Geometria i Algebra Liniowa
(dla I-go roku informatyki WMIM UW)

Leszek Plaskota
Instytut Matematyki Stosowanej i Mechaniki
Uniwersytet Warszawski

styczeń 2009

Spis treści

1 Grupy i ciała, liczby zespolone **3**

1.1 Podstawowe struktury algebraiczne . . . 3

1.1.1 Grupa . . . 3

1.1.2 Ciało . . . 5

1.2 Ciało liczb zespolonych . . . 6

1.2.1 Definicja . . . 6

1.2.2 Postać trygonometryczna . . . 7

1.2.3 Wzór de Moivre'a . . . 8

1.2.4 Pierwiastki z jedynki . . . 9

1.2.5 Sprzężenie . . . 9

1.3 Wielomiany . . . 10

1.3.1 Algorytm Hornera . . . 10

1.3.2 Zasadnicze twierdzenie algebry . . . 10

2 Macierze liczbowe **13**

2.1 Podstawowe definicje . . . 13

2.1.1 Macierze szczególnych formatów . . . 13

2.1.2 Podział blokowy . . . 14

2.2 Działania na macierzach . . . 14

2.2.1 Podstawowe działania . . . 14

2.2.2 Mnożenie macierzy . . . 15

2.2.3 Mnożenie macierzy w postaci blokowej . . . 17

2.3 Dalsze oznaczenia . . . 18

2.3.1 Macierze trójkątne i jednostkowe . . . 18

2.3.2 Układ równań jako równanie macierzowe . . . 19

2.4 Macierze nieosobliwe . . . 19

2.4.1 Grupa macierzy nieosobliwych . . . 19

2.4.2 Warunek nieosobliwości macierzy . . . 21

2.4.3 Permutacje 21

3 Normy wektorów i macierzy 25

3.1 Ogólna definicja normy 25
3.2 Normy wektorów 26
3.2.1 Normy p-te 26
3.2.2 Pożyteczne (nie)równości 27
3.3 Normy macierzy 28
3.3.1 Normy p-te 28
3.3.2 Pożyteczne (nie)równości 29
3.3.3 Norma Frobeniusa 31

4 Przestrzenie liniowe 35

4.1 Przestrzenie i podprzestrzenie 35
4.1.1 Definicja i podstawowe własności 35
4.1.2 Podprzestrzenie liniowe 36
4.2 Baza i wymiar przestrzeni 37
4.2.1 Liniowa (nie)zależność 37
4.2.2 Baza i wymiar, twierdzenie Steinitza 39
4.2.3 Przykłady 40
4.3 Sumy i sumy proste 41
4.3.1 Suma (prosta) dwóch podprzestrzeni 41
4.3.2 Suma (prosta) w ogólnym przypadku 43
4.4 Izomorfizm przestrzeni 44
4.5 Warstwy modulo $\mathcal{Y}$ 45
4.5.1 Definicja 45
4.5.2 Przestrzeń warstw 46

5 Obraz, rząd i jądro macierzy 49

5.1 Obraz i rząd macierzy 49
5.1.1 Rząd kolumnowy i rząd wierszowy 49
5.1.2 Rząd macierzy 50
5.2 Przestrzeń zerowa (jądro) macierzy 51
5.3 Rozkład względem obrazu i jądra 52

6 Funkcjonały liniowe 55

6.1 Funkcjonały 55
6.1.1 Definicja i przykłady 55

6.1.2 Przestrzeń sprzężona 56
6.2 Refleksywność 57
6.2.1 Równość $\mathcal{X}$ i $\mathcal{X}^{**}$ 57
6.2.2 Przykłady 58
6.3 Rozszerzenie rachunku macierzy 59
6.3.1 Macierze wektorów i funkcjonałów 59
6.3.2 Postać macierzowa izomorfizmów 60

7 Układy równań liniowych 63
7.1 Zbiór rozwiązań 63
7.1.1 Twierdzenie Kroneckera-Capelliego 63
7.1.2 Zbiór rozwiązań jako warstwa 64
7.1.3 Układy nieosobliwe 65
7.2 Efektywna metoda rozwiązania 65
7.2.1 Ogólny schemat 66
7.2.2 Eliminacja Gaussa 66
7.2.3 Konstrukcja rozwiązania ogólnego 68
7.3 Interpretacja macierzowa eliminacji 69
7.3.1 Analiza operacji elementarnych 69
7.3.2 Rozkład trójkątno-trójkątny macierzy 71
7.4 Eliminacja bez przestawień 72

8 Przekształcenia liniowe 75
8.1 Podstawowe pojęcia i własności 75
8.1.1 Obraz, jądro i rząd przekształcenia 75
8.1.2 Przykłady 77
8.1.3 Różnowartościowość 77
8.1.4 Przestrzeń przekształceń liniowych 78
8.2 Macierz przekształcenia liniowego 78
8.2.1 Definicja 78
8.2.2 Izomorfizm $\mathcal{L}in(\mathcal{X}, \mathcal{Y})$ i $\mathbf{K}^{m,n}$ 79
8.3 Dalsze własności macierzy przekształceń 80
8.3.1 Obraz i jądro przekształcenia/macierzy 80
8.3.2 Zmiana bazy 80
8.3.3 Złożenie przekształceń 81

9 Wyznacznik macierzy **83**
9.1 Definicja i pierwsze własności 83
9.2 Wyznacznik a operacje elementarne 84
9.2.1 Permutacja kolumn 84
9.2.2 Kombinacja liniowa kolumn 86
9.3 Dalsze własności wyznaczników 87
9.3.1 Wyznacznik iloczynu macierzy 87
9.3.2 Wyznacznik macierzy nieosobliwej i transponowanej . . 88
9.4 Definicja kombinatoryczna wyznacznika 89
9.5 Wzory Cramera 90

10 Formy dwuliniowe i kwadratowe **93**
10.1 Formy dwuliniowe 93
10.1.1 Definicja i przykłady 93
10.1.2 Macierz formy dwuliniowej 94
10.2 Twierdzenie Sylwester'a 96
10.3 Formy kwadratowe 97
10.3.1 Określoność formy kwadratowej 97
10.3.2 Kryterium Sylwester'a 98

11 Przestrzenie Euklidesowe **101**
11.1 Definicja, iloczyn skalarny i norma 101
11.2 Rzut prostopadły 102
11.2.1 Zadanie aproksymacji 102
11.2.2 Twierdzenie o rzucie prostopadłym 103
11.3 Układy ortogonalne 104
11.3.1 Macierz Grama 104
11.3.2 Ortogonalizacja Grama-Schmidta 105
11.3.3 Rozkład ortogonalno-trójkątny macierzy 107

Nota autora

Niniejszy skrypt został napisany z myślą o studentach pierwszego roku informatyki na *Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego*, uczęszczających na semestralny wykład pt. "Geometria z algebrą liniową". Skrypt powstawał równolegle z prowadzonym wykładem, a stąd zawiera treści przekazywane na wykładzie i praktycznie tylko te treści. Powinien więc, i takie było moje zamierzenie, stanowić dla studentów podstawowy przewodnik po w/w wykładzie.

Skrypt ma swoją historię. W swoim czasie prof. Andrzej Kiełbasiński prowadził na tym samym wydziale i także dla studentów informatyki wykład pt. "Algebra liniowa i jej metody obliczeniowe". Pozostałością po tym wykładzie są, m.in., obszerne odręczne notatki prowadzącego. Notatki te wydały mi się (i nie tylko mi) na tyle cenne, że stały się podstawą do przygotowania bieżącego wykładu. Ponieważ, w wyniku reformy studiów, wykład został ograniczony do jednego semestru, materiał musiał być z konieczności mocno skrócony. Jednak duch wykładu i w szczególności *oryginalna* notacja wprowadzona przez prof. Kiełbasińskiego pozostały, mam nadzieję, niezmienione.

Skrypt ma dynamiczny charakter i jest na bieżąco poprawiany i modyfikowany.

Leszek Plaskota
Warszawa, styczeń 2009

Rozdział 1

Grupy i ciała, liczby zespolone

Dla ustalenia uwagi, będziemy używać następujących oznaczeń:

$\mathbf{N} = \{ 1, 2, 3, \dots \}$ - liczby naturalne,

$\mathbf{Z} = \{ 0, \pm 1, \pm 2, \dots \}$ - liczby całkowite,

$\mathbf{W} = \left\{ \frac{m}{n} : \ m \in \mathbf{Z}, n \in \mathbf{N} \right\}$ - liczby wymierne,

$\mathbf{R} = \overline{\mathbf{W}}$ - liczby rzeczywiste,

$\mathbf{C} = \{ (a, b) : \ a, b \in \mathbf{R} \}$ - liczby zespolone.

Dwuargumentowym *działaniem wewnętrznym* '$\circ$' w zbiorze X nazywamy dowolną funkcję z iloczynu kartezjańskiego $X \times X$ w X. Wynik takiego działania na parze (x, y) będziemy oznaczać przez $x \circ y$.

1.1 Podstawowe struktury algebraiczne

Zaczniemy od przedstawienia abstrakcyjnych definicji grupy i ciała.

1.1.1 Grupa

Definicja 1.1 *Zbiór (niepusty) G wraz z wewnętrznym działaniem dwuargumentowym '$\circ$' jest* grupą *jeśli spełnione są następujące warunki (aksjomaty grupy):*

(i) $\forall a, b, c \in G \quad (a \circ b) \circ c = a \circ (b \circ c)$
(łączność działania)

(ii) $\exists e \in G \; \forall a \in G \quad a \circ e = a = e \circ a$
(istnienie elementu neutralnego)

(iii) $\forall a \in G \; \exists a' \in G \quad a \circ a' = e = a' \circ a$
(istnienie elementów przeciwnych/odwrotnych)

Jeśli ponadto

(iv) $\forall a, b \in G \quad a \circ b = b \circ a$

to grupę nazywamy przemienną *(lub* abelową*).*

Grupę będziemy oznaczać przez $\{G, \circ\}$.

Zauważmy, że już z aksjomatów grupy wynika, iż element neutralny jest wyznaczony jednoznacznie. Rzeczywiście, załóżmy, że istnieją dwa elementy neutralne, e_1 i e_2. Wtedy, z warunku (ii) wynika, że $e_1 = e_1 \circ e_2 = e_2$. Podobnie, istnieje tylko jeden element odwrotny dla każdego $a \in G$. Jeśli bowiem istniałyby dwa odwrotne, a'_1 i a'_2, to mielibyśmy

$$a'_1 = e \circ a'_1 = (a'_2 \circ a) \circ a'_1 = a'_2 \circ (a \circ a'_1) = a'_2 \circ e = a'_2,$$

przy czym skorzystaliśmy kolejno z własności (ii), (iii), (i) i ponownie (iii) i (ii).

Łatwo też pokazać, że w grupie $\{G, \circ\}$ równania

$$a \circ x = b \qquad \text{oraz} \qquad y \circ c = d$$

dla $a, b, c, d \in G$ mają jednoznaczne rozwiązania. W uzasadnieniu, ograniczymy się tylko do pierwszego równania. Łatwo sprawdzić, że $x = a' \circ b$ jest rozwiązaniem. Z drugiej strony, jeśli x jest rozwiązaniem to $a' \circ (a \circ x) = a' \circ b$, czyli $x = a' \circ b$.

Przykładami grup są:

- $\{\mathbf{Z}, +\}$, gdzie elementem neutralnym jest $e = 0$, a elementem przeciwnym do a' do a jest $-a$.
- $\{\mathbf{W} \setminus \{0\}, *\}$, gdzie $e = 1$ a $a' = a^{-1}$ jest odwrotnością a.

- Grupa obrotów płaszczyzny wokół początku układu współrzędnych, gdzie elementem neutralnym jest obrót o kąt zerowy, a elementem odwrotnym do obrotu o kąt α jest obrót o kąt $-\alpha$.

Zwróćmy uwagę na istotność wyjęcia zera w drugim przykładzie. Ponieważ 0 nie ma elementu odwrotnego, $\{\mathbf{W}, *\}$ nie jest grupą. Nie są też grupami np. $\{\mathbf{N}, *\}$ (nie ma elementów odwrotnych) oraz $\{\mathbf{R}, -\}$ (nie ma łączności oraz elementu neutralnego).

1.1.2 Ciało

Definicja 1.2 Ciałem *(a ściślej,* ciałem przemiennym*) nazywamy (co najmniej dwuelementowy) zbiór* $\mathbf{K}$ *z dwoma dwuargumentowymi działaniami wewnętrznymi, dodawaniem '+' i mnożeniem '$*$', spełniające następujące warunki (aksjomaty ciała):*

(i) $\{\mathbf{K}, +\}$ *jest grupą przemienną (w której element neutralny oznaczamy przez* 0*, a element przeciwny do a przez* $-a$*),*

(ii) $\{\mathbf{K} \setminus \{0\}, *\}$ *jest grupą przemienną (w której element neutralny oznaczamy przez* 1*, a odwrotny do a przez* a^{-1}*),*

(iii) $\forall a, b, c \in \mathbf{K} \quad a * (b + c) = a * b + a * c$
(mnożenie jest rozdzielne względem dodawania). [1]

Bezpośrednio z definicji ciała można pokazać następujące ogólne własności (uzasadnienie pozostawiamy jako proste ćwiczenie):

1. $0 \neq 1$,

2. $\forall a \in \mathbf{K} \quad 0 * a = 0 = a * 0$,

3. $\forall a \in \mathbf{K} \quad (-1) * a = -a$,

4. jeśli $a * b = 0$ to $a = 0$ lub $b = 0$,

5. jeśli $a \neq 0$ i $b \neq 0$ to $(a * b)^{-1} = b^{-1} * a^{-1}$,

[1] Przyjmujemy konwencję, że w wyrażeniach w których występują i dodawania i mnożenia najpierw wykonujemy mnożenia.

dla dowolnych $a, b \in \mathbf{K}$.

W ciele możemy formalnie zdefiniować odejmowanie i dzielenie, mianowicie

$$\begin{array}{rcll} a - b & := & a + (-b) & \forall a, b \in \mathbf{K}, \\ a/b & := & a * b^{-1} & \forall a \in \mathbf{K}, b \in \mathbf{K} \setminus \{0\}. \end{array}$$

Przykładem ciała są liczby rzeczywiste $\mathbf{R}$ z naturalnymi działaniami dodawania i mnożenia. Ciałem jest też zbiór liczb

$$\{\, a + b\sqrt{2} : \; a, b \in \mathbf{W} \,\} \subset \mathbf{R}$$

z tymi samymi działaniami.

1.2 Ciało liczb zespolonych

Ważnym przykładem ciała jest ciało liczb zespolonych, któremu poświęcimy tą część wykładu.

1.2.1 Definicja

Definicja 1.3 *Ciało* liczb zespolonych *to zbiór par uporządkowanych*

$$\mathbf{C} := \mathbf{R} \times \mathbf{R} = \{\, (a, b) : \; a, b \in \mathbf{R} \,\}$$

z działaniami dodawania i mnożenia zdefiniowanymi jako:

$$\begin{array}{rcl} (a, b) + (c, d) & = & (a + c, b + d), \\ (a, b) * (c, d) & = & (a * c - b * d, a * d + b * c), \end{array}$$

dla dowolnych $a, b, c, d \in \mathbf{R}$. [2]

Formalne sprawdzenie, że $\mathbf{C}$ ze zdefiniowanymi działaniami jest ciałem pozostawiamy czytelnikowi. Tu zauważymy tylko, że elementem neutralnym

[2] Zauważmy, że znaki dodawania i mnożenia występują tu w dwóch znaczeniach, jako działania na liczbach rzeczywistych oraz jako działania na liczbach zespolonych. Z kontekstu zawsze wiadomo w jakim znaczeniu te działania są użyte.

dodawania jest $(0,0)$, a mnożenia $(1,0)$. Elementem przeciwnym do (a,b) jest $-(a,b) = (-a,-b)$, a odwrotnym do $(a,b) \neq (0,0)$ jest

$$(a,b)^{-1} = \left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2} \right).$$

Zdefiniujemy mnożenie liczby zespolonej przez rzeczywistą w następujący (naturalny) sposób. Niech $z = (a,b) \in \mathbf{C}$ i $c \in \mathbf{R}$. Wtedy

$$c * (a,b) = (a,b) * c = (c * a, c * b).$$

Przyjmując tą konwencję, mamy

$$(a,b) = a * (1,0) + b * (0,1).$$

W końcu, utożsamiając liczbę zespoloną $(a,0)$ z liczbą rzeczywistą a, oraz wprowadzając dodatkowo oznaczenie

$$\imath := (0,1)$$

otrzymujemy

$$(a,b) = a + \imath * b. \tag{1.1}$$

$a = \Re z$ nazywa się *częścią rzeczywistą*, a $b = \Im z$ *częścią urojoną* liczby zespolonej. Samą liczbę zespoloną $\imath$ nazywamy *jednostką urojoną*. Zauważmy, że

$$\imath^2 = (-1,0) = -1.$$

1.2.2 Postać trygonometryczna

Postać (1.1) jest najbardziej rozpowszechniona. Często wygodnie jest użyć również postaci trygonometrycznej, która jest konsekwencją interpretacji liczby zespolonej (a,b) jako punktu na płaszczyźnie (tzw. *płaszczyźnie zespolonej*) o współrzędnych a i b. Dokładniej, przyjmując

$$|z| := \sqrt{a^2+b^2}$$

oraz kąt ϕ tak, że

$$\sin\phi = \frac{b}{|z|}, \qquad \cos\phi = \frac{a}{|z|},$$

otrzymujemy

$$z = |z|(\cos\phi + \imath\sin\phi). \tag{1.2}$$

Jest to właśnie *postać trygonometryczna*. Liczbę rzeczywistą $|z|$ nazywamy *modułem* liczby zespolonej z, a ϕ jej *argumentem*, $\phi = \arg z$.

Jeśli $z \neq 0$ i założymy, że $\phi \in [0, 2\pi)$ to postać trygonometryczna jest wyznaczona jednoznacznie. Piszemy wtedy $\phi = \mathrm{Arg} z$.

1.2.3 Wzór de Moivre'a

Niech $z = |z|(\cos\phi + \imath\sin\phi)$, $w = |w|(\cos\psi + \imath\sin\psi)$ będą dwoma liczbami zespolonymi. Wtedy

$$\begin{aligned} w * z &= |w||z|\left((\cos\phi\cos\psi - \sin\phi\sin\psi) + \imath(\sin\phi\cos\psi + \sin\psi\cos\phi)\right) \\ &= |w||z|\left(\cos(\phi+\psi) + \imath\sin(\phi+\psi)\right), \end{aligned}$$

a stąd

$$|w * z| = |w||z| \qquad \text{oraz} \qquad \arg(w * z) = \arg w + \arg z.$$

Właśnie w tych równościach przejawia się wygoda postaci trygonometrycznej. W szczególności mamy bowiem $z^2 = |z|^2(\cos 2\phi + \imath\sin 2\phi)$ i postępując dalej indukcyjnie otrzymujemy *wzór de Moivre'a*. Mianowicie, dla dowolnej liczby zespolonej z w postaci trygonometrycznej (1.2) mamy

$$z^n = |z|^n(\cos(n\phi) + \imath\sin(n\phi)), \qquad n = 0, 1, 2, \ldots \tag{1.3}$$

Łatwo zauważyć, że wzór (1.3) jest prawdziwy również dla $n = -1$, a stąd dla wszystkich całkowitych n. Przyjmując za $z^{1/n}$ szczególne rozwiązanie równania $w^n = z$, mianowicie

$$z^{1/n} = |z|^{1/n}\left(\cos(\phi/n) + \imath\sin(\phi/n)\right),$$

gdzie $\phi = \mathrm{Arg} z$, uogólniamy (1.3) dla wszystkich wykładników wymiernych. Stosując dalej argument z przejściem granicznym (każda liczba rzeczywista jest granicą ciągu liczb wymiernych) otrzymujemy w końcu następujący *uogólniony* wzór de Moivre'a:

$$\forall a \in \mathbf{R} \qquad z^a = |z|^a\left(\cos(a\phi) + \imath\sin(a\phi)\right).$$

Prostym wnioskiem z ostatniego wzoru jest równanie

$$z = |z| * \omega^{\phi},$$

gdzie $\omega = \cos 1 + \imath \sin 1 = 0,540302\ldots + \imath * 0,84147\ldots \in \mathbf{C}$. Jest to uogólnienie na przypadek liczb zespolonych wzoru $x = |x| * \mathrm{sgn}(x)$ znanego z przypadku liczb rzeczywistych.

1.2.4 Pierwiastki z jedynki

Rozpatrzmy rozwiązania równania

$$z^n = 1$$

dla dowolnej naturalej n. W dziedzinie rzeczywistej pierwiastkiem jest 1 jeśli n jest nieparzyste, albo 1 i (-1) jeśli n jest parzyste. W dziedzinie zespolonej mamy zawsze n pierwiastków. Rzeczywiście, ponieważ $1 = \cos(2k\pi) + \imath \sin(2k\pi)$, ze wzoru de Moivre'a dostajemy, że wszyskie pierwiastki wyrażają się wzorami

$$z_k := \cos\left(\frac{2k\pi}{n}\right) + \imath \sin\left(\frac{2k\pi}{n}\right), \qquad k = 0, 1, 2, \ldots, n-1.$$

Zauważmy, że z_j leżą na okręgu jednostkowym płaszczyzny zespolonej. Zbiór $G = \{z_k : k = 0, 1, \ldots, n-1\}$ ze zwykłym mnożeniem liczb zespolonych tworzy grupę z elementem neutralnym $z_0 = 1$.

1.2.5 Sprzężenie

Liczbę *sprzężoną* do $z = a + \imath b$ definiujemy jako

$$\overline{z} := a - \imath b.$$

Zauważmy, że $\overline{\overline{z}} = z$ oraz $z * \overline{z} = |z|^2$. Mamy też

$$\frac{z + \overline{z}}{2} = \Re z \quad \text{i} \quad \frac{z - \overline{z}}{2\imath} = \Im z.$$

I jeszcze jedna ważna własność sprzężenia. Jeśli $\diamond \in \{+, -, *, /\}$ to

$$\overline{w \diamond z} = \overline{w} \diamond \overline{z}.$$

Stosując indukcję, można ten wzór uogólnić w następujący sposób. Jeśli $f(u_1, u_2, \ldots, u_s)$ jest wyrażeniem arytmetycznym, gdzie u_j są stałymi lub zmiennymi zespolonymi, to

$$\overline{f(u_1, u_2, \ldots, u_s)} = f(\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_s).$$

1.3 Wielomiany

Definicja 1.4 Wielomianem *p nad ciałem* $\mathbf{K}$ *nazywamy funkcję zmiennej z o wartościach w ciele* $\mathbf{K}$ *daną wzorem*

$$p(z) := \sum_{j=0}^{n} a_j z^j = a_0 + a_1 z + \cdots + a_n z^n,$$

gdzie $a_j \in \mathbf{K}$, $0 \leq j \leq n$, $a_n \neq 0$, *są* współczynnikami *wielomianu. Liczbę* n *nazywamy* stopniem *wielomianu i oznaczamy*

$$n = \deg p.$$

(Przyjmujemy przy tym, że $\deg 0 = -\infty$.*)*

1.3.1 Algorytm Hornera

Każdy wielomian $p(z) = \sum_{k=0}^{n} a_k z^k$ stopnia $n \geq 1$ o współczynnikach zespolonych można podzielić przez dwumian $z - \xi$ otrzymując

$$p(z) = q(z)(z - \xi) + \eta,$$

gdzie $\deg q = n - 1$, a $\eta \in \mathbf{C}$. Dodatkowo, jeśli p ma współczynniki rzeczywiste i $\xi \in \mathbf{R}$, to q ma również współczynniki rzeczywiste i $\eta \in \mathbf{R}$.

Iloraz q oraz resztę η z dzielenia można otrzymać stosując *algorytm Hornera*:

$\{\ b_n := a_n;$
  for $k := n-1$ **downto** 0 **do** $b_k := a_k + \xi * b_{k+1};$
$\}$

Wtedy $q(z) = \sum_{k=1}^{n} b_k z^{k-1}$ oraz reszta $\eta = b_0$.

1.3.2 Zasadnicze twierdzenie algebry

Dla wielomianów zespolonych prawdziwe jest następujące ważne twierdzenie.

Twierdzenie 1.1 (Zasadnicze Twierdzenie Algebry)
Każdy wielomian zespolony p stopnia co najmniej pierwszego ma pierwiastek zespolony, tzn. równanie $p(z) = 0$ *ma rozwiązanie.*

Twierdzenie 1.1 mówi, że liczby zespolone **C** są ciałem *algebraicznie domkniętym.* (Przypomnijmy, że liczby rzeczywiste **R** *nie* są algebraicznie domknięte, bo np. równanie $x^2 + 1 = 0$ nie ma rozwiązań w **R**.)

Konsekwencją algebraicznej domkniętości **C** jest *faktoryzacja* (rozkład) wielomianu zespolonego na czynniki pierwszego stopnia. Dokładniej, stosując n-krotnie zasadnicze twierdzenie algebry oraz fakt, że jeśli ξ jest pierwiastkiem wielomianu p to reszta z dzielenia p przez $(\cdot - \xi)$ jest zerowa, otrzymujemy rozkład

$$p(z) = a_n(z - z_1)(z - z_2)\cdots(z - z_n), \tag{1.4}$$

gdzie z_j, $1 \leq j \leq n$, są pierwiastkami p. Zakładając, że tylko m pierwiastków jest parami różnych $(1 \leq m \leq n)$, możemy równoważnie napisać, że

$$p(z) = a_n(z - u_1)^{s_1}(z - u_2)^{s_2}\cdots(z - u_m)^{s_m},$$

gdzie $u_i \neq u_j$ o ile $i \neq j$, oraz $\sum_{j=1}^m s_j = n$. Przy tym zapisie, s_j nazywamy *krotnością* pierwiastka u_j.

Załóżmy teraz, że współczynniki wielomianu p są rzeczywiste, $a_j \in \mathbf{R}$, $0 \leq j \leq n$. Załóżmy też, że $p(\xi) = 0$ i $\xi \notin \mathbf{R}$. Wtedy $\overline{\xi} \neq \xi$ i

$$p(\overline{\xi}) = \sum_{j=0}^n a_j \overline{\xi}^j = \sum_{j=0}^n \overline{a}_j \overline{\xi^j} = \overline{\sum_{j=0}^n a_j \xi^j} = \overline{0} = 0,$$

tzn. jeśli ξ jest pierwiastkiem to także liczba sprzężona $\overline{\xi}$ jest pierwiastkiem; obie występują w rozwinięciu (1.4). Ale

$$(z - \xi)(z - \overline{\xi}) = z^2 - z(\xi + \overline{\xi}) + \xi\overline{\xi} = z^2 - 2z\Re\xi + |\xi|^2$$

jest trójmianem kwadratowym o współczynnikach rzeczywistych. Stąd wniosek, że wielomian rzeczywisty daje się rozłożyć na iloczyn czynników stopnia co najwyżej drugiego.

Rozdział 2

Macierze liczbowe

2.1 Podstawowe definicje

Macierzą (nad ciałem $\mathbf{K}$) nazywamy tablicę prostokątną

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix},$$

gdzie $a_{i,j} \in \mathbf{K}$, $1 \le i \le m$, $1 \le j \le n$. Będziemy mówić, że A jest macierzą *formatu* $m \times n$, tzn. macierzą o m *wierszach* i n *kolumnach.* Zbiór wszystkich takich macierzy oznaczamy przez $\mathbf{K}^{m,n}$.

2.1.1 Macierze szczególnych formatów

- $n \times n$ Macierze *kwadratowe* $\mathbf{K}^{n,n}$.

- $m \times 1$ Macierze jednokolumnowe nazywane *wektorami.*
  Zbiór wektorów oznaczamy przez $\mathbf{K}^{m,1} = \mathbf{K}^m$,

$$\mathbf{K}^m \ni A = (a_{i,1}) = \vec{a} = \hat{a} = (a_i)_{i=1}^m = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}.$$

- $1 \times n$ Macierze jednowierszowe nazywane *funkcjonałami*.
  Zbiór funkcjonałów oznaczamy przez $\mathbf{K}^{1,n} = \mathbf{K}^{nT}$ (albo $\mathbf{K}^{nH}$),
  $$\mathbf{K}^{\mathbf{n}T} \ni A = (a_{1,j}) = \vec{a}^T = \hat{a}^T = (a_j)_{j=1}^n = [a_1, \ldots, a_n] \,.$$
- 1×1 Macierze jednoelementowe, utożsamiane z $\mathbf{K}$, tzn. $\mathbf{K}^{1,1} = \mathbf{K}$.

2.1.2 Podział blokowy

Często wygodnie jest przedstawić macierz w postaci blokowej, która w ogólności wygląda następująco:

$$A = \begin{bmatrix} A_{1,1} & \ldots & A_{1,t} \\ \vdots & & \vdots \\ A_{s,1} & \ldots & A_{s,t} \end{bmatrix} \in \mathbf{K}^{m,n}, \tag{2.1}$$

gdzie $A_{p,q} \in \mathbf{K}^{m_p,n_q}$, $1 \le p \le s$, $1 \le q \le t$, $\sum_{p=1}^s m_p = m$, $\sum_{q=1}^t n_q = n$.

Na postać blokową można patrzyć jak na macierz, której elementami są macierze. Z drugiej strony, macierz liczbową można interpretować jako macierz w postaci blokowej z blokami formatu 1×1.

Ważne szczególne przypadki to podział kolumnowy macierzy,

$$A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n] \,, \qquad \text{gdzie} \quad \vec{a}_j = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix}, \quad 1 \le j \le n,$$

oraz podział wierszowy macierzy,

$$A = \begin{bmatrix} \hat{a}_1^T \\ \hat{a}_2^T \\ \vdots \\ \hat{a}_m^T \end{bmatrix}, \qquad \text{gdzie} \quad \hat{a}_i^T = [a_{i,1}, a_{i,2}, \ldots, a_{i,n}] \,, \quad 1 \le i \le m.$$

2.2 Działania na macierzach

2.2.1 Podstawowe działania

Możemy na macierzach wykonywać różne działania. Podstawowe z nich to:

$u \in \mathbf{K}, A \in \mathbf{K}^{m,n} \implies B = u * A \in \mathbf{K}^{m,n},\ b_{i,j} = u * a_{i,j}$
(mnożenie macierzy przez liczbę)

$A, B \in \mathbf{K}^{m,n} \implies C = A + B \in \mathbf{K}^{m,n},\ c_{i,j} = a_{i,j} + b_{i,j}$
(dodawanie macierzy)

$A \in \mathbf{K}^{m,n} \implies B = A^T \in \mathbf{K}^{n,m},\ b_{j,i} = a_{i,j}$ (transpozycja macierzy)

$A \in \mathbf{C}^{m,n} \implies B = A^H \in \mathbf{K}^{n,m},\ b_{j,i} = \overline{a}_{i,j}$ (sprzężenie hermitowskie)

$A \in \mathbf{C}^{m,n} \implies B = |A| \in \mathbf{C}^{m,n},\ b_{i,j} = |a_{i,j}|$ (moduł macierzy)

W szczególności, mamy też dla $u, v \in \mathbf{K} \subset \mathbf{C}$, $A, B \in \mathbf{C}^{m,n}$,

$$(u * A \pm v * B)^H = \overline{u} * A^H \pm \overline{v} * B^H,$$
$$\left(A^T\right)^T = A = \left(A^H\right)^H.$$

Zauważmy, że macierze formatu $m \times n$ z działaniem dodawania są grupą przemienną, przy czym elementem neutralnym jest macierz zerowa (gdzie $a_{i,j} = 0\ \forall i, j$), a przeciwną do $(a_{i,j})$ jest macierz $(-a_{i,j})$.

