Fizyka dla firm — Matematyka

29. Iloczyny skalarne, normy i metryki

P. F. Góra

https://zfs.fais.uj.edu.pl/pawel_gora

5 kwietnia 2023

Iloczyn skalarny ☺

Niech $\mathcal{V}$ będzie przestrzenią liniową nad ciałem liczb zespolonych, $\mathbb{C}$. Iloczynem skalarnym nazywam odwzorowanie $\mathcal{V} \times \mathcal{V} \to \mathbb{C}$ spełniające ($\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$, $\alpha, \beta \in \mathbb{C}$, $(\cdot)^*$ oznacza sprzężenie zespolone):

$$\mathbf{x} \circ \mathbf{y} = (\mathbf{y} \circ \mathbf{x})^* \tag{1a}$$

$$\mathbf{x} \circ (\alpha\mathbf{y} + \beta\mathbf{z}) = \alpha\, \mathbf{x} \circ \mathbf{y} + \beta\, \mathbf{x} \circ \mathbf{z} \tag{1b}$$

$$(\alpha\mathbf{x} + \beta\mathbf{y}) \circ \mathbf{z} = \alpha^*\, \mathbf{x} \circ \mathbf{z} + \beta^*\, \mathbf{y} \circ \mathbf{z} \tag{1c}$$

$$\mathbf{x} \circ \mathbf{x} \geqslant 0\,; \quad \mathbf{x} \circ \mathbf{x} = 0 \Leftrightarrow \mathbf{x} = 0 \tag{1d}$$

Ostatni warunek wymaga pewnego komentarza: W ogólności iloczyn skalarny dwóch wektorów zespolonych jest zespolony, ale *kwadrat skalarny* jednego wektora jest rzeczywisty i nieujemny. Tylko kwadrat skalarny wektora zerowego równa się zero.

Przestrzeń $\mathbb{R}^n$

Przestrzeń $\mathbb{R}^n$ jest szczególnym przypadkiem przestrzeni $\mathcal{V}$, o jakiej mowa była na poprzednim slajdzie. Jest to przestrzeń nad ciałem liczb rzeczywistych, więc wektory i skalary są rzeczywiste, a sprzężenie zespolone jest operacją identycznościową. W tej przestrzeni iloczynem skalarnym jest

$$\mathbf{x} \circ \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i \tag{2}$$

Innym, ale równie dobrze zdefiniowanym iloczynem skalarnym jest

$$\mathbf{x} \circ \mathbf{y} = \mathbf{x}^T \mathbf{A} \mathbf{y} \tag{3}$$

gdzie $\mathbf{A}$ jest macierzą symetryczną, rzeczywistą i *dodatnio określoną*: $\forall \mathbf{x} \neq 0 : \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$.

Przestrzeń $\mathbb{C}^n$

Przestrzeń $\mathbb{C}^n$ jest innym przykładem przestrzeni $\mathcal{V}$. W tej przestrzeni iloczynem skalarnym jest

$$\mathbf{x} \circ \mathbf{y} = \mathbf{x}^\dagger \mathbf{y} = \sum_{i=1}^{n} x_i^* y_i \,. \tag{4}$$

Innym, ale również dobrze zdefiniowanym iloczynem skalarnym jest

$$\mathbf{x} \circ \mathbf{y} = \mathbf{x}^\dagger \mathbf{A} \mathbf{y} \tag{5}$$

gdzie $\mathbf{A}$ jest macierzą hermitowską, $\mathbf{A}^\dagger = \mathbf{A}$, i *dodatnio określoną*: $\forall \mathbf{x} \neq 0 : \mathbf{x}^\dagger \mathbf{A} \mathbf{x} > 0$.

Ponieważ istnieje nieskończenie wiele różnych macierzy hermitowskich $\mathbb{C}^{n\times n}$ dodatnio określonych, w $\mathbb{C}^n$ można, formalnie rzecz biorąc, zdefiniować nieskończenie wiele iloczynów skalarnych. *Niektóre* z nich są użyteczne.

