Akademia Górniczo-Hutnicza
im. Stanisława Staszica
w Krakowie

Praca licencjacka

Działanie Maszyny Wektorów Nośnych w przestrzeniach Kreina

Kamil Bartocha

Kierunek: Matematyka

Nr albumu: 296032

Promotor
dr hab. Sergiusz Kużel

Wydział Matematyki Stosowanej

Kraków 2020

Oświadczenie studenta

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2018 r. poz. 1191 z późn. zm.): Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystyczne wykonanie albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.", a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 307 ust. 1 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 z późn. zm.) Student podlega odpowiedzialności dyscyplinarnej za naruszenie przepisów obowiązujących w uczelni oraz za czyn uchybiający godności studenta.", oświadczam, że niniejszą pracę dyplomową wykonałem osobiście i samodzielnie i nie korzystałem ze źródeł innych niż wymienione w pracy. Jednocześnie Uczelnia informuje, że zgodnie z art. 15a ww. ustawy o prawie autorskim i prawach pokrewnych Uczelni przysługuje pierwszeństwo w opublikowaniu pracy dyplomowej studenta. Jeżeli Uczelnia nie opublikowała pracy dyplomowej w terminie 6 miesięcy od dnia jej obrony, autor może ją opublikować, chyba że praca jest częścią utworu zbiorowego. Ponadto Uczelnia jako podmiot, o którym mowa w art. 7 ust. 1 pkt 1 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 z późn. zm.), może korzystać bez wynagrodzenia i bez konieczności uzyskania zgody autora z utworu stworzonego przez studenta w wyniku wykonywania obowiązków związanych z odbywaniem studiów, udostępniać utwór ministrowi właściwemu do spraw szkolnictwa wyższego i nauki oraz korzystać z utworów znajdujących się w prowadzonych przez niego bazach danych, w celu sprawdzania z wykorzystaniem systemu antyplagiatowego. Minister właściwy do spraw szkolnictwa wyższego i nauki może korzystać z prac dyplomowych znajdujących się w prowadzonych przez niego bazach danych w zakresie niezbędnym do zapewnienia prawidłowego utrzymania i rozwoju tych baz oraz współpracujących z nimi systemów informatycznych.

. . . .
(Podpis studenta)

Oświadczenie promotora

Oświadczam, że praca spełnia wymogi stawiane pracom licencjackim.

. . . .
(Podpis promotora)

Spis treści

Streszczenie

W pracy omówione zostało działanie maszyny wektorów nośnych w przestrzeniach Kreina. W tym celu przedstawiona jest idea działania MWN w przestrzeniach liniowych i w przestrzeniach Hilberta. Rozdział drugi traktuje o elementach teorii Kreina oraz o macierzach Grama. Następnie przeanalizowano możliwość zdefiniowania MWN jako rzutowania w przestrzeniach Kreina oraz zbadano konsekwencje takiego rozwiązania.

Słowa kluczowe: operator rzutowania, przestrzeń Kreina, nieoznaczony iloczyn skalarny, macierz Grama.

Key words: operator of projection, Krein space, indefinite inner product, Gram matrix.

1 Wstęp

Rozwój technologi sprawił, że jesteśmy w stanie wykorzystać komputery do wykonywania zadań, w których maszyna w porównaniu z człowiekiem będzie nieporównywalnie wydajniejsza. Jednym z takich problemów jest klasyfikowanie dużych zbiorów danych. Rozpoznawanie twarzy lub odręcznego pisma, biotechnologia, kategoryzacja tekstów, to problemy, z którymi świetnie radzi sobie maszyna wektorów nośnych opracowana przez Vladimira Vapnik'a w drugiej połowie lat dziewięćdziesiątych ubiegłego wieku.

Maszyna wektorów nośnych (SVM) to klasyfikator, który dzięki oznaczonym danym treningowym tworzy hiperpłaszczyznę zdolną oddzielać dane. Innymi słowy, bazując na danych treningowych, algorytm generuje optymalną hiperpłaszczyznę, która kategoryzuje nowe przykłady z maksymalnym marginesem separacji, czyli odległością między danymi różnego typu.

W przypadku danych liniowo separowalnych MWN to klasyfikator oddzielający dane na dwie klasy. W dwóch wymiarach przestrzennych jest to linia dzieląca płaszczyznę na dwie części. Jeżeli jednak mamy do czynienia z danymi nieseparowalnymi liniowo to za pomocą metody wektorów nośnych możemy znaleźć hiperpłaszczyznę klasyfikującą obiekty z minimalnym błędem i jednocześnie przebiegającą jak najdalej od typowych skupień dla każdej klasy. Metoda ta pozbawiana jest wad typowych dla uczenia sieci neuronowych tj. zatrzymanie procesu optymalizacji w lokalnym minimum.

Istnieje wiele wydajnych algorytmów MWN działających w przestrzeniach unormowanych, półokreślonych dodatnio, gdzie rozwiązaniem problemu jest znalezienie minimum zadania programowania kwadratowego.

Istnieją jednak dane dla których wymagana półokreśłoność nie może być spełniona, wówczas popularnym podejściem poradzenia sobie z takimi problemem jest ich modyfikacja, co oczywiście negatywnie wpływa na działanie klasyfikatora. Ciekawszym podejściem wydaje się rozwiązanie problemu w przestrzeniach Kreina, gdzie przez brak normy drogą do znalezienia rozwiązania, czyli znalezienia punktu stacjonarnego jest wykorzystanie operatora rzutowania.

Aby lepiej zrozumieć sposób działania MWN zaczniemy od opisu jego najprostszej, liniowej wersji, aby nastęnie zająć się sformułowaniem definicji maszyny wektorów nośnych w przestrzeniach Kreina.

2 Maszyna wektorów nośnych

2.1 Liniowa MWN

Rozważmy zbiór n danych uczących postaci

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n),$$

gdzie $y_i \in \{-1, 1\}$ określa, do której klasy dany punkt należy. Każdy wektor $\mathbf{x}_i$ jest p-wymiarowym wektorem o rzeczywistych współrzędnych. Chcemy znaleźć hiperpłaszczyznę o maksymalnym marginesie oddzielającą punkty $\mathbf{x}_i$ na dwie grupy tak aby zmaksymalizować odległość między hiperpłaszczyzną, a najbliższymi punktami $\mathbf{x}_i$ z obu zbiorów.

Hiperpłaszczyznę możemy zapisać jako zbiór punktów $\mathbf{x}$ spełniających równość

$$\mathbf{w} \cdot \mathbf{x} - b = 0, \tag{2.1}$$

gdzie $\mathbf{w}$ to wektor normalny względem hiperpłaszczyzny.

