Macierz Grama. Kryterium Sylvestera

Na ostatnim wykładzie omówione zostały podstawowe metody wyznaczania bazy ortogonalnej przestrzeni euklidesowej. Dziś przekonamy się, że z iloczynem skalarnym i układem wektorów przestrzeni euklidesowej $(V, \langle , \rangle)$ związać można pewne bardzo istotne macierze, tzw. macierze Grama. W języku wyznaczników związanych z tymi macierzami sformułujemy kryterium wyróżniające iloczyny skalarne spośród funkcji $f : V \times V \to \mathbb{R}$ liniowych ze względu na każdą ze zmiennych i symetrycznych, nazwane imieniem Jacoba Sylvestera.

Definicja 56. *Niech $\alpha_1, \dots, \alpha_k$ będzie układem wektorów przestrzeni euklidesowej $(V, \langle , \rangle)$.* Macierzą Grama *układu $\alpha_1, \dots, \alpha_k$ nazywamy macierz*

$$G(\alpha_1, \dots, \alpha_k) \in M_k(\mathbb{R}),$$

która w i-tym wierszu i j-tej kolumnie ma element $\langle \alpha_i, \alpha_j \rangle$, to znaczy:

$$G(\alpha_1, \dots, \alpha_k) = \begin{bmatrix} \langle \alpha_1, \alpha_1 \rangle & \langle \alpha_1, \alpha_2 \rangle & \dots & \langle \alpha_1, \alpha_k \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \langle \alpha_2, \alpha_2 \rangle & \dots & \langle \alpha_2, \alpha_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \alpha_k, \alpha_1 \rangle & \langle \alpha_k, \alpha_2 \rangle & \dots & \langle \alpha_k, \alpha_k \rangle \end{bmatrix}$$

Wyznacznikiem Grama *układu $\alpha_1, \dots, \alpha_k$ nazywamy liczbę*

$$W(\alpha_1, \dots, \alpha_k) = \det G(\alpha_1, \dots, \alpha_k) \in \mathbb{R}.$$

Zauważmy, że rozmiar macierzy Grama nie musi mieć nic wspólnego z wymiarem przestrzeni V. W n-wymiarowej przestrzeni możemy rozpatrywać układ złożony na przykład z dwóch wektorów i wyznaczyć jego macierz Grama rozmiarów 2×2. Możemy też rozważać układ złożony z $2n$ wektorów i jego macierz Grama będzie rozmiarów $2n \times 2n$. Z symetryczności iloczynu skalarnego $\langle \alpha_i, \alpha_j \rangle = \langle \alpha_j, \alpha_i \rangle$ dla każdych $1 \leq i, j \leq k$ widzimy, że macierz Grama jest symetryczna.

Warto odnotować, że macierze Grama rozważać można w ogólniejszym kontekście. Nie ma przeszkód, by rozważać $G(\alpha_1, \dots, \alpha_k)$ dla układu wektorów $\alpha_1, \dots, \alpha_k$ w przestrzeni nieskończenie wymiarowej V z iloczynem skalarnym $\langle , \rangle : V \times V \to \mathbb{R}$.

Zobaczmy przykłady macierzy Grama:

- W przestrzeni euklidesowej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym rozpatrzmy układ wektorów $\alpha_1 = (1,0,1), \alpha_2 = (1,1,2)$. Wówczas $\langle \alpha_1, \alpha_1 \rangle = 2$, $\langle \alpha_1, \alpha_2 \rangle = \langle \alpha_2, \alpha_1 \rangle = 3$, $\langle \alpha_2, \alpha_2 \rangle = 6$. Stąd:

$$G(\alpha_1, \alpha_2) = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}, \quad W(\alpha_1, \alpha_2) = \begin{vmatrix} 2 & 3 \\ 3 & 6 \end{vmatrix} = 3.$$

- W $\mathbb{R}^2$ z iloczynem skalarnym zadanym wzorem

$$\langle (x_1, x_2), (y_1, y_2) \rangle = 2x_1y_1 + x_1x_2 + x_2x_1 + x_2y_2$$

  rozpatrzmy układ wektorów

$$\alpha_1 = (1,0), \alpha_2 = (0,1), \alpha_3 = (1,1).$$

  Wówczas:

$$G(\alpha_1, \alpha_2, \alpha_3) = \begin{bmatrix} \langle \alpha_1, \alpha_1 \rangle & \langle \alpha_1, \alpha_2 \rangle & \langle \alpha_1, \alpha_3 \rangle \\ \langle \alpha_2, \alpha_1 \rangle & \langle \alpha_2, \alpha_2 \rangle & \langle \alpha_2, \alpha_3 \rangle \\ \langle \alpha_3, \alpha_1 \rangle & \langle \alpha_3, \alpha_2 \rangle & \langle \alpha_3, \alpha_3 \rangle \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix}, \quad W(\alpha_1, \alpha_2, \alpha_3) = 0.$$

  Zwróćmy uwagę na to, że $\alpha_1 + \alpha_2 = \alpha_3$, oraz dla macierzy $G(\alpha_1, \alpha_2, \alpha_3)$ pierwszy wiersz + drugi wiersz = trzeci wiersz oraz pierwsza kolumna + druga kolumna = trzecia kolumna. Widać dlaczego?

- Zauważmy, że $G(\alpha_1, \ldots, \alpha_n)$ jest macierzą diagonalną wtedy i tylko wtedy, gdy układ $\alpha_1, \ldots, \alpha_n$ jest układem prostopadłym. Diagonalna macierz Grama ma na przekątnej wyrazy dodatnie. Szczególnie istotny przypadek zachodzi, gdy $\alpha_1, \ldots, \alpha_n$ jest bazą ortonormalną przestrzeni V. Wówczas macierz Grama jest macierzą identycznościową.

Fakt 76. *Macierz Grama $G(\alpha_1, \ldots, \alpha_k)$ układu wektorów w $(V, \langle , \rangle)$ wyznacza wartości iloczynu skalarnego na $\mathrm{lin}(\alpha_1, \ldots, \alpha_k)$: dla $v = a_1\alpha_1 + \ldots + a_k\alpha_k$ oraz $w = b_1\alpha_1 + \ldots + b_k\alpha_k$ mamy:*

$$\langle a_1\alpha_1 + \ldots + a_k\alpha_k, b_1\alpha_1 + \ldots + b_k\alpha_k \rangle = \begin{bmatrix} a_1 & \ldots & a_k \end{bmatrix} \cdot G(\alpha_1, \ldots, \alpha_k) \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}.$$

W szczególności, jeśli $\beta_1, \ldots, \beta_n$ jest bazą ortonormalną $(V, \langle , \rangle)$, wówczas dla dowolnych $\alpha, \beta \in V$ mamy

$$\langle \alpha, \beta \rangle = \begin{bmatrix} \langle \alpha, \beta_1 \rangle & \ldots & \langle \alpha, \beta_n \rangle \end{bmatrix} \cdot \begin{bmatrix} \langle \beta, \beta_1 \rangle \\ \vdots \\ \langle \beta, \beta_n \rangle \end{bmatrix}.$$

Dowód. Teza pierwszej części wynika z następującego ciągu równości:

$$\begin{aligned}\langle a_1\alpha_1 + \ldots + a_n\alpha_n, b_1\alpha_1 + \ldots + b_n\alpha_n\rangle &= a_1\langle \alpha_1, b_1\alpha_1 + \ldots + b_n\alpha_n\rangle + \\ &+ a_2\langle \alpha_2, b_1\alpha_1 + \ldots + b_n\alpha_n\rangle + \\ &+ \ldots + \\ &+ a_n\langle \alpha_n, b_1\alpha_1 + \ldots + b_n\alpha_n\rangle = \\ &= \sum_{1\le i,j\le n} a_i b_j \langle \alpha_i, \alpha_j\rangle\end{aligned}$$

Uzyskany wynik jest równy $\begin{bmatrix} a_1 & \ldots & a_k\end{bmatrix} \cdot G(\alpha_1, \ldots, \alpha_k) \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_k\end{bmatrix}$.