Przestrzeń unormowana

Niech $\mathcal{V}$ będzie przestrzenią liniową nad ciałem liczb zespolonych, $\mathbb{C}$. Normą wektora nazywam odwzorowanie $\mathcal{V} \to \mathbb{R}$ spełniające
Nieujemność:

$$\forall \mathbf{x} \in \mathcal{V} : \|\mathbf{x}\| \geqslant 0 \tag{6a}$$

Norma wektorów niezerowych jest dodatnia:

$$\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0 \tag{6b}$$

Mnożenie przez skalar:

$$\forall \mathbf{x} \in \mathcal{V}\,,\ \alpha \in \mathbb{C} : \|\alpha \cdot \mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\| \tag{6c}$$

Nierówność trójkąta:

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{V} : \|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\| \tag{6d}$$

Przestrzeń liniową (wektorową), na której określono normę, nazywam przestrzenią unormowaną.

Norma zadana przez iloczyn skalarny

Twierdzenie: Jeżeli $\mathbf{x} \circ \mathbf{y}$ jest iloczynem skalarnym, wówczas

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \circ \mathbf{x}} \tag{7}$$

jest normą. Normę tę nazywam normą zadaną przez iloczyn skalarny.

Przykład 1

W $\mathbb{R}^n$ normę zadaje iloczyn skalarny (2)

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{8}$$

Normę tę nazywam normą euklidesową. Jest to najczęściej używana norma, wobec czego zapis $\|\cdot\|$ często oznacza właśnie normę euklidesową, choć *powinno się* pisać $\|\cdot\|_2$.

Przykład 2

Uogólnieniem normy euklidesowej na przestrzeń $\mathbb{C}^n$ jest

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{9}$$

Przykład 3

Istnieją normy niezadawane przez iloczyn skalarny. Na przykład w $\mathbb{C}^n$ (i odpowiednio w $\mathbb{R}^n$) istnieją normy

Norma Manhattan albo taksówkowa:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \tag{10a}$$

czy też norma maksimum lub *worst offender*:

$$\|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i| \tag{10b}$$

W przypadku tych norm na ogół *nie* unika się pisania znaczników dolnych, $\|\cdot\|_1$, $\|\cdot\|_\infty$.

Przestrzeń metryczna

Przestrzenią metryczną nazywam zbiór $\mathcal{V}$ wraz z określoną na nim metryką, czyli funkcją $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$, spełniającą

$$d(A, A) = 0 \tag{11a}$$

$$A \neq B \Rightarrow d(A, B) > 0 \tag{11b}$$

$$d(A, B) = d(B, A) \tag{11c}$$

$$d(A, B) \leqslant d(A, C) + d(C, B) \tag{11d}$$

Jeżeli $\mathcal{V}$ jest przestrzenią liniową z zadaną normą, wówczas

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \tag{12}$$

jest metryką. O metryce takiej mówię, że jest zadana przez normę.

Przykład 4

Norma euklidesowa w $\mathbb{R}^n$ zadaje metrykę euklidesową:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{13}$$

Przykład 5

Norma Manhattan zadaje metrykę Manhattan, norma maksimum zadaje metrykę maksimum (zwaną niekiedy w tym kontekście metryką Czebyszewa).

Istnieją metryki, które **nie** są zadawane przez normę, na przykład metryka dyskretna:

$$d(A,B)=\begin{cases}0 & A=B\\ 1 & A\neq B\end{cases} \tag{14}$$

Kula otwarta

Niech $\mathcal{V}$ będzie przestrzenią metryczną z metryką d. *Kulą otwartą* o środku w punkcie P i promieniu r nazywam zbiór punktów