Aby rozszerzyć działanie MWN do przypadków gdy dane nie są seperowalne liniowo wprowadźmy funkcję

$$\max(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i - b)). \tag{2.2}$$

Funkcja (2.2) przyjmuje wartość 0 jeśli warunek $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geqslant 1$ jest spełniony dla wszystkich $i \in \{1, \ldots, n\}$. Inaczej mówiąc, jeśli $\mathbf{x}_i$ leży po dobrej stronie marginesu. Natomiast dla danych leżących po złej stronie, wartość funkcji jest proporcjonalna do odległości od marginesu. Dzięki temu rozwiązaniu dopuszczamy możliwość by pewne punkty leżały po złej stornie marginesu, jednocześnie poprawnie oddzielając większość danych. W takim przypadku chcemy minimalizować

$$\left[\frac{1}{n} \sum_{i=1}^{n} \max\left(0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i - b)\right)\right] + \lambda \|\mathbf{w}\|^2, \tag{2.3}$$

gdzie λ to parametr określający jak bardzo pozwalamy błędnie zakwalifikować dane.

2.2 MWN w przestrzeniach Hilberta

Sukces maszyny wektorów nośnych w wielu dziedzinach jest możliwy dzięki istnieniu stabilnych i wydajnych algorytmów numerycznych. Zajmiemy się przestrzeniami funkcji ciągłych $\mathfrak{H}$. Niech $\mathcal{X} = \{x_i, \dots, x_l\}$ to l punktów treningowych przestrzenni $\mathcal{X}$, wraz z odpowiadającymi wskaźnikami klasy $y_i \in \{-1, 1\}$ do której dany punkt należy. Rozważamy przestrzeń $\mathcal{X}$ jako $\mathbb{R}^d$. Dla danego parametru regulującego C maszyna wektorów nośnych to minimum z funkcjonału

$$J_C(f, b) = \frac{1}{2}\|f\|_{\mathfrak{H}}^2 + C\sum_{i=1}^{l} \max(0, 1 - y_i(f(x_i) + b)). \tag{2.4}$$

Wykorzystując rozwiązanie $(f_C^*, b_C^*) := \arg\min\ J_C(f, b)$, możemy zdefiniować organicznie $\tau = \sum_{i=1}^{l} \max(0, 1 - y_i(f_C^*(x_i) + b_C^*))$ i związany z nim problem programowania kwadratowego

$$\begin{cases} \min\limits_{f\in\mathfrak{H},\ b\in\mathbb{R}} & \frac{1}{2}\|f\|_{\mathfrak{H}}^2 \\ \text{p.w.} & \sum_{i=1}^{l} \max(0, 1 - y_i(f(x_i) + b)) \leqslant \tau. \end{cases} \tag{2.5}$$

Powyższy problem kwadratowy posiada dokładnie jedno rozwiązanie $(f_\tau^*, b_\tau^*, \lambda_\tau^*)$, gdzie λ_τ^* to mnożnik Lagrange'a dla optymalnego rozwiązania. Okazuje się, że dla odpowiednio dobranego C i τ problemy (2.4) i (2.5) są równoznaczne oraz posiadają to samo, dokładnie jedno rozwiązanie $(f_C^*, b_C^*) = (f_\tau^*, b_\tau^*)$. Faktycznie, funkcjonał J_C jest wypukły, ponieważ jest sumą dwóch wypukłych funkcjonałów. Jeśli przyjmiemy $C = \lambda_\tau^*$ jako jego parametr regulacyjny, może on być interpretowany jako optymalny mnożnik Lagrange'a dla (2.5), dzięki czemu rozwiązanie (2.4) spełnia warunki Karush-Kuhn'a-Tucker'a dla (2.5), jednocześnie rozwiązanie (2.5) jest punktem stacjonarnym, a zatem jest rozwiązaniem (2.4).

Zadanie programowania kwadratowego (2.5) może być też przedstawione jako ortogonalne rzutowanie funkcji zerowej z przestrzeni $\mathfrak{H}$ na wypukły zbiór rozwiązań dopuszczalnych. Zauważmy, że stała b nie jest brana pod uwagę w rzutowaniu, ale pełni ważną rolę dla unikalności rozwiązania, dlatego też dołączamy ją w definicji zbioru wypukłego na który rzutujemy.

Oznaczmy hamiltonian

$$H(f, b) = \sum_{i=1}^{l} \max(0, 1 - y_i(f(x_i) + b)) \tag{2.6}$$

i załóżmy że $0 \in \partial H(f,b)$, gdzie $\partial H(f,b)$ jest subróżniczką funkcji $H(f,b)$ według b.

Rozpatrzmy zbiór

$$\mathcal{S} = \{f \in \mathfrak{H},\ b \in \mathbb{R} \mid H(f,b) \leqslant \tau \ \wedge\ 0 \in \partial_b H(f,b)\}, \tag{2.7}$$

załóżmy że $\mathcal{S}$ jest zbiorem wypukłym i domkniętym w $\mathfrak{H}$ (według zmiennych $f \in \mathfrak{H}$) oraz niech $h \in \mathfrak{H} \wedge h \notin \mathcal{S}$. Wówczas [5] istnieje jednoznacznie określony wektor $f \in \mathcal{S}$ taki, że

$$\parallel f - h \parallel^2 = \inf_{g \in \mathcal{S}} \parallel g - h \parallel^2 . \tag{2.8}$$

W szczególności jeśli $h = 0$, to

$$\parallel f \parallel^2 = \inf_{g \in \mathcal{S}} \parallel g \parallel^2 . \tag{2.9}$$

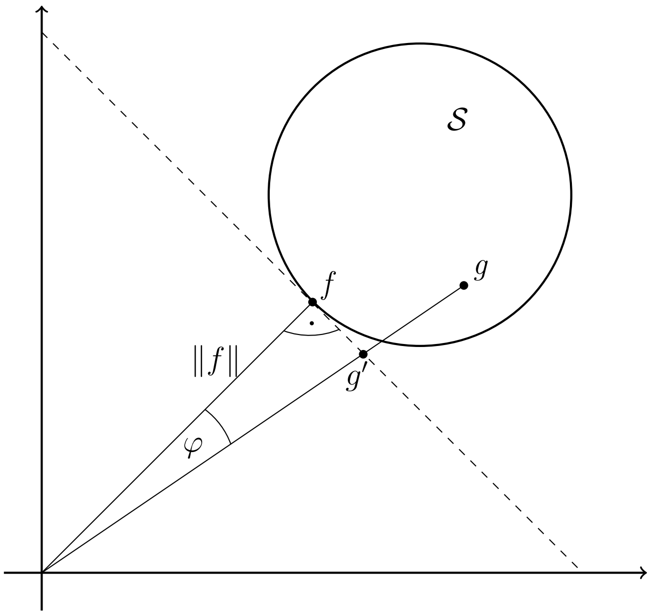
Stąd wynika, że rozwiązanie f zagadnienia (2.5) może być zapisane przez rzutowanie funkcji zerowej h na wypukły zbiór $\mathcal{S}$. Zanim napiszemy definicję MWN jako rzutowanie zwróćmy uwagę, że warunek (2.9) jest równoważny z warunkiem

$$(f, f - g) \leqslant 0 \quad \forall g \in \mathcal{S}. \tag{2.10}$$

Dowód.

