

Obliczenia neuro-rozmyte

Sławomir T. Wierchoń
Instytut Podstaw Informatyki PAN
Instytut Informatyki Uniwersytetu Gdańskiego

16 maja 2011

Spis treści

1	Wstęp	11
2	Pojęcia podstawowe	13
2.1	Podstawowe definicje	13
2.2	Funkcja przynależności	15
2.3	Liczby i przedziały rozmyte	18
2.4	Estymacja liczb rozmytych	19
2.5	Reprezentacja zbioru rozmytego	20
2.6	Zadania	22
3	Zasada rozszerzania i jej zastosowania	23
3.1	Idea rozszerzania	23
3.2	Rozmyta arytmetyka	25
3.3	Metody numeryczne dla zasady rozszerzania	28
3.3.1	Metoda wierzchołkowa	29
3.3.2	Metoda transformacji	30
3.3.3	Metoda funkcji sklepanych	33
3.4	Zastosowania	37
3.4.1	Rozmyte operatory min oraz max	37
3.4.2	Metryzacja przestrzeni liczb rozmytych	38
3.4.3	Równania liniowe	39
3.4.4	Układy równań liniowych	41
3.4.4.1	Metoda α -warstw	41
3.4.4.2	Metoda geometryczna	45
3.5	Zadania	46
4	Sterowanie i sterowniki rozmyte	49
4.1	Architektura rozmytego sterownika	50
4.1.1	Baza reguł	51
4.1.2	Blok rozmywania	52
4.1.3	Blok wnioskowania	54
4.1.4	Blok wyostrzania	57
4.1.4.1	Sterownik Mamdaniego	57
4.1.4.2	Sterownik Takagi-Sugeno	59
4.1.5	Przykłady	59
4.1.5.1	Odwrócone wahadło	59
4.1.5.2	Automatyczny pilot	61
4.2	Podstawowy algorytm generowania reguł rozmytych	63
4.2.1	Podział wejściowej i wyjściowej przestrzeni na obszary	63

4.2.2	Tworzenie reguł rozmytych na podstawie przykładów	63
4.2.3	Przyporządkowanie stopni prawdziwości regułom	64
4.2.4	Utworzenie bazy rozmytych reguł	64
4.2.5	Wyostrażanie	64
4.3	Alternatywne metody generowania reguł	65
4.3.1	Generowanie uproszczonych reguł	65
4.3.1.1	Wnioskowanie z użyciem prostych reguł rozmytych	67
4.3.1.2	Identyfikacja prostych reguł rozmytych	67
4.3.1.3	Konwersja prostych reguł rozmytych	68
4.3.2	Zastosowanie algorytmów genetycznych	69
4.3.2.1	Genetyczna regresja	69
4.3.2.2	Uczenie reguł	69
4.4	Przykładowe problemy	71
4.4.1	Zadanie parkowania ciężarówki	71
4.5	Równania różniczkowe	72
4.5.1	Pojęcia wstępne	73
4.5.1.1	Pochodne rozmytych funkcji	73
4.5.2	Rozmyte zagadnienie Cauchy'ego	74
4.5.2.1	Liniowe równania różniczkowe z rozmytym warunkiem początkowym	74
5	Sieci jednowarstwowe	77
5.1	Podstawowe określenia	77
5.2	Klasyfikacja liniowa	79
5.3	Algorytm perceptronu	80
5.4	Reguła <i>delta</i>	82
5.4.1	Reguła Widrowa-Hoffa	82
5.4.2	Wybór parametrów	85
5.4.3	Reguła delta z nieliniową funkcją aktywacji	86
5.5	Zadania	87
6	Sieci wielowarstwowe	89
6.1	Uogólniona reguła delta	89
6.2	Radialne sieci neuronowe	92
6.2.1	Uczenie wag	93
6.2.2	Wybór centrów	94
6.3	Sieci wielowarstwowe jako uniwersalne aproksymatory	95
6.4	Zadania	97
7	Sieci samoorganizujące się	99
8	Inne sieci neuronowe	101
9	Systemy neuro-rozmyte	103
10	Projekty	105
10.1	Jak przygotowywać raporty dokumentujące projekty?	105
10.1.1	Część opisowa	105
10.1.1.1	Streszczenie	105
10.1.1.2	Wprowadzenie	105

10.1.1.3	Metodologia	106
10.1.1.4	Wyniki i ich analiza	106
10.1.1.5	Konkluzje i plany dalszych badań	107
10.1.1.6	Bibliografia	107
10.1.2	Część programowa	108

Wykaz rysunków

2.1	Przykładowe funkcje przynależności wypukłego i niewypukłego zbioru rozmytego.	15
2.2	Pseudo-gaussowskie funkcje przynależności.	16
2.3	Rozmyte przedziały.	17
2.4	Funkcja przynależności liczby rozmytej reprezentowana przez rodzinę przedziałów ufności.	20
3.1	Rozmyta liczba $\Delta(-8, 2, 4)$ i jej obraz w przekształceniu $f(x) = x^2$	24
3.2	Funkcja przynależności iloczynu dwóch liczb rozmytych	27
3.3	Liczba rozmyta $\tilde{q} = \tilde{a}/\tilde{a}$	28
3.4	Ilustracja metody wierzchołkowej	30
3.5	Ilustracja metody transformacji	31
3.6	Obraz zbioru rozmytego w nieliniowym przekształceniu.	33
3.7	Wynik nieliniowego przekształcenia $g = \sin(x_1) + x_2^2 - x_3$	34
3.8	Ilustracja aproksymacyjnej metody wyznaczania funkcji przynależności dla $f(\tilde{u})$	36
3.9	Przybliżona funkcja przynależności zbioru $f(\tilde{u})$	36
3.10	Wyznaczanie rozmytego minimum i maksimum.	38
3.11	Rozwiązanie równania liniowego.	40
3.12	Przykład słabego rozwiązania.	43
3.13	Piramidalne rozwiązanie układu równań.	47
4.1	Podstawowy schemat sterowania	49
4.2	Schemat blokowy sterownika rozmytego.	51
4.3	Wykresy T-normy $\min(a, b)$ i dualnej S-normy $\max(a, b)$	53
4.4	Graficzna reprezentacja procesu wnioskowania w oparciu o dwie reguły . . .	56
4.5	Porównanie metod wyostrzania.	59
4.6	Problem sterowania odwróconym wahadłem.	60
4.7	Pożądana trajektoria lotu w przestrzeni fazowej (h, v)	62
4.8	Zależność sygnału wyjściowego od wejściowego w proporcjonalnym sterowniku rozmytym	66
4.9	Porównanie jakości aproksymacji zadanej funkcji dla dwóch różnych pokryć jej dziedziny	68
4.10	Parametry wejściowe, (x, u, ϕ) oraz parametr sterujący θ w zadaniu parkowania ciężarówki.	72
4.11	Przykładowe definicje zbiorów rozmytych wykorzystane do sterowania procesem parkowania ciężarówki	73
4.12	Dane empiryczne wraz z wygenerowanym zestawem reguł	74

5.1	Różnica między klasyfikacją a regresją.	78
5.2	Przykład zbioru separowalnego i nieseparowalnego	80
5.3	Jednowarstwowa sieć neuronowa.	82
5.4	Reguła najszybszego spadku. Wybór kolejnego punktu odbywa się w kierunku przeciwnym do kierunku pochodnej.	83
5.5	Wartości rzeczywiste, a wartości wyznaczone przez sieć neuronową dla funkcji liniowej.	85
5.6	Unipolarna i bipolarna funkcja aktywacji.	86
6.1	Struktura sieci wielowarstwowej. Warstwa wejściowa składa się z n neuronów, warstwa ukryta – z u neuronów, a w warstwie wyjściowej jest pojedynczy neuron.	89
6.2	Uczenie funkcji XOR. Na osi odciętych zaznaczono epoki, a na osi rzędnych wartość błędu	92

Wykaz tablic

3.1	Przykładowa implementacja metody transformacji	32
4.1	Normy i ko-normy trójkątne	54
4.2	Najczęściej spotykane definicje rozmytej implikacji	54
4.3	Baza rozmytych reguł	61
4.4	Pokrycie zbiorami rozmytymi dziedziny zmiennej v	62
4.5	Rozmyta pamięć asocjacyjna	62
4.6	Postać bazy rozmytych reguł	65
4.7	Lingwistyczny opis systemu wraz z danymi empirycznymi	69
5.1	Wartości wektora wagowego uzyskane w trzech losowych eksperymentach. W ostatniej kolumnie podano odległość estymowanego wektora od wektora docelowego $w^* = (-4, 3, 7)^T$	84
5.2	Często stosowane postacie funkcji aktywacji. Uwaga: przyjęto tu „posze- rzoną” reprezentację obiektów i wag, tzn. $\mathbf{x} = (x_1, \dots, x_n, R^2)^T$, $\mathbf{w} =$ $(w_1, \dots, w_n, b)^T$	86

Notacja

$\mathcal{F}(X)$ – rodzina rozmytych podzbiorów przestrzeni X

$\mathcal{K}_{\mathcal{C}}^n$ – rodzina niepustych wypukłych i zwartych podzbiorów przestrzeni \mathbb{R}^n (zamiast $\mathcal{K}(\mathbb{R}^n)$)

$\mathbb{R}_{\mathcal{F}}$ – zbiór liczb i przedziałów rozmytych

\mathbb{R}^n – n wymiarowa przestrzeń euklidesowa, tzn. $\mathbb{R}^n = \times_{i=1}^n \mathbb{R}$

Rozdział 1

Wstęp

Logical Designs can help your company to understand and use fuzzy logic in your applications. Fuzzy logic deals with information in the same indefinite terms that characterize human speech and perceptions. By developing a membership function for "speed", we can say "medium speed" instead of "55.0 MPH". This gives us a means of quantifying the rules of thumb that experts use in describing interactions in a complex system.

Many commercial applications of fuzzy logic have reached the production stage of development. Currently, most of these applications are Japanese. Applications include exposure control for cameras, scheduling for subways and, controllers for air conditioning.

Fuzzy Cognitive Maps and Fuzzy Associative Memories are techniques that bridge the gap between traditional expert systems and neural networks. These are the primary techniques used in developing fuzzy logic applications.

Development of a fuzzy logic application is quite different from a typical neural network application. Fuzzy logic applications tend to be hand designed as opposed to trained from examples. Rules and relationships known to the human expert are directly implemented in the cognitive maps and FAM Rules.

Recent work has shown that adaptive methods can be used in which system input-output data is converted into weighted fuzzy associative memory (FAM) rules. Set membership functions and FAM rules can also be adapted to optimize any desired performance measure. This implies that a system, developed by an expert, can be improved by training on the available data.

Fuzzy logic applications usually require less data than neural network applications. Fuzzy applications almost always require less processing power than the equivalent neural application.

The resulting fuzzy application has the important characteristic that the activity internal to the network can be traced and understood. By comparison, it is difficult to understand the internal workings of a typical neural application.

Logical Designs can handle all aspects of the design of your fuzzy application. We can supply custom programming or use commercially available software. Fuzzy logic has been implemented in hardware chips from several companies. Logical Designs can implement the developed network using these special chips for high speed, on standard microprocessors, or in software on your computer system.

Rozdział 2

Pojęcia podstawowe

2.1 Podstawowe definicje

Zakładamy, że wyróżniono pewne uniwersum X . Może ono reprezentować np. zakres mierzonych wartości, takich jak płace pracowników w firmach wybranej branży, temperatura ciała, itp. Dla uproszczenia zakładamy, że $X \subset \mathbb{R}$. W zbiorze X możemy teraz wyróżniać pewne podzbiory A_1, \dots, A_n grupujące elementy o określonych własnościach; w przypadku płac można mówić np. o *niskiej*, *średniej*, *wysokiej*, czy *bardzo wysokiej* płacy.

W matematyce typowym sposobem definiowania zbioru jest użycie funkcji charakterystycznej

$$\chi_{A_i}: X \rightarrow \{0, 1\} \quad (2.1)$$

która przypisuje elementom $x \in X$ należącym do zbioru A_i (a więc elementom, posiadającym własność etykietowaną tym zbiorem) wartość $\chi_{A_i}(x) = 1$, a pozostałym elementom – wartość $\chi_{A_i}(x) = 0$.

W przypadku zbiorów rozmytych dopuszcza się częściową przynależność elementów uniwersum do określonego zbioru A_i . Stopień przynależności konkretnego elementu $x \in X$ do zbioru A_i kwantyfikuje teraz funkcja przynależności

$$\mu_{A_i}: X \rightarrow [0, 1] \quad (2.2)$$

Jeżeli $\mu_{A_i}(x) = 1$ oznacza to, że element x na pewno spełnia warunek pozwalający zaliczyć go do zbioru A_i , a jeżeli $\mu_{A_i}(x) = 0$, że element x zdecydowanie nie posiada danej własności. Istnieje jednak pokaźna liczba elementów spełniających warunek A_i jedynie w pewnym stopniu. Można np. powiedzieć, że jeżeli $\mu_{A_i}(x) > \frac{1}{2}$, to ten element *raczej* należy do zbioru A_i , natomiast te elementy, dla których $\mu_{A_i}(x) < \frac{1}{2}$ – *raczej* nie należą do A_i .

Uwaga 2.1.1. W literaturze stosuje się różne konwencje dotyczące oznaczania zbiorów rozmytych:

- (a) Jedną z bardziej popularnych jest użycie symbolu falki nad symbolem zbioru; tak więc \tilde{A} oznacza zbiór rozmyty, natomiast A – zbiór ostry. Konwencję tę będziemy stosować w tych sytuacjach, gdy jednocześnie operuje się pojęciami ostrymi i rozmytymi. W sytuacjach, gdy jest oczywiste, że mówimy o zbiorze rozmytym, symbol falki będziemy – dla uproszczenia notacji – pomijać.
- (b) Zbiór rozmyty charakteryzuje się poprzez funkcje przynależności. Piszemy więc $\mu_{\tilde{A}}(x)$, ewentualnie $\mu_A(x)$.

- (c) Dalsze uproszczenie notacji polega na zastąpieniu oznaczeń $\mu_{\tilde{A}}(x)$, ewentualnie $\mu_A(x)$ prostszymi, tzn. $\tilde{A}(x)$ lub po prostu $A(x)$.

Wybór właściwej notacji dyktowany jest oszczędnością i czytelnością zapisu, a ponadto zależy od specyficznego kontekstu. Bez względu na podjętą decyzję, rodzinę rozmytych podzbiorów uniwersum X będziemy oznaczać symbolem $\mathcal{F}(X)$. \square

Wprowadźmy pewne podstawowe definicje:

1. O funkcji przynależności zakłada się zazwyczaj, że istnieje przynajmniej jeden element x^* taki, że $\mu_A(x^*) = 1$. Mówimy wówczas, że dany zbiór rozmyty jest normalny (tzn. znormalizowany). W przypadku przeciwnym mówimy o zbiorze subnormalnym.
2. Przez wysokość $h(A)$ zbioru rozmytego A rozumiemy maksymalną wartość charakteryzującą go funkcji przynależności, tzn.

$$h(A) = \sup_{x \in X} \mu_A(x) \quad (2.3)$$

Zatem zbiór rozmyty A jest zbiorem normalnym, jeżeli jego wysokość $h(A) = 1$, a zbiorem subnormalnym, jeżeli $h(A) < 1$.

3. Domknięcie zbioru $\{x \in X : \mu_A(x) > 0\}$ nazywamy nośnikiem zbioru A_i i oznaczamy symbolem $\text{supp}(A)$, tzn.

$$\text{supp}(A) = \overline{\{x \in X : \mu_A(x) > 0\}} \quad (2.4)$$

gdzie \overline{Z} oznacza operację domknięcia zbioru Z .

Przekątną nośnika $\text{supp}(A)$ nazywamy średnicą zbioru rozmytego A i oznaczamy symbolem $\text{diam}(A)$. Jeżeli $\text{supp}(A) = [a, b]$, przy czym $|ab| < \infty$, to

$$\text{diam}(A) = b - a$$

4. Zbiór rozmyty A jest pusty, co zapisujemy $A = \emptyset$, wtedy i tylko wtedy, gdy $\mu_A(x) = 0$ dla każdego $x \in X$. Zauważmy, że gdy $A = \emptyset$, to $\text{supp}(A) = \emptyset$.

Podobnie, jeżeli $\mu_A(x) = 1$ dla każdego $x \in X$, to zbiór A jest tożsamy z całym uniwersum X .