Druga część tezy jest zastosowaniem pierwszej. Skoro baza $\beta_1, \ldots, \beta_n$ jest ortonormalna, to $G(\beta_1, \ldots, \beta_n) = I$. Co więcej, korzystając z faktu o postaci współrzędnych wektora w bazie ortonormalnej mamy:

$$\begin{aligned}\alpha &= \langle \alpha, \beta_1\rangle\beta_1 + \langle \alpha, \beta_2\rangle\beta_2 + \ldots + \langle \alpha, \beta_n\rangle\beta_n, \\ \beta &= \langle \beta, \beta_1\rangle\beta_1 + \langle \beta, \beta_2\rangle\beta_2 + \ldots + \langle \beta, \beta_n\rangle\beta_n.\end{aligned}$$

□

Przykład. W przestrzeni euklidesowej $\mathbb{R}^3$ ze standardowym iloczynem skalarnym rozpatrzmy układ wektorów $\alpha_1 = (1,0,1), \alpha_2 = (1,1,2)$. Wówczas $\langle \alpha_1, \alpha_1\rangle = 2$, $\langle \alpha_1, \alpha_2\rangle = \langle \alpha_2, \alpha_1\rangle = 3$, $\langle \alpha_2, \alpha_2\rangle = 6$. Stąd:

$$G(\alpha_1, \alpha_2) = \begin{bmatrix} 2 & 3 \\ 3 & 6\end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2\end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2\end{bmatrix}.$$

Fakt 77. *Dla układu wektorów $\alpha_1, \ldots, \alpha_k$ przestrzeni euklidesowej $(V, \langle , \rangle)$ oraz dla bazy ortonormalnej $\beta_1, \ldots, \beta_n$ tej przestrzeni niech A będzie macierzą rozmiaru $n \times k$ mającą w j-tej kolumnie współrzędne wektora α_j w bazie ortonormalnej $\beta_1, \ldots, \beta_n$, to znaczy: $A = (a_{ij})$, gdzie $a_{ij} = \langle \alpha_j, \beta_i\rangle$, dla $i = 1, \ldots, k$, $j = 1, \ldots, n$. Wówczas:*

$$G(\alpha_1, \ldots, \alpha_k) = A^T A.$$

W szczególności dla $k = n$ mamy:

$$W(\alpha_1, \ldots, \alpha_n) = (\det A)^2.$$

Dowód. Z definicji macierzy transponowanej i-ty wiersz macierzy A^T ma kolejne wyrazy takie same jak kolejne wyrazy i-tej kolumny macierzy A. Czyli zawiera on współrzędne wektora α_i w bazie $\beta_1, \ldots, \beta_n$. Traktując go jako macierz $1 \times k$ mamy:

$$\begin{bmatrix} \langle \alpha_i, \beta_1\rangle & \langle \alpha_i, \beta_2\rangle & \ldots & \langle \alpha_i, \beta_n\rangle\end{bmatrix}.$$

Z drugiej strony j-ta kolumna A ma współrzędne wektora α_j w bazie ortonormalnej $\beta_1, \ldots, \beta_n$, czyli: $\langle \alpha_j, \beta_1\rangle, \ldots, \langle \alpha_j, \beta_n\rangle$. Teza wynika zatem z poprzedniego faktu.

Druga część tezy wynika z tego, że $\det A = \det A^T$, gdy $k = n$, a zatem z twierdzenia Cauchy'ego o wyznacznikach mamy $\det G(\alpha_1, \ldots, \alpha_k) = \det A^T \det A = (\det A)^2$. Dowód jest zakończony. □

Fakt 78. *Dla każdego układu $\alpha_1, \ldots, \alpha_k$ wektorów przestrzeni euklidesowej $(V, \langle, \rangle)$ mamy nierówność $W(\alpha_1, \ldots, \alpha_k) \geq 0$. Co więcej następujące warunki są równoważne:*

(a) $\alpha_1, \ldots, \alpha_k$ *jest liniowo niezależny,*

(b) $W(\alpha_1, \ldots, \alpha_k) > 0$.

Dowód. Niech $\mathcal{B} = \{\beta_1, \ldots, \beta_n\}$ będzie bazą przestrzeni euklidesowej $W = \mathrm{lin}(\alpha_1, \ldots, \alpha_k)$ z iloczynem skalarnym $\langle, \rangle|_W$. Niech A będzie ponownie macierzą rozmiaru $n \times k$ mającą w j-tej kolumnie współrzędne wektora α_j w bazie $\beta_1, \ldots, \beta_n$. Zauważmy, że jeśli układ $\mathcal{A} = \{\alpha_1, \ldots, \alpha_k\}$ jest liniowo niezależny, to jest on bazą W, czyli macierz A jest odwracalną macierzą kwadratową postaci $M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}}$. Skoro $G(\alpha_1, \ldots, \alpha_k) = A^T A$, to w tym przypadku

$$W(\alpha_1, \ldots, \alpha_k) = \det(A^T A) = \det(A)^2 > 0.$$

Jeśli $\alpha_1, \ldots, \alpha_k$ jest układem liniowo zależnym, to macierz A ma rząd mniejszy niż k. Rzeczywiście, warunkiem równoważnym liniowej zależności układu $\alpha_1, \ldots, \alpha_k$ jest to, że jeden z wektorów α_i jest kombinacją liniową wektorów α_j, dla $i \neq j$. Niech to będzie

$$\alpha_i = \sum_{j \neq i} a_j \alpha_j.$$

Wówczas i-ta kolumna macierzy A ma postać:

$$\begin{bmatrix} \cdots & \left\langle \sum_{j\neq i} a_j\alpha_j, \beta_1 \right\rangle & \cdots \\ \cdots & \left\langle \sum_{j\neq i} a_j\alpha_j, \beta_2 \right\rangle & \cdots \\ \ddots & \vdots & \ddots \\ \cdots & \left\langle \sum_{j\neq i} a_j\alpha_j, \beta_n \right\rangle & \cdots \end{bmatrix} = \sum_{j \neq i} a_j \begin{bmatrix} \cdots & \langle \alpha_j, \beta_1 \rangle & \cdots \\ \cdots & \langle \alpha_j, \beta_2 \rangle & \cdots \\ \ddots & \vdots & \ddots \\ \cdots & \langle \alpha_j, \beta_n \rangle & \cdots \end{bmatrix}.$$

A zatem wykonując ciąg operacji elementarnych typu (1) polegających na odejmowaniu od i-tej kolumny j-tej przemnożonej przez a_j uzyskujemy wyzerowanie i-tej kolumny. Skoro rząd macierzy nie zmienia się przy operacjach elementarnych, to rząd macierzy A jest mniejszy niż k. Z pierwszego semestru wiemy, że dla dowolnych macierzy X, Y, dla których istnieje iloczyn XY, mamy

To po prostu wynika stąd, że kolumny macierzy XY są kombinacjami liniowymi kolumn macierzy Y.

$$r(XY) \leq \min\{r(X), r(Y)\}.$$

Stąd $r(G(\alpha_1, \ldots, \alpha_k)) = r(A^T A) \leq \min\{r(A^T), r(A)\} = r(A) < k$. Zatem $W(\alpha_1, \ldots, \alpha_k) = 0$.