Jeśli macierze dane są w postaci blokowej (2.1) to:

$B = u * A \implies B_{p,q} = u * A_{p,q}$

$C = A + B \implies C_{p,q} = A_{p,q} + B_{p,q}$

$B = A^T \implies B_{p,q} = A^T_{q,p}$

$B = A^H \implies B_{p,q} = A^H_{q,p}$

2.2.2 Mnożenie macierzy

Jeśli $A \in \mathbf{K}^{m,l}$ i $B \in \mathbf{K}^{l,n}$ to

$$C = A * B \in \mathbf{K}^{m,n},$$

gdzie

$$c_{i,j} = \sum_{k=1}^{l} a_{i,k} * b_{k,j}, \qquad 1 \le i \le m, 1 \le j \le n.$$

Zauważmy, że mnożenie $A * B$ jest wykonalne wtedy i tylko wtedy gdy liczba kolumn macierzy A jest równa liczbie wierszy macierzy B. Jeśli A jest w postaci wierszowej, a B kolumnowej,

$$A = \begin{bmatrix} \hat{a}_1^T \\ \vdots \\ \hat{a}_m^T \end{bmatrix}, \qquad B = \left[\vec{b}_1, \ldots, \vec{b}_l\right],$$

to $c_{i,j} = \hat{a}_i^T * \vec{b}_j \ \forall i, j$.

Podstawowe własności mnożenia macierzy są następujące. (Zakładamy, że macierze są odpowiednich formatów tak, że działania są wykonalne.)

$(A + B) * C = A * C + B * C$

$C * (A + B) = C * A + C * B$
(rozdzielność mnożenia względem dodawania)

$u * (A * B) = (u * A) * B = A * (u * B) = (A * B) * u \quad (u \in \mathbf{K})$

$(A * B) * C = A * (B * C)$ (łączność mnożenia)

Dowody tych własności polegają na zwykłym sprawdzeniu. Dlatego, dla przykładu, pokażemy tu jedynie łączność. Niech macierze A, B, C będą odpowiednio formatów $m \times k$, $k \times l$, $l \times n$. (Zauważmy, że tylko wtedy odpowiednie mnożenia są wykonalne.) Mamy

$$\begin{aligned} ((A * B) * C)_{i,j} &= \sum_{s=1}^{l} (A * B)_{i,s} c_{s,j} = \sum_{s=1}^{l} \left(\sum_{t=1}^{k} a_{i,t} b_{t,s} \right) c_{s,j} \\ &= \sum_{t=1}^{k} a_{i,t} \sum_{s=1}^{l} b_{t,s} c_{s,j} = \sum_{t=1}^{k} a_{i,t} (B * C)_{t,j} \\ &= (A * (B * C))_{i,j}. \end{aligned}$$

Mamy też

$$(A * B)^T = B^T * A^T, \qquad (A * B)^H = B^H * A^H.$$

Rzeczywiście,

$$\left((A * B)^H\right)_{i,j} = \overline{(A * B)_{j,i}} = \overline{\sum_{k=1}^{l} a_{j,k} b_{k,i}}$$

$$
\begin{aligned}
&= \sum_{k=1}^{l} \overline{a}_{j,k}\overline{b}_{k,i} = \sum_{k=1}^{l} \overline{b}_{k,i}\overline{a}_{j,k} \\
&= \sum_{k=1}^{l} \left(B^H\right)_{i,k} \left(A^H\right)_{k,j} = \left(B^H * A^H\right)_{i,j}.
\end{aligned}
$$

2.2.3 Mnożenie macierzy w postaci blokowej

Jeśli macierze są podane w postaci blokowej to można je mnożyć 'blok-po-bloku' (tak jak w przypadku bloków 1×1) o ile formaty odpowiednich bloków są zgodne. Dokładniej, jeśli $A = (A_{i,k})$, $B = (B_{k,j})$, $1 \le i \le m$, $1 \le k \le l$, $1 \le j \le n$, oraz dla wszystkich i, j, k liczba kolumn bloku $A_{i,k}$ macierzy A jest równa liczbie wierszy bloku $B_{k,j}$ macierzy B to iloczyn

$$C = A * B = (C_{i,j}),$$

$1 \le i \le m$, $1 \le j \le n$, gdzie

$$C_{i,j} = \sum_{k=1}^{l} A_{i,k} * B_{k,n}.$$

Pokażemy to na przykładzie. Niech

$$
A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \\ A_{2,1} & A_{2,2} & A_{2,3} & A_{2,4} \\ A_{3,1} & A_{3,2} & A_{3,3} & A_{3,4} \end{bmatrix}, \qquad B = \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \\ B_{3,1} & B_{3,2} \\ B_{4,1} & B_{4,2} \end{bmatrix}.
$$

Wtedy

$$
C = \begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \\ C_{3,1} & C_{3,2} \end{bmatrix},
$$

gdzie

$$
\begin{aligned}
C_{1,1} &= A_{1,1} * B_{1,1} + A_{1,2} * B_{2,1} + A_{1,3} * B_{3,1} + A_{1,4} * B_{4,1}, \\
C_{1,2} &= A_{1,1} * B_{1,2} + A_{1,2} * B_{2,2} + A_{1,3} * B_{3,2} + A_{1,4} * B_{4,2}, \\
C_{2,1} &= A_{2,1} * B_{1,1} + A_{2,2} * B_{2,1} + A_{2,3} * B_{3,1} + A_{2,4} * B_{4,1}, \\
C_{2,2} &= A_{2,1} * B_{1,2} + A_{2,2} * B_{2,2} + A_{2,3} * B_{3,2} + A_{2,4} * B_{4,2}, \\
C_{3,1} &= A_{3,1} * B_{1,1} + A_{3,2} * B_{2,1} + A_{3,3} * B_{3,1} + A_{3,4} * B_{4,1}, \\
C_{3,2} &= A_{3,1} * B_{1,2} + A_{3,2} * B_{2,2} + A_{3,3} * B_{3,2} + A_{3,4} * B_{4,2},
\end{aligned}
$$

o ile formaty bloków $A_{i,k}$ i $B_{k,j}$ są zgodnie.

Bardzo ważnym przypadkiem szczególnym mnożenia blokowego jest

$$\begin{aligned} A * B &= A * \left[\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_l\right] \\ &= \left[A * \vec{b}_1, A * \vec{b}_2, \ldots, A * \vec{b}_l\right]. \end{aligned}$$

Zwróćmy jeszcze uwagę na fakt, że jeśli $\vec{a} \in \mathbf{K}^m$ oraz $\vec{b} \in \mathbf{K}^n$ to

$$C = \vec{a} * \vec{b}^T \in \mathbf{K}^{m,n}$$

jest macierzą formatu $m \times n$, nazywaną *iloczynem wewnętrznym* wektorów. Jeśli natomiast wektory są tych samych formatów, $\vec{a}, \vec{b} \in \mathbf{K}^n$, to

$$c = \vec{a}^T * \vec{b} = \vec{b}^T * \vec{a} \in \mathbf{K}$$

jest liczbą, nazywaną *iloczynem zewnętrznym.* W przypadku $\vec{a}, \vec{b} \in \mathbf{C}^n$ definiujemy również *iloczyn skalarny* wektorów jako liczbę zespoloną

$$g = \vec{b}^H * \vec{a} \in \mathbf{C}.$$

2.3 Dalsze oznaczenia

2.3.1 Macierze trójkątne i jednostkowe

Wyróżnimy następujące podzbiory macierzy formatu $m \times n$ (niekoniecznie kwadratowych):

$$\begin{aligned} \mathrm{TRIU}^{m,n} &= \{ A \in \mathbf{K}^{m,n} \,:\, \forall i > j \ a_{i,j} = 0 \}, \\ \mathrm{TRIL}^{m,n} &= \{ A \in \mathbf{K}^{m,n} \,:\, \forall i < j \ a_{i,j} = 0 \}, \\ \mathrm{DIAG}^{m,n} &= \{ A \in \mathbf{K}^{m,n} \,:\, \forall i \neq j \ a_{i,j} = 0 \}. \end{aligned}$$

Są to odpowiednio macierze *trójkątne górne*, *trójkątne dolne* i *diagonalne.* Zauważmy, że każdy z tych podzbiorów macierzy stanowi grupę ze względu na działanie dodawania macierzy (są to *podgrupy* $\{\mathbf{K}^{m,n}, +\}$), oraz

$$\mathrm{DIAG}^{m,n} = \mathrm{TRIU}^{m,n} \cap \mathrm{TRIL}^{m,n}.$$

Ponieważ macierze diagonalne $D \in \mathrm{DIAG}^{m,n}$ mają elementy niezerowe jedynie na głównej diagonali, powiedzmy d_i, $1 \le i \le \min(m,n)$, będziemy pisać

$$D = \mathrm{diag}\left(d_1, d_2, \ldots, d_{\min(m,n)}\right).$$

Szczególnie ważnymi macierzami diagonalnymi są (kwadratowe) macierze *jednostkowe*

$$I_n = \mathrm{diag}_n(\underbrace{1, 1, \ldots, 1}_{n}) \in \mathbf{K}^{n,n}.$$

Jeśli $A \in \mathbf{K}^{m,n}$ to

$$I_m * A \;=\; A \;=\; A * I_n,$$

co oznacza, że I_m i I_n są elementami neutralnymi mnożenia (odpowiednio lewostronnym i prawostronnym).

2.3.2 Układ równań jako równanie macierzowe

Rozpatrzmy następujący układ równań:

$$\left\{ \begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots & + & a_{2,n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array} \right. . \tag{2.2}$$

Mówimy, że jest to układ m równań *liniowych* z n niewiadomymi. Liczby $a_{i,j} \in \mathbf{K}$ nazywamy i *współczynnikami* układu, b_i wyrazami *wolnymi*, a x_j to *niewiadome*.

Oznaczmy

$$A = (a_{i,j}) \in \mathbf{K}^{m,n}, \quad \vec{b} = (b_i) \in \mathbf{K}^m, \quad \vec{x} = (x_j) \in \mathbf{K}^n.$$

Wtedy układ (2.2) możemy równoważnie zapisać po prostu jako równanie macierzowe

$$A * \vec{x} = \vec{b}.$$

2.4 Macierze nieosobliwe

2.4.1 Grupa macierzy nieosobliwych

W zbiorze $\mathbf{K}^{n,n}$ mnożenie macierzy jest działaniem wewnętrznym. Ponadto, jak wcześniej zauważyliśmy, mnożenie jest łączne, a macierz jednostkowa

$I_n = \mathrm{diag}(1, \ldots, 1) \in \mathbf{K}^{n,n}$ jest elementem neutralnym mnożenia,

$$\forall A \in \mathbf{K}^{n,n} \quad A * I_n = A = I_n * A.$$

(Przypomnijmy, że element neutralny, jeśli istnieje, jest tylko jeden.) Naturalnym staje się teraz pytanie, czy istnieją elementy odwrotne. Niestety, nie zawsze. Na przykład, łatwo sprawdzić, że (niezerowa) macierz

$$\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$$

nie ma odwrotności (zarówno lewostronnej jak i prawostronnej). Z drugiej strony, wiele macierzy niezerowych mają odwrotności. Na przykład, macierze

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \quad \text{oraz} \quad B = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$$

stanowią parę macierzy do siebie wzajemnie odwrotnych, $A*B = I_2 = B*A$, tak, że możemy napisać $B = A^{-1}$ i $A = B^{-1}$. (Przypomnijmy, że element odwrotny, jeśli istnieje, jest wyznaczony jednoznacznie.)

Definicja 2.1 *Macierz kwadratową $A \in \mathbf{K}^{n,n}$ dla której istnieje macierz odwrotna $A^{-1} \in \mathbf{K}^{n,n}$ nazywamy* odwracalną *albo* nieosobliwą. *Macierz, która nie posiada macierzy odwrotnej nazywamy* osobliwą.

Zwróćmy uwagę na fakt, że pojęcie macierzy (nie)osobliwej przysługuje jedynie macierzy kwadratowej.

Iloczyn macierzy nieosobliwych jest macierzą nieosobliwą. Rzeczywiście, jeśli $A, B \in \mathbf{K}^{n,n}$ to sprawdzamy bezpośrednio, że odwrotnością $C = A * B$ jest macierz

$$C^{-1} = B^{-1} * A^{-1}.$$

Stąd wniosek, że

ZBIÓR MACIERZY NIEOSOBLIWYCH FORMATU $n \times n$ Z DZIAŁANIEM MNOŻENIA MACIERZY JEST GRUPĄ (NIEPRZEMIENNĄ).

2.4.2 Warunek nieosobliwości macierzy

Twierdzenie 2.1 *Aby macierz $A \in \mathbf{K}^{n,n}$ była nieosobliwa potrzeba i wystarcza, aby dla każdego $\vec{b} \in \mathbf{K}^n$ układ równań $A * \vec{x} = \vec{b}$ miał jednoznaczne rozwiązanie $\vec{x} \in \mathbf{K}^n$.*

Dowód. (Konieczność.) Jeśli A jest nieosobliwa to łatwo sprawdzić, że $\vec{x} = A^{-1} * \vec{b}$ jest rozwiązaniem. Z drugiej strony, jeśli $\vec{x}$ jest rozwiązaniem, $A * \vec{x} = \vec{b}$, to $A^{-1} * (A * \vec{x}) = A^{-1} * \vec{b}$, czyli $\vec{x} = A^{-1} * \vec{b}$ jest rozwiązaniem jednoznacznym.

(Dostateczność.) Układy równań $A*\vec{b}_j = \vec{e}_j$, gdzie $\vec{e}_j$ jest j-tym *wersorem*,

$$\vec{e}_j = [0, \ldots, 0, 1, 0, \ldots, 0]^T,$$

(gdzie jedynka stoi na j-tym miejscu) mają jednoznaczne rozwiązania $\vec{b}_j$, $1 \leq j \leq n$. Biorąc $B = [\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n]$ mamy

$$A * B = [A * \vec{b}_1, \ldots, A * \vec{b}_n] = [\vec{e}_1, \ldots, \vec{e}_n] = I_n.$$

Pozostaje jeszcze pokazać, że $B*A = I_n$. Rzeczywiście, mamy $(A*B)*A = A$, czyli $A * (B * A) = A$. Rozwiązaniem równania $A * Z = A$ jest $Z = I_n$, a ponieważ z założenia rozwiązanie to jest jednoznaczne to $B * A = I_n$. Stąd $B = A^{-1}$, co kończy dowód.

Jednym z ważnych wniosków z tego twierdzenie jest następujący.

Wniosek 2.1 *Macierz trójkątna (górna lub dolna) $T \in \mathbf{K}^{n,n}$ jest nieosobliwa wtedy i tylko wtedy gdy wszystkie elementy na głównej diagonali są niezerowe.*

Rzeczywiście, wystarczy sprawdzić jednoznaczną rozwiązywalność odpowiedniego układu równań. Dodajmy, że macierz odwrotna do trójkątnej dolnej (górnej), jeśli istnieje, jest też trójkątna dolna (górna).

2.4.3 Permutacje

Niech $p = [p(1), p(2), \ldots, p(n)] \in \mathrm{Perm}(n)$ będzie *permutacją* n elementową. Odpowiadającą tej permutacji macierz $P = (p_{i,j}) \in \mathbf{K}^{n,n}$ zdefiniowaną jako

$$p_{i,j} = \begin{cases} 1 & \text{gdy } j = p(i), \\ 0 & \text{gdy } j \neq p(i), \end{cases}$$

nazywamy *macierzą permutacji.* Na przykład, jeśli $p = [3, 1, 4, 2] \in \mathrm{Perm}(4)$ to

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

Równoważnie, macierz kwadratowa P jest macierzą permutacji wtedy i tylko wtedy gdy w każdym wierszu i w każdej kolumnie występuje dokładnie jedna jedynka, a pozostałe elementy są zerami.

Łatwo sprawdzić, że permutacje n-elementowe $\mathrm{Perm}(n)$ tworzą grupę ze względu na ich złożenie,

$$(q \circ p)(i) = q(p(i)) \qquad 1 \le i \le n.$$

Elementem neutralnym jest permutacja identycznościowa $\mathrm{id}(i) = i \ \forall i$, a elementem odwrotnym do p jest permutacja odwrotna p' zdefiniowana równością $p'(p(i)) = i \ \forall i$.

Podobnie, macierze permutacji tworzą grupę ze względu na mnożenie macierzy, przy czym

$$P(q \circ p) = P(p) * P(q).$$

Rzeczywiście, $(P(q \circ p))_{i,j} = 1$ w.t.w. gdy $q(p(i)) = j$. Z drugiej strony, $(P(p) * P(q))_{i,j} = 1$ w.t.w gdy $(P(q))_{p(i),j} = 1$, czyli znów $q(p(i)) = j$.

Podobnie pokazujemy, że

$$P(p') = (P(p))^{-1} = (P(p))^T.$$

Zauważmy jeszcze, że jeśli $P = P(p)$, $p \in \mathrm{Perm}(n)$, to

$$P * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{p(1)} \\ \vdots \\ x_{p(n)} \end{bmatrix},$$

czyli mnożenie wektora z lewej strony przez macierz permutacji skutkuje zamianą kolejności współrzędnych. Podobnie,

$$P * \begin{bmatrix} \hat{a}_1^T \\ \vdots \\ \hat{a}_n^T \end{bmatrix} = \begin{bmatrix} \hat{a}_{p(1)}^T \\ \vdots \\ \hat{a}_{p(n)}^T \end{bmatrix}$$

powoduje przestawienie wierszy macierzy zgodnie z p. Ponieważ

$$A * P = \left((A * P)^T\right)^T = \left(P^T * A^T\right)^T,$$

dochodzimy do wniosku, że

$A * P$ permutuje kolumny A zgodnie z p',

$A * P^T$ permutuje kolumny A zgodnie z p.

Rozdział 3

Normy wektorów i macierzy

W tym rozdziale zakładamy, że

$$\mathbf{K} \subseteq \mathbf{C}.$$

3.1 Ogólna definicja normy

Niech $\psi : \mathbf{K}^{m,n} \to [0, +\infty)$ będzie przekształceniem spełniającym warunki:

(i) $\forall A \in \mathbf{K}^{m,n} \quad \psi(A) = 0 \iff A = 0$,

(ii) $\forall A \in \mathbf{K}^{m,n} \, \forall u \in \mathbf{K} \quad \psi(u * A) = |u| \cdot \psi(A)$,

(iii) $\forall A, B \in \mathbf{K}^{m,n} \quad \psi(A + B) \leq \psi(A) + \psi(B)$
(*nierówność trójkąta* albo *subaddytywność*).

Każde takie przekształcenie ψ nazywamy *normą* w $\mathbf{K}^{m,n}$ i oznaczamy

$$\psi(A) = \|A\|.$$

Norma jest miarą "wielkości" macierzy. Dlatego

$$\|A - B\|$$

uznajemy za miarę odległości między macierzami A i B.

Powiemy, że norma jest *monotoniczna* gdy warunek $|A| \leq |B|$ (tzn. gdy $|a_{i,j}| \leq |b_{i,j}| \ \forall i, j$) implikuje $\|A\| \leq \|B\|$. Jeśli norma w $\mathbf{K}^{n,n}$ spełnia

$$\|A * B\| \leq \|A\| \cdot \|B\|, \qquad \forall A, B \in \mathbf{K}^{n,n},$$

to mówimy, że norma jest *submultiplikatywna*.

3.2 Normy wektorów

3.2.1 Normy p-te

Wektory w $\mathbf{K}^n$ są szczególnymi macierzami. W tym przypadku, ważnymi przykładami norm są *normy Schura*, zdefiniowane dla danej p, $1 \le p \le \infty$, jako

$$\begin{aligned} \|\vec{x}\|_p &= \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \qquad \text{dla} \quad 1 \le p < \infty, \\ \|\vec{x}\|_\infty &= \max_{1\le i\le n} |x_i|. \end{aligned}$$

Nietrudno zauważyć, że $\|\vec{x}\|_\infty = \lim_{p\to\infty} \|\vec{x}\|_p$, $\forall \vec{x} \in \mathbf{K}^n$.

Warunki (i) i (ii) normy są trywialnie spełnione przez normy Schura. Warunek (iii) łatwo sprawdzić dla $p = 1, \infty$. Dla $p = 1$ mamy bowiem

$$\|\vec{x}+\vec{y}\|_1 = \sum_{i=1}^n |x_i+y_i| \le \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| = \|\vec{x}\|_1 + \|\vec{y}\|_1,$$

a dla $p = \infty$

$$\|\vec{x}+\vec{y}\|_\infty = \max_{1\le i\le n} |x_i+y_i| \le \max_{1\le i\le n} |x_i| + \max_{1\le i\le n} |y_i| = \|\vec{x}\|_\infty + \|\vec{y}\|_\infty.$$

(W obu przypadkach zastosowaliśmy nierówność trójkąta $|u+v| \le |u| + |v|$ dla liczb zespolonych u i v.) Dla innych wartości p warunek (iii) jest dużo trudniej pokazać. Dlatego ograniczymy się tu jedynie do przypadku $p = 2$.

Lemat 3.1 (Nierówność Schwarza)
Dla dowolnych $\vec{u}, \vec{v} \in \mathbf{K}^n$ mamy

$$|\vec{u}^H * \vec{v}| \le \|\vec{u}\|_2 \cdot \|\vec{v}\|_2.$$

Dowód. Dla $t \in \mathbf{K}$ mamy

$$\begin{aligned} 0 &\le \|\vec{u}+\vec{v}*t\|_2^2 = (\vec{u}+\vec{v}*t)^H \cdot (\vec{u}+\vec{v}*t) \\ &= \vec{u}^H * \vec{u} + \bar{t}\cdot t * \vec{v}^H * \vec{v} + \vec{u}^H * \vec{v} * t + \vec{v}^H * \vec{u} * \bar{t} \\ &= \|\vec{u}\|_2^2 + |t|^2 \cdot \|\vec{v}\|_2^2 + |t| \cdot |\vec{u}^H * \vec{v}| \cdot \left(\omega^{(\varphi+\psi)} + \omega^{-(\varphi+\psi)}\right), \end{aligned}$$

gdzie $t = |t| \cdot \omega^\psi$, $\vec{u}^H * \vec{v} = |\vec{u}^H * \vec{v}| \cdot \omega^\varphi$, $\omega = \cos 1 + \imath \cdot \sin 1$.

Biorąc teraz $\psi = -\varphi$ mamy

$$0 \leq \|\vec{u}\|_2^2 + 2|t| \cdot |\vec{u}^H * \vec{v}| + |t|^2 \cdot \|\vec{v}\|_2^2,$$

a biorąc $\psi = \pi - \varphi$ mamy

$$0 \leq \|\vec{u}\|_2^2 - 2|t| \cdot |\vec{u}^H * \vec{v}| + |t|^2 \cdot \|\vec{v}\|_2^2.$$

Stąd dla dowolnej $\tau \in \mathbf{R}$ otrzymujemy

$$0 \leq \|\vec{u}\|_2^2 + 2\tau|\vec{u}^H * \vec{v}| + \tau^2\|\vec{v}\|_2^2.$$

Ponieważ prawa strona ostatniej nierówności jest, jako funkcja τ, trójmianem kwadratowym o wartościach nieujemnych, to

$$0 \geq \Delta = 4\left(|\vec{u} * \vec{v}|^2 - \|\vec{u}\|_2^2 \cdot \|\vec{v}\|_2^2\right),$$

co implikuje $|\vec{u}^H * \vec{v}| \leq \|\vec{u}\|_2 \cdot \|\vec{v}\|_2$ i kończy dowód.

Na podstawie nierówności Schwarza mamy teraz

$$\begin{array}{rcl} \|\vec{u} + \vec{v}\|_2^2 & = & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + \vec{u}^H * \vec{v} + \vec{v}^H * \vec{u} \\ & = & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2\Re(\vec{u}^H * \vec{v}) \\ & \leq & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2|\vec{u}^H * \vec{v}| \\ & \leq & \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 + 2\|\vec{u}\|_2\|\vec{v}\|_2 \\ & = & (\|\vec{u}\|_2 + \|\vec{v}\|_2)^2, \end{array}$$

czyli nierówność trójkąta dla $\|\cdot\|_2$.

3.2.2 Pożyteczne (nie)równości

Nietrudno pokazać następujące nierówności łączące normy p-te Schura dla $p = 1, 2, \infty$. Mianowicie, dla każdego $\vec{u} \in \mathbf{K}^n$ mamy

$$\begin{array}{rcccl} \|\vec{u}\|_\infty & \leq & \|\vec{u}\|_1 & \leq & n \cdot \|\vec{u}\|_\infty, \\ \|\vec{u}\|_\infty & \leq & \|\vec{u}\|_2 & \leq & \sqrt{n} \cdot \|\vec{u}\|_\infty, \\ \|\vec{u}\|_2 & \leq & \|\vec{u}\|_1 & \leq & \sqrt{n} \cdot \|\vec{u}\|_2, \end{array}$$

przy czym ostatnia z tych nierówności jest konsekwencją nierówności Schwarza,

$$\|\vec{u}\|_1 = \sum_{i=1}^{n} |u_i| = \sum_{i=1}^{n} |u_i| \cdot |1| \leq \left(\sum_{i=1}^{n} |u_i|^2 \right)^{1/2} \left(\sum_{i=1}^{n} 1^2 \right)^{1/2} = \sqrt{n} \cdot \|\vec{u}\|_2.$$

Dodatkowo zauważamy, że nierówności tych nie można poprawić. Na przykład, dla pierwszego wersora $\vec{e}_1$ mamy $\|\vec{e}_1\|_p = 1 \ \forall p$, a dla $\vec{1} = [1, 1, \ldots, 1] \in \mathbf{K}^n$ mamy $\|\vec{1}\|_1 = \sqrt{n}\|\vec{1}\|_2 = n\|\vec{1}\|_\infty$.

Kulą jednostkową w $\mathbf{K}^n$ (ze względu na normę $\| \cdot \|$) nazywamy zbiór wektorów

$$K \ = \ \{ \, \vec{u} \in \mathbf{K}^n : \ \|\vec{u}\| \leq 1 \} \, .$$

Z podanych powyżej nierówności wynika w szczególności, że

$$K_1 \subset K_2 \subset K_\infty,$$

gdzie K_p jest kulą jednostkową w normie p-tej Schura.

Zauważmy jeszcze, że normy p-te są monotoniczne oraz, że dla dowolnej macierzy permutacji $P \in \mathbf{K}^{n,n}$ i wektora $\vec{x} \in \mathbf{K}^n$

$$\|P * \vec{x}\|_p = \|\vec{x}\|_p,$$

tzn. norma p-ta wektora jest *niezmiennicza* ze względu na przestawienia kolejności jego współrzędnych.

3.3 Normy macierzy

3.3.1 Normy p-te

Normy p-te macierzy są definiowane (indukowane) przez normy p-te wektorów w następujący sposób:

$$\begin{aligned} \|A\|_p \quad &= \quad \sup_{\vec{0} \neq \vec{x} \in \mathbf{K}^n} \frac{\|A * \vec{x}\|_p}{\|\vec{x}\|_p} \\ &= \quad \sup \{ \, \|A * \vec{x}\|_p \ : \ \vec{x} \in \mathbf{K}^n, \|\vec{x}\|_p = 1 \} \, . \end{aligned}$$

Zauważmy, że używamy tego samego oznaczenia dla norm wektora jak i macierzy. Jest to uzasadnione, gdyż norma p-ta macierzy jest uogólnieniem normy p-tej wektora. Dla $A = [u_1, \ldots, u_m]^T \in \mathbf{K}^{m,1} = \mathbf{K}^m$ mamy bowiem $\|A\|_p = \sup_{|t|=1} \|A * t\|_p = (\sum_{i=1}^m |u_i|^p)^{1/p}$. (Tutaj $t \in \mathbf{K}$!)

Wprost z definicji wynika, że normy indukowane macierzy spełniają warunek *zgodności* (z normą wektorową), tzn.

$$\forall A \in \mathbf{K}^{m,n}\, \forall \vec{x} \in \mathbf{K}^n \quad \|A * \vec{x}\|_p \leq \|A\|_p \cdot \|\vec{x}\|_p.$$

Normy te są również submultiplikatywne,

$$\forall A \in \mathbf{K}^{m,l}\, \forall B \in \mathbf{K}^{l,n} \qquad \|A * B\|_p \leq \|A\|_p \cdot \|B\|_p.$$

Rzeczywiście, dla $\vec{x} \in \mathbf{K}^l$ mamy

$$\begin{aligned} \|(A * B) * \vec{x}\|_p &= \|A * (B * \vec{x})\|_p \leq \|A\|_p \cdot \|B * \vec{x}\|_p \\ &\leq \|A\|_p \cdot \|B\|_p \cdot \|\vec{x}\|_p, \end{aligned}$$

skąd

$$\sup_{\vec{x} \neq \vec{0}} \frac{\|(A * B) * \vec{x}\|_p}{\|\vec{x}\|_p} \leq \|A\|_p \cdot \|B\|_p.$$

Dla macierzy permutacji $P \in \mathbf{K}^{m,m}$ i $Q \in \mathbf{K}^{n,n}$ mamy

$$\|P * A * Q^T\|_p = \|A\|_p,$$

co oznacza, że przestawienie kolumn i wierszy macierzy nie zmienia jej p-tej normy. Rzeczywiście, ponieważ przestawienie współrzędnych nie zmienia normy p-tej wektora, mamy

$$\sup_{\vec{x} \neq \vec{0}} \frac{\|P * A * Q^T * \vec{x}\|_p}{\|\vec{x}\|_p} = \sup_{\vec{x} \neq \vec{0}} \frac{\|A * Q^T * \vec{x}\|_p}{\|Q^T * \vec{x}\|_p} = \sup_{\vec{y} \neq \vec{0}} \frac{\|A * \vec{y}\|_p}{\|\vec{y}\|_p}.$$

3.3.2 Pożyteczne (nie)równości

Dla niektórych p, normę można wyrazić w sposób pozwalający ją łatwo obliczyć.

Lemat 3.2 *Dla dowolnej macierzy $A = (a_{i,j}) \in \mathbf{K}^{m,n}$*

(a) $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{i,j}|$,

(b) $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{i,j}|$.