5. Zbiór A jest podzbiorem zbioru B , co zapisujemy $A \subset B$, wtedy i tylko wtedy, gdy $\mu_A(x) \leq \mu_B(x)$ dla każdego $x \in X$. Jeżeli $\mu_A(x) = \mu_B(x)$, to oba zbiory są identyczne, co oznaczamy $A \equiv B$.
6. α -przekrojem lub α -warstwą, $0 \leq \alpha \leq 1$, zbioru rozmytego nazywamy ostry zbiór

$$[A]_\alpha = \{x \in X : \mu_A(x) \geq \alpha\} \quad (2.5)$$

Jeżeli $\alpha_1 < \alpha_2$ to $[A]_{\alpha_1} \supseteq [A]_{\alpha_2}$. Można sprawdzić, że jeżeli A, B są dwoma zbiorami rozmytymi, to **poprawić**

$$\begin{aligned} [A \cup B]_\alpha &= [A]_\alpha \cup [B]_\alpha \\ [A \cap B]_\alpha &= [A]_\alpha \cap [B]_\alpha \end{aligned} \quad (2.6)$$

7. Zbiór

$$\text{core}(A) = [A]_1 = \{x \in X : \mu_A(x) = 1\} \quad (2.7)$$

nazywamy jądrem zbioru rozmytego A . Jeżeli A jest zbiorem normalnym, to jego jądro zawiera przynajmniej jeden element.

8. Zbiór rozmyty $A \subset X$ nazywamy wypukłym¹ jeżeli

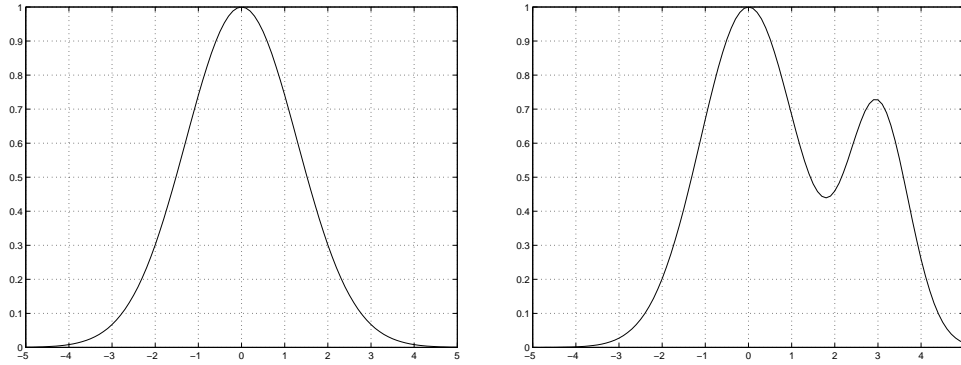
$$(\forall x_1, x_2 \in \mathbb{R})(\forall \lambda \in [0, 1]) \mu_A[\lambda x_1 + (1 - \lambda)x_2] \geq \min[\mu_A(x_1), \mu_A(x_2)] \quad (2.8)$$

Z definicji tej nie wynika, że μ_A musi być wypukłą funkcją zmiennej x . Wynika natomiast wypukłość jego α -warstw. Co więcej, wypukłość zbioru A stwierdzamy wtedy i tylko wtedy, gdy wszystkie jego α -przekroje są zbiorami wypukłymi.

Istotnie, zakładając, że wszystkie α -przekroje zbioru A są zbiorami wypukłymi wybierzmy wartość $\alpha = \mu_A(x_1) \leq \mu_A(x_2)$. Wówczas $x_1, x_2 \in [A]_\alpha$, a ponadto $\lambda x_1 + (1 - \lambda)x_2 \in [A]_\alpha$ gdyż $[A]_\alpha$ jest zbiorem wypukłym. Stąd

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \alpha = \mu_A(x_1) = \min[\mu_A(x_1), \mu_A(x_2)]$$

Na odwrót, jeżeli spełniona jest nierówność (2.8) i $\alpha = \mu_A(x_1)$, to z definicji α -warstwy wynika, że zbiór $[A]_\alpha$ jest zbiorem punktów $x_2 \in X$ spełniających warunek $\mu_A(x_2) \geq \mu_A(x_1)$. W świetle definicji (2.8) każdy punkt $x_3 = \lambda x_1 + (1 - \lambda)x_2$ jest elementem zbioru $[A]_\alpha$, a zatem $[A]_\alpha$ jest zbiorem wypukłym. Ilustracją wprowadzonych tu pojęć jest rysunek 2.1.



Rysunek 2.1: Przykładowe funkcje przynależności wypukłego i niewypukłego zbioru rozmytego

2.2 Funkcja przynależności

Typowe postacie funkcji przynależności to (dla prostoty pomijamy indeks specyfikujący zbiór rozmyty):

1. Gaussowska funkcja przynależności

$$\mu^G(x; x_0, \sigma) = \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right) \quad (2.9)$$

Z uwagi na to, że średnica zbioru opisanego taką funkcją jest nieskończona, stosuje się często funkcję pseudo-gaussowską μ^{pG} , zdefiniowaną następująco

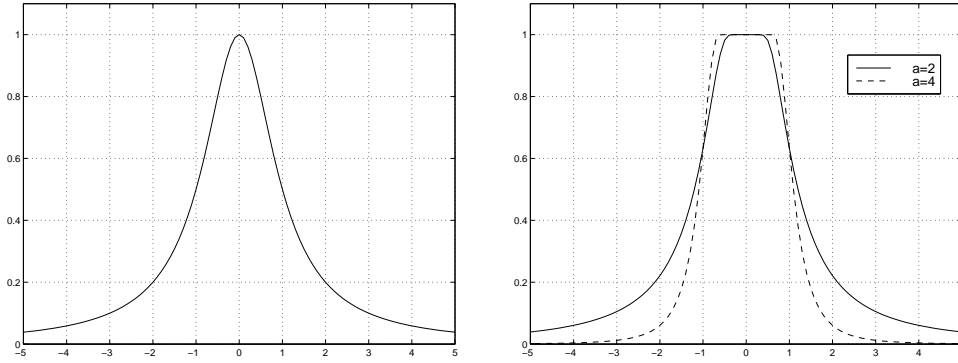
$$\mu^{pG}(x; x_0, \sigma) = \begin{cases} \mu^G(x; x_0, \sigma) & \text{gdy } |x - x_0| \leq 3\sigma \\ 0 & \text{gdy } |x - x_0| > 3\sigma \end{cases} \quad (2.10)$$

¹O ile wszystkie dotychczasowe definicje odnoszą się do przypadku, gdy uniwersum X jest dowolnym zbiorem elementów, to teraz musimy założyć, że $X \subset \mathbb{R}$.

2. Kształt zbliżony do gaussowskiego mają funkcje opisane równaniami

$$\begin{aligned}\mu_1(x) &= \left[1 + \left(\frac{x-x_0}{\sigma}\right)^2\right]^{-1} \\ \mu_2(x) &= 1 - \exp\left[-\left(\frac{\sigma}{x_0-x}\right)^a\right]\end{aligned}\quad (2.11)$$

Dodatkowy parametr a użyty w drugim równaniu pozwala sterować nachyleniem ramion krzywej. Oba typy funkcji przynależności przedstawiono na rysunku 2.2. O ile funkcja μ_1 może być użyta do reprezentowania liczb rozmytych, to funkcja μ_2 definiuje przedział rozmyty. Podobnie jak w przypadku zbiorów gaussowskich tak i tutaj średnice odpowiednich zbiorów rozmytych są nieskończone.



Rysunek 2.2: Pseudo-gaussowskie funkcje przynależności. Lewy rysunek przedstawia funkcję μ_1 opisaną pierwszym równaniem (2.11), a prawy – funkcję μ_2 zadaną drugim równaniem (2.11). Przyjęto $x_0 = 0$, $\sigma = 1$.

3. S -funkcja (przypominająca kształtem funkcję sigmoidalną); jej parametrami są lewy (x_l) i prawy (x_r) punkt przegięcia.

$$\mu^S(x; x_l, x_r) = \begin{cases} 0 & \text{gdy } x < x_l \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-x_r}{x_r-x_l}\pi\right) & \text{gdy } x_l \leq x \leq x_r \\ 1 & \text{gdy } x > x_r \end{cases}$$

Jej odbiciem jest Z -funkcja zdefiniowana następująco

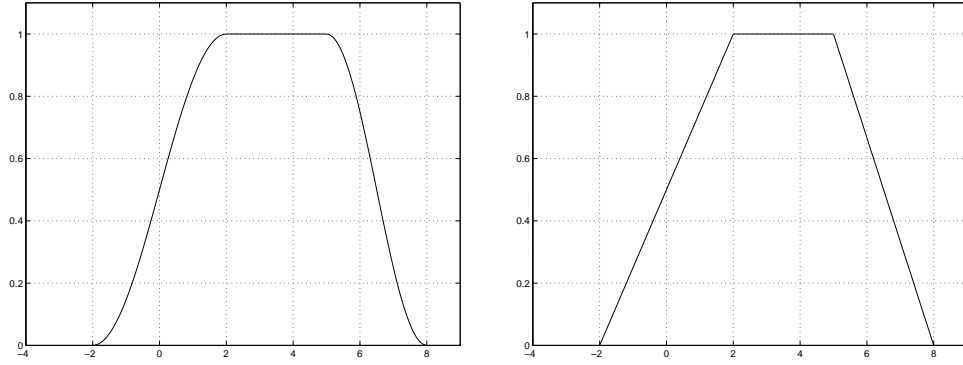
$$\mu^Z(x; x_l, x_r) = \begin{cases} 1 & \text{gdy } x < x_l \\ \frac{1}{2} + \frac{1}{2} \cos\left(\frac{x-x_l}{x_r-x_l}\pi\right) & \text{gdy } x_l \leq x \leq x_r \\ 0 & \text{gdy } x > x_r \end{cases}$$

Złożeniem obu tych funkcji przynależności jest tzw. π -funkcja przedstawiona na rysunku 2.3 i określona równaniem

$$\mu^\pi(x; x_1, x_2, x_3, x_4) = \min[\mu^S(x; x_1, x_2), \mu^Z(x; x_3, x_4)] \quad (2.12)$$

Duże znaczenie mają trapezoidalne funkcje przynależności opisane następująco (por. prawy rysunek 2.3)

$$\mu(x; a, b, c, d) = \begin{cases} \frac{x-a}{b-a} & \text{gdy } a \leq x \leq b \\ 1 & \text{gdy } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{gdy } c \leq x \leq d \\ 0 & \text{wpp} \end{cases} \quad (2.13)$$



Rysunek 2.3: Rozmyte przedziały. Lewy rysunek przedstawia funkcję przynależności typu π , a prawy – trapezoidalną funkcję przynależności. Parametry funkcji π to $x_1 = -2$, $x_2 = 2$, $x_3 = 5$, $x_4 = 8$. Dla funkcji trapezoidalnej przyjęto analogiczne wartości odpowiednich parametrów.

W przypadku gdy $b = c$ otrzymujemy *trójkątną* funkcję przynależności opisaną trójką liczb (a, b, c) (dla prostoty przyjmuje się często, że $b = \frac{a+c}{2}$, chociaż ogólnie musi być spełniony jedynie warunek $a \leq b \leq c$). Stopień przynależności elementu x do zbioru rozmytego A opisanego taką funkcją przynależności μ_A oblicza się następująco:

$$\mu_A(x) = \begin{cases} \frac{x-a}{b-a} & \text{gdy } a \leq x \leq b \\ \frac{c-x}{c-b} & \text{gdy } b \leq x \leq c \\ 0 & \text{wpp} \end{cases} \quad (2.14)$$

O ile trójka (a, b, c) , gdzie $|ab| < \infty$, specyfikuje rozmyty przedział, o tyle trójki $(-\infty, b, c)$, (a, b, ∞) określają rozmyte półprzedziały. Np. gdy zbiór B ma postać $(-\infty, b, c)$, wówczas

$$\mu_B(x) = \begin{cases} 1 & \text{gdy } x \leq b \\ \frac{c-x}{c-b} & \text{gdy } b \leq x \leq c \\ 0 & \text{gdy } x \geq c \end{cases} \quad (2.15)$$

W praktyce trójkątną funkcję przynależności będziemy oznaczać symbolem $\Delta(a, b, c)$, a zbiór rozmyty scharakteryzowany taką funkcją będziemy nazywać trójkątnym zbiorem rozmytym. Podobnie, zbiór rozmyty opisany trapezoidalną funkcją przynależności będziemy nazywali trapezoidalnym zbiorem rozmytym.

Praktyka pokazuje, że wybór kształtu funkcji przynależności jest zdecydowanie mniej istotny w porównaniu z prawidłowym określeniem jej nośnika.

Instrukcja obsługi pakietu `TILShell` [30] rekomenduje co następuje:

- Rozpoczynaj od trójkątnych funkcji przynależności. Wszystkie używane funkcje powinny być symetrycznymi trójkątami o identycznej długości podstawy. Skrajny lewy trójkąt powinien mieć postać $\Delta(-\infty, a, b)$, zaś skrajny prawy – $\Delta(a, b, \infty)$.
- Ramiona sąsiadujących trójkątów powinny przecinać się w punkcie o współrzędnych $(\xi, 0.5)$. Pokrycie dziedziny każdej zmiennej powinno być tak dobierane, aby każdy jej element, z wyjątkiem skrajnych punktów, należał do (przynajmniej) dwóch zbiorów rozmytych.

2.3 Liczby i przedziały rozmyte

Z wagi na praktyczne znaczenie, odrębny punkt poświęcamy rozmytym podzbiорom zbioru liczb rzeczywistych \mathbb{R} .

Definicja 2.3.1. Liczbą rozmytą \tilde{v} nazywamy rozmyty podzbiор zbioru \mathbb{R} , spełniający następujące warunki:

- (a) $\mu_{\tilde{v}}$ jest funkcją półciągłą z góry².
- (b) \tilde{v} jest wypukłym, w sensie definicji (2.8) zbiorem rozmytym.
- (c) \tilde{v} jest normalnym zbiorem rozmytym, przy czym istnieje dokładnie jeden element $r_0 \in \mathbb{R}$, dla którego $\mu_{\tilde{v}}(r) = 1$. Liczbę r_0 nazywa się czasem wartością średnią liczby rozmytej.
- (d) Nośnik \tilde{v} jest przedziałem domkniętym, $\text{supp}(\tilde{v}) = [a, b]$. Przedział ten bywa określany przedziałem niepewności (albo ufności) liczby rozmytej. \square

W dalszym ciągu liczby rozmyte będziemy oznaczać symbolami $\tilde{u}, \tilde{v}, \dots$, lub po prostu u, v, \dots jeśli nie będzie to budzić wątpliwości. Zbiór liczb rozmytych będziemy oznaczać symbolem $\mathbb{R}_{\mathcal{F}}$.

Uwaga 2.3.1. Z definicji 2.3.1 wynikają następujące fakty:

- (i) Warunek (a) pozwala do $\mathbb{R}_{\mathcal{F}}$ zaliczyć także liczby rzeczywiste, które opisuje funkcja charakterystyczna

$$\mu_x(r) = \begin{cases} 1 & \text{gdy } r = x \\ 0 & \text{gdy } r \neq x \end{cases} \quad (2.16)$$

- (ii) Skoro nośnik liczby \tilde{v} jest zbiorem skończonym, utożsamia się go często z zerową α -warstwą liczby \tilde{v} (pamiętajmy jednak, że jest to *wygodna* konwencja), tzn.

$$[\tilde{v}]_0 = \text{supp}(\tilde{v}) \quad (2.17)$$

Innymi słowy przyjmuje się, że

$$[\tilde{v}]_{\alpha} = \begin{cases} \{x \in \mathbb{R} : \mu_{\tilde{v}}(r) \geq \alpha\} & \text{gdy } \alpha \in (0, 1] \\ \text{supp}(\tilde{v}) & \text{gdy } \alpha = 0 \end{cases} \quad (2.18)$$

- (iii) Przyjmując w warunku (c), że $\text{core}(\tilde{v}) = [c, d] \subseteq \text{supp}(\tilde{v})$, $c \neq d$, otrzymujemy definicję przedziału rozmytego. Ponownie, z warunku (a) wynika, że do zbioru $\mathbb{R}_{\mathcal{F}}$ można także zaliczyć wszystkie przedziały domknięte $[c, d]$ opisane funkcją charakterystyczną

$$\mu_{[c,d]}(r) = \begin{cases} 1 & \text{gdy } r \in [c, d] \\ 0 & \text{gdy } r \notin [c, d] \end{cases} \quad (2.19)$$

\square

Wygodnym modelem liczby rozmytej jest trójkątna liczba rozmyta $\tilde{v} \sim \Delta(a, b, c)$. W takim wypadku jej przedziałem niepewności jest zbiór $[\tilde{v}]_0 = [a, c]$, natomiast $m_{\tilde{v}} = b$ jest jej wartością średnią.

Z kolei zdefiniowane wcześniej trapezoidalne funkcje przynależności (a, b, c, d) wykorzystuje się do modelowania przedziałów rozmytych.