W szczególności pokazaliśmy też pierwszą część tezy.

□

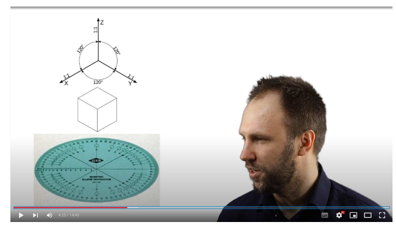
Zwieńczeniem rozważań w tej części wykładu jest niezwykle istotne (i popularne) kryterium Sylvestera. Mówi ono o tym kiedy symetryczna macierz kwadratowa może być macierzą Grama.

Kilka zdań wprowadzenia: rozpatrzmy n wymiarową przestrzeń liniową V nad ciałem $\mathbb{R}$ i jej bazę $\alpha_1, \ldots, \alpha_n$. Funkcja $f : V \times V \to \mathbb{R}$ zadana dla każdych wektorów $x_1\alpha_1 + \ldots + x_n\alpha_n$, $y_1\alpha_1 + \ldots + y_n\alpha_n$ wzorem:

$$f(x_1\alpha_1 + \ldots + x_n\alpha_n, y_1\alpha_1 + \ldots + y_n\alpha_n) = \sum_{i,j=1}^{n} a_{ij}x_iy_j$$

jest liniowa ze względu na każdą ze zmiennych oraz $f(\alpha, \beta) = f(\beta, \alpha)$, dla każdych $\alpha, \beta \in V$. Pytanie: co musimy wiedzieć o współczynnikach a_{ij}, aby powiedzieć, że spełniony jest warunek $f(\alpha, \alpha) > 0$, dla $\alpha \neq 0$? Okazuje się, że kryterium jest bardzo czytelne.

Definicja 57. *Dla dowolnej macierzy $A \in M_n(K)$ oraz dla $i = 1, \ldots, n$ niech $A^{(i)} \in M_i(K)$ oznacza macierz powstałą z A przez usunięcie ostatnich $n - i$ wierszy i ostatnich $n - i$ kolumn. Wyznacznik tej macierzy nazywamy* WIODĄCYM MINOREM GŁÓWNYM STOPNIA *i macierzy A.*



Po zapoznaniu się z dowodem kryterium, przykładem i dowodem, zachęcam też do obejrzenia poniższego filmu `https://youtu.be/92xYChexuyU`.

Fakt 79 (Kryterium Sylvestera)**.** *Niech V będzie przestrzenią liniową nad $\mathbb{R}$ i niech $\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$ będzie jej bazą. Dla dowolnej macierzy symetrycznej $A = [a_{ij}] \in M_n(\mathbb{R})$ rozpatrujemy funkcję $f : V \times V \to \mathbb{R}$ zadaną dla każdych wektorów postaci $x_1\alpha_1 + \ldots + x_n\alpha_n$, $y_1\alpha_1 + \ldots + y_n\alpha_n$ wzorem:*

$$f(x_1\alpha_1 + \ldots + x_n\alpha_n, y_1\alpha_1 + \ldots + y_n\alpha_n) = \sum_{i,j=1}^{n} a_{ij}x_iy_j.$$

Wówczas następujące warunki są równoważne.

(i) funkcja f jest iloczynem skalarnym na przestrzeni V,

(ii) $\det A^{(i)} > 0$, dla $i = 1, \ldots, n$.

Przykład: niech $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, gdzie

$$f((x_1, x_2, x_3), (y_1, y_2, y_3)) = (4 - r)x_1y_1 + x_1y_2 + x_1y_3 + x_2y_1 + x_2y_2 + x_2y_3 + x_3y_1 + x_3y_2 + rx_3y_3.$$

- Dla jakich $r \in \mathbb{R}$ funkcja f jest iloczynem skalarnym na $\mathbb{R}^3$?
- Dla jakich $r \in \mathbb{R}$ funkcja $f|_W$ jest iloczynem skalarnym na przestrzeni $W = \mathrm{lin}((1, 1, -1), (0, 1, 1))$?

Rozwiązanie. Podana funkcja jest dwuliniowa i symetryczna, a zatem możemy zastosować kryterium Sylvestera. Bierzemy bazę standardową $(\epsilon_1, \epsilon_2, \epsilon_3)$ przestrzeni $\mathbb{R}^3$ i macierz o wyrazach $f(\epsilon_i, \epsilon_j)$ postaci:

$$\begin{bmatrix} 4-r & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & r \end{bmatrix}.$$

Aby macierz ta była macierzą Grama potrzeba i wystarcza, aby:

$$4 - r > 0, \quad \begin{vmatrix} 4-r & 1 \\ 1 & 1 \end{vmatrix} > 0,$$

oraz

$$\begin{vmatrix} 4-r & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & r \end{vmatrix} > 0.$$

A zatem dostajemy układ warunków:

$$4 > r, \quad 3 > r, \quad -r^2 + 4r - 3 = -(r-1)(r-3) > 0.$$

Stąd f jest iloczynem skalarnym na $\mathbb{R}^3$ wtedy i tylko wtedy, gdy $r \in (1,3)$.

Odpowiedzmy na drugie pytanie. Bierzemy

$$\alpha_1 = (1,1,-1), \quad \alpha_2 = (0,1,1).$$

Wyznaczamy macierz o wyrazach $f(\alpha_i, \alpha_j)$. Ma ona postać:

$$\begin{bmatrix} 3 & 3-r \\ 3-r & 3+r \end{bmatrix}.$$

A zatem jest ona macierzą Grama wtedy i tylko wtedy, gdy $3(3+r) - (3-r)^2 > 0$, co daje $-r^2 + 9r > 0$, czyli $r(9-r) > 0$. A zatem w tym przypadku $r \in (0,9)$.

Dowód. Dowodzimy (i) $\Rightarrow$ (ii). Jeśli f jest iloczynem skalarnym na V, to oczywiście A jest macierzą Grama układu $(\alpha_1, \dots, \alpha_n)$. Ograniczenie $f|_{W_i}$ do podprzestrzeni $W_i = \mathrm{lin}(\alpha_1, \dots, \alpha_i)$ jest iloczynem skalarnym na przestrzeni W_i, dla każdego $i = 1, \dots, n$. Stąd, zgodnie z Faktem 78, macierze $A^{(i)} = G(\alpha_1, \dots, \alpha_i)$ mają dodatnie wyznaczniki.

Dowodzimy (ii) $\Rightarrow$ (i) przez indukcję po wymiarze n przestrzeni V. Oczywiście f spełnia trzy warunki: liniowości ze względu na każdą zmienną oraz warunek symetryczności (bo macierz A jest symetryczna) występujące w definicji iloczynu skalarnego. Trzeba jedynie pokazać, że f jest dodatnio określona, tzn. $f(\alpha, \alpha) > 0$ dla $0 \neq \alpha \in V$.