Dowód. (a) Dla $\vec{x} = [x_1, \ldots, x_n]^T \in \mathbf{K}^n$ mamy

$$\begin{aligned} \|A * \vec{x}\|_\infty &= \max_{1 \le i \le m} \left| \sum_{j=1}^n a_{i,j} \cdot x_j \right| \le \max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}| \cdot |x_j| \\ &\le \|\vec{x}\|_\infty \cdot \left(\max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}| \right). \end{aligned}$$

Z drugiej strony, weźmy $\vec{x}^* = (x_j^*)$ taki, że $x_j^* = \omega^{-\varphi_j}$, $1 \le j \le n$, gdzie φ_j jest argumentem liczby $a_{s,j}$, tzn. $a_{s,j} = |a_{s,j}|\omega^{\varphi_j}$, a s jest tym indeksem i, dla którego suma $\sum_{j=1}^n |a_{i,j}|$ jest największa. Wtedy $\|\vec{x}^*\|_\infty = 1$ oraz

$$\|A * \vec{x}^*\|_\infty \ge \left| \sum_{j=1}^n a_{s,j} \cdot x_j^* \right| = \left| \sum_{j=1}^n |a_{s,j}| \omega^{\varphi_j} \omega^{-\varphi_j} \right| = \sum_{j=1}^n |a_{s,j}|,$$

a stąd $\|A\|_\infty \ge \max_{1 \le i \le m} \sum_{j=1}^n |a_{i,j}|$.
(b) Dla dowolnego $\vec{x}$ mamy

$$\begin{aligned} \|A * \vec{x}\|_1 &= \sum_{i=1}^m \left| \sum_{j=1}^n a_{i,j} \cdot x_j \right| \le \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}| \cdot |x_j| \\ &= \sum_{j=1}^n |x_j| \cdot \sum_{i=1}^m |a_{i,j}| \le \left(\max_{1 \le j \le n} \sum_{i=1}^m |a_{i,j}| \right) \cdot \|\vec{x}\|_1. \end{aligned}$$

Z drugiej strony, dla $\vec{x}^*$ takiego, że $x_j^* = 0$ dla $j \ne s$, $x_j^* = 1$ dla $j = s$, gdzie s jest tym indeksem j dla którego suma $\sum_{i=1}^m |a_{i,j}|$ jest największa, mamy $\|\vec{x}^*\|_1 = 1$ oraz $\|A * \vec{x}\|_1 = \sum_{i=1}^m |a_{i,s}|$, a stąd $\|A\|_1 \ge \max_{1 \le j \le n} \sum_{i=1}^m |a_{i,j}|$.

Z powyższego lematu łatwo widać, że

$$\begin{aligned} \|A^T\|_\infty &= \|A^H\|_\infty = \|A\|_1, \\ \|A^T\|_1 &= \|A^H\|_1 = \|A\|_\infty. \end{aligned}$$

Szczególną rolę odgrywa norma druga $\|\cdot\|_2$, ze względów, które będą jasne później. Niestety, nie wyraża się ona w tak prosty sposób jak $\|\cdot\|_1$ i $\|\cdot\|_\infty$. W odróżnieniu od tych ostatnich, norma druga ma jednak dodatkową ważną własność; mianowicie, dla dowolnej $A \in \mathbf{K}^{m,n}$

$$\|A^T\|_2 = \|A^H\|_2 = \|A\|_2.$$

Równość ta wynika bezpośrednio z faktu, że

$$\|A\|_2 = \sup_{\vec{z}} \sup_{\vec{y}} \left|\vec{y}^H * A * \vec{z}\right|,$$

gdzie suprema wzięte są po $\vec{z} \in \mathbf{K}^n$ i $\vec{y} \in \mathbf{K}^m$ takich, że $\|\vec{z}\|_2 = 1 = \|\vec{y}\|_2$.

Rzeczywiście, dla dowolnych $\vec{y}$ i $\vec{z}$ o jednostkowych normach mamy

$$|\vec{y}^H * A * \vec{z}| \leq \|\vec{y}\|_2 \cdot \|A * \vec{z}\|_2 = \|A * \vec{z}\|_2 \leq \|A\|_2,$$

przy czym w pierwszej nierówności zastosowaliśmy nierówność Schwarza. Z drugiej strony, dla $\vec{z}$ o jednostkowej normie i takiego, że $A * \vec{z} \neq \vec{0}$ mamy

$$\|A * \vec{z}\|_2 \;=\; \frac{\|A * \vec{z}\|_2^2}{\|A * \vec{z}\|_2} \;=\; \frac{(A * \vec{z})^H * A * \vec{z}}{\|A * \vec{z}\|_2} \;\leq\; \sup_{\|\vec{y}\|_2 = 1} \left|\vec{y}^H * A * \vec{z}\right|,$$

gdzie podstawiliśmy $\vec{y} = A * \vec{z} / \|A * \vec{z}\|_2$.

3.3.3 Norma Frobeniusa

Normę *Frobeniusa* (albo *Euklidesową*) macierzy $A \in \mathbf{K}^{m,n}$ definiujemy jako

$$\|A\|_F \;=\; \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2 \right)^{1/2}.$$

Zaletą normy $\|\cdot\|_F$ jest jej łatwa "obliczalność", natomiast wadą, że nie jest to norma indukowana przez żadną normę wektorową.

Związek między normą Frobeniusa i normą drugą pokazuje następujący lemat.

Lemat 3.3 *Dla dowolnej* $A \in \mathbf{K}^{m,n}$ *mamy*

$$\|A\|_2 \;\leq\; \|A\|_F \;\leq\; \sqrt{\min(m,n)} \cdot \|A\|_2.$$

Dowód. Wykorzystując nierówność Schwarza, dla dowolnego $\vec{x} \in \mathbf{K}^n$ o jednostkowej normie drugiej mamy

$$\begin{aligned} \|A * \vec{x}\|_2^2 \;&=\; \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{i,j} \cdot x_j \right|^2 \;\leq\; \sum_{i=1}^{m} \left(\sum_{j=1}^{n} |a_{i,j}| \cdot |x_j| \right)^2 \\ &\leq\; \sum_{i=1}^{m} \left(\sum_{j=1}^{n} |a_{i,j}|^2 \right) \left(\sum_{j=1}^{n} |x_j|^2 \right) \;=\; \|A\|_F^2, \end{aligned}$$

a stąd $\|A\|_2 \le \|A\|_F$.

Z drugiej strony, przedstawiając A jako

$$A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n]\,, \qquad \vec{a}_j \in \mathbf{K}^m,$$

mamy $\|A\|_2 \ge \|A * \vec{e}_j\|_2 = \|\vec{a}_j\|_2$, gdzie $\vec{e}_j$ jest j-tym wersorem. Stąd

$$\|A\|_2^2 \;\ge\; \frac{1}{n} \cdot \sum_{j=1}^{n} \|a_j\|_2^2 \;=\; \frac{1}{n} \cdot \|A\|_F^2,$$

czyli $\|A\|_F \le \sqrt{n} \cdot \|A\|_2$. Ale również

$$\|A\|_F = \|A^T\|_F \le \sqrt{m} \cdot \|A^T\|_2 = \sqrt{m} \cdot \|A\|_2,$$

co kończy dowód.

Zauważymy jeszcze jedną własność norm p-tych macierzy. Niech macierz A będzie dana w postaci blokowej,

$$A = [A_1, A_2, \ldots, A_s]\,.$$

Wtedy

$$\begin{aligned} \|A_k\|_p &= \sup_{\|\vec{x}_k\|_p = 1} \|A_k * \vec{x}_k\|_p \;=\; \sup_{\|\vec{x}_k\|_p = 1, \vec{x}_j = \vec{0}, j \ne k} \left\| \sum_{j=1}^{s} A_j * \vec{x}_j \right\|_p \\ &\le \sup_{\|\vec{x}\|_p = 1} \|A * \vec{x}\|_p \;=\; \|A\|_p. \end{aligned}$$

Podobnie, jeśli

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_t \end{bmatrix}$$

to

$$\begin{aligned} \|A_k\|_p^p &= \sup_{\|\vec{x}\|_p = 1} \|A_k * \vec{x}\|_p^p \;\le\; \sup_{\|\vec{x}\|_p = 1} \sum_{j=1}^{t} \|A_j * \vec{x}\|_p^p \\ &= \sup_{\|\vec{x}\|_p = 1} \|A * \vec{x}\|_p^p \;=\; \|A\|_p^p. \end{aligned}$$

Stąd dostajemy wniosek, że jeśli A jest w postaci blokowej to dla każdego bloku $A_{i,j}$ mamy

$$\|A_{i,j}\|_p \leq \|A\|_p, \qquad 1 \leq p \leq \infty.$$

Oczywiście, ta własność zachodzi również dla normy Frobeniusa.

Rozdział 4

Przestrzenie liniowe

4.1 Przestrzenie i podprzestrzenie

4.1.1 Definicja i podstawowe własności

Niech $\mathcal{X}$ z działaniem dodawania '+' będzie grupą przemienną (abelową). Oznaczmy przez $\mathbf{0}$ element neutralny tej grupy, a przez $(-a)$ element przeciwny do $a \in \mathcal{X}$. Załóżmy ponadto, że w $\mathcal{X}$ zdefiniowane jest działanie '$*$' mnożenia przez *skalary*, czyli elementy pewnego ciała $\mathbf{K}$, które spełnia następujące warunki: [1]

(i) $\forall a \in \mathcal{X} \quad \forall \alpha \in \mathbf{K} \quad \alpha * a = a * \alpha \in \mathcal{X}$

(ii) $\forall a \in \mathcal{X} \quad 1 * a = a \qquad$ (gdzie 1 jest jedynką w $\mathbf{K}$)

(iii) $\forall a, b \in \mathcal{X} \quad \forall \alpha, \beta \in \mathbf{K}$

$$(\alpha + \beta) * a = \alpha * a + \beta * a$$
$$\alpha * (a + b) = \alpha * a + \alpha * b$$
$$(\alpha * \beta) * a = \alpha * (\beta * a).$$

Definicja 4.1 *Zbiór $\mathcal{X}$ z działaniami o wyżej wymienionych własnościach nazywamy* przestrzenią liniową nad ciałem $\mathbf{K}$ *i oznaczamy $\mathcal{X}_{|\mathbf{K}}$ (albo po prostu $\mathcal{X}$).*

[1] Zauważmy, że symbolu '$*$' używamy zarówno do oznaczenia mnożenia skalaru przez element z grupy jak i mnożenia skalaru przez skalar. Podobnie '+' oznacza zarówno dodawanie w ciele $\mathbf{K}$ jak i w grupie $\mathcal{X}$. Nie prowadzi to jednak do niejednoznaczności, bo z kontekstu zawsze wiadomo o jakie działanie chodzi.

Podamy kilka elementarnych własności przestrzeni liniowych:

- $\forall a \in \mathcal{X} \quad 0 * a = \mathbf{0}$
- $\forall a \in \mathcal{X} \quad (-1) * a = -a$
- $\forall \alpha \in \mathbf{K}\ \forall a \in \mathcal{X} \quad [\ \alpha * a = \mathbf{0} \iff (\alpha = 0) \text{ lub } (a = \mathbf{0})\ ]$

Pierwsza własność wynika z równości $0 * a = (0 + 0) * a = 0 * a + 0 * a$, a druga z równości $\mathbf{0} = 0 * a = (1 + (-1)) * a = a + (-1) * a$. Implikacja w lewą stronę w ostatniej własności jest oczywista. Aby pokazać implikację w prawą stronę załóżmy, że $\alpha * \mathbf{0} = \mathbf{0}$ i $\alpha \neq 0$. Wtedy

$$a = 1 * a = (\alpha^{-1} * \alpha) * a = \alpha^{-1} * (\alpha * a) = \alpha^{-1} * \mathbf{0} = \mathbf{0}.$$

Elementy przestrzeni liniowej $\mathcal{X}_{|\mathbf{K}}$ nazywamy zwykle *wektorami*, odwołując się do odpowiedniej interpretacji geometrycznej.

Przykładami przestrzeni liniowych są $\mathbf{R}^n_{|\mathbf{R}}$, $\mathbf{C}^n_{|\mathbf{R}}$, $\mathbf{C}^n_{|\mathbf{C}}$, $\mathbf{K}^{m,n}_{|\mathbf{K}}$. We wszystkich tych przykładach mnożenie wektora przez skalar zdefiniowane jest w naturalny sposób "wyraz po wyrazie". Przestrzeń liniową nad $\mathbf{R}$ (albo nad $\mathbf{C}$) tworzą też wielomiany stopnia co najwyżej $(n-1)$ o współczynnikach rzeczywistych (albo zespolonych). Oznaczamy ją przez $\mathcal{P}^n_{|\mathbf{R}}$ (albo $\mathcal{P}^n_{|\mathbf{C}}$).

4.1.2 Podprzestrzenie liniowe

Definicja 4.2 *Niech $\mathcal{X}_{|\mathbf{K}}$ będzie przestrzenią liniową. Niepusty podzbiór $\mathcal{Y} \subseteq \mathcal{X}$ nazywamy* podprzestrzenią (liniową) *przestrzeni $\mathcal{X}_{|\mathbf{K}}$, gdy $\mathcal{Y}$ jest przestrzenią liniową nad $\mathbf{K}$ (z działaniami jak w $\mathcal{X}$). Piszemy przy tym*

$$\mathcal{Y}_{|\mathbf{K}} \subseteq \mathcal{X}_{|\mathbf{K}}.$$

Twierdzenie 4.1 *Na to, aby $\mathcal{Y}_{|\mathbf{K}} \subseteq \mathcal{X}_{|\mathbf{K}}$ potrzeba i wystarcza, że:*

(i) $\forall a, b \in \mathcal{Y} \quad a + b \in \mathcal{Y}$

(ii) $\forall \alpha \in \mathbf{K} \quad \forall a \in \mathcal{Y} \quad \alpha * a \in \mathcal{Y}.$

Dowód. (i) i (ii) oznaczają, że dodawanie wektorów i mnożenie ich przez skalar nie wyprowadzają poza zbiór $\mathcal{Y}$. Pozostałe warunki bycia podprzestrzenią są w sposób oczywisty spełnione, bo są one spełnione w $\mathcal{X}$.

Szczególnymi przykładami podprzestrzeni są $\mathcal{Y} = \mathcal{X}$ (podprzestrzeń niewłaściwa) oraz $\mathcal{Y} = \{\mathbf{0}\}$ (podprzestrzeń zerowa).

Twierdzenie 4.2 *Część wspólna dowolnej rodziny podprzestrzeni przestrzeni liniowej $\mathcal{X}_{|\mathbf{K}}$ jest też podprzestrzenią $\mathcal{X}_{|\mathbf{K}}$.*

Dowód. Niech $\{\mathcal{Y}_j\}_{j\in J}$, gdzie J jest (być może nieskończonym) zbiorem indeksów, będzie dowolną rodziną podprzestrzeni. Oznaczmy

$$\mathcal{Y} = \bigcap_{j\in J} \mathcal{Y}_j.$$

Wobec twierdzenia 4.1 wystarczy pokazać, że działania dodawania i mnożenia przez skalar nie wyprowadzają poza zbiór $\mathcal{Y}$. Rzeczywiście, warunek $a, b \in \mathcal{Y}$ oznacza, że $a, b \in \mathcal{Y}_j$ dla wszystkich $j \in J$, a stąd również $a + b \in \mathcal{Y}_j$. W konsekwencji $a+b \in \cap_{j\in J}\mathcal{Y}_j = \mathcal{Y}$. Podobne uzasadnienie dla mnożenia przez skalar omijamy.

Ważnymi przykładami podprzestrzni liniowych przestrzeni macierzy $\mathbf{K}^{m,n}_{|\mathbf{K}}$ są $\mathrm{TRIL}^{m,n}$, $\mathrm{TRIU}^{m,n}$ oraz $\mathrm{DIAG}^{m,n}$. Podprzestrzeniami liniowymi w $\mathcal{P}^n_{|\mathbf{K}}$ są $\mathcal{P}^k_{|\mathbf{K}}$ z $k \leq n$, albo wielomiany w których zmienna występuje tylko w potęgach parzystych. (Przyjmujemy przy tym, że $-\infty$, czyli stopień wielomianu zerowego, jest liczbą parzystą.)

4.2 Baza i wymiar przestrzeni

4.2.1 Liniowa (nie)zależność

Niech $\{b_j\}_{j=1}^n \subset \mathcal{X}$ oraz i $\{\alpha_j\}_{j=1}^n \subset \mathbf{K}$. Element

$$b = \sum_{j=1}^n \alpha_j * b_j$$

nazywamy *kombinacją liniową* elementów $\{b_j\}$, przy czym liczby $\{\alpha_j\}$ są współczynnikami tej kombinacji.

Zauważmy, że

$$B = \mathrm{span}(b_1, b_2, \ldots, b_n) \;:=\; \Big\{ \sum_{j=1}^n \alpha_j * b_j \;:\; \{\alpha_j\}_{j=1}^n \subset \mathbf{K} \Big\},$$

czyli zbiór wszystkich kombinacji liniowych danych elementów $\{b_j\}$, jest podprzestrzenią przestrzeni $\mathcal{X}_{|\mathbf{K}}$. Mówimy, że B *jest rozpięta* na elementach $b_1, \ldots, b_n$.

Definicja 4.3 *Układ* $\{b_j\}_{j=1}^n \subset \mathcal{X}$ *jest* liniowo zależny *jeśli istnieje układ skalarów* $\{\alpha_j\}_{j=1}^n \subset \mathbf{K}$ *zawierający liczby niezerowe, dla którego*

$$\sum_{j=1}^n \alpha_j * b_j = \mathbf{0}.$$

Definicja 4.4 *Układ* $\{b_j\}_{j=1}^n \subset \mathcal{X}$ *jest* liniowo niezależny *jeśli nie jest liniowo zależny, tzn. gdy dla dowolnych skalarów* $\{\alpha_j\}_{j=1}^n$ *z równości*

$$\sum_{j=1}^n \alpha_j * b_j = \mathbf{0}$$

wynika, że $\alpha_j = 0$, $1 \le j \le n$.

Łatwo zauważyć, że dowolny (niepusty) podukład układu liniowo niezależnego jest układem liniowo niezależnym. Z drugiej strony, jeśli układ ma podukład liniowo zależny to układ wyjściowy jest liniowo zależny.

Rozpatrzmy dowolny układ $\{b_j\}_{j=1}^n$. Jeśli jest on liniowo zależny to istnieją $\{\alpha_j\}_{j=1}^n$ takie, że dla pewnego s mamy $\alpha_s \neq 0$ oraz $\sum_{j=1}^n \alpha_j * b_j = \mathbf{0}$. Wtedy

$$b_s = \sum_{s \neq j=1}^n \left(-\frac{\alpha_j}{\alpha_s} \right) * b_j,$$

czyli $b_s \in \operatorname{span}(b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_n)$, a stąd

$$\operatorname{span}(b_1, \ldots, b_s, \ldots, b_n) = \operatorname{span}(b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_n).$$

Można tak postępować dalej otrzymując w końcu układ liniowo niezależny rozpinający tą samą przestrzeń co $\{b_j\}_{j=1}^n$. (Ponieważ układ wyjściowy jest skończony, proces "wyjmowania" kolejnych wektorów musi się skończyć po co najwyżej n krokach.)

Wniosek 4.1 *Z każdego układu wektorów* $(b_1, \ldots, b_n)$ *można wyjąć podukład* $(b_{j(1)}, \ldots, b_{j(k)})$, $1 \le j(1) < \cdots < j(k) \le n$ *(*$0 \le k \le n$*) taki, że jest on liniowo niezależny oraz*

$$\operatorname{span}(b_1, \ldots, b_n) = \operatorname{span}(b_{j(1)}, \ldots, b_{j(k)}).$$

4.2.2 Baza i wymiar, twierdzenie Steinitza

Definicja 4.5 *Układ* $\{b_j\}_{j=1}^n$ *nazywamy* bazą *przestrzeni* $\mathcal{Y}_{|\mathbf{K}} \subseteq \mathcal{X}_{|\mathbf{K}}$ *gdy:*

(i) jest on liniowo niezależny,

(ii) $\mathcal{Y} = \mathrm{span}(b_1, b_2, \ldots, b_n)$.

Mamy następujące ważne twierdzenie.

Twierdzenie 4.3 *Każda przestrzeń liniowa* $\mathcal{Y}_{|\mathbf{K}}$ *ma bazę. Ponadto, wszystkie bazy są równoliczne.*

Twierdzenie to prowadzi do następującej definicji.

Definicja 4.6 *Liczbę elementów bazy danej przestrzeni* $\mathcal{Y}_{|\mathbf{K}}$ *nazywamy jej* wymiarem *i oznaczamy* $\dim(\mathcal{Y}_{|\mathbf{K}})$.

Dowód twierdzenia 4.3 o istnieniu i równoliczności baz udowodnimy teraz jedynie w przypadku przestrzeni rozpiętych na układach skończonych. Zauważmy najpierw, że z Wniosku 4.1 natychmiast wynika, iż takie przestrzenie mają bazę. Dowód równoliczności baz opiera się na następującym bardzo pożytecznym twierdzeniu.

Twierdzenie 4.4 (Steinitza o wymianie)
Niech

$$\mathrm{span}(b_1, \ldots, b_n) \subseteq \mathrm{span}(c_1, \ldots, c_m) = \mathcal{X},$$

przy czym układ $\{b_j\}_{j=1}^n$ *jest liniowo niezależny. Wtedy* $n \leq m$ *oraz* n *elementów układu* $\{c_j\}_{j=1}^n$ *można wymienić na* $\{b_j\}_{j=1}^n$ *otrzymując układ rozpinający* $\mathcal{X}$.

Dowód. (Indukcja względem n.)

Dla $n = 0$ teza jest oczywista. Załóżmy, że teza zachodzi dla $n-1$. Wtedy $n-1 \leq m$ oraz

$$\mathcal{X} = \mathrm{span}(b_1, \ldots, b_{n-1}, c_n, c_{n+1}, \ldots, c_m).$$

(Zakładamy bez zmniejszenia ogólności, że wymieniliśmy $n-1$ początkowych elementów układu $\{c_j\}_{j=1}^m$.) Ponieważ $b_n \in \mathcal{X}$ to można go przedstawić w postaci kombinacji liniowej

$$b_n = \sum_{j=1}^{n-1} \alpha_j * b_j + \sum_{j=n}^{m} \beta_j * c_j.$$

Zauważmy, że istnieje s, $n \le s \le m$, taka, że $\beta_s \neq 0$, bo w przeciwnym przypadku b_n byłby liniowo zależny od $b_1, \ldots, b_{n-1}$. Stąd $n \le m$ oraz

$$c_s = \frac{b_n}{\beta_s} - \sum_{j=1}^{n-1} \left(\frac{\alpha_j}{\beta_s}\right) * b_j - \sum_{s \neq j = n}^{m} \left(\frac{\beta_j}{\beta_s}\right) * c_j,$$

tzn. c_s jest liniową kombinacją wektorów $b_1, \ldots, b_n, c_n, \ldots, c_{s-1}, c_{s+1}, \ldots, c_m$. Wymieniając c_s na b_n dostajemy

$$\begin{aligned} \mathcal{X} = \operatorname{span}(c_1, \ldots, c_m) &= \operatorname{span}(b_1, \ldots, b_{n-1}, c_n, \ldots, c_m) \\ &= \operatorname{span}(b_1, \ldots, b_{n-1}, b_n, c_{n+1}, \ldots, c_m). \end{aligned}$$

To kończy dowód.

Biorąc teraz dwie bazy, $(b_1, \ldots, b_n)$ oraz $(c_1, \ldots, c_m)$, tej samej przestrzeni $\mathcal{Y}_{|\mathbf{K}}$ i stosując twierdzenie Steinitza otrzymujemy z jednej strony $n \le m$, a z drugiej $m \le n$. Stąd $m = n$, czyli bazy są równoliczne.

Z twierdzenia Steinitza można łatwo wywnioskować następujące własności. (Poniżej zakładamy, że $\dim(\mathcal{X}_{|\mathbf{K}}) < \infty$.)

1. Każdy układ liniowo niezależny w $\mathcal{X}$ można uzupełnić do bazy w $\mathcal{X}$.

2. Jeśli $\mathcal{Y}_{|\mathbf{K}} \subseteq \mathcal{X}_{|\mathbf{K}}$ to $\dim(\mathcal{Y}_{|\mathbf{K}}) \le \dim(\mathcal{X}_{|\mathbf{K}})$.

3. Niech $\mathcal{Y}_{|\mathbf{K}} \subseteq \mathcal{X}_{|\mathbf{K}}$. Wtedy

$$\mathcal{Y} = \mathcal{X} \iff \dim(\mathcal{Y}_{|\mathbf{K}}) = \dim(\mathcal{X}_{|\mathbf{K}}).$$

4.2.3 Przykłady

Podamy teraz kilka przykładów przestrzeni i ich baz.

- $$\mathbf{K}^m_{|\mathbf{K}} = \mathrm{span}(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m),$$

  gdzie $\vec{e}_j = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ jest j-tym wersorem (jedynka na j-tej współrzędnej). Stąd $\dim(\mathbf{K}^m_{|\mathbf{K}}) = m$.

- $$\mathbf{K}^{m,n}_{|\mathbf{K}} = \mathrm{span}(E_{i,j} :\ 1 \le i \le m, 1 \le j \le n),$$

  gdzie
  $$(E_{i,j})_{p,q} = \begin{cases} 1 & i = p, j = q, \\ 0 & \text{wpp.} \end{cases}$$
  Stąd $\dim(\mathbf{K}^{m,n}_{|\mathbf{K}}) = m \cdot n$.

- $$\mathbf{C}^{m,n}_{|\mathbf{R}} = \mathrm{span}(E_{i,j}, \imath \cdot E_{i,j} :\ 1 \le i \le m, 1 \le j \le n) \qquad (\imath = \sqrt{-1}).$$

  Stąd $\dim(\mathbf{C}^{m,n}_{|\mathbf{R}}) = 2 \cdot m \cdot n$.

- $$\mathcal{P}^n_{|\mathbf{R}} = \mathrm{span}(1, t, t^2, \ldots, t^{n-1})$$

  i $\dim(\mathcal{P}^n_{|\mathbf{R}}) = n$.

4.3 Sumy i sumy proste

4.3.1 Suma (prosta) dwóch podprzestrzeni

Niech $\mathcal{Y}$ i $\mathcal{Z}$ będą podprzestrzeniami $\mathcal{X}$. Definiujemy *iloczyn* tych podprzestrzeni jako
$$\mathcal{S} = \mathcal{Y} \cap \mathcal{Z} := \{x \in \mathcal{X} :\ x \in \mathcal{Y} \text{ i } x \in \mathcal{Z}\},$$
oraz *sumę* jako
$$\mathcal{T} = \mathcal{Y} + \mathcal{Z} := \{y + z :\ y \in \mathcal{Y}, z \in \mathcal{Z}\}.$$

Zauważmy, że suma podprzestrzeni nie jest zwykłą sumą teoriomnogościową.

Oczywiście, zarówno iloczyn $\mathcal{S}$ jak i suma $\mathcal{T}$ są podprzestrzeniami $\mathcal{X}$.

Definicja 4.7 *Jeśli iloczyn* $\mathcal{Y} \cap \mathcal{Z} = \{\mathbf{0}\}$ *to sumę* $\mathcal{Y} + \mathcal{Z}$ *nazywamy* sumą prostą *i oznaczamy*

$$\mathcal{T} = \mathcal{Y} \oplus \mathcal{Z}.$$

Podamy teraz kilka własności wymiarów sum i sum prostych.

(W1)

$$0 \le \dim(\mathcal{Y} \cap \mathcal{Z}) \le \min\left(\dim(\mathcal{Y}), \dim(\mathcal{Z})\right)$$

(W2)

$$\begin{aligned} \max\left(\dim(\mathcal{Y}), \dim(\mathcal{Z})\right) &\le \dim(\mathcal{Y} + \mathcal{Z}) \\ &\le \min\left(\dim(\mathcal{X}), \dim(\mathcal{Y}) + \dim(\mathcal{Z})\right) \end{aligned}$$

(W3)

$$\dim(\mathcal{Y} + \mathcal{Z}) = \dim(\mathcal{Y}) + \dim(\mathcal{Z}) - \dim(\mathcal{Y} \cap \mathcal{Z})$$

(W4)

$$\dim(\mathcal{Y} \oplus \mathcal{Z}) = \dim(\mathcal{Y}) + \dim(\mathcal{Z})$$

Własność (W1) jak i lewa strona (W2) wynikają po prostu z zawierania się odpowiednich podprzestrzeni, a prawa strona w (W2) z faktu, że $\mathcal{Y} + \mathcal{Z} \subseteq \mathcal{X}$ oraz, że suma teoriomnogościowa baz w $\mathcal{Y}$ i $\mathcal{Z}$ rozpina $\mathcal{Y} + \mathcal{Z}$.

Ponieważ (W4) wynika bezpośrednio z (W3), dla pełności dowodu wystarczy pokazać (W3). W tym celu bierzemy bazę $(b_1, \ldots, b_u)$ w $\mathcal{Y} \cap \mathcal{Z}$, a następnie uzupełniamy ją do bazy $(b_1, \ldots, b_u, y_{u+1}, \ldots, y_s)$ w $\mathcal{Y}$ oraz do bazy $(b_1, \ldots, b_u, z_{u+1}, \ldots, z_t)$ w $\mathcal{Z}$. Jasne jest, że

$$\mathrm{span}(y_{u+1}, \ldots, y_s) \cap \mathrm{span}(z_{u+1}, \ldots, z_t) = \{\mathbf{0}\},$$

bo inaczej wspólny element niezerowy byłby w $\mathcal{Y} \cap \mathcal{Z}$, a wówczas układ $(b_1, \ldots, b_u, y_{u+1}, \ldots, y_s)$ nie byłby liniowo niezależny.

Układ $(b_1, \ldots, b_u, y_{u+1}, \ldots, y_s, z_{u+1}, \ldots, z_t)$ jest więc liniowo niezależny i rozpina $\mathcal{Y} + \mathcal{Z}$, a więc jest też bazą tej przestrzeni. Dlatego

$$\begin{aligned} \dim(\mathcal{Y} + \mathcal{Z}) &= u + (s - u) + (t - u) \;=\; s + t - u \\ &= \dim(\mathcal{Y}) + \dim(\mathcal{Z}) - \dim(\mathcal{Y} \cap \mathcal{Z}). \end{aligned}$$

4.3.2 Suma (prosta) w ogólnym przypadku

Uogólnimy pojęcia sumy i sumy prostej na dowolną, ale skończoną, liczbę podprzestrzeni. Niech $\mathcal{Y}_j$, $1 \le j \le s$, będą podprzestrzeniami $\mathcal{X}$. Sumę tych podprzestrzeni definujemy jako

$$\begin{aligned} \mathcal{Y} &= \mathcal{Y}_1 + \mathcal{Y}_2 + \cdots + \mathcal{Y}_s = \sum_{j=1}^{s} \mathcal{Y}_j \\ &:= \{y_1 + \cdots + y_s : \; y_j \in \mathcal{Y}_j, \, 1 \le j \le s\}. \end{aligned}$$

Definicja 4.8 *Jeśli dla każdego* t, $1 \le t \le s$,

$$\mathcal{Y}_t \cap \left(\sum_{t \ne j=1}^{s} \mathcal{Y}_j \right) = \{\mathbf{0}\}$$

to sumę $\mathcal{Y}_1 + \cdots + \mathcal{Y}_s = \sum_{j=1}^{s} \mathcal{Y}_j$ *nazywamy* suma̧ prosta̧ *i oznaczamy*

$$\mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_s = \bigoplus_{j=1}^{s} \mathcal{Y}_j.$$

Twierdzenie 4.5 *Jeśli* $\mathcal{Y} = \oplus_{j=1}^{s} \mathcal{Y}_j$ *to każdy wektor* $y \in \mathcal{Y}$ *ma* jednoznaczne *przedstawienie w postaci*

$$y = y_1 + y_2 + \cdots + y_s, \qquad y_j \in \mathcal{Y}_j, \, 1 \le j \le s.$$

Dowód. (Indukcja względem s.)