²Tzn. w każdym punkcie $r_0 \in \mathbb{R}$ dla dowolnego $\varepsilon > 0$ istnieje takie sąsiedztwo U punktu r_0 , że $\mu_{\tilde{v}}(r) \leq \mu_{\tilde{v}}(r_0) + \varepsilon$ dla wszystkich $r \in U$. Np. funkcja $f(x) = \lfloor x \rfloor$ jest półciągłą z góry. Por. S. Łojasiewicz: *Wstęp do teorii funkcji rzeczywistych*. Warszawa, PWN, 1973, ss. 58-63.

Uwaga 2.3.2. W praktycznych zastosowaniach występuje potrzeba operowania rozmytymi podzbiorami n -wymiarowej przestrzeni \mathbb{R}^n , por. [?]. W takim wypadku pod pojęciem $\mathbb{R}_{\mathcal{F}}^n$ rozumiemy zbiór odwzorowań postaci $\mu: \mathbb{R}^n \rightarrow [0, 1]$ spełniających uogólnione warunki (a)-(d) definicji 2.3.1, tzn. odwzorowanie μ jest półciągłe z góry, spełnia warunek wypukłości i normalności, a ponadto wszystkie α -warstwy $\{\mathbf{r} \in \mathbb{R}^n: \mu(\mathbf{r}) \geq \alpha\}$, oraz $\text{supp}(\mu)$ są zbiorami zwartymi. \square

Uwaga 2.3.3. Rozróżnianie między liczbami, a przedziałami rozmytymi ma charakter czysto techniczny. Jego istota dotyczy bowiem wyłącznie postaci jądra odpowiedniego zbioru rozmytego. Dlatego zbiór $\mathbb{R}_{\mathcal{F}}^n$ uważamy za rodzinę normalnych i wypukłych podzbiorów rozmytych \mathbb{R}^n , które mają zwarty nośnik i półciągłą z góry funkcję przynależności.

W literaturze zbiór ten bywa często oznaczany jako \mathcal{E}^n . Naszym zdaniem symbol $\mathbb{R}_{\mathcal{F}}^n$ jest bardziej intuicyjny. \square

2.4 Estymacja liczb rozmytych

Operowanie zbiorami rozmytymi stawia istotne pytanie: skąd brać funkcje przynależności? Chociaż brak jednoznacznej i ogólnie akceptowalnej odpowiedzi na to pytanie, w praktyce problem ten okazuje się nie być aż tak krytyczny.

Najprostsza metoda pozyskiwania liczb rozmytych odwołuje się do wiedzy ekspertów. Korzystanie z takiej wiedzy ma długą historię w dziedzinie badań operacyjnych. Korzysta się z niej przy szacowaniu tzw. prawdopodobieństwa subiektywnego³ czy problemach zarządzania ryzykiem. W przypadku szacowania funkcji przynależności opisującą wartość interesującej wielkości (np. czasu realizacji pewnego przedsięwzięcia, czy też jednej z czynności wymaganych do jego przeprowadzenia), można poprosić eksperta o podanie najbardziej pesymistycznej wartości \underline{a} jaką jego zdaniem owa wielkość może przyjąć, wartości najbardziej optymistycznej \bar{a} , oraz wartości najbardziej prawdopodobnej a_m . Uzyskana w ten sposób trójka $(\underline{a}, a_m, \bar{a})$ może posłużyć do zdefiniowania trójkątnej liczby rozmytej.

Jeżeli do oszacowania wspomnianej wyżej wielkości zatrudniono grupę n ekspertów, każdy z nich przedstawia swoją propozycję $(\underline{a}^i, a_m^i, \bar{a}^i)$, $i = 1, \dots, n$, a uzyskane w ten sposób wyniki uśrednia się.

Metoda bardziej rygorystyczna polega na wstępnym zebraniu próbki o liczebności n , a następnie wyznaczeniu przedziałów ufności dla różnych wartości poziomu istotności α . Przypomnijmy, [65], że jeżeli cecha X ma rozkład normalny $N(m, \sigma^2)$, przy czym odchylenie standardowe σ jest znane, a wartość oczekiwana m jest nieznana, to jej estymator będący średnią arytmetyczną \hat{X}_n z n obserwacji ma rozkład $N(m, \sigma^2/n)$. Dla danej wartości α znajdziemy liczbę ϵ_α taką, że

$$P\left(\left|\frac{\hat{X}_n - m}{\sigma/\sqrt{n}}\right| < \epsilon_\alpha\right) = 1 - \alpha$$

gdzie symbol P oznacza prawdopodobieństwo. Wzór ten można przepisać w postaci

$$P\left(\hat{X}_n - \epsilon_\alpha \frac{\sigma}{\sqrt{n}} < m < \hat{X}_n + \epsilon_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

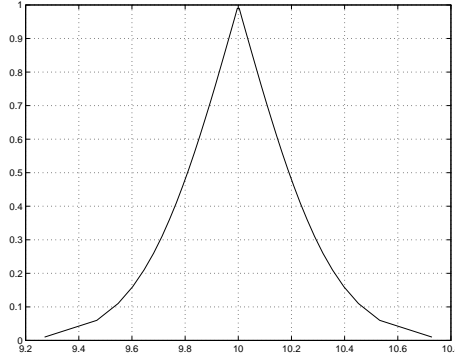
Przedział losowy

$$\left(\hat{X}_n - \epsilon_\alpha \frac{\sigma}{\sqrt{n}} < m < \hat{X}_n + \epsilon_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

³Por. D. Luce, H. Raiffa, *Gry i decyzje*, PWN, Warszawa 1964

jest przedziałem ufności dla parametru m na poziomie ufności $1-\alpha$. Ma on długość $2\epsilon_\alpha\sigma/\sqrt{n}$, która maleje ze wzrostem liczebności n i wzrostem poziomu α .

Na rysunku 2.4 przedstawiono zrekonstruowaną w ten sposób funkcję przynależności. Wartość średnią oszacowano po wykonaniu 100 obserwacji otrzymując $\hat{X}_{100} = 10$ oraz $\sigma = 2$.



Rysunek 2.4: Funkcja przynależności liczby rozmytej reprezentowana przez rodzinę przedziałów ufności.

Warto w tym miejscu odnotować, że opisana tu procedura ma charakter daleki od abstrakcyjnego pomysłu. Dała ona asumpt interesującym, probabilistycznym, interpretacjom funkcji przynależności [25].

2.5 Reprezentacja zbioru rozmytego

Reprezentacja zbioru rozmytego za pomocą funkcji przynależności nie zawsze jest wygodna. Inną możliwość gwarantuje reprezentacja parametryczna korzystająca z pojęcia α -warstwy

$$\mu_A(x) = \sup_{\alpha \in (0,1]} \min[\alpha, \chi_{[A]_\alpha}(x)] \quad (2.20)$$

Negoita i Ralescu [56] udowodnili ważne

Twierdzenie 2.5.1. (Twierdzenie o reprezentacji) Jeżeli $B_\alpha, \alpha \in [0, 1]$ jest monotoniczną rodziną zbiorów taką, że

(i) $B_{\alpha_1} \subseteq B_{\alpha_2}$ dla dowolnych $1 \geq \alpha_1 \geq \alpha_2 \geq 0$,

(ii) jeżeli $\{\alpha_n\}$ jest monotonicznie rosnącą sekwencją liczb ze zbioru $[0, 1]$ przy czym

$$\lim_{n \rightarrow \infty} \alpha_n = \alpha \text{ to } B_\alpha = \bigcap_{n=1}^{\infty} B_{\alpha_n}$$

to istnieje zbiór rozmyty \tilde{A} taki, że $[\tilde{A}]_\alpha = B_\alpha$ dla każdej wartości $\alpha \in [0, 1]$. Ponadto

$$\mu_{\tilde{A}}(x) = \begin{cases} \sup\{\alpha : x \in B_\alpha\} & \text{gdy } x \in B_0 \\ 0 & \text{wpp} \end{cases} \quad (2.21)$$

□

W inżynierskich zastosowaniach podstawową rolę odgrywa parametryczna (przedziałowa) reprezentacja liczb rozmytych postaci

$$[u]_\alpha = [\underline{u}(\alpha), \bar{u}(\alpha)] \quad (2.22)$$

gdzie u jest liczbą rozmytą o funkcji przynależności $\mu_u: \mathbb{R} \rightarrow [0, 1]$, natomiast $\underline{u}: (0, 1] \rightarrow \mathbb{R}$ i $\bar{u}: [0, 1] \rightarrow \mathbb{R}$ są odwzorowaniami takimi, że

- (a) \underline{u} jest monotonicznie rosnącą (niemalejącą) i lewostronnie ciągłą funkcją w przedziale $[0, 1]$,
- (b) \bar{u} jest monotonicznie malejącą (nierosnącą) i lewostronnie ciągłą funkcją w przedziale $[0, 1]$,
- (c) $\underline{u}(\alpha) \leq \bar{u}(\alpha)$ dla każdej wartości $\alpha \in [0, 1]$

W szczególności, gdy $\underline{u}(\alpha) = \bar{u}(\alpha)$ dla każdej wartości $\alpha \in [0, 1]$, wówczas u jest liczbą rzeczywistą. Jeżeli $\underline{u}(1) < \bar{u}(1)$, to u jest rozmytym przedziałem, a jeżeli $\underline{u}(1) = \bar{u}(1)$, to u jest liczbą rozmytą o wartości średniej $x_0 = \underline{u}(1) = \bar{u}(1)$.

Uwaga 2.5.1. Często, szczególnie wówczas, gdy dolne i górne ograniczenie jest funkcją innych (oprócz α) zmiennych, zamiast $[\underline{u}(\alpha), \bar{u}(\alpha)]$ będziemy stosować krótszą notację $[\underline{u}_\alpha, \bar{u}_\alpha]$. \square

Przykład 2.5.1. Niech $u = \Delta(a, b, c)$ będzie trójkątną liczbą rozmytą. Tutaj

$$\begin{aligned}\underline{u}(\alpha) &= a + \alpha(b - a) \\ \bar{u}(\alpha) &= c - \alpha(c - b)\end{aligned}$$

a zatem

$$[u]_\alpha = [a + \alpha(b - a), c - \alpha(c - b)]$$

Przedział $[a, c]$ jest przedziałem nieokreśloności (ufności) trójkątnej liczby u , a $\underline{u}(1) = \bar{u}(1) = b$ jest jej wartością średnią.

W przypadku trapezoidalnych przedziałów rozmytych $v = (a, b, c, d)$ mamy

$$[v]_\alpha = [a + \alpha(b - a), d - \alpha(d - c)]$$

i oczywiście $\underline{v}(1) = b < c = \bar{v}(1)$. \square

Jeżeli $\tilde{v} \in \mathbb{R}_{\mathcal{F}}$ jest liczbą rozmytą, to wartość

$$W(\tilde{v}) = \frac{1}{2} \int_0^1 [\underline{v}(\alpha) + \bar{v}(\alpha)] d\alpha \quad (2.23)$$

można traktować jako jej wyostrzenie⁴. Przykładowo, jeżeli $v = (a, b, c, d)$ jest rozmytym przedziałem, to

$$W(v) = \frac{1}{4}(a + b + c + d)$$

natomiast w przypadku trójkątnej liczby rozmytej $u = \Delta(a, b, c)$

$$W(u) = \frac{1}{4}(a + 2b + c)$$

Jeżeli u jest symetryczną liczbą trójkątną, tzn. $b = (a + c)/2$, to $W(u) = (a + c)/2$.

Wyostrzenie (2.23) pozwala zdefiniować relację porządku w zbiorze $\mathbb{R}_{\mathcal{F}}$. Mianowicie

$$\tilde{u} \leq_W \tilde{v} \quad \text{jeżeli} \quad W(\tilde{u}) \leq W(\tilde{v}) \quad (2.24)$$

Własności wyostrzenia W zebrano w poniższym lemacie⁵ (występujące tu operacje na liczbach rozmytych definiujemy w punkcie 3.2)

⁴Por. P. Fortemps, M. Roubens. Ranking and defuzzification methods based on area compensation, *Fuzzy Sets Syst.* **82**, 1996, 319–330. Problem wyostrzenia zbioru rozmytego, a więc zastępowania go pojedynczym skłarem omawiamy w rozdziale dotyczącym sterowników rozmytych.

⁵Por. S. Abbasbandy, M. Amirfakhrian. A new approach to universal approximation of fuzzy functions on a discrete set of points. *Applied Mathematical Modelling*, **30**, 2006, 1525–1534.

Lemat 2.5.1. Jeżeli $\tilde{u}, \tilde{v}, \tilde{w} \in \mathbb{R}_{\mathcal{F}}$, oraz $\alpha \in \mathbb{R}$, to:

- (a) $W(-\tilde{u}) = -W(\tilde{u})$
- (b) $W(\tilde{u} + \tilde{v}) = W(\tilde{u}) + W(\tilde{v})$
- (c) $W(\alpha\tilde{u}) = \alpha W(\tilde{u})$
- (d) jeżeli $\tilde{u} \leq_W \tilde{v} + \tilde{w}$, to $\tilde{u} - \tilde{w} \leq_W \tilde{v}$
- (e) jeżeli $\tilde{u} \leq_W \tilde{w}$, to $\tilde{u} + \tilde{w} \leq_W \tilde{v} + \tilde{w}$
- (f) jeżeli $\tilde{u} \leq_W \tilde{v}$, $-\tilde{v} \leq_W -\tilde{u}$

□

2.6 Zadania

Problem 2.6.1. Alternatywna reprezentacja trójkątnej funkcji przynależności ma postać

$$\mu_A(x) = \begin{cases} 1 - \frac{a-x}{\alpha} & \text{gdy } a - \alpha \leq x \leq a \\ 1 - \frac{x-a}{\beta} & \text{gdy } a \leq x \leq a + \beta \\ 0 & \text{wpp} \end{cases} \quad (2.25)$$

Oznaczmy zbiór rozmyty o takiej funkcji przynależności symbolem $\Delta'(a; \alpha, \beta)$.

Podać interpretację parametrów a, α, β . Narysować funkcję $\Delta'(5; 2, 3)$. Podać reprezentację parametryczną tego zbioru. Obliczyć jego nośnik.

Problem 2.6.2. Uogólnić reprezentację (2.25) na przypadek trapezoidalnych funkcji przynależności.

Problem 2.6.3. Dla trapezoidalnego zbioru rozmytego \tilde{A} postaci $(2, 4, 6, 9)$ wyznaczyć:

- (a) analityczną postać funkcji przynależności $\mu_{\tilde{A}}$,
- (b) dopełnienie \tilde{A}^c zbioru \tilde{A} , tzn. podać analityczną postać funkcji przynależności $\mu_{\tilde{A}^c}$ i sporządzić wykres tej funkcji, oraz
- (c) narysować wykresy funkcji przynależności dla zbiorów $\tilde{A} \cup \tilde{A}^c$ i $\tilde{A} \cap \tilde{A}^c$.

Jak wyglądają α -warstwy zbioru \tilde{A}^c ? Jak można skomentować wyniki z punktu (c)?

Problem 2.6.4. Zapisać liczbę rzeczywistą $r = 4$ w postaci liczby rozmytej \tilde{r} o trójkątnej funkcji przynależności. Wyznaczyć α -warstwy tej liczby.

Rozdział 3

Zasada rozszerzania i jej zastosowania

Zaproponowana przez Zadeha [84] zasada rozszerzania jest jednym z podstawowych narzędzi teorii zbiorów rozmytych. Pozwalając wyznaczać obraz zbioru rozmytego w klasycznym, tzn. nierozmytym, odwzorowaniu f , umożliwia ona dostosowywanie klasycznych pojęć matematycznych, takich jak operacje arytmetyczne, funkcje wielu zmiennych, itd., do potrzeb przetwarzania zbiorów, a w szczególności – liczb rozmytych. W niniejszym rozdziale omawiamy klasyczne sformułowanie tej zasady, wraz z metodami umożliwiającymi jej efektywne stosowanie w rozwiązywaniu praktycznych problemów. Warto uprzedzić Czytelnika, że istnieją także inne sformułowania zasady rozszerzania, które odwołują się do różnych interpretacji operatorów logicznych; szerzej temat ten jest omawiany np. w [16]. Jednakże ogólna idea pozostaje wszędzie taka sama. Pewną odległą analogią dla zasady rozszerzania jest metoda definiowania zmiennych losowych dla danej przestrzeni probabilistycznej.