Dla $n = 1$ teza jest jasna, bo $V = \mathrm{lin}(\alpha_1)$ oraz $A = [a_{11}]$, przy czym $a_{11} > 0$ (bo to jest $\det A^{(1)}$). A zatem dla każdego niezerowego $a\alpha_1 \in V$

$$f(a\alpha_1, a\alpha_1). = a^2 a_{11} > 0.$$

Zatem f jest iloczynem skalarnym. Załóżmy, że dowodzona implikacja jest prawdziwa dla $n-1$. Dowodzimy dla n. Niech $W = \mathrm{lin}(\alpha_1, \dots, \alpha_{n-1})$.

Z założenia indukcyjnego $(W, f|_W)$ jest przestrzenią euklidesową. Niech $\beta_1, \ldots, \beta_{n-1}$ będzie bazą prostopadłą W otrzymaną z bazy $\alpha_1, \ldots, \alpha_{n-1}$ metodą Grama-Schmidta. W szczególności $f(\beta_i, \beta_j) = 0$, dla $i \neq j$. Definiujemy wektor β_n postaci:

$$\beta_n = \alpha_n - \sum_{i=1}^{n-1} \frac{f(\alpha_n, \beta_i)}{f(\beta_i, \beta_i)} \beta_i.$$

Wówczas nowo zdefiniowany wektor spełnia $f(\beta_n, \beta_i) = 0$, dla wszystkich $i = 1, \ldots, n-1$ (sprawdzamy to podobnie jak w dowodach z poprzedniego wykładu, po prostu stosując liniowość ze względu na każdą zmienną oraz symetryczność f). Rozpatrzmy teraz macierz $B = [b_{ij}] \in M_n(\mathbb{R})$, gdzie $b_{ij} = f(\beta_i, \beta_j)$, dla $i, j = 1, \ldots, n$, czyli:

$$B = \begin{bmatrix} f(\beta_1, \beta_1) & 0 & \ldots & 0 \\ 0 & f(\beta_2, \beta_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & f(\beta_n, \beta_n) \end{bmatrix}.$$

Teraz trzeba udowodnić coś intuicyjnie jasnego, ale technicznego w dowodzie. Mianowicie: w istocie macierz B powstaje z macierzy A przez wykonanie operacji elementarnych typu (1) na wierszach i kolumnach (dodawanie wiersza pomnożonego przez stałą do innego wiersza, analogicznie dla kolumn). Dlaczego to wystarczy?

Otóż operacje elementarne typu (1) nie zmieniają wyznacznika. Więc $\det A = \det B$. Wiemy też, że B jest diagonalna i ma na diagonali wyrazy $f(\beta_1, \beta_1), \ldots, f(\beta_n, \beta_n)$. Ich iloczyn to wyznacznik macierzy B. Wiemy z założenia indukcyjnego, że $f(\beta_1, \beta_1), \ldots, f(\beta_{n-1}, \beta_{n-1})$ są dodatnie, bo $f|_W$ jest iloczynem skalarnym i korzystamy z (i) $\Rightarrow$ (ii). Ostatni element diagonali $f(\beta_n, \beta_n)$ też musi być jednak dodatni, bo z (ii) mamy $\det A^{(n)} = \det A > 0$. To oznacza, że f jest iloczynem skalarnym, bo biorąc $0 \neq \alpha = x_1\beta_1 + \ldots + x_n\beta_n$ i korzystając z $f(\beta_i, \beta_j) = 0$, dla $i \neq j$, mamy $f(\alpha, \alpha) = \sum_{i=1}^{n} x_i^2 f(\beta_i, \beta_i) > 0$, co zakończy dowód.

Dowód wymaga dwóch kroków. Po pierwsze zajmiemy się macierzą $X_{\mathcal{A}}$ mającą w wierszach wektory $\alpha_1, \ldots, \alpha_n$ bazy $\mathcal{A}$, a następnie dopiero popatrzymy na minory macierzy A. Po pierwsze twierdzimy, że macierz $X_{\mathcal{B}}$ mająca w kolejnych wierszach wektory $\beta_1, \ldots, \beta_n$ powstaje z macierzy $X_{\mathcal{A}}$ poprzez operacje typu (1) na wierszach.

Mała delikatność: jeszcze nie wiemy czy układ $\beta_1, \ldots, \beta_n$ jest bazą V, bo nie wiemy czy f to iloczyn skalarny na V.

Pokażmy przez indukcję ze względu na m, że sprowadzenie pierwszych m wierszy macierzy $X_{\mathcal{A}}$ do pierwszych m wierszy macierzy $X_{\mathcal{B}}$ wymaga stosowania jedynie operacji (1) na pierwszych m wierszach.

Dla $m = 1$ mamy $\beta_1 = \alpha_1$, więc teza jest jasna, bo nie trzeba wykonywać żadnych operacji. Natomiast jeśli wiemy, że jest tak dla $m-1$, to bierzemy macierz $X_{\mathcal{A}}$. Zamieniamy pierwsze $m-1$ wierszy z założenia indukcyjnego na $\beta_1, \ldots, \beta_{m-1}$ operacjami (1) wykonywanymi w obrębie $m-1$ pierwszych wierszy. Dostajemy macierz $X_{\mathcal{A}'}$ której pierwsze $m-1$ wierszy to $\beta_1, \ldots, \beta_{m-1}$, a kolejne to $\alpha_m, \ldots, \alpha_n$.

Intuicja: wykonujemy na kolejnych wierszach $X_{\mathcal{A}}$ te same operacje, co przy ortogonalizacji G-S, choć nie wiemy czy f to iloczyn skalarny. Na i-tym wierszu stosujemy $i-1$ operacji (1), co zamienia α_i na $\beta_i = \alpha_i - c_1\beta_1 - \ldots - c_{i-1}\beta_{i-1}$.

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{m-1} \\ \alpha_m \\ \vdots \\ \alpha_n \end{bmatrix} \xrightarrow{\text{m-1 operacji}} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{m-1} \\ \alpha_m \\ \vdots \\ \alpha_n \end{bmatrix}$$

Z twierdzenia o ortogonalizacji oraz z definicji wektora β_n mamy

$$\beta_m = \alpha_m - c_1\beta_1 - c_2\beta_2 - \ldots - c_{m-1}\beta_{m-1}, \text{ gdzie } c_i = f(\alpha_m, \beta_i)/f(\beta_i, \beta_i).$$

A zatem od wiersza m-tego $X_{\mathcal{A}'}$ odejmujemy c_1 razy pierwszy wiersz. Od uzyskanego m-tego wiersza odejmujemy c_2 razy drugi wiersz itd, na końcu odejmując c_{m-1} razy wiersz $m-1$ od wiersza m-tego. W rezultacie dostajemy macierz $X_{\mathcal{B}'}$, która w pierwszych m wierszach ma wektory $\beta_1, \ldots, \beta_m$, uzyskane przez stosowanie operacji (1) na pierwszych m wierszach, a dalej (jeśli są) wektory $\alpha_{m+1}, \ldots, \alpha_n$. To kończy dowód indukcyjny, że $X_{\mathcal{B}}$ można uzyskać z $X_{\mathcal{A}}$ przez operacje (1).