$$\{X \in \mathcal{V} : d(P, X) < r\} \tag{15}$$

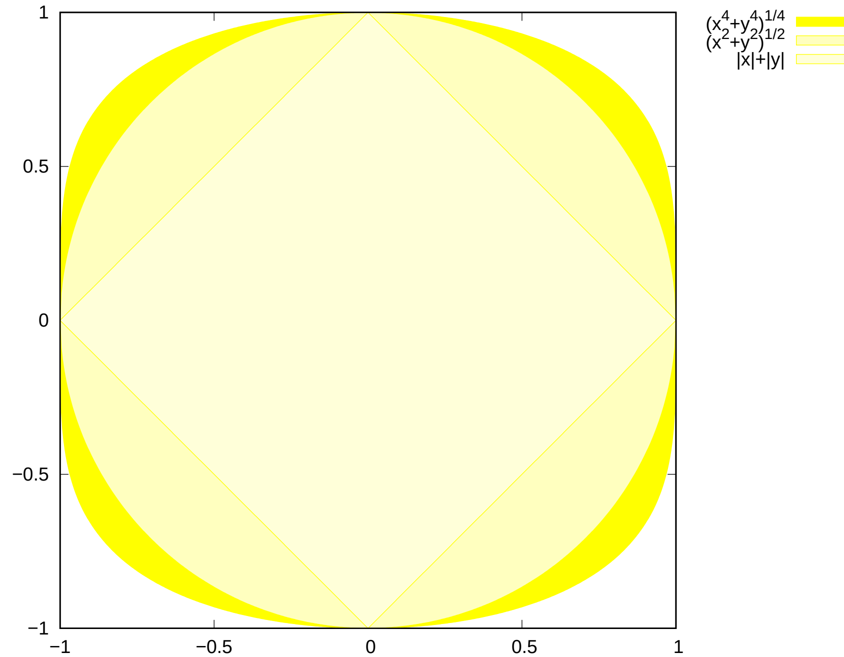
Przykład 6

W przestrzeni $\mathbb{R}^2$ kulą otwartą o środku w punkcie $P(x_0, y_0)$ i promieniu r jest wnętrze okręgu

$$(x - x_0)^2 + (y - y_0)^2 = r \tag{16}$$

W tej samej przestrzeni z metryką Manhattan kulą otwartą są punkty spełniające

$$|x - x_0| + |y - y_0| < r \tag{17}$$



Kule jednostkowe na płaszczyźnie w różnych metrykach: Manhattan, euklidesowej, $\sqrt[4]{x^4 + y^4}$. *Cały* kwadrat jednostkowy jest kulą w normie maksimum.

Norma indukowana macierzy

Niech $\mathbf{A} \in \mathbb{R}^{n\times n}$ lub $\mathbf{A} \in \mathbb{C}^{n\times n}$. Macierze takie tworzą przestrzeń liniową nad $\mathbb{R}$ lub, odpowiednio, $\mathbb{C}$ (macierze są tam “wektorami niegeometrycznymi”).

Niech $\|\cdot\|$ będzie normą wektorów w $\mathbb{R}^n$ lub, odpowiednio, $\mathbb{C}^n$ — najczęściej, choć nie wyłącznie, mamy na myśli normę euklidesową. Poniżej będę pomijał zapis “lub $\mathbb{C}^n$”, gdyż uogólnienie na przypadek zespolony jest oczywiste.

Definicja: Normą macierzy indukowaną przez normę wektorów nazywam wielkość

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| \tag{18}$$

Po prawej stronie (18) stoją normy wektorów, a więc wielkości znane. Ponieważ $\forall\, \mathbf{x} \neq 0 \;\; \exists \mathbf{e} : \mathbf{x} = \|\mathbf{x}\| \cdot \mathbf{e} \;\wedge\; \|\mathbf{e}\| = 1$

$$\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|\mathbf{x}\| \cdot \|\mathbf{A}\mathbf{e}\|}{\|\mathbf{x}\| \cdot \|\mathbf{e}\|} = \|\mathbf{A}\mathbf{e}\| \tag{19}$$

i druga równość w (18) staje się oczywista.

Można pokazać, że wyrażenie (18) jest normą w $\mathbb{R}^{n\times n}$.