$$\begin{aligned}
(f, f - g) &\leqslant 0 \\
(f, f) - (f, g) &\leqslant 0 \\
(f, f) &\leqslant (f, g) \\
\parallel f \parallel^2 &\leqslant (f, g) \\
\parallel f \parallel^2 &\leqslant \parallel f \parallel \cdot \parallel g \parallel \cdot \cos\varphi \qquad \varphi \text{ - kąt między } f \text{ i } g \\
\parallel f \parallel &\leqslant \parallel g \parallel \cdot \cos\varphi \\
\parallel g \parallel &> \parallel g' \parallel \\
\parallel g \parallel \cdot \cos\varphi &\geqslant \parallel g' \parallel \cdot \cos\varphi = \parallel f \parallel
\end{aligned}$$

■



Rys. 1

Definicja 1. *(MWN jako rzutowanie) Dla wypukłego zbioru*

$$\mathcal{S} = \{f \in \mathfrak{H},\ b \in \mathbb{R} \mid H(f,b) \leqslant \tau \ \wedge\ 0 \in \partial_b H(f,b)\},$$

maszyną wektorów nośnych nazywamy jednoznacznie określone $(f,b) \in \mathcal{S}$ *spełniające warunek*

$$\forall (g,a) \in \mathcal{S} \quad (f, f-g)_{\mathfrak{H}} \ \leqslant\ 0.$$

Celem niniejszej pracy jest analiza możliwości zdefiniowania MWN jako rzutowania w przestrzeniach Kreina. Krótko mówiąc, przestudiujemy jakie konsekwencje pojawią się, jeśli zdefiniujemy MWN w przestrzeniach Kreina podobnie do Definicji 1.

3 MWN w przestrzeniach Kreina

3.1 Preliminaria. Elementy teorii przestrzeni Kreina

Przestrzeń ($\mathfrak{H}$, $[\cdot\ ,\ \cdot]$) nazywana jest przestrzenią Kreina, gdy zespolona przestrzeń liniowa $\mathfrak{H}$ z nieoznaczonym iloczynem skalarnym $[\cdot\ ,\ \cdot]$ poddaje się dekompozycji:

$$\mathfrak{H} = \mathfrak{H}_+[\dot{+}]\mathfrak{H}_-, \tag{3.1}$$

która jest ortogonalną sumą prostą dwóch przestrzeni Hilberta ($\mathfrak{H}_+$, $[\cdot\ ,\ \cdot]$) oraz ($\mathfrak{H}_-$, -$[\cdot\ ,\ \cdot]$).

Dekompozycja (3.1) nazywana jest *fundamentalną* i indukuje dodatni iloczyn skalarny

$$(f, g) := [f_+, g_+] - [f_-, g_-], \quad f = f_+ + f_-, \quad g = g_+ + g_-, \quad f_\pm, g_\pm \in \mathfrak{H}$$

W powiązanej przestrzeni Hilberta ($\mathfrak{H}$, $(\cdot\ ,\ \cdot)$) operator

$$\mathcal{P}f = f_+ - f_-, \quad f = f_+ + f_-$$

jest unitarny oraz samosprzężony i jest nazwany *operatorem symetrii fundamentalnej* w przestrzeni Kreina ($\mathfrak{H}$, $[\cdot\ ,\ \cdot]$). Fundamentalna symetria $\mathcal{P}$ pozwala wyrażać $(\cdot\ ,\ \cdot)$ za pomocą metryki nieoznaczonej: $(\cdot\ ,\ \cdot) = [\mathcal{P}\cdot\ ,\ \cdot]$ oraz odwrotnie $[\cdot\ ,\ \cdot\ ] = (\mathcal{P}\cdot\ ,\ \cdot)$.

Niezerowy wektor $f \in \mathfrak{H}$ nazywamy *neutralnym, ujemnym, dodatnim* jeżeli kolejno: $[f, f] = 0, \quad [f, f] < 0, \quad [f, f] > 0$.

Domkniętą podprzestrzeń $\mathfrak{L}$ przestrzeni Hilberta ($\mathfrak{H}$, $(\cdot\ ,\ \cdot)$) nazywamy *neutralną, ujemną, dodatnią*, jeżeli wszystkie niezerowe elementy $f \in \mathfrak{L}$ są neutralne, ujemne lub dodatnie.

Wektory f, g w przestrzeni Kreina ($\mathfrak{H}$, $[\cdot\ ,\ \cdot]$) nazywamy $\mathcal{P}$ *− ortogonalnymi* i oznaczamy przez $f[\perp]g$ jeśli $[f, g] = 0$. Analogicznie, podprzestrzenie $\mathfrak{L}_1, \mathfrak{L}_2$ przestrzeni $\mathfrak{H}$ nazywamy $\mathcal{P}$*−ortogonalnymi* i oznaczamy przez $\mathfrak{L}_1[\perp]\mathfrak{L}_2$ jeśli dla dowolnych $f \in \mathfrak{L}_1$ oraz $g \in \mathfrak{L}_2$ $[f, g] = 0$.

Dopełnienie $\mathcal{P}$ *− ortogonalne* podprzestrzeni $\mathfrak{L}$ z $\mathfrak{H}$ zdefiniowane jest jako

$$\mathfrak{L}^{[\perp]} = \{g \in \mathfrak{H} : [f, g] = 0, \quad \forall f \in \mathfrak{L}\}$$

i jest ono domkniętą podprzestrzenią liniową przestrzeni Hilberta ($\mathfrak{H}$, $(\cdot\ ,\ \cdot)$).

$\mathfrak{L}_0$ nazywamy *izotropową częścią* domkniętej podprzestrzeni $\mathfrak{L}$ jeśli $\mathfrak{L}_0 = \mathfrak{L} \cap \mathfrak{L}^{[\perp]}$, zaś jej niezerowe elementy nazywamy *wektorami izotropowymi*. Istnienie izotropowych wektorów zależy od wyboru podprzestrzeni $\mathfrak{L}$. W szczególności dowolny element neutralnej podprzestrzeni $\mathfrak{L}$ jest izotropowy. Z drugiej strony, nie istnieją izotropowe elementy całej przestrzeni $\mathfrak{H}$. Faktycznie, jeśli $f \in \mathfrak{H}$ to $[f, g] = 0$ dla dowolnego $g \in \mathfrak{H}$. Następnie dla $g = \mathcal{P}f$ otrzymujemy

$$0 = [f, \mathcal{P}f] = [\mathcal{P}f, \mathcal{P}f] = (\mathcal{P}^2 f, f) = (f, f) =\| f \|^2,$$

Wtedy $f = 0$.