Niech A' będzie f-macierzą Grama układu $\beta_1, \ldots, \beta_{m-1}, \alpha_m, \ldots, \alpha_n$ oraz B' niech będzie f-macierzą Grama układu $\beta_1, \ldots, \beta_m, \alpha_{m+1}, \ldots, \alpha_n$. Pisząc f-macierz Grama układu wektorów $v_1, \ldots, v_k$ mamy tu na myśli macierz o wyrazach $f(v_i, v_j)$, dla $1 \leq i, j \leq k$.

Do dowodu, że B jest uzyskiwana z A przez operacje elementarne wierszowe i kolumnowe typu (1) wystarczy pokazać, że B' można uzyskać z A' przez operacje typu (1) zmieniające jedynie wiersze i kolumny o indeksie m. Wynika to z faktu, że macierze A' oraz B' mają identyczne wyrazy poza m-tym wierszem i m-tą kolumną, bo układy $\beta_1, \ldots, \beta_{m-1}, \alpha_m, \ldots, \alpha_n$ oraz $\beta_1, \ldots, \beta_m, \alpha_{m+1}, \ldots, \alpha_n$ różnią się tylko m-tym elementem. Zobaczmy:

$$A' = \begin{bmatrix} & \ldots & f(\beta_1, \beta_{m-1}) & f(\beta_1, \alpha_m) & \ldots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f(\beta_{m-1}, \beta_1) & \ldots & f(\beta_{m-1}, \beta_{m-1}) & f(\beta_{m-1}, \alpha_m) & \vdots & f(\beta_{m-1}, \alpha_n) \\ f(\alpha_m, \beta_1) & \ldots & f(\alpha_m, \beta_{m-1}) & f(\alpha_m, \alpha_m) & \vdots & f(\alpha_m, \alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \ldots & f(\alpha_n, \beta_{m-1}) & f(\alpha_n, \alpha_m) & \ldots & \end{bmatrix}$$

$$B' = \begin{bmatrix} & \cdots & f(\beta_1, \beta_{m-1}) & f(\beta_1, \beta_m) & \cdots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f(\beta_{m-1}, \beta_1) & \cdots & f(\beta_{m-1}, \beta_{m-1}) & f(\beta_{m-1}, \beta_m) & \vdots & f(\beta_{m-1}, \alpha_n) \\ f(\beta_m, \beta_1) & \cdots & f(\beta_m, \beta_{m-1}) & f(\beta_m, \beta_m) & \vdots & f(\beta_m, \alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \cdots & f(\alpha_n, \beta_{m-1}) & f(\alpha_n, \beta_m) & \cdots & \end{bmatrix}$$

Wiemy, że $\beta_m = \alpha_m - c_1\beta_1 - c_2\beta_2 - \ldots - c_{m-1}\beta_{m-1}$, dla znanych c_i. Macierz $X_{\mathcal{B}'}$ uzyskiwaliśmy z $X_{\mathcal{A}'}$ przez wykonanie $m-1$ operacji odjęcia od m-tego wiersza wielokrotności c_i wiersza i-tego. Aby uzyskać macierz B' z A' trzeba wykonać $2m-2$ operacji typu (1). Pierwsze dwie to: odjęcie od m-tego wiersza c_1 razy wiersz pierwszy, potem od kolumny m-tej trzeba odjąć c_1 razy kolumnę pierwszą. Oto rezultat:

$$\begin{bmatrix} & \cdots & f(\beta_1, \beta_{m-1}) & f(\beta_1, \alpha_m - c_1\beta_1) & \cdots & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f(\beta_{m-1}, \beta_1) & \cdots & f(\beta_{m-1}, \beta_{m-1}) & f(\beta_{m-1}, \alpha_m - c_1\beta_1) & \vdots & f(\beta_{m-1}, \alpha_n) \\ f(\alpha_m - c_1\beta_1, \beta_1) & \cdots & f(\alpha_m - c_1\beta_1, \beta_{m-1}) & f(\alpha_m - c_1\beta_1, \alpha_m - c_1\beta_1) & \vdots & f(\alpha_m - c_1\beta_1, \alpha_n) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & \cdots & f(\alpha_n, \beta_{m-1}) & f(\alpha_n, \alpha_m - c_1\beta_1) & \cdots & \end{bmatrix}$$

Następne dwie operacje to: od m-tego wiersza odjąć c_2 razy drugi wiersz i potem od m-tej kolumny odjąć c_2 razy kolumnę drugą:

$$\begin{bmatrix} & \cdots & f(\beta_1, \alpha_m - c_1\beta_1 - c_2\beta_2) & \cdots & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ f(\alpha_m - c_1\beta_1 - c_2\beta_2, \beta_2) & \cdots & f(\alpha_m - c_1\beta_1 - c_2\beta_2, \alpha_m - c_1\beta_1 - c_2\beta_2) & \vdots & f(\alpha_m - c_1\beta_1 - c_2\beta_2, \alpha_n) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ & \cdots & f(\alpha_n, \alpha_m - c_1\beta_1 - c_2\beta_2) & \cdots & \end{bmatrix},$$

i tak dalej. W ten sposób, korzystając z liniowości f względem każdej ze zmiennych dostajemy żądaną postać B'. Zauważmy wszak, że m-ty wiersz macierzy B' to kombinacja liniowa c_1 razy pierwszy wiersz + c_2 razy drugi wiersz + $\ldots + c_{m-1}$ razy $m-1$-wszy wiersz. Podobnie dla kolumn. Zatem pokazaliśmy, że B można dostać z A operacjami elementarnymi typu (1), co kończy dowód implikacji (ii) $\Rightarrow$ (i). $\square$

Dowód był stosunkowo zawiły. Wynika to z pewnego ubóstwa narzędzi, którymi w tym momencie dysponujemy. Na kolejnych wykładach dowiemy się więcej o macierzy Grama. Za kilka tygodni poznamy uogólnienie kryterium Sylvestera zwane twierdzeniem o bezwładności.

Uzupełnienie. Macierzowy dowód kryterium Sylvestera

Dowód kryterium Sylvestera zaprezentowany podczas wykładu można uczytelnić za pomocą argumentów macierzowych. Przypomnijmy: przyglądamy się macierzy symetrycznej $G(f;\mathcal{A})$ o wyrazach $f(\alpha_i,\alpha_j)$, gdzie $f : V \times V \to \mathbb{R}$, zaś $\mathcal{A} = \{\alpha_1,\dots,\alpha_n\}$ jest bazą V. Kluczowa uwaga jest następująca: jeśli $\mathcal{B} = (\beta_1,\dots,\beta_n)$ jest również bazą V, to

$$G(f;\mathcal{B}) = \left(M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}}\right)^T \cdot G(f;\mathcal{A}) \cdot M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}}. \qquad (\ddagger)$$

Formułę tą wykażemy ponownie (i nieco inaczej) w ramach wykładu o formach dwuliniowych w maju.