Dla $s = 1$ twierdzenie jest w oczywisty sposób prawdziwe. Załóżmy, że jest ono prawdziwe dla $s - 1$. Niech

$$y = y_1 + \cdots + y_s = y'_1 + \cdots + y'_s.$$

Wtedy

$$\mathcal{Y}_s \ni y_s - y'_s = \sum_{j=1}^{s-1} (y'_j - y_j) \in \mathcal{Y}_1 + \cdots + \mathcal{Y}_{s-1},$$

a ponieważ $\mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_{s-1} \oplus \mathcal{Y}_s$ to $y_s = y'_s$ i $y_1 + \cdots + y_{s-1} = y'_1 + \cdots + y'_{s-1}$. Wobec tego, że $\mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_{s-1}$, co wynika wprost z definicji sumy prostej, możemy teraz skorzystać z założenia indukcyjnego, aby wywnioskować, że $y_j = y'_j$ dla $1 \le j \le s - 1$. To kończy dowód.

Zauważmy, że jeśli $\mathcal{Y} = \mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_s$ to suma teoriomnogościowa baz w $\mathcal{Y}_j$, $1 \le j \le s$, jest bazą $\mathcal{Y}$. W szczególnym przypadku, gdy $(b_1, \ldots, b_n)$ jest bazą $\mathcal{X}$ to

$$\mathcal{X} = \mathrm{span}(b_1) \oplus \cdots \oplus \mathrm{span}(b_n).$$

Ponadto, każdemu wektorowi $x \in \mathcal{X}$ można jednoznacznie przyporządkować współczynniki α_j, $1 \le j \le n$, takie, że

$$x = \sum_{j=1}^{n} \alpha_j * b_j.$$

4.4 Izomorfizm przestrzeni

Definicja 4.9 *Przestrzeń $\mathcal{X}_{|\mathbf{K}}$ jest* izomorficzna *z $\mathcal{Y}_{|\mathbf{K}}$ (obie przestrzenie nad tym samym ciałem) gdy istnieje wzajemnie jednoznaczne (różnowartościowe i "na") odwzorowanie*

$$f : \mathcal{X} \to \mathcal{Y}$$

zachowujące kombinacje liniowe, tzn. $\forall x_1, x_2 \in \mathcal{X}\ \forall \alpha_1, \alpha_2 \in \mathbf{K}$

$$f(\alpha * x_1 + \alpha_2 * x_2) = \alpha_1 * f(x_1) + \alpha_2 * f(x_2).$$

Odwzorowanie f nazywamy izomorfizmem.

Zauważmy, że jeśli $f : \mathcal{X} \to \mathcal{Y}$ jest izomorfizmem to $f(\mathbf{0}) = \mathbf{0}$ (bo $f(\mathbf{0}) = f(\mathbf{0} + \mathbf{0}) = f(\mathbf{0}) + f(\mathbf{0})$). Izomorfizm zachowuje też liniową (nie)zależność wektorów, co wynika z faktu, że warunek $\sum_{j=1}^{s} \alpha_j * f(b_j) = \mathbf{0}$ jest równoważny $f(\sum_{j=1}^{s} \alpha_j * b_j) = \mathbf{0}$, czyli $\sum_{j=1}^{s} \alpha_j * b_j = \mathbf{0}$. Stąd mamy prosty wniosek, że izomorfizm f przeprowadza bazę $(b_1, \ldots, b_n)$ przestrzeni $\mathcal{X}$ na bazę $(f(b_1), \ldots, f(b_n))$ przestrzeni $\mathcal{Y}$.

Ponadto mamy:

(i) każda przestrzeń jest izomorficzna ze sobą,

(ii) jeśli $\mathcal{X}$ jest izomorficzna z $\mathcal{Y}$ to $\mathcal{Y}$ jest izomorficzna z $\mathcal{X}$,

(iii) jeśli $\mathcal{X}$ jest izomorficzna z $\mathcal{Y}$ oraz $\mathcal{Y}$ jest izomorficzna z $\mathcal{Z}$ to $\mathcal{X}$ jest izomorficzna z $\mathcal{Z}$.

Aby pokazać (i) wystarczy zauważyć, że przekształcenie identycznościowe w $\mathcal{X}$ ustala izomorfizm $\mathcal{X}$ z $\mathcal{X}$. Dla (ii) wykażemy, że odwzorowanie odwrotne $f^{-1} : \mathcal{Y} \to \mathcal{X}$ ustala izomorfizm $\mathcal{Y}$ z $\mathcal{X}$. Rzeczywiście, jeśli $y_1, y_2 \in \mathcal{Y}$ to istnieją $x_1, x_2 \in \mathcal{X}$ takie, że $y_1 = f(x_1)$ i $y_2 = f(x_2)$. Stąd

$$\begin{aligned} &f^{-1}(\alpha_1 * y_1 + \alpha_2 * y_2) \\ &\quad = f^{-1}(\alpha_1 * f(x_1) + \alpha_2 * f(x_2)) = f^{-1}(f(\alpha_1 * x_1 + \alpha_2 * x_2)) \\ &\quad = \alpha_1 * x_1 + \alpha_2 * x_2 = \alpha_1 * f^{-1}(y_1) + \alpha_2 * f^{-1}(y_2). \end{aligned}$$

W końcu, aby pokazać (iii) zauważmy, że jeśli f i g są odpowiednio izomorfizmami $\mathcal{X}$ w $\mathcal{Y}$ oraz $\mathcal{Y}$ w $\mathcal{Z}$ to złożenie $h(\cdot) := g(f(\cdot))$ jest izomorfizmem $\mathcal{X}$ w $\mathcal{Z}$. Rzeczywiście,

$$\begin{aligned} &h(\alpha_1 * x_1 + \alpha_2 * x_2) \\ &\quad = g(f(\alpha_1 * x_1 + \alpha_2 * x_2)) = g(\alpha_1 * f(x_1) + \alpha_2 * f(x_2)) \\ &\quad = \alpha_1 * g(f(x_1)) + \alpha_2 * g(f(x_2)) = \alpha_1 * h(x_1) + \alpha_2 * h(x_2). \end{aligned}$$

Własności (i)-(iii) pokazują, że relacja "bycia przestrzeniami izomorficznymi" jest zwrotna, symetryczna i przechodnia, a więc jest *relacją równoważności.* Stąd, zbiór wszystkich przestrzeni liniowych nad ustalonym ciałem można podzielić na rozłączne podzbiory będące *klasami abstrakcji* tej relacji. Do tej samej klasy należą przestrzenie wzajemnie izomorficzne.

Wniosek 4.2 *Każda przestrzeń liniowa $\mathcal{X}_{|\mathbf{K}}$ wymiaru n jest izomorficzna z $\mathbf{K}^n_{|\mathbf{K}}$.*

Rzeczywiście, wybierając dowolną bazę $(b_1, \ldots, b_n)$ w $\mathcal{X}_{|\mathbf{K}}$ i definiując odwzorowanie $f : \mathcal{X} \to \mathcal{Y}$ jako

$$f\Big(\sum_{j=1}^{n} \alpha_j * b_j\Big) := \sum_{j=1}^{n} \alpha_j * \vec{e}_j$$

(gdzie $\vec{e}_j$ jest j-tym wersorem) otrzymujemy izomorfizm przestrzeni $\mathcal{X}_{|\mathbf{K}}$ w $\mathbf{K}^n_{|\mathbf{K}}$.

4.5 Warstwy modulo $\mathcal{Y}$

4.5.1 Definicja

Niech $\mathcal{Y}$ będzie podprzestrzenią przestrzeni $\mathcal{X}$ i niech $x_0 \in \mathcal{X}$.

Definicja 4.10 *Zbiór wektorów*

$$W(x_0,\mathcal{Y}) := \{\, x_0 + y : y \in \mathcal{Y}\,\}$$

nazywamy warstwą modulo $\mathcal{Y}$ *przez* x_0 *(albo* hiperpłaszczyzną *równoległą do* $\mathcal{Y}$ *przez punkt* x_0*).*

Zauważmy, że jeśli $x_1 - x_2 \in \mathcal{Y}$ to warstwy $W(x_1,\mathcal{Y})$ i $W(x_2,\mathcal{Y})$ zawierają te same wektory. Rzeczywiwiście, jeśli $x = x_1 + y \in W(x_1,\mathcal{Y})$ to $x = x_2 + ((x_1 - x_2) + y) \in W(x_2,\mathcal{Y})$. Podobnie, jeśli $x \in W(x_2,\mathcal{Y})$ to $x \in W(x_1,\mathcal{Y})$.

Z drugiej strony, jeśli $x \in W(x_1,\mathcal{Y}) \cap W(x_2,\mathcal{Y})$ to $x = x_1 + y_1 = x_2 + y_2$ dla pewnych $y_1, y_2 \in \mathcal{Y}$. Stąd $x_1 - x_2 = y_2 - y_1 \in \mathcal{Y}$ i w konsekwencji $W(x_1,\mathcal{Y}) = W(x_2,\mathcal{Y})$.

Na podstawie powyższej analizy możemy stwierdzić, że dwie warstwy, $W(x_1,\mathcal{Y})$ i $W(x_2,\mathcal{Y})$, są sobie równe (gdy $x_1 - x_2 \in \mathcal{Y}$) albo rozłączne (gdy $x_1 - x_2 \notin \mathcal{Y}$). Dlatego warstwy $W(x_1,\mathcal{Y})$ i $W(x_2,\mathcal{Y})$ takie, że $x_1 - x_2 \in \mathcal{Y}$ będziemy utożsamiać.

Trywialnymi przykładami warstw są $W(x_0,\mathcal{X}) = \mathcal{X}$ oraz $W(x_0,\{\mathbf{0}\}) = \{x_0\}$.

4.5.2 Przestrzeń warstw

W zbiorze wszystkich warstw modulo $\mathcal{Y}$ ($\mathcal{Y} \subseteq \mathcal{X}$) wprowadzimy działania dodawania warstw i mnożenia przez skalar $\alpha \in \mathbf{K}$ w następujący sposób:

(i) $W(x_1,\mathcal{Y}) + W(x_2,\mathcal{Y}) := W(x_1 + x_2,\mathcal{Y})$,

(ii) $\alpha * W(x,\mathcal{Y}) := W(\alpha * x,\mathcal{Y})$.

Działania te są dobrze zdefiniowane, bo jeśli

$$W(x_1,\mathcal{Y}) = W(x_1',\mathcal{Y}) \qquad \text{i} \qquad W(x_2,\mathcal{Y}) = W(x_2',\mathcal{Y})$$

to $x_1 - x_1' \in \mathcal{Y}$ i $x_2 - x_2' \in \mathcal{Y}$, a stąd $(x_1 - x_1') + (x_2 - x_2') \in \mathcal{Y}$, czyli $W(x_1 + x_2,\mathcal{Y}) = W(x_1' + x_2',\mathcal{Y})$. Podobnie, jeśli $W(x,\mathcal{Y}) = W(x',\mathcal{Y})$ to $\alpha * x - \alpha * x' = \alpha * (x - x') \in \mathcal{Y}$, czyli $W(\alpha * x,\mathcal{Y}) = W(\alpha * x',\mathcal{Y})$.

Łatwo sprawdzić, że zbiór warstw modulo $\mathcal{Y}$ z powyżej zdefiniowanymi działaniami jest przestrzenią liniową nad $\mathbf{K}$. Aby znaleźć bazę tej przestrzeni, zapiszemy $\mathcal{X}$ jako sumę prostą $\mathcal{X} = \mathcal{Y} \oplus \mathcal{Z}$ (gdzie $\mathcal{Z}$ jest oczywiście wyznaczona niejednoznacznie) i weźmiemy dowolną bazę $(z_1, z_2, \ldots, z_k)$ w $\mathcal{Z}$ (gdzie

$k = \dim(\mathcal{Z})$). Okazuje się, że przestrzeń warstw jest izomorficzna z $\mathcal{Z}$, a układ

$$(W(z_1, \mathcal{Y}), \ldots, W(z_k, \mathcal{Y}))$$

jest jej bazą. Aby się o tym przekonać, wystarczy pokazać, że odwzorowanie

$$f(z) = W(z, \mathcal{Y}), \qquad z \in \mathcal{Z},$$

jest izomorfizmem. Rzeczywiście, z definicji dodawania warstw i mnożenia przez skalar wynika, że f zachowuje kombinacje liniowe. Jest ono również różnowartościowe, bo jeśli $f(z_1) = f(z_2)$ to $z_1 - z_2 \in \mathcal{Y}$, a ponieważ $\mathcal{Y}$ i $\mathcal{Z}$ tworzą sumę prostą to $z_1 - z_2 = \mathbf{0}$ i $z_1 = z_2$. W końcu, f jest przekształceniem "na", bo dla dowolnej warstwy $W(x, \mathcal{Y})$, $x \in \mathcal{X}$, mamy $W(x, \mathcal{Y}) = f(z)$, gdzie z pochodzi z (jednoznacznego) rozkładu $x = y + z$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$.

W szczególności pokazaliśmy również, że przestrzeń warstw modulo $\mathcal{Y}$ ma wymiar $\dim(\mathcal{X}) - \dim(\mathcal{Y})$.

Na przykład, jeśli $\mathcal{Y} = \mathcal{X}$ to przestrzeń warstw jest izomorficza z przestrzenią zerową, a jeśli $\mathcal{Y} = \{\mathbf{0}\}$ to jest ona izomorficzna z $\mathcal{X}$.

Rozdział 5

Obraz, rząd i jądro macierzy

5.1 Obraz i rząd macierzy

5.1.1 Rząd kolumnowy i rząd wierszowy

Niech $A \in \mathbf{K}^{m,n}$ będzie dana w postaci blokowej,

$$A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n], \quad \vec{a}_j \in \mathbf{K}^m,\ 1 \le j \le n.$$

Obraz macierzy A definiujemy jako

$$\mathcal{R}(A) \;:=\; \{\, A * \vec{x} :\ \vec{x} \in \mathbf{K}^n \,\} \;=\; \mathrm{span}(\vec{a}_1, \vec{a}_2 \ldots, \vec{a}_n) \subseteq \mathbf{K}^m.$$

Dalej, *rząd kolumnowy* macierzy A definiujemy jako

$$\mathrm{rz}_k(A) \;:=\; \dim\left(\mathcal{R}(A)\right).$$

Oczywiście, $0 \le \mathrm{rz}_k(A) \le \min(m,n)$. Przedstawiając z kolei A jako wektory-wiersze (funkcjonały),

$$A = \begin{bmatrix} \hat{a}_1^T \\ \vdots \\ \hat{a}_m^T \end{bmatrix},$$

definiujemy *rząd wierszowy* macierzy A jako

$$\mathrm{rz}_w(A) \;=\; \dim\left(\mathcal{R}(A^T)\right) \;=\; \dim\left(\mathrm{span}(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n)\right).$$

Podobnie jak dla rzędu kolumnowego, $0 \le \mathrm{rz}_w(A) \le \min(m,n)$.

5.1.2 Rząd macierzy

Mamy następujące ważne twierdzenie.

Twierdzenie 5.1 *Dla dowolnej macierzy* $A \in \mathbf{K}^{m,n}$

$$\mathrm{rz}_k(A) = \mathrm{rz}_w(A).$$

Dowód. Oznaczmy

$$k = \mathrm{rz}_k(A) \quad \text{oraz} \quad w = \mathrm{rz}_w(A).$$

Zauważmy najpierw, że permutacja kolumn macierzy nie zmienia ani jej rzędu kolumnowego (bo to tylko zmiana kolejności wektorów) ani jej rzędu wierszowego (bo to tylko przenumerowanie współrzędnych, identyczne dla każdego z wektorów). Podobnie rzędów nie zmienia permutacja wierszy.

Dokonajmy więc, dla uproszczenia, takiej permutacji kolumn, a potem wierszy, aby otrzymana macierz $\hat{A}$ była postaci

$$\hat{A} = \begin{bmatrix} A_I & A_{II} \end{bmatrix},$$

gdzie $A_I \in \mathbf{K}^{m,k}$, $A_{II} \in \mathbf{K}^{m,n-k}$, $\mathrm{rz}_k(A_I) = k$, oraz

$$A_I = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

przy czym $A_1 \in \mathbf{K}^{w_1,k}$, $A_2 \in \mathbf{K}^{m-w_1,k}$, $w_1 := \mathrm{rz}_w(A_I) = \mathrm{rz}_w(A_1)$. Oczywiście

$$w_1 \le w,$$

bo wiersze A_1 są "obciętymi" wierszami $\hat{A}$.

Ponieważ wektory-wiersze macierzy A_2 są liniowo zależne od wektorów-wierszy macierzy A_1 to istnieje macierz $B \in \mathbf{K}^{w_1,m-w_1}$ taka, że $A_2^T = A_1^T * B$ (gdzie kolejne kolumny B są współczynnikami odpowiednich kombinacji liniowych), czyli $A_2 = B^T * A_1$. Dla dowolnego $\vec{x} \in \mathbf{K}^k$ mamy więc

$$A_I * \vec{x} = \begin{bmatrix} A_1 * \vec{x} \\ A_2 * \vec{x} \end{bmatrix} = \begin{bmatrix} A_1 * \vec{x} \\ B^T * A_1 * \vec{x} \end{bmatrix}.$$

Stąd, $A_1 * \vec{x} = \vec{0}$ wtedy i tylko wtedy gdy $A_I * \vec{x} = \vec{0}$, a ponieważ kolumny macierzy A_I są liniowo niezależne, oznacza to także liniową niezależność kolumn macierzy A_1. A jeśli tak to ich liczba k nie może przekroczyć w_1, czyli wymiaru przestrzeni do której należą.

Otrzymaliśmy więc, że

$$\mathrm{rz}_k(A) = \mathrm{rz}_k(\hat{A}) = k \le w_1 \le w = \mathrm{rz}_w(\hat{A}) = \mathrm{rz}_w(A).$$

Przeprowadzając podobne rozumowanie dla macierzy A^T otrzymujemy $\mathrm{rz}_w(A) \le \mathrm{rz}_k(A)$, a stąd ostatecznie $\mathrm{rz}_w(A) = \mathrm{rz}_k(A)$, co należało pokazać.

Na podstawie twierdzenia 5.1 poprawna jest następująca definicja rzędu macierzy.

Definicja 5.1 Rzędem *macierzy* A *nazywamy liczbę*

$$\mathrm{rz}(A) := \mathrm{rz}_k(A) = \mathrm{rz}_w(A).$$

5.2 Przestrzeń zerowa (jądro) macierzy

Dla $A \in \mathbf{K}^{m,n}$ zbiór

$$\mathcal{N}(A) := \left\{ \vec{x} \in \mathbf{K}^n : \; A * \vec{x} = \vec{0} \right\}$$

nazywamy *jądrem* macierzy A.

Niech $k = \mathrm{rz}(A)$. Załóżmy, że kolumny macierzy A zostały tak przestawione, że otrzymana macierz $\hat{A}$ ma postać

$$\hat{A} = \left[\begin{array}{cc} A_I & A_{II} \end{array} \right],$$

gdzie $A_I \in \mathbf{K}^{m,k}$, $A_{II} \in \mathbf{K}^{m,n-k}$, oraz $\mathrm{rz}(A_I) = \mathrm{rz}(\hat{A})$ $(= \mathrm{rz}(A))$. Jeśli tak to kolumny macierzy A_{II} są liniowo zależne od kolumn macierzy A_I. W konsekwencji $A_{II} = A_I * B$ dla pewnej $B \in \mathbf{K}^{k,n-k}$. Załóżmy teraz, że $\vec{x} \in \mathcal{N}(\hat{A})$. Przedstawiając $\vec{x}$ w postaci

$$\vec{x} = \begin{bmatrix} \vec{x}_I \\ \vec{x}_{II} \end{bmatrix},$$

$\vec{x}_I \in \mathbf{K}^k$, $\vec{x}_{II} \in \mathbf{K}^{n-k}$, mamy

$$\begin{aligned} \vec{0} \;&=\; \hat{A} * \vec{x} = \left[\begin{array}{cc} A_I & A_{II} \end{array} \right] \begin{bmatrix} \vec{x}_I \\ \vec{x}_{II} \end{bmatrix} = A_I * \vec{x}_I + A_{II} * \vec{x}_{II} \\ &=\; A_I * \vec{x}_I + A_I * B + \vec{x}_{II} \;=\; A_I * (\vec{x}_I + B * \vec{x}_{II}). \end{aligned}$$

Ostatnie wyrażenie jest liniową kombinacją kolumn macierzy A_I, a ponieważ kolumny te są liniowo niezależne to kombinacja ta daje wektor zerowy tylko wtedy gdy $\vec{x}_I + B * \vec{x}_{II} = \vec{0}$, czyli $\vec{x}_I = -B * \vec{x}_{II}$. Stąd

$$\begin{aligned} \mathcal{N}(\hat{A}) &= \left\{ \begin{bmatrix} -B * \vec{x}_{II} \\ \vec{x}_{II} \end{bmatrix} : \vec{x}_{II} \in \mathbf{K}^{n-k} \right\} \\ &= \left\{ \begin{bmatrix} -B \\ I_{n-k} \end{bmatrix} * \vec{x}_{II} : \vec{x}_{II} \in \mathbf{K}^{n-k} \right\}. \end{aligned}$$

Przedstawiając B kolumnowo, $B = [\vec{b}_1, \ldots, \vec{b}_{n-k}]$, otrzymujemy ostatecznie

$$\mathcal{N}(\hat{A}) = \mathcal{R}\left(\begin{bmatrix} -B \\ I_{n-k} \end{bmatrix} \right) = \operatorname{span}\left(\begin{bmatrix} -\vec{b}_1 \\ \vec{e}_1 \end{bmatrix}, \ldots, \begin{bmatrix} -\vec{b}_{n-k} \\ \vec{e}_{n-k} \end{bmatrix} \right),$$

gdzie jak zwykle $\vec{e}_j \in \mathbf{K}^{n-k}$ jest j-tym wersorem. Ponieważ $\vec{e}_1, \ldots, \vec{e}_{n-k}$ są liniowo niezależne to liniowo niezależne są też wektory w powyższym “span”. Stąd $\dim(\mathcal{N}(\hat{A})) = n - k = n - \mathrm{rz}(A)$. Wobec równości $\dim(\mathcal{N}(\hat{A})) = \dim(\mathcal{N}(A))$ (bo permutacja kolumn skutkuje jedynie przestawieniem współrzędnych w jądrze) dostajemy następujący wniosek.

Wniosek 5.1 *Dla dowolnej macierzy $A \in \mathbf{K}^{m,n}$*

$$\dim(\mathcal{N}(A)) + \dim(\mathcal{R}(A)) = n.$$

5.3 Rozkład względem obrazu i jądra

Zatrzymajmy się na chwilę na przypadku gdy $\mathbf{K} \subseteq \mathbf{C}$. Ponieważ wtedy

$$\overline{\left(\sum_{j=1}^{n} \vec{a}_j * x_j \right)} = \sum_{j=1}^{n} \overline{\vec{a}}_j * \overline{x}_j$$

(gdzie sprzężenie wektora oznacza sprzężenie “po współrzędnych”) to wektory $(\vec{a}_1, \ldots, \vec{a}_n)$ oraz $(\overline{\vec{a}}_1, \ldots, \overline{\vec{a}}_n)$ są jednocześnie albo liniowo niezależne, albo liniowo zależne. Stąd $\mathrm{rz}(\overline{A}) = \mathrm{rz}(A)$ (gdzie znów sprzężenie macierzy oznacza sprzężenie “po współrzędnych”). W konsekwencji,

$$\mathrm{rz}(A^H) = \mathrm{rz}(\overline{A}^T) = \mathrm{rz}(A^T) = \mathrm{rz}(A).$$

Łatwo można też wywnioskować inną własność; mianowicie, jeśli

$$A = B * C,$$

$A \in \mathbf{K}^{m,n}$, $B \in \mathbf{K}^{m,k}$, $C \in \mathbf{K}^{k,n}$, to

$$\mathrm{rz}(A) \le \min(\mathrm{rz}(B), \mathrm{rz}(C)).$$

Rzeczywiście, równość $A = B * C$ oznacza, że kolumny macierzy A są liniową kombinacją kolumn macierzy B, a stąd $\mathcal{R}(A) \subseteq \mathcal{R}(B)$ i w konsekwencji $\mathrm{rz}(A) \le \mathrm{rz}(B)$. Biorąc z kolei transpozycję mamy $A^T = C^T * B^T$ i to samo rozumowanie daje $\mathcal{R}(A^T) \subseteq \mathcal{R}(C^T)$ oraz

$$\mathrm{rz}(A) = \mathrm{rz}(A^T) \le \mathrm{rz}(C^T) = \mathrm{rz}(C).$$

Na koniec jeszcze jedno istotne twierdzenie.

Twierdzenie 5.2 *Niech* $\mathbf{K} \subseteq \mathbf{C}$ *i* $A \in \mathbf{K}^{m,n}$. *Wtedy*

$$\begin{aligned} \mathbf{K}^m &= \mathcal{R}(A) \oplus \mathcal{N}(A^H) \\ \mathbf{K}^n &= \mathcal{R}(A^H) \oplus \mathcal{N}(A). \end{aligned}$$

Dowód. Wystarczy pokazać pierwszą z tych równości. W tym celu najpierw uzasadnimy, że suma jest sumą prostą. Rzeczywiście, jeśli $\vec{y} \in \mathcal{R}(A) \cap \mathcal{N}(A^H)$ to $A^H * \vec{y} = \vec{0}$ oraz istnieje $\vec{x} \in \mathbf{K}^n$ taki, że $A * \vec{x} = \vec{y}$. Stąd

$$\|\vec{y}\|_2^2 \;=\; \vec{y}^H * \vec{y} \;=\; (A * \vec{x})^H * \vec{y} \;=\; \vec{x}^H * (A^H * \vec{y}) \;=\; \vec{0},$$

czyli $\vec{y} = \vec{0}$ i suma podprzestrzeni jest prosta.

Pozostaje pokazać, że wymiar sumy prostej wynosi m. Korzystając z wniosku 5.1 mamy bowiem

$$\begin{aligned} \dim\left(\mathcal{R}(A) \oplus \mathcal{N}(A^H)\right) &= \dim\left(\mathcal{R}(A)\right) + \dim\left(\mathcal{N}(A^H)\right) \\ &= \dim\left(\mathcal{R}(A)\right) + \left[m - \dim\left(\mathcal{R}(A^H)\right)\right] \\ &= \dim\left(\mathcal{R}(A)\right) + \left[m - \dim\left(\mathcal{R}(A)\right)\right] \\ &= m. \end{aligned}$$

Rozdział 6

Funkcjonały liniowe

6.1 Funkcjonały

6.1.1 Definicja i przykłady

Niech $\mathcal{X}_{|\mathbf{K}}$ będzie przestrzenią liniową, $\dim(\mathcal{X}_{|\mathbf{K}}) < \infty$.

Definicja 6.1 *Odwzorowanie*

$$s : \mathcal{X} \to \mathbf{K}$$

nazywamy funkcjonałem (liniowym) *na $\mathcal{X}_{|\mathbf{K}}$ gdy dla dowolnych $a, b \in \mathcal{X}$ i $\alpha, \beta \in \mathbf{K}$*

$$s(a * \alpha + b * \beta) = s(a) * \alpha + s(b) * \beta.$$

Zbiór wszystkich funkcjonałów (liniowych) na $\mathcal{X}_{|\mathbf{K}}$ oznaczamy przez $\mathcal{X}^*$.

Podamy teraz kilka przykładów funkcjonałów.

W przestrzeni wektorów $\mathbf{K}^n_{|\mathbf{K}}$ funkcjonałami są przekształcenia postaci

$$s(\vec{x}) = \hat{a}^T * \vec{x}, \qquad \forall \vec{x} \in \mathbf{K}^n,$$

gdzie $\hat{a} \in \mathbf{K}^n$ jest ustalonym wektorem. (Tu wyjaśnia się tajemnica nazwania wcześniej funkcjonałem macierzy jednowierszowej.)

W przestrzeni macierzy $\mathbf{K}^{m,n}_{|\mathbf{K}}$ funkcjonałami są np. $s_1(A) = a_{2,3}$, $s_2(A) = \mathrm{tr}(A) := \sum_{j=1}^{\min(m,n)} a_{j,j}$ (jest to *ślad* macierzy), przy czym $A = (a_{i,j}) \in \mathbf{K}^{m,n}$.

W przestrzeni wielomianów $\mathcal{P}^n_{|\mathbf{R}}$ funkcjonałami są np. $s_1(p) = p(2)$, $s_2(p) = 3 * p(-1) - 7 * p(3)$,

$$s_3(p) = \frac{d^2 p}{dt^2}\Big|_{t=1} = p''(1), \qquad s_4(p) = \int_0^1 p(t)dt,$$

przy czym $p \in \mathcal{P}^n$.

6.1.2 Przestrzeń sprzężona

Na zbiorze $\mathcal{X}^*$ możemy w naturalny sposób zdefiniować dodawanie funkcjonałów $s_1, s_2 \in \mathcal{X}^*$,

$$(s_1 + s_2)(a) := s_1(a) + s_2(a), \qquad \forall a \in \mathcal{X},$$

oraz mnożenie funkcjonału $s \in \mathcal{X}^*$ przez skalar $\alpha \in \mathbf{K}$,

$$(\alpha * s)(a) := \alpha * s(a), \qquad \forall a \in \mathcal{X}.$$

Twierdzenie 6.1 *Zbiór $\mathcal{X}^*$ z powyżej zdefiniowanymi działaniami dodawania funkcjonałów i mnożenia przez skalar jest przestrzenią liniową nad $\mathbf{K}$.*

Dowód tego twierdzenia polega na bezpośrednim sprawdzeniu warunków bycia przestrzenią liniową. Tutaj zauważymy tylko, że elementem zerowym $\mathcal{X}^*_{|\mathbf{K}}$ jest funkcjonał zerowy, $\mathbf{0}^*(a) = 0\ \forall a \in \mathcal{X}$, a elementem przeciwnym do $s \in \mathcal{X}^*$ jest funkcjonał $(-s)$ zdefiniowany jako $(-s)(a) = -s(a)\ \forall a \in \mathcal{X}$.

Definicja 6.2 *Przestrzeń $\mathcal{X}^*_{|\mathbf{K}}$ nazywamy* przestrzenią sprzężoną *do $\mathcal{X}_{|\mathbf{K}}$.*

Skoro $\mathcal{X}^*_{|\mathbf{K}}$ jest przestrzenią liniową to możemy spytać o jej wymiar i bazę.

Twierdzenie 6.2 *Mamy*

$$\dim\left(\mathcal{X}^*_{|\mathbf{K}}\right) = \dim\left(\mathcal{X}_{|\mathbf{K}}\right).$$

Ponadto, jeśli układ wektorów $(a_1, a_2, \ldots, a_n)$ jest bazą $\mathcal{X}_{|\mathbf{K}}$ to układ funkcjonałów $(s_1, s_2, \ldots, s_n)$ zdefiniowany warunkami

$$s_k(a_j) = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

gdzie $1 \leq j, k \leq n$, jest bazą $\mathcal{X}^*_{|\mathbf{K}}$.

