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 $y\in\mathcal{Y}$*

$$y\neq x_{\mathcal{Y}} \implies \|x-x_{\mathcal{Y}}\| < \|x-y\|.$$

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.