Przeprowadźmy krótkie rozumowanie. Z definicji funkcji f wiemy, że

$$f(\alpha,\alpha') = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \cdot G(f;\mathcal{A}) \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

gdzie $(a_1,\dots,a_n)$, $(b_1,\dots,b_n)$ są wektorami współrzędnych wektorów α,α' w bazie $\mathcal{A}$. W szczególności, biorąc $\alpha = \beta_i$ oraz $\alpha' = \beta_j$ mamy:

$$f(\beta_i,\beta_j) = \begin{bmatrix} a'_1 & \dots & a'_n \end{bmatrix} \cdot G(f;\mathcal{A}) \cdot \begin{bmatrix} b'_1 \\ \vdots \\ b'_n \end{bmatrix},$$

gdzie $(a'_1,\dots,a'_n)$ oraz $(b'_1,\dots,b'_n)$ są wektorami wspólrzędnych wektorów β_i,β_j w bazie $\mathcal{A}$, czyli tak naprawdę odpowiednio: i-tą oraz j-tą kolumną macierzy $M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}}$. Dostajemy więc $(\ddagger)$.

Zakładamy, że dla bazy $\mathcal{A}$ macierz $G(f;\mathcal{A})$ ma dodatnie główne minory wiodące. Rozważmy postać blokową powyższej macierzy postaci:

Innymi słowy pokazujemy tu tylko trudniejszą implikację kryterium.

$$G(f;\mathcal{A}) = \begin{bmatrix} G(f;\mathcal{A})^{(n-1)} & \alpha \\ \alpha^T & b \end{bmatrix}, \text{ gdzie } \alpha \in \mathbb{R}^{n-1}, b \in \mathbb{R}.$$

Główne minory wiodące $G(f;\mathcal{A})^{(n-1)}$ są głównymi minorami wiodącymi $G(f;\mathcal{A})$. Zatem z założenia indukcyjnego $g = f|_{V_{n-1}\times V_{n-1}}$ jest iloczynem skalarnym na przestrzeni V_{n-1}, gdzie $V_{n-1} = \mathrm{lin}(\alpha_1,\dots,\alpha_{n-1})$.

Jeśli $\mathcal{C} = (\beta_1,\dots\beta_{n-1})$ jest bazą V_{n-1}, to w bazie $\mathcal{B} = (\beta_1,\dots\beta_{n-1},\alpha_n)$ przestrzeni V macierz $G(f;\mathcal{B})$ ma postać:

$$G(f;\mathcal{B}) = \begin{bmatrix} G(g;\mathcal{C}) & \gamma \\ \gamma^T & b \end{bmatrix}, \text{ gdzie } \gamma \in \mathbb{R}^{n-1}$$

i skoro $G(f;\mathcal{B}) = C^T G(f;\mathcal{A}) C$, dla pewnej macierzy odwracalnej C, to wobec założenia $\det G(f;\mathcal{A}) > 0$ (to jeden z minorów) mamy $\det G(f;\mathcal{B}) > 0$. Przyjmijmy dalej, że baza $\mathcal{C}$ jest ortonormalna.

Wobec tego macierz funkcji f w bazie $(\beta_1, \ldots, \beta_{n-1}, \alpha_n)$ ma postać

$$G(f;\mathcal{B}) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & a_1 \\ 0 & 1 & 0 & \ldots & 0 & a_2 \\ 0 & 0 & 1 & \ldots & 0 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & a_{n-1} \\ a_1 & a_2 & a_3 & \ldots & a_{n-1} & b \end{bmatrix}.$$

Niech P^T będzie macierzą powstałą przez zamianę ostatniego wiersza w macierzy identycznościowej $I_n \in M_n(\mathbb{R})$ na wiersz

$$\begin{bmatrix} -a_1 & -a_2 & -a_3 & \ldots & -a_{n-1} & 1 \end{bmatrix}.$$

Łatwo sprawdzić (patrząc jakie operacje elementarne związane są z mnożeniem przez P oraz P^T), że istnieje $d \in \mathbb{R}$, że:

$$P^T G(f;\mathcal{B}) P = \begin{bmatrix} I_{n-1} & 0 \\ 0 & d \end{bmatrix}.$$

Skoro zaś uzyskana macierz jest to kolejna macierz funkcji f (znowu dokonaliśmy zmiany bazy), to musi mieć dodatni wyznacznik. Zatem $d > 0$ i f jest w sposób oczywisty iloczynem skalarnym.

Kluczowe jest zrozumienie, że w istocie zaprezentowany dowód jest identyczny z przedstawionym na wykładzie. Macierz P ma, jak widać, wyznacznik 1, a zatem mnożenie przez P^T czy P nie zmienia wyznacznika. Co ta macierz robi? Otóż macierz $P^T G(f;\mathcal{B})$ to macierz, w której od ostatniego wiersza macierzy $G(f;\mathcal{B})$ odjęliśmy kolejno:

- a_1 razy wiersz pierwszy,
- a_2 razy wiersz drugi,
- $\ldots$
- a_{n-1} razy wiersz $n-1$-wszy.

Analogiczne działania na kolumnach macierzy $G(f;\mathcal{B})$ wykonane są przez mnożenie $G(f;\mathcal{B})P$. Ta operacja wymienia układ $\beta_1, \ldots, \beta_{n-1}, \beta_n$ na taki układ $\beta_1, \ldots, \beta_{n-1}, \beta_n'$, że β_n' jest prostopadły do wektorów $\alpha_1, \ldots, \alpha_{n-1}$. Czy Czytelnik widzi, że ma to związek z ortogonalizacją?

Wymiana bazy w macierzy Grama A pewnego iloczynu skalarnego polegająca na wzięciu iloczynu $P^T A P$ to niezwykle istotny zabieg algebraiczny, który zobaczymy w teorii fukcjonałów dwuliniowych symetrycznych (to są funkcje takie, jak wyjściowa funkcja f, bez własności dodatniej określoności). Pozwoli on na wprowadzenie drugiej, obok podobieństwa, kluczowej relacji równoważności na zbiorze macierzy kwadratowych (nie tylko nad $\mathbb{R}$) – relacji kongruencji.

Dodatek. Macierze $A^T A$ i metoda najmniejszych kwadratów

W ramach tego wykładu kilkukrotnie korzystamy istotnie z rozmaitych własności iloczynu $A^T A$. Powiemy o jednym z najważniejszych jego zastosowań, czyli metodzie najmniejszych kwadratów.

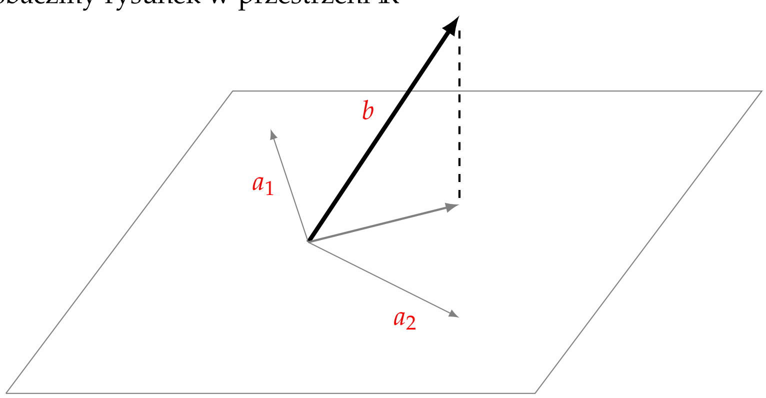
Jest to oczywiście opowiadanie o najprostszej jej formie – jest to rozbudowana teoria, której elementy poznają Państwo m.in. na matematyce obliczeniowej.