3.1 Idea rozszerzania

Niech $f: \mathbb{R} \rightarrow \mathbb{R}$ będzie funkcją rzeczywistą zmiennej rzeczywistej. Obrazem liczby rozmytej \tilde{x} w tym odwzorowaniu jest liczba rozmyta $f(\tilde{x})$ o funkcji przynależności [84]

$$\mu_{f(\tilde{x})}(y) = \begin{cases} \sup_{\xi \in f^{-1}(y)} \mu_{\tilde{x}}(\xi) & \text{gdy } f^{-1}(y) \neq \emptyset \\ 0 & \text{wpp} \end{cases} \quad (3.1)$$

Jeżeli f jest funkcją ściśle monotoniczną, to obrazem zbioru \tilde{x} w odwzorowaniu f jest

$$\mu_{f(\tilde{x})}(y) = \begin{cases} \mu_{\tilde{x}}(f^{-1}(y)) & \text{gdy } f^{-1}(y) \neq \emptyset \\ 0 & \text{wpp} \end{cases} \quad (3.2)$$

Powyższa konstrukcja pokazuje, że funkcja f indukuje rozmyte odwzorowanie $F: \mathbb{R}_{\mathcal{F}} \rightarrow \mathbb{R}_{\mathcal{F}}$ o dobrze określonych własnościach formalnych [56]. Zauważmy w szczególności, że spełniona jest następująca nierówność

$$\forall \xi \in X, \mu_{f(\tilde{x})}(f(\xi)) \geq \mu_{\tilde{x}}(\xi) \quad (3.3)$$

gdzie X jest dziedziną funkcji f .

Przykład 3.1.1. Niech $f(x) = \lambda x$, $\lambda \neq 0$. Jeżeli \tilde{x} jest liczbą rozmytą o funkcji przynależności $\mu_{\tilde{x}}$, to

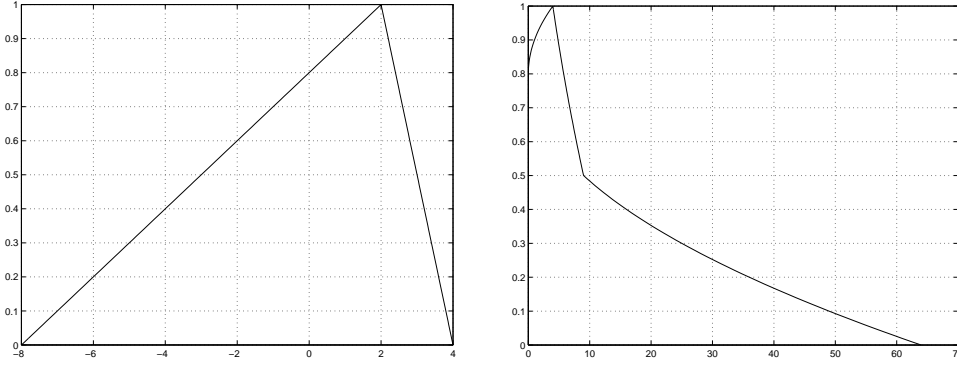
$$\mu_{f(\tilde{x})}(y) = \sup_{\lambda\xi=y} \mu_{\tilde{x}}(\xi) = \mu_{\tilde{x}}(y/\lambda)$$

W szczególności, gdy $\lambda = -1$ to obrazem liczby \tilde{x} jest przeciwna do niej liczba $-\tilde{x}$ o parametrycznej reprezentacji $[-\bar{x}_\alpha, -\underline{x}_\alpha]$. \square

Przykład 3.1.2. Jeżeli $f(x) = x^2$, a \tilde{x} jest liczbą rozmytą o funkcji przynależności $\mu_{\tilde{x}}$, to

$$\begin{aligned} \mu_{f(\tilde{x})}(y) &= \begin{cases} \sup_{\xi^2=y} \mu_{\tilde{x}}(\xi) & \text{gdy } y \geq 0 \\ 0 & \text{wpp} \end{cases} \\ &= \begin{cases} \max[\mu_{\tilde{x}}(-\sqrt{y}), \mu_{\tilde{x}}(\sqrt{y})] & \text{gdy } y \geq 0 \\ 0 & \text{wpp} \end{cases} \end{aligned} \quad (3.4)$$

Na rysunku 3.1 przedstawiono obraz liczby $\Delta(-8, 2, 4)$ w powyższym przekształceniu. \square



Rysunek 3.1: Rozmyta liczba $\Delta(-8, 2, 4)$ i jej obraz w przekształceniu $f(x) = x^2$.

Twierdzenie 3.1.1. Odwzorowanie $F: \mathbb{R}_{\mathcal{F}} \rightarrow \mathbb{R}_{\mathcal{F}}$ indukowane przez funkcję f zgodnie z regułą (3.6) jest monotonicznie rosnące, tzn.

$$(\forall \tilde{x}_1, \tilde{x}_2 \in \mathbb{R}_{\mathcal{F}})(\tilde{x}_1 \subset \tilde{x}_2 \Rightarrow F(\tilde{x}_1) \subset F(\tilde{x}_2)) \quad (3.5)$$

Dowód: Jeżeli \tilde{x}_1, \tilde{x}_2 są dwiema liczbami rozmytymi takimi, że $\tilde{x}_1 \subset \tilde{x}_2$, to zgodnie z definicją (3.2) otrzymujemy

$$\mu_{f(\tilde{x}_1)}(y) = \sup_{\xi \in f^{-1}(y)} \mu_{\tilde{x}_1}(\xi) \leq \sup_{\xi \in f^{-1}(y)} \mu_{\tilde{x}_2}(\xi) = \mu_{f(\tilde{x}_2)}(y)$$

dla dowolnej wartości y należącej do przeciwdziedziny funkcji f . \square

Niech teraz $f: \mathbb{R}^n \rightarrow \mathbb{R}$ będzie funkcją rzeczywistą n zmiennych. Zastępując rzeczywiste wartości x_1, \dots, x_n liczbami rozmytymi $\tilde{x}_1, \dots, \tilde{x}_n$, otrzymamy pewną liczbę rozmytą, a zatem – jak poprzednio – można przyjąć, że funkcja f indukuje rozmyte odwzorowanie $F: \mathbb{R}_{\mathcal{F}}^n \rightarrow \mathbb{R}_{\mathcal{F}}$. Funkcję przynależności zbioru rozmytego $\tilde{B} = F(\tilde{x}_1, \dots, \tilde{x}_n)$ wyznaczamy z uogólnionej zasady (3.1) postaci [84]

$$\mu_{\tilde{B}}(y) = \begin{cases} \sup_{(\xi_1, \dots, \xi_n) \in f^{-1}(y)} \min[\mu_{\tilde{x}_1}(\xi_1), \dots, \mu_{\tilde{x}_n}(\xi_n)] & \text{gdy } f^{-1}(y) \neq \emptyset \\ 0 & \text{wpp} \end{cases} \quad (3.6)$$

Uwaga 3.1.1. W ogólnym przypadku rozważa się odwzorowanie $f: X_1 \times \cdots \times X_n \rightarrow Y$, gdzie X_1, \dots, X_n oznaczają dziedziny odpowiednich argumentów, a Y jest zbiorem wartości funkcji f . Interesuje nas wówczas rozszerzenie postaci $F: \mathcal{F}(X_1) \times \cdots \times \mathcal{F}(X_n) \rightarrow \mathcal{F}(Y)$, gdzie $\mathcal{F}(X_i), i = 1, \dots, n$, oraz $\mathcal{F}(Y)$ są rodzinami rozmytych podzbiorów odpowiednich przestrzeni. W ten sposób można np. konstruować rozmyte relacje porządku na pewnej przestrzeni $\mathcal{F}(X)$, w szczególności na zbiorze liczb rozmytych $\mathbb{R}_{\mathcal{F}}$, czy też bardziej zaawansowane struktury algebraiczne – por. [16], [56]. \square

Jedną z podstawowych własności indukowanych w powyższy sposób zbiorów rozmytych jest zachowanie zgodności odpowiednich α -warstw, tzn.

Twierdzenie 3.1.2. Jeżeli $f: X_1 \times \cdots \times X_n \rightarrow Y$ jest funkcją ciągłą, to F jest dobrze określonym odwzorowaniem rozmytym, przy czym

$$[F(\tilde{x}_1, \dots, \tilde{x}_n)]_{\alpha} = f([\tilde{x}_1]_{\alpha}, \dots, [\tilde{x}_n]_{\alpha}) \quad (3.7)$$

gdzie $f(A) = \{f(a) : a \in A\}$ dla dowolnego ostrego zbioru A . \square

Twierdzenie to, udowodnione przez Nguyena [58], odgrywa ważną rolę w praktycznych zastosowaniach zasady rozszerzania.

Przykład 3.1.3. (Rozmyta odległość między zbiorami rozmytymi) Niech X będzie przestrzenią metryczną z (pseudo-)metryką d , tzn. odwzorowaniem $d: X \times X \rightarrow \mathbb{R}_+ \cup \{0\}$ spełniającym dla dowolnych $p, q, r \in \mathbb{R}$ warunki

- (i) $d(p, q) = 0$, gdy $p = q$,
- (ii) $d(p, q) = d(q, p)$,
- (iii) $d(p, r) \leq d(p, q) + d(q, r)$.

Stosując zasadę rozszerzania, odległość $D(\tilde{A}, \tilde{B})$ między zbiorami rozmytymi \tilde{A}, \tilde{B} definiuje się następująco

$$\mu_{D(\tilde{A}, \tilde{B})}(\delta) = \sup_{\substack{(p, q) \in X \\ d(p, q) = \delta}} \min[\mu_{\tilde{A}}(p), \mu_{\tilde{B}}(q)]$$

Korzystając z parametrycznej reprezentacji przepisujemy powyższą równość w postaci

$$\mu_{D(\tilde{A}, \tilde{B})}(\delta) = \sup\{\alpha \in (0, 1] : \delta \in d([\tilde{A}]_{\alpha}, [\tilde{B}]_{\alpha})\}$$

Tak zdefiniowana funkcja D modeluje odległość między „plamami” \tilde{A} i \tilde{B} . Gdy są to zbiory wypukłe, $D(\tilde{A}, \tilde{B})$ jest przedziałem, którego krańce reprezentują najmniejszą i największą odległość między punktami $p \in \text{supp}(\tilde{A})$ oraz $q \in \text{supp}(\tilde{B})$. \square

W ogólnym przypadku reguła (3.6) reprezentuje sobą pewne zadanie programowania nieliniowego, którego rozwiązanie wiąże się ze znacznymi kosztami. Metody przybliżonego rozwiązywania tego zadania omawiamy w punkcie 3.3. W pewnych jednak przypadkach zadanie to można rozwiązać w prosty sposób, co ilustruje kolejny podrozdział.

3.2 Rozmyta arytmetyka

Załóżmy, że $u, v \in \mathbb{R}_{\mathcal{F}}$ są dwiema liczbami rozmytymi, natomiast \square oznacza jedną z czterech operacji arytmetycznych. Funkcja przynależności liczby rozmytej $w = u \square v$ ma postać

$$\mu_w(z) = \sup_{z = x \square y} \min\{\mu_u(x), \mu_v(y)\} \quad (3.8)$$

Pamiętając o skróconej konwencji oznaczania zbiorów rozmytych (por. punkt (c) na stronie 13) powyższe równanie można przepisać w postaci

$$w(z) = \sup_{z=x \square y} \min\{u(x), v(y)\} \quad (3.9)$$

W ogólności istnieje niemal nieskończona liczba kombinacji liczb x, y gwarantujących wynik z ; jeżeli np. \square jest operacją mnożenia, to każda para $(x, z/x), x \neq 0$ daje w wyniku liczbę z i praktyczne stosowanie zasady rozszerzania jest raczej kłopotliwe.

Aby móc efektywnie wyznaczyć funkcję przynależności charakteryzującą wynik, stosuje się albo dekompozycję odpowiednich zbiorów rozmytych zadaną równaniem (2.21), albo też dokonuje się dyskretyzacji ciągłych funkcji przynależności – por. [27] lub [16]. W istocie obie te metody prowadzą do identycznych wyników.

Przykład 3.2.1. Przypuśćmy, że u, v są dwiema trójkątnymi liczbami rozmytymi, powiedzmy $u = \Delta(1, 4, 7), v = \Delta(2, 5, 8)$. Niech $s_1 = [1, 7], s_2 = [2, 8]$ będą nośnikami tych liczb. Podzielny każdy nośnik na m równych podprzedziałów i oznaczmy $x_i = 1 + i * h_u$, $y_j = 2 + j * h_v$, gdzie $h_u = s_1/m$, $h_v = s_2/m$. W przypadku, gdy \square jest operacją mnożenia, wprowadzimy symbol \odot na oznaczenie mnożenia liczb rozmytych. Wówczas

$$(u \odot v)(z) = \max_{x_i \cdot y_j = z} \min\{u(x_i), v(y_j)\} = \max_{i=1, \dots, m} \min\{u(x_i), v(z/y_j)\}$$

Zauważmy, że tutaj zmienna z przyjmuje skończoną liczbę wartości nie przekraczającą m^2 . Można sprawdzić, że w tym konkretnym przykładzie liczba różnych wartości jest równa 28. Wynikową funkcję przynależności przedstawiono w lewej części rysunku 3.2. W prawej części tego rysunku pokazano rezultat uzyskany przy zwiększeniu liczby przedziałów do $m = 80$.

Jak widać w trakcie obliczeń otrzymuje się wiele punktów stanowiących elementy poszczególnych α -warstw. Ich rola jest raczej nieistotna, interesuje nas bowiem wyłącznie obwiednia reprezentująca skrajne punkty każdej α -warstwy. Uzyskanie obwiedni jest stosunkowo proste: w zbiorze punktów o identycznej wartości funkcji przynależności należy wyróżnić skrajne punkty. Taki sposób postępowania narzuca inną organizację obliczeń. Można podzielić odcinek $[0, 1]$ na m podprzedziałów, co wymaga określenia $m + 1$ wartości $0 = \alpha_0 < \alpha_1 < \dots < \alpha_m - 1 < \alpha_m = 1$, a następnie należy wybrać punkty odpowiadające krańcom poszczególnych α -warstw. Przemieszczając się od warstwy o najwyższej wartości $\alpha = 1$ do warstwy o najniższej wartości α wykonuje się stosowne obliczenia.

W przypadku, gdy korzystamy z α -warstw to dolne i górne ograniczenia obu argumentów mają postać $\underline{u}(\alpha) = 1 + 3\alpha$, $\bar{u}(\alpha) = 7 - 3\alpha$, $\underline{v}(\alpha) = 2 + 3\alpha$, $\bar{v}(\alpha) = 8 - 3\alpha$. Można sprawdzić, że $\underline{z}(\alpha) = \underline{u}(\alpha)\underline{v}(\alpha)$ oraz $\bar{z}(\alpha) = \bar{u}(\alpha)\bar{v}(\alpha)$ wyznaczają monotoniczną rodzinę zbiorów, którą w oparciu o twierdzenie o reprezentacji można potraktować jako rodzinę α -warstw liczby rozmytej $z = u \odot v$. \square

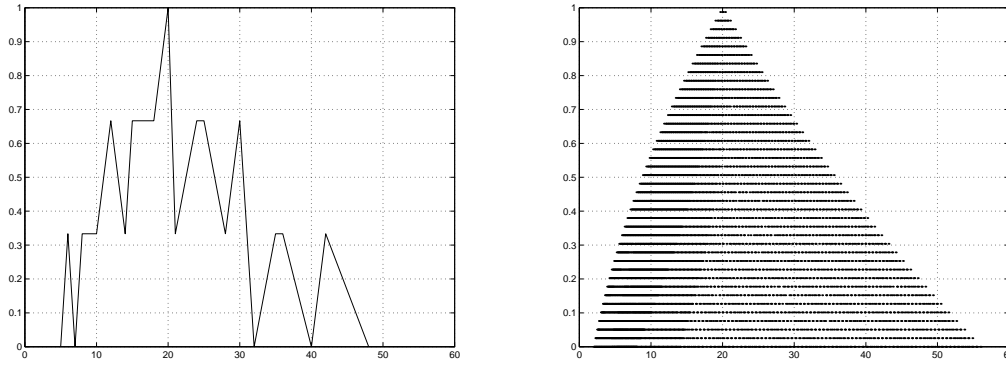
Z uwagi na to, że operacje arytmetyczne są monotoniczne, zasada rozszerzania prowadzi do prostych wyników. Mianowicie, jeżeli u, v są dwiema liczbami rozmytymi, to α -warstwy liczby $u \square v$ gdzie $\square \in \{+, -, \cdot, : \}$ mają następującą postać:

(a) dodawanie:

$$[u + v]_\alpha = [\underline{u}(\alpha) + \underline{v}(\alpha), \bar{u}(\alpha) + \bar{v}(\alpha)] \quad (3.10)$$

(b) odejmowanie:

$$[u - v]_\alpha = [\underline{u}(\alpha) - \bar{v}(\alpha), \bar{u}(\alpha) - \underline{v}(\alpha)] \quad (3.11)$$



Rysunek 3.2: Funkcja przynależności iloczynu dwóch liczb rozmytych. Nośniki obu zbiorów podzielono na: 7 części (lewy rysunek) oraz (80 części)

