2\alpha\sin\alpha - \sin^3\alpha.$

**Zadanie 1.5.** *Proszę wyliczyć* $\sin 5\alpha$ *za pomocą potęg* $\cos\alpha$ *i* $\sin\alpha$*.*

# Rozdział 2

# Geometria danych

**Streszczenie.** Do pracy z danymi przyda nam się pojęcie przestrzeni $\mathbb{R}^D$ (uogólnienie płaszczyzny i przestrzeni) wraz z iloczynem skalarnym, który pozwala nam w szczególności mierzyć odległość między punktami. Zobaczymy także pojęcie macierzy.

## 2.1 Płaszczyzna, czyli $\mathbb{R}^2$

Stosujemy układ Kartezjański (pojawił się już wcześniej). Zwyczajowo punkty na płaszczyźnie zapisuje się za pomocą dwóch współrzędnych - konwencja $(x_1, x_2)$.

**Uwaga 2.1.** *Zapis $(x, y)$ na punty na płaszczyźnie był częściej stosowany w szkole, ja raczej będę używał zapisu $(x_1, x_2)$.*

Czasami stosuje się także zapis w nawiasach kwadratowych (macierzowy), w postaci

$$\text{wierszowej: } \begin{bmatrix} x_1 & x_2 \end{bmatrix} \text{ lub kolumnowej: } \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Tu już po raz pierwszy pojawią się pewne elementy geometrii.
Operacje w $\mathbb{R}^2$:

- $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$ (analogicznie dla odejmowania)
- $\alpha \cdot (x_1, x_2) = (\alpha x_1, \alpha x_2)$

Odległość od zera z Tw. Pitagorasa (czasami mówi się norma / metryka euklidesowa):

$$\|x\| = \|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}.$$

Wtedy oczywiście odległość $v$ od $w$ jest dana wzorem $\|v - w\|$.

**Zadanie 2.1.** *Udowodnij tw. cosinusów: w trójkącie o bokach długości $a, b, c$ mamy*

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma), \tag{2.1}$$

*gdzie $\gamma$ kąt przeciwległy do $c$. Zastanów się, jak wyglądałoby tw. odwrotne, i czy jest prawdziwe.*

Iloczyn skalarny wektorów $x, y \in \mathbb{R}^2$ definiuje się następująco:

$$(x_1, x_2) \circ (y_1, y_2) = x_1 y_1 + x_2 y_2.$$

Inne notacje: czasami zamiast $x \circ y$ pisze się $\langle x, y\rangle$. Oczywiście mamy

$$\|v\| = \sqrt{v \circ v}.$$

**Obserwacja 2.1.** *Udowodnij następujące własności iloczynu skalarnego:*

1. $u \circ (v+w) = u \circ v + u \circ w$;
2. $u \circ v = v \circ u$, $(\alpha v) \circ w = \alpha(v \circ w)$ *dla* $\alpha \in \mathbb{R}$.

**Zadanie 2.2.** *Korzystając jedynie z powyższej obserwacji, proszę udowodnić, że*

$$(v-w) \circ (v-w) = v \circ v - 2v \circ w + w \circ w,\ (v-w) \circ (v+w) = v \circ v - w \circ w.$$

Jedną z najbardziej istotnych własności jest:

**Twierdzenie 2.1.** *Mamy*

$$v \circ w = \|v\| \cdot \|w\| \cos(\angle v, w).$$

*Dowód.* Rozpatrzmy trójkąt zaczepiony w zerze, którego pozostałym wierzchołkami są punkty $v$ i $w$ (długość $v$ oznaczmy przez $a$ i długość $w$ przez $b$). Możemy teraz odległość $v$ od $w$ policzyć na dwa sposoby. Z jednej strony, bazując na powyższych własnościach iloczynu skalarnego dostajemy

$$\|v-w\|^2 = (v-w) \circ (v-w) = v \circ v + w \circ w - 2v \circ w = \|v^2\| + \|w\|^2 - 2v \circ w.$$

Łącząc powyższy wzór z tw. cosinusów (2.1), dostajemy tezę. □

Ważną konsekwencją powyższego, jest to, że aby sprawdzić, czy wektory są prostopadłe, wystarczy policzyć, czy ich iloczyn skalarny jest równy zero.

**Zadanie 2.3.** *Proszę policzyć kąt pomiędzy wektorami* $(1,2)$ *i* $(5,3)$.

**Zadanie 2.4.** *Proszę sprawdzić dla jakiego parametru* $a$ *wektory* $(1,a)$ *i* $(a-2,a)$ *są do siebie prostopadłe.*

Przyda się nam jeszcze pojęcie prostej. Prostą zazwyczaj zadajemy jedną z następujących postaci:

- $x_2 = ax_1 + b$, $x_2 = a(x_1 - \bar{x})$;
- $x_1 = a_1 t + b_1$, $x_2 = a_2 t + b_2$ gdzie $a_1 \neq 0$ lub $a_2 \neq 0$ parametrycznie;
- $ax_1 + bx_2 + c = 0$ lub $a(x_1 - \bar{x}_1) + b(x_2 - \bar{x}_2) = 0$ gdzie $a \neq 0$ lub $b \neq 0$ (w sposób uwikłany)

Przyda nam się:

**Twierdzenie 2.2.** *Prosta* $P$ *przechodząca przez punkt* $(\bar{x}_1, \bar{x}_2)$ *i prostopadła do niezerowego wektora* $(v_1, v_2)$ *wyraża się wzorem* $v_1(x_1 - \bar{x}_1) + v_2(x_2 - \bar{x}_2) = 0$.

*Dowód.* Mamy

$$\begin{aligned} P &= \{x \in \mathbb{R}^2 : (x - \bar{x}) \perp v\} = \{x \in \mathbb{R}^2 : v \circ (x - \bar{x}) = 0\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : v_1(x_1 - \bar{x}_1) + v_2(x_2 - \bar{x}_2) = 0\}. \end{aligned}$$

□

Następne ważne twierdzenie, będzie dotyczyło obserwacji które punkty są bliżej wybranym punktom.

**Twierdzenie 2.3.** *Rozpatrzmy punkty* $v, w \in \mathbb{R}^2$. *Wtedy zbiór punktów na płaszczyźnie równo odległych od* $v$ *i* $w$ *to jest dokładnie prosta przechodząca przez* $(v+w)/2$ *i prostopadła do wektora* $w - v$.

*Co więcej*

- *punkt* $x$ *jest bliżej* $w$ *o ile* $\langle x - \frac{v+w}{2}, w - v \rangle > 0$;

- *punkt $x$ jest bliżej $v$ o ile $\langle x - \frac{v+w}{2}, w-v\rangle < 0$.*

*Dowód.* Mamy

$$\begin{aligned}
&\{x \in \mathbb{R}^2 : \|x-w\| < \|x-v\|\} = \{x : \langle x-w, x-w\rangle < \langle x-v, x-v\rangle\}\\
&= \{x : \langle x,x\rangle - 2\langle x,w\rangle + \langle w,w\rangle < \langle x,x\rangle - 2\langle x,v\rangle + \langle v,v\rangle\}\\
&= \{x : 2\langle x, w-v\rangle > \langle w,w\rangle - \langle v,v\rangle\} = \{x : 2\langle x, w-v\rangle > \langle w+v, w-v\rangle\}\\
&= \{x : \langle x, w-v\rangle - \langle \tfrac{w+v}{2}, w-v\rangle\}
\end{aligned}$$

□

**Uwaga 2.2.** *Powyższe twierdzenie jest WAŻNE: przyda się nam to potem zarówno w metodzie k-means (analiza skupień, diagram Voronoi) jak i w dyskryminacji.*

**Zadanie 2.5.** *Niech dana będzie prosta $a_1x_1 + a_2x_2 + b = 0$ w $\mathbb{R}^2$ i punkt $\bar{x} = (\bar{x}_1, \bar{x}_2) \in \mathbb{R}^2$. Udowodnij, że odległość tego punktu od prostej wynosi*

$$\frac{|a_1\bar{x}_1 + a_2\bar{x}_2 + b|}{\sqrt{a_1^2 + a_2^2}}.$$

Możliwe są następujące położenia dwóch prostych: albo się przecinają dokładnie w jednym punkcie, albo są równoległe. Jeżeli są równoległe, to albo się pokrywają (są identyczne), albo się nie przecinają w żadnym punkcie.

Więcej informacji o prostej można znaleźć na wikipedii (zachęcam do samodzielnego przeglądnięcia): `http://pl.wikipedia.org/wiki/Prosta`

KONIEC WYKŁADU 2

## 2.2 Przestrzeń trójwymiarowa, czyli $\mathbb{R}^3$

Przestrzeń trójwymiarowa jest pod kątem operacji bardzo podobna do $\mathbb{R}^2$, dlatego będę tylko skrótowo wymieniał pewne własności.

**Obserwacja 2.2.** *Korzystając z Tw. Pitagorasa widać, że odległość punktu $x = (x_1, x_2, x_3)$ od zera wyraża się wzorem $\sqrt{x_1^2 + x_2^2 + x_3^2}$.*

Tak więc analogicznie jak dla płaszczyzny przez normę euklidesową $x$ rozumiemy odległość $x$ od zera, czyli

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

Iloczyn skalarny dwóch wektorów $v, w \in \mathbb{R}^3$ dany jest wzorem

$$v \circ w = v_1w_1 + v_2w_2 + v_3w_3.$$

**Zadanie 2.6.** *Niech $u = (1, 2, -1)$, $v = (3, 1, -5)$. Korzystając z własności iloczynu sprawdź, czy wektory te są prostopadłe i policz długość wektora $2u - 3v$.*

Przyda się nam jeszcze pojęcie prostej i płaszczyzny w $\mathbb{R}^3$. Prostą zazwyczaj zadajemy w postaci

- (jako funkcję $x$) $y = ax + b$, $z = cx + d$ (analogicznie można jako funkcję $y$ lub $z$)
- (parametrycznie) $x = a_1t + b_1$, $y = a_2t + b_2$, $z = a_3t + b_3$
- (krawędziowo, jako przecięcie dwóch płaszczyzn)

Płaszczyznę zazwyczaj zadajemy jedną z następujących postaci:

- ($z$ jako funkcje zmiennych $x$ i $y$) $z = ax + by + c$ (analogicznie można dla $x$ lub $y$)
- (parametrycznie) $x = a_1t + b_1s + c_1$, $y = a_2t + b_2s + c_2$, $z = a_3t + b_3s + c_3$, gdzie wektory $a$ i $b$ nie są równoległe
- $Ax + By + Cz + D = 0$ lub $A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$ gdzie $(A, B, C) \neq 0$.

Proszę spróbować zobaczyć co może wyjść z przecięcia 3 płaszczyzn w $\mathbb{R}^3$ (i zrobić odpowiednie rysunki).

**Zadanie 2.7.** *Wypisz równanie płaszczyzny przechodzącej przez punkty $A(1, 0, 1), B(2, 0, -1), C(3, 1, 2)$ (to zadanie można łatwo zrobić korzystając z iloczynu wektorowego).*

**Zadanie 2.8.** *Niech dana będzie płaszczyzna $Ax+By+Cz+D = 0$ w $\mathbb{R}^3$ i punkt $(x_0, y_0, z_0) \in \mathbb{R}^3$. Udowodnij, że odległość tego punktu od płaszczyzny wynosi*

$$d = \frac{|Ax_0 + By_0 + Cz_0 + D|}{\sqrt{A^2 + B^2 + C^2}}.$$

Przez *iloczyn mieszany* $[a, b, c]$ trzech wektorów $a, b, c$ z $\mathbb{R}^3$ rozumiem $a \circ (b \times c)$.

**Zadanie 2.9.** *Pokaż, że objętość równoległościanu którego wierzchołkiem jest początek układu współrzędnych oraz punkty $a, b, c$ (są jeszcze inne wierzchołki, ale wystarczy chyba podać te do opisu) wynosi $|[a, b, c]|$.*

## 2.3 Przestrzeń $\mathbb{R}^n$

Podstawowe operacje w $\mathbb{R}^n$:

- $(x_1, \dots, x_n) + (y_1, \dots, y_n) := (x_1 + y_1, \dots, x_n + y_n)$,
- $\alpha \cdot (x_1, \dots, x_n) := (\alpha x_1, \dots, \alpha x_n)$, gdzie $\alpha \in \mathbb{R}$ skalar.

Norma euklidesowe i iloczyn skalarny definiuje się analogicznie jak w przestrzeni $\mathbb{R}^2$ i $\mathbb{R}^3$ (w przestrzeni $\mathbb{R}^n$ dla $n > 3$ nie ma bezpośredniego odpowiednika iloczynu wektorowego dla dwóch wektorów). Analogicznie także definiujemy pojęcie wypukłości.

**Przykład 2.1.** *Muzyka*

## 2.4 Macierze

Macierze $M_{n\times k}(\mathbb{R})$ o współczynnikach z $\mathbb{R}$, zapisuje się też jako $\mathbb{R}^{n\times k}$. Zapis:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix}$$

Dodawanie, odejmowanie i mnożenie przez skalary po współrzędnych. Iloczyn skalarny analogicznie

$$\langle A, B \rangle = \sum_{i,j} a_{ij} b_{ij} \text{ dla } A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{n\times k}$$

**Przykład 2.2.** *naturalny model: zdjęcia czarno-białe, przykłady dodawania, odejmowania, mnożenia i interpretacje.*

*kolorowe: RGB (trzy macierze)*

# Rozdział 3

# Kompresja i dyskretyzacja

**Streszczenie.** Przedstawimy podstawy zastosowań przestrzeni $\mathbb{R}^N$ w informatyce (kompresja i dykretyzacja) jak i analizie danych (metoda k-means – podstawowa metoda analizy skupień = klastrowania/klasteryzacji). Naturalnym modelem jest przetwarzanie zdjęć.

## 3.1 Kompresja - postawienie problemu w $\mathbb{R}$

Kompresja stratna i bezstratna.

Dane zajmowałyby za dużo pamięci.

**Problem 3.1.** Zastąpić grupę punktów/danych $X = \{x_i\}_{i=1..k} \subset \mathbb{R}$ za pomocą jednego, tak by był minimalny błąd.

Dwa pytania:

- co rozumiemy przez błąd?
- jak znaleźć ten jeden punkt (minimum)?

**Dygresja 3.1.** *Funkcja $ax^2 + bx + c$, gdzie $a > 0$, osiąga minimum w punkcie $-b/(2a)$. Wartość minimalna wynosi $-\Delta/(4a)$.*

Jednym z najczęściej stosowanych sposobów pomiaru błędu jest tak zwany *błąd kwadratowy* (**se**: squarred error) popełniany przy zastąpieniu każdego punktu z zestawu danych $X$ przez jeden punkt $v$:

$$\mathbf{se}(X, v) = \sum_i |x_i - v|^2.$$

Łatwo widać, że

$$\mathbf{se}(X, v) = k[v^2 - 2(\tfrac{1}{k}\sum_i x_i)v + \tfrac{1}{k}\sum_i x_i^2].$$

W konsekwencji otrzymujemy, że minimum jest uzyskiwane dla $v$ równego średniej:

$$v = E(X) = \tfrac{1}{k}\sum_i x_i,$$

zaś wartość tego minimalnego błędu jest równa

$$\mathbf{se}(X, E(X)) = \sum_i |x_i - E(X)|^2 = \sum_i (x_i^2 - E(X))^2 = kE(X^2) - kE(X)^2.$$

gdzie pierwszą równość jedno dostajemy bezpośrednio wstawiając, a drugie z wzoru, że minimum funkcji kwadratowej wynosi $-\Delta/4a$. Jeżeli weźmiemy pod uwagę wartość tego błędu uśrednionego (t.j. podzielonego przez ilość elementów $k$ zbioru $X$) to dostajemy definicję tak zwanej *wariancji*:

$$\mathrm{Var}(X) = \frac{1}{k}\sum_i |x_i - E(X)|^2 = E(X^2) - E(X)^2 = \tfrac{1}{k}\sum_i x_i^2 - (\tfrac{1}{k}\sum_i x_i)^2. \tag{3.1}$$

Często piszemy $EX$ zamiast $E(X)$ o ile nie powoduje to nieporozumień. Przez $\sigma$ oznaczamy *odchylenie standardowe* dane wzorem $\sigma^2 = \mathrm{Var}(X)$ (pierwiastek z wariancji). Z (3.1) mamy

$$\sigma_X^2 = \mathrm{Var}(X) = E((X - EX)^2) = EX^2 - (EX)^2.$$

Koniec Wykładu 3

**Uwaga 3.1.** *Dobre: szybko się liczy (liniowo względem ilości danych), naturalna motywacja.*

*Złe: otrzymujemy często wynik który nie istnieje (dla przykładu dla koloru), duża czułość na wartości oddalone (outliers=zaburzenia/błędy w danych).*

*:-) Presja inflacyjna :-) - większość ludzi zarabia poniżej średniej.*

Zobaczymy teraz co by się stało gdybyśmy rozważali błąd innego typu.

Załóżmy, że interesuje nas błąd dany przez:

$$\sum_i |x_i - \bar{x}|.$$

**Lemat 3.1.** *Załóżmy dodatkowo, że $X$ jest posortowany, to znaczy $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{k-1} \leqslant x_k$. Rozpatrzmy funkcję*

$$f : \bar{x} \to \sum_{i=1}^{k} |x_i - \bar{x}|.$$

*Wtedy $f'(x) = 2i - k$ dla $x \in (x_i, x_{i+1})$ (gdzie $x_0$ interpretujemy jako $-\infty$ a $x_{k+1}$ jako $+\infty$).*

*Dowód.* Zauważmy, że funkcja $x \to |x_i - x|$ ma pochodną w punkcie $x$ równą $-1$ o ile $x_i > x$ i 1 o ile $x_i < x$. Oznacza to, że pochodna funkcji $f$ w punkcie $x$ jest równa

$$\mathrm{card}\{i : x_i < x\} - \mathrm{card}\{i : x_i > x\} = \mathrm{card}\{i : x_i < x\} - (k - \mathrm{card}\{i : x_i < x\}),$$

co daje tezę. $\square$

**Wniosek 3.1.** *Przy założeniu jak wyżej (zbiór $X$ posortowany), funkcja $f$ ma następujące własności:*

- *$k$ parzyste: silnie maleje na przedziale $(-\infty, x_{k/2}]$; jest stała na przedziale $[x_{k/2}, x_{k/2+1}]$; silnie rośnie na przedziale $[x_{k/2+1}, \infty)$.*
- *$k$ nieparzyste: silnie maleje na przedziale $(-\infty, x_{(k+1)/2}]$; silnie rośnie na przedziale $[x_{(k+1)/2}, \infty)$.*

**Definicja 3.1.** W ten sposób otrzymujemy definicję mediany – dowolny punkt taki, że ilość punktów ze zbioru silnie mniejszych jest równa ilości silnie większych. Jak nieparzysta ilość danych, to jednoznacznie zdefiniowana, jeżeli parzysta, to przedział. Jak widzimy ważne jest by dane wstępnie posortować.

**Uwaga 3.2.** *Dobre: w miarę szybko się liczy ($n \log n$ względem ilości danych), naturalna motywacja, zawsze otrzymujemy realną daną, mniejszy wpływ outliersów.*

*Złe: nie ma wzoru dla sytuacji wyżej wymiarowej (nawet na płaszczyźnie)!*

## 3.2 Sytuacja wyżej wymiarowa

Pokażemy, że jest pełna analogia z przypadkiem jednowymiarowym. Zastąpienie grupy danych $X = \{x_i\} \subset \mathbb{R}^N$ przy pomocy jednej $v$ - kompresja minimalizowanie wartości:

$$v \to \sum_i \|x_i - v\|^2 \tag{3.2}$$

**Twierdzenie 3.1.** *Rozpatrzmy funkcję*

$$g : x \to a\langle x, x\rangle + \langle b, x\rangle + c.$$

*Wtedy*

$$g(x) = a\langle x + b/(2a), x + b/(2a)\rangle - (\langle b, b\rangle - 4ac)/(4a) = a\|x + b/(2a)\|^2 - \Delta/(4a).$$

*Oznacza, że jeżeli $a > 0$ to minimum jest przyjmowane dla $x = -b/(2a)$ i wynosi $-\Delta/(4a)$ (bo $\|x+b/(2a)\|^2$ jest minimalizowane dla $x = -b/(2a)$).*

*Dowód.* Taki sam jak dla jednowymiarowego przypadku (sprowadzanie do postaci kanonicznej). □

Wzór który wylicza (3.2) do jednej postaci – kluczowa jest obserwacja, że suma funkcji kwadratowych jest funkcją kwadratową!

**Stwierdzenie 3.1.** *Mamy wzór:*

$$\frac{1}{k}\sum_i \|x_i - v\|^2 = \|v\|^2 - 2\langle E(X), v\rangle + \frac{1}{k}\sum_i \|x_i\|^2 = \|v - E(X)\|^2 + (\tfrac{1}{k}\sum_i \|x_i - E(X)\|^2).$$

Widzimy kiedy się ten obiekt minimalizuje! W konsekwencji, podobnie jak w przypadku jednowymiarowym, dla zbioru danych $X = \{x_i\}_{i=1..k}$ w sposób naturalny zdefiniowaliśmy średnią:

$$E(X) = \tfrac{1}{k}\sum_i x_i,$$

oraz uogólniony odpowiednik wariancji (na potrzeby wykładu używam tego samego oznaczenia):

$$\mathrm{Var}(X) = \tfrac{1}{k}\sum_i \|x_i - E(X)\|^2.$$

**Ćwiczenie 3.1.** *Proszę pokazać, że*

$$\mathrm{Var}(X) = \tfrac{1}{2k^2}\sum_{i,j} \|x_i - x_j\|^2.$$

**Ćwiczenie 3.2** (do metody otsu)**.** *Korzystając z poprzedniego ćwiczenia, proszę pokazać, że przy rozbiciu $X$ na dwa rozłączne podzbiory $X_1, X_2$ mamy*

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \tfrac{k_1 k_2}{k^2}(EX_1 - EX_2)^2.$$

## 3.3 Dyskretyzacja, diagram Voronoi i k-means

**Przykład 3.1** (binaryzacja zdjęć)**.** *Typowe zdjęcie czarno-białe ma kolory zapisywane za pomocą jednego bajtu, czyli mają zakres od* 0 *(kolor czarny, piksel nie jest zapalony) do* 255 *(biały, piksel jest zapalony). Okazuje się, że z punktu widzenia niektórych metod przetwarzania obrazu (pracy z obrazem) jest łatwiej i szybciej pracować ze zdjęciem zbinaryzowanym, czyli takim które ma tylko dwa kolory: czarny i biały (a nie ma żadnych odcieni szarości). Typowy sposób postępowania, to przyjęcie, pewnego progu* $K \in \{0..255\}$ *binaryzacji, i piksele o kolorze słabo mniejszym od progu zastępujemy kolorem czarnym (zero), a silnie większym przez kolor czarny (255).*

*Oczywiście, pojawia się naturalne pytanie jaki próg jest optymalny, ale na to da nam odpowiedź próg Otsu (metoda k-means).*

**Przykład 3.2** (praca ze zdjęciem)**.** *Często dokonujemy obróbki zdjęcia czarno białego, i w konsekwencji otrzymujemy wartości rzeczywiste. Musimy następnie wrócić do zakresu* 0..255 *– jest dość oczywiste jak to zrobić, ale przydałoby się jakieś uzasadnienie.*

**Problem 3.2.** Postawienie problemu: Mamy dany zestaw możliwych punktów których używamy do dyskretyzacji $V = \{v_1, \dots, v_k\} \subset \mathbb{R}^N$.

Chcemy znaleźć, dla zestawu danych $X \subset \mathbb{R}^N$, przyporządkowanie punktom indeksu

$$X \ni x \to j(x) \in \{1, \dots, k\}$$

tak by zminimalizować całkowity (kwadratowy) błąd popełniony przy dyskretyzacji

$$\sum_i \|x_i - v_{j(x_i)}\|^2.$$

Widać, że wystarczy nam się zająć tym, którym punktem ze zbioru $V$ należy przybliżyć $x$, aby błąd był możliwie najmniejszy. Oczywiście tym które minimalizuje odległość od $x$, czyli kładziemy

$$j(x) = \operatorname*{argmin}_{j \in \{1,\dots,k\}} \|x - v_j\|.$$

Czyli inaczej mówiąc, przybliżamy $x$ najbliższym elementem ze zbioru $V$. Powyższe oznacza, na podstawie wcześniejszych wyliczeń, że płaszczyzna (przestrzeń) rozbija się na wielokąty (wielościany), reprezentujące zbiory punktów dla których dany element $v \in V$ jest najbliższy – to jest tak zwany *diagram Voronoi*.

Diagram Voronoi można zobaczyć w: `http://alexbeutel.com/webgl/voronoi.html`

Konkludując wtedy całkowity błąd kwadratowy (squarred error) przy możliwości zastąpienia punktu $x$ przez dowolny element z $V$ wynosi

$$\mathbf{se}(X, V) = \sum_i d^2(x; V),$$

gdzie $d(x; V)$ oznacza odległość $x$ od zbioru $V$ i jest dana przez $d(x; V) = \min\{\|x - v\| : v \in V\}$.

Natomiast wyobraźmy sobie, że możemy dobrać $V$ mające $k$ punktów dowolnie. Prowadzi nas to do

**Problem 3.3.** Niech $X$ będzie ustalonym zbiorem i niech $k$ będzie liczbą naturalną. Znaleźć zbiór $V$ składający się z $k$ punktów $v_1, \dots, v_k$ tak by zminimalizować błąd $\mathbf{se}(X, V)$ popełniony przy dykretyzacji $X$ za pomocą elementów $V$.

Okazuje się, że powyższy problem nie daje się efektywnie rozwiązać (w informatyce mówi się, że jest NP-trudny), natomiast prowadzi do metody *k-means*:

1. początkowo (kładziemy $l = 0$) jako $V^l = \{v_1^l, \dots, v_k^l\}$ wybieramy losowe/dowolne elementy zbioru $X$;
2. dokonujemy dyskretyzacji $X$ za pomocą $V^l$, wtedy $X$ rozdziela się nam na podzbiory $X_j^l$ punktów które będą zastąpione (inaczej mówiąc którym najbliżej do) przez $v_j^l$;

3. zauważmy, że z tego co pokazaliśmy wcześniej, błąd kwadratowy zmiejeszymy, jeżeli zamiast dyskretyzacji $X_j^l$ przez $v_j^l$ zastąpimy go przez jego średnią, czyli kładziemy $v_j^{l+1} = E(X_j^l)$ i $V^{l+1} = \{v_1^{l+1}, \ldots, v_k^{l+1}\}$;
4. zwiększamy $l$ o jeden, i o ile zmieniło się choć jedno $v_j$ (w stosunku do poprzedniego kroku), skaczemy do punktu 2, w przeciwnym razie kończymy procedurę.

Widać, że powyższa procedura za każdym krokiem w sposób gwarantowany minimalizuje nam błąd kwadratowy. Nie mamy oczywiście natomiast żadnej gwarancji, że znajdziemy w ten sposób globalne minimum (aby zwiększyć szanse by tak było, zazwyczaj startuje się wielokrotnie wybierając różne punkty początkowe na start).

Koniec Wykładu 4

## 3.4 Zastosowania dyskretyzacji i metody k-means w analizie zdjęć

zdjęcia w odcieniach szarości (grayscale)

binaryzacja - progowanie

histogram?

jeden ze standartów binaryzacji - automatyczne szukanie progu - metoda Otsu (binaryzacja Otsu)

segmentacja obrazów - metoda k-means dla zbioru danych w $\mathbb{R}^5$ powstałego przez [kolor+współrzędne]$\subset \mathbb{R}^3 \times \mathbb{R}^2 = \mathbb{R}^5$ (kolor ma zazwyczaj 3 współrzędne), aby zrównoważyć ewentualną różną wielkość (rozdzielczość w obrazie), dajemy współrzędne przestrzenne z wagą $w$:

$$\text{pixel} \to (\text{pixel\_colors}, w \cdot \text{pixel\_coordinates})$$

# Rozdział 4

# Podstawy równań liniowych

## 4.1 Motywacja: Google Pagerank, geometria

Chciałbym teraz skrótowo opisać jedno z najważniejszych (i najbardziej intratnych) zastosowań algebry liniowej, czyli sposób wyszukiwania informacji przez firmę Google[1] (aby się dowiedzieć więcej polecam książkę: [A. Langville, C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings,* Princeton University Press, 2006]).

Załóżmy, że chcemy znaleźć strony które zawierają słowo „krowa”, i poszeregować je względem ważności (popularności). O ile wyszukanie stron na których występuje dane słowo jest (przynajmniej teoretycznie) sprawą prostą, to problem się pojawia przy tym aby sprecyzować, co rozumiemy przez „ważność” danej strony. Kiedyś (przed Google) był z tym problem: uznawano zazwyczaj, że ważność danej strony jest określona przez ilość stron które na nią wskazują. Niestety, takie sortowanie wyników nie dawało to dobrych rezultatów: jeżeli na moją stronę wskazuje choćby tylko jedna strona którą wszyscy znają, to moja strona powinna znaczyć dużo więcej niż gdyby na nią wskazywało dużo stron których prawie nikt nie zna.

Sergey Brin i Lary Page, twórcy algorytmu PageRank spróbowali zastosować prosty model opisujący jak będziemy się poruszać po internecie - jak znajdujemy się na danej stronie, to wybieramy losowo link, i skaczemy do strony na którą wskazujemy[2]. Obserwujemy jak użytkownik będzie się poruszał po stronach, mierzymy jaki czas przebywał na każdej ze stron, i rangujemy strony na podstawie tego czasu.

Takie podejście jest fajne, ale ma pewne problemy:

- nie wiemy jak długo powinniśmy skakać aby mieć w miarę ustalony wynik?
- czy wynik zależy (nie chcielibyśmy aby tak było) od tego na której stronie zaczęliśmy?
- nie jest na razie jasne jak to policzyć efektywnie numerycznie?

Postaram się pokazać odpowiedź na prostym przykładzie. Rozpatrzmy następujący graf który opisuje strony $S1, S2, S3, S4$ (skąd dokąd są linki po których się poruszamy):

$$S1 \to \{S2, S3\}, S2 \to S3, S3 \to S4, S4 \to \{S1, S2, S3\}.$$

Wtedy korzystając z zasady otrzymujemy prawdopodobieństwa przeskoczenia ze strony do strony.

$$S1 \stackrel{1/2}{\to} \{S2, S3\}, S2 \stackrel{1}{\to} S3, S3 \stackrel{1}{\to} S4, S4 \stackrel{1/3}{\to} \{S1, S2, S3\}.$$

[1] nazwa Google wzięła sie od terminu matematycznego „Googol” oznaczającego 1 ze 100 zerami, wymyślonego przez Miltona Sirotta

[2] W rzeczywistości algorytm Google Pagerank dopuszczał możliwość teleportacji - chodziło o to, że człowiek jak był znudzony, to z pewnym ustalonym z góry prawdopodobieństwem mógł skoczyć do dowolnie losowo wybranej strony. Powód zrobienia teleportacji był także teoretyczno-numeryczny – mamy wtedy twierdzenie które gwarantuje jedyność rozkładu stacjonarnego – będzie za chwilkę.

Rozpatrzmy teraz stronę na którą patrzy dany człowiek, rozpatrzmy prawdopodobieństwa $p_1^t, .., p_4^t$, że w chwili $t$ znajduje się na stronach $S_1, .., S_4$. Zobaczmy, jak te prawdopodobieństwa będą wyglądać w następnym kroku (patrzymy na to, w jaki sposób możemy na daną stronę skoczyć). W tym celu przyjrzyjmy się dla przykładu prawdopodobieństwu $p_1^{t+1}$, że będziemy w chwili $t+1$ znajdować się na stronie $S1$: zgodnie z tym co jest napisane powyżej, do strony $S1$ możemy skoczyć tylko ze strony $S4$ z prawdopodobieństwem $1/3$. Oznacza to, że

$$p_1^{t+1} = \tfrac{1}{3}p_4^t.$$

Przeprowadzając analogiczne rozumowanie dla każdej ze stron, otrzymujemy formułę rekurencyjną (równanie różnicowe jak w oryginalnym problemie Fibonacciego) opisującą prawdopodobieństwo w chwili $t+1$ o ile znamy rozkład w chwili $t$:

$$\begin{cases} p_1^{t+1} &= \tfrac{1}{3}p_4^t, \\ p_2^{t+1} &= \tfrac{1}{2}p_1^t + \tfrac{1}{3}p_4^t, \\ p_3^{t+1} &= \tfrac{1}{2}p_1^t + 1p_2^t + \tfrac{1}{3}p_4^t, \\ p_4^{t+1} &= 1p_3. \end{cases}$$

I teraz cały pomysł polega na tym aby wiedzieć (teoria), że interesuje nas sytuacja graniczna, czyli te rozkłady muszą się ustabilizować. Oznacza to, że możemy zmazać literkę $t$, i dostaniemy układ równań:

$$\begin{cases} p_1 &= \tfrac{1}{3}p_4, \\ p_2 &= \tfrac{1}{2}p_1 + \tfrac{1}{3}p_4, \\ p_3 &= \tfrac{1}{2}p_1 + 1p_2 + \tfrac{1}{3}p_4, \\ p_4 &= 1p_3. \end{cases} \tag{4.1}$$

Jedyny czego w tym układzie nie ma to informacji o tym, że to są prawdopodobieństwa (widzimy, że przemnożenie rozwiązania przez liczbę daje nam rozwiązanie). W konsekwencji dorzucamy jeszcze do (4.1) równanie mówiące o tym, że to sa prawdopodobieństwa (czyli, że sumują się do jedynki):

$$p_1 + p_2 + p_3 + p_4 = 1. \tag{4.2}$$

Dostajemy teraz fajny układ równań, który będzie nam dość łatwo rozwiązać: od razu możemy za pomocą $p_4$ podstawić $p_1 = \frac{1}{3}p_4$, $p_3 = p_4$, wstawiamy do drugiego równania dostając $p_2 = \frac{1}{6}p_4 + \frac{1}{3}p_4 = \frac{1}{2}p_4$. Podstawiając w (4.2) dostajemy

$$\tfrac{1}{3}p_4 + \tfrac{1}{2}p_4 + p_4 + p_4 = 1,$$

czyli $p_4 = \frac{6}{17}$ i w konsekwencji otrzymujemy wynik:

$$p_1 = \tfrac{2}{17}, p_2 = \tfrac{3}{17}, p_3 = \tfrac{6}{17}, p_4 = \tfrac{6}{17}.$$

I rangi stron w kolejności to $I(S3, S4), II(S2), III(S1)$.