Ten dodatek napisany jest na podstawie znakomitej książki Gilberta Stranga „Linear algebra and its applications".

Zacznijmy od układów równań. Jeśli $A \in M_{m\times n}(R)$ oraz $b \in \mathbb{R}^m$, to układ $Ax = b$ może nie mieć rozwiązania $x \in \mathbb{R}^n$. Jeśli jednak rozważamy strukturę przestrzeni euklidesowej $(\mathbb{R}^n, \langle\,,\rangle)$, to możemy zastanawiać się czy jest jakiś wektor, który byłby „najbliżej" rozwiązania. Rozważmy przykład takiego równania:

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \iff x_1 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_2 \cdot \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Nietrudno widzieć, że równanie to nie ma rozwiązań, bo wektor $(4,5,6)$ nie jest kombinacją liniową wektorów $a_1 = (1,1,0), a_2 = (2,3,0)$. Zobaczmy rysunek w przestrzeni $\mathbb{R}^3$



Zaznaczona „płaszczyzna" to przestrzeń rozpięta przez kolumny a_1, a_2 macierzy naszego układu. Każdy wektor na tej płaszczyźnie jest postaci $x_1 a_1 + x_2 a_2$, dla pewnych $x_1, x_2 \in \mathbb{R}$. To są właśnie wektory Ax. Linią przerywaną oznaczony jest wektor opisujący $Ax - b$. Niestety może on nigdy nie być zerowy. Ale skoro jesteśmy w przestrzeni euklidesowej, to możemy próbować minimalizować wyrażenie $\|Ax - b\|$.

Problem ten przypomina Fakt 70, który mówił, że jeśli W jest podprzestrzenią przestrzeni euklidesowej V oraz v' jest rzutem prostopadłym wektora $v \in V$ na W, wówczas $\|v - v'\| = \min_{w\in W} \|v - w\|$.

W naszym przypadku podprzestrzeń $W = \{Ax, x \in \mathbb{R}^3\}$ jest przestrzenią kolumnową macierzy A. Zatem wektor $\hat{x}$, który minimalizuje $\|Ax - b\|$ jest rzutem wektora b na podprzestrzeń kolumnową A.

Jak wyznaczyć ten rzut? Popatrzmy na tzw. wektor błędu $b - A\hat{x}$.

Wektor ten musi być prostopadły do każdej kolumny $a_1, \ldots, a_n$ macierzy A, co możemy zapisać schematycznie w postaci:

$$\begin{cases} \langle a_1^T, b - A\hat{x} \rangle & = 0 \\ \langle a_2^T, b - A\hat{x} \rangle & = 0 \\ & \vdots \\ \langle a_n^T, b - A\hat{x} \rangle & = 0 \end{cases},$$

co w przypadku standardowego iloczynu skalarnego przepisać można do postaci:

$$A^T(b - A\hat{x}) = 0.$$

Innymi słowy, wektor $\hat{x}$ jest w tym przypadku rozwiązaniem tzw. równania normalnego:

$$A^T A\hat{x} = A^T b,$$

powstającego przez przemnożenie wyjściowego równania przez A^T. Zachodzi następujące twierdzenie.

Fakt 80. *Jeśli równanie $Ax = b$ nie ma rozwiązań w przestrzeni euklidesowej $(\mathbb{R}^n, \langle\,,\rangle_{st})$, to $\|Ax - b\|$ jest minimalna dla $x = \hat{x}$, będącego rozwiązaniem równania:*

$$A^T A\hat{x} = A^T b.$$

Macierz $A^T A$ jest odwracalna wtedy i tylko wtedy, gdy kolumny A są liniowo niezależne. Wówczas:

$$\hat{x} = (A^T A)^{-1} A^T b.$$

Zachęcam Czytelnika do dowodu tego faktu. Proszę przy okazji pokazać, że jednorodne układy równań liniowych zadane macierzami A oraz $A^T A$ mają ten sam zbiór rozwiązań.

Przejdźmy na koniec do naszego przykładu. Rzut prostopadły (standardowy) wektora $b = (4,5,6)$ na przestrzeń kolumnową macierzy o kolumnach $(1,1,0)^T$ oraz $(2,3,0)^T$ to po prostu rzut prostopadły na przestrzeń $\mathrm{lin}((1,0,0),(0,1,0))$, i jest to oczywiście $(4,5,0)$. Potwierdza się to przy wyliczeniach proponowanych przez fakt wyżej:

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

$$A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}.$$

Co to wszystko ma wspólnego z zastosowaniami matematyki? Załóżmy, że wykonaliśmy serię eksperymentów i oczekujemy, że wynik b będzie funkcją liniową argumentu t. Szukamy zatem prostej o równaniu $C + Dt = b$. Przykłady:

- Wyznaczamy odległość satelity od Marsa. W tym przypadku t jest zmienną czasową, a b – dystansem od planety. O ile nie wyłączy się silnika i o ile grawitacja nie będzie bardzo silna, satelita powinien poruszać się z mniej więcej stałą prędkością v, czyli $b = b_0 + vt$.
- W teorii elastyczności badamy rozszerzenie b struny obciążonej określoną masą t. O ile obciążenie nie jest nadmierne, rozważa się liniową zależność $b = C + Dt$.
- Koszt produkcji t książek jest niemal liniową funkcją $b = C + Dt$, gdzie C jest kosztem przygotowania, a D jest kosztem każdej dodatkowej książki

Jak wyznaczyć C i D? Jeśli nie ma błędu, wówczas każde dwa pomiary b powinny określać prostą $b = C + Dt$. Ale jeśli jest błąd, wówczas musimy być gotowi na znajdowanie „optymalnej" prostej przybliżającej te rozwiązania. Wykonujemy te pomiary dostając układ równań:

$$\begin{cases} C + Dt_1 & = b_1 \\ C + Dt_2 & = b_2 \\ & \vdots \\ C + Dt_m & = b_m. \end{cases}.$$

Ten układ zawiera nadmiarowe informacje: mamy m równań i tylko dwie niewiadome C, D. Jeśli wystąpiły błędy, układ ten nie ma rozwiązania. Jeśli wystąpiły, to najlepsze rozwiązanie $\hat{x} = (\hat{C}, \hat{D})^T$, przybliżające błąd (w normie $\|, \|_{st}$) ma minimalizować wyrażenie:

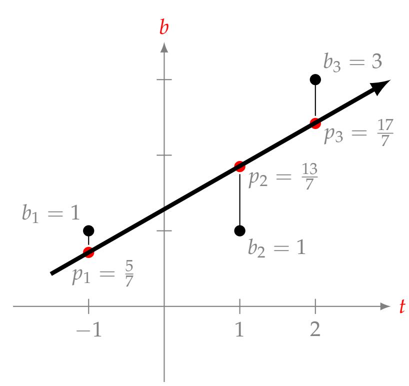
$$\|Ax - b\| = (b_1 - C - Dt_1)^2 + \ldots + (b_m - C - Dt_m)^2.$$

Wektor $p = (p_1, p_2, p_3) = A\hat{x}$ jest „tak blisko" $b = (b_1, b_2, b_3)$, jak to możliwe, tzn. suma kwadratów różnic $p_i - b_i$ jest minimalna.