(c) mnożenie:

$$[uv]_{\alpha} = [\underline{uv}(\alpha), \overline{uv}(\alpha)] \quad (3.12)$$

przy czym

$$\begin{aligned} \underline{uv}(\alpha) &= \min\{\underline{u}(\alpha)\underline{v}(\alpha), \underline{u}(\alpha)\overline{v}(\alpha), \overline{u}(\alpha)\underline{v}(\alpha), \overline{u}(\alpha)\overline{v}(\alpha)\} \\ \overline{uv}(\alpha) &= \max\{\underline{u}(\alpha)\underline{v}(\alpha), \underline{u}(\alpha)\overline{v}(\alpha), \overline{u}(\alpha)\underline{v}(\alpha), \overline{u}(\alpha)\overline{v}(\alpha)\} \end{aligned}$$

(d) dzielenie: jeżeli $0 \notin [v(0), \overline{v}(0)]$, to

$$\left[\frac{u}{v}\right]_{\alpha} = [\underline{u/v}(\alpha), \overline{u/v}(\alpha)] \quad (3.13)$$

przy czym

$$\begin{aligned} \underline{u/v}(\alpha) &= \min\left\{\frac{\underline{u}(\alpha)}{\underline{v}(\alpha)}, \frac{\underline{u}(\alpha)}{\overline{v}(\alpha)}, \frac{\overline{u}(\alpha)}{\underline{v}(\alpha)}, \frac{\overline{u}(\alpha)}{\overline{v}(\alpha)}\right\} \\ \overline{u/v}(\alpha) &= \max\left\{\frac{\underline{u}(\alpha)}{\underline{v}(\alpha)}, \frac{\underline{u}(\alpha)}{\overline{v}(\alpha)}, \frac{\overline{u}(\alpha)}{\underline{v}(\alpha)}, \frac{\overline{u}(\alpha)}{\overline{v}(\alpha)}\right\} \end{aligned}$$

W szczególności, gdy v jest skalar, oznaczmy go przez k , to równanie (3.12) przyjmuje postać

$$[ku]_{\alpha} = \begin{cases} [k\underline{u}(\alpha), k\overline{u}(\alpha)] & \text{gdy } k \geq 0 \\ [k\overline{u}(\alpha), k\underline{u}(\alpha)] & \text{gdy } k < 0 \end{cases} \quad (3.14)$$

Zauważmy, że w przypadku liczb rozmytych $\tilde{a} - \tilde{a} \neq 0$ jak również $\tilde{a}/\tilde{a} \neq 1$. Mamy natomiast

$$\begin{aligned} [\tilde{a} - \tilde{a}]_{\alpha} &= [\underline{a}(\alpha) - \overline{a}(\alpha), \overline{a}(\alpha) - \underline{a}(\alpha)] \\ \left[\frac{\tilde{a}}{\tilde{a}}\right]_{\alpha} &= \left[\frac{\underline{a}(\alpha)}{\overline{a}(\alpha)}, \frac{\overline{a}(\alpha)}{\underline{a}(\alpha)}\right] \end{aligned} \quad (3.15)$$

W szczególności parametrami liczby $\tilde{r} = \tilde{a} - \tilde{a}$ są: wartość średnia $m_{\tilde{r}} = 0$ oraz średnica $diam(\tilde{r}) = 2diam(\tilde{a})$. Oznacza to, że wartość średnia oraz średnica liczby rozmytej zachowują się podobnie jak wartość średnia $m_R = \mathbb{E}(X - Y)$ i odchylenie standardowe $\sigma_R = \sqrt{\mathbb{V}(X - Y)}$ różnicy niezależnych zmiennych losowych: $m_R = m_X - m_Y$, $\sigma_R = \sigma_X + \sigma_Y$, pod warunkiem, że $X = Y$.

Analogia ta idzie nieco dalej: można sprawdzić, że obrazem liczby rozmytej \tilde{x} o parametrach $m_{\tilde{x}}$ oraz $diam(\tilde{x})$ w przekształceniu liniowym $f(x) = ax + b$ jest liczba \tilde{y} o parametrach

$$m_{\tilde{y}} = a \cdot m_{\tilde{x}} + b, \quad \text{diam}(\tilde{y}) = a \cdot \text{diam}(\tilde{x}) \quad (3.16)$$

gdyż, parametrami zmiennej losowej $Y = aX + b$ są: $m_Y = a \cdot m_X + b$ oraz $\sigma_Y = a \cdot \sigma_X$.

Dalsze analogie z rachunkiem prawdopodobieństwa załamują się. W przypadku liczb postaci $\tilde{q} = \tilde{a}/\tilde{a}$ mamy bowiem $m_{\tilde{q}} = 1$, oraz, gdy $\underline{a}(0) \cdot \bar{a}(0) > 0$

$$\text{diam}(\tilde{q}) = \frac{\bar{a}^2(0) - \underline{a}^2(0)}{\underline{a}(0)\bar{a}(0)} = \frac{\text{diam}(\tilde{a})(\underline{a}(0) + \bar{a}(0))}{\underline{a}(0)\bar{a}(0)}$$

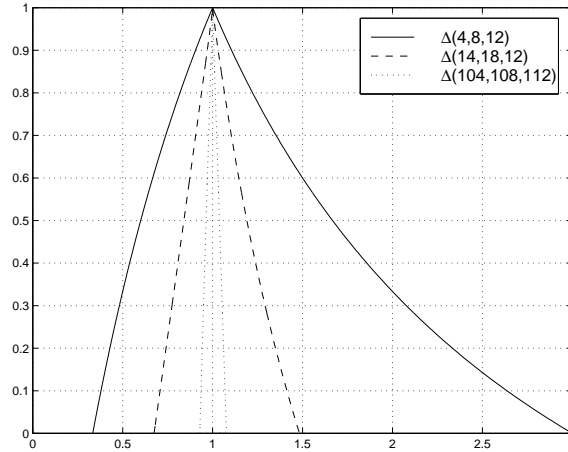
skąd wynika m.in., że gdy $\underline{a}(0) \rightarrow \bar{a}(0)$, to $\text{diam}(\tilde{q}) \rightarrow \infty$.

Jeżeli $\underline{a}(0) \cdot \bar{a}(0) < 0$, to – jak wynika z poprzednich rozważań – α -warstwy liczby \tilde{q} mają postać

$$\begin{aligned} \underline{q}(\alpha) &= \min[\underline{a}(\alpha)/m_{\tilde{a}}, \bar{a}(\alpha)/m_{\tilde{a}}] \\ \bar{q}(\alpha) &= \max[\underline{a}(\alpha)/m_{\tilde{a}}, \bar{a}(\alpha)/m_{\tilde{a}}] \end{aligned}$$

skąd wynika m.in., że $\text{diam}(\tilde{q}) \rightarrow \infty$ gdy $m_{\tilde{a}} \rightarrow 0$.

Na rysunku 3.3 pokazano wartość ilorazu $\tilde{q}_i = \tilde{a}_i/\tilde{a}_i$ dla trzech trójkątnych liczb rozmytych: $\tilde{a}_1 \sim \Delta(4, 8, 12)$, $\tilde{a}_2 \sim \Delta(14, 18, 12)$, oraz $\tilde{a}_3 \sim \Delta(104, 108, 112)$. Chociaż $\text{diam}(\tilde{a}_1) = \text{diam}(\tilde{a}_2) = \text{diam}(\tilde{a}_3) = 8$, to średnica liczb \tilde{q}_i maleje ze wzrostem wartości $\underline{a}_i(0)$: $\text{diam}(\tilde{q}_1) = 2.6667$, $\text{diam}(\tilde{q}_2) = 0.8084$, $\text{diam}(\tilde{q}_3) = 0.0787$.



Rysunek 3.3: Wartość ilorazu $\tilde{q}_i = \tilde{a}_i/\tilde{a}_i$ dla trzech trójkątnych liczb rozmytych: $\tilde{a}_1 \sim \Delta(4, 8, 12)$, $\tilde{a}_2 \sim \Delta(14, 18, 12)$, oraz $\tilde{a}_3 \sim \Delta(104, 108, 112)$

Natomiast postać rozkładu zmiennej losowej $Z = X/Y$, gdzie X, Y są niezależnymi zmiennymi o identycznym rozkładzie, zależy od rozkładu zmiennej X . Np. jeżeli $X \sim N(0, 1)$, to Z ma (standardowy) rozkład Cauchy'ego charakteryzujący się tym, że jego wartość średnia jest nieskończona, a wariancja nie istnieje.

3.3 Metody numeryczne dla zasady rozszerzania

W ogólnym przypadku bezpośrednie stosowanie zasady rozszerzania jest dosyć kłopotliwe – musimy wielokrotnie rozwiązywać zadanie programowania nieliniowego. Z kolei stosowanie omówionej w poprzednim punkcie reprezentacji przedziałowej prowadzi do niejednoznaczności wyników, co ilustruje poniższy

Przykład 3.3.1. [26]. Wyznaczmy wartość funkcji

$$f(x) = x^3 - x^2 - 17x - 15 \quad (3.17)$$

w rozmytym punkcie $\tilde{x} = [4, 4.5, 5]$. Stosując bezpośrednio zasady rozmytej arytmetyki dostajemy np.

$$\begin{aligned} \text{supp}(f(\tilde{x})) &= [4, 5]^3 - [4, 5]^2 - 17[4, 5] - 15 \\ &= [64, 125] + [-25, -16] + [-85, -68] - 15 \\ &= [-61, 26] \end{aligned}$$

Jeżeli do obliczeń wykorzysta się schemat Hornera

$$f(\tilde{x}) = ((\tilde{x} - 1)\tilde{x} - 17)\tilde{x} - 15$$

to otrzyma się

$$\begin{aligned} \text{supp}(f(\tilde{x})) &= (([4, 5] - 1) \cdot [4, 5] - 17) \cdot [4, 5] - 15 \\ &= ([3, 4] \cdot [4, 5] - 17) \cdot [4, 5] - 15 \\ &= ([12, 20] - 17) \cdot [4, 5] - 15 = [-5, 3] \cdot [4, 5] - 15 \\ &= [-25, 15] - 15 = [-40, 0] \end{aligned}$$

Natomiast przedstawiając prawą stronę wyrażenia (3.17) w postaci

$$f(\tilde{x}) = (\tilde{x} + 3)(\tilde{x} + 1)(\tilde{x} - 5)$$

to otrzymamy

$$\begin{aligned} \text{supp}(f(\tilde{x})) &= ([4, 5] + 3)([4, 5] + 1)([4, 5] - 5) \\ &= [7, 8] \cdot [5, 6] \cdot [-1, 0] \\ &= [35, 48] \cdot [-1, 0] = [-48, 0] \end{aligned}$$

Tymczasem poprawny wynik to $[-35, 0]$. □

3.3.1 Metoda wierzchołkowa

Korzystając z reprezentacji obrazu zbioru rozmytego poprzez α -warstwy (por. twierdzenie 3.1.2), Dong i Shah [15] zauważyli, że w przypadku funkcji jednej zmiennej $f(\tilde{u})$, gdzie \tilde{u} jest liczbą rozmytą, krańce odpowiedniej α -warstwy są równe

$$[f(\tilde{u})]_\alpha = [\min\{f(\underline{u}_\alpha), f(\bar{u}_\alpha)\}, \max\{f(\underline{u}_\alpha), f(\bar{u}_\alpha)\}] \quad (3.18)$$

gdzie $\underline{u}_\alpha = \underline{u}(\alpha)$, $\bar{u}_\alpha = \bar{u}(\alpha)$ to krańce α -warstwy $[\tilde{u}]_\alpha$.

Równanie to jest prawdziwe pod warunkiem, że f jest funkcją ciągłą i ściśle monotoniczną (a więc nie posiadającą ekstremów lokalnych) w zbiorze $\text{supp}(\tilde{u})$. Jako przykład może tu służyć funkcja (3.17), która na odcinku $[4, 5]$ jest ściśle rosnąca. W tym wypadku, gdy $\tilde{u} = \Delta(4, 4.5, 5)$ to

$$\text{supp}(f(\tilde{u})) = [\min\{f(4), f(5)\}, \max\{f(4), f(5)\}] = [-35, 0]$$

Jeżeli f jest funkcją n zmiennych, należy wyznaczyć jej wartości we wszystkich $N = 2^n$ punktach wierzchołkowych w_j kostki $[\underline{u}_{1,\alpha}, \bar{u}_{1,\alpha}] \times \dots \times [\underline{u}_{n,\alpha}, \bar{u}_{n,\alpha}]$, tzn.

$$[f(\tilde{u})]_\alpha = [\min_{j=1,\dots,N} f(w_j), \max_{j=1,\dots,N} f(w_j)] \quad (3.19)$$

Na przykład, gdy \tilde{u}, \tilde{v} są rozmytymi argumentami funkcji dwóch zmiennych $h(x, y)$, to $N = 4$ oraz

$$[h(\tilde{u}, \tilde{v})]_\alpha = \left[\min[h(\underline{u}_\alpha, \underline{v}_\alpha), h(\underline{u}_\alpha, \bar{v}_\alpha), h(\bar{u}_\alpha, \underline{v}_\alpha), h(\bar{u}_\alpha, \bar{v}_\alpha)], \right. \\ \left. \max[h(\underline{u}_\alpha, \underline{v}_\alpha), h(\underline{u}_\alpha, \bar{v}_\alpha), h(\bar{u}_\alpha, \underline{v}_\alpha), h(\bar{u}_\alpha, \bar{v}_\alpha)] \right]$$

Jeżeli natomiast funkcja f nie jest monotoniczna i w interesującym nas przedziale osiąga m ekstremów w punktach e_k , $k = 1, \dots, m$, to punkty te należy potraktować jako dodatkowe punkty wierzchołkowe i zmodyfikować równanie (3.19) do postaci

$$[f(\tilde{u})]_\alpha = \left[\min_{\substack{j=1, \dots, N \\ k=1, \dots, m}} \{f(w_j), f(e_k)\}, \max_{\substack{j=1, \dots, N \\ k=1, \dots, m}} \{f(w_j), f(e_k)\} \right] \quad (3.20)$$

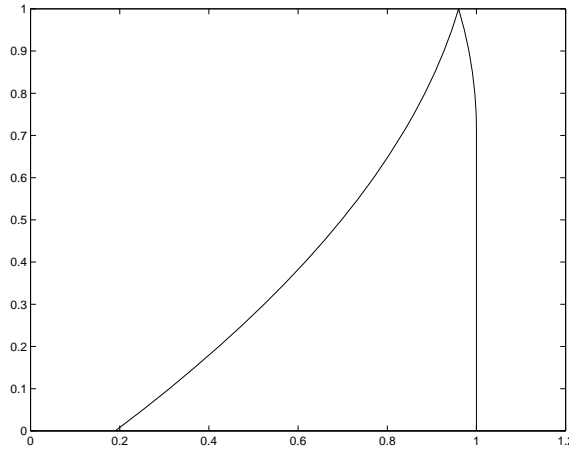
Przykład 3.3.2. Rozważmy kwadratową funkcję $f(x) = 2x - x^2$ i wyznaczmy obraz zbioru $\tilde{u} = \Delta(0.5, 1.2, 1.9)$. Funkcja f osiąga w przedziale $[0.5, 1.9]$ wartość ekstremalną (maksimum) w punkcie $x^* = 1$ i punkt ten należy do α -warstw $[0.5 + 0.7\alpha, 1.9 - 0.7\alpha]$ dla $\alpha \leq 5/7$. Skoro $e_1 = f(x^*) = 1$, to zgodnie z równaniem (3.20) mamy

$$[f(\tilde{u})]_\alpha = [\min\{f(0.5 + 0.7\alpha), f(1.9 - 0.7\alpha), 1\}, \max\{f(0.5 + 0.7\alpha), f(1.9 - 0.7\alpha), 1\}]$$

gdy $\alpha \in [0, 5/7]$ oraz

$$[f(\tilde{u})]_\alpha = [\min\{f(0.5 + 0.7\alpha), f(1.9 - 0.7\alpha)\}, \max\{f(0.5 + 0.7\alpha), f(1.9 - 0.7\alpha)\}]$$

gdy $\alpha \in (5/7, 1]$. Wynikową funkcję przynależności zbioru rozmytego $f(u)$ przedstawiono na rysunku 3.4. \square



Rysunek 3.4: Ilustracja metody wierzchołkowej: obraz liczby trójkątnej $\Delta(0.5, 1.2, 1.9)$ w odwzorowaniu $f(x) = 2x - x^2$.

3.3.2 Metoda transformacji

Poszukiwanie ekstremów lokalnych funkcji n zmiennych bywa kłopotliwe, a przede wszystkim wymaga dodatkowych obliczeń. Dlatego Hanss [26] zaproponował w pełni zalgorytmizowaną implementację metody wierzchołkowej.