**Uwaga 4.1.** *Zauważmy istotną różnicę w stosunku do podejścia pre-Google: jeżeli do nas ($S4$) prowadzi link od strony bardzo popularnej ($S3$) i dodatkowo $S3$ nie wskazuje na żadne inne strony, to prawdopodobieństwo znalezienia się na $S4$ jest takie same (lub większe) jak na $S3$ bo znajdując się na stronie $S3$ musimy skoczyć na stronę $S4$.*

**Geometria:** czytelna intuicja geometryczna (przecięcia prostych w $\mathbb{R}^2$; płaszczyzn w $\mathbb{R}^3$, rysunki). Zobaczymy co się może wydarzyć na paru najprostszych przykładach.

**Przykład 4.1.** *dwie proste równoległe: $x + y = 1, x + y = 2$ – układ sprzeczny (brak rozwiązań);*
*trzy proste przecinające się w jednym punkcie: $x = 1, y = 1, x + y = 2$ – dokładnie jedno rozwiązanie;*
*trzy proste które nie mają wspólnego punktu: $x = 1, y = 1, x + y = 0$;*
*jedna prosta: $x + y = 3$ nieskończenie wiele rozwiązań ($y = 3 - x$);*
*proszę sobie spróbować zrobić analogony w $\mathbb{R}^3$.*

Zazwyczaj (najczęściej) jako przecięcie dwóch prostych na płaszczyźnie otrzymujemy punkt, jako przecięcie dwóch płaszczyzn w przestrzeni trójwymiarowej dostajemy prostą, jako przecięcie trzech płaszczyzn (lub płaszczyzny i prostej) dostajemy punkt.

## 4.2 Metoda eliminacji Gaussa: tablice (macierze)

Na razie tablice (macierze) jest traktowane jako forma zapisu układów równań liniowych.

Pokażemy rozwiązywanie metoda eliminacji Gaussa przez sprowadzanie do postaci schodkowej za pomocą operacji elementarnych.

Typowy układ równań liniowych:

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1, \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m. \end{cases}.$$

Geometrycznie rzecz biorąc, szukamy przecięcia $m$ „hiperpłaszczyzn kowymiaru 1" (cokolwiek by to nie znaczyło) w przestrzeni $n$-wymiarowej $\mathbb{R}^n$. W sensie analitycznym mamy $m$ równań w przestrzeni $\mathbb{R}^n$.

Zaczniemy sobie od tego w jaki sposób można (potencjalnie) uprościć, ale bez zmieniania zbioru rozwiązań.

**Twierdzenie 4.1.** *Wykonanie poniższych operacji (tak zwane* operacje elementarne*):*

- *zamiana ze sobą dwóch równań (interpretowane jako wiersze w tablicy/macierzy):* $w_i \leftrightarrow w_j$*;*
- *pomnożenie danego równania (wiersza w macierzy) przez niezerowy skalar* $w_i \mathrel{*}= \alpha$ *(gdzie* $\alpha \in \mathbb{R} \setminus \{0\}$*);*
- *dodanie bądź odjęcie do danego równania innego pomnożonego przez dowolny (niekoniecznie niezerowy) skalar* $\alpha \in \mathbb{R}$*:* $w_i \mathrel{+}= \alpha \cdot w_j$ *(gdzie* $i \neq j$*) [analogicznie można odejmować];*

*nie zmienia zbioru rozwiązań.*

*Dowód.* Wszystkie powyższe operacje są odwracalne:

- $w_k \leftrightarrow w_j$: operacja odwrotna $w_k \leftrightarrow w_j$;
- $w_k \mathrel{*}= \alpha$: operacja odwrotna $w_k \mathrel{*}= \frac{1}{\alpha}$;
- $w_k \mathrel{+}= \alpha \cdot w_j$: operacja odwrotna $w_k \mathrel{-}= \alpha \cdot w_j$;

i w konsekwencji każde wykonanie działanie możemy „cofnąć". □

Skoro tak, to możemy za pomocą powyższych operacji starać się uprościć dany układ równań liniowych. Jeżeli chcemy zapisywać, to wygodniej / szybciej będzie nam stosować zapis macierzowy, czyli będziemy pisać:

$$\left[\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right]$$

Okazuje się, że przy rozwiązywaniu dobrze jest sprowadzać do postaci schodkowej (row echelon form z angielskiego) lub do uproszczonej postaci schodkowej (simplified row echelon form). Zaczniemy sobie od przykładów. Pokażemy potem, że taki sposób postępowania można zapisać w postaci algorytmu (czyli precyzyjnie opisanej procedury postępowania).

**Przykład 4.2.** *Rozpatrzmy równanie:*

$$\begin{cases} 2u & + & v & + & w & = & 5 \\ 4u & - & 6v & & & = & -2 \\ -2u & + & 7v & + & 2w & = & 9 \end{cases}$$

*Dla skrócenia zapisu, zapisujemy w postaci macierzowej:*

$$\left[\begin{array}{rrr|r} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

*Musimy pamiętać, że wiersze odpowiadają równaniom (k-ty wiersz oznaczamy przez $w_k$).*

*Usuwamy pod dwójką w 1 kolumnie: $w_2 \mathrel{+}= -2 \cdot w_1$:*

$$\left[\begin{array}{rrr|r} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ -2 & 7 & 2 & 9 \end{array}\right],$$

*następnie $w_3 \mathrel{+}= w_1$ i dostajemy*

$$\left[\begin{array}{rrr|r} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{array}\right]$$

*Usuwamy pod $-8$, $w_3 \mathrel{+}= w_2$:*

$$\left[\begin{array}{rrr|r} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

*Czyli $w = 2$. I potem postępujemy „wstecz”, czyli znając $w$ wyznaczamy $v$: $-8v - 2w = -12$, $-8v = -8$, czyli $v = 1$. Teraz wyznaczamy $u$: $2u + v + w = 5$, czyli $2u = 5 - 2 - 1$, czyli $u = 1$.*

*Macierz schodkowa* – macierz, której pierwsze niezerowe elementy kolejnych niezerowych wierszy, znajdują się w coraz dalszych kolumnach, a powstałe wiersze zerowe umieszcza się jako ostatnie.

**Twierdzenie 4.2.** *Każda macierz może zostać przekształcona do postaci schodkowej za pomocą operacji elementarnych, w szczególności metody Gaussa.*

Koniec Wykładu 5

**Definicja 4.1.** Macierz *schodkowa zredukowana* to macierz schodkowa, która spełnia następujące warunki:

- jej pierwszym niezerowym elementem kolejnych wierszy (współczynnikiem wiodącym) jest jedynka,
- jeśli wyraz $a_{ij}$ znajduje się w tej samej kolumnie, co pewien współczynnik wiodący i w wierszu powyżej tego współczynnika, to $a_{ij} = 0$.

Sprowadzanie do postaci schodkowej zredukowanej (wersja eliminacji Gaussa) nazywa się czasem metodą Gaussa-Jordana.

Aby sprowadzić macierz do postaci schodkowej zredukowanej, najpierw sprowadzamy do postaci schodkowej, a następnie idąc od schodków najbardziej na prawo do tych najbardziej na lewo:

1. najpierw dzielimy wiersz w którym znajduje się schodek przez wartość macierzy w „schodku” – wtedy wartość w schodku staje się równa jeden;
2. następnie zerujemy wszystkie wartości poza schodkiem znajdujące się w kolumnie.

Macierz schodkowa zredukowana jest wygodna, jeżeli chcemy rozwiązywać równania liniowe na komputerze, gdyż algorytm doprowadza do postaci w której w zasadzie mamy gotowe rozwiązanie. Rozpatrujemy macierz rozszerzoną którą sprowadzamy do postaci schodkowej.

- jeżeli w ostatniej kolumnie (odpowiadającej wyrazom wolnym) mamy schodek, to znaczy, że układ nie ma rozwiązań;
- w przeciwnym razie, zmienne które odpowiadają schodkom wyznaczają się bezpośrednio za pomocą parametrów, które odpowiadają zmiennym w którym nie ma schodków.

## 4.3 Przykłady

Pokażemy teraz parę przykładów sprowadzania macierzy do postaci schodkowej i do postaci schodkowej zredukowanej. Następnie rozwiążemy przykładowe równania.

**Przykład 4.3.** *Sprowadzimy macierz do postaci schodkowej za pomocą metody eliminacji Gaussa:*

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

*Ponieważ pierwsza kolumna jest cała zerowa, patrzymy na drugą. W drugiej pierwszy element niezerowy jest dopiero w trzecim wierszu, musimy więc zamienić wiersz pierwszy z trzecim ($w_1 \leftrightarrow w_3$):*

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

*Teraz redukujemy za pomocą operacji elementarnych wszystkie niezerowe wyrazy znajdujące się poniżej* $1$, $w_4 -= 2w_1$*:*

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

*Pod jedynką są teraz same zera, możemy więc przejść do drugiego wiersza. Za pomocą* $3$ *redukujemy* $1$ *w wierszu trzecim,* $w_3 -= w_2/3$*:*

$$\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 5/3 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

*Teraz zostało jedynie wyzerować 3 w wierszy czwartym za pomocą* $5/3$ *z wiersza trzeciego,* $w_4 -= \frac{9}{5} w_3$*:*

$$\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 5/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

*Dostaliśmy postać schodkową naszej macierzy. Jeszcze zaznaczę za pomocą* $*$ *„schodki” (czyli współczynniki wiodące):*

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

**Przykład 4.4.** *Sprowadzenie do postaci schodkowej (bez drobiazgowego komentarza):*

- $\begin{bmatrix} 1 & 0 & 3 & 1 \\ 2 & 0 & 1 & 2 \end{bmatrix}$ $(w_2 -= 2w_1)$ $\begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 0 & -5 & 0 \end{bmatrix}$, *schodki:* $\begin{bmatrix} * & & & \\ 0 & 0 & * & \end{bmatrix}$;

- $\begin{bmatrix} 0 & 1 \\ 3 & 1 \\ -1 & 0 \\ 1 & 2 \end{bmatrix} (w_2 \leftrightarrow 2w_1) \begin{bmatrix} 3 & 1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 2 \end{bmatrix} (w_3 \mathrel{+}= w_1/3; w_4 \mathrel{-}= w_1/3) \begin{bmatrix} 3 & 1 \\ 0 & 1 \\ 0 & 1/3 \\ 0 & 7/3 \end{bmatrix}$

  $(w_3 \mathrel{-}= w_2/3; w_4 \mathrel{-}= \frac{7}{3}w_2) \begin{bmatrix} 3 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, *schodki* $\begin{bmatrix} * & 1 \\ 0 & * \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

**Przykład 4.5.** *Sprowadzimy teraz macierz do postaci schodkowej zredukowanej:*

$$\begin{bmatrix} 2 & 1 & 3 & 1 \\ 0 & 0 & -5 & 0 \\ 2 & 2 & 3 & 1 \end{bmatrix} (w_1 \mathrel{*}= \frac{1}{2}) \begin{bmatrix} 1 & 1/2 & 3/2 & 1/2 \\ 0 & 0 & -5 & 0 \\ 2 & 2 & 3 & 1 \end{bmatrix} (w_3 \mathrel{-}= 2w_1) \begin{bmatrix} 1 & 1/2 & 3/2 & 1/2 \\ 0 & 0 & -5 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$(w_2 \leftrightarrow w_3) \begin{bmatrix} 1 & 1/2 & 3/2 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -5 & 0 \end{bmatrix} (w_1 \mathrel{-}= w_2/2; w_3 \mathrel{*}= \frac{-1}{5}; w_1 \mathrel{-}= \frac{3}{2}w_3) \begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Ponieważ umiemy już sprowadzać macierz do postaci schodkowej i schodkowej zredukowanej, pokażemy na paru przykładach jak to się przydaje do rozwiązywania równań liniowych.

**Przykład 4.6.** *Rozpatrzmy układ równań*

$$\begin{cases} x + y + z = 1 \\ 2x - 3y - z = 2 \\ x + 6y + 4z = 1 \end{cases}$$

*Geometrycznie rzec biorąc przecinamy 3 płaszczyzny w przestrzeni trójwymiarowej. Jak się okaże, rozwiązaniem będzie prosta.*

*Macierz stowarzyszoną z tym układem sprowadzamy do postaci schodkowej (pierwsza kolumna odpowiada zmiennej $x$, druga $y$, trzecia $z$)*

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & -3 & -1 & 2 \\ 1 & 6 & 4 & 1 \end{array}\right] (w_2 \mathrel{-}= 2w_1; w_3 \mathrel{-}= w_1) \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -5 & -3 & 0 \\ 0 & 5 & 3 & 0 \end{array}\right] (w_3 \mathrel{+}= w_2) \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -5 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

*Zobaczmy jeszcze gdzie są „schodki”*

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

*Ponieważ nie ma „schodka” w ostatniej kolumnie, więc układ ma rozwiązanie. Równanie będzie zależało od parametru $z$ (parametrem będzie każda zmienna dla której w odpowiadającej kolumnie nie ma schodka).*

*Wyliczamy więc „od końca” kolejne zmienne: $-5y - 3z = 0$, czyli $y = -\frac{5}{3}z$, i dalej $x + y + z = 1$, czyli $x = 1 - (-\frac{5}{3}z) - z = 1 + \frac{2}{3}z$.*

**Zadanie 4.1.** *Proszę rozwiązać (przy pomocy metody eliminacji Gaussa) układy równań:*

1. $x = 1$, $x = 2$;

2. $x + y = 1$, $x + 2y = 2$, $2x + 3y = 3$;

3. $x + y = 1$, $x + 2y = 2$, $2x + 3y = 0$;

*4.* $3x + y = 1$, $x + 2y = 5$;

*5.* $x_2 + x_3 + x_4 + x_5 = 1$, $x_2 - x_3 + x_6 = 0$, $x_1 - x_2 + 2x_6 = -1$, $x_4 - x_5 = 2$, $x_1 - x_3 + x_4 = 0$;

*6.* $a + b + c = 2$, $-2a + 2b + 3c = 1$, $-a + 3b + 4c = 3$;

*7.* $w + v = 1$, $u + z = 2$, $v + u + w = 3$, $w - u = 1$.

KONIEC WYKŁADU 6

# Rozdział 5

# Operacje macierzowe

## 5.1 Typy wektorów i macierzy

KONWENCJA ZAPISU WEKTORÓW:

W przypadku $\mathbb{R}^n$ przez $e_i$ będę rozumiał wektory które na $i$-tym miejscu mają jedynkę, a poza tym same zera (tak zwane wektory bazy kanonicznej):

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, e_{n-1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}, e_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Ponieważ jak widać tak zapisane wektory zajmują dużo miejsca, często zapisuje się używając operacji transpozycji (operacja zamienia w macierzy wiersze z kolumnami), i wtedy na przykład $e_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T$. W niektórych programach (n.p. Matlab) jest przyjęta jeszcze konwencja, że jak zapisujemy macierze, to przecinek oznacza postawienie macierzy obok siebie, a średnik oznacza pod (wygodne jeżeli używamy macierzy budując je z bloków innych macierzy). Wtedy w tej konwencji zapis $e_1$ mógłby być $e_1 = [1; 0; \dots; 0] = [1, 0, \dots, 0]^T$.

PODSTAWOWE TYPY MACIERZY:

- Macierz zerowa oznaczamy przez 0 (jest tu pewna kolizja oznaczeń). Zdefiniowana jest dla dowolnych macierzy.

- Macierze kwadratowe to są macierze które mają równą liczbę wierszy i kolumn.

- Macierz identycznościowa $I$ jest zdefiniowana tylko dla macierzy kwadratowych (jej wielkość, czyli ilość kolumn i wierszy jest zazwyczaj wyciągana z kontekstu), i składa się z jedynek na przekątnej, zer poza:

$$I = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & & \vdots \\ \vdots & 0 & 1 & \ddots & \\ \vdots & & \ddots & \ddots & \\ 0 & \dots & & & \end{bmatrix}$$

- Inne, często spotykane macierze kwadratowe: diagonalna, górnie trójkątna (spotyka się także nazwę górnie przekątniowa): to taka macierz, która poniżej głównej przekątnej ma same zera (dualnie: dolnie trójkątna, to taka która powyżej przekątnej ma same zera). Z sympatycznych macierzy jest jeszcze macierz trójdiagonalna, która jest potencjalnie niezerowa tylko na głównej przekątnej i „przekątną poniżej i powyżej”:

$$\begin{bmatrix} a_{11} & a_{12} & 0 & \dots & \dots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & & \vdots \\ 0 & a_{32} & a_{33} & a_{34} & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & \end{bmatrix}.$$

Formalnie mówiąc, macierz kwadratowa $A = [a_{ij}]$ jest trójdiagonalna o ile $a_{ij} = 0$ dla wszystkich $i, j$, takich, że $|i - j| > 1$.

- Są jeszcze ważne typy macierzy kwadratowych (symetryczne, dodatnio określone), ale jeszcze troszkę za wcześnie o nich.

- Macierze blokowe – ważne z punktu zastosowań praktycznych (szybsze obliczenia), elementami zamiast liczb rzeczywistych są macierze.

- Macierze rzadkie (sparse matrices) – są to takie macierze w których elementów niezerowych jest znacznie mniej niż zerowych (występują często w praktyce, choćby w NLP=natural Language Processing, czy macierzach sąsiedztwa dla grafów).

## 5.2 Wektory i macierze: działania

Podstawowy zapis macierzy:

$$A := \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

Wtedy $A$ ma $m$-wierszy i $n$-kolumn. Będziemy takie macierze oznaczać przez $(a_{ij})_{mn}$ albo skrótowo $(a_{ij})$ (zbiór macierzy $m \times n$ o współczynnikach z $\mathbb{R}$ oznaczamy przez $M_{mn}(\mathbb{R})$). Przez przekątną główną macierzy rozumiemy elementy $(a_{ii})$.

Mówimy, że macierz jest *kwadratowa* jeżeli ma tyle samo kolumn co wierszy.

Dodawanie, odejmowanie i przemnażanie macierzy przez skalar (liczbę) – tak samo jak było dla wektorów, proste. Dodawanie: możemy dodawać macierze posiadające tę samą ilość kolumn i wierszy (jeżeli $A$ i $B$ to macierze, to dodajemy je (A+B)). Jeżeli $A = (a_{ij})$, $B = (b_{ij})$, to

$$A + B = (a_{ij} + b_{ij}).$$

Analogicznie definiujemy odejmowanie macierzy (jest także określone dla macierzy tych samych wymiarów):

$$A - B = (a_{ij} - b_{ij}).$$

Łatwo można sprawdzić, że $0 + A = A + 0 = A$ (mówimy, że 0 jest elementem neutralnym dodawania). Dodawanie macierzy jest przemienne ($A + B = B + A$) i łączne, to znaczy $(A + B) + C = A + (B + C)$.

Macierze możemy mnożyć przez skalary za pomocą wzoru

$$\alpha \cdot A = (\alpha a_{ij}).$$

Teraz zajmijmy się mnożeniem wektora przez macierz. Wrócimy tutaj do przykładu z Fibonacciego (oryginalny problem!). Przypominam jak wygląda równanie:

$$\begin{cases} stare_{n+1} = 0 \cdot stare_n + 1 \cdot mlode_n, \\ mlode_{n+1} = 1 \cdot stare_n + 1 \cdot mlode_n. \end{cases} \tag{5.1}$$

Prawą stronę równania zapiszmy w postaci macierzowej (jak w równaniach):

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

I teraz chcemy tak zdefiniować mnożenie, by powyższe równanie można było zapisać w postaci:

$$\begin{bmatrix} stare_{n+1} \\ mlode_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} stare_n \\ mlode_n \end{bmatrix} = F \cdot \begin{bmatrix} stare_n \\ mlode_n \end{bmatrix}.$$

I teraz widzimy, że aby to uzyskać musimy przemnożyć (jakby za pomocą iloczynu skalarnego) każdy wiersz macierzy przez wektor $\begin{bmatrix} stare_n \\ mlode_n \end{bmatrix}$.

**Definicja 5.1.** Rozpatrzmy macierz $A = [a_{ij}]_{i=1..n,j=1..k} \in \mathbb{R}^{n \times k}$ o $n$ wierszach i $k$ kolumnach:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix},$$

oraz wektor $v \in \mathbb{R}^k$ zapisany w postaci kolumnowej

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix} \in \mathbb{R}^k.$$

Wtedy w wyniku działania $A \cdot v$ (często zapisywane jako $Av$) jest element przestrzeni $\mathbb{R}^n$ dany wzorem

$$A \cdot v = \overbrace{\begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix}}^{k} \cdot \left.\begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}\right\} k = \begin{bmatrix} a_{11}v_1 + \dots a_{1k}v_k \\ \vdots \\ a_{n1}v_1 + \dots + a_{nk}v_k \end{bmatrix} \in \mathbb{R}^n.$$

UWAGA [bardzo ważne]: jak widać z powyższej procedury, nie ma można (operacja nie da się poprawnie zdefiniować) mnożyć macierzy przez wektor, jeżeli macierz nie ma tylu kolumn jaki jest wymiar wektora. Inaczej mówiąc, macierz $n \times k$ możemy przemnożyć tylko przez wektor o $k$ współrzędnych, i w efekcie powstaje nam wektor o $n$ współrzędnych.

Inaczej mówiąc można powiedzieć, że każdej macierzy $A \in \mathbb{R}^{n \times k}$ odpowiada odwzorowanie (liniowe, zaraz wyjaśnię) z $\mathbb{R}^k$ do $\mathbb{R}^n$:

$$A(v + w) = Av + Aw, \; A(\alpha v) = \alpha Av.$$

Zwyczajowo się utożsamia to odwzorowanie z macierzą.

ZASTOSOWANIA: obroty, rzutowania prostopadłe, zgniecenia, obroty lustrzane, etc.

Postaram się teraz na przykładach w $\mathbb{R}^2$ i $\mathbb{R}^3$ zobrazować geometrycznie (na wykładzie było bardzo dużo przykładów wraz z rysowaniem, tu tylko wypiszę parę przykładów).

**Przykład 5.1.** *Przykładowe operacje na płaszczyźnie które można uzyskać przez macierze:*

- *zgniatanie i rozszerzanie:*

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

- *symetria względem prostej $x = y$*

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

- *obrót o kąt $\alpha$:*

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

Orientacja przestrzeni.

**Przykład 5.2.** *Wstęga Möbiusa.*

**Przykład 5.3.** *Transport obuwia z Polski do Rosji.*

**Przykład 5.4.** *My jesteśmy lewoskrętni.*
*[źródło: `http://sciaga.pl/slowniki-tematyczne/2601/zastosowanie-znaczenie-funkcje/`]*
*Może zdarzyć się także, że jeden z izomerów leku ma działanie lecznicze, a drugi jest bardzo szkodliwy dla organizmu. Najbardziej znanym lekiem takiego typu jest talidomid. Talidomid był środkiem uspokajającym podawanym kobietom ciężarnym. Był popularny w latach 50. i 60. XX wieku w Europie. Lekarze często przepisywali talidomid, jednak po pewnym czasie od momentu rozpoczęcia stosowania tego leku zauważono wzrost urodzeń dzieci z ciężkimi wadami wrodzonymi (braki rąk, nóg).*

*Po przeprowadzonym dochodzeniu okazało się, że tabletki talidomidu zawierały mieszaninę izomerów R i S. Izomer R miał działanie terapeutyczne, natomiast za powstawanie wad wrodzonych u dzieci odpowiedzialny był enancjomer S talidomidu. Od tego czasu bardzo dokładnie bada się działanie wszystkich enancjomerów danego leku, zanim zostanie on wprowadzony na rynek farmaceutyczny. Odkrycie odmiennego działania enancjomerów niektórych leków spowodowało, że opracowano bardzo szczegółowe sposoby rozdziału mieszanin enancjomerów.*

## 5.3 Mnożenie macierzy

**Uwaga 5.1.** *Często jeżeli mamy zestaw danych (formalnie ciąg punktów) $v_1, \ldots, v_n$ z przestrzeni $\mathbb{R}^k$, to często zapisujemy w postaci macierzy*

$$V = [v_1, \ldots, v_n] \in \mathbb{R}^{k \times n}.$$

*Wtedy mnożenie $A \cdot V$, gdzie $A \in \mathbb{R}^{j \times k}$ będzie odpowiadało przemnożeniu każdego wektora składającego się na $V$ przez $A$, i w konsekwencji dostaniemy $n$ wektorów z przestrzeni $\mathbb{R}^j$. Taka interpretacja prowadzi nas w sposób jednoznaczny do definicji mnożenia dwóch macierzy.*

Jeżeli $A$ i $B$ są macierzami, to możemy je wymnożyć o ile ilość kolumn $A$ odpowiada ilości wierszy $B$. Wtedy w macierzy $C = AB$ punkt $c_{ij}$ powstaje „nieprecyzyjnie mówiąc" przez wymnożenie $i$-tego wiersza z $A$ przez $j$-tą kolumnę z $B$ (czyli $c_{ij} = \sum_k a_{ik} b_{kj}$). Na początku najlepiej nauczyć się mnożyć $A$ przez $B$ „układając" $B$ „u góry". Łatwo sprawdzić, że

$$AI = IA = A,$$

gdzie $I$ jest macierzą jednostkową odpowiedniej wielkości dobranej tak by mnożenie miało sens.

Jeżeli $A$ jest macierzą kwadratową, to przez $A^n$ (gdzie $n \in \mathbb{N}$) rozumiem $n$-tą potęgę $A$, czyli (definicja indukcyjna):

$$A^0 = I,\ A^{n+1} = A \cdot A^n.$$

Jeżeli mamy dany wielomian $W(x) = \sum_i a_i x^i$ oraz macierz kwadratową $A$, to przez $W(A)$ rozumiemy $\sum_i a_i A^i$.

Dla prostoty w poniższych zadaniach zapisuję macierze jak w Matlabie (Scilabie), to znaczy średnik oznacza przejście do następnego wiersza.

**Zadanie 5.1.** *Zadania na mnożenie macierzy:*

1. *Niech* $A = [1,1;0,1]$. *Proszę policzyć* $I + A^2$; $W(A)$, *gdzie* $W(x) = 1 - x^3$; $A^n$ *dla* $n \in \mathbb{N}$.
2. *Niech* $A = [0,1;1,1]$. *Policz* $A^n$ *dla* $n \in \mathbb{N}$.
3. *Niech* $A = [1,2,3;3,-1,0]$, $B = [1,2,0;2,0,1;3,-2,0]$, $C = [1,2,-1;3,-2,1]$, $D = [2,1;1,0]$, $E = [1,0;2,-1;0,1]$, $F = [1,2,0]$, $G = [1;2;3]$. *Proszę wymnożyć (o ile to możliwe)*

$$AB, AC, BC, CD, DE, AE, BE, EB, GF, AG, GA, GE, GC, EA.$$

**Uwaga 5.2.** *Mnożenie macierzy nie jest przemienne, to znaczy zazwyczaj* $AB \neq BA$. *Jako przykład wystarczy rozpatrzyć:*

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

*Z drugiej strony można pokazać (nie będziemy tego robić w tym momencie, bo jest dużo pisaniny), że mnożenie macierzy jest łączne, to znaczy*

$$A(BC) = (AB)C,$$

*dla dowolnych macierzy* $A, B, C$ *które można wymnożyć. W związku z tym będziemy pisać po prostu* $ABC$ *zamiast* $A(BC)$ *czy* $(AB)C$.

*Warto zauważyć, że łączność jest ważna – dodawanie liczb na komputerze nie jest łączne, należy w miarę możliwości dodawać od najmniejszych do największych (czyli dodawać liczby tego samego rzędu wielkości).*

**Zadanie 5.2.** *Niech* $A, B, C \in M_{22}(\mathbb{R})$ *(macierze* $2 \times 2$*). Proszę pokazać, że*

$$A(BC) = (AB)C.$$

ZŁOŻONOŚĆ: Mnożenie macierzy kwadratowych: trywialne złożoność taka bezpośrednia jest $O(n^3)$ (patrzymy na ilość mnożeń): zauważmy, że jak mnożymy $A = [a_{ij}]$ przez $B = [b_{ij}]$, to (dla przykładu) popatrzmy w którym mnożeniu występuje $a_{ij}$, jego będziemy mnożyć przez dowolne $b_{jk}$, gdzie $k$ dowolne z zakresu $1..n$. Oznacza to, że mamy $n^2 \cdot n$ mnożeń (bo możemy wybrać $ij$ na $n^2$ sposobów). Zobaczmy jak możemy uzyskać tą złożoność za pomocą macierzy blokowych.

Bezpośrednie zastosowanie najprostsza zasada dziel i rządź ma $n^3$: jeżeli zastosujemy w postaci blokowej:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

gdzie $A, B$ są macierzami $n \times n$, a $A_{ij}$ oraz $B_{ij}$ są macierzami $n_1 \times n_1$ (gdzie $n = 2n_1$), wtedy mamy

$$AB = C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

gdzie $C_{ij}$ są macierzami $n_1 \times n_1$ takimi, że $C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j}$. Wzory na $C_{ij}$ pokazują, że aby je wyliczyć potrzebujemy 8 mnożeń macierzy $n_1 \times n_1$ i 4 dodawania. W konsekwencji, przy dwukrotnym wzroście wielkości macierzy ilość operacji (przypominam, zliczamy tylko mnożenie) wzrasta 8-krotnie. Oznacza to, jak łatwo zauważyć, że złożoność algorytmu jest sześcienna – $O(n^3)$.

Strassen [Gaussian Elimination is Not Optimal, Numer. Math, 1969, vol. 4 pp 354-356] znalazł wzory które wymagają tylko 7 mnożeń:

$$\begin{array}{rcl} I &=& (A_{11} + A_{22})(B_{11} + B_{22}) \\ II &=& (A_{21} + A_{22})B_{11} \\ III &=& A_{11}(B_{12} - B_{22}) \\ IV &=& A_{22}(-B_{11} + B_{21}) \\ V &=& (A_{11} + A_{12})B_{22} \\ VI &=& (-A_{11} + A_{21})(B_{11} + B_{12}) \\ VII &=& (A_{12} - A_{22})(B_{21} + B_{22}) \end{array},$$

oraz

$$\begin{array}{l} C_{11} = I + IV - V + VII \\ C_{21} = II + IV \\ C_{12} = III + V \\ C_{22} = I + III - II + VI \end{array},$$

co oznacza, że jego algorytm ma złożoność $O(n^{\log_2 7} = n^{2,8..})$. Zostało to potem poprawione, ale hipoteza, że złożoność wynosi $O(n^{2+\varepsilon})$ jest o ile wiem, ciągle otwarta.

Więcej informacji można znaleźć na wikipedii: `https://en.wikipedia.org/wiki/Matrix_multiplication_algorithm`.

Koniec Wykładu 7

## 5.4 Macierz odwrotna

W niniejszej sekcji zajmiemy się definicją i liczeniem macierzy odwrotnej do danej.