Dla przykładu: biorąc $b_1 = 1, t_1 = 1$, $b_2 = 1, t_2 = 1, b_3 = 3, t_3 = 2$ widzimy, że punkty (b_i, t_i) nie leżą na jednej prostej. Istnieje jednak prosta $\hat{C} + \hat{D}t = \frac{9}{7} + \frac{4}{7}t$, która „najlepiej" przybliża rozwiązania układu

$$\begin{cases} C - D & = 1 \\ C + D & = 1 \\ C + 2D & = 3 \end{cases}.$$

Rysunek obok przedstawia uzyskaną prostą i „punkty pomiarowe".



Trivia. Kluby Nieparzystowa

Miasteczko Nieparzystów liczy sobie n mieszkańców. Ich głównym zajęciem jest formowanie rozmaitych klubów, co grozi wielkim politycznym chaosem. Rada Nieparzystowa rozporządziła jednak, aby każdy klub miał nieparzystą liczbę członków. Co więcej każde dwa kluby muszą mieć parzystą liczbę wspólnych członków. Pokazać, że nie jest możliwe stworzenie więcej klubów, niż wynosi ludność Nieparzystowa.

Liczba potencjalnych klubów, czyli podzbiorów zbioru n-elementowego wynosi 2^n, a więc więcej niż n. Ile wśród nich jest klubów o nieparzystej liczbie członków? Wydaje się, że również możliwe jest, żeby było ich więcej niż n. Gdy $n = 3$ mamy możliwe trzy kluby o jednym członku i jeden klub o trzech członkach, a więc więcej niż 3. Łatwo jednak sprawdzić, że konfiguracja taka nie może wystąpić w przypadku, gdy każde dwa kluby muszą mieć parzystą liczbę wspólnych członków. Jak to pokazać dla każdego n?

Dowód. Ponumerujmy mieszkańców liczbami $1, 2, \ldots, n$, a kluby nazwijmy $C_1, \ldots, C_m$. Definiujemy macierz $A \in M_{n\times m}(\mathbb{R})$, w której kolumnach stoją „listy przynależności" do klubów. Dokładniej $A = (a_{ij})$:

$$a_{ij} = \begin{cases} 1, & \text{jeśli mieszkaniec } i \text{ należy do } C_j \\ 0, & \text{jeśli mieszkaniec } i \text{ nie należy do } C_j. \end{cases}$$

Przyjrzyjmy się teraz iloczynowi $A^T A$. Jest to macierz rozmiarów $m \times m$, której wyraz w wierszu i-tym oraz w kolumnie j-tej dostajemy ze wzoru: $\sum\limits_{k=1}^{n} a_{ki}a_{kj}$. Co ta liczba oznacza?

Zauważmy, że i-ty wiersz macierzy A^T liczy sobie dokładnie n wyrazów i wyraz a_{ki} koduje przynależność k-tego mieszkańca do klubu C_i. Wyraz ten wynosi 1, jeśli k-ty mieszkaniec należy do klubu C_i oraz 0, jeśli nie należy. Mnożenie i-tego wiersza A^T oraz j-tego wiersza A to po prostu standardowy iloczyn skalarny „list przynależności" do klubów C_i i C_j. Proszę zauważyć, że iloczyn ten ma składniki równe 0 lub 1. Składnik $a_{ki}a_{kj}$ jest równy 1 wtedy i tylko wtedy gdy k-mieszkaniec jest członkiem zarówno klubu C_i jak i klubu C_j. A zatem wyraz macierzy $A^T A$ stojący w i-tym wierszu i j-tej kolumnie będący sumą takich $a_{ki}a_{kj}$ opisuje dokładnie liczbę wspólnych członków klubu i-tego i j-tego! To jest dokładnie macierz Grama układu wektorów opisujących „listy przynależności" do klubów od 1 do m. Zauważmy jeszcze, że można mówić o „prostopadłości" wektorów z listami klubów. Oznacza ona po prostu, że kluby nie mają wspólnych członków.

Co to oznacza dla naszego problemu? Prawo w Nieparzystowie implikuje, że na przekątnej naszej macierzy Grama $A^T A$ stoi nieparzysta liczba członków i-tego klubu (bo i-ty klub ma wszystkich członków wspólnych z i-tym klubem), a wszędzie indziej stoją liczby parzyste. Nietrudno pokazać, że macierz $A^T A$ rozmiaru $m \times m$ jest pełnego rzędu m. Wynika to stąd, że wyznacznik macierzy Grama jest liczbą całkowitą (wszystkie wyrazy są całkowite) zaś wyznacznik macierzy o współczynnikach całkowitych ma resztę z dzielenia przez 2 taką samą jak wyznacznik macierzy powstałej przez podmiany pierwotnych wyrazów przez ich reszty modulo 2.

Można to pokazać na wiele sposobów. Natychmiast wynika to na przykład z wzoru permutacyjnego na wyznacznik. Np. dla zbadania parzystości wyznacznika macierzy całkowitoliczbowej:

$$\begin{vmatrix} 102495 & 550428 & 873298 & 660694 \\ 370628 & 909093 & 127450 & 925600 \\ 835044 & 601178 & 624653 & 263392 \\ 663780 & 487252 & 292276 & 593107 \end{vmatrix}$$

wystarczy wziąć każdy ze współczynników modulo 2 i dostajemy macierz I_4, której wyznacznik jest nieparzysty.

W przypadku naszej „klubowej" macierzy Grama. Po wzięciu jej współczynników modulo 2 dostajemy I_m, a więc macierz $m \times m$ o niezerowym wyznaczniku 1. Zatem $r(A^T A) = m$. Na wykładzie przypomnieliśmy nierówność

$$m = r(A^T A) \leq r(A) = r(A^T) \leq n.$$

Klubów jest zatem nie więcej, niż mieszkańców. $\square$

Podobnych zadań wykorzystujących „wektory przynależności" (lub wektory incydencji) jest znacznie więcej. Poniżej wypisałem kilka. Może Czytelnik zechce się z nimi zmierzyć?

Zadanie 1. Studenci wychodzą na lody w grupach co najmniej dwuosobowych. Po tym jak $k > 1$ grup poszło na lody okazało się, że każdych dwóch studentów było razem na lodach co najmniej raz. Wykaż, że wszystkich studentów jest nie więcej niż k.

Zadanie 2. W Różnicowie mieszka n mieszkańców. Stworzyli oni n parami różnych klubów $A_1, \ldots, A_n$ (czyli podzbiorów zbioru mieszkańców). Pokaż, że istnieje w Różnicowie mieszkaniec x taki, że zbiory $A_1 \setminus \{x\}, \ldots, A_n \setminus \{x\}$ są również parami różne.

Zadanie 3. Uczestnicy Turnieju Matematycznego w miejscowości Bezpunktów rozwiązywali n zadań. Każde zadanie było warte pewną całkowitą dodatnią liczbę punktów, określaną przez Jury. Uczestnik dostawał 0 punktów za złe rozwiązanie albo wszystkie możliwe punkty za poprawne rozwiązanie. Jury zorientowało się, że w zależności od tego jak dobierze punktację za poszczególne zadania, uzyska wszystkie możliwe ścisłe listy rankingowe uczestników (czyli takie, gdzie dwóch uczestników nie ma takiego samego łącznego wyniku). Znajdź największą możliwą liczbę uczestników tych zawodów.