Dowód. Zauważmy najpierw, że s_k są formalnie dobrze zdefiniowanymi funkcjonałami. Dla dowolnego $a = \sum_{j=1}^n a_j * \alpha_j \in \mathcal{X}$ mamy bowiem

$$s_k(a) = s_k\left(\sum_{j=1}^n a_j * \alpha_j\right) = \sum_{j=1}^n s_k(a_j) * \alpha_j = \alpha_k.$$

Stąd s_k "zwraca" k-tą współrzędną rozwinięcia wektora a w bazie wektorów $(a_1, \ldots, a_n)$.

Pokażemy najpierw liniową niezależność funkcjonałów s_k, $1 \leq k \leq n$. W tym celu załóżmy, że liniowa kombinacja $s := \sum_{k=1}^n s_k * \alpha_k = \mathbf{0}^*$. Wtedy, w szczególności, dla każdego j mamy $s(a_j) = 0$, a ponieważ

$$s(a_j) = \left(\sum_{k=1}^n s_k * \alpha_k\right)(a_j) = \sum_{k=1}^n s_k(a_j) * \alpha_k = \alpha_j$$

to $\alpha_j = 0$.

Pozostaje pokazać, że funkcjonały s_k, $1 \leq k \leq n$, rozpinają $\mathcal{X}^*$. Rzeczywiście, dla dowolnego $s \in \mathcal{X}^*$ oraz $a = \sum_{j=1}^n a_j * \alpha_j \in \mathcal{X}$ mamy

$$\begin{aligned} s(a) &= s\left(\sum_{j=1}^n a_j * \alpha_j\right) = \sum_{j=1}^n s(a_j) * \alpha_j \\ &= \sum_{j=1}^n \sigma_j * s_j(a) = \left(\sum_{j=1}^n \sigma_j * s_j\right)(a), \end{aligned}$$

gdzie $\sigma_j = s(a_j)$. Stąd $s = \sum_{j=1}^n \sigma_j * s_j$ jest kombinacją liniową funkcjonałów s_j i w konsekwencji $\mathcal{X}^* = \mathrm{span}(s_1, \ldots, s_n)$.

6.2 Refleksywność

6.2.1 Równość $\mathcal{X}$ i $\mathcal{X}^{**}$

Dla wygody wprowadzimy oznaczenie

$$s \cdot a := s(a), \qquad \forall s \in \mathcal{X}^* \; \forall a \in \mathcal{X}.$$

Zauważmy, że zapis $s \cdot a$ możemy traktować jako działanie funkcjonału s na wektor a, ale też odwrotnie, jako działanie wektora a na funkcjonał s. Ponieważ dodatkowo dla dowolnych $s_1, s_2 \in \mathcal{X}^*$ i $\alpha_1, \alpha_2 \in \mathbf{K}$ mamy

$$(\alpha_1 * s_1 + \alpha_2 * s_2) \cdot a = \alpha_1 * (s_1 \cdot a) + \alpha_2 * (s_2 \cdot a),$$

możemy traktować wektor a jako funkcjonał na $\mathcal{X}^*_{|\mathbf{K}}$, tzn. $a \in \mathcal{X}^{**} := (\mathcal{X}^*)^*$. Mamy więc $\mathcal{X} \subseteq \mathcal{X}^{**}$, a ponieważ na podstawie twierdzenia 6.2

$$\dim\left(\mathcal{X}_{|\mathcal{K}}\right) = \dim\left(\mathcal{X}^*_{|\mathcal{K}}\right) = \dim\left(\mathcal{X}^{**}_{|\mathcal{K}}\right)$$

to

$$\mathcal{X} = \mathcal{X}^{**}.$$

Ostatnia własność nazywa się *refleksywnością*. [1]

Doadajmy jeszcze, że jeśli $(s_1, \ldots, s_n)$ jest bazą $\mathcal{X}^*$ sprzężoną z bazą $(a_1, \ldots, a_n)$ to również odwrotnie, $(a_1, \ldots, a_n)$ jest bazą $\mathcal{X}^{**} = \mathcal{X}$ sprzężoną do $(s_1, \ldots, s_n)$. Wynika to bezpośrednio z faktu, że $s_j \cdot a_k$ wynosi 1 dla $j = k$ oraz zero dla $j \neq k$.

6.2.2 Przykłady

Podamy teraz przykłady baz i baz sprzężonych.

Niech $\mathcal{X}_{|\mathbf{K}} = \mathbf{K}^n_{|\mathbf{K}}$. Bazą sprzężoną do bazy $(\vec{e}_1, \ldots, \vec{e}_n)$ przestrzeni wektorów $\mathbf{K}^n_{|\mathbf{K}}$ jest $(\vec{e}_1^{\,T}, \ldots, \vec{e}_n^{\,T})$. Stąd w szczególności wynika że

$$(\mathbf{K}^n)^* = (\mathbf{K}^n)^T.$$

W ogólnym przypadku, bazą sprzężoną do dowolnej bazy $(\vec{a}_1, \ldots, \vec{a}_n)$ jest $(\hat{a}_1^T, \ldots, \hat{a}_n^T)$, gdzie wektory $\hat{a}_j$ są tak dobrane, że transpozycja macierzy $\hat{A} := [\hat{a}_1, \ldots, \hat{a}_n]$ jest lewą odwrotnością macierzy $A := [\vec{a}_1, \ldots, \vec{a}_n]$, tzn. $\hat{A}^T * A = I_n$. (Macierz $\hat{A}$ istnieje, bo istnieje baza sprzężona.)

Niech $\mathcal{X}_{|\mathbf{K}} = \mathcal{P}^n_{|\mathbf{R}}$. Wtedy bazę sprzężoną do bazy potęgowej wielomianów $(1, t, t^2, \ldots, t^{n-1})$ tworzą funkcjonały $(s_1, \ldots, s_n)$ zdefiniowane jako

$$s_k(p) = \frac{1}{(k-1)!} \frac{d^{k-1}p}{dt^{k-1}}\Big|_{t=0} = \frac{p^{(k-1)}(0)}{(k-1)!}, \qquad \forall p \in \mathcal{P}^n.$$

Jeśli zaś $s_k(p) = p(t_k)$, $1 \le k \le n$, gdzie $t_1 < t_2 < \cdots < t_n$ są ustalonymi punktami, to bazę sprzężoną do bazy funkcjonałów $(s_1, \ldots, s_n)$ tworzą wielomiany *Lagrange'a* $(l_1, \ldots, l_n)$ zdefiniowane jako

$$l_j(t) = \prod_{j \neq i = 1}^{n} \frac{t - t_i}{t_j - t_i}. \tag{6.1}$$

[1] Pokazaliśmy, że przestrzenie skończenie wymiarowe są refleksywne. Warto dodać, że własność ta w ogólności *nie* zachodzi dla przestrzeni nieskończenie wymiarowych.

Rzeczywiście, łatwo sprawdzić, że

$$s_k(l_j) = l_j(t_k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

6.3 Rozszerzenie rachunku macierzy

6.3.1 Macierze wektorów i funkcjonałów

W tym miejscu rozszerzymy nieco formalizm rachunku macierzy na macierze nieliczbowe, których elementami są wektory, a nawet funkcjonały. Pomoże nam to uprościć pewne rachunki na macierzach.

Niech $\mathcal{X}_{|\mathbf{K}}$ będzie przestrzenią liniową i $a_j \in \mathcal{X}$, $1 \leq j \leq k$. Wtedy możemy formalnie zdefiniować macierz jednowierszową wektorów

$$\mathbb{A} = [a_1, \ldots, a_k] \in \mathcal{X}^{1,k}.$$

Dla $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} \in \mathbf{K}^k$ definiujemy w naturalny sposób mnożenie

$$\mathbb{A} * \vec{\alpha} := \sum_{j=1}^{k} a_j * \alpha_j,$$

będące skrótowym zapisem kombinacji liniowej wektorów a_j.

Podobnie, mając dane $s_j \in \mathcal{X}^*$, $1 \leq j \leq l$, możemy zdefiniować macierz jednokolumnową funkcjonałów

$$\mathbb{S} = \begin{bmatrix} s_1 \\ \vdots \\ s_l \end{bmatrix} \in (\mathcal{X}^*)^{l,1}.$$

Dla $x \in \mathcal{X}$ definiujemy w naturalny sposób mnożenie

$$\mathbb{S} \cdot x := \begin{bmatrix} s_1 \cdot x \\ \vdots \\ s_l \cdot x \end{bmatrix} \in \mathbf{K}^{l,1}.$$

Co więcej, iloczyn $\mathbb{S} \cdot \mathbb{A}$ możemy również w naturalny sposób zdefiniować jako macierz

$$\mathbb{S} \cdot \mathbb{A} := (s_i \cdot a_j) \in \mathbf{K}^{l,k}.$$

Rzeczywiście, tak właśnie mnożymy macierz jednowierszową przez macierz jednokolumnową w przypadku macierzy liczbowych. Ponadto, dla dowolnego $\vec{\alpha} \in \mathbf{K}^k$ spełnona jest równość

$$\mathbb{S} \cdot (\mathbb{A} * \vec{\alpha}) = (\mathbb{S} \cdot \mathbb{A}) * \vec{\alpha}.$$

Idąc dalej możemy zapytać, czy ma sens mnożenie $\mathbb{A}$ przez $\mathbb{S}$. W przypadku macierzy liczbowych, mnożenie wektora-wiersza przez wektor-kolumnę jest możliwe tylko wtedy gdy wektory te mają tyle samo współrzędnych. Tak jest też teraz. Dokładniej jeśli $k = l$ to

$$\mathbb{A} * \mathbb{S} := \sum_{j=1}^{k} a_j * s_j$$

i interpretujemy ten zapis jako przekształcenie $\mathcal{X}$ w $\mathcal{X}$ zdefiniowane wzorem

$$(\mathbb{A} * \mathbb{S})(x) := \mathbb{A} * (\mathbb{S} \cdot x) = \sum_{j=1}^{k} a_j * s_j \cdot x.$$

W szczególności, $a \cdot s$ dla $a \in \mathcal{X}$ i $s \in \mathcal{X}^*$ jest przekształceniem “zwracającym” wektor a pomnożony przez wartość funkcjonału s.

6.3.2 Postać macierzowa izomorfizmów

Załóżmy teraz, że $\dim\big(\mathcal{X}_{|\mathbf{K}}\big) = n$ oraz $(a_1, \ldots, a_n)$ jest bazą $\mathcal{X}$. Niech $\mathbb{A} = [a_1, \ldots, a_n] \in \mathcal{X}^{1,n}$. Jasne jest, że wtedy odwzorowanie $f : \mathbf{K}^n \to \mathcal{X}$ zdefiniowane wzorem

$$f(\vec{\alpha}) := \mathbb{A} * \vec{\alpha}, \qquad \forall \vec{\alpha} \in \mathbf{K}^n,$$

jest izomorfizmem $\mathbf{K}^n$ w $\mathcal{X}$. Ponadto, izomorfizm odwrotny $f^{-1} : \mathcal{X} \to \mathbf{K}^n$ dany jest wzorem

$$f^{-1}(x) = \mathbb{S} \cdot x, \qquad \forall x \in \mathcal{X},$$

gdzie $\mathbb{S} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \in (\mathcal{X}^*)^{n,1}$ oraz $(s_1, \ldots, s_n)$ jest bazą sprzężoną do bazy $(a_1, \ldots, a_n)$.

Sprawdzamy, że w tym przypadku

$$\mathbb{S} \cdot \mathbb{A} = (s_i \cdot a_j)_{i,j=1}^n = I_n$$

jest identycznością w $\mathbf{K}^n$, oraz że dla dowolnego $x = \mathbb{A} * \vec{\alpha}$ z $\vec{\alpha} \in \mathbf{K}^n$

$$(\mathbb{A} * \mathbb{S})(x) = (\mathbb{A} * \mathbb{S})(\mathbb{A} * \vec{\alpha}) = \mathbb{A} * (\mathbb{S} \cdot \mathbb{A}) * \vec{\alpha} = \mathbb{A} * \vec{\alpha} = x,$$

czyli $\mathbb{A} * \mathbb{S}$ jest identycznością w $\mathcal{X}$.

Możemy więc uznać, że $\mathbb{S}$ jest odwrotnością $\mathbb{A}$, jak również, że $\mathbb{A}$ jest odwrotnością $\mathbb{S}$, tj.

$$\mathbb{S} = \mathbb{A}^{-1} \quad \text{oraz} \quad \mathbb{A} = \mathbb{S}^{-1}.$$

Rozdział 7

Układy równań liniowych

W tym rozdziale zajmiemy się rozwiązywaniem układów równań liniowych (2.2). Stosując zapis macierzowy zadanie formułujemy następująco. Dla danej macierzy (*współczynników*) $A \in \mathbf{K}^{m,n}$ oraz wektora (*wyrazów wolnych*) $\vec{b} \in \mathbf{K}^m$ należy znaleźć wszystkie wektory (*niewiadome*) $\vec{x}$ spełniające równość

$$A * \vec{x} = \vec{b}. \tag{7.1}$$

7.1 Zbiór rozwiązań

7.1.1 Twierdzenie Kroneckera-Capelliego

Mamy trzy możliwości:

(i) $\forall \vec{x} \in \mathbf{K}^n \quad A * \vec{x} \neq \vec{b} \quad \Longrightarrow$ układ jest *sprzeczny*

(ii) $\exists \vec{x} \in \mathbf{K}^n \quad A * \vec{x} = \vec{b} \quad \Longrightarrow$ układ jest *niesprzeczny*

(iii) $!\exists \vec{x} \in \mathbf{K}^n \quad A * \vec{x} = \vec{b} \quad \Longrightarrow$ układ jest *oznaczony* [1]

Twierdzenie 7.1 (KRONECKERA-CAPELLIEGO)
Układ $A * \vec{x} = \vec{b}$ jest niesprzeczny wtedy i tylko wtedy gdy

$$\mathrm{rz}(A) = \mathrm{rz}([A, \vec{b}]),$$

tzn. rząd macierzy A jest równy rzędowi A rozszerzonej o wektor $\vec{b}$.

[1] Symbol "$!\exists$" czytamy jako "istnieje dokładnie jeden".

Dowód. Jeśli $\mathrm{rz}([A,\vec{b}]) = \mathrm{rz}(A)$ to wektor $\vec{b}$ należy do przestrzeni rozpiętej przez wektory-kolumny macierzy A. To zaś oznacza, że $\vec{b}$ jest liniową kombinacją tych wektorów. Współczynniki tej kombinacji tworzą rozwiązanie $\vec{x}$ układu.

Z drugiej strony, jeśli istnieje $\vec{x} \in \mathbf{K}^n$ taki, że $A*\vec{x} = \vec{b}$ to $\vec{b}$ jest kombinacją liniową wektorów-kolumn macierzy A, czyli $\vec{b}$ należy do przestrzeni rozpiętej na tych wektorach. To zaś implikuje że $\mathrm{rz}([A,\vec{b}]) = \mathrm{rz}(A)$ i kończy dowód.

Możemy równoważnie stwierdzić, że układ $A * \vec{x} = \vec{b}$ jest niesprzeczny wtedy i tylko wtedy gdy $\vec{b} \in \mathcal{R}(A)$, czyli wektor wyrazów wolnych leży w obrazie macierzy współczynników.

7.1.2 Zbiór rozwiązań jako warstwa

Niech

$$\mathcal{L}(A,\vec{b}) \;=\; \{\, \vec{x} \in \mathbf{K}^n : \; A * \vec{x} = \vec{b} \,\}$$

będzie zbiorem wszystkich rozwiązań układu $A * \vec{x} = \vec{b}$.

Definicja 7.1 *Powiemy, że dwa układy,* $A_1 * \vec{x} = \vec{b}_1$ *oraz* $A_2 * \vec{x} = \vec{b}_2$*, są* równoważne *gdy mają ten sam zbiór rozwiązań, tzn. gdy*

$$\mathcal{L}(A_1,\vec{b}_1) = \mathcal{L}(A_2,\vec{b}_2).$$

Twierdzenie 7.2 *Jeśli układ* $A * \vec{x} = \vec{b}$ *jest niesprzeczny to zbiór rozwiązań*

$$\mathcal{L}(A,\vec{b}) \;=\; \{\, \vec{x}_0 + \vec{y} : \; \vec{y} \in \mathcal{N}(A) \,\} \;=\; W\left(\vec{x}_0, \mathcal{N}(A)\right),$$

gdzie $\vec{x}_0$ *jest dowolnym rozwiązaniem szczególnym układu.*

Dowód. Jeśli $\vec{x}_0$ jest rozwiązaniem szczególnym i $\vec{y} \in \mathcal{N}(A)$ to

$$A * (\vec{x}_0 + \vec{y}) = A * \vec{x}_0 + A * \vec{y} = \vec{b} + \vec{0} = \vec{b},$$

czyli $\vec{x}_0 + \vec{y}$ jest też rozwiązaniem. To zaś implikuje że $W\left(\vec{x}_0, \mathcal{N}(A)\right) \subseteq \mathcal{L}(A,\vec{b})$.

Z drugiej strony, jeśli $A*\vec{x} = \vec{b}$ to $A*(\vec{x}-\vec{x}_0) = \vec{b}-\vec{b} = \vec{0}$, czyli $(\vec{x}-\vec{x}_0) \in \mathcal{N}(A)$. A więc $\vec{x} = \vec{x}_0 + (\vec{x} - \vec{x}_0)$ jest z jednej strony rozwiązaniem układu, a z drugiej elementem warstwy $W\left(\vec{x}_0, \mathcal{N}(A)\right)$. Stąd $\mathcal{L}(A,\vec{b}) \subseteq W\left(\vec{x}_0, \mathcal{N}(A)\right)$.

7.1.3 Układy nieosobliwe

Rozpatrzmy przez chwilę układy z macierzami kwadratowymi.

Twierdzenie 7.3 *Macierz kwadratowa $A \in \mathbf{K}^{n,n}$ jest nieosobliwa wtedy i tylko wtedy gdy* $\mathrm{rz}(A) = n$.

Dowód. Wobec nierówności

$$n = \mathrm{rz}(I_n) = \mathrm{rz}(A * A^{-1}) \leq \min\big(\mathrm{rz}(A), \mathrm{rz}(A^{-1})\big)$$

mamy, że jeśli A jest nieosobliwa to $\mathrm{rz}(A) = n = \mathrm{rz}(A^{-1})$. Z drugiej strony, jeśli $\mathrm{rz}(A) = n$ to kolumny A są wektorami liniowo niezależnymi. Stąd istnieje macierz $X \in \mathbf{K}^{n,n}$ taka, że $A * X = I_n$. Podobnie, istnieje $Y \in \mathbf{K}^{n,n}$ taka, że $A^T * Y = I_n$, czyli $Y^T * A = I_n$. Ponadto

$$Y^T = Y^T * I_n = (Y^T * A) * X = I_n * X = X,$$

tzn. odwrotności lewostronna i prawostronna istnieją i są sobie równe, $A^{-1} = X = Y^T$. To kończy dowód.

Wiemy, że jeśli macierz kwadratowa A jest nieosobliwa to rozwiązaniem układu $A*\vec{x} = \vec{b}$ jest $\vec{x}^* := A^{-1}*\vec{b}$ i jest to jedyne rozwiązanie, tzn. układ jest oznaczony. Ale też odwrotnie, jeśli układ $A * \vec{x} = \vec{b}$ z macierzą kwadratową A jest oznaczony dla pewnego $\vec{b}$ to macierz A jest nieosobliwa. Rzeczywiście, wtedy jądro $\mathcal{N}(A) = \{\vec{0}\}$. To znaczy, że wektory-kolumny macierzy tworzą układ liniowo niezależny i $\mathrm{rz}(A) = n$. Na podstawie twierdzenia 2.1 wnioskujemy że A jest nieosobliwa.

Wniosek 7.1 *Niech A będzie macierzą kwadratową. Układ $A * \vec{x} = \vec{b}$ jest oznaczony wtedy i tylko wtedy gdy A jest nieosobliwa.*

7.2 Efektywna metoda rozwiązania

Dla danej macierzy $A \in \mathbf{K}^{m,n}$ i wektora $\vec{b} \in \mathbf{K}^m$ chcemy zbadać, czy układ (7.1) jest niesprzeczny, a jeśli tak to znaleźć zbiór jego rozwiązań (warstwę)

$$\mathcal{L}(A, \vec{b}) = \vec{x}_0 + \mathcal{N}(A),$$

gdzie $\mathcal{N}(A) = \mathrm{span}(\vec{z}_{s+1}, \vec{z}_{s+2}, \ldots, \vec{z}_n)$ i $s = \mathrm{rz}(A)$.

7.2.1 Ogólny schemat

Najpierw wyznaczymy $s = \mathrm{rz}(A)$ i $t = \mathrm{rz}([A, \vec{b}])$, a następnie w przypadku $s = t$ skonstruujemy rozwiązanie szczególne $\vec{x}_0$ oraz bazę $\vec{z}_{s+1}, \ldots, \vec{z}_n$ jądra $\mathcal{N}(A)$.

Ogólny schemat postępowania prowadzący do postaci równania, z którego można już stosunkowo łatwo odczytać rozwiązanie jest następujący. Startując z układu wyjściowego (7.1), który oznaczymy przez (U_0), kolejno dla $k = 1, 2, \ldots, s$ konstruujemy "prostsze" i (prawie) równoważne (U_0) układy (U_k) z macierzami $A^{(k)}$ tego samego formatu co A. Macierz $A^{(s)}$ układu docelowego (U_s) okaże się trójkątną górną.

Dopuszczamy przy tym następujące operacje na układnie równań:

(i) zamiana kolejności równań (wierszy macierzy),

(ii) zamiana kolejności niewiadomych (kolumn macierzy),

(iii) dodanie do jednego z równań innego równania pomnożonego przez skalar.

Zauważmy, że operacje (i) oraz (iii) prowadzą do układów równoważnych, zaś (ii) powoduje jedynie zmianę kolejności niewiadomych. W szczególności, rząd macierzy układu nie ulega zmianie.

Schemat sprowadzania układu wyjściowego do postaci z macierzą trójkątną górną przy pomocy zdefiniowanych operacji, który teraz dokładniej opiszemy, nazywamy ELIMINACJĄ GAUSSA.

7.2.2 Eliminacja Gaussa

Załóżmy, że wykonaliśmy już k przekształceń układu wyjściowego,

$$(U_0) \to (U_1) \to \ldots \to (U_k),$$

gdzie $0 \le k \le s-1$, otrzymując układ

$$A^{(k)} * \vec{x}^{(k)} = \vec{b}^{(k)}$$

z macierzą

$$A^{(k)} = \begin{bmatrix} R_k & T_k \\ 0 & V_k \end{bmatrix},$$

gdzie $R_k \in \mathrm{TRIU}^{k,k}$ jest kwadratową i nieosobliwą macierzą trójkątną górną. Oczywiście, założenie to jest spełnione dla $k = 0$, czyli dla układu wyjściowego, bowiem wtedy $A^{(0)} = V_0 = A$.

Krok $(k + 1)$-szy eliminacji

1. Wybieramy w V_k element różny od zera, powiedzmy $v_{p,q} \neq 0$, $k + 1 \leq p \leq m$, $k + 1 \leq q \leq n$. (Taki element istnieje, bo inaczej mielibyśmy $\mathrm{rz}(A) = \mathrm{rz}(A^{(k)}) = k < s = \mathrm{rz}(A)$.)

2. Przestawiamy wiersze $(k + 1)$-szy z p-tym oraz kolumny (niewiadome) $(k + 1)$-szą z q-tą.

3. Dokonujemy eliminacji $(k + 1)$-szej niewiadomej z równań od $(k + 1)$-szego do m-tego odejmując od równania i-tego równanie $(k + 1)$-sze pomnożone przez

$$l_{i,k+1} := a^{(k)}_{i,k+1}/a^{(k)}_{k+1,k+1}, \quad \text{dla } i = k + 1, k + 2, \ldots, m.$$

(Tutaj $a^{(k)}_{i,j}$ oznacza element aktualnie stojący na przecięciu i-tego wiersza i j-tej kolumny macierzy układu).

Po wykonaniu $(k + 1)$-szego kroku metody dostajemy układ z macierzą $A^{(k+1)}$, która ma wyzerowane współczynniki o indeksach (i, j) dla $1 \leq j \leq k + 1$, $j < i \leq m$.

Po wykonaniu $s = \mathrm{rz}(A)$ kroków dostajemy układ (U_s) postaci

$$A^{(s)} * \vec{x}^{(s)} = \vec{b}^{(s)},$$

przy czym

$$A^{(s)} = \begin{bmatrix} R_s & T_s \\ 0 & 0 \end{bmatrix},$$

a wektor $\vec{x}^{(s)}$ różni się od $\vec{x}^{(0)}$ jedynie przestawieniem (permutacją) współrzędnych. Rzeczywiście, gdyby odpowiednia podmacierz V_s macierzy $A^{(s)}$ była niezerowa to mielibyśmy $\mathrm{rz}(A) = \mathrm{rz}(A^{(s)}) > s$.

Dodajmy jeszcze, że w przypadkach $s = 0$ i $s = \min(m, n)$ niektóre bloki macierzy $A^{(s)}$ są puste.

7.2.3 Konstrukcja rozwiązania ogólnego

Przyjmując

$$\vec{x}^{(s)} = \begin{bmatrix} \vec{x}_I^{(s)} \\ \vec{x}_{II}^{(s)} \end{bmatrix}, \qquad \vec{b}^{(s)} = \begin{bmatrix} \vec{b}_I^{(s)} \\ \vec{b}_{II}^{(s)} \end{bmatrix},$$

gdzie $\vec{x}_I^{(s)}, \vec{b}_I^{(s)} \in \mathbf{K}^s$, $\vec{x}_{II}^{(s)} \in \mathbf{K}^{n-s}$, $\vec{b}_{II}^{(s)} \in \mathbf{K}^{m-s}$, układ (U_s) możemy zapisać jako

$$\left\{ \begin{array}{rcl} R_s * \vec{x}_I^{(s)} \; + \; T_s * \vec{x}_{II}^{(s)} & = & \vec{b}_I^{(s)} \\ \vec{0} & = & \vec{b}_{II}^{(s)} \end{array} \right. .$$

Jeśli teraz $\vec{b}_{II}^{(s)} \neq \vec{0}$ to układ jest sprzeczny i nie ma rozwiązań. Jeśli zaś $\vec{b}_{II}^{(s)} = \vec{0}$ to układ (U_s) redukuje się do

$$R_s * \vec{x}_I^{(s)} \; + \; T_s * \vec{x}_{II}^{(s)} \; = \; \vec{b}_I^{(s)}.$$

Przyjmując $\vec{x}_{II}^{(s)} = \vec{0}$ otrzymujemy rozwiązanie szczególne

$$\vec{x}_0^{(s)} = \begin{bmatrix} \vec{u}_0 \\ \vec{0} \end{bmatrix},$$

gdzie $\vec{u}_0 \in \mathbf{K}^s$ jest rozwiązaniem nieosobliwego układu

$$R_s * \vec{u}_0 = \vec{b}_I^{(s)}$$

z macierzą trójkątną dolną R_s.

Bazę jądra macierzy $A^{(s)}$,

$$\mathcal{N}(A^{(s)}) = \mathcal{N}([R_s, T_s]),$$

znajdujemy rozwiązując $(n-s)$ liniowo niezależnych rozwiązań układów jednorodnych

$$R_s * \vec{x}_I^{(s)} \; + \; T_s * \vec{x}_{II}^{(s)} \; = \; \vec{0}$$

kładąc kolejno za $\vec{x}_{II}^{(s)}$ wersory $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{n-s} \in \mathbf{K}^{n-s}$. Oznaczając

$$T_s = [\vec{t}_{s+1}, \vec{t}_{s+2}, \ldots, \vec{t}_n]$$

otrzymujemy

$$\vec{z}_j^{(s)} = \begin{bmatrix} \vec{u}_j \\ \vec{e}_{j-s} \end{bmatrix}, \qquad \text{gdzie} \quad R_s * \vec{u}_j^{(s)} = -\vec{t}_j,$$

$s + 1 \le j \le n$. Ostatecznie dostajemy

$$\mathcal{L}(A^{(s)}, \vec{b}^{(s)}) = \vec{x}_0 + \mathrm{span}(\vec{z}^{(s)}_{s+1}, \ldots, \vec{z}^{(s)}_n).$$

Są to również wszystkie rozwiązania układu wyjściowego, z dokładnością do odpowiedniej permutacji współrzędnych.

7.3 Interpretacja macierzowa eliminacji

Zobaczymy teraz jak proces eliminacji Gaussa wygląda z punktu widzenia rachunku macierzy.

7.3.1 Analiza operacji elementarnych

Załóżmy najpierw, dla uproszczenia, że nie musimy wykonywać przestawień ani wierszy ani kolumn układu (tzn. w każdym kroku element k-ty na głównej przekątnej jest niezerowy). Wtedy eliminacja w k-tym kroku odpowiada mnożeniu macierzy $A^{(k-1)}$ z lewej strony przez macierz kwadratową

$$L_k := I_m - \vec{l}_k * \vec{e}_k^T = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & -l_{k+1,k} & \ddots & & \\ & & & \vdots & & \ddots & \\ & & & -l_{m,k} & & & 1 \end{bmatrix} \in \mathbf{K}^{m,m},$$

gdzie $\vec{l}_k := [0, \ldots, 0, l_{k+1,k}, \ldots, l_{m,k}]^T$ oraz

$$l_{i,k} := \frac{a^{(k-1)}_{i,k}}{a^{(k-1)}_{k,k}}, \qquad k + 1 \le i \le m. \tag{7.2}$$

Dlatego

$$\begin{aligned} A^{(1)} &= L_1 * A, \\ A^{(2)} &= L_2 * A^{(1)} = L_2 * L_1 * A, \\ &\cdots \\ A^{(s)} &= L_s * A^{(s-1)} = \cdots = L_s * L_{s-1} * \cdots * L_1 * A, \end{aligned}$$

a stąd $A = (L_1^{-1} * \cdots * L_s^{-1}) * A^{(s)}$.

Zauważmy teraz, że

$$L_i^{-1} = (I_m - \vec{l}_i * \vec{e}_i^T)^{-1} = (I_m + \vec{l}_i * \vec{e}_i^T).$$

Rzeczywiście, wobec tego, że $\vec{e}_i^T * \vec{l}_i = 0$ mamy bowiem

$$(I_m - \vec{l}_i * \vec{e}_i^T) * (I_m + \vec{l}_i * \vec{e}_i^T) = I_m - \vec{l}_i * (\vec{e}_i^T * \vec{l}_i) * \vec{e}_i^T = I_m.$$

Stąd

$$\begin{aligned} L_1^{-1} * \cdots * L_s^{-1} &= (I_m + \vec{l}_1 * \vec{e}_1^T) * \cdots * (I_m + \vec{l}_s * \vec{e}_s^T) \\ &= I_m + \vec{l}_1 * \vec{e}_1^T + \cdots + \vec{l}_s * \vec{e}_s^T. \end{aligned}$$

Oznaczając $\hat{L} := L_1^{-1} * \cdots * L_s^{-1}$ oraz $\hat{R} := A^{(s)}$ otrzymujemy ostatecznie rozkład (faktoryzację) macierzy,

$$A = \hat{L} * \hat{R},$$

gdzie

$$\hat{L} = \begin{bmatrix} L & 0 \\ H & I \end{bmatrix} \in \mathbf{K}^{m,m}, \qquad \hat{R} = \begin{bmatrix} R & T \\ 0 & 0 \end{bmatrix} \in \mathbf{K}^{m,n},$$

$L \in \mathrm{TRIL}^{s,s}$ jest kwadratową macierzą trójkątną dolną z jedynkami na głównej przekątnej oraz $R \in \mathrm{TRIU}^{s,s}$ jest macierzą trójkątną górną.