**Definicja 5.2.** Niech $A \in M_{nn}(\mathbb{R})$. Mówimy, że $B$ jest *macierzą odwrotną do* $A$ (oznaczamy macierz odwrotną przez $A^{-1}$) o ile

$$AB = BA = \mathrm{I}.$$

Jeżeli dla danej macierzy istnieje macierz odwrotna, to tą macierz nazywamy *macierzą odwracalną.*

**Obserwacja 5.1.** *Jeżeli dla danej macierzy istnieje macierz odwrotna, to jest ona jednoznacznie wyznaczona.*

*Dowód.* Załóżmy, że mamy macierze $B_1, B_2$ takie, że

$$AB_1 = B_1A = \mathrm{I} = B_2A = AB_2.$$

Wtedy $B_1 = B_1\mathrm{I} = B_1(AB_2) = (B_1A)B_2 = \mathrm{I}B_2 = B_2$. □

**Zadanie 5.3.** *Proszę pokazać, że macierz*

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

*nie jest odwracalna. Wsk.: proszę najpierw zauważyć, że jeżeli macierz $B$ jest odwracalna, to $Bv \neq 0$ dla każdego wektora $v \neq 0$, a następnie znaleźć takie $v \in \mathbb{R}^2$, że $Av = 0$.*

Spróbujmy więc policzyć macierz odwrotną „bezmyślnie” (proszę tak nie rozwiązywać!!!). Załóżmy, że mamy daną macierz $A = [a_{ij}]$ $n \times n$, i szukamy takiej macierzy $B = [b_{ij}]$, by $AB = \mathrm{I}$. Wtedy

$$\sum_{k=1}^{n} a_{ik}b_{kj} = \delta_{ij} \quad \text{dla } i, j \in \{1, \ldots, n\},$$

gdzie $\delta_{ij} = 1$ jeśli $i = j$ i 0 gdy $i \neq j$. W konsekwencji otrzymujemy $n^2$ równań o zmiennych $b_{ij}$, który możemy spróbować rozwiązać. Jak pokażę u dołu, dla macierzy $2 \times 2$ można się jeszcze w ten sposób bawić (mamy 4 zmienne), ale dla $3 \times 3$ (9 zmiennych) czy $4 \times 4$ (16 zmiennych) to już się robi makabra.

**Przykład 5.5.** *Spróbujmy w powyższy sposób znaleźć macierz odwrotną do*

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

*Niech*

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

*Chcemy, aby $AB = \mathrm{I}$, czyli by*
$$\begin{bmatrix} b_{11}+b_{21} & b_{12}+b_{22} \\ b_{11} & b_{12} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.2}$$
*W konsekwencji dostajemy układ równań*
$$\begin{cases} b_{11}+b_{21} = 1 \\ b_{12}+b_{22} = 0 \\ b_{11} = 0 \\ b_{12} = 1 \end{cases},$$
*czyli otrzymujemy, że*
$$B = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}.$$
*Wprost z tego jak wyprowadziliśmy mamy, że $AB = \mathrm{I}$. Aby sprawdzić, czy $B$ jest rzeczywiście macierzą odwrotną do $A$, musimy jeszcze sprawdzić czy $BA = \mathrm{I}$ (tak jest).*

Widzimy, że powyższa metoda jest w zasadzie niestosowalna. Ale spróbujmy zobaczyć troszkę bardziej inteligentnie jak ona działa. Zobaczmy, że (5.2) można potraktować jako dwa równania:
$$\begin{cases} b_{11}+b_{21} = 1 \\ b_{11} = 0 \end{cases} \text{oraz} \begin{cases} b_{12}+b_{22} = 0 \\ b_{12} = 1 \end{cases}$$
Macierze rozszerzone dla powyższych układów są dane
$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 0 & 0 \end{array}\right] \text{oraz} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 0 & 1 \end{array}\right].$$
Rozwiążmy więc układ pierwszy (sprowadzając do postaci schodkowej zredukowanej). W tym celu wystarczy dokonać operacji $w_2 -= w_1$; $w_2 *= -1$, $w_1 -= w_2$ aby dostać
$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right],$$
czyli $b_{11} = 0$, $b_{21} = 1$. Analogicznie rozwiązujemy drugi układ (proszę zauważyć, że dokonujemy dokładnie tych samych operacji elementarnych co wcześniej) za pomocą $w_2 -= w_1$; $w_2 *= -1$, $w_1 -= w_2$
$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & -1 \end{array}\right],$$
czyli $b_{12} = 1$, $b_{22} = -1$. Skoro dokonywaliśmy dokładnie tych samych operacji, to po co mamy to powtarzać? Możemy zamiast tego utworzyć taką „dużą” rozszerzoną macierz $[A|I]$ daną przez
$$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right],$$
i spróbować sprowadzić $A$ do postaci schodkowej zredukowanej, równocześnie modyfikując $I$ do macierzy odwrotnej do $A$. Widzimy, że stosując tą metodę można relatywnie szybko znaleźć macierz odwrotną. Pokazuje to poniższe twierdzenie.

**Twierdzenie 5.1.** *Niech $A$ będzie daną macierzą kwadratową. Rozpatrujemy macierz rozszerzoną $[A|I]$. Niech $[C|B]$ oznacza macierz uzyskaną w wyniku sprowadzenia $A$ (w macierzy $[A|I]$) do postaci schodkowej zredukowanej za pomocą operacji elementarnych.*

*Wtedy $A$ jest macierzą odwracalną wtedy i tylko wtedy gdy $C = I$. Co więcej, jeżeli $C = I$, to $A^{-1} = B$.*

**Zadanie 5.4.** *Proszę znaleźć (o ile to możliwe) macierze odwrotne do następujących macierzy* $[1]$, $[1,2;2,1]$, $[1,2,3;2,0,1]$, $[1,0,0;0,1,0;0,0,2]$, $[1,0,0,1;1,0,0,2;0,1,0,0;0,0,1,0]$ *(stosuję zapis macierzy jak w Scilabie).*

# Rozdział 6

# Wyznacznik macierzy

## 6.1 Definicja wyznacznika

**Przykład 6.1.** *GLOBUS vs Mapa kwadratowa vs Mapa eliptyczna vs Mapa pocięta (siatka).*

*Ważna jest umiejętność odzyskania oryginalengo pola, jeżeli znamy na mapie kwadratowej i wzór przekształcenia. Lokalnie (analiza) każde rozsądne odwzorowanie można przybliżyć za pomocą odwzorowania liniowego, co znaczy, że aby móc policzyć, jak się zmieniło pole, musimy wiedzieć, jak odwzorowania liniowe zmieniają pole.*

Rozpatrzmy wektory $v, w \in \mathbb{R}^2$ (zapisane w postaci kolumnowej). Jeżeli ustawimy je obok siebie, to powstaje „macierz" $2 \times 2$:

$$[v, w] = \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix}.$$

**Definicja 6.1.** Zobaczmy teraz jak wyraża się pole równoległoboku rozpiętego na wektorach $v$ i $w$. Załóżmy dla prostoty, że wektory są w I ćwiartce, mają niezerowe współrzędne, i że $w$ leży „nad" wektorem $v$ (tzn. $v_1, v_2, w_1, w_2 \geqslant 0$ oraz $w_2/w_1 \geqslant v_2/v_1$). Wtedy po łatwych wyliczeniach („zanurzamy" ten równoległobok w prostokącie o bokach $v_1 + w_1$ i $v_2 + w_2$) dostajemy, że pole równoległoboku wynosi $v_1w_2 - w_1v_2$.

Można pokazać, że (niezależnie od tego gdzie te wektory są położone), pole równoległoboku rozpiętego na nich wynosi $|v_1w_2 - w_1v_2|$). Tak doszliśmy do definicji *wyznacznika* (z angielskiego determinant) dla takiej macierzy $2 \times 2$:

$$\begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} = v_1w_2 - w_1v_2.$$

Widzimy więc, że moduł wyznacznika daje nam pole. Okazuje się, znak mówi czy macierz zmienia orientację, czyli czy lewy but pod wpływem działania macierzy zmienia się na prawy.

Postaramy się teraz wyprowadzić definicję wyznacznika w ogólnym przypadku. Dla macierzy $A = [a_{ij}] \in \mathbb{R}^{2\times 2}$ przez

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \det A$$

rozumiem „zorientowane pole" równoległoboku rozpiętego na wektorach danych przez pierwszą i drugą kolumnę $A$ (nie precyzuję tego dokładniej, ale chodzi o to, że potencjalnie „zorientowane pole" może być ujemne, aby odzyskać normalne pole trzeba wziąć moduł).

**Uwaga 6.1.** *Spróbujmy wypisać jakie własności chcielibyśmy aby to zorientowane pole miało:*

- *pole równoległoboku rozpiętego na wektorach jednostkowych układu współrzędnych (czyli inaczej mówiąc jednostkowego kwadratu) jest równe* 1*, czyli*

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1; \tag{6.1}$$

- *pole równoległoboku o podstawie $v$ i drugim boku równym $w_1+w_2$ jest równy sumie pól równoległoboków o podstawie $v$ i boku $w_1$ oraz $w_2$:*

$$\det[v,(w_1+w_2)] = \det[v,w_1] + \det[v,w_2]; \tag{6.2}$$

  *Powyższy fakt łatwo widać, jak za podstawę przyjmiemy $v$ i oznaczymy przez $h_1$ ($h_2$) wysokość równoległoboku rozpiętego na $[v,w_1]$ (odp. $[v,w_2]$). Wtedy z rysunku (był na wykładzie) widać, że wysokość równoległoboku rozpiętego na $[v,w_1+w_2]$ to $h_1+h_2$. Oraz analogicznie*

$$\det[(v_1+v_2),w] = \det[v_1,w] + \det[v_2,w]; \tag{6.3}$$

- *przemnożenie boku przez $\alpha \in \mathbb{R}$ przemnaża pole przez $\alpha$*

$$\det[(\alpha v),w] = \alpha \det[v,w] = \det[v,(\alpha w)]; \tag{6.4}$$

- *pole równoległoboku rozpiętego na $v$ i $v$ (to jest de facto patyczek) jest równe zero:*

$$\det[v,v] = 0. \tag{6.5}$$

*Zauważmy, że nie wiemy, czy da się zdefiniować funkcję, która takie warunki w rzeczywistości spełnia (potencjalnie mogłoby się okazać, że nie). Można pokazać, że tak jest, natomiast naszym celem będzie założenie, że taka funkcja istnieje, i będziemy starali się ją wyliczać.*

Oczywiście, bezpośrednio z powyższego dostajemy, że wyznacznik macierzy diagonalnej jest równy iloczynowi wyrazów na przekątnej:

$$\det \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} = a_{11}a_{22}.$$

Koniec Wykładu 8

Korzystając z powyższych wzorów dostajemy, że

$$0 = \det[(v+w),(v+w)] = \det[v,v] + \det[v,w] + \det[w,v] + \det[w,w],$$

czyli zamiana kolejności kolumn zmienia znak na przeciwny

$$\det[w,v] = -\det[v,w] \text{ (antysymetryczność)}. \tag{6.6}$$

Jako konsekwencję otrzymujemy, że przy dodaniu do jednej kolumny drugiej przemnożonej przez jakąś stałą, wartość wyznacznika się nie zmienia:

$$\det[v+\alpha w,w] = \det[v,w] + \alpha\det[w,w] = \det[v,w]. \tag{6.7}$$

**Wniosek 6.1.** *Jeżeli $i$-ta kolumna jest zerowa, to wyznacznik mamy równy zero bo*

$$0 = 0 \det A \overset{k_i * = 0}{=} \det A.$$

*W konsekwencji, jeżeli dwie kolumny się powtarzają, też mamy zero (można od jednej odjąć drugą).*

**Wniosek 6.2.** *Dla macierzy górnie (dolnie) trójkątnej:*

$$\det \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = a_{11}a_{22}.$$

*Dowód.* Robimy dla górnie (dualnie można dla dolnie). Jeżeli $a_{11} = 0$, wtedy jest kolumna zerowa, czyli mamy zero. W przeciwnym razie korzystamy z

$$\det \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} \overset{k2 - = \frac{a_{12}}{a_{11}} k_1}{=} \det \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} = a_{11} a_{22}.$$

□

**Stwierdzenie 6.1.** *Mamy*

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11} a_{22} - a_{21} a_{12}.$$

*Dowód.* Mamy

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \det \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} + \det \begin{bmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Teraz pierwszy jest górnie przekątniowy, a w drugiej po zamianie kolumn dostajemy dolnie przekątniową, i w konsekwencji otrzymujemy $a_{11} a_{22} - a_{21} a_{12}$. □

**Przykład 6.2.** *Policzmy pole równoległoboku rozpiętego na wektorach*

$$v := \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ w = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

*Możemy skorzystać, z wzoru powyżej, ale postaramy się zastosować inną strategię, a mianowicie wersję metody eliminacji Gaussa (z tym, że uwaga, nie pracujemy na wierszach lecz na kolumnach).*

*Mamy*

$$\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} \overset{k1+}{=}$$

*Otrzymaliśmy więc, że pole tego równoległoboku wynosi* 5.

**Zadanie 6.1.** *Proszę analogicznie policzyć pole równoległoboku rozpiętego na wektorach* $v := \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *oraz* $w = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$.

## 6.2 Wyznacznik w $\mathbb{R}^3$

Widzimy, że liczenie wyznacznika dla macierzy $2 \times 2$ wygląda tak: przekątną ↘ bierzemy z plusem i przekątną ↗ z minusem. Okaże się, że podobny wzór zachodzi dla macierzy $3 \times 3$ (ale już nie dla macierzy wyższych rzędów!).

Musimy najpierw uogólnić sytuację z $\mathbb{R}^2$ na $\mathbb{R}^N$: Przyjmuję następującą konwencję:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, e_N = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Dla $N$ wektorów (kolumnowych) $v_1, \ldots, v_N \in \mathbb{R}^N$, przez

$$\det[v_1, \ldots, v_N]$$

rozumiem „zorientowaną $N$-wymiarową objętość" równoległościanu rozpiętego na tych wektorach (nie precyzuję dokładniej co przez to rozumiemy), ale chodzi o to, że „zorientowana objętość" może być ze znakiem $-$, aby odzyskać klasyczną objętość trzeba wziąć moduł). Spróbujmy wypisać jakie własności chcielibyśmy aby to zorientowane pole miało:

- objętość $N$-wymiarowej kostki rozpiętej na wektorach $e_1, \ldots, e_N$ jest równe 1, czyli
$$\det I = \det[e_1, \ldots, e_N] = 1; \tag{6.8}$$

- mamy rozdzielność mnożenia względem dodawania:
$$\begin{aligned}&\det[v_1, \ldots, v_{k-1}, (w_k + w'_k), v_{k+1}, \ldots, v_n] \\ &= \det[v_1, \ldots, v_{k-1}, w_k, v_{k+1}, \ldots, v_n] + \det[v_1, \ldots, v_{k-1}, w'_k, v_{k+1}, \ldots, v_n];\end{aligned} \tag{6.9}$$

- przemnożenie boku przez $\alpha \in \mathbb{R}$ przemnaża objętość przez $\alpha$
$$\det[v_1, \ldots, v_{k-1}, (\alpha v_k), v_{k+1}, \ldots, v_n] = \alpha \det[v_1, \ldots, v_{k-1}, v_k, v_{k+1}, \ldots, v_n]; \tag{6.10}$$

- jeżeli dwa boki się powtarzają, to objętość się zeruje:
$$\det[v_1, \ldots, w, \ldots, w, \ldots, v_N] = 0. \tag{6.11}$$

Korzystając z powyższych wzorów dostajemy, że zmiana kolejności dwóch dowolnych elementów w iloczynie powoduje zmianę znaku
$$\det[\ldots, v, \ldots, w, \ldots] = -\det[\ldots, w, \ldots, v, \ldots] \text{ skośnie-symetryczność} \tag{6.12}$$
(można spotkać się z nazwą antysymetryczne).

Postaramy się teraz policzyć wyznacznik macierzy $3 \times 3$:

**Stwierdzenie 6.2.** *Rozpatrzmy macierz*
$$A := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$
*Aby policzyć wyznacznik tej macierzy postępujemy następująco: dopisujemy dwa pierwsze wiersze u dołu*
$$\begin{array}{c} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ \begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{matrix} \end{array},$$
*(analogicznie można dwie pierwsze kolumny po prawej) i liczymy z plusem przekątne $\searrow$ a z minusem $\nearrow$:*
$$\det A = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}.$$

*Dowód.* Wyliczając powyższy iloczyn, zauważmy, że czynnik w którym się dwa razy powtarza któryś $e_k$ (dla $k = 1, 2, 3$) zeruje się, w konsekwencji więc zostają jedynie czynniki w którym występują (w pewnej kolejności) wszystkie czynniki $e_1, e_2, e_3$. Każdy z tych czynników możemy sprowadzić przy pomocy pewnej liczby transpozycji (zamiany ze sobą dwóch czynników) do wyznacznika kostki jednostkowej $\det[e_1, e_2, e_3]$ (pamiętając oczywiście o tym, że każda wymiana, na podstawie antysymetryczności, zmienia nam znak). Dla przykładu mamy
$$\det[e_2, e_3, e_1] = [\text{zamieniamy } e_2 \text{ i } e_1 \text{ miejscami}] = (-1)\det[e_1, e_3, e_2]$$
$$= [\text{zamieniamy } e_3 \text{ i } e_2 \text{ miejscami} = (-1)^2 \det[e_1, e_2, e_3].$$

Korzystając z powyższych obserwacji liczymy:

$$\begin{aligned}\det A = \det[(a_{11}e_1 + a_{21}e_2 + a_{31}e_3), (a_{12}e_1 + a_{22}e_2 + a_{32}e_3), (a_{13}e_1 + a_{23}e_2 + a_{33}e_3)]\\
= \quad \det[a_{11}e_1, a_{22}e_2, a_{33}e_3] + \det[a_{11}e_1, a_{32}e_3, a_{23}e_2]\\
+\det[a_{21}e_2, a_{12}e_1, a_{33}e_3] + \det[a_{21}e_2, a_{32}e_3, a_{13}e_1]\\
+\det[a_{31}e_3, a_{12}e_1, a_{23}e_2] + \det[a_{31}e_3, a_{22}e_2, a_{13}e_1]\\
= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}.\end{aligned}$$

□

## 6.3 Wyznacznik w $\mathbb{R}^N$: metoda eliminacji Gaussa

Korzystamy z faktu, że za pomocą operacji elementarnych na wierszach zawsze można dojść do postaci schodkowej (u dołu schodów zera, i bezpośrednio nad schodkami zera). Analogicznie/dualnie, za pomocą operacji elementarnych na kolumnach można dojść do schodkowej dualnej, w której na „prawo” schodów zera, i na lewo od schodków zera.

**Twierdzenie 6.1.** *Wyznacznik dolnie/górnie-trójkątnej to iloczyn wyrażeń na przekątnej.*

*Dowód.* Dowód indukcyjny, robimy sprowadzenie do postaci dolnie trójkątnej. Jeżeli $a_{11} \neq 0$, to usuwamy w tym wierszu wszystkie na prawo, i jeżeli potem po przejściu do postaci schodkowej mamy wszędzie na głównej przekątnej niezerowe, to wtedy możemy do diagonalnej. Jeżeli nie, to na końcu kolumna ostatnia jest zerowa. Jeżeli $a_{11}$ jest zerowe, to wtedy po przejściu do postaci schodkowej mamy znowu w ostatniej kolumnie zero, czyli dostajemy zero. □

Jeżeli chcemy efektywnie liczyć wyznacznik najlepiej zastosować metodę eliminacji Gaussa (względem kolumn). To znaczy, przy dodaniu/odjęciu od innej wyznacznik się nie zmienia, natomiast przy zamianie kolumn zmieniamy znak na przeciwny. I to nam wystarcza, by szybko sprowadzić macierz do postaci dolnietrójkątnej, i w konsekwencji wtedy wystarczy już przemnożyć wyrażenia na przekątnej.

## 6.4 Wyznacznik w $\mathbb{R}^N$: permutacje

Teraz postaramy się wyliczyć wzór (jest istotny z punktu widzenia teorii, natomiast dla „konkretnej” macierzy nie należy go stosować bo ma złożoność silnii) na wyznacznik ogólnej macierzy kwadratowej $A \in M_{NN}(\mathbb{R})$:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \dots & a_{NN} \end{bmatrix}.$$

Zgodnie z naszym podstawowym wzorem mamy:

$$\det A = \det[(a_{11}e_1 + \dots a_{N1}e_N), \dots, (a_{1N}e_1 + \dots + a_{NN}e_N)]. \tag{6.13}$$

Podobnie jak poprzednio, przy wyliczaniu tego iloczynu wyzerują się składniki w których powtarza się jakiś czynnik dwukrotnie. Aby to opisać matematycznie przyda się nam pojęcie permutacji.

**Definicja 6.2.** Przez *permutację* rozumiem dowolną bijekcję zbioru $N$-elementowego w siebie (standardowo, jeżeli nie jest wyraźnie napisane inaczej bierzemy pod uwagę zbiór $\{1, \dots, N\}$). Zbiór wszystkich permutacji zbioru $N$-elementowego oznaczamy przez $S_N$.

Łatwo widać, że ilość permutacji zbioru $N$-elementowego wynosi $N!$. Parę uwag:

- permutacje oznaczamy często literką $\sigma$ (sigma);
- niech $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$; wtedy $\sigma$ jest permutacją wtw. gdy $\sigma$ jest iniekcją wtw. gdy $\sigma$ jest surjekcją;
- permutacje zazwyczaj zapisujemy w postaci
$$\begin{pmatrix} 1 & \dots & N \\ \sigma(1) & \dots & \sigma(N) \end{pmatrix},$$
tak więc
$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$
oznacza permutację, taką, że $\sigma(1) = 3$, $\sigma(2) = 1$, $\sigma(3) = 2$;
- rozważa się jeszcze *cykle*, czyli dla przykładu (3412) oznacza taką permutację, że $3 \to 4$, $4 \to 1$, $1 \to 2$, $2 \to 3$, szczególnym przypadkiem cyklu jest cykl dwuelementowy (*transpozycja).*

Wróćmy teraz do naszego wzoru (6.13). Na potrzeby poniższego rozumowania, przyjmijmy, że $O_N$ oznacza wszystkie odwzorowania zbioru $\{1, \dots, N\}$ w siebie. Wtedy

$$\begin{aligned} \det A &= (a_{11}e_1 + \dots a_{N1}e_N) \wedge \dots \wedge (a_{1N}e_1 + \dots + a_{NN}e_N) \\ &= \sum_{p \in O_N} (a_{p(1)1}e_{p(1)}) \wedge \dots \wedge (a_{p(N)N}e_{p(N)}) \\ &= \sum_{p \in O_N} (a_{p(1)1} \dots a_{p(N)N}) (e_{p(1)} \wedge \dots \wedge e_{p(N)}). \end{aligned}$$

Na podstawie naszych założeń wiemy, że jeśli w $\det[e_{p(1)}, \dots, e_{p(N)}]$ któryś czynnik powtarza się dwukrotnie, to ten iloczyn się zeruje. Ale tak będzie, w.t.w. gdy $p$ nie jest różnowartościowe. W konsekwencji możemy się zawęzić w naszej sumie do tych odwzorowań które są różnowartościowe, czyli do permutacji:

$$\det A = \sum_{p \in S_N} (a_{p(1)1} \dots a_{p(N)N}) \det[e_{p(1)}, \dots, e_{p(N)}].$$

Jedyne co nam teraz zostało, to pozbyć się czynnika $e_{p(1)} \wedge \dots \wedge e_{p(N)}$. Znowu korzystając z założeń wiemy, że ten iloczyn można za pomocą pewnej ilości transpozycji sprowadzić do $\pm(e_1 \wedge \dots \wedge e_N)$, przy czym znak będzie + jeżeli ilość tych zamian jest parzysta (wtedy mówimy o permutacji parzystej), a minus jeżeli nieparzysta (wtedy mówimy o permutacji nieparzystej). Oznaczmy dla danej permutacji $p$ ten znak[1] przez $\mathrm{sign}(p)$. W konsekwencji dostajemy

**Twierdzenie 6.2.** *Wyznacznik macierzy $A = [a_{ij}] \in M_{NN}(\mathbb{R})$ wyraża się wzorem*

$$\det A := \sum_{p \in S_N} \mathrm{sign}(p) \cdot a_{p(1)1} \dots a_{p(N)N}.$$

Poniższe twierdzenie (nie robię dowodu) pokazuje, że znak permutacji jest dobrze określonym pojęciem:

**Twierdzenie 6.3.** *Zachodzą następujące fakty:*

- *każdą permutację można przedstawić w postaci iloczynu transpozycji, tzn. dla każdej permutacji $\sigma$ istnieją transpozycje $\tau_1, \dots, \tau_r$ takie, że*
$$\sigma = \tau_1 \cdot \ldots \cdot \tau_r;$$

[1] proszę zauważyć, że ja nie udowodniłem, że pojęcie znaku permutacji jest dobrze zdefiniowane, sformułuję odpowiednie twierdzenia później

- *jeżeli mamy dwa przedstawienia w powyższym rozkładzie, tzn.*

$$\sigma = \tau_1 \cdot \ldots \cdot \tau_r = \tau_1' \cdot \ldots \cdot \tau_{r'}'$$

  *to* $r \equiv r' \mod 2$ *(czyli parzystość bądź nieparzystość nie zależy od rozkładu);*

- *w konsekwencji wartość* $\mathrm{sign}(\sigma) := (-1)^r$ *jest dobrze zdefiniowana;*

- *znak ma następujące własności:*

$$\mathrm{sign}(\sigma^{-1}) = \mathrm{sign}(\sigma),\ \mathrm{sign}(\sigma\sigma') = \mathrm{sign}(\sigma) \cdot \mathrm{sign}(\sigma').$$

**Zadanie 6.2.** *Niech będzie dany cykl $k$-elementowy $\sigma$. Policz* $\mathrm{sign}(\sigma)$*.*

**Zadanie 6.3.** *(z gwiazdką) Niech będzie dana permutacja $\sigma \in S_n$. Wtedy*

$$\mathrm{sign}(\sigma) = (-1)^{N(\sigma)},$$

*gdzie $N(\sigma)$ oznacza liczbę inwersji w $\sigma$, czyli ilość par $k, l \in \{1, \ldots, n\}$ takich, że $k < l$ i $\sigma(k) > \sigma(l)$ (ta własność jest wygodna do liczenia znaku permutacji na komputerze).*

*Dowód.* Jeżeli tak zdefiniujemy wyznacznik, to to jest dobrze zdefiniowana funkcja, i wystarczy potem się przypatrzyć, czy rzeczywiście znak permutacji jest ok? □

## 6.5 Własności wyznaczników

Przyda się nam dla badania macierzy pojęcie macierzy transponowanej: jeżeli mamy macierz $A = [a_{ij}]_{ij} \in M_{mn}(\mathbb{R})$, to przez $A^T$, *macierz transponowaną do $A$,* rozumiem macierz $A^T := [a_{ij}]_{ji} \in M_{nm}$, tzn. jeśli

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix},$$

to

$$A^T = \begin{bmatrix} a_{11} & \ldots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \ldots & a_{mn} \end{bmatrix}.$$

Okazuje się, że operacja transponowania macierzy kwadratowej nie zmienia wyznacznika:

**Stwierdzenie 6.3.** *Mamy*

$$\det A = \det A^T.$$

*Dowód.* Niech $A = [a_{ij}] \in M_{nn}(\mathbb{R})$. Oznaczmy przez $a_{ij}^T := a_{ji}$ (wtedy $A^T = [a_{ij}^T]$). Wtedy

$$\begin{aligned} \det A &= \textstyle\sum_{p \in S_n} \mathrm{sign}(p) \cdot a_{p(1)1} \ldots a_{p(n)n} \\ &= \textstyle\sum_{p \in S_n} \mathrm{sign}(p^{-1}) \cdot a_{1p^{-1}(1)} \ldots a_{np^{-1}(n)} \\ &= \textstyle\sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{1\sigma(1)} \ldots a_{n\sigma(n)} \\ &= \textstyle\sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{\sigma(1)1}^T \ldots a_{\sigma(n)n}^T = \det A^T. \end{aligned}$$

□

Jako wniosek otrzymujemy, że możemy tak samo pracować na wierszach jak i na kolumnach. Teraz postaramy się pokazać pewne ważne własności wyznaczników. Zaczniemy od przypomnienia:

- zamiana kolumn → zmienia znak;
- dodanie kolumny innej przemnożonej przez skalar → nie zmienia;
- jeżeli dwie kolumny są równe → zero;
- przemnożenie kolumny przez skalar → przemnaża wyznacznik przez skalar;

**Stwierdzenie 6.4.** *Jeżeli macierz jest górnie (dolnie) trójkątna, to wyznacznik jest równy iloczynowi współczynników na przekątnej głównej.*

*Dowód.* Zrobimy dowód dla górnie trójkątnej. Ponieważ $\det A = \sum_{\sigma \in S_n} \mathrm{sign}\sigma a_{\sigma(1)1} \dots a_{\mathrm{sign}(n)n}$, to aby czynnik był niezerowy musimy mieć $\sigma(i) \leqslant i$ dla każdego $i$ (czyli $\sigma(1) \leqslant 1$, czyli $\sigma(1) = 1$, analogicznie $\sigma(2) \leqslant 2$, ale ponieważ nie może być jeden bo już było, dostajemy $\sigma(2) = 2$, itd.). W konsekwencji $\sigma(i) = i$. □

Korzystając z powyższego stwierdzenia i uwagi wcześniej liczy się zazwyczaj wyznacznik dla macierzy więcej niż $4 \times 4$, a mianowicie sprowadza się do prostszej postaci (możemy pracować zarówno na kolumnach jak i na wierszach).

**Zadanie 6.4.** *Jeżeli macierz jest blokowa postaci*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

*gdzie $A_{ik}$ to macierze, oraz $A_{12} = 0$ lub $A_{21} = 0$, to wyznacznik $A$ jest równy iloczynowi wyznaczników $A_{11}$ i $A_{22}$.*

*Wsk.: (rozpatruję przypadek gdy $A_{12} = 0$) niech $K$ oznacza ilość wierszy macierzy kwadratowej $A_{11}$; proszę zauważyć, że dla dowolnej permutacji $\sigma$, jeżeli $\sigma(k) > K$ dla pewnego $k \in \{1, \dots, K\}$, to $a_{\sigma(k)k} = 0$; oznacza to, że wystarczy rozpatrywać tylko te permutacje dla których $\sigma(k) \in \{1, \dots, K\}$ dla $k \in \{1, \dots, K\}$.*

Można pokazać, że iloczyn wyznaczników to wyznacznik iloczynu.

**Twierdzenie 6.4.** *Niech $A, B \in M_{NN}(\mathbb{R})$. Wtedy*

$$\det(AB) = \det A \cdot \det B.$$

## 6.6 Rozwinięcie Laplace'a

Naszym głównym celem będzie teraz wyznaczenie wzoru Laplace'a, który pozwala liczyć wyznacznik rekurencyjnie względem wiersza lub kolumny. Dla prostoty zapisu pokażę wyprowadzenie dla macierzy $3 \times 3$.