Każdy wektor izotropowy f podprzestrzeni $\mathfrak{L}$ jest *neutralny* (jeśli f należy do $\mathfrak{L}$ i jest $\mathcal{P}$ – *ortogonalne* do wszystkich elementów $\mathfrak{L}$). Przeciwna implikacja

$$\text{neutralny wektor} \quad \mathfrak{L} \quad \Rightarrow \quad \text{izotropowy wektor} \quad \mathfrak{L}$$

nie zawsze jest prawdziwa. Jeśli $\mathfrak{L}$ jest podprzestrzenią nieujemną, to podzbiór wektorów neutralnych i izotropowych jest równoważny. Fakt ten wynika z nierówności Cauchy'ego-Schwartza

$$|[f, g]|^2 \leqslant [f, f] \cdot [g, g] \quad (f, g \in \mathfrak{L}).$$

Domkniętą podprzestrzeń $\mathfrak{L}$ nazywamy *niezdegenerowaną* jeśli jej izotropowa część jest trywialna, $\mathfrak{L}_0 = 0$. W innych przypadkach nazywamy ją *degenerowalną*. Każda neutralna podprzestrzeń jest degenerowalna, zaś skończone podprzestrzenie są niezdegenerowane. Jeśli $\mathfrak{L}$ jest niezdegenerowana, to $\mathfrak{L}^{[\perp]} \cap \mathfrak{L} = \{0\}$, wtedy możemy rozważać sumę prostą $\mathfrak{L}^{[\perp]} \dot{+} \mathfrak{L}$ tych podprzestrzeni.

Twierdzenie 1. *[3] Jeśli podprzestrzeń $\mathfrak{L}$ przestrzeni Kreina $\mathfrak{H}$ jest skończenie wymiarowa, to następujące zdania są równoważne:*

1. Dla każdego $\mathbf{z} \in \mathfrak{H}$ istnieją i są określone jednoznacznie wektory $\hat{\mathbf{z}} \in \mathfrak{L}$, $\tilde{\mathbf{z}} \in \mathfrak{L}^{[\perp]}$, takie że:

$$\mathbf{z} = \hat{\mathbf{z}} + \tilde{\mathbf{z}}; \tag{3.2}$$

2. Podprzestrzeń $\mathfrak{L}$ nie zawiera wektorów izotropowych;

3. Podprzestrzeń $\mathfrak{L}$ jest niezdegenerowana.

Uwaga 1. *Operator $P_{\mathfrak{L}}\mathbf{z} = \hat{\mathbf{z}}$, gdzie $\hat{\mathbf{z}}$ jest z* (3.2) *będziemy nazywali operatorem rzutowania ortogonalnego na podprzestrzeń $\mathfrak{L}$ przestrzeni Kreina $\mathfrak{H}$.*

Wniosek 3.1. *Operator rzutowania ortogonalnego na skończenie wymiarową podprzestrzeń $\mathfrak{L}$ przestrzeni Kreina $\mathfrak{H}$ istnieje wtedy i tylko wtedy, gdy (co najmniej) jeden z warunków Twierdzenia 1 jest spełniony.*

3.2 Macierze Grama w przestrzeniach Kreina

Skupmy się na liniowych podprzestrzeniach $\mathfrak{H}$. Oznaczmy przez $\mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ liniową podprzestrzeń $\mathfrak{H}$, generowaną przez elementy $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\} \in \mathfrak{H}$. *Macierz Grama* zbioru elementów $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ definiujemy przez macierz o wymiarach $(\mathbf{N}+1) \times (\mathbf{N}+1)$

$$R_y = [[\mathbf{y}_i, \mathbf{y}_j]]_{i,j=0:N}.$$

Z własności iloczynu skalarnego $[\mathbf{x}, \mathbf{y}] = [\mathbf{y}, \mathbf{x}]^*$, gdzie * jest sprzężeniem zespolonym wynika, że macierz Grama jest macierzą Hermit'a. Wprowadźmy oznaczenie wektora kolumnowego elementów $\mathbf{y}_i$ jako

$$\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N]^T,$$

oraz macierz Grama jako

$$R_y = [\mathbf{y}, \mathbf{y}].$$

Rozważmy dwa zbiory elementów $\{\mathbf{z}_0, \ldots, \mathbf{z}_N\}$ i $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ oraz oznaczmy wektory kolumn przez

$$\mathbf{z} = [\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_N]^T \quad \mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_N]^T,$$

wtedy *krzyżową macierzą Grama* nazywamy macierz rozmiaru $(\mathbf{M}+1) \times (\mathbf{N}+1)$

$$R_{zy} = [[\mathbf{z}_i, \mathbf{y}_j]]_{i=0:M, j=0:N} = [\mathbf{z}, \mathbf{y}],$$

zauważmy, że spełnia ona własność

$$R_{zy} = R_{yz}^*$$

Lemat 1. *(Dodatnio i ujemnie określone przestrzenie liniowe) Niech $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ to liniowo niezależne elementy przestrzeni $\mathfrak{H}$. Wtedy $\mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ jest dodatnio(ujemnie) określoną podprzestrzenią $\mathfrak{H}$ wtedy i tylko wtedy, gdy*

$$R_y > 0 \ \ (R_y < 0).$$

Dowód. Weźmy dowolne, niezerowe $\mathbf{z} \in \mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$. Elementy $\mathbf{y}_i$ są liniowo niezależne, więc istnieje dokładnie jedno $k \in \mathcal{C}^{N+1}$ spełniające równanie $\mathbf{z} = k^*\mathbf{y}$. Następnie

$$[\mathbf{z}, \mathbf{z}] = [k^*\mathbf{y}, k^*\mathbf{y}] = k^*[\mathbf{y}, k^*\mathbf{y}] = k^*(k^*[\mathbf{y}, \mathbf{y}]) = k^*([\mathbf{y}, \mathbf{y}]k^*)^* = k^*[\mathbf{y}, \mathbf{y}]k = k^*R_y k$$

k^*k jest zawsze dodatnie zatem $[\mathbf{z}, \mathbf{z}] > 0$ dla dowolnego $\mathbf{z} \in \mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ wtedy i tylko wtedy, gdy $R_y > 0$. Dowód analogiczny dla $R_y < 0$.

■

Uwaga 2. *Z dowodu Lematu 1 wynika, że podprzestrzeń* $\mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ *przestrzeni Kreina* $\mathfrak{H}$ *jest niezdegenerowana* $\iff \det R_y \neq 0$. *Wektor* $\mathbf{z} \in \mathfrak{L}\{\mathbf{y}\}$ *jest wektorem izotropowym dla* $\mathfrak{L}\{\mathbf{y}\} \iff R_y\mathbf{z} = 0$. *Macierz Grama* R_y *nazywamy singularną jeśli* $\det R_y = 0$.

3.3 Rzutowanie w przestrzeniach Kreina

Definicja 2. *(Rzutowanie w przestrzeni Kreina) Dla danego wektora* $\mathbf{z}$ *w przestrzeni Kreina* $\mathfrak{H}$ *oraz dla wektorów* $\{\mathbf{y}_0, \ldots, \mathbf{y}_N\} \in \mathfrak{H}$, *definiujemy* $\hat{\mathbf{z}}$ *jako rzutowanie* $\mathbf{z}$ *na podprzestrzeń* $\mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ *jeśli*

$$\mathbf{z} = \hat{\mathbf{z}} + \tilde{\mathbf{z}},$$

gdzie $\hat{\mathbf{z}} \in \mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ *oraz* $\tilde{\mathbf{z}}$ *spełnia warunek*

$$\tilde{\mathbf{z}} \perp \mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}, \ \textit{czyli} \ \forall_{i \in \{0,\ldots,N\}} [\tilde{\mathbf{z}}, \mathbf{y}_i] = 0.$$

W przestrzeniach Hilberta rzutowanie istnieje zawsze i jest określone jednoznacznie. Jednak podobne własności nie muszą być spełnione w przestrzeniach Kreina.