Założmy, że chcemy obliczyć wartość funkcji n zmiennych x_1, \dots, x_n w punkcie o rozmytych współrzędnych $\tilde{u}_1, \dots, \tilde{u}_n$. Metoda transformacji polega na wyznaczeniu m α -warstw każdego rozmytego argumentu funkcji f i wybraniu z nich odpowiednich punktów. Jeżeli f jest funkcją monotoniczną, wystarczy (jak poprzednio) ograniczyć się do krańców $\underline{u}_{i,\alpha}, \bar{u}_{i,\alpha}$ odpowiednich α -warstw $[\tilde{u}_i]_\alpha$, $i = 1, \dots, n$, a w przypadku przeciwnym należy uwzględnić dodatkowo pewną liczbę punktów z wnętrza odcinka $[\underline{u}_{i,\alpha}, \bar{u}_{i,\alpha}]$. Hanss [26] sugeruje, aby dzielić przedział $[0, 1]$ na m równych odcinków, tzn. przyjąć $\alpha_j = j/m$, $j = 1, \dots, m$, i wyróżnić $m+1$ przedziałów

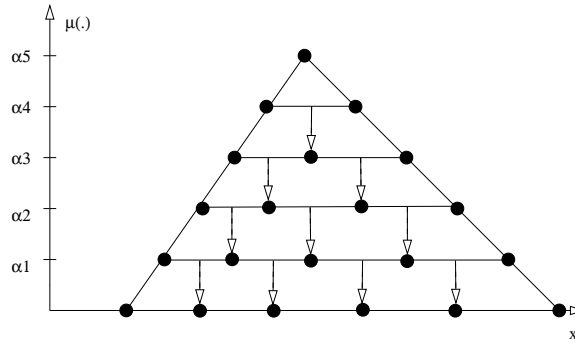
$$[u_i]_{\alpha_m} \subset [u_i]_{\alpha_{m-1}} \subset \dots \subset [u_i]_{\alpha_1} \subset [u_i]_0$$

gdzie $[u_i]_0 = \text{supp}(\tilde{u}_i)$ jest nośnikiem liczby rozmytej \tilde{u}_i , $i = 1, \dots, n$. Zauważmy, że skoro \tilde{u} jest liczbą rozmytą, to $[\tilde{u}]_{\alpha_m}$ jest zbiorem jednoelementowym dla każdego $i = 1, \dots, n$.

W przypadku niemonotonicznej funkcji f zbiór punktów E_j przeznaczonych do testowania j -tej α -warstwy $[\tilde{u}]_{\alpha_j}$ zawiera $m+1-j$ elementów $e_{j,l}$ postaci

$$e_{j,l} = \begin{cases} \underline{u}_{\alpha_j} & \text{gdy } l = 1 \\ \frac{1}{2}(e_{j+1,l-1} + e_{j+1,l}) & \text{gdy } l = 2, \dots, m-j \\ \bar{u}_{\alpha_j} & \text{gdy } l = m-j+1 \end{cases} \quad (3.21)$$

Na rysunku 3.5 pokazano sposób wybierania punktów $e_{j,l}$ dla każdej α -warstwy.



Rysunek 3.5: Ilustracja metody transformacji. Kropkami zaznaczono punkty wybierane do testowania w każdej α -warstwie.

W kolejnym kroku wyznacza się wartość funkcji

$$f_{j,l_1,\dots,l_n} = f(e_{1,j,l_1}, \dots, e_{n,j,l_n})$$

dla każdej kombinacji wartości l_1, \dots, l_n (symbol e_{i,j,l_i} wskazuje na l_i -ty element j -tej α -warstwy i -tego argumentu). Minimum z tych wartości¹ odpowiada skrajnej lewej, \underline{f}_{α_j} , a maksimum – skrajnej prawej \bar{f}_{α_j} wartości α -warstwy $[f]_{\alpha_j}$.

W przypadku, gdy n jest niezbyt dużą wartością, powiedzmy $n \leq 3$, taki sposób postępowania można z łatwością zaimplementować stosując n pętli **for**. Założmy dla ilustracji, że $n = 2$, a elementy zbiorów E_j obu argumentów zapisane są w 2-wymiarowych tablicach E1 oraz E2. j -ty wiersz tablicy zawiera elementy odpowiedniego zbioru E_j . Z uwagi na sposób prowadzenia obliczeń od wartości $\alpha = 1$ do $\alpha = 0$, wygodnie jest przyjąć, że pierwszy

¹W istocie jest to minimum z tych wartości oraz $\underline{f}_{\alpha_{j+1}}$. Podobna uwaga dotyczy wyznaczania maksimum – por. [26], s. 281.

wiersz² i -tej tablicy zawiera pojedynczy element będący zawartością warstwy $[\tilde{u}_i]_{\alpha_m=1}$, natomiast ostatni wiersz zawiera $m+1$ elementów ze zbioru $\text{supp}(\tilde{u}_i)$. Wyniki przechowywane są w tablicy W o $m+1$ wierszach i dwóch kolumnach; pierwsza kolumna zawiera lewe, a druga – prawe ograniczenie odpowiedniej α -warstwy zbioru $f(\tilde{u}_1, \dots, \tilde{u}_n)$. Program (w języku Java) implementujący omówiony sposób postępowania przedstawiono w tablicy 3.1.

```
void evaluate() {
    double min, max, y;

    W[0][1] = fun(E1[0][0], E2[0][0]);
    W[0][2] = Q[0][1];
    for (int j=1; j<=m; j++) {
        min = Q[j-1][1];
        max = Q[j-1][2];
        for (int l1=0; l1<X1[j].length; l1++) {
            for (int l2=0; l2<X2[i].length; l2++) {
                y = fun(E1[j][l1], E2[j][l2]);
                if (y<min) min = y;
                if (y>max) max = y;
            }
        }
        W[j][1] = min;
        W[j][2] = max;
    }
}
```

Tablica 3.1: Przykładowa implementacja metody transformacji dla wyznaczania rozmytej wartości funkcji dwóch (rozmytych) zmiennych.

W przypadku większej liczby argumentów można postępować w sposób bardziej wyrafinowany, a który zaproponował Klimke [41].

Odnotujmy też, że jeżeli korzystamy wyłącznie z krańców warstw $[\tilde{u}]_\alpha$, to mamy do czynienia ze zredukowaną metodą transformacji, a w przypadku przeciwnym – z pełną metodą transformacji. Hanss ([27], punkt 4.3) wprowadził także pojęcie rozszerzonej metody transformacji, polegającej na wstępnej ocenie, które z argumentów funkcji f odpowiedzialne są za jej niemonotoniczność, a które nie. W przypadku argumentów należących do pierwszej grupy, nazywanych argumentami typu g , do zbioru E_j zaliczamy wszystkie punkty wyznaczone zgodnie ze wzorem 3.21, a w przypadku argumentów drugiej grupy, nazywanych argumentami typu r , zbiór $E_j = \{\underline{u}_{\alpha_j}, \bar{u}_{\alpha_j}\}$.

Zilustrujmy metodę transformacji przykładem.

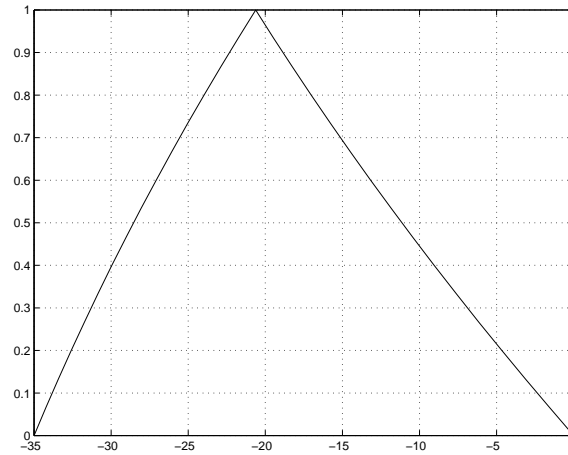
Przykład 3.3.3. Rozważmy funkcję trzech zmiennych (por. [27], s. 130)

$$g(x_1, x_2, x_3) = \sin(x_1) + x_2^2 - x_3$$

i załóżmy, że jej argumenty $\tilde{u}_1, \tilde{u}_2, \tilde{u}_3$ są gaussowskimi liczbami rozmytymi o funkcjach przynależności

$$\mu_i(\xi) = \begin{cases} \exp \left[-\frac{(\xi - m_i)^2}{2\sigma_i^2} \right] & \text{gdy } 0 \leq \xi \leq 3\sigma_i \\ 0 & \text{w p.p.} \end{cases} \quad (3.22)$$

²Odnotujmy, że w większości języków programowania, np. Java, C, czy C++, wiersz ten ma on numer 0.



Rysunek 3.6: Obraz zbioru rozmytego $\Delta(4, 4.5, 5)$ w przekształceniu $f(x) = x^3 - x^2 - 17x - 15$.

i parametrach (m_i, σ_i) , $i = 1, 2, 3$, równych odpowiednio $(\pi/3, \pi/9)$, $(0.5, 0.5)$ oraz $(3, 1/3)$. Stosując pełną metodę transformacji dostajemy liczbę rozmytą \tilde{g} o funkcji przynależności przedstawionej w lewej części rysunku 3.7. Jej wyznaczenie wymagało wykonania $e(m) = \sum_{j=1}^{m+1} (m+1-j)^3$ ewaluacji funkcji f , np. $e(10) = 4356$, $e(20) = 53361$.

Założmy teraz dla kontrastu, że argumentami funkcji g są zmienne losowe X_1, X_2, X_3 o rozkładach normalnych z podanymi wyżej parametrami. Funkcja gęstości dla $g(X_1, X_2, X_3)$ wyznaczona po wygenerowaniu jednego miliona trójek (x_1, x_2, x_3) ma postać pokazaną w prawej części rysunku 3.7.

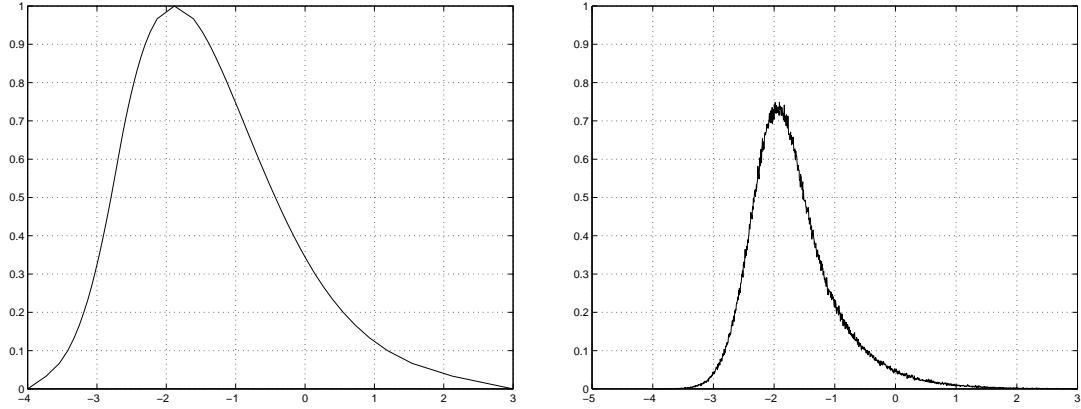
Lewy i prawy wykres mają podobne cechy (pomijając efekt skali). Można jednak zauważyć, że:

- (a) Formalizm rozmyty wymaga mniejszej liczby obliczeń. W istocie, stosując rozszerzoną metodę transformacji, liczbę koniecznych ocen funkcji g można zredukować do $e_r(10) = 1012$, gdyż \tilde{u}_1, \tilde{u}_2 są g parametrami, zaś \tilde{u}_3 jest r parametrem (por. [27], s. 131).
- (b) Zakres zmienności (rozmytej) liczby $g(\tilde{u}_1, \tilde{u}_2, \tilde{u}_3)$, równy $[-4, 3]$, jest szerszy niż ma to miejsce w ujęciu probabilistycznym. Szacowane skrajne wartości funkcji gęstości zmiennej losowej $g(X_1, X_2, X_3)$ to $p(-3.9975) \approx 0.0014$ oraz $p(\xi) \leq 0.0006$ dla $\xi \geq 2.8$. Efekt ten pogłębia się wraz ze wzrostem liczby zmiennych. Można więc powiedzieć, że z inżynierskiego punktu widzenia metoda rozmyta jest bardziej wiarygodna (jeżeli tylko interesują nas możliwe do zaobserwowania skrajne wartości). \square

3.3.3 Metoda funkcji sklepanych

Metoda transformacji przeznaczona jest do operowania na liczbach rozmytych, ale nie na przedziałach rozmytych. Wady tej nie posiada omawiana obecnie metoda korzystająca z funkcji sklepanych, a opisana w [9]. Jej istotą jest wyznaczenie przybliżonego oszacowania $\hat{F}_m(\tilde{x}_1, \dots, \tilde{x}_n)$ funkcji n -zmiennych $f(x_1, \dots, x_n)$ poprzez:

- (a) rozbiecie nośników zbiorów rozmytych odpowiadających argumentom $\tilde{x}_1, \dots, \tilde{x}_n$ na m podprzedziałów I_1, \dots, I_m o równej długości, a następnie



Rysunek 3.7: Wynik nieliniowego przekształcenia $g = \sin(x_1) + x_2^2 - x_3$. Lewa część rysunku przedstawia efekt zastosowania metody transformacji, a prawa to funkcja gęstości sumy zmiennych losowych o rozkładzie normalnym.

- (b) aproksymację funkcji f liniowymi funkcjami \hat{f}_i określonymi na odpowiednich przedziałach I_i , $i = 1, \dots, m$.

Skoro operujemy funkcjami liniowymi, to znalezienie rozmytej wartości $\hat{f}_i(\tilde{x}_1, \dots, \tilde{x}_n)$ odbywa się szybko poprzez korzystanie z reguł arytmetyki przedziałowej, tzn.

$$\hat{f}_i(\tilde{x}_1, \dots, \tilde{x}_n) = a_0 \oplus a_1 \tilde{x}_1 \oplus \dots \oplus a_n \tilde{x}_n$$

gdzie a_0, a_1, \dots, a_m są odpowiednio dobranymi współczynnikami.

Omawianie tej metody rozpoczynamy od przypadku jednowymiarowego [8]. Niech \tilde{u} będzie zbiorem rozmytym o nośniku $\text{supp}(\tilde{u}) = [a, b]$ i niech $f(x): A \rightarrow \mathbb{R}$, gdzie $A \supset [a, b]$, będzie funkcją ciągłą i różniczkowalną w (a, b) .

Podzielmy przedział $[a, b]$ na m podprzedziałów o równej długości, $[x_i, x_{i+1}]$, przy czym $x_i = a + (i-1)(b-a)/m$, $i = 1, m+1$, i określmy m rozmytych zbiorów \tilde{u}_i o funkcjach przynależności³

$$\tilde{u}_i(x) = \begin{cases} \tilde{u}(x) & \text{gdy } x \in [x_i, x_{i+1}] \\ 0 & \text{gdy } x \notin [x_i, x_{i+1}] \end{cases}, \quad i = 1 \dots, m$$

W ten sposób liczbę rozmytą \tilde{u} (a ściślej: jej funkcję przynależności) można przedstawić w postaci

$$\tilde{u}(x) = \max [\tilde{u}_1(x), \tilde{u}_2(x), \dots, \tilde{u}_n(x)] \quad (3.23)$$

Niech dalej

$$x_i^* = \frac{x_i + x_{i+1}}{2}, i = 1, \dots, m$$

Jako funkcję f_i przyjmujemy teraz styczną do $f(x)$ w punkcie $(x_i^*, f(x_i^*))$ i określoną na zbiorze $[x_i, x_{i+1}]$. Zatem funkcja f_i ma postać

$$\begin{aligned} f_i(x) &= f'(x_i^*)(x - x_i^*) + f(x_i^*) \\ &= [f(x_i^*) - f'(x_i^*)x_i^*] + f'(x_i^*)x \\ &= a_0 + a_1 x \end{aligned} \quad (3.24)$$

³Przypominamy – por. strona 13 – że $\tilde{u}_i(x)$ jest równoważne pełnemu zapisowi $\mu_{u_i}(x)$.

Gdy jej argument jest zbiorem rozmytym \tilde{u}_i , to $f_i(\tilde{u}_i)$ jest zbiorem rozmytym o funkcji przynależności

$$\mu_{f_i(\tilde{u}_i)}(y) = \begin{cases} \mu_{\tilde{u}_i}\left(\frac{y-f(x_i^*)}{f'(x_i^*)} + x_i^*\right) & \text{gdy } f'(x_i^*) \neq 0 \\ \mu_{\tilde{u}_i}(x_i^*) & \text{gdy } f'(x_i^*) = 0 \end{cases} \quad (3.25)$$

Ostatecznie, zbiór rozmyty $f(\tilde{u})$ aproksymuje funkcja przynależności $\mu_{\hat{f}_m(\tilde{u})}$ postaci

$$\mu_{\hat{f}_m(\tilde{u})}(y) = \max_{i=1,\dots,m} \mu_{f_i(\tilde{u}_i)}(y) \quad (3.26)$$

Do jej wyznaczenia wystarczy określić nośnik zbioru $f(\tilde{u})$, a następnie dla wybranych wartości y z tego nośnika wyznaczyć, zgodnie ze wzorem (3.25) wartości funkcji przynależności $\mu_{f_i(\tilde{u}_i)}(y)$ dla $i = 1, \dots, m$ i wybrać spośród nich wartość maksymalną.