Rozpatrzmy macierz

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Naszym celem jest wyliczenie wyznacznika tej macierzy, to jest takiej liczby $\det A$, że

$$(a_{11}e_1 + a_{21}e_2 + a_{31}e_3) \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) = \det A \cdot (e_1 \wedge e_2 \wedge e_3).$$

Lewa strona powyższej równości wynosi:

$$L = \begin{array}{l} a_{11} \cdot e_1 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \\ +a_{21}e_2 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \\ +a_{31} \cdot e_3(a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \end{array}.$$

Zajmijmy się pierwszym składnikiem powyższej sumy. Z własności iloczynu zewnętrznego wiemy, że możemy pominąć te czynniki w których dwukrotnie występuje $e_1$, czyli

$$a_{11} \cdot e_1 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) = a_{11} \cdot e_1 \wedge (a_{22}e_2 + a_{32}e_3) \wedge (a_{23}e_2 + a_{33}e_3).$$

Ale z definicji wyznacznika, wiemy, że

$$(a_{22}e_2 + a_{32}e_3) \wedge (a_{23}e_2 + a_{33}e_3) = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \cdot e_2 \wedge e_3.$$

W konsekwencji dostajemy

$$\begin{aligned} &a_{11}e_1 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \\ &= a_{11} \cdot \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \cdot e_1 \wedge e_2 \wedge e_3 \end{aligned}$$

Analogicznie

$$\begin{aligned} &a_{21}e_2 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \\ &= a_{21}e_2 \wedge \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} (e_1 \wedge e_3) \\ &= (-1)^1 a_{21} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} \cdot (e_1 \wedge e_2 \wedge e_3), \end{aligned}$$

oraz

$$\begin{aligned} &a_{31}e_3 \wedge (a_{12}e_1 + a_{22}e_2 + a_{32}e_3) \wedge (a_{13}e_1 + a_{23}e_2 + a_{33}e_3) \\ &= a_{31}e_3 \wedge \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \cdot (e_1 \wedge e_2) \\ &= (-1)^2 a_{31} \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \cdot (e_1 \wedge e_2 \wedge e_3). \end{aligned}$$

W konsekwencji otrzymaliśmy rozwinięcie macierzy $A$ względem pierwszej kolumny:

$$\det A = (-1)^{1+1} a_{11} \cdot \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{2+1} a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{3+1} a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

Dla skrócenia zapisu niech $A_{ij}$ oznacza macierz powstałą z macierzy $A$ przez wykreślenie $i$-tego wiersza i $j$-tej kolumny (wyznacznik $A_{ij}$ nazywamy minorem i często oznaczamy $M_{ij}$). Tak więc dostaliśmy

$$\det A = (-1)^{1+1} a_{11} \det A_{11} + (-1)^{2+1} a_{21} \det A_{21} + (-1)^{3+1} a_{31} \det A_{31}.$$

Stosując analogiczne rozumowanie względem $k$-tej kolumny dla ogólnej macierzy $N \times N$, rozwijając dostajemy tak zwane rozwinięcie Laplace'a (drugi wzór jest konsekwencją pierwszego oraz tego, że transponowanie nie zmienia wartości wyznacznika):

**Twierdzenie 6.5** (rozwinięcie Laplace'a)**.** *Niech $A \in M_{NN}(\mathbb{R})$.*
*a) [względem $j$-tej kolumny] Niech $j \in \{1, \ldots, N\}$ będzie dowolne. Wtedy*

$$\det A = \sum_{i=1}^{N} (-1)^{i+j} a_{ij} \det A_{ij}. \tag{6.14}$$

*b) [względem $i$-tego wiersza] Niech $i \in \{1, \ldots, N\}$ będzie dowolne. Wtedy*

$$\det A = \sum_{j=1}^{N} (-1)^{i+j} a_{ij} \det A_{ij}. \tag{6.15}$$

**Zadanie 6.5.** *Niech* $f_n = \det A$, *gdzie gdzie* $A = [a_{ij}] \in M_{nn}(\mathbb{R})$, *a współczynniki* $a_{ij}$ *zadane są wzorami*

$$a_{i(i+1)} = 3,\ a_{(i+1)i} = 5 \text{ dla } i = 1, \dots, n-1,$$

*oraz* $a_{ij} = 0$ *w pozostałych przypadkach. Proszę wyliczyć* $f_n$ *za pomocą jawnego wzoru.*

**Zadanie 6.6.** *Niech* $A = [a_{ij}]$ *będzie macierzą* $n \times n$, *taką, że* $a_{ii} = 0$ *dla* $i = 2, \dots, n$ *oraz* $a_{ij} = 1$ *w pozostałym przypadku. Proszę wyliczyć wyznacznik A. Wsk.: proszę odjąć od pierwszego wiersza drugi (wyznacznik się nie zmieni), i następnie zastosować rozwinięcie Laplace'a.*

**Zadanie 6.7.** *Niech* $f_n = \det[a_{ij}]$, *gdzie* $[a_{ij}] \in M_{nn}(\mathbb{R})$ *dane jest*

$$a_{ii} = 2 \text{ dla } i = 1, \dots, n,\ a_{i(i+1)} = 1,\ a_{(i+1)i} = 1 \text{ dla } i = 1, \dots, n-1,$$

*oraz* $a_{ij} = 0$ *o ile* $|i - j| > 1$. *Proszę znaleźć jaką formułę rekurencyjną spełnia* $f_n$ *a następnie ją rozwiązać, czyli wyliczyć* $f_n$ *za pomocą jawnego wzoru.*

Pokażemy teraz, że z rozwinięcia Laplace'a łatwo można wyliczyć wzór na macierz odwrotną. Rozpatrzmy macierz (nazywamy ją macierzą dołączoną do macierzy $A$):

$$B = [(-1)^{i+j} \det A_{ij}]^T = \begin{bmatrix} +\det A_{11} & -\det A_{21} & +\det A_{31} & \dots \\ -\det A_{12} & +\det A_{22} & - & \\ +\det A_{13} & -\det A_{23} & + & \\ - & + & - & \\ \vdots & & & \end{bmatrix}$$

Naszym celem będzie obserwacja, że

$$AB = BA = (\det A) \cdot \mathrm{I}. \tag{6.16}$$

Na podstawie wzorów (6.14) i (6.15) otrzymujemy trywialnie, że zarówno $AB$ jak i $BA$ na przekątnej głównej mają wartość $\det A$. Dowód powyższej równości kończy

**Zadanie 6.8.** *Proszę pokazać, że* $AB$ *i* $BA$ *są macierzami diagonalnymi. Wsk.: niech* $C = BA$; *spróbujmy sprawdzić czy* $c_{12} = 0$; *oczywiście*

$$c_{12} = \det(A_{11})a_{12} - \det(A_{21})a_{22} + \det(A_{31})a_{32} \pm \dots,$$

*proszę zauważyć, że powyższy wzór możemy otrzymać stosując wzór Laplace'a względem pierwszej kolumny na wyznacznik macierzy powstałej z A przez zastąpienie pierwszej kolumny drugą (a jak skądinąd wiemy ten wyznacznik jest równy zero, bo ta macierz ma dwie kolumny równe).*

Dostajemy w konsekwencji:

**Twierdzenie 6.6.** *Niech* $A \in M_{NN}(\mathbb{R})$. *Wtedy NWSR:*

- *macierz A jest macierzą odwracalną;*
- $\det A \neq 0$*;*
- *istnieje* $B \in M_{NN}(\mathbb{R})$, *takie, że* $AB = \mathrm{I}$ *(i wtedy* $A^{-1} = B$*);*
- *istnieje* $B \in M_{NN}(\mathbb{R})$, *takie, że* $BA = \mathrm{I}$ *(i wtedy* $A^{-1} = B$*).*

*Co więcej, jeżeli A jest odwracalna, to*

$$A^{-1} = \frac{1}{\det A}[(-1)^{i+j} \det A_{ij}]^T.$$

*Dowód.* Jeżeli $\det A \neq 0$, to wzór na macierz odwrotną, to oczywiście (6.16).

Niech teraz $\det A = 0$. Pokażemy, że $A$ nie jest odwracalna. Dla dowodu nie wprost załóżmy, że $A$ jest odwracalna, czyli, że istnieje $B$ takie, że $AB = \mathrm{I}$. Wtedy

$$1 = \det \mathrm{I} = \det(AB) = \det A \cdot \det B = 0 \cdot \det B = 0,$$

sprzeczność.

Załóżmy teraz, że znaleźliśmy takie $B$, że $AB = \mathrm{I}$. Pokażę, że wtedy $A$ jest odwracalne i, że $B = A^{-1}$. Skoro znaleźliśmy takie $B$ to w konsekwencji stosując powyższe rozumowanie otrzymujemy, że $\det A \neq 0$, czyli $A$ jest odwracalna. Teraz mnożąc obie strony równości $AB = \mathrm{I}$ lewostronnie przez $A^{-1}$ dostajemy

$$\begin{aligned}&A^{-1}(AB) = A^{-1}(\mathrm{I}),\\ &(A^{-1}A)B = A^{-1},\\ &B = A^{-1}.\end{aligned}$$

Pozostałe implikacje są analogiczne. □

Ważnym zastosowaniem jest znalezienie wielomianu możliwie najniższego stopnia przechodzącego przez punkty $(x_0, y_0), \ldots, (x_n, y_n)$ (zakładamy, że $x_i$ są parami różne). Pokażę, że istnieje dokładnie jeden wielomian $W$ stopnia co najwyżej $n$ przechodzący przez te punkty, to znaczy taki, że $W(x_i) = y_i$. Niech

$$W(x) = a_0 + a_1 x^1 + \ldots + a_n x^n.$$

Wtedy $W$ musi spełniać

$$W(x_i) = a_0 + a_1 x_i^1 + \ldots + a_{n-1} x_i^{n-1} = y_i \quad \text{dla } i = 0, \ldots, n,$$

co oznacza, że współczynniki $a_k$ spełniają

$$\begin{bmatrix} 1 & x_0^1 & \ldots & x_0^n \\ \vdots & & \vdots & \\ 1 & x_n^1 & \ldots & x_n^n \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$$

Korzystając z następującego faktu:

**Zadanie 6.9.** *(wyznacznik Vandermonda) Zachodzi równość:*

$$\begin{vmatrix} 1 & x_0^1 & \ldots & x_0^n \\ \vdots & & \vdots & \\ 1 & x_n^1 & \ldots & x_n^n \end{vmatrix} = \prod_{i<j} (x_j - x_i).$$

*Wsk.: proszę najpierw odjąć od wierszy* $2 : n+1$ *wiersz pierwszy (wtedy z rozwinięcia Laplace'a sytuacja redukuje się do wyznacznika macierzy* $n \times n$*); a następnie odejmować kolejno od ostatniej kolumny przedostatnią, od przedostatniej przed-przedostatnią, itd; a potem się chwilę zastanowić i skorzystać z założenia indukcyjnego.*

otrzymujemy, że wyznacznik Vandermonda jest niezerowy, czyli macierz jest odwracalna, czyli współczynniki $a_k$ są jednoznacznie wyznaczone przez:

$$\begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 1 & x_0^1 & \ldots & x_0^n \\ \vdots & & \vdots & \\ 1 & x_n^1 & \ldots & x_n^n \end{bmatrix}^{-1} \cdot \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$$

**Zadanie 6.10** (Scilab). *Proszę dla danego ciągu skończonego ciągu par $(x_i, y_i)$ (dane jako macierz o dwóch wierszach, zakładamy, że dane są poprawne, tzn $x_i \neq x_j$ dla $i \neq j$) znaleźć wzór wielomianu najniższego stopnia przechodzącego przez te punkty.*

**Zadanie 6.11.** *Wielkość*

$$((a_1, \ldots, a_n)) := \det(-diag(ones(1, n-1), -1) + diag([a_1, \ldots, a_n]) + diag(ones(1, n-1), 1))$$

*wiąże się z ułamkami łańcuchowymi, tzn*

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_{n-1} + \frac{1}{a_n}}}} = \frac{((a_1, \ldots, a_n))}{((a_2, \ldots, a_n))}.$$

## 6.7 Wzory Cramera

Teraz pokażemy, że z (6.16) wynika wzór na rozwiązanie równania liniowego (Wzory Cramera). Rozpatrzmy równanie

$$Ax = b,$$

gdzie $A$ jest odwracalną macierzą kwadratową $N \times N$. Postaramy się wyliczyć $x = [x_1, \ldots, x_N]^T$. Oczywiście $x = A^{-1}b$, co oznacza, że

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} (-1)^{1+1} \det A_{11} & \ldots & (-1)^{N+1} \det A_{N1} \\ \vdots & & \vdots \\ (-1)^{1+N} \det A_{1N} & \ldots & (-1)^{N+N} \det A_{NN} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

czyli

$$x_k = \frac{1}{\det A}((-1)^{1+k} \det(A_{1k})b_1 + \ldots (-1)^{k+N} \det(A_{Nk})b_N).$$

Zauważmy teraz, że $W_k = (-1)^{1+k} \det(A_{1k})b_1 + \ldots (-1)^{k+N} \det(A_{Nk})b_N$ to po prostu wyznacznik macierzy powstałej z $A$ przez zamienienie $k$-tej kolumy $b$. Oznaczając $W = \det A$, dostajemy wzór na $x_k$:

$$x_k = W_k/W.$$

Powyższy wzór jest szczególnie wygodny jeśli chodzi o rozwiązywanie układów równań liniowych z parametrem.

**Przykład 6.3.** *Rozpatrzmy układ z parametrem $a$:*

$$\begin{cases} ax + y = 1, \\ x - y = 1. \end{cases}$$

*Rozpatrzmy macierz odpowiadającą temu układowi:*

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

*Wyznacznik tej macierzy wynosi $W = -a-1$. Jeżeli ten wyznacznik będzie różny od zera (czyli gdy $-a-1 \neq 0$, czyli $a \neq -1$), to nasz układ będzie miał dokładnie jedno rozwiązanie dane wzorem*

$$x = W_x/W = \frac{1}{-a-1} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}, \; y = W_y/W = \frac{1}{-a-1} \begin{vmatrix} a & 1 \\ 1 & 1 \end{vmatrix}.$$

*Pozostaje jeszcze rozwiązanie w przypadku gdy $a = -1$. Wtedy układ przybiera postać*

$$\begin{cases} -x + y = 1, \\ x - y = 1. \end{cases},$$

*i jak widać nie ma on rozwiązań.*

*GEOMETRYCZNA INTERPRETACJA: dla $a \neq -1$ mamy dwie proste które nie są równoległe, czyli przecinają się w dokładnie jednym punkcie, a dla $a = -1$ mamy dwie proste równoległe które się nie przecinają.*

# Rozdział 7

# Regresja liniowa: podstawy nauczanie maszynowego

**Streszczenie.** Uczenie nienadzorowane (klastrowanie/analiza skupień) i nadzorowane.
Regresja, znamy przykładowe wartości $(x_i, y_i)$, ktoś nam podaje nowy punkt $\bar{x}$ i chcemy przewidzieć wartość $\bar{y}$.
Klasyfikacja następuje wtedy, gdy z góry wiemy, że $y$ może należeć tylko do pewnej skończonej ilości klas, zazwyczaj zakładamy, że to $\pm 1$ (i tak problem wtedy jest wystarczająco skomplikowany). Przykłady klasyfikacji:

- przewidzenie chory/zdrowy
- lek/nie lek
- klasyfikacja obiektów na zdjęciu (klasy)

W bieżącej sekcji zajmiemy się omówieniem podstaw regresji liniowej, w następnych rozdziałach wrócimy jeszcze do klasyfikacji.

## 7.1 Regresja liniowa: postawienie problemu i propozycja rozwiązania

**Historia 7.1** (Metoda Najmniejszych Kwadratów). *W roku 1801 astronomowie zgubili z oczu asteroidę, i chodziło o to by odszukać ją z powrotem na niebie. Gauss stworzył MNK właśnie w celu by ja odszukać, co mu się udało – znalazła się dokładnie tam, gdzie Gauss przewidział, że będzie.*

Dokonaliśmy pomiarów pewnej funkcji:

$$\begin{array}{c|ccc} x_i & 1 & 2 & 4 \\ \hline y_i & 1 & 2 & 3 \end{array}.$$

Podejrzewamy, że dane mogą być dobrze przybliżone za pomocą funkcji liniowej $y = ax + b$. W związku z tym szukamy takich parametrów $a, b$ aby przybliżenie

$$y_i \approx ax_i + b \text{ dla } i = 1, .., n$$

(gdzie w naszym przypadku $n = 3$) było optymalne. Inaczej mówiąc szukamy takich $a, b$ aby

$$\begin{cases} 1a + b \approx 1 \\ 2a + b \approx 2 \\ 4a + b \approx 3 \end{cases}$$

co w zapisie macierzowym możemy przedstawić:

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Jak poprzednio, pojawia się problem co to znaczy optymalne, oraz (jak to już doprecyzujemy) jak to optymalne znaleźć. Precyzyjniej mówiąc potrzebujemy dookreślić jaką funkcję kosztu tu dopasujemy, która będzie nam mówiła ile „kosztuje” nas dany błąd (w zależności od doboru parametrów $a, b$). Naturalnym wydawałoby się posumowanie modułów błędów

$$\text{error}(a, b) = \sum_i |y_i - (ax_i + b)|.$$

Tak się czasami robi, ale takie podejście ma wadę (patrz analogiczna sytuacja dla dyskretyzacji opisana we wcześniejszym rozdziale), że nie da się tych współczynników wyliczyć jawnym wzorem. W związku z tym dokonamy naturalnej modyfikacji, zastępując moduł kwadratem:

$$se(a, b) = \text{squarred error}(a, b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2.$$

**Uwaga 7.1.** *Interpretacja geometryczna: szukamy takiego punktu* $(a, b)$ *na płaszczyźnie, aby jego obraz po odwzorowaniu za pomocą macierzy* $A$ *leżał najbliżej punktu* $(1, 2, 3)$.

I teraz zajmiemy się problemem jak znaleźć $a, b$ które realizują minimum.

Zauważmy, że funkcja $se(a, b)$ po uproszczeniu przyjmuje postać:

$$se(a, b) = \sum_i y_i^2 + a^2 \sum_i x_i^2 + b^2 \sum_i 1 - 2a \sum_i y_i x_i - 2b \sum_i y_i + 2ab \sum_i x_i.$$

I teraz mamy funkcję dwóch zmiennych, która przy ustaleniu każdej daje zwykłą funkcję kwadratową. Możemy teraz skorzystać z faktu, że wiemy, że funkcja kwadratowa $w \to a_0 w^2 + a_1 w + a_2$ przyjmuje minimum dla $w = -a_1/(2a_0)$. W konsekwencji, jeżeli zafiksujemy (to znaczy potraktujemy jako stałą) $b$, to funkcja

$$a \to se(a, b) = (\sum_i x_i^2) a^2 + (-2 \sum_i y_i x_i + 2b \sum_i x_i) a + \text{const}$$

przyjmuje minimum dla $a$ spełniającego warunek

$$a = \frac{\sum_i y_i x_i - b \sum_i x_i}{\sum_i x_i^2}.$$

Analogicznie, po zafiksowaniu $a$ funkcja

$$b \to se(a, b) = (\sum_i 1) b^2 + (-2 \sum_i y_i + 2a \sum_i x_i) b + \text{const}$$

przyjmuje minimum dla $b$ takiego, że

$$b = \frac{\sum_i y_i - a \sum_i x_i}{\sum_i 1}.$$

W konsekwencji otrzymaliśmy układ równań jaki muszą spełnić parametry $a, b$, aby dawały minimum:

$$\begin{cases} a = \frac{\sum_i y_i x_i - b \sum_i x_i}{\sum_i x_i^2}, \\ b = \frac{\sum_i y_i - a \sum_i x_i}{\sum_i 1}. \end{cases}$$

Korzystając z

| | | | | $\sum$ |
|---|---|---|---|---|
| $x_i$ | 1 | 2 | 4 | 7 |
| $y_i$ | 1 | 2 | 3 | 6 |
| $y_i x_i$ | 1 | 4 | 12 | 17 |
| $x_i^2$ | 1 | 4 | 16 | 21 |

otrzymujemy układ

$$\begin{cases} a = \frac{17-7b}{21}, \\ b = \frac{6-7a}{3}, \end{cases}$$

który łatwo można rozwiązać uzyskując $a = 9/14, b = 1/2$.

## 7.2 Lepsza metoda

PRZYPOMNIENIE: będziemy potrzebować następujących faktów dotyczących macierzy transponowanej:

- $(AB)^T = B^T A^T$ (proszę sprawdzić),
- $u^T v = \langle u, v\rangle = \langle v, u\rangle = v^T u$,
- $\|u\|^2 = \langle u, u\rangle$.

Jak widzieliśmy, trzeba było się dosyć namęczyć, by nawet sformułować układ równań jaki nas interesuje. Pokażę inne (ogólniejsze rozwiązanie), które daje gotowy wzór. Chcemy znaleźć $x_0$ które daje najlepsze przybliżenie $Ax = b$. Inaczej mówiąc, nasz problem możemy sformułować (ogólnie) jako problem znalezienia

$$x_0 = \mathrm{argmin}\{\|Ax - b\|^2 : x\}.$$

Jeżeli teraz $x_0$ spełnia powyższe, to znaczy, że

$$\|A(x_0 + h) - b\|^2 \geqslant \|Ax_0 - b\|^2.$$

Po uproszczeniu dostajemy

$$2\langle Ah, Ax_0 - b\rangle + \langle Ah, Ah\rangle \geqslant 0.$$

Czyli

$$\langle h, A^T Ax_0 - A^T b\rangle \geqslant -\frac{1}{2}\langle Ah, Ah\rangle \text{ dla dowolnego } h.$$

Podstawiając $h = \alpha h$ dla $\alpha > 0$ w powyższym wzorze i dzieląc przez $\alpha$ dostajemy

$$\langle h, A^T Ax_0 - A^T b\rangle \geqslant -\frac{\alpha}{2}\langle Ah, Ah\rangle \text{ dla dowolnego } h, \alpha > 0.$$

Przechodząc z $\alpha \to 0$ (czyli biorąc $\alpha$ dowolnie małe), otrzymujemy, że lewa strona jest większa od dowolnie bliskiej zero liczby ujemnej, czyli musi zachodzić

$$\langle h, A^T Ax_0 - A^T b\rangle \geqslant 0 \text{ dla dowolnego } h.$$

Podkładając $h = -(A^T Ax_0 - A^T b)$ dostajemy

$$-\|A^T Ax_0 - A^T b\|^2 \geqslant 0 \text{ czyli } \|A^T Ax_0 - A^T b\| \leqslant 0,$$

co w konsekwencji znaczy, że $x_0$ musi spełniać równanie

$$A^T Ax_0 = A^T b.$$

Można pokazać, że powyższy warunek to tak naprawdę wkw. Proszę zauważyć, że ten warunek jest łatwy w zapamiętaniu, jeżeli chcemy znaleźć najlepsze przybliżenie (nieistniejącego) rozwiązania równania $Ax = b$, wystarczy, że przemnożymy obie strony przez $A^T$ i rozwiążemy powstałe równanie, które zawsze będzie miało rozwiązanie.

Inaczej mówiąc, zachodzi twierdzenie:

**Twierdzenie 7.1.** *Punkt $x_0$ spełnia*

$$x_0 = \operatorname{argmin} \|Ax - b\|^2 : x$$

*wtw. gdy*

$$(A^T A)x_0 = A^T b.$$

**Przykład 7.1.** *Wróćmy do przykładu z poprzedniej sekcji*

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} \approx \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

*Aby znaleźć optymalne przybliżenie, mnożymy obie strony przez $A^T$ dostając równanie*

$$\begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

*czyli rozwiązujemy*

$$\begin{bmatrix} 21 & 7 \\ 7 & 3 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 17 \\ 6 \end{bmatrix}.$$

*Zauważmy, że co nie jest nadmiernie zaskakujące, otrzymaliśmy dokładnie ten sam układ równań co poprzednio (tylko troszkę inaczej zapisany), i w konsekwencji rozwiązaniem jest*

$$a = 9/14, b = 1/2.$$

**Przykład 7.2.** *Załóżmy, że mamy dane ciało fizyczne na którego nie działa żadna siła, i w konsekwencji wiemy, że porusza się ruchem jednostajnym. Udało nam się zmierzyć (oczywiście w pewnym przybliżeniu) $b$ – położenie tego ciała w następujących chwilach czasowych:* $\begin{array}{c|c|c|c|c} t & 1 & 3 & 4 & 7 \\ \hline b & 2 & 4 & 6 & 9 \end{array}$*. Interesuje nas jakie będzie przewidywane położenie tego ciała w chwili $t = 12$.*

*Spróbujmy troszkę podejść abstrakcyjnie: załóżmy, że znamy $b_i$, położenie ciała w chwili $t_i$ (potencjalnie z pewnym błędem wynikającym z dokładności naszych narzędzi pomiarowych), i wiemy, że $b$ jest funkcją liniową, to znaczy wyraża się wzorem $b(t) = x_1 + x_2 t$ dla pewnych $x_1, x_2$. Chcemy rozwiązać (względem $x_1, x_2$) układ równań*

$$x_1 + t_i x_2 = b_i \text{ dla } i = 1, \ldots, n,$$

*co w zapisie macierzowym wyraża się*

$$Ax = b,$$

*gdzie*

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

*Wróćmy więc do naszego początkowego równania, wtedy*

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \\ 1 & 7 \end{bmatrix},\ b = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 9 \end{bmatrix}.$$

*Dostajemy równanie $A^T Ax = A^T b$:*

$$\begin{bmatrix} 4 & 15 \\ 15 & 75 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 21 \\ 101 \end{bmatrix},$$

*które łatwo już możemy rozwiązać otrzymując $x_1 = 4/5$, $x_2 = 89/75$. Czyli dla $t = 12$ przewidujemy, że nasze ciało znajdować się będzie w punkcie $4/5 + 12 \cdot 89/75 = 15\frac{1}{25}$.*

**Zadanie 7.1** (MNK). *Stosując metodę najmniejszych kwadratów proszę znaleźć funkcję kwadratową (a precyzyjniej jej współczynniki $a, b, c$)*

$$y = a + bx + cx^2$$

*która najlepiej przybliża dane*

$$y(0) = 2,\ y(1) = 0,\ y(3) = 0,\ y(4) = 1.$$

## 7.3 Walidacja krzyżowa

Jak się uczyć do egzaminu: załóżmy, że mamy zadania z $k$ poprzednich lat, przerobiliśmy je wszystkie, i umiemy zrobić wszystkie zadania. Czy to znaczy, że rozwiążemy nowy egzamin bardzo dobrze? Otóż niekoniecznie – naszym celem bowiem powinno być nie przygotowanie się do poprzednich egzaminów, lecz do egzaminu którego nie znamy. W związku z tym, dużo bardziej wiarygodne byłoby uczenie się z egzaminów z $k-1$ poprzednich lat, i po nauczeniu się zweryfikowanie naszych umiejętności na pozostałym.

Taki jest cel walidacji krzyżowej – ona ma w szczególności pomóc nam ocenić, czy nasz model się nie przeuczył (overfitting), czyli nie dopasował do danych.

**Dygresja 7.1.** *OVERFITTING: załóżmy, że mamy 100 punktów $x_i$ oraz wartości $y_i$, które mniej więcej układają się wzdłuż prostej. Wtedy na postawie wcześniejszych wyników możemy znaleźć wielomian $W$ stopnia co najwyżej 99, który będzie przechodził przez te punkty, to znaczy*

$$W(x_i) = y_i \text{ dla } i = 0, .., 99.$$

*Weźmy nowy punkt $x_{100}$ i spytajmy się czy $W(x_{100}) \approx y_{100}$? Otóż tak raczej nie będzie, gdyż wielomian stopnia 99 bardzo szybko będzie uciekał do $\pm\infty$ dla dużych $x$, podczas gdy nasza funkcja (jak wiemy z założeń) zachowuje się w przybliżeniu jak funkcja liniowa.*

Walidacja krzyżowa (cross validation). Dzielimy zestaw danych na $n = 5$ lub 10 pakietów. Nasz model uczymy na $n-1$ pakietach, testujemy (względem funkcji kosztu która nas interesuje) na pozostałym, po czym przechodzimy każdy kolejny, i wyniki uśredniamy.

Koniec Wykładu OSTATNIEGO W I SEMESTRZE

# Rozdział 8

# Przestrzenie wektorowe

## 8.1 Typy danych

$\mathbb{R}^N$

muzyka: ciągi z zakresu $[-1, 1]$ o dowolnej liczbie elementów, zawarte w przestrzeni wszystkich ciągów równych zero od pewnego momentu

zdjęcia ustalonej wielkości - funkcje z dziedziny w $[0, 255]^3$ (lub więcej) zawarte w przestrzeni funkcji z dziedziny w $\mathbb{R}$, można interpretować jako zestaw 3 macierzy o tych samych wymiarach

macierze ustalonej wielkości

macierze kwadratowe, symetryczne

wielomiany ustalonego (od góry) stopnia

wielomiany dowolnego stopnia

NLP (natural language processing - przetwarzanie języka naturalnego), w naturalny sposób można stworzyć bag of words (worek słów), dla danego tekstu jest to funkcja ze zbioru słów w ilość wystąpień

## 8.2 Podstawowe definicje

$\mathbb{F} = \mathbb{R}, \mathbb{C}$ (domyślnie).

**Definicja 8.1.** Mówimy, że $X$ jest przestrzenią wektorową (nad ciałem $\mathbb{F}$) jeżeli w $X$ mamy operację dodawania taką, że $(X, +)$ jest grupą oraz działanie mnożenia (przez skalary) $\cdot : \mathbb{F} \times X \to X$ takie, że

- $\alpha \cdot (x + y) = \alpha x + \alpha y$ dla $\alpha \in \mathbb{F}$, $x, y \in X$;
- $(\alpha + \beta) \cdot x = \alpha x + \beta x$ dla $\alpha, \beta \in \mathbb{F}$, $x \in X$;
- $1 \cdot x = x$.

Często elementy ciała nazywamy skalarami, a elementy $X$ wektorami.

**Przykład 8.1.** *Najczęściej spotykane przestrzenie wektorowe to:*

- $\mathbb{F}^n$*;*
- *ciągi nieskończone o współczynnikach z* $\mathbb{F}$*;*
- *ciągi nieskończone o współczynnikach z* $\mathbb{F}$ *które są od pewnego momentu równe zero;*
- *ciągi nieskończone o współczynnikach z* $\mathbb{F}$ *które są od pewnego momentu stałe;*
- *wielomiany o współczynnikach z* $\mathbb{F}$*;*

- *wielomiany stopnia co najwyżej $n$ o współczynnikach z $\mathbb{F}$;*
- *przestrzeń macierzy $M_{m\times n}(\mathbb{F})$;*
- *funkcje ze zbioru $A$ w $\mathbb{F}$.*

**Zadanie 8.1.** *Proszę sprawdzić, że powyższe zbiory z naturalnymi działaniami są przestrzeniami wektorowymi nad ciałem $\mathbb{F}$.*

**Definicja 8.2.** Niech $V \subset \mathbb{K}^N$. Wtedy $V$ nazywamy podprzestrzenią liniową (wektorową) $\mathbb{R}^N$ o ile

- $v + w \in V$ dla dowolnych $v, w \in V$;
- $\alpha v \in V$ dla dowolnych $v \in V$, $\alpha \in \mathbb{K}$.

**Stwierdzenie 8.1.** *Dowolna kombinacja liniowa elementów $V$ należy do $V$, tzn. dla dowolnych $v_1, \ldots, v_n \in V$, $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$*

$$\alpha_1 v_1 + \ldots + \alpha_n v_n \in V.$$

*Dowód.* Dowód indukcyjny ze względu na $n$ (był na wykładzie). □

**Zadanie 8.2.** *Sprawdzić które z poniższych zbiorów są podprzestrzeniami liniowymi: a) $\{(x, y) \in \mathbb{R}^2 : x+y = 0\}$; b) $\{(x, y, z) \in \mathbb{R}^3 : x + y + z = 1\}$; c) $\{(x, y) \in \mathbb{R}^2 : x = 0 \wedge y = 0\}$.*

**Zadanie 8.3.** *Niech $V, W$ będą podprzestrzeniami liniowymi $\mathbb{R}^N$. Proszę pokazać, że $V \cap W$ jest podprzestrzenią liniową.*

**Zadanie 8.4.** *Proszę pokazać przykład dwóch podprzestrzeni liniowych $\mathbb{R}^2$, takich, że ich suma mnogościowa jest podprzestrzenią liniową.*

**Zadanie 8.5.** *Niech $V, W$ będą podprzestrzeniami liniowymi $\mathbb{R}^N$. Proszę pokazać, że $V \cup W$ jest podprzestrzenią liniową wtw. gdy $V \subset W$ lub $W \subset V$.*

**Zadanie 8.6.** *Niech $v_1, \ldots, v_k \in \mathbb{R}^N$. Proszę udowodnić, że $\mathrm{lin}(v_1, \ldots, v_k)$ jest podprzestrzenią liniową $\mathbb{R}^N$.*

## 8.3 Rząd, liniowa niezależność i Tw. Kroneckera-Capellego

W tej sekcji wrócimy jeszcze na chwilkę do równań (motywacja do liniowej niezależności i rzędu).

Rozpatrzmy układ $n$-równań z $k$-niewiadomymi

$$\begin{cases} a_{11}x_1 + \ldots + a_{1k}x_k = b_1, \\ \vdots \\ a_{n1}x_1 + \ldots + a_{nk}x_k = b_n. \end{cases}$$

Interesuje nas czy on ma rozwiązanie, czyli czy istnieją skalary $x_1, \ldots, x_k$ takie, że

$$x_1 \cdot \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} + \ldots + x_k \cdot \begin{bmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Jeżeli przez $v_l$ oznaczymy $l$-ta kolumnę macierzy $A$, to znaczy $v_l = [a_{1l}, \ldots, a_{nl}]^T$, to nasze pytanie sprowadza się do pytania, czy istnieją skalary $x_1, \ldots, x_k$, takie, że

$$x_1 v_1 + \ldots + x_k v_k = b.$$

Tak więc w naturalny sposób otrzymaliśmy następującą definicję:

**Definicja 8.3.** Niech $v_1, \ldots, v_k, b \in \mathbb{R}^n$ będą dane. Jeżeli istnieją skalary $x_1, \ldots, x_k \in \mathbb{R}$, takie, że

$$x_1 v_1 + \ldots + x_k v_k = b.$$

to mówimy, że $b$ jest kombinacją liniową wektorów $v_1, \ldots, v_k$.

**Przykład 8.2.** *Analogicznie (dualnie) mozna spytać się czy jedno z równań da się uzyskać w postaci kombinacji pozostałych – wtedy można go wykreślić.*

Zbiór wszystkich kombinacji liniowych wektorów $v_1, \ldots, v_n$ oznaczamy przez $\mathrm{lin}(v_1, \ldots, v_n)$, tzn:

$$\mathrm{lin}(v_1, \ldots, v_n) := \{x_1 v_1 + \ldots + x_k v_k : x_1, \ldots, x_k \in \mathbb{R}\}.$$

Tak więc $b$ jest kombinacją liniową $v_1, \ldots, v_n$ wtw. $b \in \mathrm{lin}\{v_1, \ldots, v_k\})$.

Będzie nas interesowało uzyskanie prostego i efektywnego kryterium pozwalającego rozstrzygnąć powyższy problem.