Z Tw. 1 dostajemy, że rzutowanie dowolnego wektora $\mathbf{z} \in \mathfrak{H}$ na posprzestrzeń $\mathfrak{L}$ istnieje i jest określone jednoznacznie $\iff$ macierz Grama R_y jest macierzą niesingularną (tj. $\det R_y \neq 0$). W tym przypadku istnieje operator rzutowania $P_{\mathfrak{L}}\mathbf{z}$ określony w Uwadze do Tw. 1.

Jeśli $\mathfrak{L}$ jest zdegenerowana (tj. istnieją wektory izotropowe $\iff \det R_y = 0$), to dla ustalonego $\mathbf{z} \in \mathfrak{H} \wedge \mathbf{z} \notin \mathfrak{L}$ istnieją dwie możliwości:

1° rzutowanie $\hat{\mathbf{z}}$ istnieje, ale nie jest określone jednoznacznie.
Sprawdźmy, niech $\mathbf{z} = P_{\mathfrak{L}}\mathbf{z}$ oraz

$$\mathbf{z} = \hat{\mathbf{z}} + \tilde{\mathbf{z}},$$

jeśli $\hat{\mathbf{z}} \in \mathfrak{L}$ oraz $\mathbf{z_0} \in \mathfrak{L}$ to $\hat{\mathbf{z}} + \mathbf{z_0} \in \mathfrak{L}$ oraz $[\tilde{\mathbf{z}} - \mathbf{z_0}, y] = [\tilde{\mathbf{z}}, y] - [\mathbf{z_0}, y] = 0$ otrzymujemy równość

$$\mathbf{z} = \hat{\mathbf{z}} + \mathbf{z_0} + \tilde{\mathbf{z}} - \mathbf{z_0}.$$

2° rzutowanie $\hat{\mathbf{z}}$ nie istnieje.
Przykład: Niech $\mathbf{z} = P_{\mathfrak{L}}\mathbf{z}_0$, gdzie $\mathbf{z}_0$ jest wektorem izotropowym. Załóżmy, że $\mathbf{z} = \hat{\mathbf{z}} + \tilde{\mathbf{z}}$. Zauważmy, że

$$(\mathbf{z}, \hat{\mathbf{z}}) = (\mathcal{P}\mathbf{z_0}, \hat{\mathbf{z}}) = [\mathbf{z_0}, \hat{\mathbf{z}}] = 0 \quad (\text{ponieważ } \hat{\mathbf{z}} \in \mathfrak{L}),$$
$$(\mathbf{z}, \tilde{\mathbf{z}}) = (\mathcal{P}\mathbf{z_0}, \tilde{\mathbf{z}}) = [\mathbf{z_0}, \tilde{\mathbf{z}}] = 0 \quad (\text{ponieważ } \tilde{\mathbf{z}} \perp \mathfrak{L} \quad \text{oraz} \quad \mathbf{z_0} \in \mathfrak{L})$$

zatem

$$(\mathbf{z},\mathbf{z}) = (\mathbf{z},\hat{\mathbf{z}}) + (\mathbf{z},\tilde{\mathbf{z}}) = 0 \Rightarrow \mathbf{z} = 0,$$

otrzymujemy sprzeczność.

Przypadki 1° i 2° możemy łatwo opisać za pomocą macierzy Grama.

Twierdzenie 2. *1. Jeśli macierz Grama* $R_y = [\mathbf{y},\mathbf{y}]$ *jest niesingularna, to rzutowanie wektora* $\mathbf{z}$ *na podprzestrzeń* $\mathfrak{L}\{\mathbf{y}\}$ *jest określone jednoznacznie i zadane przez*

$$\hat{\mathbf{z}} = [\mathbf{z},\mathbf{y}][\mathbf{y},\mathbf{y}]^{-1}\mathbf{y} = R_{zy}R_y^{-1}\mathbf{y}.$$

2. Jeśli macierz Grama $R_y = [\mathbf{y},\mathbf{y}]$ *jest singularna to występują dwa przypadki:*

(a) $\mathcal{R}(R_{yz}) \subseteq \mathcal{R}(R_y)$, *gdzie* $\mathcal{R}(A)$ *oznacza powłokę liniową utworzoną z kolumn macierzy* A. *Rzutowanie* $\hat{\mathbf{z}}$ *istnieje, ale nie jest wyznaczone jednoznacznie. Rzeczywiście,* $\hat{\mathbf{z}} = k_0\mathbf{y}$, *gdzie* k_0 *jest dowolnym rozwiązaniem równania*

$$R_y k_0 = R_{yz}.$$

(b) $\mathcal{R}(R_{yz}) \nsubseteq \mathcal{R}(R_y)$, *rzutowanie* $\hat{\mathbf{z}}$ *nie istnieje.*

3.4 Definicja MWN w przestrzeni Kreina

Omówmy ogólny przypadek definicji Maszyny Wektorów Nośnych w przestrzeniach Hilberta i Kreina. Załóżmy, że zbiór $\mathcal{S}$ w definicji 1 MWN ma wymiar skończony (jest to naturalne założenie, ponieważ dysponujemy skończonym zbiorem treningowym). Dodatkowo, niech zbiór $\mathcal{S} = \mathfrak{L}\{\mathbf{y}_0, \dots, \mathbf{y}_N\}$, gdzie podprzestrzeń $\mathfrak{L}$ jest opisana w rozdziale 3.2.

Niech wektor $\mathbf{z}$ będzie ustalony w przestrzeni Hilberta, taki że $\mathbf{z} \in \mathfrak{H}$ oraz $\mathbf{z} \notin \mathfrak{L}\{\mathbf{y}_0, \dots, \mathbf{y}_N\}$. Z rozdziału 2.3 wynika, że jeśli $\hat{\mathbf{z}} \in \mathfrak{L}$ będzie reprezentantem maszyny wektorów nośnych, to $\hat{\mathbf{z}}$ jest rzutowaniem wektora $\mathbf{z}$ na podprzestrzeń $\mathfrak{L}$. Oznacza to, że $\mathbf{z} = \hat{\mathbf{z}} + \tilde{\mathbf{z}}$, gdzie $\hat{\mathbf{z}} \in \mathfrak{L}$, $\tilde{\mathbf{z}} \perp \mathfrak{L}\{\mathbf{y}_0, \dots, \mathbf{y}_N\}$ oraz

$$\|\mathbf{z} - \hat{\mathbf{z}}\|^2 = \min_{\mathbf{k} \in \mathbb{C}^{N+1}} \|\mathbf{z} - \mathbf{k}\mathbf{y}\|^2,$$

gdzie $\mathbf{k} = (\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_N), \mathbf{k}_i \in \mathbb{C}$ dla $i \in \{0, \dots, N\}, \mathbf{k}\mathbf{y} = \mathbf{k}_0\mathbf{y}_0 + \mathbf{k}_1\mathbf{y}_1 + \dots + \mathbf{k}_N\mathbf{k}_N$. Zapiszmy formę kwadratową