Dodajmy jeszcze, że współczynniki $l_{i,k}$ macierzy $\hat{L}$ są dla $1 \leq k \leq s$, $k < i \leq m$, zdefiniowane przez (7.2).

Rozpatrzmy teraz ogólny przypadek, gdy dokonujemy przestawień wierszy i/lub kolumn macierzy. Przypomnijmy, że przestawienie wierszy i-tego z j-tym jest równoważne pomnożeniu macierzy przez elementarną macierz permutacji (transpozycję) $T_{i,j}$, natomiast pomnożenie macierzy z prawej strony przez $T_{i,j}$ jest równoważne przestawieniu kolumny i-tej z j-tą. Dlatego w obecności przestawień dostajemy

$$\begin{aligned} A^{(1)} &= L_1 * T_{1,p_1} * A * T_{1,q_1}, \\ A^{(2)} &= L_2 * T_{2,p_2} * A^{(1)} * T_{2,q_2} = L_2 * T_{2,p_2} * L_1 * T_{1,p_1} * A * T_{1,q_1} * T_{2,q_2} \\ &\cdots \\ A^{(s)} &= L_s * T_{s,p_s} * \cdots * T_{2,p_2} * L_1 * T_{1,p_1} * A * T_{1,q_1} * T_{2,q_2} * \cdots * T_{s,q_s}, \end{aligned}$$

przy czym zawsze $i \leq p_i$, $j \leq q_j$, $1 \leq i, j \leq s$.

Zauważmy, że

$$T_{2,p_2} * L_1 * T_{1,p_1} = (T_{2,p_2} * L_1 * T_{2,p_2}) * T_{2,p_2} * T_{1,p_1}.$$

Dalej

$$\begin{aligned} T_{2,p_2} * L_1 * T_{2,p_2} &= T_{2,p_2} * (I_m - \vec{l}_1 * \vec{e}_1^T) * T_{2,p_2} \\ &= I_m - (T_{2,p_2} * \vec{l}_1) * (\vec{e}_1^T * T_{2,p_2}) \\ &= I_m - \vec{l}_1' * \vec{e}_1^T \\ &=: L_1^{(1)}, \end{aligned}$$

gdzie $L_1^{(1)}$ różni się od L_1 jedynie przestawieniem wyrazów 2-go i p_2-go w pierwszej kolumnie. Uogólniając to rozumowanie dostajemy

$$\begin{aligned} & L_s * T_{s,p_s} * \cdots * T_{2,p_2} * L_1 * T_{1,p_1} \\ &= L_s * T_{s,p_s} * \cdots * L_2 * L_1^{(1)} * T_{2,p_2} * T_{1,p_1} \\ &= L_s * T_{s,p_s} * \cdots * (T_{3,p_3} * L_2 * T_{3,p_3}) * (T_{3,p_3} * L_1^{(1)} * T_{3,p_3}) \\ &\qquad\qquad * T_{3,p_3} * T_{2,p_2} * T_{1,p_1} \\ &= L_s * T_{s,p_s} * \cdots * L_3 * L_2^{(2)} * L_1^{(2)} * T_{3,p_3} * T_{2,p_2} * T_{1,p_1} \\ & \ldots \\ &= (L_s^{(s-1)} * L_{s-1}^{(s-1)} * \cdots * L_3^{(s-1)} * L_2^{(s-1)} * L_1^{(s-1)}) * (T_{s,p_s} * \cdots * T_{1,p_1}). \end{aligned}$$

Definiując macierze permutacji

$$Q^{-1} = Q^T := T_{1,q_1} * \cdots * T_{s,q_s}, \qquad P := T_{s,p_s} * \cdots * T_{1,p_1},$$

otrzymujemy ostatecznie

$$A^{(s)} = L_s^{(s-1)} * \cdots * L_1^{(s-1)} * P * A * Q^T = \hat{R},$$

czyli pożądany rozkład

$$P * A * Q^T = \hat{L} * \hat{R},$$

$\hat{L} = (L_1^{(s-1)})^{-1} * \cdots * (L_s^{(s-1)})^{-1}$, $\hat{R} = A^{(s)}$.

7.3.2 Rozkład trójkątno-trójkątny macierzy

Wynikiem naszej analizy jest następujące twierdzenie.

Twierdzenie 7.4 (O ROZKŁADZIE TRÓJKĄTNO-TRÓJKĄTNYM MACIERZY) *Dla dowolnej macierzy $A \in \mathbf{K}^{m,n}$ rzędu s istnieją (na ogół niejednoznaczne) macierze permutacji $P \in \mathbf{K}^{m,m}$ i $Q \in \mathbf{K}^{n,n}$ takie, że macierz $\hat{A} = P * A * Q^T$ ma jednoznaczny rozkład trójkątno-trójkątny*

$$\hat{A} = \hat{L} * \hat{R},$$

gdzie $\hat{L} \in \mathbf{K}^{m,m}$, $\hat{R} \in \mathbf{K}^{m,n}$, $\hat{l}_{i,i} = 1 \ \forall i$,

$$\hat{L} = \begin{bmatrix} L & 0 \\ H & I \end{bmatrix}, \qquad \hat{R} = \begin{bmatrix} R & T \\ 0 & 0 \end{bmatrix},$$

$L \in \mathrm{TRIL}^{s,s}$, $R \in \mathrm{TRIU}^{s,s}$.

Dowód. Istnienie rozkładu udowodniliśmy przez konstrukcję macierzy $\hat{L}$ i $\hat{R}$ (za pomocą eliminacji Gaussa). Aby pokazać jednoznaczność, przedstawimy $\hat{A}$ w postaci blokowej,

$$\hat{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}, \qquad A_{1,1} \in \mathbf{K}^{s,s}.$$

Wtedy dla danego rozkładu $\hat{A} = \hat{L} * \hat{R}$ mamy

$$A_{1,1} = L * R, \quad A_{1,2} = L * T, \quad A_{2,1} = H * R, \quad A_{2,2} = H * T.$$

Gdyby istniał inny rozkład z macierzami $\overline{L}$ i $\overline{R}$ to mielibyśmy $\overline{L} * \overline{R} = L * R$, czyli

$$L^{-1} * \overline{L} = R * \overline{R}^{-1}.$$

Po lewej stronie ostatniej równości mamy macierz trójkątną dolną z jedynkami na przekątnej, a z prawej trójkątną górną. To wymusza $L^{-1} * \overline{L} = I_s = R * \overline{R}^{-1}$, czyli $\overline{L} = L$ i $\overline{R} = R$. Jednoznaczność pozostałych bloków w rozkładzie wynika z równości $T = L^{-1} * A_{1,2}$ i $H = A_{2,1} * R^{-1}$.

7.4 Eliminacja bez przestawień

Podamy teraz warunek konieczny i dostateczny na to, aby istniał rozkład trójkątno-trójkątny oryginalnej macierzy A, a tym samym, aby eliminacja Gaussa była wykonalna bez przestawień wierszy i/lub kolumn.

Twierdzenie 7.5 *Aby macierz $A = (a_{i,j}) \in \mathbf{K}^{m,n}$ rzędu s miała rozkład trójkątno-trójkątny bez przestawień wierszy i/lub kolumn potrzeba i wystarcza, że wszystkie macierze kątowe $A_k = (a_{i,j})_{i,j=1}^{k} \in \mathbf{K}^{k,k}$ dla $k = 1, 2, \ldots, s$ są nieosobliwe.*

Dowód. Jeśli macierz ma rozkład $A = \hat{L}*\hat{R}$ to $A_k = \hat{L}_k*\hat{R}_k$ jest nieosobliwa dla $1 \le k \le s$. Dowód w drugą stronę podamy przez indukcję po s. Dla $s = 1$ twierdzenie jest oczywiste, bo wtedy $a_{1,1} \ne 0$ i eliminacja pierwszej kolumny jest wykonalna. Załóżmy, że twierdzenie jest prawdziwe dla $s - 1$. Oznaczając

$$A_s = \begin{bmatrix} A_{s-1} & \vec{a} \\ \vec{c}^T & a_{s,s} \end{bmatrix},$$

mamy z założenia indukcyjnego, że istnieje rozkład $A_{s-1} = L^{(s-1)} * R^{(s-1)}$ dla pewnych nieosobliwych macierzy

$$L^{(s-1)} \in \mathrm{TRIL}^{s-1,s-1} \quad \text{oraz} \quad R^{(s-1)} \in \mathrm{TRIU}^{s-1,s-1}.$$

Oznaczając

$$L^{(s)} = \begin{bmatrix} L^{(s-1)} & \vec{0} \\ \vec{g}^T & 1 \end{bmatrix}, \qquad R^{(s)} = \begin{bmatrix} R^{(s-1)} & \vec{b} \\ \vec{0}^T & r_{s,s} \end{bmatrix},$$

i rozwiązując równanie $A^{(s)} = L^{(s)} * R^{(s)}$ otrzymujemy

$$\begin{aligned} \vec{a} &= L^{(s-1)} * \vec{b}, \\ \vec{c}^T &= \vec{g}^T * R^{(s-1)} \qquad (\text{albo} \quad (R^{(s-1)})^T * \vec{g} = \vec{c}), \\ a_{s,s} &= \vec{g}^T * \vec{b} + r_{s,s}, \end{aligned}$$

skąd kolejno wyliczamy $\vec{b}$, $\vec{g}$ i $r_{s,s}$. Rozkład trójkątno-trójkątny macierzy kątowej $A^{(s)}$ w oczywisty sposób implikuje już rozkład całej macierzy A.

Na koniec podamy ważne klasy macierzy, dla których eliminacja Gaussa jest możliwa bez przestawień wierszy i/lub kolumn. Są to macierze:

(a) *diagonalnie dominujące wierszami*

$$\mathrm{WDD}^{n,n} = \left\{ A = (a_{i,j}) \in \mathbf{K}^{n,n} : \ |a_{i,i}| > \sum_{i \ne j=1}^{n} |a_{i,j}|, \ 1 \le i \le n \right\}.$$

(b) *diagonalnie dominujące kolumnami*

$$\mathrm{KDD}^{n,n} = \{A \in \mathbf{K}^{n,n} : \ A^T \in \mathrm{WDD}^{n,n}\}.$$

(c) *hermitowskie dodatnio określone*

$$\mathrm{HPD}^{n,n} = \{A \in \mathbf{K}^{n,n} : \ A^H = A \text{ oraz } \forall \vec{x} \neq \vec{0} \ \vec{x}^H * A * \vec{x} > 0\}.$$

(d) *hermitowskie ujemnie określone*

$$\mathrm{HND}^{n,n} = \{A \in \mathbf{K}^{n,n} : \ (-A) \in \mathrm{HPD}^{n,n}\}.$$

Aby pokazać, że w tych przypadkach przestawienie wierszy/kulumn nie jest konieczne, wykażemy, że spełnione są założenia twierdzenia 7.5.

W przypadku (a) zauważamy, że jeśli $A \in \mathrm{WDD}^{n,n}$ to A jest nieosobliwa, $\mathcal{N}(A) = \{\vec{0}\}$. Jeśli bowiem $A * \vec{x} = \vec{0}$ i $\vec{x} \neq \vec{0}$ to dla p takiego, że $|x_p| = \|\vec{x}\|_\infty$ mamy $a_{p,p}x_p + \sum_{j\neq p} a_{p,j}x_j = 0$, a stąd

$$|a_{p,p}| \leq \sum_{j\neq p} |a_{p,j}| \left|\frac{x_j}{x_p}\right| \leq \sum_{j\neq p} |a_{p,j}|,$$

co przeczy dominacji głównej diagonali macierzy. Uzasadnienie uzupełnia obserwacja, że jeśli $A \in \mathrm{WDD}^{n,n}$ to również macierze kątowe $A_k \in \mathrm{WDD}^{k,k}$, $1 \leq k \leq n$.

Dla przypadku (b) wystarczy zauważyć, że jeśli $A \in \mathrm{KDD}^{n,n}$ to $A^T \in \mathrm{WDD}^{n,n}$, oraz wykorzystać fakt, że nieosobliwość A jest równoważna nieosobliwości A^T.

W przypadku (c) (i zupełnie podobnie w (d)) zauważamy, że każda macierz $A \in \mathrm{HPD}^{n,n}$ jest nieosobliwa. Jeśli bowiem $\vec{x} \neq \vec{0}$ i $A * \vec{x} = \vec{0}$ to $\vec{x}^H * A * \vec{x} = 0$. Ponadto, wszystkie macierze kątowe $A_k \in \mathrm{HPD}^{k,k}$, bo dla dowolnego niezerowego $\vec{y} \in \mathbf{K}^k$ mamy

$$\vec{y}^H * A_k * \vec{y} = \begin{bmatrix} \vec{y} \\ \vec{0} \end{bmatrix}^H * A * \begin{bmatrix} \vec{y} \\ \vec{0} \end{bmatrix} > 0.$$

Rozdział 8

Przekształcenia liniowe

8.1 Podstawowe pojęcia i własności

Niech $\mathcal{X}_{|\mathbf{K}}$ i $\mathcal{Y}_{|\mathbf{K}}$ będą dwiema przestrzeniami liniowymi nad tym samym ciałem $\mathbf{K}$.

Definicja 8.1 *Przekształcenie $f : \mathcal{X} \to \mathcal{Y}$ nazywamy* przekształceniem liniowym *$\mathcal{X}$ w $\mathcal{Y}$ jeśli $\forall x, y \in \mathcal{X}$ $\forall \alpha, \beta \in \mathbf{K}$ zachodzi równość*

$$f(x * \alpha + y * \beta) = f(x) * \alpha + f(y) * \beta.$$

8.1.1 Obraz, jądro i rząd przekształcenia

Dla $\mathcal{X}_1 \subseteq \mathcal{X}$, zbiór

$$f(\mathcal{X}_1) := \{f(x) : x \in \mathcal{X}_1\}$$

nazywamy *obrazem* zbioru $\mathcal{X}_1$.

Jeśli $\mathcal{X}_1$ jest podprzestrzenią $\mathcal{X}$ to $f(\mathcal{X}_1)$ jest podprzestrzenią $\mathcal{Y}$. Rzeczywiście, jeśli $y_1, y_2 \in f(\mathcal{X}_1)$ to dla pewnych $x_1, x_2 \in \mathcal{X}_1$ mamy $y_1 = f(x_1)$ i $y_2 = f(x_2)$. Stąd dla dowolnych $\alpha_1, \alpha_2 \in \mathbf{K}$ mamy

$$y_1 * \alpha_1 + y_2 * \alpha_2 = f(x_1) * \alpha_1 + f(x_2) * \alpha_2 = f(x_1 * \alpha_1 + x_2 * \alpha_2) \in f(\mathcal{X}_1).$$

W szczególności, $f(\mathcal{X})$ oraz $f(\{\mathbf{0}\}) = \{\mathbf{0}\}$ są podprzestrzeniami.

Łatwo również sprawdzić, że obrazem warstwy $W(x_0, \mathcal{X}_1) \subseteq \mathcal{X}$ jest warstwa $W(f(x_0), f(\mathcal{X}_1)) \subseteq \mathcal{Y}$. A więc bycie podprzestrzenią, elementem zerowym albo warstwą są *niezmiennikami* przekształceń liniowych.

Podobnie jak dla macierzy definiujemy *obraz* przekształcenia liniowego f

$$\mathrm{im}(f) := f(\mathcal{X}) = \{f(x) : x \in \mathcal{X}\} \subseteq \mathcal{Y},$$

jego *jądro*

$$\ker(f) := \{x \in \mathcal{X} : f(x) = \mathbf{0}\} \subseteq \mathcal{X},$$

oraz *rząd*

$$\mathrm{rank}(f) := \dim(\mathrm{im}(f)).$$

Oczywiście, jądro jest też podprzestrzenią.

Twierdzenie 8.1 *Dla dowolnego przekształcenia liniowego f mamy*

$$\dim(\mathcal{X}) = \dim(\mathrm{im}(f)) + \dim(\ker(f)).$$

Dowód. Niech $\mathcal{X}_1$ będzie tak zdefiniowane, że

$$\mathcal{X} = \mathcal{X}_1 \oplus \ker(f).$$

Wtedy $\dim(\mathcal{X}) = \dim(\mathcal{X}_1) + \dim(\ker(f))$. Pokażemy, że $\dim(\mathrm{im}(f)) = \dim(\mathcal{X}_1)$. W tym celu zauważmy, że każdy $x \in \mathcal{X}$ można jednoznacznie przedstawić jako $x = x_1 + x_0$, gdzie $x_1 \in \mathcal{X}_1$ i $x_0 \in \ker(f)$. Stąd

$$\mathrm{im}(f) = \{f(x_1 + x_0) : x_1 \in \mathcal{X}_1, x_0 \in \ker(f)\} = \{f(x_1) : x_1 \in \mathcal{X}_1\}.$$

Teraz wystarczy pokazać, że $\dim(\mathcal{X}_1) = \dim(f(\mathcal{X}_1))$. Rzeczywiście, niech

$$\mathbb{A} = [x_1, \ldots, x_s] \in \mathcal{X}^{1,s}$$

będzie bazą $\mathcal{X}_1$ ($s = \dim(\mathcal{X}_1)$) oraz

$$\mathbb{B} = [f(x_1), \ldots, f(x_s)] \in \mathcal{Y}^{1,s}.$$

Wtedy $f(\mathcal{X}_1) = \mathrm{span}(f(x_1), \ldots, f(x_s))$ oraz układ $\{f(x_j)\}_{j=1}^{s}$ jest liniowo niezależny. Jeśli bowiem $\mathbb{B} * \vec{\alpha} = \mathbf{0}$ to również $f(\mathbb{A} * \vec{\alpha}) = \mathbf{0}$. Ponieważ $\mathbb{A} * \vec{\alpha} \notin \ker(f) \setminus \{\mathbf{0}\}$ to $\mathbb{A} * \vec{\alpha} = \mathbf{0}$ i z liniowej niezależności $\{x_j\}_{j=1}^{s}$ dostajemy $\vec{\alpha} = \vec{0}$. Otrzymaliśmy, że $\mathbb{B}$ jest bazą $f(\mathcal{X}_1)$ i $\dim(f(\mathcal{X}_1)) = s = \dim(\mathcal{X}_1)$.

8.1.2 Przykłady

- Każda macierz $A \in \mathbf{K}^{m,n}$ może być identyfikowana z przekształceniem liniowym $f : \mathbf{K}^n \to \mathbf{K}^m$ danym wzorem

$$f(\vec{x}) = A * \vec{x}, \qquad \vec{x} \in \mathbf{K}^n.$$

  Wtedy $\mathrm{im}(f) = \mathcal{R}(A)$, $\ker(f) = \mathcal{N}(A)$ oraz $\mathrm{rank}(f) = \mathrm{rz}(A)$. Twierdzenie 8.1 sprowadza się w tym przypadku do wniosku 5.1.

  W szczególności, funkcjonały liniowe są przekształceniami liniowymi. Wtedy $A \in \mathbf{K}^{1,n}$ oraz $\mathcal{Y} = \mathbf{K}$.

- Niech $f : \mathcal{P}^{10}_{|\mathbf{R}} \to \mathcal{P}^{10}_{|\mathbf{R}}$, $f(p) = p''$ (druga pochodna). Wtedy $\ker(f) = \mathcal{P}^{2}_{|\mathbf{R}}$ i $\mathrm{im}(f) = \mathcal{P}^{8}_{|\mathbf{R}}$.

- Jeśli zaś w poprzednim przykładzie $f(p) = p' - p$ to $\mathrm{im}(f) = \mathcal{P}^{10}_{|\mathbf{R}}$ oraz $\ker(f) = \mathcal{P}^{0}_{|\mathbf{R}} = \{\mathbf{0}\}$.

8.1.3 Różnowartościowość

Twierdzenie 8.2 *Na to, aby przekształcenie liniowe $f : \mathcal{X} \to \mathcal{Y}$ było różnowartościowe potrzeba i wystarcza, że* $\ker(f) = \{\mathbf{0}\}$.

Dowód. Jeśli f jest różnowartościowe to tylko dla $x = \mathbf{0}$ mamy $f(x) = \mathbf{0}$, czyli $\ker(f) = \{\mathbf{0}\}$. Z drugiej strony, jeśli $\ker(f) = \{\mathbf{0}\}$ i $f(x_1) = f(x_2) = \mathbf{0}$ to $f(x_1 - x_2) = \mathbf{0}$, a stąd $x_1 - x_2 = \mathbf{0}$ i $x_1 = x_2$, co kończy dowód.

Z ostatniego twierdzenia wynika, że jeśli $\ker(f) = \{\mathbf{0}\}$ to istnieje przekształcenie "odwrotne" $f^{-1} : \mathrm{im}(f) \to \mathcal{X}$ takie, że $\forall x \in \mathcal{X}\ f^{-1}(f(x)) = x$ oraz $\forall y \in \mathrm{im}(f)\ f(f^{-1}(y)) = y$. Ponadto f^{-1} jest liniowe, bo jeśli $y_1, y_2 \in \mathrm{im}(f)$ to definiując $x_1 = f^{-1}(y_1)$ i $x_2 = f^{-1}(y_2)$ mamy

$$\begin{aligned} f^{-1}(y_1 * \alpha_1 + y_2 * \alpha_2) &= f^{-1}\left(f(x_1) * \alpha_1 + f(x_2) * \alpha_2\right) \\ &= f^{-1}\left(f(x_1 * \alpha_1 + x_2 * \alpha_2)\right) \\ &= x_1 * \alpha_1 + x_2 * \alpha_2 \\ &= f^{-1}(y_1) * \alpha_1 + f^{-1}(y_2) * \alpha_2. \end{aligned}$$

Mówiąc inaczej, każde różnowartościowe przekształcenie liniowe $f : \mathcal{X} \to \mathcal{Y}$ ustala izomorfizm pomiędzy $\mathcal{X}$ i swoim obrazem $\mathrm{im}(f) \subseteq \mathcal{Y}$.

8.1.4 Przestrzeń przekształceń liniowych

Zbiór wszystkich przekształceń liniowych z $\mathcal{X}$ w $\mathcal{Y}$ tworzy przestrzeń liniową nad $\mathbf{K}$, jeśli działania dodawania przekształceń i mnożenia przez skalar zdefiniowane są w naturalny sposób jako:

$$(\alpha * f)(x) = \alpha * f(x), \qquad (f+g)(x) = f(x) + g(x).$$

Przestrzeń tą oznaczamy $(\mathcal{X} \to \mathcal{Y})_{|\mathbf{K}}$ albo $\mathcal{L}in(\mathcal{X}, \mathcal{Y})$. Oczywiście, elementem neutralnym (zerowym) tej przestrzeni jest przekształcenie zerowe, a przeciwnym do f jest $(-f)$.

Podobnie jak dla funkcjonałów, dla wygody będziemy często stosować zapis

$$f \cdot x := f(x), \qquad \forall f \in \mathcal{L}in(\mathcal{X}, \mathcal{Y}) \quad \forall x \in \mathcal{X}.$$

Uwaga. Zauważmy, że wobec równości

$$(\alpha * f + \beta * g) \cdot x = \alpha * (f \cdot x) + \beta * (g \cdot x)$$

każdy wektor $x \in \mathcal{X}$ może być traktowany jako element przestrzeni

$$\mathcal{L}in\left(\mathcal{L}in(\mathcal{X}, \mathcal{Y}), \mathcal{Y}\right).$$

Jednak w ogólności nie mamy równości pomiędzy $\mathcal{L}in\left(\mathcal{L}in(\mathcal{X}, \mathcal{Y}), \mathcal{Y}\right)$ i $\mathcal{X}$, tak jak jest w przypadku $\mathcal{Y} = \mathbf{K}$.

8.2 Macierz przekształcenia liniowego

8.2.1 Definicja

Niech $\dim(\mathcal{X}) = n$, $\dim(\mathcal{Y}) = m$. Niech

$$\mathbb{A} = [x_1, \ldots, x_n] \in \mathcal{X}^{1,n}, \qquad \mathbb{B} = [y_1, \ldots, y_m] \in \mathcal{Y}^{1,m}$$

będą odpowiednio bazami $\mathcal{X}$ i $\mathcal{Y}$. Wtedy

$$\mathcal{X} = \{\mathbb{A} * \vec{a} : \vec{a} \in \mathbf{K}^n\}, \qquad \mathcal{Y} = \{\mathbb{B} * \vec{b} : \vec{b} \in \mathbf{K}^m\}.$$

Przypomnijmy, że $\mathbb{B}^{-1}$ jest wektorem funkcjonałów,

$$\mathbb{B}^{-1} = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \in (\mathcal{Y}^*)^{m,1},$$

gdzie $r_j \in \mathcal{Y}^*$, $1 \leq j \leq m$, tworzą bazę $\mathcal{Y}^*$ sprzężoną do $\mathbb{B}$.

Niech $f : \mathcal{X} \to \mathcal{Y}$ będzie przekształceniem liniowym i $y = f \cdot x$. Przyjmując $x = \mathbb{A} * \vec{a}$ i $y = \mathbb{B} * \vec{b}$ mamy

$$\begin{aligned} \vec{b} &= \mathbb{B}^{-1} \cdot y \;=\; \mathbb{B}^{-1} \cdot (f \cdot x) \\ &= \mathbb{B}^{-1} \cdot (f \cdot (\mathbb{A} * \vec{a})) \;=\; \big(\mathbb{B}^{-1} \cdot f \cdot \mathbb{A}\big) * \vec{a} \\ &= F * \vec{a}, \end{aligned}$$

gdzie $F \in \mathbf{K}^{m,n}$,

$$F \;=\; \mathbb{B}^{-1} \cdot f \cdot \mathbb{A},$$

jest macierzą o wyrazach $f_{i,j} = r_i(f(x_j))$, tzn. w j-tej kolumnie macierzy F stoją współczynniki rozwinięcia wektora $f(x_j)$ w bazie $[y_1, \ldots, y_m]$.

Definicja 8.2 *Macierz liczbową* $F = \mathbb{B}^{-1} \cdot f \cdot \mathbb{A}$ *nazywamy* macierzą przekształcenia $f : \mathcal{X} \to \mathcal{Y}$ w bazach $\mathbb{A}$ *i* $\mathbb{B}$ *odpowiednio przestrzeni* $\mathcal{X}$ *i* $\mathcal{Y}$.

8.2.2 Izomorfizm $\mathcal{L}in(\mathcal{X}, \mathcal{Y})$ i $\mathbf{K}^{m,n}$

Niech $\Phi : \mathcal{L}in(\mathcal{X}, \mathcal{Y}) \to \mathbf{K}^{m,n}$,

$$\Phi(f) = \mathbb{B}^{-1} \cdot f \cdot \mathbb{A}, \qquad \forall f \in \mathcal{L}in(\mathcal{X}, \mathcal{Y}).$$

Odwzorowanie Φ przyporządkowujące przekształceniu liniowemu jego macierz jest liniowe (zachowuje kombinacje liniowe). Jeśli bowiem $\Phi(f) = F$ i $\Phi(g) = G$ to

$$\begin{aligned} \Phi(\alpha * f + \beta * g) &= \mathbb{B}^{-1} \cdot (\alpha * f + \beta * g) \cdot \mathbb{A} \\ &= \alpha * (\mathbb{B}^{-1} \cdot f \cdot \mathbb{A}) + \beta * (\mathbb{B}^{-1} \cdot g \cdot \mathbb{A}) \\ &= \alpha * \Phi(f) + \beta * \Phi(g). \end{aligned}$$

Ponadto, łatwo sprawdzić, że Φ jest wzajemnie jednoznaczne i odwzorowanie odwrotne $\Phi : \mathbf{K}^{m,n} \to \mathcal{L}in(\mathcal{X}, \mathcal{Y})$ wyraża się wzorem

$$\Phi^{-1}(F) \;=\; \mathbb{B} * F * \mathbb{A}^{-1}, \qquad \forall F \in \mathbf{K}^{m,n}.$$

Stąd Φ jest izomorfizmem a przestrzenie $\mathcal{L}in(\mathcal{X}, \mathcal{Y})$ i $\mathbf{K}^{m,n}$ są izomorficzne.

Ponieważ dla przestrzeni macierzy mamy $\dim(\mathbf{K}^{m,n}) = m \cdot n$, otrzymujemy w szczególności wniosek, że

$$\dim\left(\mathcal{L}in(\mathcal{X}, \mathcal{Y})\right) \;=\; \dim(\mathcal{X}) \cdot \dim(\mathcal{Y}).$$

Przykładową bazę $\mathcal{L}in(\mathcal{X},\mathcal{Y})$ tworzą przekształcenia $\varphi_{i,j}$, $1 \le i \le m$, $1 \le j \le n$, dane wzorem $\varphi_{i,j} = \Phi^{-1}(E_{i,j})$ (gdzie $E_{i,j}$ ma jedynkę na przecięciu i-tego wiersza i j-tej kolumny, a poza tym zera). Dokładniej, dla $x = \mathbb{A} * \vec{a}$, $\vec{a} = [\alpha_1, \ldots, \alpha_n]^T$, mamy

$$f_{i,j} \cdot x = (\mathbb{B} * E_{i,j} * \mathbb{A}^{-1}) * \mathbb{A} * \vec{a} = \mathbb{B} * (E_{i,j} * \vec{a}) = (\mathbb{B} * \vec{e}_i) * \alpha_j = \vec{y}_i * \alpha_j.$$

8.3 Dalsze własności macierzy przekształceń

8.3.1 Obraz i jądro przekształcenia/macierzy

Twierdzenie 8.3 *Mamy*

$$\begin{aligned} \mathrm{im}(f) &= \mathbb{B} * \mathcal{R}(F) := \{\mathbb{B} * \vec{b} : \vec{b} \in \mathcal{R}(F)\}, \\ \ker(f) &= \mathbb{A} * \mathcal{N}(F) := \{\mathbb{A} * \vec{a} : \vec{a} \in \mathcal{N}(F)\}. \end{aligned}$$

Dowód. Bezpośrednio sprawdzamy, że

$$\begin{aligned} \mathrm{im}(f) &= \{f \cdot x : x \in \mathcal{X}\} = \{f \cdot \mathbb{A} * \vec{a} : \vec{a} \in \mathbf{K}^n\} \\ &= \{\mathbb{B} * (\mathbb{B}^{-1} \cdot f \cdot \mathbb{A}) * \vec{a} : \vec{a} \in \mathbf{K}^n\} = \{\mathbb{B} * F * \vec{a} : \vec{a} \in \mathbf{K}^n\} \\ &= \{\mathbb{B} * \vec{b} : \vec{b} \in \mathcal{R}(F)\}, \end{aligned}$$

oraz

$$\begin{aligned} \ker(f) &= \{x \in \mathcal{X} : f \cdot x = \mathbf{0}\} = \{\mathbb{A} * \vec{a} \in \mathcal{X} : f \cdot \mathbb{A} * \vec{a} = \mathbf{0}\} \\ &= \{\mathbb{A} * \vec{a} : \mathbb{B} * (\mathbb{B}^{-1} \cdot f \cdot \mathbb{A}) * \vec{a} = \mathbf{0}\} \\ &= \{\mathbb{A} * \vec{a} : \mathbb{B} * F * \vec{a} = \mathbf{0}\} = \{\mathbb{A} * \vec{a} : F * \vec{a} = \mathbf{0}\} \\ &= \{\mathbb{A} * \vec{a} : \vec{a} \in \mathcal{N}(F)\}. \end{aligned}$$

Na podstawie twierdzenia 8.3 możemy powiedzieć, że $\mathbb{B}$ ($\mathbb{B}^{-1}$) jest izomorfizmem $\mathcal{R}(F)$ w $\mathrm{im}(f)$ ($\mathrm{im}(f)$ w $\mathcal{R}(F)$), a $\mathbb{A}$ ($\mathbb{A}^{-1}$) jest izomorfizmem $\mathcal{N}(F)$ w $\ker(f)$ ($\ker(f)$ w $\mathcal{N}(F)$).