Posuńmy się w naszych rozważaniach troszkę dalej: załóżmy, że tak jest (czyli istnieje rozwiązanie). Jak rozstrzygnąć które z równań są nam zbędne? Zauważmy, że jeżeli jakieś równanie (a konkretniej jego reprezentacja w postaci wierszu w macierzy rozszerzonej) jest kombinacją liniową pozostałych, to możemy go odrzucić/skreślić bez zmieniania zbioru rozwiązań. Możemy tak postępować aż do momentu, gdy już żadnego równania nie da się wyrzucić, czyli inaczej mówiąc, gdy żadne z równań nie będzie kombinacją liniową pozostałych. Dostajemy w konsekwencji następującą definicję[1]:

**Definicja 8.4.** Rozpatrzmy wektory $v_1, \ldots, v_k \in \mathbb{R}^n$. Mówimy, że te wektory są *liniowo zależne*, gdy jeden da się przedstawić jako kombinacja liniowa pozostałych, tzn. istnieje $i \in \{1, \ldots, k\}$ oraz skalary $\alpha_1, \ldots, \hat{\alpha}_i, \ldots, \alpha_k$ takie, że

$$v_i = \sum_{l \neq i} \alpha_l v_l.$$

Wektory są *liniowo niezależne*, jeżeli nie są liniowo zależne.

**Obserwacja 8.1.** *NWSR:*

- *wektory $v_1, \ldots, v_k$ są liniowo niezależne;*
- *zachodzi implikacja $\alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \implies v_1 = \ldots = v_k = 0$.*

*Dowód.* Dowód zostaje jako ćwiczenie (był na wykładzie). □

Jesteśmy teraz gotowi do:

**Definicja 8.5.** Przez *rząd* macierzy $A$ (oznaczam $\mathrm{rank}(A)$) rozumiem maksymalną ilość kolumn w $A$ liniowo niezależnych.

**Obserwacja 8.2.** *Jeżeli mamy wektory $v_1, \ldots, v_n$, to aby sprawdzić czy są liniowo niezależne wystarczy „ułożyć” z nich macierz (nazwijmy ją $A$), a następnie sprawdzić czy* $\mathrm{rank}(A) = n$.

Rząd macierzy się liczy bardzo łatwo – okazuje się, że operacje elementarne zarówno na wierszach, jak i na kolumnach nie zmieniają rzędu macierzy (ćwiczenie). Ponieważ dla macierzy schodkowej jej rząd to po prostu ilość schodków (ćwiczenie), jako wniosek dostajemy: rząd $A$ to ilość schodków po sprowadzeniu $A$ za pomocą operacji elementarnych na wierszach do postaci schodkowej.

Łatwo można także pokazać, że

[1] oznaczenie: element który usuwam z ciągu $\alpha_1, \ldots, \alpha_k$ oznaczam przez $\hat{\alpha}_i$

- dualnie, gdy sprowadzimy $A$ do postaci takiej jakby „dolnie schodkowej” za pomocą operacji elementarnych na kolumnach (to samo jakbyśmy sprowadzali $A^T$ do postaci schodkowej za pomocą operacji elementarnych na wierszach), to rząd $A$ to ilość „pionowych” schodków;
- w konsekwencji $\text{rank}A = \text{rank}A^T$;
- rząd $A$ to wielkość (tzn. ilość kolumn czy wierszy) maksymalnej (w sensie ilości wierszy czy kolumn) odwracalnej macierzy kwadratowej powstałej z $A$ w wyniku skreślenia pewnej ilości kolumn i wierszy (minor).

Widzimy więc, że jeżeli mamy macierz $m \times n$, to maksymalnie rząd tej macierzy wynosi $\min(m, n)$.

Przydatny w liczeniu rzędu może być także następujący

**Fakt 8.1.** *Załóżmy, że mamy daną macierz w postaci blokowej*

$$D := \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

*Wtedy jeżeli albo rząd $A$ albo rząd $B$ jest maksymalny, to* $\text{rank}(D) = \text{rank}(A) + \text{rank}(C)$*.*

Jako bezpośrednią konsekwencję otrzymujemy:

**Twierdzenie 8.1** (Tw. Kroneckera-Capellego)**.** *Układ równań*

$$Ax = b$$

*ma niezerowe rozwiązanie, wtw. gdy* $\text{rank}(A) = \text{rank}(A|b)$ *(rząd macierzy rozszerzonej) (to jest równoważne temu, że $b \in \text{lin}\{col(A)\}$).*

*Dowód.* Sprowadzamy $[A|b]$ do postaci schodkowej za pomocą operacji elementarnych na wierszach, wtedy rzędy się nie zmieniają. Ponieważ rząd to ilość schodków, to: jeżeli nie występuje schodek w ostatniej kolumnie (czyli gdy $\text{rank}(A) = \text{rank}(A|b)$), to układ ma rozwiązanie; jeżeli zaś występuje (czyli gdy $\text{rank}(A) < \text{rank}(A|b)$), to oznacza, że układ nie ma rozwiązania. □

To pozwala nam teraz rozwiązywać układy równań liniowych.

SCHEMAT (formalizacja):

- rozpatrujemy układ równań zapisanych w postaci macierzowej $Ax = b$;
- sprawdzamy, czy $r(A|b) = r(A) = K$;
- jeżeli nie, to nie ma;
- jeżeli tak, to znajdujemy w $A$ minor $M$ rzędu $K := r(A|b) = r(A)$, taki, że $\det M \neq 0$, i wtedy wykreślamy pozostałe równania, a pozostałe traktujemy jako parametry (czyli mamy ich $k - K$);
- w konsekwencji nasze równanie zależy od $k - K$ parametrów, i możemy rozwiązać układ między innymi korzystając z wzorów Cramera.

**Zadanie 8.7.** *Sprawdź czy wektory są liniowo niezależne: a)* $[1, 2], [2, 1], [3, 0]$*; b)* $[1, 2, 3], [2, 3, 1], [3, 1, 0]$.

**Zadanie 8.8.** *Policz rząd macierzy* $[2, 3, 1, 3; 3, 1, 2, 0; 2, 1, 0, -3]$*;* $[1, 0; 0, 1; 2, 1; 3, 2; 3, 4]$.

## 8.4 Bazy

**Definicja 8.6.** Niech $X$ będzie przestrzenią wektorową. Mówimy, że zbiór wektorów $\{v_i\}_{i\in I} \subset X$ jest liniowo niezależny gdy dla dowolnych $i_1, \ldots, i_n \in I$, $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ zachodzi implikacja:

$$\alpha_1 v_{i_1} + \ldots + \alpha_n v_{i_n} = 0 \implies \alpha_1 = \ldots = \alpha_n = 0.$$

**Definicja 8.7.** Niech $X$ będzie przestrzenią wektorową. Mówimy, że zbiór wektorów $\{v_i\}_{i\in I} \subset X$ jest generujący (generuje $X$) gdy dla dowolnego $x \in X$ istnieją $i_1, \ldots, i_n \in I$, $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ takie, że:

$$x = \alpha_1 v_{i_1} + \ldots + \alpha_n v_{i_n}.$$

Zbiór wektorów który jest liniowo niezależny i generujący nazywamy bazą. Ustalenie jakiejś bazy w danej przestrzeni wektorowej to intuicyjnie rzecz biorąc przyjęcie jakiegoś układu współrzędnych (współrzędne danego punktu $x$ to po prostu współczynniki z przedstawienia tego punktu za pomocą kombinacji liniowej elementów z bazy).

**Zadanie 8.9.** *Proszę sprawdzić, czy dany zbiór wektorów jest liniowo niezależny/generujący.*

**Przykład 8.3.** *Tak zwaną* bazą kanoniczną $\mathbb{R}^N$ *nazywamy bazę zadaną przez* $e_1 = [1, 0, \ldots, 0]^T, \ldots, e_N = [0, \ldots, 0, 1]^T$. *Oczywiście ten zbiór jest liniowo niezależny, i każdy punkt* $[x_1, \ldots, x_N]^T$ *z* $\mathbb{R}^N$ *da się przedstawić w postaci kombinacji liniowej elementów z bazy:* $[x_1, \ldots, x_N]^T = x_1 e_1 + \ldots + x_N e_N$.

**Uwaga 8.1.** *(intuicja) Zbiór wektorów w $V$ możemy traktować jak zbiór kierunków w które możemy się poruszać. Wtedy to, że $v_1, \ldots, v_n$ generuje $V$ oznacza, że do każdego punktu z $V$ możemy dojść, zaś to, że te wektory są liniowo niezależne oznacza to, że to dojście jest jednoznaczne („każdy punkt ma co najwyżej jeden adres).*

**Przykład 8.4.** *Pokażemy, że* $[1, 2]^T, [-1, 0]^T$ *jest bazą* $\mathbb{R}^2$. *Najpierw sprawdzimy te wektory są liniowo niezależne, w tym celu wystarczy sprawdzić czy rząd macierzy*

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

*jest dwa, ale jak widać tak jest (macierz $P$ często nazywamy* macierzą przejścia*). Aby sprawdzić, czy te wektory generują* $\mathbb{R}^2$, *musimy sprawdzić, czy dla dowolnych* $[x_1, x_2]^T$ *układ równań*

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

*ma rozwiązanie. Tak jest, gdyż macierz $P$ jest macierzą odwracalną (bo jest to macierz $2 \times 2$ która ma rząd 2).*

*Jeżeli chcemy znaleźć współrzędne punktu* $[x_1, x_2]^T \in \mathbb{R}^2$ *w nowym układzie współrzędnych (w nowej bazie), to musimy rozwiązać powyższy układ, jak łatwo widać*

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = P^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

**Uwaga 8.2.** *Niech $v_1, \ldots, v_N$ będzie bazą $\mathbb{R}^N$. Jeżeli chcemy znaleźć współrzędne punktu $x \in \mathbb{R}^N$ w tej bazie, tzn. skalary $\alpha_1, \ldots, \alpha_N$ takie, że*

$$\alpha_1 v_1 + \ldots + \alpha_N v_N = x,$$

*to musimy odwrócić macierz przejścia $P = [v_1, \ldots, v_N]$, tzn.*

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = P^{-1} x.$$

**Zadanie 8.10.** *a) Pokaż, że* $(1, 2)$, $(2, 1)$ *jest bazą* $\mathbb{R}^2$*. Znajdź współrzędne punktu* $(x_1, x_2)$ *w tej bazie.*

*b) Pokaż, że* $(1, 0, \ldots, 0), (1, 1, 0, \ldots, 0), (1, \ldots, 1, 0), (1, \ldots, 1) \in \mathbb{R}^N$ *tworzą bazę* $\mathbb{R}^N$*. Znajdź współrzędne punktu* $(x_1, \ldots, x_N)$ *w tej bazie.*

**Twierdzenie 8.2.** *Każdy zbiór wektorów liniowo niezależnych można rozszerzyć do bazy.*

*Każdy generujący zbiór wektorów można zmniejszyć do bazy.*

*Dowód.* Pokażę, że każdy zbiór wektorów liniowo niezależnych można rozszerzyć do bazy. Niech $V \subset X$ będzie ustalonym zbiorem wektorów liniowo niezależnych. Niech

$$\mathcal{W} = \{W \subset X \mid V \subset W, W \text{ liniowo niezależny}\}.$$

Oczywiście jest to zbiór niepusty, gdyż $V$ do niego należy. Na $\mathcal{W}$ definiujemy porządek

$$W_1 \prec W_2 :\Leftrightarrow W_1 \subset W_2.$$

Będziemy chcieli zastosować Lemat Kuratowskiego-Zorna. W tym celu musimy pokazać, że każdy łańcuch w $\mathcal{W}$ ma majorantę. Niech $\mathcal{P}$ będzie dowolnym łańcuchem w $\mathcal{W}$. Rozważmy zbiór

$$\bar{P} = \bigcup \mathcal{P}.$$

Będę chciał pokazać, że $\bar{P}$ jest majorantą łańcucha $\mathcal{P}$. Oczywiście wprost z definicji każdy element z $\mathcal{P}$ jest podzbiorem $\bar{P}$. Musimy w takim razie pokazać, że $\bar{P} \in \mathcal{V}$, czyli trzeba sprawdzić, czy $\bar{P}$ jest zbiorem liniowo niezależnym. Niech więc $p_1, \ldots, p_n \in \bar{P}$ będą dowolne i niech $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ będą takie, że

$$\alpha_1 p_1 + \ldots + \alpha_n p_n = 0. \tag{8.1}$$

Oczywiście $p_1 \in P_1, \ldots, p_n \in P_n$ dla pewnych $P_i \in \mathcal{P}$. Ponieważ $\mathcal{P}$ jest łańcuchem, więc istnieje w nim element większy od wszystkich $P_1, \ldots, P_n$, nazwijmy go $P$. Wtedy $P$ jest liniowo niezależny i $p_1, \ldots, p_n \in P$. Ale to oznacza na podstawie (8.1), że $\alpha_1 = \ldots = \alpha_n = 0$.

Zachodzą więc założenia Lematu Kuratowskiego-Zorna, i w konsekwencji zbiór $\mathcal{W}$ można znaleźć element maksymalny (nazwijmy go $\bar{W}$). Pokażemy, że $W$ jest generujący. Załóżmy, dla dowodu nie wprost, że tak nie jest. Wtedy istnieje punkt $x \in X$ który nie da się wygenerować za pomocą elementów ze zbioru $\bar{W}$. Pokażę, że wtedy $\bar{W} \cup \{x\}$ jest także zbiorem liniowo niezależnym (dostanę wtedy sprzeczność z maksymalnością $\bar{W}$).

Weźmy w takim razie dowolne parami różne $w_1, \ldots, w_n \in \bar{W} \cup \{x\}$ oraz $\alpha_1, \ldots, \alpha_n$ takie, że

$$\alpha_1 w_1 + \ldots + \alpha_n w_n = 0. \tag{8.2}$$

Jeżeli wszystkie $w_i$ należą do $W$, to na podstawie liniowej niezależności $W$ dostajemy, że $\alpha_1 = \ldots = \alpha_n = 0$. Rozpatrzmy teraz przypadek, gdy któryś $w_i$ jest równy $x$ (dla prostoty przyjmijmy, że $w_n = x$). Jeżeli $\alpha_n = 0$, to wtedy oczywiście mamy

$$\alpha_1 w_1 + \ldots + \alpha_{n-1} w_{n-1} = 0$$

czyli konsekwentnie z liniowej niezależności $W$ dostajemy $\alpha_1 = \ldots = \alpha_{n-1} = 0$. Pozostał nam jedynie przypadek, gdy $\alpha_n \neq 0$. Ale wtedy na podstawie (8.2) mamy

$$x = -\frac{\alpha_1}{\alpha_n} w_1 - \ldots - \frac{\alpha_{n-1}}{\alpha_n} w_{n-1},$$

otrzymaliśmy sprzeczność, gdyż to oznacza, że $x$ da się wygenerować za pomocą elementów ze zbioru $W$.

Jako ćwiczenie proszę pokazać, że każdy generujący zbiór wektorów można zmniejszyć do bazy. □

**Dygresja 8.1.** *Jak widzimy w powyższym twierdzeniu wykorzystujemy pewnik wyboru. Okazuje się (dowód został zrobiony dopiero w 1984 roku), że to, że każda przestrzeń wektorowa ma bazę jest tak naprawdę równoważne pewnikowi wyboru.*

## 8.5 Wymiar przestrzeni wektorowej

Okazuje się, że każde dwie bazy danej przestrzeni wektorowej są równoliczne. To pozwala nam zdefiniować pojęcie *wymiaru* danej przestrzeni wektorowej, jako ilość elementów dowolnie wybranej bazy (oznaczamy przez $\dim X$, od angielskiego dimension). Mówimy wtedy, że przestrzeń wektorowa ma skończony wymiar, jeżeli ma bazę która ma skończoną liczbę elementów.

**Twierdzenie 8.3.** *Każde dwie bazy danej przestrzeni wektorowej mają tyle samo elementów.*

Dowód powyższego twierdzenia w przypadku skończenie-wymiarowym wynika z poniższego stwierdzenia:

**Stwierdzenie 8.2.** *Niech $v_1, \ldots, v_n$ będzie dowolnym zbiorem generującym $X$ i niech $w_1, \ldots, w_k \in X$ będzie dowolnym ciągiem wektorów. Jeżeli $k > n$, to ciąg $(w_i)$ nie jest liniowo niezależny.*

*Dowód.* Każdy punkt $w_l$ możemy przedstawić za pomocą kombinacji elementów $v_i$

$$w_l = a_{1l}v_1 + \ldots + a_{nl}v_n.$$

Te współczynniki możemy teraz zapisać w postaci macierzy (kolejne kolumny odpowiadają współczynnikom kolejnych wektorów):

$$\begin{bmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nk} \end{bmatrix}$$

Ale powyższa macierz jest macierzą $n \times k$, czyli jej rząd jest ograniczony z góry przez $n$. To oznacza, kolumny nie są liniowo niezależne, czyli któraś z kolumn w tej macierzy da się przedstawić za pomocą kombinacji liniowej pozostałych (załóżmy, że to jest $i$-ta kolumna). Ale to oznacza, że wektor $w_i$ także da się przedstawić za pomocą kombinacji liniowej pozostałych wektorów. □

**Wniosek 8.1.** *Jeżeli $X$ jest przestrzenią $N$ wymiarową, to każdy zbiór $N$ wektorów liniowo niezależnych tworzy bazę. Każdy zbiór $(N+1)$ wektorów jest liniowo zależny.*

**Wniosek 8.2.** *Rozpatrzmy $K$ wektorów w przestrzeni $\mathbb{R}^N$. Wtedy wektory te tworzą bazę, wtw. gdy $K = N$ oraz rząd macierzy powstałej z zestawienia tych wektorów wynosi $N$ (wynika to z faktu, że przy tym założeniu wektory te są liniowo niezależne, a jak wiemy $N$ wektorów liniowo niezależnych w przestrzeni $\mathbb{R}^N$ tworzy bazę).*

UWAGA: Powyższy wniosek jest bardzo wygodny do sprawdzania, czy dany zbiór wektorów w $\mathbb{R}^N$ tworzy bazę!

Teraz powiemy sobie o *macierzy przejścia* (zakładamy, że przestrzeń jest skończenie wymiarowa). Załóżmy, że mamy ustaloną bazę $e_1, \ldots, e_n$ w przestrzeni $X$, i wiemy, że nasz punkt $x$ ma współrzędne w tej bazie $\lambda_1, \ldots, \lambda_n$. Załóżmy, że zmieniamy bazę na $e'_1, \ldots, e'_n$. Pytanie: w jaki sposób wyliczyć współczynniki $x$ w tej nowej bazie? Przedstawmy w tym celu każdy $e'_i$ za pomocą naszej starej bazy:

$$e'_i = p_{1i}e_1 + \ldots + p_{ni}e_n.$$

Układając $p$ w macierz kolumnowo otrzymujemy tak zwaną macierz przejścia (z bazy $e_i$ do $e'_i$)

$$P_X = \begin{bmatrix} p_{11} & \ldots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \ldots & p_{nn} \end{bmatrix}$$

Jeżeli chcemy teraz wyliczyć współczynniki punktu $x$ w nowej bazie, to musimy przedstawić każdy z elementów starej bazy za pomocą nowej, i w konsekwencji jak się temu przypatrzeć (było wyprowadzenie na wykładzie), to sprowadza się to do znalezienia macierzy odwrotnej do $P_X$. Konkludując, współczynniki $x$ w nowej bazie możemy obliczać biorąc

$$(P_X)^{-1}\begin{bmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{bmatrix}.$$

**Przykład 8.5.** *Rozpatrzmy w $\mathbb{R}^2$ bazę kanoniczną $e_1 = [1,0]^T$, $e_2 = [0,1]^T$ punkt*

$$x = \begin{bmatrix}2\\3\end{bmatrix}.$$

*Interesują nas współrzędne tego punkty w bazie*

$$f_1 = [2,5]^T,\ f_2 = [3,-1]^T.$$

*W tym celu tworzymy macierz*

$$P = \begin{bmatrix}2 & 3\\ 5 & -1\end{bmatrix},$$

*znajdujemy macierz do niej odwrotną*

$$P^{-1} = \frac{1}{-17}\begin{bmatrix}-1 & -3\\ -5 & 2\end{bmatrix},$$

*i wyliczamy*

$$P^{-1}\cdot\begin{bmatrix}2\\3\end{bmatrix}.$$

## 8.6 Odwzorowania liniowe

**Definicja 8.8.** Niech $X, Y$ będą przestrzeniami nad ciałem $\mathbb{K}$. Mówimy, że $A : X \to Y$ jest odwzorowaniem liniowym, jeżeli

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay \text{ dla } x, y \in X, \alpha, \beta \in \mathbb{K}.$$

**Obserwacja 8.3.**

- *wystarczy zadać na bazie*
- *z odzorowaniem liniowym, przy ustalonych bazach, stowarzyszamy macierz*
- *z macierzą stowarzyszamy odwzorowanie, gdzie za bazę bierzemy bazę kanoniczną*

**Przykład 8.6.** *Przestrzeń wielomianów stopnia $\leqslant 2$, odwzorowanie pochodnej; wielomiany st. $\leqslant 2 \to \mathbb{R}^2$, wartości w trzech punktach*

Zobaczymy teraz jak zmienia się macierz odwzorowania przy zmianie bazy.

**Przykład 8.7.** *Niech $X = \mathbb{R}^2$. Rozpatrzmy $A$ odwzorowanie zadane przez macierz*

$$\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}.$$

*Wiemy, że wektory własne tej macierzy to $v_1 = [1,1]^T$ i $v_2 = [1,-1]^T$. Zobaczymy, jak będzie wyglądać macierz tego odwzorowania jeśli w $X$ zmienimy bazę na $v_1$ i $v_2$ (inaczej mówiąc zmieniamy w $X$ układ współrzędnych).*

*Zobaczmy więc na który wektor $A$ odwzorowuje $v_1$ i $v_2$:*

$$Av_1 = 3\begin{bmatrix}1\\1\end{bmatrix} = 3v_1 + 0v_2,$$

*analogicznie*

$$Av_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0v_1 + 1v_2.$$

*Widzimy więc, że w macierz odwzorowania* $A$ *w nowej bazie to*

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Zmiana bazy (układu współrzędnych). Porysujemy sobie.

**Przykład 8.8.** *Zobaczymy co się stanie z kwadratem* $K = [-1,1] \times [-1,1]$ *gdy zmienimy bazę na* $v_1 = [1,1]^T$, $v_2 = [0,1/2]^T$. *Mamy*

$$K = \{(x_1, x_2) : x_1 \in [-1,1], x_2 \in [-1,1]\}$$

$$\{x_1 e_1 + x_2 e_2 : x_1 \in [-1,1], x_2 \in [-1,1]\}$$

$$\{x_1(v_1 - 2v_2) + x_2(2v_1) : x_1 \in [-1,1], x_2 \in [-1,1]\}$$

$$\{(x_1 - 2x_2)v_1 - 2x_1 v_1\}...$$

*Jak widzimy po zmianie bazy z kwadratu zrobił się równoległobok. Analogicznie, z koła może się zrobić elipsa.*

**Przykład 8.9.** *Przydatne do preprocessingu danych (whitening, dekorelacja, wybielanie, będzie później).*

# Rozdział 9

# Iloczyn skalarny

## 9.1 Przestrzenie z iloczynem skalarnym (przestrzenie unitarne)

Przypomnijmy sobie pojęcie iloczynu skalarnego.

**Definicja 9.1.** W $\mathbb{R}^N$ przez (standardowy) iloczyn skalarny rozumiemy

$$x \circ y = \sum_i x_i y_i.$$

W $\mathbb{C}^N$ kładziemy

$$x \circ y = \sum_i x_i \bar{y}_i$$

(to jest konsekwencja tego, że chcemy by $x \circ x \geqslant 0$, a dla liczb zespolonych mamy $|z|^2 = z\bar{z}$).

WŁASNOŚCI (dla $\mathbb{R}^N$):

- $(\alpha x) \circ y = \alpha \cdot (x \circ y)$, $x \circ (\alpha y) = \alpha \cdot (x \circ y)$;
- $x \circ (y + z) = x \circ y + x \circ z$;
- $x \circ y = y \circ x$;
- $x \circ x = \|x\|^2$.

WŁASNOŚCI (dla $\mathbb{C}^N$):

- $(\alpha x) \circ y = \alpha \cdot (x \circ y)$, $x \circ (\alpha y) = \bar{\alpha} \cdot (x \circ y)$;
- $x \circ (y + z) = x \circ y + x \circ z$;
- $x \circ y = \overline{y \circ x}$;
- $x \circ x = \|x\|^2$.

Mówimy, że wektory $u, v$ są prostopadłe (ortogonalne), oznaczamy przez $u \perp v$, gdy kąt między nimi jest prosty, tzn. gdy $u \circ v = 0$. Wektor nazywamy normalnym, gdy jego długość jest równa jeden (normalizacja wektora to dzielenie przez normę).

Naszym celem będzie dorzucenie do abstrakcyjnej przestrzeni wektorowej („wyabstrahowanie” własności przestrzeni $\mathbb{R}^N$) iloczynu skalarnego.

**Definicja 9.2.** Niech $X$ będzie rzeczywistą przestrzenią wektorową. Mówimy, że odwzorowanie$\langle , \rangle : X \times X \to \mathbb{R}$ jest iloczynem skalarnym jeżeli jest

- dwulinowe (to znaczy liniowe ze względu na każdą ze zmiennych z osobna):

$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,\ \langle x, \alpha y + \beta z\rangle = \alpha\langle x, y\rangle + \beta\langle x, z\rangle.$$

- symmetryczne (przemienne):

$$\langle y, x\rangle = \langle x, y\rangle.$$

- dodatnio określone:

$$\langle x, x\rangle \geqslant 0,\ \langle x, x\rangle = 0 \text{ wtw gdy } x = 0.$$

**Zadanie 9.1** (iloczyn skalarny)**.** *Proszę sprawdzić które z poniższych funkcji zadają iloczyn skalarny na* $\mathbb{R}^2$ *(gdzie* $x = (x_1, x_2)$*,* $y = (y_1, y_2)$*):*

- $\langle x, y\rangle = x_1y_2 + x_2y_1$;
- $\langle x, y\rangle = x_1y_1 + x_2y_2 + x_1y_2 + x_2y_1$;
- $\langle x, y\rangle = x_1y_1 + x_1y_2 - x_2y_1$;
- $\langle x, y\rangle = |x_1y_1| + |x_2y_2|$.

W przypadku zespolonym definicja modyfikuje się do

**Definicja 9.3.** Niech $X$ będzie zespoloną przestrzenią wektorową (to jest przestrzenią wektorową nad ciałem $\mathbb{C}$). Mówimy, że odwzorowanie$\langle , \rangle : X \times X \to \mathbb{C}$ jest iloczynem skalarnym jeżeli jest

- półtoralinowe:

$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,\ \langle x, \alpha y + \beta z\rangle = \bar{\alpha}\langle x, y\rangle + \bar{\beta}\langle x, z\rangle.$$

- spełnia warunek

$$\langle y, x\rangle = \overline{\langle x, y\rangle}.$$

- dodatnio określone.

**Przykład 9.1.** *Na przestrzeni* $C([a, b], \mathbb{R})$ *funkcji ciągłych na przedziale* $[a, b]$ *o wartościach w* $\mathbb{R}$ *definiujemy iloczyn skalarny za pomocą wzoru*

$$\langle f, g\rangle := \int_a^b f(x)g(x)dx,$$

*gdzie* $\int_a^b h(x)dx$ *oznacza całkę z* $h$ *w przedziale* $[a, b]$*, czyli uproszczając pole pod wykresem.*

**Przykład 9.2.** *Na przestrzeni* $M_{m\times n}(\mathbb{R})$ *macierzy* $m \times n$ *zadajemy często iloczyn skalarny za pomocą wzoru*

$$\langle A, B\rangle = \mathrm{tr} A^T B = \sum_{i,j} a_{ij}b_{ij},$$

*gdzie przypominam* tr *oznacza ślad macierzy, czyli suma wyrażeń na przekątnej. cw.: Proszę sprawdzić tą drugą równość (wystarczy na macierzch* $2 \times 2$*).*

Dla iloczynu skalarnego przez odległość od zera (normę) rozumiemy $\|x\| = \sqrt{\langle x, x\rangle}$.

**Stwierdzenie 9.1** (nierówność Cauchy'ego-Schwarza)**.** *Zachodzi nierówność Cauchy'ego-Schwarza:*

$$|\langle x, y\rangle| \leqslant \|x\|\|y\|$$

*Dowód.* Robimy dowód w przypadku rzeczywistym. Wprost z definicji, funkcja kwadratowa

$$t \to \|x - ty\|^2 = \langle x, x\rangle - 2t\langle x, y\rangle + t^2\langle y, y\rangle$$

jest zawsze nieujemna, co oznacza, że jej wyróżnik $\Delta$ jest niedodatni, czyli

$$4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 \leqslant 0,$$

co dowodzi tezę. □

**Obserwacja 9.1.** *Norma w przestrzeni unitarnej spełnia nierówność trójkąta:*

$$\|x + y\| \leqslant \|x\| + \|y\| \text{ dla } x, y \in X.$$

**Zadanie 9.2.** *Proszę sprawdzić, że norma pochodząca od iloczynu skalarnego spełnia warunek równoległoboku:*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{9.1}$$

*Geometrycznie powyższy fakt jest uogólnieniem tw. Pitagorasa, i mówi, że dla danego równoległoboku, suma kwadratów długości przekątnych jest równa sumie kwadratów długości boków.*

**Zadanie 9.3.** *Niech $X$ rzeczywista przestrzeń unitarna. Proszę pokazać, że dla dowolnych $x, y \in X$*

$$\langle x, y\rangle := \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2). \tag{9.2}$$

**Zadanie 9.4.** *Niech $X$ rzeczywista przestrzeń unitarna. Mamy dwa wektory $x, y$ takie, że*

$$\|x + y\| = 5,\ \|x - y\| = 3.$$

*Proszę wyliczyć $\langle x, y\rangle$.*

## 9.2 Bazy ortonormalne i ortogonalne

Opowiemy sobie teraz o porządnych bazach, które nam nie „zniekształcają” geometrii (jak widzieliśmy, przy zmianie bazy kwadrat może stać się równoległobokiem, a koło elipsą), a precyzyjniej mówiąc nie zmieniają odległości (iloczynu skalarnego). Takie bazy to tak zwane bazy ortonormalne.

**Definicja 9.4.** Mówimy, że ciąg wektorów jest *ortonormalny*, jeżeli jest

- ortogonalny (wszystkie elementy są parami ortogonalne)
- normalny (wektory mają długość jeden)

Okazuje się, że bazy które sa ortogonalne / ortonormalne są bardzo przydatne i wygodne w użyciu. W szczególności łatwo i szybko znajduje się współrzędne w nowej bazie (nie trzeba tak jak poprzednio odwracać macierzy przejścia, co jest czasochłonne). W szczególności dlatego w formacie .jpg używa się właśnie baz ortogonalnych.

**Obserwacja 9.2.** *Załóżmy, że mamy daną bazę ortogonalną $v_1, \ldots, v_N$ w $\mathbb{R}^N$. Chcemy przedstawić punkt $x \in \mathbb{R}^N$ w naszej bazie, tzn. wyliczyć $\alpha_1, \ldots, \alpha_N$ takie, że*

$$x = \alpha_1 v_1 + \ldots + \alpha_N v_N.$$

*TRIK (sprytny): przemnażamy prawostronnie przez $v_k$:*

$$x \circ v_k = (\alpha_1 v_1 + \ldots + \alpha_N v_N) \circ v_k,$$

*ale z ortogonalności $v_i \circ v_j = \delta_{ij}$, co oznacza, że*

$$x \circ v_k = \alpha_k (v_k \circ v_k).$$

*W konsekwencji*

$$\alpha_k = \frac{x \circ v_k}{\|v_k\|^2}.$$

*Gdy dodatkowo nasza baza jest ortonormalna, to $\|v_k\| = 1$, co oznacza, że*

$$\alpha_k = x \circ v_k.$$

**Przykład 9.3.** *Rozpatrzmy sobie jakąś prostą bazę ortogonalną w $\mathbb{R}^2$:*

$$v_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}, v_2 = \begin{bmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}.$$

*Chcemy przedstawić punkt $x = [1, 2]^T$ w naszej bazie. W tym celu wystarczy policzyć*

$$\alpha_1 = x \circ v_1 = 3\sqrt{2}/2, \; \alpha_2 = x \circ v_2 = \sqrt{2}/2$$

*(łatwo sprawdzić, że tak rzeczywiście jest).*

**Zadanie 9.5.** *Niech będzie dany ciąg niezerowych wektorów ortogonalnych. Proszę pokazać, że są one liniowo niezależne. Wsk.: proszę założyć, że tak nie jest, i dojść do sprzeczności modyfikując ideę z poprzedniej obserwacji.*

Dla nas ważna będzie umiejętność konstruowania / budowania ciągów ortonormalnych / ortogonalnych z danych. Jest to tak zwana **Ortonormalizacja / ortogonalizacja Grama-Schmidta**

Opiszemy proces ortogonalizacji (ortonormalizacje najlepiej zrobić najpierw ortogonalizaując, a potem tylko normalizując). Załóżmy, że mamy wektory liniowo niezależne $v_1, \ldots, v_n$. Postaramy się je tak kolejno modyfikować, aby dostać ciąg wektorów ortogonalnych.