Przykład 3.3.4. Obliczmy wartość funkcji $f(x) = -2x^2 + 4x$ w punkcie \tilde{u} opisanym funkcją przynależności (por. rysunek 3.8)

$$u(x) = \begin{cases} \sqrt{x} & \text{gdy } x \in [0.0, 1.0] \\ 1 & \text{gdy } x \in [1.0, 1.5] \\ 4(x-2)^2 & \text{gdy } x \in [1.5, 2.0] \end{cases}$$

W tym celu podzielimy przedział $[a, b] = \text{supp}(\tilde{u})$ na cztery podprzedziały o długości 0.5 każdy i określmy na nich liniowe funkcje aproksymujące o współczynnikach obliczanych zgodnie z równaniem (3.24) (por. prawy rysunek 3.8)

$$\begin{aligned} f_1(x) &= 3x + 0.125, & x \in [0.0, 0.5] \\ f_2(x) &= x + 1.125, & x \in [0.5, 1.0] \\ f_3(x) &= -x + 3.125, & x \in [1.0, 1.5] \\ f_4(x) &= -3x + 6.125, & x \in [1.5, 2.0] \end{aligned}$$

Założmy dalej, że chcemy wyznaczyć wartość funkcji przynależności $\mu_{\hat{f}_4(\tilde{u})}$ w punkcie $y = 0.4$. Dla $i = 1$ mamy $y_1 = \frac{y-0.125}{3} = 0.09167$ oraz $\mu_{\tilde{u}}(y_1) \approx 0.3$. Podobnie stwierdzamy, że $y_2 = 0.4 - 1.125 = -0.725$ skąd $\mu_{\tilde{u}}(y_1) = 0$, $\mu_{\tilde{u}}(y_3 = 2.725) = 0$ oraz $\mu_{\tilde{u}}(y_4 = 1.9083) = 0.336$. Ostateczne więc $\mu_{\hat{f}_4(\tilde{u})}(0.4) = \max\{0.3, 0, 0, 0.336\} = 0.3$. Wykres funkcji przynależności zbioru $f(\tilde{u})$ przedstawiono na rysunku 3.9. \square

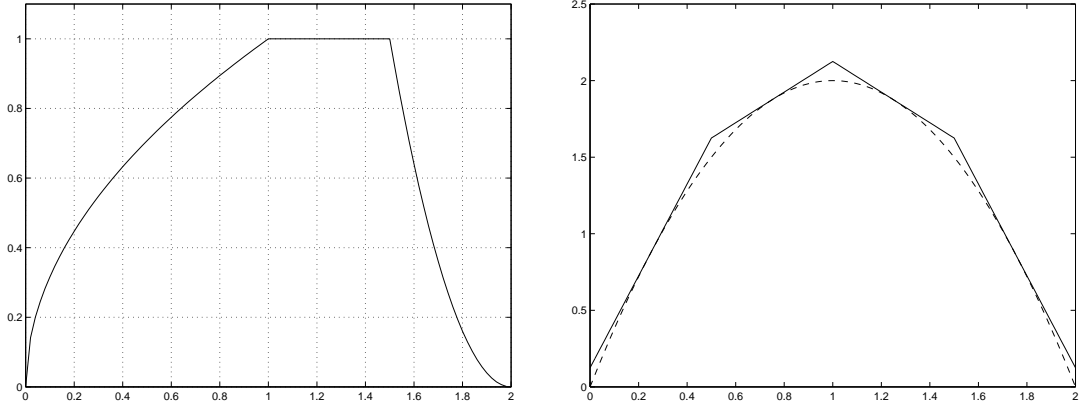
Następujące twierdzenia podane w [8] charakteryzują własności aproksymacji $\hat{f}_m(\tilde{u})$.

Twierdzenie 3.3.1. Niech \tilde{u} będzie zbiorem rozmytym o nośniku $[a, b]$ i niech $f: A \rightarrow \mathbb{R}$, $A \supset [a, b]$, będzie funkcją ciągłą i różniczkowalną w (a, b) . Niech $\hat{f}_m(\tilde{u})$ będzie zbiorem rozmytym o funkcji przynależności postaci (3.26). Wówczas:

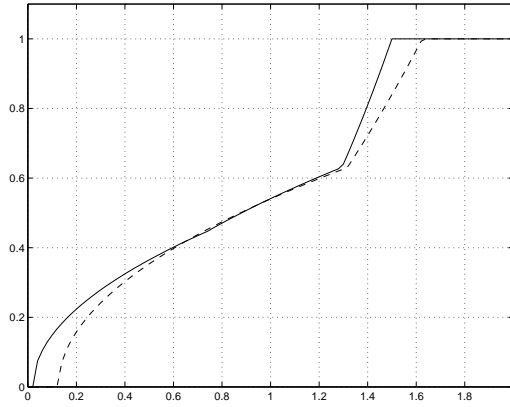
- (a) $(\forall \alpha \in [0, 1])[\hat{f}_m(\tilde{u})]_\alpha = \bigcup_{i=1}^m f_i([\tilde{u}_i]_\alpha)$,
- (b) $\hat{f}_m(\tilde{u})$ jest normalnym zbiorem rozmytym,
- (c) $\hat{f}_m(\tilde{u}): \mathbb{R}_{\mathcal{F}} \rightarrow \mathbb{R}_{\mathcal{F}}$ jest dobrze określoną funkcją.

Dowód:

- (a) Wynika z twierdzenia Nguyena 3.1.2.
- (b) Skoro \tilde{u} jest rozmytym przedziałem, to istnieje przynajmniej jeden element $x_0 \in [a, b]$ taki, że $\mu_{\tilde{u}}(x_0) = 1$. Założmy, że $x_0 \in [x_i, x_{i+1}]$ dla pewnego $i \in \{1, \dots, m\}$ i niech $f_i(x_0) = y_0$. Wówczas $\mu_{f_i(\tilde{u}_i)}(y_0) = 1$ i, z uwagi na (3.26), także $\mu_{\hat{f}_m(\tilde{u})}(y_0) = 1$.



Rysunek 3.8: Ilustracja aproksymacyjnej metody wyznaczania funkcji przynależności dla $f(\tilde{u})$. Lewy rysunek: funkcja przynależności przedziału \tilde{u} , prawy – aproksymacja funkcji $f(x)$ (jej wykres zaznaczono linią przerywaną) czterema funkcjami liniowymi.



Rysunek 3.9: Przybliżona funkcja przynależności zbioru $f(\tilde{u})$. Linią przerywaną zaznaczono aproksymację uzyskaną dla $m = 4$, a linią ciągłą – dla $m = 10$.

- (c) Należy pokazać, że dla każdego $\alpha \in [0, 1]$ zbiór $[\hat{f}_m(\tilde{u})]_\alpha$ jest niepusty i domknięty. Jego domkniętość wynika (a), natomiast niepustość z (b). Powinniśmy też sprawdzić, że $[\hat{f}_m(\tilde{u})]_{\alpha_1} \subseteq [\hat{f}_m(\tilde{u})]_{\alpha_2}$ dla dowolnych $\alpha_1 > \alpha_2 \in [0, 1]$. \square

Lemat 3.3.1. Niech \tilde{u} będzie zbiorem rozmytym o nośniku $[a, b]$ i niech $f: A \rightarrow \mathbb{R}$, $A \supset [a, b]$, będzie ciągłą funkcją posiadającą w (a, b) ograniczoną pochodną, tzn. $\sup_{x \in (a, b)} |f'(x)| = K_1 < \infty$. Jeżeli przedział $[a, b]$ podzielimy na m podprzedziałów $[x_i, x_{i+1}]$ i na każdym z nich określimy zgodnie ze wzorem (3.24) funkcję $f_i, i = 1, \dots, m$, to istnieje taka stała $K > 0$, że

$$\max_{i=1, \dots, m} \sup_{x_i \in [x_i, x_{i+1}]} |f(x) - f_i(x)| \leq K/m$$

Dowód:

\square

Twierdzenie 3.3.2. Niech \tilde{u} będzie zbiorem rozmytym o nośniku $[a, b]$ i niech $f: A \rightarrow \mathbb{R}$, $A \supset [a, b]$, będzie funkcją ciągłą i różniczkowalną w (a, b) , przy czym $\sup_{x \in (a, b)} |f'(x)| = K_1 < \infty$. Niech $\hat{f}_m(\tilde{u})$ będzie zbiorem rozmytym o funkcji przynależności postaci (3.26).

Wówczas istnieje taka stała $K > 0$, że

$$D_H(f(\tilde{u}), \hat{f}_m(\tilde{u})) \leq \frac{K}{m} \quad (3.27)$$

Dowód: Zgodnie z definicją (??) odległość Hausdorffa między dwoma zbiorami rozmytymi jest równa

$$D_H(f(\tilde{u}), \hat{f}_m(\tilde{u})) = \sup_{\alpha \in [0,1]} d_H(f([\tilde{u}]_\alpha), \hat{f}_m([\tilde{u}]_\alpha))$$

□

3.4 Zastosowania

Chociaż zasada rozszerzania stanowi ważne narzędzie umożliwiające adaptację klasycznych pojęć do potrzeb analizy rozmytej, to sam proces formułowania rozmytych odpowiedników takich pojęć bywa niejednoznaczny. Dotyczy to np. rozmytych równań różniczkowych. Nawet tak proste pojęcia, jak równania liniowe, można także interpretować na wiele sposobów uzyskując (bądź nie) znacznie różniące się rozwiązania. Zagadnienia te ilustrujemy w niniejszym punkcie.

3.4.1 Rozmyte operatory min oraz max

Jeżeli $\tilde{a}, \tilde{b} \in \mathbb{R}_{\mathcal{F}}$, to zgodnie z zasadą rozszerzania rozmyte minimum z obu tych liczb jest liczbą rozmytą, o funkcji przynależności postaci (zastosowaliśmy tu konwencję (c) z uwagi 2.1.1)

$$\min(\tilde{a}, \tilde{b})(z) = \sup_{\substack{(x,y) \in \mathbb{R}^2 \\ z = \min(x,y)}} \min(\tilde{a}(x), \tilde{b}(y)) \quad (3.28)$$

Zgodnie z zasadą rozszerzania mamy

$$[f(\tilde{a}, \tilde{b})]_\alpha = \min([\tilde{a}]_\alpha, [\tilde{b}]_\alpha) = [\min(\underline{a}_\alpha, \underline{b}_\alpha), \min(\bar{a}_\alpha, \bar{b}_\alpha)]$$

Podobnie definiujemy rozmyte maksimum

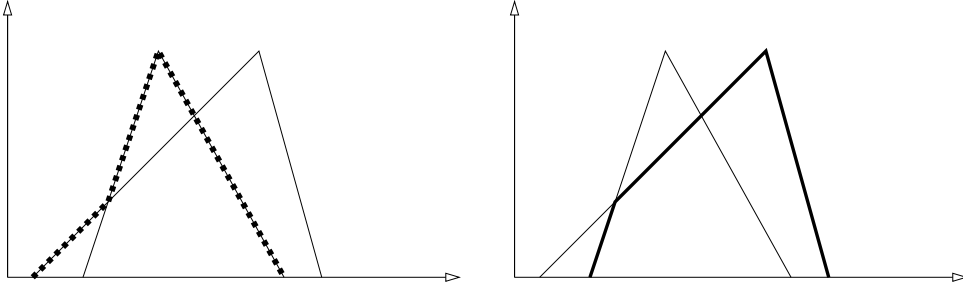
$$\max(\tilde{a}, \tilde{b})(z) = \sup_{\substack{(x,y) \in \mathbb{R}^2 \\ z = \max(x,y)}} \min(\tilde{a}(x), \tilde{b}(y)) \quad (3.29)$$

Sposób wyznaczania obu tych wielkości zilustrowano na rysunku 3.10.

Dysponując tak zdefiniowanymi operatorami, można wprowadzić relację porządku w zbiorze liczb rozmytych, np. liczba

$$v(\tilde{a} \leq \tilde{b}) = \sup_{\substack{(x,y) \in \mathbb{R}^2 \\ x \leq y}} \min[\tilde{a}(x), \tilde{b}(y)] \quad (3.30)$$

określa stopień, w jakim liczba \tilde{a} jest mniejsza bądź równa liczbie \tilde{b} .



Rysunek 3.10: Wyznaczanie rozmytego minimum i maksimum. Cienkimi liniami zaznaczono funkcje przynależności dwóch liczb rozmytych. Pogrubioną linią kropkowaną zaznaczono rozmyte minimum (lewy rysunek), a pogrubioną linią ciągłą – rozmyte maksimum (prawy rysunek) z obu tych liczb.

3.4.2 Metryzacja przestrzeni liczb rozmytych

W wielu zastosowaniach dotyczących np. rozmytych zmiennych losowych [?], rozmytych równań różniczkowych [?] czy układów dynamicznych istnieje potrzeba metryzacji przestrzeni $\mathbb{R}_{\mathcal{F}}^n$. Najczęściej w przestrzeni \mathbb{R}^n definiuje się w metrykę d_{∞} stanowiącą uogólnienie odległości Hausdorffa. Mianowicie

Definicja 3.4.1. Niech X będzie zwartą przestrzenią metryczną z metryką d , niech

$$d_H(A, B) = \max \left[\sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y) \right] \quad (3.31)$$

oznacza odległość Hausdorffa między zbiorami $A \subseteq X, B \subseteq X$.

W przypadku zbiorów rozmytych $\tilde{u}, \tilde{v} \in \mathbb{R}_{\mathcal{F}}^n$ odległość d_{∞}

$$d_{\infty}(\tilde{u}, \tilde{v}) = \sup_{\alpha \in [0,1]} d_H([\tilde{u}]_{\alpha}, [\tilde{v}]_{\alpha}) \quad (3.32)$$

jest uogólnioną odległością Hausdorffa. □

Jeżeli wspomniane wyżej zbiory A i B są przedziałami postaci $A = [u_1, u_2]$, $B = [v_1, v_2]$ wówczas (3.31) przyjmuje postać

$$d_H(A, B) = \max[|u_1 - v_1|, |u_2 - v_2|]$$

Zatem, gdy $n = 1$, to stosując parametryczną reprezentację (2.22) wzór (3.32) przyjmuje postać

$$d_{\infty}(\tilde{u}, \tilde{v}) = \sup_{\alpha \in [0,1]} \max(|\underline{u}_{\alpha} - \underline{v}_{\alpha}|, |\bar{u}_{\alpha} - \bar{v}_{\alpha}|) \quad (3.33)$$

Lemat 3.4.1. Jeżeli $u, v, w, x \in \mathbb{R}_{\mathcal{F}}^n$ oraz $\lambda \in \mathbb{R}$, to odległość d_{∞} posiada m.in. następujące własności:

- (d1) $d_{\infty}(u, v) = d_{\infty}(v, u)$, $u, v \in \mathbb{R}_{\mathcal{F}}$,
- (d2) $d_{\infty}(u + w, v + w) = d_{\infty}(u, v)$, $u, v, w \in \mathbb{R}_{\mathcal{F}}$,
- (d3) $d_{\infty}(\lambda \cdot u, \lambda \cdot v) = |\lambda| \cdot d_{\infty}(u, v)$, $u, v \in \mathbb{R}_{\mathcal{F}}$, $\lambda \in \mathbb{R}$,
- (d4) $d_{\infty}(u + v, w + x) \leq d_{\infty}(u, w) + d_{\infty}(v, x)$, $u, v, w, x \in \mathbb{R}_{\mathcal{F}}$. □

Pokazuje się [14], że $(\mathbb{R}_{\mathcal{F}}^n, d_{\infty})$ jest zupełną przestrzenią metryczną. Można ją więc traktować jako uogólnienie przestrzeni metrycznej $\mathcal{K}(\mathbb{R}^n)$ niepustych wypukłych i zwartych podzbiorów przestrzeni \mathbb{R}^n z metryką d_{∞} . Jednakże metryka d_{∞} nie jest pełnym rozszerzeniem

topologii indukowanej metryką Hausdorffa. W szczególności $(\mathbb{R}_{\mathcal{F}}^n, d_{\infty})$ nie jest przestrzenią ośrodkową⁴.

Definicja 3.4.2. Odwzorowanie $F: \mathbb{R}_{\mathcal{F}}^n \rightarrow \mathbb{R}_{\mathcal{F}}$ jest ciągle w punkcie $\tilde{\mathbf{x}}^* \in \mathbb{R}_{\mathcal{F}}^n$ jeżeli dla dowolnego $\epsilon > 0$ istnieje $\delta > 0$ takie, że $d_{\infty}(F(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}^*) < \epsilon$ jeżeli tylko $d_{\infty}(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}^*) < \delta$.