8.3.2 Zmiana bazy

Zastanówmy się jak wygląda zależność pomiędzy współczynnikami rozwinięcia danego wektora $x \in \mathcal{X}$ w dwóch różnych bazach $\mathbb{A}$ i $\mathbb{B}$ przestrzeni $\mathcal{X}$.

Formalnie możemy rozpatrzyć macierz przekształcenia identycznościowego $f = \mathrm{id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$, $\mathrm{id}_{\mathcal{X}}(x) = x$. Zapisując x z jednej strony jako $x = \mathbb{A} * \vec{a}$, a z drugiej jako $x = \mathbb{B} * \vec{b}$ otrzymujemy

$$\vec{b} = \left(\mathbb{B}^{-1} \cdot \mathbb{A}\right) * \vec{a}.$$

Macierz $F = \mathbb{B}^{-1} \cdot \mathbb{A} \in \mathbf{K}^{n,n}$ o współczynnikach $f_{i,j} = r_i \cdot x_j$ nazywa się *macierzą zmiany bazy* z $\mathbb{A}$ na $\mathbb{B}$.

Oczywiście, macierz zmiany bazy jest nieosobliwa.

Podamy teraz charakterystyczny przykład zmiany bazy. Niech $\mathcal{X}_{|\mathbf{K}} = \mathcal{P}^n_{|\mathbf{R}}$ będzie przestrzenią wielomianów stopnia co najwyżej $n-1$. Rozpatrzmy bazę potęgową $\mathbb{A} = [1, t, t^2, \ldots, t^{n-1}]$ oraz bazę $\mathbb{B} = [l_1, \ldots, l_n]$, gdzie l_i są wielomianami Lagrange'a zdefiniowanymi w (6.1) dla ustalonych węzłów $t_1 < t_2 < \cdots < t_n$. Wtedy funkcjonały r_k, $1 \le k \le n$, tworzące macierz $\mathbb{B}^{-1}$, dane są wzorem $r_k(p) = p(t_k)$ $\forall p \in \mathcal{P}^n_{|\mathbf{R}}$. Stąd współczynniki macierzy przejścia $F = \mathbb{B}^{-1} \cdot \mathbb{A} \in \mathbf{K}^{n,n}$ wynoszą $f_{i,j} = (t_i)^j$, czyli

$$F = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix}.$$

Jest to *macierz Vandermonde'a*. Zauważmy, że "przy okazji" pokazaliśmy, iż macierz Vandermonde'a, jako macierz zmiany bazy, jest nieosobliwa.

8.3.3 Złożenie przekształceń

Niech $f \in \mathcal{L}in(\mathcal{X}, \mathcal{Y})$ i $g \in \mathcal{L}in(\mathcal{Y}, \mathcal{Z})$. Wtedy złożenie (superpozycja)

$$g \circ f : \mathcal{X} \to \mathcal{Z},$$

$(g \circ f)(x) = g(f(x))$ $\forall x$ jest też liniowe, tzn. $(g \circ f) \in \mathcal{L}in(\mathcal{X}, \mathcal{Y})$. Rzeczywiście,

$$\begin{aligned} (g \circ f)(x_1 * \alpha_1 + x_2 * \alpha_2) &= g(f(x_1) * \alpha_1 + f(x_2) * \alpha_2) \\ &= (g \circ f)(x_1) * \alpha_1 + (g \circ f)(x_2)\alpha_2. \end{aligned}$$

Twierdzenie 8.4 *Niech $\mathbb{A}$, $\mathbb{B}$ i $\mathbb{C}$ będą odpowiednio bazami przestrzeni $\mathcal{X}$, $\mathcal{Y}$ i $\mathcal{Z}$. Niech $f \in \mathcal{L}in(\mathcal{X},\mathcal{Y})$, $g \in \mathcal{L}in(\mathcal{Y},\mathcal{Z})$, a F, G będą odpowiednio macierzami przekształceń f i g w podanych bazach. Wtedy macierz złożenia $h = g \circ f \in \mathcal{L}in(\mathcal{X},\mathcal{Z})$ wynosi*

$$H = G * F.$$

Dowód. Rzeczywiście, mamy bowiem

$$\begin{aligned} H &= \mathbb{C}^{-1} \cdot h \cdot \mathbb{A} = \mathbb{C}^{-1} \cdot g \cdot f \cdot \mathbb{A} \\ &= \left(\mathbb{C}^{-1} \cdot g \cdot \mathbb{B}\right) * \left(\mathbb{B}^{-1} \cdot f \cdot \mathbb{A}\right) \\ &= G * F. \end{aligned}$$

Rozdział 9

Wyznacznik macierzy

9.1 Definicja i pierwsze własności

Niech A będzie macierzą kwadratową nad ciałem $\mathbf{K}$,

$$A = (a_{i,j})_{i,j=1}^{n} \in \mathbf{K}^{n,n}.$$

Definicja 9.1 (PRZEZ ROZWINIĘCIE LAPLACE'A)
Wynacznikiem *macierzy kwadratowej* $n \times n$ *nazywamy funkcję*

$$\det_n : \mathbf{K}^{n,n} \to \mathbf{K},$$

zdefiniowaną rekurencyjnie w następujący sposób:

$$(n = 1) \quad \det_1(A) := \det_1([a_{1,1}]) = a_{1,1},$$

$$(n \geq 2) \quad \det_n(A) := \textstyle\sum_{i=1}^{n} (-1)^{i+n} a_{i,n} \cdot \det_{n-1}(A_{i,n}),$$

gdzie $A_{i,n} \in \mathbf{K}^{n-1.n-1}$ *jest macierzą powstałą z* A *poprzez usunięcie z niej* i*-tego wiersza i* n*-tej kolumny.*

Zgodnie z definicją mamy

$$\begin{aligned}
\det_2(A) &= a_{1,1}a_{2,2} - a_{1,2}a_{2,1}, \\
\det_3(A) &= a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} \\
&\quad - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,1}a_{3,2}, \\
\det_4(A) &= \ldots .
\end{aligned}$$

Wprost z definicji rekurencyjnej łatwo również zauważyć, że dla macierzy identycznościowej mamy $\det_n(I_n) = 1$. Ogólniej, jeśli A jest macierzą rójkątną dolną lub trójkątną górną, $A \in \mathrm{TRIL}^{n,n} \cup \mathrm{TRIU}^{n,n}$, to

$$\det_n(A) = \prod_{i=1}^{n} a_{i,i}.$$

Jeśli format macierzy jest znany lub nieistotny to dalej będziemy dla uproszczenia pisać $\det(A)$ zamiast $\det_n(A)$.

Twierdzenie 9.1 *Wyznacznik jest funkcją liniową ze względu na dowolną kolumnę macierzy, tzn.*

$$\begin{aligned} &\det([\vec{a}_1, \ldots, \vec{a}_p * \alpha + \vec{a}'_p * \alpha', \ldots, \vec{a}_n]) \\ &\quad = \det([\vec{a}_1, \ldots, \vec{a}_p, \ldots, \vec{a}_n]) * \alpha + \det([\vec{a}_1, \ldots, \vec{a}'_p, \ldots, \vec{a}_n]) * \alpha', \end{aligned}$$

$1 \leq p \leq n$.

Dowód. Rzeczywiście, równość w oczywisty sposób zachodzi dla $n = 1$, a dla $n \geq 2$ wystarczy osobno rozpatrzyć dwa przypadki, $p = n$ i $1 \leq p \leq n-1$, oraz skorzystać z definicji rekurencyjnej.

Z twierdzenia 9.1 mamy od razu, że $\det([\ldots, \vec{0}, \ldots]) = 0$. Natomiast stosując twierdzenie 9.1 kolejno do każdej z kolumn macierzy otrzymujemy, że dla dowolnej macierzy diagonalnej $D = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$

$$\det(A * D) = \det([\vec{a}_1 * \alpha_1, \ldots, \vec{a}_n * \alpha_n]) = \det(A) \cdot \prod_{i=1}^{n} \alpha_i. \tag{9.1}$$

W szczególności,

$$\det_n(\alpha * A) = \alpha^n \cdot \det_n(A) \qquad \text{oraz} \qquad \det_n(-A) = (-1)^n \cdot \det_n(A).$$

9.2 Wyznacznik a operacje elementarne

9.2.1 Permutacja kolumn

Twierdzenie 9.2 *Przestawienie różnych kolumn macierzy zmienia znak wyznacznika, tzn. dla dowolnej transpozycji $T_{p,q}$, $p \neq q$,*

$$\det(A * T_{p,q}) = -\det(A).$$

Dowód. (Indukcja względem n.)
Dla $n = 1, 2$ wzór sprawdzamy bezpośrednio z definicji. Dla $n \geq 3$ rozpatrujemy trzy przypadki.
(a) $1 \leq p < q \leq n - 1$.
Korzystając z założenia indukcyjnego mamy

$$\begin{aligned}
\det_n(A * T_{p,q}) &= \sum_{i=1}^{n} (-1)^{i+n} a_{i,n} \det_{n-1}((A * T_{p,q})_{i,n}) \\
&= -\sum_{i=1}^{n} (-1)^{i+n} a_{i,n} \det_{n-1}(A_{i,n}) \\
&= -\det_n(A).
\end{aligned}$$

(b) $p = n - 1$, $q = n$.
Stosując dwukrotnie rozwinięcie Laplace'a dostajemy

$$\begin{aligned}
\det_n(A) &= \sum_{i=1}^{n} (-1)^{i+n} a_{i,n} \det_{n-1}(A_{i,n}) \\
&= \sum_{i=1}^{n} (-1)^{i+n} \Big(\sum_{k=1}^{i-1} (-1)^{k+(n-1)} a_{k,n-1} \det_{n-2}(A_{\{i,k\}\{n-1,n\}}) \\
&\quad + \sum_{k=i+1}^{n} (-1)^{(k-1)+(n-1)} a_{k,n-1} \det_{n-2}(A_{\{i,k\}\{n-1,n\}}) \Big) \\
&= -\sum_{k<i} (-1)^{i+k} a_{i,n} a_{k,n-1} \det_{n-2}(A_{\{i,k\}\{n-1,n\}}) \\
&\quad + \sum_{i<k} (-1)^{i+k} a_{i,n} a_{k,n-1} \det_{n-2}(A_{\{i,k\}\{n-1,n\}}),
\end{aligned}$$

gdzie $A_{\{i,k\}\{n-1,n\}}$ jest macierzą powstałą z A poprzez usunięcie wierszy i-tego i k-tego oraz kolumn $(n-1)$-szej i n-tej. Wykonując to samo dla macierzy $A * T_{p,q}$ otrzymujemy ten sam wzór, ale z odwróconymi znakami przed symbolami sumowania.
(c) $1 \leq p \leq n - 2$, $q = n$.
W tym przypadku wystarczy zauważyć, że

$$A * T_{p,n} = A * T_{p,n-1} * T_{n-1,n} * T_{p,n-1}$$

i skorzystać dwukrotnie z (a) i raz (b).

Z twierdzenia 9.2 wynika w szczególności, że wyznacznik macierzy transpozycji $T_{p,q}$ z $p \neq q$ wynosi -1.

Wyznacznik można rozwijać nie tylko względem ostatniej, ale również względem dowolnej kolumny.

Twierdzenie 9.3 *Dla dowolnego $n \geq 2$ i $1 \leq j \leq n$ mamy*

$$\det_n(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det_{n-1}(A_{i,j}).$$

Dowód. Jeśli $j = n-1$ to

$$\begin{aligned}
\det_n(A) &= -\det_n(A * T_{n-1,n}) \\
&= -\sum_{i=1}^{n} (-1)^{i+n} a_{i,n-1} \cdot \det_{n-1}(A_{i,n-1}) \\
&= \sum_{i=1}^{n} (-1)^{i+n-1} a_{i,n-1} \cdot \det_{n-1}(A_{i,n-1}).
\end{aligned}$$

Dalej, korzystając z prawdziwości rozwinięcia dla $j = n-1$, pokazujemy podobnie prawdziwość rozwinięcia dla $j = n-2$, itd., aż do $j = 1$.

9.2.2 Kombinacja liniowa kolumn

Z twierdzenia 9.2 od razu otrzymujemy

$$\det([\ldots, \vec{a}, \ldots, \vec{a}, \ldots]) = 0.$$

Stąd i z liniowości wyznacznika względem dowolnej kolumny wynika, że wyznacznik nie ulegnie zmianie gdy do kolumny dodamy inną kolumnę pomnożoną przez skalar, tzn.

$$\begin{aligned}
&\det([\vec{a}_1, \ldots, \vec{a}_{p-1}, \vec{a}_p + \vec{a}_q * m, \vec{a}_{p+1}, \ldots, \vec{a}_n]) \\
&\quad = \det([\vec{a}_1, \ldots, \vec{a}_{p-1}, \vec{a}_p, \vec{a}_{p+1}, \ldots, \vec{a}_n]).
\end{aligned}$$

Uogólnieniem ostatniej własności jest następująca.

Twierdzenie 9.4 *Jeśli do p-tej kolumny dodamy kombinację liniową pozostałych kolumn to wyznacznik macierzy nie ulegnie zmianie, tzn.*

$$\begin{aligned}
&\det\Big(\Big[\vec{a}_1, \ldots, \vec{a}_{p-1}, \vec{a}_p + \sum_{j \neq p} \vec{a}_j * m_j, \vec{a}_{p+1}, \ldots, \vec{a}_n\Big]\Big) \\
&\quad = \det([\vec{a}_1, \ldots, \vec{a}_{p-1}, \vec{a}_p, \vec{a}_{p+1}, \ldots, \vec{a}_n]).
\end{aligned}$$

Zauważmy, że ostatnią równość można symbolicznie zapisać jako

$$\det(A * (I + \vec{m} * \vec{a}_p^T)) = \det(A), \quad \text{o ile } \vec{e}_p^T * \vec{m} = 0.$$

Wniosek 9.1 *Jeśli macierz A jest osobliwa to* $\det(A) = 0$.

Dowód. Jeśli A nie jest pełnego rzędu to jedna z kolumn, powiedzmy p, jest kombinacją liniową pozostałych kolumn. Odejmując od p-tej kolumny tą kombinację liniową otrzymujemy macierz A' o tym samym wyznaczniku co A i o zerowej p-tej kolumnie. Stąd $\det(A) = \det(A') = 0$.

9.3 Dalsze własności wyznaczników

9.3.1 Wyznacznik iloczynu macierzy

Jak wiemy, każdą macierz trójkątną dolną $L \in \mathrm{TRIL}^{n,n}$ z jedynkami na głównej przekątnej można przedstawić jako iloczyn

$$L = I_n + \vec{l}_1 * \vec{e}_1^T + \cdots + \vec{l}_{n-1} * \vec{e}_{n-1}^T = (I_n + \vec{l}_1 * \vec{e}_1^T) * \cdots * (I_n + \vec{l}_{n-1}\vec{e}_{n-1}^T),$$

gdzie $\vec{l}_j = [\underbrace{0, \ldots, 0}_{j}, l_{j+1,j}, \ldots, l_{n,j}]^T$, $1 \le j \le n-1$. Na podstawie twierdzenia 9.4 mamy więc, że

$$\det(A * L) = \det(A). \tag{9.2}$$

Podobnie, wyznacznik nie ulegnie zmianie gdy macierz pomnożymy z prawej strony przez macierz trójkątną górną z jedynkami na głównej przekątnej.

Niech teraz $W \in \mathrm{TRIL}^{n,n} \cup \mathrm{TRIU}^{n,n}$. Jeśli wszystkie wyrazy na przekątnej są niezerowe, $w_{i,i} \neq 0$, $1 \le i \le n$, to

$$W = W_1 * \mathrm{diag}(w_{1,1}, \ldots, w_{n,n}),$$

gdzie $W_1 \in \mathrm{TRIL}^{n,n} \cup \mathrm{TRIU}^{n,n}$ z jedynkami na głównej przekątnej. Stosując kolejno (9.1) i (9.2) (z macierzą odpowiednio trójkątną górną albo trójkątną dolną) dostajemy

$$\det(A * W) = \det(A * W_1) \cdot \prod_{i=1}^{n} w_{i,i} = \det(A) \cdot \prod_{i=1}^{n} w_{i,i}. \tag{9.3}$$

Jeśli zaś $w_{k,k} = 0$ dla pewnego k to W jest osobliwa, a stąd osobliwa jest również macierz $A * W$ i równanie $\det(A * W) = \det(A) \cdot \prod_{i=1}^{n} w_{i,i}$ pozostaje w mocy.

Możemy teraz pokazać następujące twierdzenie

Twierdzenie 9.5 *Dla dowolnych macierzy* $A, B \in \mathbf{K}^{n,n}$

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

Dowód. Skorzystamy z twierdzenia, że dla dowolnej macierzy B istnieje rozkład trójkątno-trójkątny $P * B * Q^T = L * R$, czyli

$$B = P^T * L * R * Q,$$

gdzie $P = T_{1,p(1)} * \cdots * T_{n-1,p(n-1)}$ i $Q = T_{1,q(1)} * \ldots * T_{n-1,q(n-1)}$ są macierzami permutacji, L jest trójkątna dolna z jedynkami na przekątnej, a R trójkątna górna. Jasne, że $\det(P) = (-1)^s$, gdzie s jest liczbą właściwych przestawień w p (tzn. liczbą tych i dla których $i \neq p(i)$), oraz podobnie $\det Q = (-1)^t$, gdzie t jest liczbą właściwych przestawień w q. Wykorzystując wielokrotnie twierdzenie 9.2 oraz wzór (9.3) otrzymujemy

$$\begin{aligned}
\det(A * B) &= \det(A * P^T * L * R) \cdot (-1)^t \\
&= \det(A * P^T * L)(-1)^t \cdot \prod_{i=1}^n r_{i,i} \\
&= \det(A * P^T)(-1)^t \cdot \prod_{i=1}^n r_{i,i} \\
&= \det(A)(-1)^{s+t} \cdot \prod_{i=1}^n r_{i,i} \\
&= \det(A) * \det(B),
\end{aligned}$$

co należało pokazać.

9.3.2 Wyznacznik macierzy nieosobliwej i transponowanej

Jak zauważyliśmy wcześniej w dowodzie twierdzenia 9.5, rozkład macierzy $A = P^T * L * R * Q$ implikuje równość

$$\det(A) = (-1)^{s+t} \cdot \prod_{i=1}^n r_{i,i},$$

która z kolei daje dwa następujące ważne wnioski.

Wniosek 9.2 *Macierz A jest nieosobliwa, tzn.* $\mathrm{rz}(A) = n$, *wtedy i tylko wtedy gdy* $\det(A) \neq 0$.

Wniosek 9.3 *Dla dowolnej macierzy kwadratowej A mamy*

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

Ostatni wniosek oznacza, że wszystkie własności wyznacznika dotyczące kolumn macierzy przysługują również jej wierszom. W szczególności, wyznacznik można rozwijać względem dowolnego wiersza,

$$\det_n(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det_{n-1}(A_{i,j}).$$

9.4 Definicja kombinatoryczna wyznacznika

Każda macierz permutacji $P \in \mathbf{K}^{n,n}$ może być rozłożona na wiele sposobów na iloczyn transpozycji, np.

$$P = T_{1,i_1} * T_{2,i_2} * \cdots * T_{n-1,i_{n-1}}. \tag{9.4}$$

Ponieważ

$$\det(T_{p,q}) = \begin{cases} 1, & p = q \quad \text{(transpozycja niewłaściwa)}, \\ -1, & p \neq q \quad \text{(transpozycja właściwa)}, \end{cases}$$

to

$$\det(P) = (-1)^{\sigma(p)},$$

gdzie $\sigma(p) = 0$ gdy liczba transpozycji właściwych w rozkładzie (9.4) jest parzysta, oraz $\sigma(p) = 1$ gdy liczba transpozycji właściwych w (9.4) jest nieparzysta. Pokazaliśmy więc następujące twierdzenie.

Twierdzenie 9.6 *W rozkładzie macierzy permutacji na iloczyn transpozycji liczba transpozycji właściwych jest zawsze parzysta, albo zawsze nieparzysta.*

Parzystość lub nieparzystość permutacji jest więc własnością permutacji (niezależną od rozkładu).

Definicja Laplace'a wyznacznika jest równoważna następującej definicji *kombinatorycznej*:

$$\det_n(A) = \sum_{p=[p(1),\ldots,p(n)]} (-1)^{\sigma(p)} \prod_{j=1}^{n} a_{p(j),j},$$

albo

$$\det_n(A) = \sum_{q=[q(1),\ldots,q(n)]} (-1)^{\sigma(q)} \prod_{i=1}^{n} a_{i,q(i)}.$$

Indukcyjny dowód równoważności tych definicji pomijamy. (Tutaj p i q są permutacjami ciągu $[1, 2, \ldots, n]$, przy czym $p \circ q = q \circ p = \mathrm{Id} = [1, 2, \ldots, n]$. Wtedy $\sigma(p) = \sigma(q)$.)

9.5 Wzory Cramera

Pokażemy teraz, że układy równań liniowych można, przynajmniej teoretycznie, rozwiązywać za pomocą liczenia odpowiednich wyznaczników.

Definicja 9.2 *Macierz $C(A) := (\gamma_{i,j}) \in \mathbf{K}^{n,n}$, gdzie*

$$\gamma_{i,j} = (-1)^{i+j} \det_{n-1}(A_{i,j}),$$

nazywamy macierzą komplementarną *do danej macierzy $A \in \mathbf{K}^{n,n}$.*

Zauważmy, że na podstawie rozwinięcia Laplace'a mamy

$$p_{j,k} := \sum_{i=1}^{n} \gamma_{i,j} a_{i,k} = \begin{cases} \det_n(A), & k = j, \\ 0, & k \neq j, \end{cases}$$

a stąd

$$P = (p_{j,k})_{j,k=1}^{n} = \det_n(A) * I_n = (C(A))^T * A.$$

Zatem jeśli $\mathrm{rz}(A) = n$ to

$$A^{-1} = \frac{(C(A))^T}{\det_n(A)} = \left(\frac{(-1)^{i+j} \det_{n-1}(A_{j,i})}{\det_n(A)} \right)_{i,j=1}^{n}.$$

Rozpatrzmy teraz układ równań $A * \vec{x} = \vec{b}$ z kwadratową i nieosobliwą macierzą $A \in \mathbf{K}^{n,n}$. Wtedy jego rozwiązanie

$$\vec{x} = (x_j)_{j=1}^n \; = \; A^{-1} * \vec{b} \; = \; \frac{(C(A))^T * \vec{b}}{\det_n(A)},$$

czyli

$$x_j \; = \; \frac{\sum_{i=1}^n \gamma_{i,j} * b_i}{\det_n(A)} \; = \; \frac{\sum_{i=1}^n (-1)^{i+j} \det_{n-1}(A_{i,j}) \cdot b_i}{\det_n(A)},$$

albo równoważnie

$$x_j \; = \; \frac{\det_n([\vec{a}_1, \ldots, \vec{a}_{j-1}, \vec{b}, \vec{a}_{j+1}, \ldots, \vec{a}_n])}{\det_n([\vec{a}_1, \ldots, \vec{a}_{j-1}, \vec{a}_j, \vec{a}_{j+1}, \ldots, \vec{a}_n])},$$

dla $1 \leq j \leq n$. Ostatnie formuły zwane są WZORAMI CRAMERA.

Uwaga. Wzory Cramera mają dla dużych n znaczenie jedynie teoretyczne, gdyż, jak łatwo się przekonać, koszt liczenia wyznacznika macierzy wprost z definicji jest proporcjonalny do $n!$ W takich przypadkach lepiej stosować eliminację Gaussa, której koszt obliczeniowy jest proporcjonalny do n^3.

Rozdział 10

Formy dwuliniowe i kwadratowe

10.1 Formy dwuliniowe

10.1.1 Definicja i przykłady

Niech $\mathcal{X}_{|\mathbf{K}}$ będzie przestrzenią liniową nad ciałem $\mathbf{K}$, $\dim(\mathcal{X}_{|\mathbf{K}}) = n$.

Definicja 10.1 *Przekształcenie* $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ *nazywamy* formą dwuliniową *na przestrzeni* $\mathcal{X}_{|\mathbf{K}}$ *jeśli*

(i) $\forall x, y_1, y_2 \in \mathcal{X}$, $\forall \alpha_1, \alpha_2 \in \mathbf{K}$

$$\varphi(x, y_1 * \alpha_1 + y_2 * \alpha_2) = \varphi(x, y_1) * \alpha_1 + \varphi(x, y_2) * \alpha_2$$

(liniowość ze względu na drugą zmienną),

(ii) $\forall x, y \in \mathcal{X} \quad \varphi(x, y) = \varphi(y, x)$ *(forma zwykła)*
albo
$\forall x, y \in \mathcal{X} \quad \varphi(x, y) = \overline{\varphi(y, x)}$ *(forma* hermitowska*).*

Oczywiście, o formach hermitowskich możemy mówić tylko wtedy gdy $\mathbf{K} \subseteq \mathbf{C}$. Dalej, dla uproszczenia, będziemy rozpatrywać jedynie formy hermitowskie.

Zauważmy, że $\forall x_1, x_2, y \in \mathcal{X}$, $\forall \beta_1, \beta_2 \in \mathbf{K}$,

$$\begin{aligned}\varphi(x_1 * \beta_1 + x_2 * \beta_2, y) &= \overline{\varphi(y, x_1 * \beta_1 + x_2 * \beta_2)} \\ &= \overline{\varphi(y, x_1)} * \overline{\beta}_1 + \overline{\varphi(y, x_2)} * \overline{\beta}_2 \\ &= \varphi(x_1, y) * \overline{\beta}_1 + \varphi(x_2, y) * \overline{\beta}_2.\end{aligned}$$

Dość oczywistym jest fakt, że zbiór wszystkich form dwuliniowych na $\mathcal{X}_{|\mathbf{K}}$ jest przestrzenią liniową nad $\mathbf{R}$ (ale nie nad $\mathbf{C}$!) z naturalnymi działaniami:

$$\begin{aligned}(\alpha * \varphi)(x, y) &:= \alpha * \varphi(x, y), \\ (\varphi_1 + \varphi_2)(x, y) &:= \varphi_1(x, y) + \varphi_2(x, y).\end{aligned}$$

Przykładami form dwuliniowych na $\mathcal{X}_{|\mathbf{K}} = \mathbf{K}^n_{|\mathbf{K}}$ ($\mathbf{K} \subseteq \mathbf{C}$) są:

$$\begin{aligned}\varphi(\vec{x}, \vec{y}) &= \sum_{i=1}^{n} \overline{x}_i * y_i * \rho_i, \quad \text{gdzie } \rho_i \in \mathbf{R},\ 1 \le i \le n, \\ \varphi(\vec{x}, \vec{y}) &= \vec{x}^H * A * \vec{y}, \quad \text{gdzie } A \in \mathbf{K}^{n,n},\ A = A^H,\end{aligned}$$

a na $\mathcal{P}^n_{|\mathbf{R}}$:

$$\begin{aligned}\varphi(p, q) &= \sum_{i=1}^{n} \overline{p^{(i)}(t_i)} \cdot q^{(i)}(t_i) \cdot \rho_i, \qquad \rho_i \in \mathbf{R},\ 1 \le i \le n, \\ \varphi(p, q) &= \int_0^1 \overline{p(t)} \cdot q(t) \cdot \rho(t)\, dt, \qquad \rho : \mathbf{R} \to \mathbf{R}.\end{aligned}$$

10.1.2 Macierz formy dwuliniowej

Dalej wygodnie nam będzie rozszerzyć działanie danej formy dwuliniowej $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ na $\varphi : \mathcal{X}^{1,s} \times \mathcal{X}^{1,t} \to \mathbf{K}^{s,t}$ w następujący sposób. Niech $\mathbb{A} = [x_1, \ldots, x_s]$ i $\mathbb{B} = [y_1, \ldots, y_t]$. Wtedy

$$\varphi(\mathbb{A}, \mathbb{B}) := (\varphi(x_i, y_j))_{i,j} \in \mathbf{K}^{s,t}.$$

W szczególności, macierz $\varphi(\mathbb{A}, \mathbb{A}) = (\varphi(x_i, x_j))_{i,j}$ jest kwadratowa i hermitowska, $\varphi(\mathbb{A}, \mathbb{A}) \in \mathrm{Herm}^{n,n}$. Mamy też

$$\begin{aligned}\forall \varphi\ \forall \alpha \in \mathbf{R} \qquad (\alpha * \varphi)(\mathbb{A}, \mathbb{B}) &= \alpha * \varphi(\mathbb{A}, \mathbb{B}), \\ \forall \varphi, \psi \qquad (\varphi + \psi)(\mathbb{A}, \mathbb{B}) &= \varphi(\mathbb{A}, \mathbb{B}) + \psi(\mathbb{A}, \mathbb{B}).\end{aligned}$$

Pożyteczne będą też następujące wzory rachunkowe:

$$\begin{array}{llcl} \forall \vec{b} \in \mathbf{K}^t & \varphi(\mathbb{A}, \mathbb{B} * \vec{b}) & = & \varphi(\mathbb{A}, \mathbb{B}) * \vec{b}, \\ \forall \vec{a} \in \mathbf{K}^s & \varphi(\mathbb{A} * \vec{a}, \mathbb{B}) & = & \vec{a}^H * \varphi(\mathbb{A}, \mathbb{B}). \end{array}$$

Rzeczywiście,

$$\varphi(\mathbb{A}, \mathbb{B} * \vec{b}) \;=\; \varphi\Big(\mathbb{A}, \sum_{j=1}^t y_j * \beta_j\Big) \;=\; \sum_{j=1}^t \varphi(\mathbb{A}, y_j) * \beta_j \;=\; \varphi(\mathbb{A}, \mathbb{B}) * \vec{b},$$

gdzie $\vec{b} = [\beta_1, \ldots, \beta_t]^T$, oraz

$$\varphi(\mathbb{A} * \vec{a}, \mathbb{B}) \;=\; (\varphi(\mathbb{B}, \mathbb{A} * \vec{a}))^H \;=\; \vec{a}^H * (\varphi(\mathbb{B}, \mathbb{A}))^H \;=\; \vec{a}^H * \varphi(\mathbb{A}, \mathbb{B}).$$

Uogólniając te wzory mamy

$$\begin{array}{llcl} \forall B \in \mathbf{K}^{t,r} & \varphi(\mathbb{A}, \mathbb{B} * B) & = & \varphi(\mathbb{A}, \mathbb{B}) * B, \\ \forall A \in \mathbf{K}^{s,r} & \varphi(\mathbb{A} * A, \mathbb{B}) & = & A^H * \varphi(\mathbb{A}, \mathbb{B}). \end{array}$$

Mamy bowiem

$$\begin{array}{rcl} \varphi(\mathbb{A}, \mathbb{B} * B) & = & \varphi(\mathbb{A}, [\mathbb{B} * \vec{b}_1, \ldots, \mathbb{B} * \vec{b}_r]) \\ & = & [\varphi(\mathbb{A}, \mathbb{B} * \vec{b}_1), \ldots, \varphi(\mathbb{A}, \mathbb{B} * \vec{b}_r)] \\ & = & [\varphi(\mathbb{A}, \mathbb{B}) * \vec{b}_1, \ldots, \varphi(\mathbb{A}, \mathbb{B}) * \vec{b}_r] \\ & = & \varphi(\mathbb{A}, \mathbb{B}) * B, \end{array}$$

gdzie $B = [\vec{b}_1, \ldots, \vec{b}_r]$, oraz

$$\begin{array}{rcl} \varphi(\mathbb{A} * A, \mathbb{B}) & = & (\varphi(\mathbb{B}, \mathbb{A} * A))^H \;=\; (\varphi(\mathbb{B}, \mathbb{A}) * A)^H \\ & = & A^H * (\varphi(\mathbb{B}, \mathbb{A}))^H \;=\; A^H * \varphi(\mathbb{A}, \mathbb{B}). \end{array}$$

Definicja 10.2 *Niech $\mathbb{A} = [x_1, \ldots, x_n]$ będzie bazą $\mathcal{X}$, a $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ formą dwuliniową na $\mathcal{X}$. Macierz hermitowską*

$$\Phi_{\mathbb{A}} \;:=\; \varphi(\mathbb{A}, \mathbb{A}) \;=\; (\varphi(x_i, x_j))_{i,j=1}^n$$

nazywamy macierzą formy *φ w bazie $\mathbb{A}$.*

Znaczenie macierzy formy wynika z następującej równości. Niech $x = \mathbb{A} * \vec{a}$ i $y = \mathbb{A} * \vec{b}$. Wtedy

$$\begin{aligned}\varphi(x,y) &= \varphi(\mathbb{A} * \vec{a}, \mathbb{A} * \vec{b}) = \vec{a}^H * \varphi(\mathbb{A},\mathbb{A}) * \vec{b} \\ &= \vec{a}^H * \Phi_{\mathbb{A}} * \vec{b} = (\mathbb{A}^{-1} \cdot x)^H * \Phi_{\mathbb{A}} * (\mathbb{A}^{-1} \cdot y).\end{aligned}$$

Przy ustalonej bazie $\mathbb{A}$, każdej formie hermitowskiej $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ można przyporządkować jej macierz $\Phi_{\mathbb{A}} = \varphi(\mathbb{A},\mathbb{A})$, która jest hermitowska. Ale też odwrotnie, każda macierz hermitowska Φ definiuje formę hermitowską zgodnie ze wzorem $\varphi(x,y) = (\mathbb{A}^{-1} \cdot x)^H * \Phi * (\mathbb{A}^{-1} \cdot y)$. Mamy przy tym, że jeśli $\gamma = \varphi + \psi$ to $\Gamma_{\mathbb{A}} = \Phi_{\mathbb{A}} + \Psi_{\mathbb{A}}$ oraz jeśli $\gamma = \alpha * \varphi$, $\alpha \in \mathbf{R}$, to $\Gamma_{\mathbb{A}} = \alpha * \Phi_{\mathbb{A}}$. Stąd przestrzeń wszystkich form hermitowskich nad $\mathbf{R}$ jest izomorficzna z przestrzenią macierzy hermitowskich nad $\mathbf{R}$. W przypadku $\mathbf{K} = \mathbf{C}$ jej wymiar wynosi n^2.