KROK 1. Zaczynamy od wektora $v_1$. Ponieważ chcemy mieć wektory o długości jeden, tu nie ma za dużo do wyboru:

$$e_1 = v_1.$$

KROK 2. Teraz mamy już wektor $e_1$. Chcemy więc zmodyfikować wektor $v_2$ aby był prostopadły do $e_1$, w tym celu rozpatrzmy wektor $v_2 - \alpha_1 e_1$. Jeżeli ma być prostopadły do $e_1$, to musi zachodzić

$$(v_2 - \alpha_1 e_1) \circ e_1 = 0,$$

czyli dostajemy $v_2 \circ e_1 - \alpha_1 (e_1 \circ e_1) = 0$, i w konsekwencji

$$\alpha_1 = \frac{v_2 \circ e_1}{e_1 \circ e_1}.$$

Otrzymaliśmy, że wektor $e_2 = v_2 - (v_2 \circ e_1) e_1$ jest prostopadły do $e_1$.

KROK 3. Postępujemy analogicznie dla $v_3$. Chcemy od niego odjąć odpowiednią kombinację liniową $e_1$ i $e_2$ tak aby powstały wektor $v_3 - \alpha_1 e_1 - \alpha_2 e_2$ (uwaga: to $\alpha_1$ jest inne niż poprzednio) był prostopadły zarówno do $e_1$ jak i $e_2$:

$$(v_3 - \alpha_1 e_1 - \alpha_2 e_2) \circ e_1 = 0, \; (v_3 - \alpha_1 e_1 - \alpha_2 e_2) \circ e_2 = 0.$$

Trywialnie dostajemy

$$\alpha_1 = \frac{v_3 \circ e_1}{e_1 \circ e_1}, \; \alpha_2 = \frac{v_3 \circ e_2}{e_2 \circ e_2}.$$

Tak więc otrzymaliśmy, że wektor $e_3 = v_3 - (v_3 \circ e_1) e_1 - (v_3 \circ e_2) e_2$ jest prostopadły zarówno do $e_1$ jak i $e_2$. Postępujemy analogicznie dla kolejnych wektorów otrzymując ciąg wektorów ortonormalnych $e_1, \ldots, e_n$.

**Zadanie 9.6.** *Dokonaj ortonormalizacji Grama-Schmidta wektorów* $(1,1,1),(1,1,0),(1,0,0)$.

**Zadanie 9.7.** *Proszę zortogonalizować ciąg* $1, x, x^2, x^3$ *w przestrzeni* $C([-1,1],\mathbb{R})$. *[informacja* $\int_a^b x^k dx = \frac{1}{k+1}(b^{k+1} - a^{k+1}$*]*

**Zadanie 9.8** (iloczyn skalarny)**.** *Proszę zortogonalizować w przestrzeni* $M_{2\times 2}(\mathbb{R})$ *ciąg macierzy* $[1,2;0,1]$, $[0,1;1,0]$, $[-1,0;0,0]$.

**Zadanie 9.9.** *Udowodnij, że wektory* $e_1 = (1/2,-1/2,1/2,-1/2)$, $e_2 = (-1/2,1/2,1/2,-1/2)$ *są ortonormalne, a następnie rozszerz do bazy ortonormalnej na* $\mathbb{R}^4$.

SZEREGI FOURIERA [tylko informacyjnie]: $1, \cos kx, \sin kx$, ortogonale w przestrzeni $C([-\pi,\pi],\mathbb{R})$ i zupełne, tzn każdą funkcję da się za pomocą ich przedstawić.

## 9.3 Dyskretna transformata Fouriera

Jednym ze sposobów analizy i kompresji danych / obrazów jest próba rozbijania na interesujące nas czynniki (coś w rodzaju zmiany sposobu patrzenia). Bardzo często w tym celu patrzymy na współczynniki w nowej bazie która pomaga wyłapywać interesujące nas zachowanie (za tym stoi idea pojęcia transformaty). Jedną z najważniejszych (jeżeli nie najważniejszą) transformatą jest transformata Fouriera, i w szczególności dyskretna transformata Fouriera. Używa się ją (a także jej liczne odmiany) między innymi do kompresji muzyki (mp3), obrazów (jpeg), usuwania szumów w dźwiękach, oraz szeroko rozumianej analizy danych. Idea dyskretnej transformaty Fouriera polega na wyłapaniu (rozkładaniu na) częstotliwości.

Baza Fouriera jest „fajną" bazą $\mathbb{C}^N$.

**Konstrukcja 9.1** (Baza Fouriera)**.** *Zaczniemy od bazy Fouriera (nazwa na potrzeby wykładu). Domyślnie chodzi o badanie ciągów okresowych o okresie* $N$. *Jak widać tę przestrzeń można utożsamić z przestrzenią* $\mathbb{C}^N$ *– na potrzeby tej sekcji elementy* $\mathbb{C}^N$ *oznaczamy przez* $x = (x_0, \dots, x_{N-1})$.

*Chcemy docelowo łapać wszystkie ciągi okresowe o okresie* $N$. *Naszą bazę jak zazwyczaj spróbujemy zbudować z ciągów geometrycznych o okresie* $N$, *czyli szukamy takich* $q \in \mathbb{C}$ *aby ciąg* $(q^k)_{k\in\mathbb{Z}}$ *był okresowy o okresie* $N$, *czyli gdy* $q^N = 1$. *Dostajemy więc, że* $q_l$ *jest* $l$*-tym pierwiastkiem* $N$*-tego stopnia z jedynki, czyli*

$$q_l = e^{2\pi i l/N} = \cos(2\pi l/N) + i\sin(2\pi l/N) \text{ dla } l = 0,\dots,N-1.$$

*W konsekwencji dostajemy ciągi*

$$(e^{2\pi i k l/N})_{l=0,\dots,N-1} \text{ dla } k = 0,\dots,N-1.$$

*Z powodów zwyczajowych dzielimy jeszcze każdy element przez* $N$. *Dostajemy więc następującą bazę naszej przestrzeni:*

$$f_l = \frac{1}{N}(e^{2\pi i 0 l/N}, e^{2\pi i 1 l/N}, \dots, e^{2\pi i (N-1) l/N})$$

$$= \frac{1}{N}(\cos(2\pi 0 l/N), \cos(2\pi 1 l/N), \dots, \cos(2\pi(N-1)l/N)) + i\frac{1}{N}(\sin(2\pi 0 l/N), \sin(2\pi 1 l/N), \dots, \sin(2\pi(N-1)l/N)).$$

*dla* $l = 0,\dots,N-1$. *Pierwszym elementem naszej bazy będzie punkt* $\frac{1}{N}(1,\dots,1)$ *(czyli on będzie „wyłapywał" zawartość czynnika stałego w* $x$*).*

*Pierwszą istotną informacją (proszę sprawdzić jako ćwiczenie), jest to, że ciąg* $f_k$ *jest ciągiem ortonormalnym w przestrzeni* $\mathbb{C}^N$. *W konsekwencji jest on liniowo niezależny, i ponieważ ma* $N$*-elementów, tworzy bazę* $\mathbb{C}^N$ *(patrz następny rozdział).*

**Obserwacja 9.3.** *Można narysować coś w rodzaju zegara na płaszczyźnie zespolonej.*

Biorąc punkt $x \in \mathbb{C}^N$ możemy go przedstawić w sposób jednoznaczny w postaci kombinacji elementów bazy Fouriera (powstałe współczynniki $X = (X_k)$ to tak zwana transformata Fouriera ciągu $x$):

$$x = \sum_{k=0}^{N-1} X_k f_k. \tag{9.3}$$

Ponieważ ciąg $f_k$ jest ortonormalny, łatwo dostajemy wzór na $X_k$:

$$X_k = (x \circ f_k)/\|f_k\|^2 = \sum_{l=0}^{N-1} x_l e^{-2\pi i k l/N}.$$

Przy wyliczaniu na komputerze transformaty Fouriera stosuje się zazwyczaj tak zwaną szybką transformatę Fouriera (z angielskiego FFT – fast Fourier transform), ona zazwyczaj wymaga założenia, że $N$ jest potęgą dwójki.

W transformacie Fouriera troszkę nie-intuicyjne jest to, że dostajemy współczynniki zespolone (trudno także rysować na wykresie czy dokonywać interpretacji). W związku z tym często rozpatruje / rysuje się spektrogram, który mówi nam na ile dana częstotliwość jest ważna, definiujemy prosto wzorem

$$A_k = |X_k|.$$

Przydatna jest także odwrotna transformata Fouriera, która mając dany ciąg $X$ zwraca nam $x$, czyli wtedy zgodnie z (9.3) dostajemy

$$x_l = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2\pi i k l/N}.$$

Na wykładzie był zrobiony referat odnośnie zastosowania transformaty Fouriera do rozróżniania silnika dobrego od zepsutego (na podstawie danych z firmy ABB).

Więcej informacji o zastosowaniach transformaty Fouriera można znaleźć na wikipedii: oraz na `https://ccrma.st`

**Zadanie 9.10.** *Niech $x = (1, 0, 1, 0)$. Znajdź transformatę Fouriera ciągu $x$, i zaznacz jego spektrogram.*

**Zadanie 9.11** (Scilab)**.** *Niech $N = 128$. Dla $\varphi \in [0, 2\pi)$ definiujemy ciąg $x_\varphi \in \mathbb{C}^N$ wzorem*

$$x_\varphi = (\cos(k\varphi/N))_{k=0,\ldots,N-1}.$$

*Dla paru wybranych $\varphi$ (w szczególności dla $\varphi = 0, \pi/N, \pi/3, \pi/2$) policz transformatę Fouriera ciągu $x^\varphi$ i narysuj spektrogram.*

## 9.4 Podejście geometryczne: rzutowania/projekcje ortogonalne

Nasze podejście z poprzedniej sekcji możemy wyrazić następująco: mamy odwzorowanie $A : \mathbb{R}^M \to \mathbb{R}^N$ oraz punkt $b \in \mathbb{R}^N$. Chcemy znaleźć przybliżone rozwiązanie równania

$$Ax = b.$$

Jeżeli $b$ należy do obrazu $A$, to oczywiście powyższe równanie można rozwiązać dokładnie. Jeżeli $b$ nie należy do obrazu $A$, to znajdujemy punkt w obrazie $A$ który jest najbliższy $b$, i potem już rozwiązujemy dokładnie.

Jeżeli obraz $A$ oznaczymy przez $M$, to nasz problem sprowadza się do następującego problemu (uwaga: zmiana oznaczeń!):

*Załóżmy, że mamy podprzestrzeń $M \subset \mathbb{R}^N$ oraz punkt $x \in \mathbb{R}^N$. Chcemy znaleźć punkt $x_M \in M$ który jest najbliższy punktowi $x$.*

Weźmy w tym celu bazę ortonormalną $M$ $e_1, \ldots, e_K$ i rozszerzmy ją do bazy $\mathbb{R}^N$. Załóżmy, że nasz punkt $x$ ma w tej bazie współrzędne

$$\alpha_1 e_+ \ldots + \alpha_K e_K + \alpha_{K+1} e_{K+1} + \ldots + \alpha_N e_N.$$

Łatwo można teraz zauważyć (patrz stwierdzenie poniżej), że punkt w $M$ najbliższy punktowi $x$ wyraża się wzorem

$$x_M = \alpha_1 e_+ \ldots + \alpha_K e_K. \tag{9.4}$$

**Stwierdzenie 9.2.** *Niech $M$ będzie podprzestrzenią $\mathbb{R}^N$ i niech $e_l$ będzie bazą ortonormalną $M$. Wtedy punkt najbliższy do $x$ w $M$ wyraża się wzorem:*

$$x_M = \langle x, e_1 \rangle e_1 + \ldots + \langle x, e_K \rangle e_K \tag{9.5}$$

*Dowód.* Roszerzmy bazę $M$ do bazy ortonormalnej $\mathbb{R}^N$. Weźmy

$$x = \langle x, e_1 \rangle e_1 + \ldots + \langle x, e_k \rangle e_k + \langle x, e_{k+1} \rangle e_{k+1} + \ldots \langle x, e_N \rangle e_N.$$

Zauważmy najpierw, że $x_M = \langle x, e_1 \rangle e_1 + \ldots + \langle x, e_k \rangle e_k$ jest elementem przestrzeni $M$. Oczywiście $x - x_M$ jest prostopadłe do $M$ (a precyzyjniej do dowolnego elementu $M$). W konsekwencji z twierdzenia Pitagorosa

$$\|x - (x_M + \sum_{i=1}^{k} \beta_i e_k)\|^2 = \|x - x_M\|^2 + \sum_{i=k+1}^{N} |\beta_i|^2,$$

co znaczy, że najbliższym punktem jest rzeczywiście $x_M$. □

Jest to o tyle wygodne, że aby wyliczyć punkt najbliższy $x$ wystarczy znaleźć bazę ortonormalną $M$ (i nie trzeba jej rozszerzać do bazy $\mathbb{R}^N$).

Rzutowanie ortogonalne na $M$ oznaczamy zazwyczaj $p_M$. Zauważmy, że z powyższego stwierdzenia wynika, że jest to odwzorowanie liniowe!

**Definicja 9.5** (przestrzeń prostopadła)**.** Tutaj w sposób naturalny nam się pojawia *przestrzeń ortogonalna* (prostopadła) do $M$:

$$M_\perp := \{x \,|\, x \perp M\},$$

gdzie $x \perp M$ o ile $x$ jest prostopadły do każdego punktu z $M$. Jeżeli $e_1, \ldots, e_K$ jest bazą ortonormalną $M$ którą rozszerzyliśmy do bazy ortonormalnej $e_1, \ldots, e_N$ całej przestrzeni, to wtedy $e_{K+1}, \ldots, e_N$ stanowi bazę $M_\perp$.

**Zadanie 9.12** (przestrzeń prostopadła)**.** *Niech $M = \{t[1;0;1] \,:\, t \in \mathbb{R}\} \subset \mathbb{R}^3$. Proszę znaleźć bazę ortonormalną przestrzeni $M_\perp$.*

Wzór (9.4) oznacza, że jeżeli chcemy znaleźć odwzorowanie $p_M$ (nazywamy go *projekcją ortogonalną na $M$*) które punktowi $x$ przyporządkowuje punkt $x_M \in M$, to możemy postąpić następująco: znajdujemy współczynniki $x$ w bazie $P = e_1, \ldots, e_N$, pierwsze $K$ zostawiamy bez zmian, pozostałe zerujemy i wracamy z powrotem do naszej bazy. Zapisując to w postaci operatorowej dostajemy

$$p_M = P \begin{bmatrix} I_K & 0 \\ 0 & 0 \end{bmatrix} P^{-1} x.$$

**Zadanie 9.13** (projekcje ortogonalne)**.** *Rozpatrujemy przestrzeń $M = \{s[1;0;1;0] + t[1;1;-1;1] \,:\, s,t \in \mathbb{R}\} \subset \mathbb{R}^4$.*

*a) Proszę znaleźć punkt $x_M \in M$ najbliższy punktowi $[0;1;2;3]$.*

*b) Proszę znaleźć macierz $p_M$ projekcji ortogonalnej na $M$.*

Umiejętność znajdowania przestrzeni prostopadłej do danej jest ważna, w szczególności gdy podprzestrzeń jest obrazem pewnego odwzorowania liniowego.

**Twierdzenie 9.1.** *Niech $A : \mathbb{R}^M \to \mathbb{R}^N$ będzie danym odwzorowaniem liniowym. Wtedy*

$$(\mathrm{image}A)_\perp = \mathrm{kernel}A^T.$$

*Dowód.* Zauważmy, że

$$x \perp (\mathrm{image}A) \Leftrightarrow x \perp Ay \quad \text{dla } y \in \mathbb{R}^K \Leftrightarrow \langle x, Ay \rangle = 0 \quad \text{dla } y \in \mathbb{R}^K$$

$$\Leftrightarrow x^T Ay = 0 \quad \text{dla } y \in \mathbb{R}^K \Leftrightarrow (A^T x)^T y = 0 \quad \text{dla } y \in \mathbb{R}^K \Leftrightarrow \langle A^T x, y \rangle = 0 \quad \text{dla } y \in \mathbb{R}^K.$$

Ale teraz $\langle A^T x, y \rangle$ dla $y \in \mathbb{R}^K$ wtedy i tylko wtedy gdy $A^T x = 0$ (aby to zauważyć wystarczy podstawić $y = A^T x$). Czyli gdy $x \in \mathrm{kernel}A^T$. □

Teraz jesteśmy już gotowi pokazać najważniejsze twierdzenie w tej sekcji, które mówi w jaki sposób można znaleźć najlepsze rozwiązanie równania $Ax = b$.

**Twierdzenie 9.2.** *Niech $A : \mathbb{R}^K \to \mathbb{R}^N$, $b \in \mathbb{R}^N$ będzie dane. Wtedy dla $\bar{x} \in \mathbb{R}^K$ NWSR:*

*i) $\bar{x}$ minimalizuje funkcję $x \mapsto \|Ax - b\|$;*

*ii) $\bar{x}$ spełnia równanie $A\bar{x} = p_{\mathrm{image}A} b$;*

*iii) $A^T A\bar{x} = A^T b$.*

*Dowód.* Równoważność i) i ii) wynika z wcześniejszych rozumowań. Pokażemy równoważność ii) i iii):

$$A\bar{x} = p_{\mathrm{image}A} b \Leftrightarrow p_{\mathrm{image}A} A\bar{x} = p_{\mathrm{image}A} b \Leftrightarrow p_{\mathrm{image}A}(A\bar{x} - b) = 0$$

$$\Leftrightarrow (A\bar{x} - b) \perp \mathrm{image}A \Leftrightarrow (A\bar{x} - b) \in (\mathrm{image}A)_\perp$$

$$\Leftrightarrow (A\bar{x} - b) \in \mathrm{kernel}A^T \Leftrightarrow A^T(A\bar{x} - b) = 0 \Leftrightarrow A^T A\bar{x} = A^T b.$$

□

## 9.5 Nadobowiązkowe: Szeregi Fouriera i wielomiany Legendre'a

Jak widzimy w zasadzie największy problem (obliczeniowy) leży w znalezieniu bazy ortogonalnej/ortonormalnej interesującej nas przestrzeni. Matematycy i fizycy się tym problemem zajęli i znaleźli wzory (mniej lub bardziej skomplikowane) na różne bazy ortogonalne.

Zacznę od opisania najprostszej, i równocześnie najważniejszej (tak zwane szeregi Fouriera). Jest to baza podprzestrzeni $C([-\pi, \pi], \mathbb{R})$ składającej się z funkcji trygonometrycznych postaci $\sum_k a_k \cos kx + \sum_k b_k \sin kx$.

**Twierdzenie 9.3.** *Powyższy ciąg jest ortogonalny. Co więcej*

$$\int_{-\pi}^{\pi} 1^2 dx = 2\pi \text{ oraz } \int_{-\pi}^{\pi} \cos^2(mx) dx = \int_{-\pi}^{\pi} \sin^2(mx) dx = \pi.$$

*Dowód.* Dowód prosty, można zrobić nawet bez znajomości całkowania. Wystarczy tak naprawdę zobaczyć, że

$$\int_{-\pi}^{\pi} \sin^2(mx) dx = \int_{-\pi}^{\pi} \cos^2(mx) dx \text{ oraz } \int_{-\pi}^{\pi} \sin^2(mx) dx + \int_{-\pi}^{\pi} \cos^2(mx) dx = \int_{-\pi}^{\pi} 1^2 dx.$$

□

Okazuje się, że każdą funkcję z $C([\pi,\pi],\mathbb{R})$ można przedstawić za pomocą sumy nieskończonego szeregu Fouriera (mówi się, że ciąg składający się z funkcji trygonometrycznych jest zupełny). Inaczej mówiąc w przestrzeni unitarnej $C([-\pi,\pi],\mathbb{R})$, mamy

$$f(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos kx + \sum_{k=1}^{\infty} b_k \sin kx,$$

gdzie

$$a_0 = \frac{\langle f,1\rangle}{\langle 1,1\rangle} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx,\ a_k = \frac{\langle f,\cos(kx)\rangle}{\langle \cos(kx),\cos(kx)\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)dx$$

oraz

$$b_k = \frac{\langle f,\sin(kx)\rangle}{\langle \sin(kx),\sin(kx)\rangle} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(kx)dx.$$

**Przykład 9.4.** *Można łatwo wyliczyć współczynniki rozwinięcia funkcji $f(x) = x$ w szereg Fouriera.*

Wielomiany Legendre'a służą analogicznemu celowi, tylko są to wielomiany ortogonalne w przestrzeni unitarnej $C([-1,1],\mathbb{R})$. Stanowią one bazę podrzestrzeni $C([-1,1],\mathbb{R})$ składającej się z wielomianów dowolnego stopnia.

Przedstawię podstawowe fakty bez dowodów. Najpierw zobaczmy podstawowe cztery sposoby definiowania wielomianów Legendre'a:

- przy pomocy pochodnej
$$P_l(x) = \frac{1}{2^l l!}\frac{\partial^l}{\partial x^l}[(x^2-1)^l];$$
- przy pomocy funkcji generujących
$$\frac{1}{\sqrt{1-2tx+t^2}} = \sum_{l=0}^{\infty} t^l P_l(x);$$
- bezpośrednio przy pomocy wzoru
$$P_l(x) = \sum_{r=0}^{\lfloor l/2\rfloor} (-1)^r \frac{(2l-2r)!}{2^l r!(l-r)!(l-2r)!} x^{l-2r};$$
- rekurencyjnie
$$P_0(x) = 1,\ P_1(x) = x;\ P_{l+1}(x) = \frac{2l+1}{l+1} x P_l(x) - \frac{l}{l+1} P_{l-1}(x).$$

**Zadanie 9.14** (wielomiany Legendre'a)**.** *Korzystając z każdego z czterech wzorów definiujących wielomiany Legendre'a proszę wypisać pierwsze pięć wielomianów Legendre'a.*

**Zadanie 9.15** (wielomiany Legendre'a)**.** *Niech $p > 0$. Proszę znaleźć pierwsze trzy współczynniki rozwinięcia funkcji a) $|x|^p$ b) $\operatorname{sign}(x)\cdot|x|^p$ w szereg Legendre'a.*

Dalsze własności wielomianów Legendre'a:

- $P_l(1) = 1$;
- $P_l$ jest funkcją (nie)parzystą dla (nie)parzystego $l$;
- $P_l$ ma w przedziale $[-1,1]$ $l$ miejsc zerowych (jednokrotnych);

- $P_l(0) = 0$ dla $l$ nieparzystych, $P_{2l}(0) = (-1)^l \frac{(2l)!}{2^{2l}(l!)^2}$.

Kluczowym faktem jest to, że zbiór wielomianów Legendre'a jest ciągiem ortogonalnym zupełnym w przestrzeni $C([-1,1],\mathbb{R})$ (przypominam: zupełność oznacza, że każdą funkcję z tej przestrzeni możemy przedstawić w postaci sumy nieskończonej wielomianów Legendre'a).

**Twierdzenie 9.4.** *Mamy*

$$\int_{-1}^{1} P_l(x)P_m(x)dx = \begin{cases} 0 \ \textit{dla } l \neq m, \\ \frac{2}{2l+1} \ \textit{dla } l = m. \end{cases}$$

**Zadanie 9.16** (wielomiany Legendre'a)**.** *Niech $p > 0$. Proszę znaleźć pierwsze trzy współczynniki rozwinięcia funkcji a) $|x|^p$ b) $\operatorname{sign}(x) \cdot |x|^p$ w szereg Legendre'a.*

# Rozdział 10

# Wektory i wartości własne

## 10.1 Motywacja

Zajmijmy się teraz raz jeszcze zadaniem postawionym przez Fibonacciego. Niech $s_n$ – ilość starych królików w $n$-tym miesiącu, zaś $m_n$ – młodych (przyjmujemy warunek początkowy $s_0 = 1$, $m_0 = 0$). Wtedy jak pamiętamy mamy

$$\begin{cases} s_{n+1} = m_n, \\ m_{n+1} = s_n + m_n, \end{cases}$$

czyli

$$\begin{bmatrix} s_{n+1} \\ m_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} s_n \\ m_n \end{bmatrix}.$$

W konsekwencji

$$\begin{bmatrix} s_n \\ m_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} s_0 \\ m_0 \end{bmatrix}.$$

Widzimy więc, że do wyliczenia ilości królików w $n$-tym miesiącu potrzebujemy wyliczyć wzór na $n$-tą potęgę odpowiedniej macierzy (a dokładniej na wektorze $[s_0, n_0]^T$). Można to łatwo zrobić korzystając z faktu, że znamy rozwiązanie ciągu Fibonacciego (ćwiczenie), ale spróbujmy to zrobić bez tej znajomości.

Tak więc CEL: znaleźć

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

Spróbujemy powtórzyć to jak rozwiązywaliśmy Fibonacciego, a mianowicie spróbujmy zobaczyć, czy da się znaleźć wektory, na których nasza macierz zachowuje się „po ludzku”, czyli takie, że dla pewnego $\lambda \in \mathbb{C}$

$$Av = \lambda v.$$

Taki wektor nazywamy *wektorem własnym odpowiadającym wartości własnej* $\lambda$:

**Definicja 10.1.** Niech $A$ będzie macierzą kwadratową $n \times n$. Mówimy, że $\lambda \in \mathbb{C}$ jest *wartością własną* $A$, jeżeli istnieje takie $w \neq 0$, $w \in \mathbb{C}^n$ (które nazywamy *wektorem własnym odpowiadającym wartości własnej* $\lambda$), że

$$Aw = \lambda w.$$

Kluczowa jest obserwacja, że potęgowanie macierzy na wartościach własnych to to samo co potęgowanie wartości własnych.

**Obserwacja 10.1.** *Jeżeli $v$ jest wektorem własnym odpowiadającym wartości własnej $\lambda$, to*

$$A^n v = \lambda^n v \quad \text{dla } n \in \mathbb{N}$$

*Dowód.* Trywialny dowód indukcyjny. □

Szukanie wartości własnych i wektorów własnych będzie jednym z ważniejszych zajęć w tym rozdziale.

**Definicja 10.2.** Zbiór wszystkich własności własnych $A$ nazywamy *widmem* macierzy $A$ (mówi się też o spektrum) i oznacza $\sigma(A)$:
$$\sigma(A) := \{\lambda \in \mathbb{C} \,|\, \exists v \neq 0 : Av = \lambda v\}.$$

## 10.2 Wartości i wektory własne

Postaramy się teraz pokazać w jaki sposób szukać wartości własnych i wektorów własnych. Chcemy znaleźć $\lambda \in \mathbb{C}$ i $v \in \mathbb{C}^N$, $v \neq 0$, takie, że
$$Av = \lambda v,$$
czyli, że $(A - \lambda \mathrm{I})v = 0$. Gdyby macierz $A - \lambda \mathrm{I}$ była odwracalna, to oczywiście dostalibyśmy $v = 0$, czyli przypadek nas nie interesujący. To oznacza, że ważny jest przypadek gdy $A - \lambda \mathrm{I}$ nie jest odwracalne. Takie $\lambda \in C$ nazywamy *wartością własną* macierzy $A$. Zbiór wszystkich wartości własnych macierzy $A$ nazywamy *spektrum* lub *widmem* macierzy $A$ (często oznaczamy przez $\sigma(A)$).

Oczywiście powstaje pytanie, jak takie wartości własne znajdować. Zauważmy, że macierz jest nieodwracalna wtw. gdy wyznacznik się zeruje. Dla danej macierzy kwadratowej $A$ zdefiniujmy jej *wielomian charakterystyczny* wzorem
$$W_A(\lambda) = \det(A - \lambda \mathrm{I}) \quad \text{dla } \lambda \in \mathbb{C}.$$
Wtedy $\lambda$ jest wartością własną macierzy $A$, wtw. gdy jest zerem jej wielomianu charakterystycznego.

Przez $\mathrm{tr}A$ rozumiem *ślad* macierzy $A$, czyli suma współczynników na przekątnej: $\mathrm{tr}A := \sum_i a_i i$.

**Zadanie 10.1.** *Proszę pokazać, że* $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

**Obserwacja 10.2.** *Niech $A = [a_{ij}]$ będzie macierzą kwadratową $n \times n$. Wtedy*
$$W_A(\lambda) = (-1)^n(\lambda^n - a_1\lambda^{n-1} \pm \ldots + (-1)^n a_n),$$
*gdzie $a_1 = \mathrm{tr}A$, $a_n = \det A$.*

Jako bezpośredni wniosek z powyższego, otrzymujemy wzór na wielomian charakterystyczny macierzy $2 \times 2$:
$$W_A(\lambda) = \lambda^2 - \mathrm{tr}A\lambda + \det A.$$

**Stwierdzenie 10.1.** *Niech $\lambda$ będzie wartością własną macierzy $A \in M_{n\times n}(\mathbb{C})$. Wtedy istnieje wektor własny odpowiadający $\lambda$, tzn. taki wektor $v \in \mathbb{C}^n$, $v \neq 0$, że $Av = \lambda v$.*

*Dowód.* Konsekwencja rzędu. □

Pokażmy teraz na przykładach w jaki sposób można liczyć wartości własne i wektory własne macierzy.

**Przykład 10.1.** *Rozpatrzmy macierz*
$$A := \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$
*Wtedy oczywiście*
$$W_A(\lambda) = \lambda^2 - \mathrm{tr}A\lambda + \det A = \lambda^2 - 4\lambda + 3.$$
*Oznacza to, że wartości własne $A$ to $\lambda_1 = 1$, $\lambda_2 = 3$.*

*Znajdziemy teraz wektory własne. Aby znaleźć wektor własny odpowiadający wartości własnej* $1$*, musimy znaleźć niezerowe rozwiązanie równania*

$$\begin{cases} 2x + y = x, \\ x + 2y = y. \end{cases}$$

*Wiemy, że jedno z powyższych równań jest kombinacją pozostałych, co oznacza, że przy rozwiązywaniu możemy (jeżeli go znajdziemy) odrzucić. Tu nie ma problemu, gdyż te równania są tylko dwa. Tak więc nasz układ jest równoważny równaniu*

$$2x + y = x.$$

*Rozwiązanie będzie zależało od parametru (możemy przyjąć, że parametrem będzie* $y$*). Ponieważ wystarczy nam znaleźć dowolne niezerowe rozwiązanie, możemy przyjąć* $y = 1$*, i wtedy dostajemy* $x = -1$*. Tak więc wektorem własnym odpowiadającym wartości własnej* $1$ *jest wektor* $[-1, 1]^T$*.*

*Znajdziemy teraz wektor odpowiadający wartości własnej* $3$*. Postępując analogicznie jak poprzednio dostajemy równanie*

$$2x + y = 3x,$$

*czyli kładąc* $y = 1$ *otrzymujemy* $x = 1$*, czyli dostajemy wektor własny równy* $[1, 1]^T$*.*

**Zadanie 10.2.** *Proszę znaleźć wartości i wektory własne macierzy a)* $[2, 1; 1; 2]$*; b)* $[0, 1; -1; 0]$*; c)* $[1, 2, 3; 3, 1, 0; -2, 1$

**Zadanie 10.3** (Scilab)**.** *Mamy dane* $5$ *stron i linki:* $1 \to 1$*,* $1 \to 2$*,* $1 \to 3$*,* $1 \to 5$*,* $2 \to 2$*,* $2 \to 3$*,* $3 \to 1$*,* $3 \to 5$*,* $4 \to 1$*,* $4 \to 4$*,* $5 \to 1$*. Proszę znaleźć wagi odpowiednich stron powstałe w wyniku zastosowania algorytmu Google PageRank.*

**Obserwacja 10.3.** *Dla macierzy górnie (dolnie) przekątniowej wartości własne znajdują się na przekątnej.*

## 10.3 Zastosowanie – rozmnażanie królików

Korzystając z wiadomości z poprzedniej sekcji spróbujmy znaleźć wektory własne i wartości własne interesującej nas macierzy Fibonacciegi. Wielomian charakterystyczny:

$$W_A(\lambda) = \lambda^2 - \lambda - 1,$$

czyli ma wartości własne dane przez

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

a odpowiadają im wartości własne: $[1, \lambda_1]^T$, $[1, \lambda_2]^T$.

Chcemy więc obliczyć

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Ponieważ wiemy w jaki sposób zachowują się iteracji naszej macierzy na wektorach własnych, spróbujmy działać jak poprzednio i przedstawić nasz wektor $[1, 0]^T$ za pomocą kombinacji wektorów własnych:

$$[1, 0]^T = \alpha_1 [1, \lambda_1]^T + \alpha_2 [1, \lambda_2]^T,$$

czyli $\alpha_1 + \alpha_2 = 1$, $\alpha_1 \lambda_1 + \alpha_2 \lambda_2 = 0$. Jako ćwiczenie proszę policzyć $\alpha_1$, $\alpha_2$ i wyliczyć w konsekwencji wzory na $s_n$, $m_n$.