Norma indukowana macierzy mówi, jak bardzo dana macierz może rozciągnąć dowolny wektor jednostkowy. Ponieważ każdy wektor daje się przedstawić jako kombinacja liniowa wektorów jednostkowych, norma indukowana jest miarą największychmożliwych rozciągnięć wektorów poddanych działaniu tej macierzy.

Jeżeli w definicji normy indukowanej użyjemy innej normy wektorów niż norma euklidesowa, $\|\cdot\|_2$, dostaniemy inną normę indukowaną, ale jej sens pozostanie taki sam: będzie to miara największych możliwych rozciągnięć wektora przez daną macierz, tyle tylko, że "rozciągnięcia" mierzy się wedle innej normy ☺

Można też definiować inne normy w tej przestrzeni, nieindukowane przez normy wektorów, ale mają one mniejsze znaczenie.

Przykład 7

Niech $\mathbf{A} \in \mathbb{R}^{n\times n}$ będzie macierzą symetryczną, rzeczywistą. Obliczmy jej normę indukowaną przez normę euklidesową wektorów.

Niech $\{\mathbf{e}_i\}_{i=1}^{n}$ będą unormowanymi wektorami własnymi macierzy $\mathbf{A}$, a $\{\lambda_i\}$ odpowiadającymi im wartościami własnymi. Wiadomo, że wektory $\{\mathbf{e}_i\}$ stanowią bazę ortonormalną w $\mathbb{R}^n$. Zatem $\forall\, \mathbf{x} \in \mathbb{R}^n$ istnieje rozkład

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i \tag{20a}$$

$$\begin{aligned} \|\mathbf{x}\|_2^2 &= \mathbf{x}^T\mathbf{x} = \left(\sum_{i=1}^{n} \alpha_i \mathbf{e}_i^T\right)\left(\sum_{j=1}^{n} \alpha_j \mathbf{e}_j\right) \\ &= \sum_{i,j=1}^{n} \alpha_i\alpha_j\, \mathbf{e}_i^T\mathbf{e}_j = \sum_{i,j=1}^{n} \alpha_i\alpha_j\, \delta_{ij} = \sum_{i=1}^{n} \alpha_i^2 \end{aligned} \tag{20b}$$

A zatem $\|\mathbf{x}\|_2 = 1 \Leftrightarrow \|\mathbf{x}\|_2^2 = 1 \Leftrightarrow \sum\limits_{i=1}^{n} \alpha_i^2 = 1$ gdzie α_i są współczynnikami rozkładu (20a).

Niech $\|\mathbf{x}\|_2 = 1$. Obliczam

$$
\begin{aligned}
\|\mathbf{Ax}\| &= \left\| \mathbf{A} \sum_{i=1}^{n} \alpha_i \mathbf{e}_i \right\|_2 \\
&= \left\| \sum_{i=1}^{n} \alpha_i \mathbf{A} \mathbf{e}_i \right\|_2 = \left\| \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{e}_i \right\|_2 \\
&= \sqrt{\sum_{i=1}^{n} \alpha_i^2 \lambda_i^2} \leqslant \sqrt{\sum_{i=1}^{n} \alpha_i^2 \max_{j=1,\dots,n} \lambda_j^2} \\
&= \sqrt{\left(\max_{j=1,\dots,n} \lambda_j^2\right) \cdot \underbrace{\left(\sum_{i=1}^{n} \alpha_i^2\right)}_{=1}} = \max_{j=1,\dots,n} |\lambda_j| \qquad \text{(20c)}
\end{aligned}
$$

A więc

$$\|\mathbf{A}\| = \max_{j=1,\dots,n} |\lambda_j| \tag{20d}$$

Ten wynik można bardzo łatwo uogólnić na przypadek macierzy hermitowskich.

Normą macierzy rzeczywistej, symetrycznej lub hermitowskiej jest największy moduł jej wartości własnych.