$$Q(\mathbf{k}) = (\mathbf{z} - \mathbf{k}\mathbf{y}, \mathbf{z} - \mathbf{k}\mathbf{y}) = \|\mathbf{z} - \mathbf{k}\mathbf{y}\|^2,$$

otrzymujemy następującą nierówność

$$Q(\mathbf{k}) \geqslant Q(\hat{\mathbf{k}}),$$

gdzie $\hat{\mathbf{z}} - \hat{\mathbf{k}}\mathbf{y}$, $\hat{\mathbf{k}} = (\hat{\mathbf{k}}_0, \hat{\mathbf{k}}_1, \dots, \hat{\mathbf{k}}_N)$ to oznacza, że rzutowanie $\mathbf{z}$ na podprzestrzeń $\mathfrak{L}$ przedstawione jako wektor $\hat{\mathbf{z}} = \hat{\mathbf{k}}\mathbf{y}$ minimizuje formę kwadratową $Q(\mathbf{k})$ oraz jest punktem stacjonarnym $Q(\mathbf{k})$, ponieważ

$$\left.\frac{\partial Q(\mathbf{k})}{\partial \mathbf{k}_0}\right|_{\mathbf{k}=\hat{\mathbf{k}}_0} = \left.\frac{\partial Q(\mathbf{k})}{\partial \mathbf{k}_1}\right|_{\mathbf{k}=\hat{\mathbf{k}}_1} = \dots = \left.\frac{\partial Q(\mathbf{k})}{\partial \mathbf{k}_N}\right|_{\mathbf{k}=\hat{\mathbf{k}}_N} = 0.$$

Przypadek przestrzeni Kreina, ze względu na brak normy, jest zupełnie inny. Rozpatrujemy formę kwadratową

$$Q(\mathbf{k}) = [\mathbf{z} - \mathbf{k}\mathbf{y}, \mathbf{z} - \mathbf{k}\mathbf{y}],$$

załóżmy, że rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ istnieje (patrz Tw. 2). Następnie zastanówmy się, czy wektor $\hat{\mathbf{z}} = \hat{\mathbf{k}}\mathbf{y}$ jest rozwiązaniem zagadnienia

$$Q(\hat{\mathbf{k}}) = [\mathbf{z} - \hat{\mathbf{k}}\mathbf{y}, \mathbf{z} - \hat{\mathbf{k}}\mathbf{y}] = \min_{\mathbf{k} \in \mathbb{C}^{N+1}} Q(\mathbf{k}).$$

Innymi słowy, czy rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ w przestrzeni Kreina minimizuje formę kwadratową $Q(\mathbf{k})$?

Twierdzenie 3. *Niech podprzestrzeń $\mathfrak{L}$ jest niezdegenerowaną. Wówczas rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ minimizuje formę kwadratową $Q(\mathbf{k})$ wtedy i tylko wtedy, gdy $\mathfrak{L}$ jest podprzestrzenią dodatnią.*

Dowód. Załóżmy, że $\mathfrak{L}$ jest podprzestrzenią dodatnią.

$$\begin{aligned} Q(\mathbf{k}) &= [\mathbf{z}-\mathbf{k}\mathbf{y}, \mathbf{z}-\mathbf{k}\mathbf{y}] \\ &= [\mathbf{z}-\hat{\mathbf{z}}+\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \mathbf{z}-\hat{\mathbf{z}}+\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}] \\ &= [\mathbf{z}-\hat{\mathbf{z}}, \mathbf{z}-\hat{\mathbf{z}}]+[\mathbf{z}-\hat{\mathbf{z}}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}]+[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \mathbf{z}-\hat{\mathbf{z}}]+[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}]. \end{aligned}$$

Zauważmy, że $\hat{\mathbf{z}} \in \mathfrak{L}$ oraz $\mathbf{k}\mathbf{y} \in \mathfrak{L}$, zatem $\hat{\mathbf{z}}-\mathbf{k}\mathbf{y} \in \mathfrak{L}$. Natomiast $\mathbf{z}-\hat{\mathbf{z}} \perp \mathfrak{L}\{\mathbf{y}_0, \dots, \mathbf{y}_N\}$ co oznacza, że $[\mathbf{z}-\hat{\mathbf{z}}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}] = 0$ oraz $[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \mathbf{z}-\hat{\mathbf{z}}] = 0$. Otrzymujemy równość

$$Q(\mathbf{k}) = [\mathbf{z}-\hat{\mathbf{z}}, \mathbf{z}-\hat{\mathbf{z}}]+[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}].$$

Pierwszy składnik $[\mathbf{z}-\hat{\mathbf{z}}, \mathbf{z}-\hat{\mathbf{z}}]$ nie zależy od $\mathbf{k}$, oznaczmy go jako c, zaś drugi ze względu na metrykę jest nieujemny. Dlatego forma $Q(\mathbf{k})$ osiągnie minimum dla zerowego $\hat{\mathbf{z}}-\mathbf{k}\mathbf{y} \Rightarrow (\hat{\mathbf{z}} = \hat{\mathbf{k}}\mathbf{y},\ \hat{\mathbf{z}}-\hat{\mathbf{k}}\mathbf{y} = 0)$ oraz zachodzi równość

$$Q(\hat{\mathbf{k}}) = c+[0,0] = c,$$

czyli $\hat{\mathbf{z}}-\hat{\mathbf{k}}\mathbf{y}$ minimizuje formę kwadratową $Q(\mathbf{k})$.

Przejdźmy do odwrotnej implikacji. Zakładamy, że rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ minimizuje formę kwadratową $Q(\mathbf{k})$. Oznacza to, że $Q(\mathbf{k})$ przyjmuje wartość minimalną dla $\mathbf{k} = \hat{\mathbf{k}}$.

Przeprowadźmy dowód nie wprost. Załóżmy, że $\mathfrak{L}$ jest podprzestrzenią ujemną. Analogicznie do powyższego rozumowania, możemy zapisać formę jako

$$Q(\hat{\mathbf{k}}) = [\mathbf{z}-\hat{\mathbf{k}}\mathbf{y}, \mathbf{z}-\hat{\mathbf{k}}\mathbf{y}] = c+[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}].$$

Zauważmy, że jeśli $\mathfrak{L}$ jest podprzestrzenią ujemną, to $[\hat{\mathbf{z}}-\mathbf{k}\mathbf{y}, \hat{\mathbf{z}}-\mathbf{k}\mathbf{y}] < 0 \Rightarrow \hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ nie minimizuje formy kwadratowej.

Otrzymujemy sprzeczność, czyli $\mathfrak{L}$ jest podprzestrzenią dodatnią.