**Dygresja 10.1.** *Jeżeli mamy macierz* $B$ *oraz* $e_i$ *to baza kanoniczna, to*

$$B = [Be_1, \dots, Be_n].$$

*W konsekwencji, by policzyć* $A^n$ *wystarczy policzyć* $A^n e_i$ *dla* $i = 1, \dots, N$*.*

**Zadanie 10.4.** *Policz* $A^n$ *dla a)* $A = [2]$*; b)* $A = [2, 1; 1, 2]$*;* $A = [1, 0, 0; 0, 2, 0; 0, 0, 3]$.

## 10.4 Macierze symetryczne

Dlaczego bardzo ważne: MNK produkuje macierz symetryczną! Specjalne metody pracy z macierzami symetrycznymi, ale w szczególności też mają one specyficzne własności.

Okazuje się, że wartości własne macierzy symetrycznej są rzeczywiste, co więcej, można dobrać wektory własne ortonormalne.

**Twierdzenie 10.1.** *Niech $A$ będzie macierzą symetryczną rzeczywistą. Wtedy istnieje baza składająca się z wektorów własnych ortonormalnych. Inaczej mówiąc, można dobrać takie wektory własne ortonormalne $v_1, \ldots, v_n$, że biorąc $P = [v_1, \ldots, v_n]$ (czyli macierz przejścia), macierz*

$$PAP^{-1}$$

*jest macierzą diagonalną.*

My udowodnimy to twierdzenie w pewnym szczególnym przypadku gdy macierz ma różne wartości własne. Przyda nam się do tego następujące twierdzenie (przypominam, że macierz jest symetryczna o ile $A^T = A$).

**Obserwacja 10.4.** *Niech $A$ będzie macierzą hermitowską. Wtedy wartości własne $A$ są rzeczywiste;*

*Dowód.* Kiedyś wspominałem (łatwo można sprawdzić bezpośrednio), że

$$\langle Au, v\rangle = \langle u, A^T v\rangle$$

Weźmy teraz wartość własną $\lambda \in \mathbb{C}$ oraz odpowiadający jej wektor własny $v$. Wtedy mamy

$$\langle Av, v\rangle = \langle \lambda v, v\rangle = \lambda \|v\|^2,$$

zaś z drugiej strony z tego, że $A$ jest hermitowska dostajemy

$$\langle Av, v\rangle = \langle v, A^\dagger v\rangle = \bar{\lambda}\langle v, v\rangle.$$

W konsekwencji $\lambda = \bar{\lambda}$, czyli $\lambda \in \mathbb{R}$. □

**Obserwacja 10.5.** *Wektory własne macierzy symetrycznej odpowiadające różnym wartościom własnym są ortogonalne.*

*Dowód.* Rozpatrzmy teraz wektory własne $v_1, v_2$ odpowiadające wartościom własnym $\lambda_1, \lambda_2$, $\lambda_1 \neq \lambda_2$. Wtedy mamy

$$\langle Av_1, v_2\rangle = \langle \lambda_1 v_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle,$$

oraz (bo wartości własne rzeczywiste)

$$\langle Av_1, v_2\rangle = \langle v_1, A^T v_2\rangle = \langle v_1, Av_2\rangle = \langle v_1, \lambda_2 v_2\rangle = \bar{\lambda}_2\langle v_1, v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

W konsekwencji $\langle v_1, v_2\rangle = 0$. □

# Rozdział 11

# Formy kwadratowe

Pojęcie formy kwadratowej jest uogólnieniem funkcji kwadratowej z $\mathbb{R}$. Pokażemy najpierw sytuację w $\mathbb{R}^2$ i $\mathbb{R}^3$ (w celu rozwinięcia intuicji), a następnie uogólnimy na $\mathbb{R}^N$ i przestrzenie skończenie wymiarowe.

MOTYWACJA 1: zobaczmy, jak wygląda obraz koła przez odwzorowanie liniowe. Odleglosc inna!

## 11.1 Dwie, trzy i $n$ zmiennych

Przez formę (funkcję) kwadratową na $\mathbb{R}^2$ rozumiem funkcję postaci

$$ax^2 + by^2 + cxy.$$

Zaczniemy od rozpatrzenia/narysowania podstawowych przykładów (były dokładne rysunki z omówieniem na wykładzie):

- $x^2 + y^2$
- $-(x^2 + y^2)$
- $x^2$
- $-y^2$
- i funkcja stale równa zero.

**Obserwacja 11.1** (było dokładnie na wykładzie)**.** *Okazuje się, że każda forma kwadratowa na $\mathbb{R}^2$ jest podobna do którejś z powyższych. Precyzyjniej mówiąc, każda forma kwadratowa na płaszczyźnie da się przedstawić w postaci sumy/różnicy co najwyżej dwóch przeskalowanych kwadratów jakiś wyrażeń (jak ktoś chce, to można zawsze to przeskalowanie wrzucić do środka).*

*Dowód sobie zobaczymy na przykładach. Weźmy*

$$f(x, y) = 2x^2 + y^2 + xy.$$

*Wtedy ponieważ $2x^2 + xy = 2(x + y/4)^2 - y^2/8$, możemy zapisać*

$$2x^2 + y^2 + xy = 2(x + y/4)^2 - y^2/8 + y^2 = 2(x + y/4)^2 + \frac{7}{8}y^2.$$

*Widzimy więc, że nasza funkcja jest „podobna" do $x^2 + y^2$. Jeżeli w naszej formie kwadratowej występuje jakaś zmienna w kwadracie, to procedura jest analogiczna.*

*Pozostaje jedynie rozważyć przypadek, gdy nie ma żadnego kwadratu:*

$$f(x, y) = xy.$$

*Wtedy mamy*

$$xy = (x + y)^2/4 - (x - y)^2/4.$$

Rozpatrzmy teraz sytuację dla trzech zmiennych. Przez formę kwadratową na $\mathbb{R}^3$ rozumiem

$$f(x,y,z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz.$$

**Obserwacja 11.2.** *I znowu możemy dowolną formę kwadratową sprowadzić do sumy/różnicy co najwyżej trzech kwadratów. Pokażemy schemat na przykładach.*

*Niech*

$$f(x,y,z) = 2y^2 - 4z^2 + xy - 4xz + 6yz.$$

*Widzimy, że w powyższym wzorze są zmienne które występują w kwadracie ($y$ i $z$). Weźmy sobie dowolną z nich (dla przykładu $z$), i postarajmy się „wrzucić" do kwadratu $z$ z wszystkie czynniki w których w powyższym wzorze występuje $z$:*

$$-4z^2 - 4xz + 6yz = -4(z + x/2 - \frac{3}{4}y)^2 + x^2 + \frac{9}{4}y^2.$$

*To oznacza, że*

$$2y^2 - 4z^2 + xy - 4xz + 6yz = [-4(z + x/2 - \frac{3}{4}y)^2 + x^2 + \frac{9}{4}y^2] + 2y^2 + xy$$

$$= -4(z + x/2 - \frac{3}{4}y)^2 + \mathit{reszta},$$

*gdzie*

$$\mathit{reszta} = x^2 + \frac{17}{4}y^2 + xy.$$

*Zauważmy teraz, że „reszta" zależy tylko od zmiennych $x, y$ i w konsekwencji możemy dla nich zastosować procedurę dla dwóch zmiennych.*

*A co się dzieje, jeżeli nie ma żadnego kwadratu? Weźmy dla przykładu*

$$f(x,y,z) = 2xy - 4xz + 6yz.$$

*Pamiętamy, że $2xy = (x+y)^2/2 - (x-y)^2/2$ (przedstawienie w postaci różnicy kwadratów). Wprowadźmy sobie nowe zmienne $u = x + y$ i $v = x - y$, i przedstawmy naszą formą przy pomocy tych zmiennych (zamiast $x$ i $y$):*

$$f(u,v,z) = u^2/2 - v^2/2 - 4((u+v)/2)z + 6((u-v)/2)z = u^2/2 - v^2/2 - 2uz - 2vz + 3uz - 3vz$$

$$= u^2/2 - v^2/2 + uz - 5vz.$$

*Widzimy, że w nowych współrzędnych mamy kwadrat, i teraz postępujemy jak w poprzednim przypadku:*

$$f(u,v,z) = [u^2/2 + uz] - v^2/2 - 5vz = [\frac{1}{2}(u+z)^2 - z^2/2] - v^2/2 - 5vz$$

$$= \frac{1}{2}(u+z)^2 - v^2/2 - 5vz - z^2/2 = \frac{1}{2}(u+z)^2 - \frac{1}{2}(v+5z)^2 - \frac{25}{2}z^2 - z^2/2$$

$$= \frac{1}{2}(u+z)^2 - \frac{1}{2}(v+5z)^2 + 12z^2.$$

*I teraz możemy wrócić do naszych podstawowych zmiennych*

$$f(x,y,z) = \frac{1}{2}(x+y+z)^2 - \frac{1}{2}(x-y+5z)^2 + 12z^2.$$

**Zadanie 11.1** (formy kwadratowe)**.** *Proszę przy pomocy powyższego algorytmu przedstawić dane formy kwadratowe w postaci sumy/różnicy kwadratów a) $x^2 + xy + y^2$; b) $4xy$; c) $x^2 + xz + y^2 + yz + z^2$; d) $xy + yz + yz$.*

Zajmiemy się teraz sytuacją dla $n$ zmiennych. Mówimy, że $f : \mathbb{R}^N \to \mathbb{R}$ jest formą kwadratową jeżeli jest postaci

$$f(x) = \sum_{i,j=1}^{N} a_{ij} x_i x_j,$$

dla $x = [x_1, \ldots, x_n]^T$. Proszę zauważyć, że z korzystając z przemienności mnożenia możemy się zawęzić do sytuacji gdy $a_{ij} = a_{ji}$, gdyż $x_i x_j = x_j x_i$ (kładziemy $a_{ij} = a_{ji} = \frac{a_{ij}+a_{ji}}{2}$). Od tej pory zawsze zakładamy, że $a_{ij} = a_{ji}$.

**Zadanie 11.2** (formy kwadratowe)**.** *Proszę uogólnić algorytm sprowadzania do postaci kanonicznej dla $n$ zmiennych i następnie przedstawić formy kwadratowe w postaci sumy/różnicy kwadratów czynników liniowo niezależnych (precyzyjnie nazywa się to sprowadzeniem do postaci kanonicznej) a)$x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1$; b) $x_1^2 + 4x_1x_2 - 4x_1x_3 + 6x_1x_4 + 2x_3x_4$.*

Tworzymy teraz macierz kwadratową symetryczną odpowiadającą danej formie kwadratowej wzorem

$$A_f = [a_{ij}]_{i,j=1..N}$$

Jak łatwo widać mając daną macierz symetryczną odtwarzamy formę kwadratową za pomocą wzoru

$$f(x) = x^T A_f x.$$

Schemat/algorytm przedstawiony w poprzedniej sekcji możemy uogólnić na dowolny skończony wymiar. Okazuje się [tak zwane Twierdzenie Sylwestra], że po sprowadzeniu formy kwadratowej $f$ do postaci kanonicznej[1] ilość współczynników $w_f$ większych od zera i ilość $m_f$ mniejszych od zera nie zależy od sposobu redukcji. W konsekwencji dobrze zdefiniowanym pojęciem jest *sygnatura formy kwadratowej* $(w_f, m_f)$.

**Zadanie 11.3.** *Policz sygnaturę formy kwadratowej $x_1x_2 + x_1x_3 + x_2x_3$.*

W wielu przypadkach można wyliczyć sygnaturę bez sprowadzania formy do postaci kanonicznej. Niech $A$ będzie macierzą symetryczną danej formy kwadratowej $f$. Poniżej przedstawiam schemat (bez dowodu na wykładzie, może zrobię na następnym?):

- niech $D_k = \det[a_{ij}]_{i,j=1..k}$ dla $k = 1..n$;
- zakładamy, że każdy $D_k \neq 0$;
- wtedy $w_f$ to ilość elementów dodatnich w ciągu $D_1, D_2/D_1, \ldots, D_n/D_{n-1}$, a $m_f$ ilość współczynników ujemnych.

**Zadanie 11.4.** *Korzystając z powyższego schematu (jeżeli się da) policzyć sygnaturę formy $x_1^2 + 4x_1x_2 - x_2^2 + 6x_1x_3 + 2x_1x_3 - 2x_3^2$.*

## 11.2 Przestrzenie skończenie wymiarowe

Teraz zajmiemy się sytuacją w ogólnych przestrzeniach skończenie wymiarowych. Niech $X$ będzie przestrzenią skończenie wymiarową, i niech $\mathcal{E} = \{e_1, \ldots, e_n\}$ będzie ustaloną bazą $X$. Mówimy, że $f : X \to \mathbb{R}$ jest formą kwadratową, jeżeli dla pewnych $a_{ij}$

$$f(x) = \sum_{i,j=1}^{N} a_{ij} \alpha_i \alpha_j \text{ gdzie } x = \alpha_1 e_1 + \ldots + \alpha_n e_n.$$

Macierz współczynników w bazie $\mathcal{E}$ oznaczmy przez $A_{\mathcal{E}} = [a_{ij}]$.

[1] czyli po przedstawieniu w postaci sumy/różnicy kwadratów czynników liniowo niezależnych

**Twierdzenie 11.1.** *Niech $f : X \to \mathbb{R}$ będzie formą kwadratową. Niech $\mathcal{E}, \mathcal{E}'$ będą danymi bazami $X$, i niech $P$ oznacza macierz przejścia z $\mathcal{E}$ do $\mathcal{E}'$*[2]

*Wtedy*

$$A_{\mathcal{E}} = P^T A_{\mathcal{E}'} P.$$

*Dowód.* Niech $x \in X$ będzie dowolne. Wtedy

$$f(x) = [\alpha_1, \ldots, \alpha_n]^T A_{\mathcal{E}} [\alpha_1, \ldots, \alpha_n] = [\alpha'_1, \ldots, \alpha'_n]^T A_{\mathcal{E}'} [\alpha'_1, \ldots, \alpha'_n], \tag{11.1}$$

gdzie $\alpha_i$ to współrzędne $x$ w bazie $\mathcal{E}$, a $\alpha'_i$ to współrzędne $x$ w bazie $\mathcal{E}'$. Skoro

$$[\alpha'_1, \ldots, \alpha'_n]^T = P[\alpha_1, \ldots, \alpha_n]^T,$$

to otrzymujemy z (11.1) dostajemy

$$\begin{aligned}[\alpha_1, \ldots, \alpha_n]^T A_{\mathcal{E}} [\alpha_1, \ldots, \alpha_n] &= (P[\alpha_1, \ldots, \alpha_n])^T A_{\mathcal{E}'} (P[\alpha_1, \ldots, \alpha_n]) \\ &= [\alpha_1, \ldots, \alpha_n]^T P^T A_{\mathcal{E}'} P [\alpha_1, \ldots, \alpha_n]\end{aligned}$$

Skoro $x$ było dowolne, to dostajemy

$$A_{\mathcal{E}} = P^T A_{\mathcal{E}'} P.$$

□

**Twierdzenie 11.2.** *Niech będzie dana forma kwadratowa na przestrzeni skończenie wymiarowej $X$. Wtedy istnieje taka baza $\mathcal{V} = \{v_1, \ldots, v_n\}$ w $X$, że*

$$f(x) = \sum_{i=1}^{n} a_i \beta_i^2 \quad \text{dla } x = \beta_1 v_1 + \ldots + \beta_n v_n.$$

*Inaczej mówiąc w pewnej bazie macierz formy kwadratowej jest diagonalna.*

*Dowód.* Wybierzmy dowolną bazę $\mathcal{E}$ w $X$, w tej bazie $f$ ma wzór

$$f(\sum_i \alpha_i e_i) = \sum_{i,j} a_{ij} \alpha_i \alpha_j,$$

gdzie $a_{ij} = a_{ji}$. Będziemy rekurencyjnie tę bazę modyfikować aby otrzymać właściwą.

Rozpatrzymy teraz dwie sytuacje.

Przypadek 1. Któryś $a_{ii} \neq 0$. Dla prostoty rozważmy przypadek, gdy $a_{11} \neq 0$. Wtedy stosując algorytm z poprzedniej sekcji możemy zapisać (dla pewnych $c_i$ i $\tilde{a}_{ij}$):

$$f(\sum_i \alpha_i e_i) = a_{11}(\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n)^2 + \sum_{i,j \geqslant 2} \tilde{a}_{ij} \alpha_i \alpha_j.$$

Tworzymy nową bazę $\tilde{e}_i$ za pomocą wzoru

$$\tilde{e}_1 = e_1, \, \tilde{e}_i = e_i - c_i e_1 \quad \text{dla } i \geqslant 2.$$

Wtedy dla $x = \sum_i \tilde{\alpha}_i \tilde{e}_i$ mamy

$$f(\sum_i \tilde{\alpha}_i \tilde{e}_i) = f((\tilde{\alpha}_1 - c_2\tilde{\alpha}_2 - \ldots - c_n\tilde{\alpha}_n)e_1 + \tilde{\alpha}_2 e_2 + \ldots + \tilde{\alpha}_n e_n)$$

[2] przypominam: oznacza to, że jeżeli $x \in X$ ma w bazie $\mathcal{E}$ współrzędne $[\alpha_1, \ldots, \alpha_n]^T$, to w bazie $\mathcal{E}'$ jego współrzędne wyrażają się wzorem $P[\alpha_1, \ldots, \alpha_n]^T$.

$$= a_1 1\tilde{\alpha}_1^2 + \sum_{i,j\geqslant 2} \tilde{a}_{i,j}\tilde{\alpha}_i\tilde{\alpha}_j.$$

I wtedy możemy się zawęzić do bazy $\tilde{e}_2, \ldots, \tilde{e}_n$ i postępować rekurencyjnie.

Przypadek 2. Wszystkie $a_{ii} = 0$. Wtedy oczywiście któryś $a_{ij} \neq 0$ (w przeciwnym razie forma jest zerowa). Sprowadzimy do Przypadku 1. Dla prostoty załóżmy, że $a_{12} \neq 0$. Wprowadzamy teraz bazę

$$\tilde{e}_1 = e_1 + e_2,\ \tilde{e}_2 = e_1 - e_2,\ \tilde{e}_i = e_i \quad \text{dla } i \geqslant 3.$$

Wtedy

$$\begin{aligned} f(\sum_i \tilde{\alpha}_i\tilde{e}_i) &= f((\tilde{\alpha}_1 + \tilde{\alpha}_2)e_1 + (\tilde{\alpha}_1 - \tilde{\alpha}_2)e_2 + \sum_{i\geqslant 3}\tilde{\alpha}_i e_i) \\ &= 2a_{12}[(\tilde{\alpha}_1 + \tilde{\alpha}_2)(\tilde{\alpha}_1 - \tilde{\alpha}_2)] + \sum_{(i,j)\notin\{(1,2),(2,1)\}} \tilde{a}_{ij}\tilde{\alpha}_i\tilde{\alpha}_j \\ &= 2a_{12}\tilde{\alpha}_1^2 - 2a_{12}\tilde{\alpha}_2^2 + \sum_{(i,j)\notin\{(1,2),(2,1)\}} \tilde{a}_{ij}\tilde{\alpha}_i\tilde{\alpha}_j \end{aligned}$$

I widzimy teraz, że współczynnik przy $\tilde{\alpha}_1$ jest w kwadracie. □

**Przykład 11.1** (zrobiony na wykładzie). $f(x_1, x_2) = x_1^2 + 4x_1x_2 - x_2^2$.

**Zadanie 11.5.** *Proszę wskazać taką bazę w $\mathbb{R}^n$, że macierz odpowiadająca w tej bazie danej formie kwadratowej staje się diagonalna: a) $f(x_1, x_2) = x_1x_2$; b) $f(x_1, x_2, x_2) = x_1^2 + x_1x_2 + x_2x_3$.*

Poniższe zadanie czasami się przydaje gdyż pozwala na wyliczanie współczynników formy kwadratowej w danej bazie.

**Zadanie 11.6.** *Niech $X$ będzie przestrzenią skończenie wymiarową i niech $e_i$ będzie bazą $X$. Niech $f$ będzie formą kwadratową której macierz (w bazie $e_i$) oznaczmy przez $A$.*

*a) Definiujemy*

$$F(x, y) := \frac{1}{4}(f(x + y) - f(x - y)) \quad \text{dla } x, y \in \mathbb{R}^n.$$

*Wtedy $F$ jest dwuliniowe (to znaczy funkcje $F(\cdot, y)$ i $F(x, \cdot)$ są liniowe dla dowolnych ustalonych $x$ i $y$) i symetryczne (to znaczy $f(x, y) = f(y, x)$).*

*b) Proszę pokazać, że*

$$F(x, y) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^T A \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{dla } x = x_1e_1 + \ldots + x_ne_n, y = y_1e_1 + \ldots + y_ne_n \in X.$$

*c) Proszę pokazać, że*

$$a_{ij} = F(e_i, e_j) \quad \text{dla } i, j = 1, \ldots, n.$$

**Twierdzenie 11.3.** *Niech $X$ będzie przestrzenią skończenie wymiarową z bazą $\mathcal{E}$. Niech $A$ będzie macierzą symetryczną (w bazie $\mathcal{E}$) danej formy kwadratowej $f$. Niech $D_0 = 1$,*

$$D_k = \det[a_{ij}]_{i,j=1..k} \neq 0 \quad \text{dla } k = 1..n.$$

*Wtedy*

- *wtedy $w_f$ to ilość elementów dodatnich w ciągu $D_1/D_0, D_2/D_1, \ldots, D_n/D_{n-1}$, a $m_f$ ilość współczynników ujemnych.*

- *f jest w postaci kanoniczej w bazie*

$$v_i = \frac{1}{D_{i-1}} \begin{vmatrix} a_{11} & \dots & a_{1i} \\ \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,i} \\ e_1 & \dots & e_i \end{vmatrix}.$$

*Dowód.* Niech $F$ będzie zdefiniowane jak w zadaniu 11.6 (jest to forma dwuliniowa „odpowiadająca" danej formie kwadratowej). Łatwo można sprawdzić, że

$$F(v_i, e_j) = \frac{1}{D_{i-1}} \begin{vmatrix} a_{11} & \dots & a_{1i} \\ \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,i} \\ a_{j,1} & \dots & a_{j,i} \end{vmatrix} = \begin{cases} 0 & \text{dla } j < i, \\ D_i/D_{i-1} & \text{dla } j = i. \end{cases}$$

Oznacza to, że $F(v_i, v_j) = 0$ dla $i < j$, i w konsekwencji $v$ diagonalizuje $f$. Oczywiście także

$$F(v_i, v_i) = D_i/D_{i-1}.$$

Ale to są współczynniki przy kwadratach! □

**Wniosek 11.1.** *Każdą formę kwadratową można za pomocą ortonormalnej zmiany zmiennych sprowadzić do postaci kanonicznej.*

*był na wykładzie.* Wystarczy rozważyć bazę ortonormalną dla macierzy symetrycznej odpowiadającej formie $A$. Wtedy macierz przejścia (czyli te wektory ułożone) jest ortonormalna, i w konsekwencji $P^{-1} = P$.

Stosujemy wcześniejsze Twierdzenie 11.1 które mówi, że

$$A_{\mathcal{E}'} = (P^T)^{-1} A_{\mathcal{E}} P^{-1} = P A_{\mathcal{E}} P^{-1}.$$

Ale to oznacza, że $A_{\mathcal{E}'}$ jest macierzą diagonalną. Czyli dostaliśmy to co chcieliśmy. □

**Przykład 11.2.** *Rozpatrzmy formę kwadratową*

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2.$$

*Chcemy znaleźć taką bazę ortonormalną, że w tej bazie dana forma jest w postaci kanonicznej. Forma $f$ jest zadana przez macierz*

$$A = [1, 1; 1, 2].$$

*Wystarczy nam w tym celu znaleźć bazę $\mathbb{R}^2$ składającą się z ortonormalnych wartości własnych (co jak wiemy, można zrobić gdyż $A$ jest macierzą symetryczną).*

*Wielomian charakterystyczny $A$ to*

$$W_A(\lambda) = \lambda^2 - \mathrm{tr} A\lambda + \det A = \lambda^2 - 3\lambda - 1.$$

*Wartości własne to*

$$\lambda = \frac{3 \pm \sqrt{5}}{2}.$$

*Wystarczy więc znaleźć wektory własne odpowiadające tym wartościom własnym i je znormalizować (gdyż skoro ma dwa różne wartości własne to wektory własne im odpowiadające są prostopadłe).*

**Zadanie 11.7.** *Proszę znaleźć bazy ortonormalne w których dana forma przyjmuje postać kanoniczną: a) $f(x_1, x_2) = x_1^2 + 2x_1x_2$; b) $f(x_1, x_2, x_3) = 2x_1x_3$.*

# Rozdział 12

# Grupy, pierścienie, ciała

## 12.1 Grupy

Jedną z najważniejszych matematycznych struktur jest grupa – dzieje się tak dzięki temu, że działanie w niej jest łączne. Zobaczmy najpierw co się dzieje gdy nie mamy łączności.

### 12.1.1 Jak dodawać na komputerze?

Przez zapis zmiennoprzecinkowy (floating point) danej liczby rzeczywistej rozumiem jej przedstawienie w postaci

$$\pm a \cdot 10^n, \text{ gdzie } a \in [0,1), n \in \mathbb{Z}.$$

Upraszczając, na komputerze zazwyczaj pamiętamy pierwsze 8 (lub 16) cyfr znaczących liczby $a$, zaś $n \in \{-128, \dots, 127\}$ (reprezentację numeryczną danej liczby oznaczamy przez $\mathrm{fl}(a)$ – od floating point). Liczby tej postaci będziemy nazwać reprezentowalnymi.

**Przykład 12.1.** *Dla prostoty załóżmy, że pamiętamy pierwsze trzy cyfry po przecinku. Wtedy dla przykładu* $\mathrm{fl}(\pi) = 0.314 \cdot 10^1$.

*Spróbujmy więc dodać dwie reprezentowalne liczby* $a = 0.123 \cdot 10^0$, $b = 0.232 \cdot 10^{-2}$, *do siebie:*

$$a + b = 0.123 \cdot 10^0 + 0.232 \cdot 10^{-2} = 0.12532 \cdot 10^0,$$

*i w konsekwencji wynikiem dodawania na komputerze będzie liczba reprezentowalna*

$$\mathrm{fl}(0.12532 \cdot 10^0) = 0.125 \cdot 10^0.$$

*Widzimy więc, że* $\mathrm{fl}(a+b) = (1+\varepsilon) \cdot (a+b)$, *gdzie* $\varepsilon$ *jest małe (zazwyczaj, tak jak w powyższym przypadku jest do liczba ujemna). Łatwo zauważyć, że najmniejszy błąd generujemy dodając do siebie liczby tego samego rzędu (wielkości).*

**Zadanie 12.1.** *Proszę pokazać, że dodawanie na komputerze nie jest łączne.*

W naszych rozważaniach dla prostoty zawężamy się do liczb nieujemnych. Jeżeli przyjrzymy się jak dodaje komputer, to w przybliżeniu jest to opisane za pomocą poniższego dziwnego dodawania (z pewnym $K \approx 1$).

**Definicja 12.1.** Niech $K > 0$ będzie dowolnie ustalone. Dla $a, b \in \mathbb{R}$ kładziemy

$$a \oplus_K b = K(a+b).$$

**Przykład 12.2.** *Załóżmy dla prostoty, że $K = 2$ (wtedy łatwiej wykonywać dodawanie): dodając do siebie dwie liczby uzyskujemy „dwa razy więcej” niż prawdziwy wynik.*

*Chcemy dodać do siebie pewną ilość liczb. Zauważmy, że im mniej uzyskamy w wyniku tego dodawania tym lepiej (tzn. bardziej jesteśmy zbliżeni do prawdziwego wyniku).*

*Dla przykładu: chcemy dodać do siebie liczby $1, 2, 3, 4$. Możemy je dodać następująco*

$$((1 \oplus 4) \oplus 2) \oplus 3 = (10 \oplus 2) \oplus 3 = 24 \oplus 3 = 54,$$

*zaś możemy dodać tak*

$$(1 \oplus 2) \oplus (3 \oplus 4) = 6 \oplus 14 = 40.$$

*Widzimy więc, że wynik dodawania zależy od kolejności w której dodajemy.*

Można pokazać następujące twierdzenie.

**Twierdzenie 12.1.** *Niech $K > 0$ będzie dowolne. Aby dodać do siebie za pomocą $\oplus_K$ z najmniejszym błędem liczby $x_1, \ldots, x_n > 0$, należy kolejno dodawać do siebie liczby najmniejsze.*

**Ćwiczenie 12.1** (myślowe)**.** *Załóżmy, że ktoś Państwu każe z możliwie małym/rozsądnym błędem dodać do siebie dużo liczb. Najprościej oczywiście dodawać kolejno, ale w momencie gdy ilość liczb jest za duża to może się okazać, że suma pierwszych $n$ liczb jest na tyle duża, że dodanie następnej nic nie zmienia (dla przykładu proszę sobie sprawdzić, jaki wynik otrzymamy dodając $10^6$ razy do siebie liczbę $1$, gdy mamy reprezentację jak w Przykładzie 12.1, to znaczy gdy pamiętamy trzy cyfry znaczące). Pytanie jak dodawać, aby uzyskać (w miarę) dobry wynik?*

*Wsk: rozsądnie jest dodawać kolejne pary, utworzyć nowy ciąg, i postępować rekurencyjnie (zamiast dwóch kolejnych liczb można oczywiście dodawać więcej, ważne aby nie przekroczyć tej ilości która generuje takie błędne efekty jak wyżej).*

### 12.1.2 Liczby Catalana

Zajmiemy się teraz pytaniem – ile jest potencjalnych sposobów dodania do siebie $n$ liczb za pomocą operacji $\oplus$ (nie zakładamy łączności) w momencie gdy nie zmieniamy kolejności ustawienia liczb.

**Przykład 12.3.** *Oznaczmy przez $N_n$ ilość możliwych wyników dodania do siebie $n$ liczb.*

*Gdy mamy zero liczb, to oczywiście niczego do siebie nie możemy dodać (mamy zero możliwych wyników), czyli przyjmujemy $c_0 = 0$. Oczywiście jedną liczbę możemy dodać na jeden sposób:*

$$a_1,$$

*czyli $c_1 = 1$. Gdy mamy dwie liczby, to oczywiście też możemy je dodać na jeden sposób: $a_1 \oplus a_2$, czyli $c_2 = 1$. Dla trzech liczb mamy już dwie możliwości, to znaczy*

$$(a_1 \oplus a_2) \oplus a_3 \text{ i } a_1 \oplus (a_2 \oplus a_3).$$

Tutaj łatwo można zauważyć, że kolejne wartości $c_n$ można wyliczyć rekurencyjnie: ciąg $a_1, \ldots, a_n$ możemy podzielić na dwa krótsze ciągi rozbijając następująco dla dowolnego $k \in \{1, \ldots, n-1\}$

$$(a_1, \ldots, a_k) \oplus (a_{k+1}, \ldots, a_n).$$

Pierwszy możemy posumować na $N_k$ sposobów, zaś drugi na $N_{n-k}$. W konsekwencji otrzymujemy

$$N_n = N_1 \cdot N_{n-1} + \ldots + N_{n-1} N_1. \tag{12.1}$$

Z powyższego wzoru łatwo wynika, że $N_4 = N_1 N_3 + N_2 N_2 + N_3 N_1 = 1 \cdot 2 + 1 \cdot 1 + 2 \cdot 1 = 5$.

Aby wyliczyć jawny wzór na $N_n$ będziemy potrzebować pojęcia funkcji generującej (bardzo fajny wykład na ten temat jest dostępny w internecie, `??? generatingfunctionology`).

Dla danego ciągu $N_n$ tworzymy funkcję generującą, czyli następujący nieskończony szereg potęgowy

$$y = N_0 + N_1 x + N_2 x^2 + N_3 x^3 + \ldots$$

Ponieważ $N_0 = 0$, $N_1 = 1$, na podstawie (12.1) otrzymujemy, że

$$y^2 = N_1 N_1 x^2 + (N_1 N_2 + N_2 N_1) x^3 + (N_1 N_3 + N_2 N_2 + N_3 N_1) x^4 + \ldots = N_2 x^2 + N_3 x^3 + N_3 x^4 + \ldots = y - N_1 x = y - x.$$

Czyli

$$y^2 - y - x = 0.$$

Otrzymaliśmy więc równanie na $y$ (to już połowa sukcesu)! Jedyne co pozostało, to na podstawie powyższego wzoru, wyliczyć jakim szeregiem potęgowym dane jest $y$. Oczywiście

$$y = \frac{1 \pm \sqrt{1-4x}}{2}.$$

Ponieważ wiemy, że $y(0) = 0$, dostajemy

$$y = \frac{1 - \sqrt{1-4x}}{2}.$$

Skorzystamy teraz z dwumianu Newtona który zachodzi także dla potęg rzeczywistych (i dla $a$ na moduł mniejszych od 1):

$$(1+a)^p = 1 + \binom{p}{1} a + \binom{p}{2} a^2 + \ldots,$$

czyli

$$(1-4x)^{1/2} = 1 + \frac{1/2}{1}(-4x) + \frac{1/2(1/2-1)}{1 \cdot 2}(-4x)^2 + \frac{1/2(1/2-1)}{1 \cdot 2 \cdot 3}(-4x)^3 + \ldots.$$

Dostajemy więc

$$y = \frac{1 - \sqrt{1-4x}}{2} = \sum_1^\infty \frac{1 \cdot 3 \cdot \ldots (2n-3)}{1 \cdot \ldots \cdot n} 2^{n-1} x^n.$$

W formie bardziej zwartej dostajemy

$$N_n = \frac{(2n-2)!}{n!(n-1)!} \quad \text{dla } n \geqslant 1. \tag{12.2}$$

Tak naprawdę to $N_n$ to przesunięte o jeden indeks liczby Catalana $(c_n)$(patrz `http://pl.wikipedia.org/wiki/Licz` lub `http://en.wikipedia.org/wiki/Catalan_number`), czyli $c_n = N_{n+1}$.