Przykład 8

Wartościami własnymi macierzy

$$\mathbf{B} = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix} \tag{21}$$

są liczby $\lambda_1 = 6$ i $\lambda_2 = 3$ (podwójnie zdegenerowana). Wobec tego norma macierzy $\mathbf{B}$ indukowana przez normę euklidesową w $\mathbb{R}^3$ wynosi $\|\mathbf{B}\| = 6$.

Przykład 9

Znajdźmy normę macierzy

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 2 \end{bmatrix} \tag{22a}$$

indukowaną przez zwykły, Euklidesowy iloczyn skalarny w $\mathbb{R}^3$.

Ponieważ funkcja kwadratowa dla argumentów dodatnich jest ściśle rosnąca, zamiast badać normę, możemy badać kwadrat normy. Niech $\mathbf{x} = [a, b, c]$. Warunek $\|\mathbf{x}\|^2 = 1$ oznacza, że

$$a^2 + b^2 + c^2 = 1 \tag{22b}$$

Mamy

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2a+b+c \\ -a+2b+c \\ -a-b+2c \end{bmatrix} \tag{22c}$$

$$\begin{aligned} \mathcal{N}^2 = \|\mathbf{A}\mathbf{x}\|^2 &= (2a+b+c)^2 + (-a+2b+c)^2 + (-a-b+2c)^2 \\ &= 6(a^2+b^2+c^2) + 2(ab-ac+bc) \end{aligned} \tag{22d}$$

Zadanie polega na znalezieniu maksimum wyrażenia (22d) pod warunkiem (22b). Poszukiwanie normy indukowanej jest zadaniem na poszukiwanie ekstremów warunkowych funkcji wielu zmiennych ☺.

W tym celu tworzę funkcję

$$\mathcal{F} = \mathcal{N}^2 - \lambda(a^2+b^2+c^2-1) \tag{22e}$$

obliczam pochodne $\frac{\partial \mathcal{F}}{\partial a}, \frac{\partial \mathcal{F}}{\partial b}, \frac{\partial \mathcal{F}}{\partial c}$ i przyrównuję je do zera, otrzymując układ równań

$$
\begin{aligned}
(6-\lambda)a + b - c &= 0 \\
a + (6-\lambda)b + c &= 0 \\
-a + b + (6-\lambda)c &= 0
\end{aligned}
\tag{22f}
$$

Układ równań (22f) jest jednorodnym układem równań liniowych, a więc jego rozwiązaniem jest $(a, b, c) = (0, 0, 0)$, co nie spełnia warunku (22b). Układ ten może mieć niezerowe rozwiązanie wtedy i tylko wtedy, gdy **wyznacznik główny tego układu równa się zero**. Oznaczając $6 - \lambda = z$ otrzymuję

$$
\det \begin{bmatrix} z & 1 & -1 \\ 1 & z & 1 \\ -1 & 1 & z \end{bmatrix} = z^3 - 3z - 2 = (z+1)^2(z-2) = 0 \tag{22g}
$$

W przypadku $z = 2$, z układu (22f) otrzymujemy $a = c = -b$, a z warunku (22b) $b = \pm 1/\sqrt{3}$. Wówczas

$$\mathcal{N}^2 = 6 \cdot 1 + 2\left(-\frac{1}{3} - \frac{1}{3} - \frac{1}{3}\right) = 6 - 2 = 4 \tag{22h}$$

W przypadku $z = -1$ z układu (22f) otrzymujemy $a = b - c$. Po podstawieniu do (22b) dostajemy

$$b^2 - bc + c^2 = \frac{1}{2} \tag{22i}$$

natomiast

$$\mathcal{N}^2 = 6 \cdot 1 + 2((b-c)b - (b-c)c + bc) = 2 + (b^2 - bc + c^2) = 6 + 2 \cdot \frac{1}{2} = 7 \tag{22j}$$

Ostatecznie

$$\|\mathbf{A}\| = \sqrt{\mathcal{N}^2} = \sqrt{7}\,. \tag{22k}$$