10.2 Twierdzenie Sylwester'a

Definicja 10.3 *Powiemy, że macierz* $A \in \mathbf{K}^{n,n}$ przystaje *do macierzy* $B \in \mathbf{K}^{n,n}$ *gdy istnieje macierz nieosobliwa* $C \in \mathbf{K}^{n,n}$ *taka, że*

$$B = C^H * A * C.$$

Niech $\mathbb{A}$ i $\mathbb{B}$ będą dwiema bazami $\mathcal{X}_{|\mathbf{K}}$. Niech $C = \mathbb{A}^{-1} \cdot \mathbb{B} \in \mathbf{K}^{n,n}$ będzie macierzą zmiany bazy z $\mathbb{A}$ na $\mathbb{B}$ tak, że

$$\mathbb{B} = \mathbb{A} * C.$$

Jeśli $\Phi_{\mathbb{A}}$ jest macierzą danej formy $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ w bazie $\mathbb{A}$ to macierz φ w bazie $\mathbb{B}$ można wyrazić wzorem

$$\begin{aligned}\Phi_{\mathbb{B}} &= \varphi(\mathbb{B},\mathbb{B}) = \varphi(\mathbb{A} * C, \mathbb{A} * C) \\ &= C^H * \varphi(\mathbb{A},\mathbb{A}) * C = C^H * \Phi_{\mathbb{A}} * C.\end{aligned}$$

Stąd, w klasie macierzy hermitowskich $\mathrm{Herm}^{n,n}$ macierz A przystaje do B gdy obie są macierzami tej samej formy (ale być może w różnych bazach).

Relacja przystawania macierzy jest zwrotna (bo $A = I^H * A * I$), symetryczna (bo jeśli $B = C^H * A * C$ to $A = (C^{-1})^H * B * C^{-1}$) oraz przechodnia (bo jeśli $A_2 = C_1^H * A_1 * C_1$ i $A_3 = C_2^H * A_2 * C_2$ to $A_3 = (C_1 * C_2)^H * A_1 * (C_1 * C_2)$).

Jest to więc relacja równoważności. A jeśli tak, to zbiór wszystkich macierzy hermitowskich można przedstawić jako rozłączną sumę macierzy do siebie wzajemnie przystających (klas abstrakcji relacji przystawania, albo jeszcze inaczej, macierzy tej samej formy, ale w różnych bazach).

Ile jest klas abstrakcji relacji przystawania w klasie macierzy hermitowskich? Odpowiedź daje natępujące twierdzenie, które podajemy bez dowodu.

Twierdzenie 10.1 (SYLWESTER'A)
Dla dowolnej macierzy hermitowskiej $A = A^H \in \mathbf{K}^{n,n}$ istnieje macierz nieosobliwa $C \in \mathbf{K}^{n,n}$ taka, że

$$C^H * A * C = \mathrm{diag}(I_\pi, -I_\nu, 0_\xi),$$

gdzie wymiary π, ν, ξ $(\pi + \nu + \xi = n)$ są wyznaczone jednoznacznie.

Stąd klas abstrakcji relacji przystawania jest tyle ile macierzy diagonalnych z elementami na diagonali kolejno $1, -1, 0$, czyli

$$\sum_{k=0}^{n} (k+1) = \frac{(n+1)(n+2)}{2}.$$

Z twierdzenia Sylwester'a wynika również następujący ważny wniosek.

Wniosek 10.1 *Dla dowolnej formy dwuliniowej $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ istnieje baza $\mathbb{A}$ w $\mathcal{X}$, w której forma ma postać*

$$\varphi(x, y) = \sum_{k=1}^{\pi} \overline{a}_k * b_k - \sum_{k=\pi+1}^{\pi+\nu} \overline{a}_k * b_k,$$

gdzie $x = \mathbb{A} * \vec{a}$, $y = \mathbb{A} * \vec{b}$.

10.3 Formy kwadratowe

10.3.1 Określoność formy kwadratowej

Każdej formie dwuliniowej $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbf{K}$ odpowiada *forma kwadratowa* $h : \mathcal{X} \to \mathbf{R}$ zdefiniowana wzorem

$$h(x) = \varphi(x, x) \qquad x \in \mathcal{X}.$$

Jeśli dla wszystkich $x \neq \mathbf{0}$ mamy $h(x) = \varphi(x,x) > 0$ to formę kwadratową h (i odpowiednio formę dwuliniową φ) nazywamy *dodatnio określoną* i piszemy $h > 0$ (odpowiednio $\varphi > 0$). Podobnie, forma h jest określona

- ujemnie, gdy $h(x) < 0 \ \forall x \neq 0 \quad (h < 0)$,
- niedodatnio, gdy $h(x) \leq 0 \ \forall x \quad (h \leq 0)$,
- nieujemnie, gdy $h(x) \geq 0 \ \forall x \quad (h \geq 0)$.

We wszystkich pozostałych przypadkach forma jest *nieokreślona.*

Z równości

$$h(x) = \vec{a}^H * \Phi_{\mathbb{A}} * \vec{a} \qquad (x = \mathbb{A} * \vec{a})$$

wynika, że określoność formy jest taka sama jak określoność jej macierzy (w dowolnej bazie!). W szczególności, stosując notację z twierdzenia Sylwester'a mamy:

$$\begin{array}{ll} h > 0 \iff \pi = n, & h \geq 0 \iff \nu = 0, \\ h < 0 \iff \nu = n, & h \leq 0 \iff \pi = 0. \end{array}$$

10.3.2 Kryterium Sylwester'a

Twierdzenie 10.2 *Niech* $A = A^H = (a_{i,j})_{i,j=1}^n \in \mathrm{Herm}^{n,n}$ *oraz* $A^{(k)} = (a_{i,j})_{i,j=1}^k$, $1 \leq k \leq n$, *będą odpowiednimi macierzami kątowymi. Wtedy*

(i) A jest dodatnio określona $\iff \det(A^{(k)}) > 0$ *dla* $1 \leq k \leq n$,

(ii) A jest ujemnie określona $\iff (-1)^k \cdot \det(A^{(k)}) > 0$ *dla* $1 \leq k \leq n$.

Dowód. Przypomnijmy (twierdzenie 7.5), że dla macierzy o nieosobliwych macierzach kątowych (a takimi są macierze dodatnio/ujemnie określone) można przeprowadzić eliminację Gaussa bez przestawień wierszy/kolumn. Dlatego A można przedstawić jako

$$A = L * R = L * D * L^H,$$

gdzie $L \in \mathrm{TRIL}^{n,n}$, $l_{i,i} = 1 \ \forall i$, $D = \mathrm{diag}(r_{1,1}, \ldots, r_{n,n})$. Podstawiając $\vec{y} := L^H * \vec{x}$, mamy

$$\begin{aligned} \vec{x}^H * A * \vec{x} &= \vec{x}^H * L * D * L^H * \vec{x} = (L^H * \vec{x})^H * D * (L^H * \vec{x}) \\ &= \vec{y}^H * D * \vec{y} = \sum_{i=1}^n |y_i|^2 \cdot r_{i,i}. \end{aligned}$$

Stąd $A > 0$ wtedy i tylko wtedy gdy $r_{i,i} > 0$ $\forall i$, oraz $A < 0$ wtedy i tylko wtedy gdy $r_{i,i} < 0$ $\forall i$.

Dowód uzupełnia spostrzeżenie, że

$$A^{(k)} = L^{(k)} * R^{(k)} \; = \; L^{(k)} * D^{(k)} * (L^{(k)})^H$$

oraz

$$\det(A^{(k)}) = |\det(L^{(k)})|^2 \cdot \det(D^{(k)}) = \prod_{i=1}^{k} r_{i,i}.$$

Rozdział 11

Przestrzenie Euklidesowe

11.1 Definicja, iloczyn skalarny i norma

Definicja 11.1 Przestrzenią Euklidesową *nazywamy parę*

$$\{\mathcal{X}_{|\mathbf{K}}, \varphi\},$$

gdzie $\mathcal{X}_{|\mathbf{K}}$ jest przestrzenią liniową nad $\mathbf{K}$, a φ formą dwuliniową (hermitowską) <u>*dodatnio*</u> *określoną na $\mathcal{X}_{|\mathbf{K}}$, zwaną* iloczynem skalarnym.

Dla uproszczenia, będziemy dalej pisać (x, y) zamiast $\varphi(x, y)$ oraz $(\mathbb{A}, \mathbb{B})$ zamiast $\varphi(\mathbb{A}, \mathbb{B})$.

Własności formy implikują następujące własności iloczynu skalarnego:

(1) $(x, y_1 * \alpha_1 + y_2 * \alpha_2) = (x, y_1) * \alpha_1 + (x, y_2) * \alpha_2 \quad \forall x, y_1, y_2 \in \mathcal{X}$ $\forall \alpha_1, \alpha_2 \in \mathbf{K}$,

(2) $(x, y) = \overline{(y, x)}, \quad \forall x, y \in \mathcal{X}$

(3) $(x, x) \geq 0 \ \ \forall x \in \mathcal{X}$, oraz $(x, a) = 0 \Longleftrightarrow x = \mathbf{0}$.

Zdefiniujmy $\gamma(x) = (x, x)^{1/2}$, $x \in \mathcal{X}$. Wtedy funkcja γ ma następujące własności:

(i) $\gamma(x) \geq 0 \ \ \forall x \in \mathcal{X}$, oraz $\gamma(x) = 0 \Longleftrightarrow x = \mathbf{0}$.

(ii) $\gamma(x * \alpha) = \gamma(x) * |\alpha| \quad \forall x \in \mathcal{X} \ \forall \alpha \in \mathbf{K}$,

(iii) $\gamma(x + y) \leq \gamma(x) + \gamma(y) \quad \forall x, y \in \mathcal{X}$.

Własności (i) oraz (ii) są oczywiste. Aby pokazać (iii) zauważmy, że

$$\begin{aligned} \gamma(x+y)^2 &= (x+y,x+y) \;=\; (x,x)+(y,x)+(x,y)+(y,y) \\ &= (x,x)+2\cdot\Re(x,y)+(y,y) \end{aligned}$$

oraz

$$\begin{aligned} (\gamma(x)+\gamma(y))^2 &= ((x,x)^{1/2}+(y,y)^{1/2})^2 \\ &= (x,x)+2\cdot(x,x)^{1/2}\cdot(y,y)^{1/2}+(y,y). \end{aligned}$$

Własność (iii) wynika teraz z nierówności

$$\Re(x,y) \;\le\; |\Re(x,y)| \;\le\; |(x,y)| \;\le\; (x,x)^{1/2}\cdot(y,y)^{1/2},$$

przy czym ostatnia z nich to nierówność Schwarza, którą znamy już z lematu 3.1. (Co prawda, wtedy rozpatrywany był szczególny przypadek $\mathcal{X}_{|\mathbf{K}} = \mathbf{K}^n_{\mathbf{K}}$ i $(\vec{x},\vec{y}) = \vec{y}^H * \vec{x}$, ale w ogólnym przypadku dowód jest niemal identyczny.)

Własności (i)-(iii) są ogólnymi warunkami normy w przestrzeni liniowej. (Wcześniej, w rozdziale 3.1 podaliśmy te warunki dla szczególnego przypadku $\mathcal{X} = \mathbf{K}^{m,n}$.) Stąd

$$\|x\| \;:=\; (x,x)^{1/2}$$

definiuje *normę* w $\mathcal{X}_{|\mathbf{K}}$ (generowaną przez iloczyn skalarny $(\cdot,\cdot)$). Przypomnijmy jeszcze raz *nierówność Schwarza* (w przestrzeni Euklidesowej):

$$|(x,y)| \;\le\; \|x\|\cdot\|y\| \qquad \forall x,y\in\mathcal{X}.$$

Dokładniejsze prześledzenie dowodu tej nierówności (patrz znów dowód lematu 3.1) pokazuje, że powyżej równość zachodzi wtedy i tylko wtedy gdy x i y są liniowo zależne.

11.2 Rzut prostopadły

11.2.1 Zadanie aproksymacji

Następujące twierdzenie rozwiązuje zadanie aproksymacji (przybliżania) elementów przestrzeni $\mathcal{X}$ elementami jej podprzestrzeni.

Twierdzenie 11.1 *Niech $\mathcal{Y}\subseteq\mathcal{X}$ będzie podprzestrzenią. Wtedy dla każdego $x\in\mathcal{X}$ istnieje dokładnie jeden $x_{\mathcal{Y}}\in\mathcal{Y}$ taki, że dla wszystkich $y\in\mathcal{Y}$*

$$y\neq x_{\mathcal{Y}} \;\Longrightarrow\; \|x-x_{\mathcal{Y}}\| < \|x-y\|.$$

Dowód. Niech $s = \dim(\mathcal{Y})$ i $\mathbb{Y} \in \mathcal{X}^{1,s}$ będzie bazą $\mathcal{Y}$. Pokażemy, że $x_{\mathcal{Y}}$ wyraża się wzorem

$$x_{\mathcal{Y}} = \mathbb{Y} * \vec{a}^*, \qquad \text{gdzie} \quad \vec{a}^* := (\mathbb{Y}, \mathbb{Y})^{-1} * (\mathbb{Y}, x) \in \mathbf{K}^s. \tag{11.1}$$

Rzeczywiście, jeśli $y \in \mathcal{Y}$ i $y \neq x_{\mathcal{Y}}$ to $y = \mathbb{Y} * \vec{a}$ dla pewnego $\vec{a} \neq \vec{a}^*$. Wtedy

$$\begin{aligned} \|x - y\|^2 &= (x - y, x - y) = ((x - x_{\mathcal{Y}}) + (x_{\mathcal{Y}} - y), (x - x_{\mathcal{Y}}) + (x_{\mathcal{Y}} - y)) \\ &= \|x - x_{\mathcal{Y}}\|^2 + 2 \cdot \Re(x_{\mathcal{Y}} - y, x - x_{\mathcal{Y}}) + \|x_{\mathcal{Y}} - y\|^2. \end{aligned}$$

Wobec tego, że

$$(\mathbb{Y}, x_{\mathcal{Y}}) = (\mathbb{Y}, \mathbb{Y} * \vec{a}^*) = (\mathbb{Y}, \mathbb{Y}) * \vec{a}^* = (\mathbb{Y}, x),$$

mamy

$$(x_{\mathcal{Y}} - y, x - x_{\mathcal{Y}}) = (\mathbb{Y} * (\vec{a}^* - \vec{a}), x - x_{\mathcal{Y}}) = (\vec{a}^* - \vec{a})^H * (\mathbb{Y}, x - x_{\mathcal{Y}}) = 0.$$

Stąd

$$\|x - y\|^2 = \|x - x_{\mathcal{Y}}\|^2 + \|x_{\mathcal{Y}} - y\|^2 > \|x - x_{\mathcal{Y}}\|^2.$$

Uwaga. Z jednoznaczności najlepszej aproksymacji wynika, że $x_{\mathcal{Y}}$ we wzorze (11.1) nie zależy od wyboru bazy $\mathbb{Y}$. Można to również łatwo sprawdzić bezpośrednio. Jeśli bowiem weźmiemy inną bazę, powiedzmy $\mathbb{Z}$, podprzestrzeni $\mathcal{Y}$ to $\mathbb{Y} = \mathbb{Z} * C$ dla pewnej nieosobliwej macierzy C, a stąd

$$\begin{aligned} \mathbb{Z} * (\mathbb{Z}, \mathbb{Z})^{-1} * (\mathbb{Z}, x) &= \mathbb{Y} * C * (\mathbb{Y} * C, \mathbb{Y} * C)^{-1} * (\mathbb{Y} * C, x) \\ &= \mathbb{Y} * C * (C^H * (\mathbb{Y}, \mathbb{Y}) * C)^{-1} * C^H * (\mathbb{Y}, x) \\ &= \mathbb{Y} * (\mathbb{Y}, \mathbb{Y})^{-1} * (\mathbb{Y}, x). \end{aligned}$$

11.2.2 Twierdzenie o rzucie prostopadłym

Definicja 11.2 *Powiemy, że dwa elementy x i y danej przestrzeni Euklidesowej $\mathcal{X}_{|\mathbf{K}}$ z iloczynem skalarnym $(\cdot,\cdot)$ są* prostopadłe *(albo* ortogonalne*), co zapisujemy $x \perp y$, gdy ich iloczyn skalarny wynosi zero, tzn.*

$$x \perp y \iff (x, y) = 0.$$

Zauważmy, że jeśli wektory $x, y \in \mathcal{X}$ są prostopadłe, $x \perp y$, to zachodzi równość

$$\|x + y\|^2 \;=\; \|x\|^2 + \|y\|^2, \tag{11.2}$$

którą odczytujemy jako (znane ze szkoły w szczególnym przypadku) TWIERDZENIE PITAGORASA. Oczywiście, zachodzi również twierdzenie odwrotne, tzn. równość (11.2) implikuje protopadłość wektorów x i y.

Najlepsza aproksymacja w podprzestrzeni $\mathcal{Y}$ posiada dodatkową ważną własność związaną z pojęciem prostopadłości.

Twierdzenie 11.2 (O RZUCIE PROSTOPADŁYM)
Niech $x_{\mathcal{Y}}$ będzie najlepszą aproksymacją elementu $x \in \mathcal{X}$ w podprzestrzeni $\mathcal{Y} \subseteq \mathcal{X}$. Wtedy

$$(y, x - x_{\mathcal{Y}}) \;=\; 0 \qquad\qquad \forall y \in \mathcal{Y} \tag{11.3}$$

tzn. $x - x_{\mathcal{Y}}$ jest prostopadły do podprzestrzeni $\mathcal{Y}$.

Ponadto, $x_{\mathcal{Y}}$ jest jedynym elementem w $\mathcal{Y}$ spełniającym (11.3).

Dowód. Wykorzystując notację z dowodu twierdzenia 11.1, dla dowolnego $y \in \mathcal{Y}$ mamy

$$(y, x - x_{\mathcal{Y}}) \;=\; \vec{a}^H * (\mathbb{Y}, x - x_{\mathcal{Y}}) \;=\; \vec{a}^H * \vec{0} \;=\; 0.$$

Jeśli zaś $y_0 = \mathbb{Y} * \vec{a}_0$ i dla każdego $\vec{a}$ mamy $(\mathbb{Y} * \vec{a}, x - \mathbb{Y} * \vec{a}_0) = 0$, to $(\mathbb{Y}, x - \mathbb{Y} * \vec{a}_0) = 0$, a stąd

$$\vec{a}_0 = (\mathbb{Y}, \mathbb{Y})^{-1} * (\mathbb{Y}, x) = \vec{a}^*$$

i $y_0 = x_{\mathcal{Y}}$.

Ze względu na twierdzenie 11.2, element $x_{\mathcal{Y}}$ najlepszej aproksymacji nazywany jest również *rzutem prostopadłym (ortogonalnym)* elementu x na podprzestrzeń $\mathcal{Y}$.

11.3 Układy ortogonalne

11.3.1 Macierz Grama

Definicja 11.3 *Niech $\mathbb{A} = [y_1, y_2, \ldots, y_s] \in \mathcal{X}^{1,s}$. Macierz*

$$(\mathbb{A}, \mathbb{A}) \in \mathrm{Herm}^{s,s}$$

nazywamy macierzą Grama *układu $\{y_i\}_{i=1}^s$.*

Wobec równości

$$\vec{a}^H * (\mathbb{A}, \mathbb{A}) * \vec{a} \;=\; (\mathbb{A} * \vec{a}, \mathbb{A} * \vec{a}) \;=\; \|\mathbb{A} * \vec{a}\|^2 \;\geq\; 0$$

mamy natychmiast, że macierz Grama jest zawsze nieujemnie określona. Ponadto, jest ona dodatnio określona wtedy i tylko wtedy gdy układ $\{y_i\}_{i=1}^s$ jest liniowo niezależny.

Jeśli $(\mathbb{A}, \mathbb{A}) = \mathrm{diag}(\delta_1, \ldots, \delta_s)$, przy czym $\delta_i = (y_i, y_i) > 0\ \forall i$ to układ $\{y_i\}_{i=1}^s$ nazywamy *ortogonalnym*. Jeśli ponadto $(y_i, y_i) = 1\ \forall i$, czyli gdy $(\mathbb{A}, \mathbb{A}) = I_s$, to układ ten nazywamy *ortonormalnym*.

Załóżmy teraz, że układ $\mathbb{Y} = [y_1, \ldots, y_s]$ jest liniowo niezależny, oraz niech

$$\mathcal{Y} = \mathrm{span}(y_1, y_2, \ldots, y_s).$$

Wtedy, jak wiemy z twierdzenia 11.1, rzut prostopadły $x \in \mathcal{X}$ na podprzestrzeń $\mathcal{Y}$ wyraża się wzorem

$$x_{\mathcal{Y}} \;=\; \mathbb{Y} * (\mathbb{Y}, \mathbb{Y})^{-1} * (\mathbb{Y}, x).$$

Wzór ten ma szczególnie prostą postać gdy baza $\mathbb{Y}$ tworzy układ ortogonalny. Wtedy

$$x_{\mathcal{Y}} \;=\; \sum_{i=1}^{s} y_i * \frac{(y_i, x)}{(y_i, y_i)}.$$

Jeśli zaś baza tworzy układ ortonormalny to

$$x_{\mathcal{Y}} \;=\; \sum_{i=1}^{s} y_i * (y_i, x).$$

Z tych względów pożądane jest posiadanie baz ortogonalnych podprzestrzeni.

11.3.2 Ortogonalizacja Grama-Schmidta

Okazuje się, że dowolną bazę podprzestrzeni można stosunkowo łatwo przekształcić w bazę ortogonalną. Służy temu proces zwany *ortogonalizacją Grama-Schmidta*.

Niech $y_1, y_2, \ldots, y_s$ będą liniowo niezależne oraz

$$\mathbb{Y}_k := [y_1, \ldots, y_k], \qquad \mathcal{Y}_k = \mathrm{span}(y_1, \ldots, y_k), \qquad 1 \leq k \leq s.$$

Oczywiście $\dim(\mathcal{Y}_k) = k\ \forall k$ oraz

$$\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \cdots \subset \mathcal{Y}_s \subseteq \mathcal{X}.$$

Twierdzenie 11.3 (Ortogonalizacja Grama-Schmidta)
Następujący proces:

$$
\begin{array}{l}
\{\ z_1 := y_1; \\
\quad \textbf{for}\ k := 2\ \textbf{to}\ s\ \textbf{do} \\
\qquad z_k := y_k - \langle\ \textit{rzut prostopadły}\ y_k\ \textit{na}\ \mathcal{Y}_{k-1}\ \rangle \\
\}
\end{array}
$$

produkuje układy ortogonalne $\mathbb{Z}_k = [z_1, \ldots, z_k]$ *takie, że*

$$\mathrm{span}(z_1, z_2, \ldots, z_k) = \mathcal{Y}_k, \qquad 1 \le k \le s.$$

Dowód. (Indukcja względem k.)
Dla $k = 1$ twierdzenie jest oczywiste. Niech $k \ge 2$. Wtedy, wobec założenia indukcyjnego, mamy $\mathcal{Y}_{k-1} = \mathrm{span}(z_1, \ldots, z_{k-1})$ oraz układ $\{z_i\}_{i=1}^{k-1}$ jest ortogonalny. Jeśli teraz r_k jest rzutem ortogonalnym y_k na $\mathcal{Y}_{k-1}$ to z twierdzenia o rzucie ortogonalnym mamy, że $z_k = y_k - r_k \neq 0$ jest prostopadły do $\mathcal{Y}_{k-1}$, a stąd układ $\{z_i\}_{i=1}^{k}$ jest też ortogonalny. Oczywiście, przy tym $\mathrm{span}(z_1, \ldots, z_k) = \mathcal{Y}_k$, co kończy dowód.

Ortogonalizację Grama-Schmidta możemy zapisać jako algorytm generujący układ $\{z_i\}_{i=1}^{s}$ z układu $\{y_i\}_{i=1}^{s}$, w następujący sposób:

$$
\begin{array}{ll}
\{ & z_1 := y_1; \quad \delta_1 := (z_1, z_1); \\
 & \textbf{for}\ k := 2\ \textbf{to}\ s\ \textbf{do} \\
 & \{\ \textbf{for}\ j := 1\ \textbf{to}\ k-1\ \textbf{do}\ \ c_{j,k} := (z_j, y_k)/\delta_j; \\
 & \quad z_k := y_k - \sum_{j=1}^{k-1} z_j * c_{j,k}; \quad \delta_k := (z_k, z_k) \\
 & \} \\
\} &
\end{array}
$$

Algorytm ten produkuje "po drodze" współczynniki $c_{j,k}$ dla $2 \le k \le s$, $1 \le j \le k-1$. Jeśli dodatkowo zdefiniujemy $c_{k,k} = 1$ dla $1 \le k \le s$, oraz $c_{j,k} = 0$ dla $1 \le k \le s-1$, $k+1 \le j \le s$, to dostaniemy

$$y_k \ = \ \sum_{j=1}^{k} c_{j,k} * z_j,$$

czyli

$$\mathbb{Y}_k \ = \ \mathbb{Z}_k * C_k, \qquad \text{gdzie} \quad C_k = (c_{i,j})_{i,j=1}^{k},$$

albo, po normalizacji bazy $z_1, \ldots, z_s$,

$$\mathbb{Y}_k = \hat{\mathbb{Z}}_k * \hat{C}_k,$$

gdzie $\hat{\mathbb{Z}}_k = \mathbb{Z} * D_k^{-1}$, $\hat{C}_k = D_k * C_k$, $D_k = \mathrm{diag}(\delta_1^{1/2}, \ldots, \delta_k^{1/2})$.

Zauważmy, że macierze C_k i $\hat{C}_k$ są trójkątne górne.

11.3.3 Rozkład ortogonalno-trójkątny macierzy

Ważnym przypadkiem szczególnym jest $\mathcal{X}_{|\mathbf{K}} = \mathbf{K}^n_{|\mathbf{K}}$, $\mathbf{K} \subseteq \mathbf{C}$, ze "zwykłym" iloczynem skalarnym $(\vec{x}, \vec{y}) = \vec{y}^H * \vec{x}$. Niech

$$A = [\vec{a}_1, \ldots, \vec{a}_n] \in \mathbf{K}^{m,n},$$

gdzie

$$\mathrm{rank}(A) = n \leq m,$$

tzn. kolumny macierzy są liniowo niezależne. Wtedy, przeprowadzając ortonormalizację (czyli ortogonalizację, a następnie normalizację) kolumn macierzy A otrzymujemy macierz $Q \in \mathbf{K}^{m,n}$, $Q = [\vec{q}_1, \ldots, \vec{q}_n]$, której kolumny $\vec{q}_j$ tworzą układ ortonormalny,

$$Q^H * Q = I_n,$$

oraz macierz trójkątną górną $R \in \mathrm{TRIU}^{n,n}$ takie, że

$$A = Q * R.$$

Jest to *rozkład ortogonalno-trójkątny* macierzy.