Pytanie które sobie warto postawić, to jakiego rzędu wielkość to $N_n$, albo inaczej mówiąc jakie jest asymptotyczne zachowanie ciągu $N_n$ (dla przykładu to nam da odpowiedź na pytanie: czy jesteśmy w stanie na komputerze sprawdzić wszystkie możliwe wyniki dla $n = 100$, czy też nie?). Do tego przyda się nam wzór Stirlinga:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

gdzie $\approx$ oznacza, że iloraz jednego czynnika przez drugi zmierza w nieskończoności do zera. Podstawiając do (12.2) dostajemy

$$\begin{aligned} N_n &\approx \frac{\sqrt{2\pi(2n-2)}((2n-2)/e)^{2n-2}}{\sqrt{2\pi n}(n/e)^n \sqrt{2\pi(n-1)}((n-1)/e)^{n-1}} \\ &\approx \frac{2^{2n-2}((n-1)/n)^{n-1}}{\sqrt{\pi n} e n} \approx \frac{1}{4\sqrt{\pi n} e^2} 4^n. \end{aligned}$$

Widzimy więc, że dla $n = 100$ sprawdzenie wszystkich możliwych wyników przewyższa możliwości nawet najszybszych komputerów. To powoduje, że tak ważną własnością jest łączność – gwarantuje ona, że wynik nie zależy od „ustawienia nawiasów”.

### 12.1.3 Łączność, czyli półgrupy i grupy

Niech dany będzie zbiór $G$ z działaniem dwuargumentowym $\circ : G \times G \to G$. Mówimy, że działanie $\circ$ jest łączne, jeżeli
$$(a \circ b) \circ c = a \circ (b \circ c) \quad \text{dla } a, b, c \in G.$$

Zbiór z działaniem łącznym nazywamy półgrupą. Jeżeli dodatkowo działanie to jest przemienne, to mówimy o półgrupie przemiennej (wtedy zazwyczaj stosujemy zapis addytywny). Kluczowym fakt mówiący o ważności półgrup jest:

**Twierdzenie 12.2.** *Niech $G$ będzie półgrupą i niech $a_1, \ldots, a_n \in G$. Wtedy wynik działania*

$$a_1 \circ \ldots \circ a_n$$

*nie zależy od sposobu ustawienia nawiasów.*

*Dowód.* Dowód robimy indukcyjnie ze względu na $n$. Dla $n = 1, 2$ nie ma problemu, bo $N_1 = 1$, $N_2 = 1$ (jest tylko jeden możliwy wynik). Jeżeli $n = 3$, to jest bezpośredni wniosek z definicji łączności. Załóżmy, że teza zachodzi dla wszystkich $k \leqslant n$ (gdzie $n \geqslant 3$), i pokażemy, że zachodzi dla wszystkich $k \leqslant n + 1$. To, że teza zachodzi dla $k \leqslant n$, oznacza, że możemy poprawnie zapisać $a_1 \circ \ldots \circ a_k$ (wartość wyniku nie zależy od poprawnego ustawienia nawiasów).

Weźmy więc ciąg długości $n+1$. Wtedy dla pewnego $k$ wykonanie działania możemy zapisać następująco

$$(a_1 \circ \ldots \circ a_k) \circ (a_{k+1} \circ \ldots \circ a_{n+1}),$$

gdzie nawiasy wewnątrz nie są potrzebne na podstawie założenia indukcyjnego. Ale teraz dla dowolnego $k' \neq k$ (dla prostoty przyjmujemy, że $k' < k$) mamy na podstawie łączności

$$(a_1 \circ \ldots \circ a'_k) \circ (a_{k'+1} \circ \ldots \circ a_{n+1}) = (a_1 \circ \ldots \circ a'_k) \circ ((a_{k'+1} \circ \ldots a_k) \circ (a_{k+1} \circ a_{n+1})) =$$

$$((a_1 \circ \ldots \circ a'_k) \circ (a_{k'+1} \circ \ldots a_k)) \circ (a_{k+1} \circ a_{n+1}) = (a_1 \circ \ldots \circ a_k) \circ (a_{k+1} \circ \ldots \circ a_{n+1}).$$

□

Jeżeli w półgrupie jest taki element $e$, że

$$a \circ e = e \circ a = a \quad \text{dla } a \in G,$$

to nazywamy go elementem neutralnym.

**Definicja 12.2.** Niech $G$ będzie półgrupą z elementem neutralnym. Jeżeli dla każdego elementu $a \in G$ istnieje element odwrotny, to jest takie $b \in G$ (oznaczamy przez $a^{-1}$), że

$$ab = ba = e,$$

to mówimy, że $G$ jest grupą. Jeżeli działanie jest przemienne, to mówimy, że $G$ jest grupą abelową.

**Zadanie 12.2.** *Zachodzą następujące własności:*

- *element odwrotny jest jednoznacznie wyznaczony, tzn. jeżeli $b$ i $b'$ są takie, że $ab = e = ab'$ i $b'a = e = ba$, to $b = b'$;*
- *$(ab)^{-1} = b^{-1}a^{-1}$.*

**Przykład 12.4.** *Poznaliśmy już dużo grup (większość abelowych), dla przykładu wymienię część z nich: $(\mathbb{R}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}_p, +_p)$, permutacje (ze składaniem), $(\mathbb{R} \setminus \{0\}, \cdot)$, odwracalne macierze kwadratowe z mnożeniem, macierze ustalonego wymiaru z dodawaniem, wielomiany z dodawaniem.*

Przyda się nam jeszcze pojęcie homomorfizmu grup: mówimy, że $f : (G, \circ) \to (H, \bullet)$ jest homomorfizmem, jeżeli zachowuje działanie, to znaczy:

$$f(g_1 \circ g_2) = f(g_1) \bullet f(g_2) \quad \text{dla } g_1, g_2 \in G.$$

Jeżeli $f$ dodatkowo jest surjekcją/iniekcją/bijekcją, to nazywamy go epi/mono/izo-morfizmem. Taka oczywista uwaga: jeżeli dwie grupy są izomorficzne, to znaczy, że mają tą samą liczbę elementów.

**Zadanie 12.3.** *Proszę sprawdzić, że homomorfizm grup przeprowadza element neutralny na element neutralny.*

**Zadanie 12.4.** *Proszę sprawdzić, czy grupy a) $\mathbb{Z}_4$ i $\mathbb{Z}_2 \oplus \mathbb{Z}_2$; b) $\mathbb{Z}_6$ i $\mathbb{Z}_2 \oplus \mathbb{Z}_3$; są izomorficzne.*

Ogólnie trudno sprawdzić, że coś jest grupą – najgorzej jest ze sprawdzeniem warunku łączności. Na szczęście często można tego uniknąć, jeżeli nasz obiekt jest podzbiorem jakiejś grupy. Niech $(G, \cdot)$ będzie grupą, mówimy, że $(H, \cdot)$ jest podgrupą $(G, \cdot)$ jeżeli $H \subset G$ i jeżeli $H$ jest grupą ze względu na to samo działanie które jest w $G$.

**Obserwacja 12.1.** *Niech $G$ będzie grupą i niech $H \subset G$. Wtedy $H$ jest podgrupą $G$ wtedy i tylko wtedy gdy*

$$ab \in H, a^{-1} \in H \quad \text{dla } a, b \in H.$$

*Czyli aby sprawdzić, że coś jest podgrupą, nie musimy sprawdzać łączności!*

*Jako przykład zastosowania, załóżmy, że mamy jako zadanie pokazać, że $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ tworzy grupę ze względu na mnożenie. Ponieważ wiemy, że $(\mathbb{C} \setminus \{0\}, \cdot)$ jest grupą, wystarczy sprawdzić, czy $|z_1 z_1| = 1$ i $|z_1|^{-1} = 1$ dla $z_1, z_2 \in S^1$, ale to jest oczywiście prawda.*

Mając dwie grupy $(G, +), (H, +)$ możemy utworzyć ich sumę prostą, zapisujemy $G \oplus H$, gdzie jest to po prostu iloczyn kartezjański tych grup z działaniem po współrzędnych, czyli $G \times H$ z działaniem

$$(g_1, h_1) + (g_2, h_2) := (g_1 + g_2, h_1 + h_2).$$

Oczywiście widać, że tak zdefiniowany obiekt jest grupą.

## 12.2 Pierścienie

W pierścieniu oprócz dodawania mamy jeszcze operację mnożenia. Formalnie, pierścień to $(P, +, \cdot)$, gdzie $(P, +)$ to grupa abelowa, $(P, \cdot)$ to półgrupa (często zakłada się domyślnie przemienność) i spełnione są warunki wiążące ze sobą dodawanie i mnożenie, a mianowicie rozdzielność mnożenia względem dodawania:

$$(a + b)c = ac + bc, \; a(b + c) = ab + ac.$$

Przykłady pierścieni (z naturalnymi działaniami): $\mathbb{Z}$, $\mathbb{Z}_n$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}[\sqrt{p}]$ gdzie $p$-pierwsza, wielomiany, macierze kwadratowe ustalonego wymiaru. Bardzo ważną rolę w szeroko rozumianej teorii aproksymacji odgrywają pierścienie wielomianów, dlatego postaramy się je dokładnie omówić w tej sekcji.

Analogicznie jak w przypadku grup, mówimy, że dane odwzorowanie jest homomorfizmem pierścieni, jeżeli zachowuje operację dodawania i mnożenia, to znaczy

$$f(a + b) = f(a) + f(b) \text{ i } f(ab) = f(a)f(b).$$

### 12.2.1 Wielomiany

Wielomian $W$ o współczynnikach z pierścienia $P$ zapisujemy

$$W(x) = a_0 + a_1x + \ldots + a_nx^n.$$

Będę czasami oznaczał wielomiany o współczynnikach z $P$ przez $P[x]$. Jeżeli $a_n \neq 0$, to mówimy, że stopień tego wielomianu jest równy $n$ (zapisujemy $\deg W = n$). Wielomiany dodajemy dodając współczynniki przy tej samej potędze (mnożymy mnożąc każdy współczynnik z każdym). W konsekwencji przy standardowym mnożeniu dwóch wielomianów stopnia $n$ wykonujemy $n^2$ razy operację mnożenia[1]

Mówimy, że pierścień jest z jedynką, jeżeli istnieje element neutralny względem mnożenia (oznaczamy często jako 1). Jeżeli mamy pierścień z jedynką to mówimy, że element $p \in P$ jest odwracalny, jeżeli istnieje takie $q$, że $pq = qp = 1$.

## 12.3 Pierścień (ciało) kwaternionów

Kwaterniony są chyba najważniejszym przykładem pierścienia nieprzemiennego z dzieleniem (czasami mówi się o ciele nieprzemiennym). W swojej idei są bardzo podobne do liczb zespolonych, sa jedynie bardziej skomplikowane. Są przydatne w grafice do wykonywania i zapisywania obrotów (w związku z tym klasa obsługująca kwaterniony jest zaimplementowana w bibliotece DirectX). Kwaternion $w$ to czwórka $w_0 + w_1i + w_2j + w_3k$, gdzie $w_i \in \mathbb{R}$. Dodaje się kwaterniony po współrzędnych. Aby wymnożyć kwaterniony potrzebne są następujące wzory (tak zwane wzory Hamiltona):

$$i^2 = j^2 = k^2 = ijk = -1.$$

Na ich podstawie łatwo wyliczamy pozostałe wzory

$$ij = k, jk = i, ik = -j,\ ji = -k, kj = -i, ki = j.$$

Dla kwaterniona $w = w_0 + w_1i + w_2j + w_3k$, przez jego część rzeczywistą $\mathrm{real}w$ rozumiem $w_0$, zaś przez część wektorową $\mathrm{vect}w$ rozumiem $w_1i + w_2j + w_3k$ (utożsamiam z punktem w $\mathbb{R}^3$ o współrzędnych $(w_1, w_2, w_3)$). Mówię, że kwaternion jest rzeczywisty, jeżeli $\mathrm{vect}w = 0$, i wektorowy, jeżeli $\mathrm{real}v = 0$.

Przez $\bar{w}$, kwaternion sprzężony do $w$, rozumiem kwaternion $w_0 - w_1i - w_2j - w_3k = \mathrm{real}w - \mathrm{vect}w$. Przez normę (moduł) kwaternionu $w$ rozumiem $\sqrt{w_0^2 + w_1^2 + w_2^2 + w_3^2}$ Ponieważ jak łatwo sprawdzić dla kwaternionów zachodzą wzory

$$|vw| = |v| \cdot |w|, v\bar{v} = \bar{v}v = |v|^2,$$

więc $v^{-1} = \bar{v}/|v|^2$. Mówimy, że kwaternion jest znormalizowany jeżeli ma normę jeden (normalizacja to dzielenie wektora przez jego normę). Proszę zauważyć, że jeżeli kwaternion wektorowy $u$ jest znormalizowany, to dla każdego $\varphi \in \mathbb{R}$ kwaternion $\cos\varphi + \sin\varphi u$ jest także znormalizowany (trywialna konsekwencja z $\cos^2 + \sin^2 = 1$).

**Zadanie 12.5** (kwaterniony)**.** *Proszę wyliczyć a)* $(1 + i + j + k)(1 - i + j - k)$*; b)* $(1 - j)(1 + i + k)^{-1}$ *c) dokonaj normalizacji kwaterniona* $2i + j + 2k$.

Mówimy, że dwa kwaterniony są równoległe (prostopadłe), jeżeli ich części wektorowe są równoległe (prostopadłe). Łatwo można zauważyć, że mnożenie dwóch równoległych kwaternionów jest przemienne.

Proszę sprawdzić, że dla wektorowych kwaternionów $v$ i $w$ zachodzi wzór

$$v \cdot w = -v \circ w + v \times w, \tag{12.3}$$

[1]można to znacznie przyspieszyć: najstarszy algorytm został wymyślony w 1960 roku przez Karatsubę i ma złożoność $n^{\log_2 3}$, bardzo polecam przeczytać `en.wikipedia.org/Karatsuba_algorithm.htm`, najszybsze algorytmy mnożenia wielomianów używają FFT.

gdzie $v \circ w$ to iloczyn skalarny, zaś $v \times w$ to iloczyn wektorowy wektorów $v$ i $w$, czyli

$$v \times w = \begin{vmatrix} i & j & k \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Na podstawie wzoru (12.3) widzimy, że jeżeli dwa wektorowe kwaterniony $v, w$ są prostopadłe, to

$$v \cdot w = -w \cdot v.$$

**Obserwacja 12.2** (nie było na wykładzie i nie obowiązuje). *Niech $u$ będzie wektorowym kwaternionem znormalizowany i niech kąt $\varphi \in \mathbb{R}$ będzie dowolnie wybrany.*

*Weźmy dowolny wektorowy kwaternion $v$ który jest prostopadły do $u$ i pomnóżmy go z lewej strony przez $\cos\varphi + \sin\varphi u$. Wtedy na podstawie wzoru* (12.3)

$$(\cos\varphi + \sin\varphi u)v = \cos\varphi \cdot v + \sin\varphi \cdot (u \times v).$$

*Ale $u \times v$ to jest wektor prostopadły do $u$ i $v$ o długości równej długości $v$. W konsekwencji powyższy wzór oznacza, że dokonaliśmy (w płaszczyźnie prostopadłej do $u$) obrotu wektora $v$ o kąt $\varphi$.*

*Analogicznie, jeżeli pomnożymy $v$ z prawej strony przez wektor sprzężony (i równocześnie odwrotny) do $\cos\varphi + \sin vau$, czyli przez $\cos\varphi - \sin vau$ to dostaniemy*

$$v(\cos\varphi - \sin\varphi u) = \cos\varphi \cdot v - \sin\varphi \cdot (v \times u) = \cos\varphi \cdot v + \sin\varphi \cdot (u \times v).$$

*Czyli także dokonamy obrotu $v$ względem prostej rozpiętej na $u$ o kąt $\varphi$!*

*Weźmy teraz dowolny wektor (kwaternion wektorowy) $w$ (nie zakładamy prostopadłości do $w$). Wtedy możemy go przedstawić (proszę sobie sprawdzić!) w postaci sumy $w = w_{||} + w_{\perp}$, gdzie $w_{||}$ jest kwaternionem równoległym do $u$ a $w_{\perp}$ jest kwaternionem wektorowym prostopadłym do $u$. Pomnóżmy teraz $w$ lewostronnie przez* $(\cos\varphi + \sin\varphi u)$ *i prawostronnie przez* $(\cos\varphi - \sin\varphi u)$ *(korzystamy w poniższych rachunkach z tego, że mnożenie kwaternionów równoległych jest przemienne):*

$$(\cos\varphi + \sin\varphi u)w(\cos\varphi - \sin\varphi u) = (\cos\varphi + \sin\varphi u)(w_{||} + w_{\perp})(\cos\varphi - \sin\varphi u) =$$

$$(\cos\varphi + \sin\varphi u)w_{\perp}(\cos\varphi - \sin\varphi u) + (\cos\varphi + \sin\varphi u)w_{||}(\cos\varphi - \sin\varphi u) ==$$

$$(\cos\varphi + \sin\varphi u)w_{\perp}(\cos\varphi - \sin\varphi u) + w_{||}.$$

*Teraz* $(\cos\varphi + \sin\varphi u)w_{\perp}(\cos\varphi - \sin\varphi u)$ *to oczywiście na podstawie wcześniejszych rozważań obrót $w_{\perp}$ o kąt $2\varphi$ względem prostej rozpiętej na wektorze $u$ (czyli część równoległą do $u$ nie zmieniliśmy, zaś część prostopadłą obróciliśmy o kąt $2\varphi$).*

*Podsumowując, dla znormalizowanego kwaternionu wektorowego $u$ i dowolnego kąta $\varphi$, operacja*

$$w \to (\cos\varphi + \sin\varphi u)w(\cos\varphi - \sin\varphi u) \tag{12.4}$$

*dokonuje obrotu $w$ o kąt $2\varphi$ względem prostej rozpiętej na $u$.*

**Zadanie 12.6** (kwaterniony). *Proszę dokonać obrotu punktu* $(1, 0, 0)$ *względem prostej przechodzącej przez punkt* $(2, -1, 2)$ *o kąt* $\varphi = 2\pi/3$.

## 12.4 Ciała

Ciałem nazywamy zbiór $\mathbb{F}$ z działaniami $+$ i $\cdot$, takimi, że $(\mathbb{F},+)$ jest grupą abelową, $(\mathbb{F},\cdot)$ jest przemienną półgrupą z jedynką taką, że każdy element niezerowy jest odwracalny oraz zachodzi rozdzielność mnożenia względem dodawania. Przykładami ciał są $\mathbb{Q}$, $\mathbb{Q}[\sqrt{p}]$ (gdzie $p$ jest pierwsze) $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}_p$ ($p$ – liczba pierwsza). Ciała można budować z pierścieni przemiennych bez dzielników zera tworząc tak zwane ciała ułamków (w ten sposób na przykład można zdefiniować $\mathbb{Q}$ jako ciało ułamków pierścienia $\mathbb{Z}$ czy funkcje wymierne powstałe jako ciało ułamków pierścienia wielomianów).

W kryptografii ważne zastosowanie mają ciała skończone, a w szczególności ciało $\mathbb{Z}_p$ (umiejętność znajdowania liczby odwrotnej w tym ciele jest przydatna w rozkładaniu liczb na czynniki pierwsze).

**Zadanie 12.7.** *Proszę znaleźć a)* $3^{-1}$ *w* $\mathbb{Z}_7$*; b)* $5^{-1}$ *w* $\mathbb{Z}_{23}$*.*

Znajomość informacji o ciałach przydatna jest także w usuwaniu niewymierności z mianownika. Ponieważ teoria którą się stosuje wymaga większego wprowadzenia którego nie będę robił, opiszę tylko sposób postępowania na jednym przykładzie.

**Przykład 12.5.** *Załóżmy, że chcemy się pozbyć niewymierności z mianownika w ułamku*

$$\frac{1}{1+\sqrt{2}+\sqrt{3}}.$$

*Postępujemy następująco: znajdujemy najmniejsze ciało zawierające liczby wymierne i* $\sqrt{2}$ *i* $\sqrt{3}$ *(oznaczam je przez* $\mathbb{Q}[\sqrt{2},\sqrt{3}]$*). Aby je znaleźć, wymnażam przez siebie kolejne pierwiastki do momentu gdy za każdym razem trafię do czegoś co już potrafię wygenerować za pomocą kombinacji liniowej poprzednich o współczynnikach z* $\mathbb{Q}$*. W naszym przypadku łatwo widać, że wystarczy rozpatrzeć tylko* $\sqrt{2}$, $\sqrt{3}$, $\sqrt{6}$*. Okazuje się, że każdy element z* $\mathbb{Q}[\sqrt{2},\sqrt{3}]$ *można zapisać w postaci*

$$q_0 + q_1\sqrt{2} + q_2\sqrt{3} + q_3\sqrt{6} \quad dla\ q_i \in \mathbb{Q}.$$

*I w szczególności także dla pewnych* $q_i \in \mathbb{Q}$

$$\frac{1}{1+\sqrt{2}+\sqrt{3}} = q_0 + q_1\sqrt{2} + q_2\sqrt{3} + q_3\sqrt{6}.$$

*Po wymnożeniu obu stron przez* $1+\sqrt{2}+\sqrt{3}$ *dostajemy równanie (jeżeli się nie pomyliłem)*

$$1 = (q_0+2q_1+3q_2)1 + (q_1+q_0+3q_3)\sqrt{2} + (q_2+2q_3+q_0)\sqrt{3} + (q_3+q_2+q_1)\sqrt{6}.$$

*Aby wyliczyć te współczynniki wystarczy teraz (oczywiście jeżeli się nam uda) rozwiązać układ 4 równań z 4 niewiadomymi*

$$\begin{cases} 1 = q_0+2q_1+3q_2 \\ 0 = q_1+q_0+q_3 \\ 0 = q_2+2q_3+q_0 \\ 0 = q_3+q_2+q_1 \end{cases}.$$

**Zadanie 12.8.** *Proszę pozbyć się niewymierności z mianownika:*

$$1/(1+2\sqrt[3]{2}+\sqrt[3]{4}),\ 1/(1+2\sqrt{2}+\sqrt{3}).$$

# Rozdział 13

# Operatory

## 13.1 Odwzorowania liniowe

Niech $X, Y$ będą przestrzeniami wektorowymi nad ciałem $\mathbb{F}$. Mówimy, że $A : X \to Y$ jest liniowe jeżeli jest addytywne i jednorodne:

$$A(x+y) = A(x) + A(y) \text{ dla } x, y \in X,\ A(\alpha x) = \alpha A(x) \text{ dla } x \in X, \alpha \in \mathbb{F}.$$

Mówimy, że odwzorowanie $A$ jest

- monomorfizmem, jeżeli jest iniekcją;
- epimorfizmem, jeżeli jest surjekcją;
- izomorfizmem, jeżeli jest bijekcją;
- endomorfizmem, jeżeli $X = Y$;
- funkcjonałem liniowym, jeżeli $Y = \mathbb{F}$.

Odwzorowania liniowe możemy w sposób naturalny dodawać do siebie i mnożyć przez skalary. W konsekwencji, $L(X, Y)$, zbiór wszystkich odwzorowań liniowych z $X$ do $Y$ jest przestrzenią wektorową.

**Zadanie 13.1.** *Proszę pokazać, że wykres (ang. graph) odwzorowania liniowego $A \in L(X, Y)$ jest podprzestrzenią wektorową przestrzeni $X \times Y$.*

**Twierdzenie 13.1.** *Jeżeli dwa odwzorowania liniowe pokrywają się na ustalonym zbiorze generującym to są sobie równe.*

**Twierdzenie 13.2.** *Odwzorowanie liniowe wystarczy zadać na bazie. To znaczy jeżeli $e_i$ baza przestrzeni $X$, a $f_i$ dowolne elementy z $Y$, wtedy istnieje dokładnie jedno liniowe $A : X \to Y$ takie, że $Ae_i = f_i$.*

**Zadanie 13.2** (odwz. liniowe)**.** *Niech $A : W_2(\mathbb{R}) \to \mathbb{R}^3$ będzie dany wzorem*

$$A(w) = (w(-1), w(0), w(1)).$$

*Czy $A$ jest odwzorowaniem liniowym? Jeżeli tak, proszę znaleźć jądro i obraz. Czy to jest epi/mono/izomorfizm? Jeżeli izomorfizm, to proszę wskazać odwzorowanie odwrotne.*

## 13.2 Macierz danego odwzorowania

Niech $X, Y$ i $A : X \to Y$. Zakładamy, że przestrzenie są skończenie wymiarowe i $e_1, \ldots, e_n$, $f_1, \ldots, f_m$ - bazy odpowiednio $X$ i $Y$. Wtedy $Ae_i$ możemy przedstawić w postaci

$$Ae_i = \alpha_{1i} f_1 + \ldots + a_{mi} f_m \text{ dla } i = 1, \ldots, m.$$

Jeżeli te współczynniki ustawimy kolumnowo w macierz (będzie to macierz o $n$-kolumnach odpowiadających wektorom $Ae_1, \ldots, Ae_n$ i $m$-wierszach):

$$M_A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}$$

to dostaniemy macierz danego odwzorowania (przy ustalonych bazach w $X$ i $Y$).

UWAGA: jeżeli $A$ jest endomorfizmem, to domyślnie (jeżeli nie jest powiedziane inaczej) to bierzemy w $X$ jedną bazę (zarówno w dziedzinie jak i obrazie)!!!

**Zadanie 13.3** (odwz. liniowe)**.** *Niech $X = W_4(\mathbb{R})$, $Y = W_1(\mathbb{R})$. Dla $w \in X$ definiujemy $Aw$ jako resztę z dzielenia $w$ przez $1+x+x^2$. Czy $A$ jest operatorem liniowym? Jeżeli tak, czy $A$ jest epi/mono/izo-morfizmem? Proszę znaleźć macierz $A$ gdy w $X$ mamy bazę $1, 1+x, 1+x+x^2, 1+x+x^2+x^3+x^4$, a w $Y$ $1, x$.*

**Zadanie 13.4** (odwz. liniowe)**.** *Niech $X = M_{2\times 2}(\mathbb{R})$ (macierze $2 \times 2$ o współczynnikach z $\mathbb{R}$). Definiujemy $L : X \to X$ za pomocą wzoru*

$$L(A) = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \cdot A \quad \textit{dla } A \in M_{2\times 2}(\mathbb{R}).$$

*Proszę pokazać, że $L$ jest odwzorowaniem liniowym. Czy jest to epi/mono/izo-morfizm? Jeżeli izomorfizm, to proszę znaleźć odwzorowanie odwrotne.*

*Proszę znaleźć macierz $L$ w bazie $e_i$ danej przez*

$$e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, e_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$e_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, e_4 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

**Zadanie 13.5** (zmiana bazy)**.** *Niech $d : X = W^2(\mathbb{R}) \to W^1(\mathbb{R}) = Y$ będzie operatorem pochodnej. W $X$ bierzemy bazę $e_i$ daną przez $1, 1+x, 1+x+x^2$ a w $Y$ $f_j$ daną przez $1-x, 1+x$. Proszę znaleźć macierz $d$ w tych bazach.*

**Zadanie 13.6.** *Niech $A : W_2(\mathbb{R}) \to \mathbb{R}^3$ będzie dany wzorem*

$$A(w) = (w(0), w'(0), w''(0)).$$

*Czy $A$ jest epi/mono/izo-morfizmem? Jeżeli izomorfizm, proszę znaleźć odwzorowanie odwrotne.*

*W $W_2(\mathbb{R})$ jako bazę bierzemy $1, x, x^2$, w $\mathbb{R}^3$: $(1,1,1)$, $(1,1,0)$, $(1,0,0)$. Proszę wypisać macierz tego odwzorowania w tych bazach.*

**Zadanie 13.7.** *Niech $X = \mathbb{C}^3$ i rozpatrzmy endomorfizm $A : X \to X$ zadany w bazie kanonicznej przez macierz*

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

*Rozpatrujemy w $X$ bazę $e_i = (q_i^0, q_i^1, q_i^2, q_i^3)$, gdzie $q_i$ to kolejne pierwiastki czwartego stopnia z 1. Proszę znaleźć macierz endomorfizmu $A$ w tej bazie.*

**Dygresja 13.1** (zmiana bazy)**.** *Załóżmy, że mamy dane bazy $e_i$ w $X$ i $f_j$ w $Y$ i znamy $M_A$ macierz danego odwzorowania $A : X \to Y$ w tych bazach. Załóżmy, że (z jakichś powodów) bierzemy nowe bazy $e_i'$ w $X$ i $f_j'$ w $Y$. Pytanie w jaki sposób uzyskać wzór na $M_A'$ macierz odwzorowania $A$ w tych nowych bazach? Przy pomocy macierzy przejścia jest to dosyć łatwe. Niech $P_X$ oznacza macierz przejścia z bazy $e_i$ do $e_i'$, a $P_Y$ macierz przejścia z $f_i$ do $f_i'$. Bierzemy dowolny punkt $x$, załóżmy, że w bazie $e_i'$ ma współrzędne $[\lambda_1, \ldots, \lambda_n]^T$. Wtedy po przejściu do bazy $e_1, \ldots, e_n$ ma współrzędne $P_X[\lambda_1, \ldots, \lambda_n]^T$ (gdyż $P$ wyraża $e_i'$ za pomocą kombinacji $e_i$). Ponieważ znamy teraz współczynniki $x$ w bazie $e_i$, znając $M_A$ możemy wyliczyć współczynniki $Ax$ w bazie $f_i$ za pomocą wzoru $M_A P_X[\lambda_1, \ldots, \lambda_n]^T$. Teraz pozostało znaleźć współczynniki tego punktu w bazie $f_i'$, ale to jak wiemy wcześniej jest wzięcie $P_Y^{-1} M_A P_X[\lambda_1, \ldots, \lambda_n]^T$. Konkludując: macierz $M_A'$ wyraża się wzorem*

$$M_A' = P_Y^{-1} M_A P_X.$$

**Zadanie 13.8.** *W $\mathbb{R}^2$ i $\mathbb{R}^3$ bierzemy bazy kanoniczne. Macierz $M_A$ naszego odwzorowania $A : \mathbb{R}^2 \to \mathbb{R}^3$ w tych bazach jest dana wzorem*

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

*W $\mathbb{R}^2$ rozpatrujemy teraz bazę $[1,1]^T, [1,-1]^T$, a w $\mathbb{R}^3$ bazę $[1,1,1]^T, [1,1,0]^T, [1,0,0]^T$. Proszę znaleźć $M_A'$ macierz odwzorowania $A$ w tych nowych bazach (na dwa sposoby: korzystając z powyższego rozumowania, oraz bezpośrednio).*

## 13.3 Jądro i obraz

Niech $A : X \to Y$ będzie liniowe. Definicja jądra $A$:

$$\mathrm{kernel} A := \{x \in X : Ax = 0\}.$$

Obraz odwzorowania $A$ to

$$\mathrm{image} A := \{y \in Y \mid \exists x \in X : y = Ax.$$

Jądro - liczenie jest łatwe - jest to de facto rozwiązywania układu równań liniowych (najlepiej sprowadzić do postaci schodkowej zredukowanej przy pomocy operacji elementarnych na wierszach).

Obraz odwzorowania liczymy sprowadzając macierz tego odwzorowania do postaci schodkowej przy pomocy operacji elementarnych na kolumnach (wtedy dostajemy od razu bazę obrazu).

**Zadanie 13.9** (baza)**.** *Niech $A : X = \mathbb{R}^3 \to \mathbb{R}^3 = Y$ będzie zadany w bazie kanonicznej przy pomocy macierzy*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

*a) Proszę znaleźć bazę* $\mathrm{kernel} A = \{x \in X : Ax = 0\}$ *i bazę* $\mathrm{image} A = \{Ax : x \in X\}$.
*b) Proszę znaleźć bazę ortogonalną* $\mathrm{kernel} A$ *i* $\mathrm{image} A$.

Bardzo ważne jest twierdzenie które łączy nas obraz z jądrem.

**Twierdzenie 13.3.** *Niech $A : X \to Y$ będzie odwzorowaniem linowym. Wtedy*

$$X/\mathrm{kernel} A \text{ jest izomorficzne z } \mathrm{image} A.$$