■

Wniosek 3.2. *Twierdzenie podobne do Twierdzenia 3 również jest prawdziwe, tj. dla niezdegenerowanej podprzestrzeni $\mathfrak{L}$ rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ maksymalizuje formę kwadratową $Q(\mathbf{k})$ wtedy i tylko wtedy, gdy $\mathfrak{L}$ jest podprzestrzenią ujemną.*

W przypadku gdy podprzestrzeń $\mathfrak{L}$ jest niezdegenerowa, rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ istnieje, ale wektor $\hat{\mathbf{z}} - \hat{\mathbf{k}}\mathbf{y}$ nie minimizuje i nie maksymalizuje formy kwadratowej $Q(\mathbf{k})$.

Z twierdzenia 2.4.1[4] dostajemy, że rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ $(\hat{\mathbf{z}} - \hat{\mathbf{k}}\mathbf{y})$ daje jednoznacznie określony punkt stacjonarny $\hat{\mathbf{k}}$ formy kwadratowej $Q(\mathbf{k})$.

Podobnie do rozdziału 2.3 spróbujemy podać definicję Maszyny Wektorów Nośnych w przestrzeni Kreina.

Definicja 3. *(MWN jako rzutowanie) Dla niezdegenerowanej podprzestrzeni $\mathcal{S} = \mathfrak{L}\{\mathbf{y}_0, \ldots, \mathbf{y}_N\}$ i ustalonego $\mathbf{z} \notin \mathcal{S}$ maszyną wektorów nośnych nazywamy jednoznacznie określone rzutowanie $\hat{\mathbf{z}} = P_{\mathfrak{L}}\mathbf{z}$ (w przestrzeni Kreina).*

Powyższe rozważania pociągają za sobą następujące konsekwencje:

Wniosek 3.3. *Maszyna wektorów nośnych $\hat{\mathbf{z}}$ jest określona jednoznacznie i zachodzi wzór na $\hat{\mathbf{z}}$ podany w Tw. 2 za pomocą macierzy Grama.*

Wniosek 3.4. *Gdy $\mathcal{S}$ jest podprzestrzenią dodatnią, to MWN zadany jako $\hat{\mathbf{z}} = \hat{\mathbf{k}}\mathbf{y}$ minimizuje formę kwadratową błędu $Q(\mathbf{k}) = [\mathbf{z} - \mathbf{k}\mathbf{y}, \mathbf{z} - \mathbf{k}\mathbf{y}]$(wynika z Tw.3).*

Wniosek 3.5. *Gdy $\mathcal{S}$ jest podprzestrzenią ujemną, to MWN $\hat{\mathbf{z}} - \hat{\mathbf{k}}\mathbf{y}$ maksymalizuje formę kwadratową błędu $Q(\mathbf{k})$.*

Wniosek 3.6. *Przy założeniu, że $\mathcal{S}$ ma wektory ujemne i wektory dodatnie dostajemy, że MWN nie minimizuje i nie maksymizuje formy kwadratowej błędu $Q(\mathbf{k})$, ale jest jednoznacznie określonym punktem stacjonarnym dla $Q(\mathbf{k})$.*

3.5 Przykłady

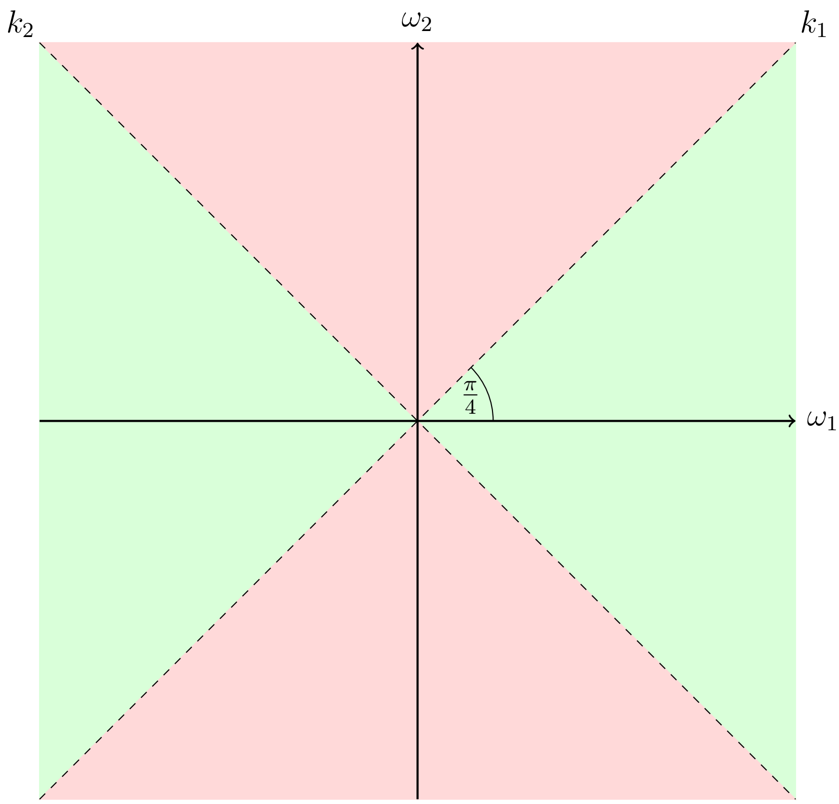
Rozważmy przestrzeń Hilberta

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} : \omega_1, \omega_2 \in \mathbb{R} \right\} \quad \text{z iloczynem skalarnym} \quad \left(\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}, \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \right) = \omega_1^2 + \omega_2^2$$

oraz przestrzeń Kreina $\mathbb{R}^2$ z metryką nieoznaczoną

$$\left[\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}, \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \right] = \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}, \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \right) = \omega_1^2 - \omega_2^2,$$

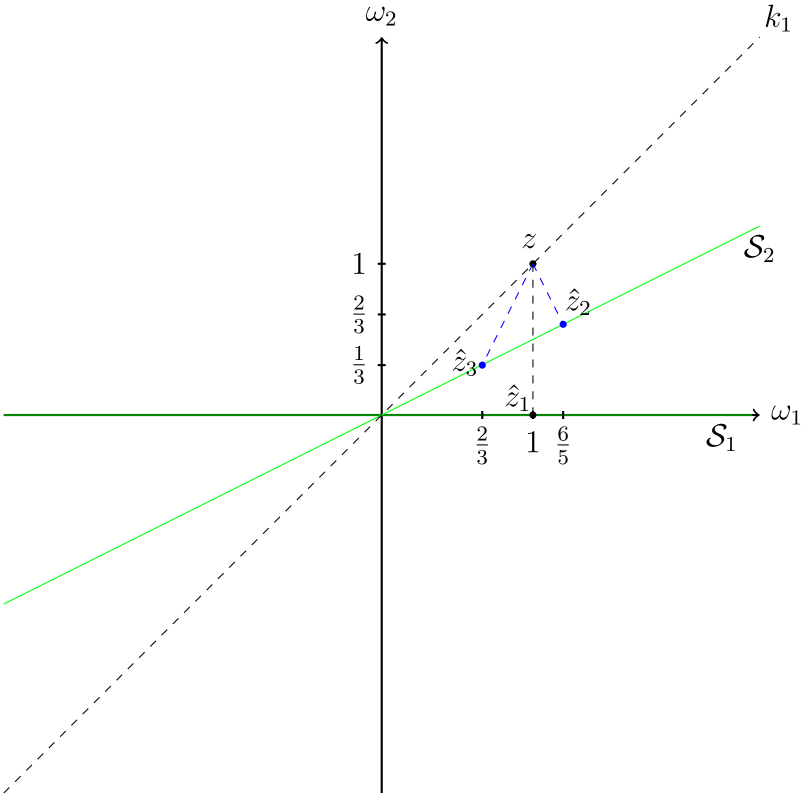
gdzie $\mathcal{P} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ to operator symetrii fundamentalnej.



Rys. 2

Każdy punkt płaszczyzny euklidesowej odpowiada wektorowi w przestrzeni Kreina. Wektory neutralne reprezentowane są przez proste k_1 oraz k_2. Obszar wektrów dodatnich zaznaczony jest na zielono, zaś ujemnych na czerwowo.

Niech $z = (1,1) \in \mathbb{R}^2$



Rys. 3

Przykład 1. Podprzestrzeń wektorów dodatnich $\mathcal{S}_1 = \{k\mathbf{y}_0 : \mathbf{y}_0 = (1,0), k \in \mathbb{R}\}$ Wtedy MWN to rzutowanie z na $\mathcal{S}_1 \Rightarrow \hat{z}_1 = (1,0)$. Zauważmy że MWN w przestrzeni Kreina to ten sam punkt.

Przykład 2. Niech $\mathcal{S}_2 = \{k\mathbf{y}_0 : \mathbf{y}_0 = (1, \frac{1}{2}), k \in \mathbb{R}\}$. W przestrzeni Hilberta szukamy z minimizującego formę $Q(k) = \|z - k\mathbf{y}_0\|^2 = (z - k\mathbf{y}_0, z - k\mathbf{y}_0)$

$$\begin{aligned} Q(k) &= \left(\begin{bmatrix} 1-k \\ 1-\frac{1}{2}k \end{bmatrix}, \begin{bmatrix} 1-k \\ 1-\frac{1}{2}k \end{bmatrix} \right) \\ &= (1-k)^2 + (1-\frac{1}{2}k)^2 \\ &= 1 - 2k + k^2 + 1 - k + \frac{1}{4}k^2 \\ &= 2 - 3k + \frac{5}{4}k^2. \end{aligned}$$

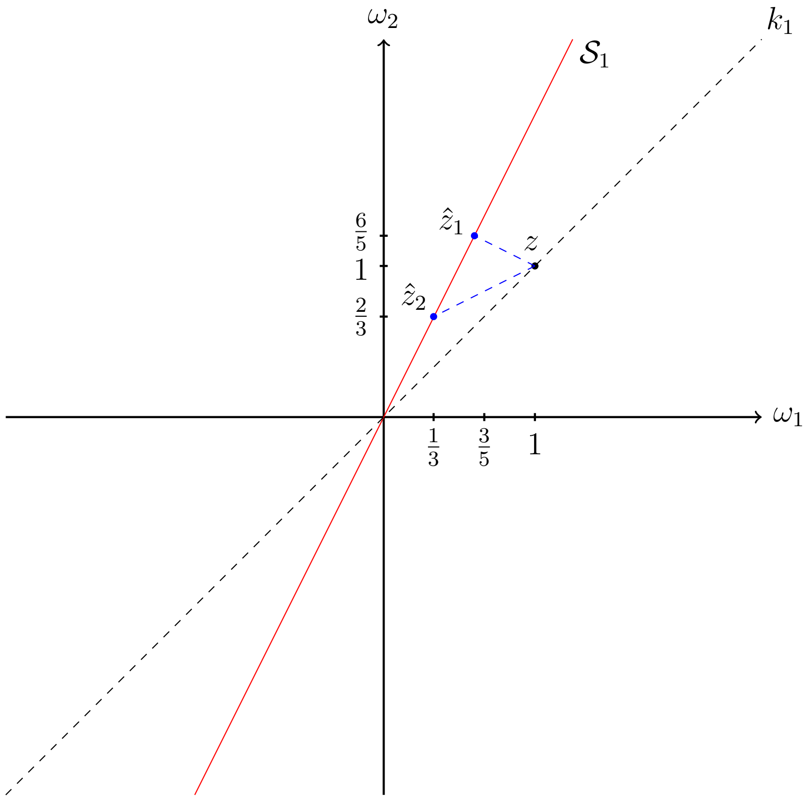
$Q(k)$ przyjmuje wartość minimalną gdy $Q'(k) = 0 \iff \frac{5}{2}k = 3 \Rightarrow \hat{k} = \frac{6}{5}$. Następnie $\hat{z} = \hat{k}\mathbf{y}_0$, stąd MWN w przestrzeni Hilberta $\hat{z}_1 = (\frac{6}{5}, \frac{3}{5})$ (Patrz Rys. 3)

Wyliczmy analogicznie z minimizujące formę $Q(k)$ w przestrzeni Kreina

$$\begin{aligned} Q(k) &= \left[\begin{bmatrix} 1-k \\ 1-\frac{1}{2}k \end{bmatrix}, \begin{bmatrix} 1-k \\ 1-\frac{1}{2}k \end{bmatrix} \right] \\ &= (1-k)^2 - (1-\frac{1}{2}k)^2 \\ &= 1-2k+k^2-1+k-\frac{1}{4}k^2 \\ &= \frac{3}{4}k^2 - k. \end{aligned}$$

$Q(k)$ przyjmuje wartość minimalną gdy $Q'(k) = 0 \iff \frac{3}{2}k = 1 \Rightarrow \hat{k} = \frac{2}{3}$ oraz $\hat{z} = \hat{k}\mathbf{y}_0$, zatem MWN w przestrzeni Hilberta $\hat{z}_2 = (\frac{2}{3}, \frac{1}{3})$ (Patrz Rys. 3). Ten sam wynik otrzymamy za pomocą macierzy Grama z Twierdzenia 2.

Przykład 3. Podprzestrzeń wektorów ujemnych $\mathcal{S}_1 = \{k\mathbf{y}_0 : \mathbf{y}_0 = (\frac{1}{2}, 1), k \in \mathbb{R}\}$. W przestrzeni Hilberta $\hat{z}_1 = (\frac{3}{5}, \frac{6}{5})$, natomiast w przestrzeni Kreina $\hat{z}_2 = (\frac{1}{3}, \frac{2}{3})$ (Patrz Rys. 4).



Rys. 4

4 Bibliografia

Literatura

[1] Gaëlle Loosli *et al., "Learning SVM in Krein Spaces"*, 2015.

[2] T. Ando, "Projections in Krein spaces," *Linear Algebra andits Applications,* vol. 431, no. 12, pp. 2346 – 2358, 2009.

[3] T. Y. Azizov and I. S. Iokhvidov, *Linear Operators in Spaces withan Indefinite Metric.* Wiley, 1989.

[4] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-quadratic estimation and control: a unified approach to H2 and H [infinity] theories.* SIAM, 1999.

[5] N. Akhiezer, I Glazman, *Theory of Linear Operators in Hilbert Spaces* Pitman Publishing Inc. 1981.