Jeżeli F jest odwzorowaniem ciągłym w każdym punkcie $\mathbf{x} \in U \subset \mathbb{R}_{\mathcal{F}}^n$, to mówimy, że jest ono ciągle w U . \square

Jeżeli f jest funkcją ciągłą to indukowane przez nią odwzorowanie F jest również ciągle w sensie podanej wyżej definicji [67].

3.4.3 Równania liniowe

Zajmijmy się teraz liniowym równaniem z rozmytymi współczynnikami postaci

$$\tilde{a}\tilde{x} + \tilde{b} = \tilde{c} \quad (3.34)$$

Korzystając z własności mówiącej, że równość zbiorów rozmytych implikuje równość ich α -warstw, napiszmy

$$[\tilde{a}\tilde{x} + \tilde{b}]_{\alpha} = [\tilde{c}]_{\alpha}$$

Zakładając parametryczną reprezentację odpowiednich zbiorów rozmytych, np. $[\tilde{a}]_{\alpha} = [\underline{a}(\alpha), \bar{a}(\alpha)]$, zapisujemy powyższe równanie w postaci

$$[\underline{a}x(\alpha) + \underline{b}(\alpha), \bar{a}x(\alpha) + \bar{b}(\alpha)] = [\underline{c}(\alpha), \bar{c}(\alpha)]$$

gdzie

$$\underline{a}x(\alpha) = \min[\underline{a}(\alpha)\underline{x}(\alpha), \underline{a}(\alpha)\bar{x}(\alpha), \bar{a}(\alpha)\underline{x}(\alpha), \bar{a}(\alpha)\bar{x}(\alpha)]$$

$$\bar{a}x(\alpha) = \max[\underline{a}(\alpha)\underline{x}(\alpha), \underline{a}(\alpha)\bar{x}(\alpha), \bar{a}(\alpha)\underline{x}(\alpha), \bar{a}(\alpha)\bar{x}(\alpha)]$$

Ostatecznie więc rozwiązanie równania (3.34) znajdujemy rozwiązując układ równań

$$\begin{cases} \underline{a}x(\alpha) + \underline{b}(\alpha) = \underline{c}(\alpha) \\ \bar{a}x(\alpha) + \bar{b}(\alpha) = \bar{c}(\alpha) \end{cases} \quad (3.35)$$

Jeżeli uzyskane w ten sposób przedziały reprezentują liczbę rozmytą \tilde{x} , to mówimy, że \tilde{x} jest silnym rozwiązaniem równania (3.34). W przypadku przeciwnym, konstruujemy zbiór rozmyty

$$[\tilde{y}]_{\alpha} = [\min(\underline{x}(\alpha), \bar{x}(\alpha)), \min(\underline{x}(\alpha), \bar{x}(\alpha))] \quad (3.36)$$

Jeżeli $\tilde{y} \in \mathbb{R}_{\mathcal{F}}$, to mówimy, że \tilde{y} jest słabym rozwiązaniem równania (3.34). Nomenklaturę tę wprowadzono w pracy [1], por. też [1].

Przykład 3.4.1. Załóżmy, że parametry równania (3.34) są trójkątnymi liczbami rozmytymi $\tilde{a} \sim \Delta(1, 2, 3)$, $\tilde{b} \sim \Delta(-3, -2, -1)$, $\tilde{c} \sim \Delta(3, 4, 5)$, tzn. ich parametryczną reprezentacją są przedziały $[\tilde{a}]_{\alpha} = [1 + \alpha, 3 - \alpha]$, $[\tilde{b}]_{\alpha} = [-3 + \alpha, -1 - \alpha]$, oraz $[\tilde{c}]_{\alpha} = [3 + \alpha, 5 - \alpha]$. Skoro \tilde{a} i \tilde{c} są dodatnimi liczbami rozmytymi, a \tilde{b} jest ujemną liczbą rozmytą, to rozwiązanie \tilde{x} powinno być dodatnią liczbą rozmytą. Obserwacja ta pozwala sformułować następujący układ równań

$$\begin{aligned} \underline{a}(\alpha)\underline{x}(\alpha) + \underline{b}(\alpha) &= \underline{c}(\alpha) \\ \bar{a}(\alpha)\bar{x}(\alpha) + \bar{b}(\alpha) &= \bar{c}(\alpha) \end{aligned}$$

⁴Fakt ten odnotowano w pracy E.P. Klement, M.L. Puri, D.A. Ralescu. Limit theorems for fuzzy random variables. *Proc. Roy. Soc. London Ser. A*, **407**, 1986, 171-182.

Jego rozwiązanie ma postać

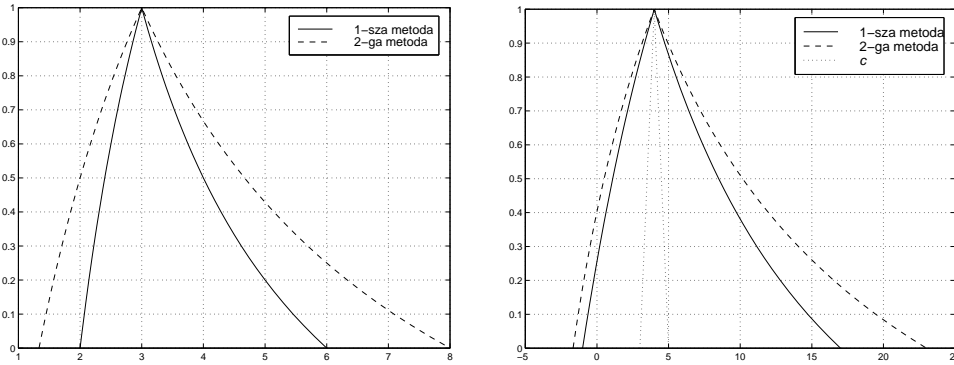
$$\underline{x}(\alpha) = \frac{6}{1+\alpha}, \quad \bar{x}(\alpha) = \frac{6}{3-\alpha}$$

Przedział o tak określonych granicach nie reprezentuje liczby rozmytej, gdyż jego lewe ograniczenie jest malejącą, a prawe – rosnącą funkcją parametru α . Oznacza to, że rozmyte równanie liniowe nie ma w tym przypadku mocnego rozwiązania. Łatwo jednak sprawdzić, że ma on słabe rozwiązanie \tilde{y} przedstawione linią ciągłą w lewej części rysunku 3.11. Jednakże konsekwencją „słabości” rozwiązania, jest to, że $a\tilde{y} + b \supseteq \tilde{c}$, co przedstawiono w prawej części tego rysunku. \square

Druga metoda rozwiązywania równania (3.34) polega na „rozmyciu” odpowiadającego mu równania deterministycznego, tzn.

$$\tilde{x} = \frac{\tilde{c} - \tilde{b}}{\tilde{a}} \quad (3.37)$$

Jej wadą jest to, że uzyskane w ten sposób rozwiązanie nie spełnia oryginalnego równania, co wynika z faktu, że $\tilde{a} \cdot 1/\tilde{a} \neq 1$. Ilustruje to rysunek 3.11.



Rysunek 3.11: Rozwiązanie równania liniowego z przykładu 3.4.1. W lewej części przedstawiono słabe rozwiązanie \tilde{y} uzyskane pierwszą metodą, a więc przez rozwiązanie układu równań (3.35) oraz rozwiązanie \tilde{x}^* uzyskane drugą metodą z zastosowaniem wzoru (3.37). W prawej porównano liczby rozmyte $\tilde{c}_1 = a\tilde{y} + \tilde{b}$, $\tilde{c}_2 = \tilde{a}\tilde{x}^* + \tilde{b}$, oraz oryginalną liczbę \tilde{c} reprezentującą prawą stronę równania (3.34).

Gdy współczynnik kierunkowy jest liczbą rzeczywistą a , to równanie (3.37) przyjmuje rozsądną parametryczną postać

$$[\tilde{x}]_\alpha = [\underline{B}(\alpha)/a, \overline{B}(\alpha)/a] \quad (3.38)$$

gdzie przez $[\underline{B}(\alpha), \overline{B}(\alpha)]$ oznaczono parametryczną reprezentację liczby $\tilde{B} = \tilde{c} - \tilde{b}$. Jeżeli przez x_m oznaczmy rozwiązanie (deterministycznego) równania $ax + b_m = c_m$, gdzie b_m i c_m są wartościami średnimi odpowiednich liczb rozmytych, to łatwo zauważyć, że przedział $[\tilde{x}]_\alpha$ można przedstawić w postaci

$$[\tilde{x}]_\alpha = x_m + \Xi_\alpha \quad (3.39)$$

gdzie

$$\Xi_\alpha = \{c: \underline{B}(\alpha) - b_m \leq c \leq \overline{B}(\alpha) - b_m\}$$

Inne metody rozwiązywania równania (3.34) zawiera praca [6].

3.4.4 Układy równań liniowych

Rozważmy teraz układ n równań liniowych z n niewiadomymi

$$\mathbf{Ax} = \mathbf{y} \quad (3.40)$$

gdzie \mathbf{A} jest macierzą o elementach rzeczywistych, natomiast \mathbf{x} i \mathbf{y} są wektorami, których składowymi są liczby rozmyte.

Tak sformułowany problem badany był w pochodzącej z 1998 roku pracy [21]. Późniejsze modyfikacje sposobu jego rozwiązywania inspirowane są tą właśnie pracą. Zarówno oryginalna jak i modyfikowana, a omawiana tu za [1], metoda polega na rozwiązaniu układu równań otrzymanego przez przyrównanie odpowiednich α -warstw lewej i prawej strony układu (3.40). Inny, nawiązujący do geometrycznej interpretacji roli macierzy \mathbf{A}^{-1} , sposób pochodzi z pracy [23]. Poniżej przedstawiamy szczegóły obu tych podejść.

3.4.4.1 Metoda α -warstw

Skoro składowe wektorów \mathbf{x} i \mathbf{y} są liczbami rozmytymi, to układ równań (3.40) można przepisać w postaci

$$\sum_{j=1}^n a_{ij} \cdot [\underline{x}_j(\alpha), \bar{x}_j(\alpha)] = [\underline{y}_i(\alpha), \bar{y}_i(\alpha)], \alpha \in [0, 1], i = 1, \dots, n \quad (3.41)$$

Zgodnie z regułami rozmytej arytmetyki, gdy $a_{ij} < 0$, to

$$a_{ij}[\underline{x}_j(\alpha), \bar{x}_j(\alpha)] = |a_{ij}|[-\bar{x}_j(\alpha), -\underline{x}_j(\alpha)]$$

Oczywiście składnik $|a_{ij}|(-\bar{x}_j(\alpha))$ znajdzie się w lewej stronie wyrażenia opisującego wartość \underline{y}_i , natomiast składnik $|a_{ij}|(-\underline{x}_j(\alpha))$ – w lewej stronie wyrażenia opisującego wartość \bar{y}_i . Zauważmy, że jeżeli prawą stronę układu (3.40) będziemy wyrażać za pomocą par postaci $(\underline{y}_i(\alpha), -\bar{y}_i(\alpha))$, to macierz współczynników takiego układu będzie nieujemna [21].

Pomijając, dla uproszczenia zapisu, wskazanie konkretnej wartości α , układ równań (3.40) można sprowadzić do postaci

$$\mathbf{SX} = \mathbf{Y} \quad (3.42)$$

gdzie \mathbf{X} i \mathbf{Y} są wektorami zawierającymi $2n$ składowych postaci

$$\mathbf{X} = \begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \\ -\bar{x}_1 \\ \vdots \\ -\bar{x}_n \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_n \\ -\bar{y}_1 \\ \vdots \\ -\bar{y}_n \end{bmatrix} \quad (3.43)$$

natomiast $\mathbf{S} = [s_{ij}]_{2n \times 2n}$ jest macierzą kwadratową o wyrazach

$$\begin{aligned} a_{ij} \geq 0 \quad & s_{ij} = a_{ij}, s_{i+n,j} = a_{ij} \\ a_{ij} < 0 \quad & s_{i,j+n} = -a_{ij}, s_{i+n,j} = -a_{ij} \end{aligned} \quad (3.44)$$

i pozostałych elementach równych zeru.

Przykład 3.4.2. Układ równań z dwiema niewiadomymi

$$\begin{cases} \tilde{x}_1 - \tilde{x}_2 = \tilde{y}_1 \\ \tilde{x}_1 + 3\tilde{x}_2 = \tilde{y}_2 \end{cases}$$

przyjmuje postać

$$\begin{cases} \{1 \cdot \underline{x}_1 + 1 \cdot (-\bar{x}_2), 1 \cdot (-\bar{x}_1) + 1 \cdot \underline{x}_2\} = \{\underline{y}_1, -\bar{y}_1\} \\ \{1 \cdot \underline{x}_1 + 3 \cdot \underline{x}_2, 1 \cdot (-\bar{x}_1) + 3 \cdot (-\bar{x}_2)\} = \{\underline{y}_2, -\bar{y}_2\} \end{cases}$$

której odpowiada macierzowe równanie

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 3 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ -\bar{x}_1 \\ -\bar{x}_2 \end{bmatrix} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ -\bar{y}_1 \\ -\bar{y}_2 \end{bmatrix}$$

□

Rozwiązaniem układu (3.42) jest oczywiście wektor

$$\mathbf{X} = \mathbf{S}^{-1} \mathbf{Y} \quad (3.45)$$

o ile tylko macierz $\mathbf{S}^{-1} = [s_{ij}^*]$ jest nieujemna, tzn. $s_{ij}^* \leq 0$, [21]. W przypadku przeciwnym, $\underline{x}_i > \bar{x}_i$ dla przynajmniej pewnych $i \in \{1, \dots, n\}$ i pewnych liczb tworzących prawą stronę układu równań.

Przykład 3.4.3. W przypadku układu równań z przykładu 3.4.2 macierz odwrotna \mathbf{S}^{-1} ma postać

$$\mathbf{S}^{-1} = \frac{1}{8} \begin{bmatrix} 9 & -1 & 3 & -3 \\ -3 & 3 & -1 & 1 \\ 3 & -3 & 9 & -1 \\ -1 & 1 & -3 & 3 \end{bmatrix}$$

Nie jest to więc macierz nieujemna.

Założmy, że $\tilde{y}_1 \sim \Delta(0, 1, 2)$, $\tilde{y}_2 \sim \Delta(4, 6, 8)$. Parametryczna reprezentacja obu tych liczb ma postać $[\tilde{y}_1]_\alpha = [\alpha, 2 - \alpha]$, $[\tilde{y}_2]_\alpha = [4 + 2\alpha, 8 - 2\alpha]$. W tym wypadku

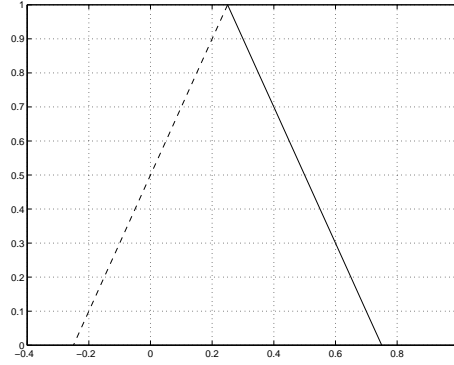
$$\mathbf{X} = \mathbf{S}^{-1} \mathbf{Y} = \frac{1}{8} \begin{bmatrix} 14 + 4\alpha \\ 6 + 4\alpha \\ -22 + 4\alpha \\ -14 + 4\alpha \end{bmatrix}$$

czyli $[\tilde{x}_1]_\alpha = [1.75 + 0.5\alpha, 2.75 - 0.5\alpha]$ i $[\tilde{x}_2]_\alpha = [0.75 + 0.5\alpha, 1.75 - 0.5\alpha]$. Sprawdzamy, że $\underline{x}_i(\alpha) \leq \bar{x}_i(\alpha)$ dla dowolnej wartości $\alpha \in [0, 1]$ oraz $i = 1, 2$.

Założmy teraz, że $\tilde{y}_1 \sim \Delta(-2, 0, 2)$, $\tilde{y}_2 \sim \Delta(0, 1, 2)$. W tym wypadku $\underline{x}_2(\alpha) = 0.75 - 0.5\alpha$, natomiast $\bar{x}_2(\alpha) = -0.25 + 0.5\alpha$, a więc $\underline{x}_2(\alpha)$ jest malejącą, a $\bar{x}_2(\alpha)$ – rosnącą funkcją α – por. rysunek 3.12. A więc rozważany tu układ równań nie ma dla takiego \mathbf{y} rozwiązania. □

Podobnie jak w przypadku równania liniowego, jeżeli składowe wektora $\mathbf{X} = \mathbf{S}^{-1} \mathbf{Y}$ reprezentują liczby rozmyte $\tilde{x}_1, \dots, \tilde{x}_n$, to mówimy, że układ równań (3.40) posiada silne rozwiązanie. W przypadku przeciwnym wyznaczamy α -warstwy

$$\begin{aligned} \underline{u}_i(\alpha) &= \min[\underline{x}_i(\alpha), \bar{x}_i(\alpha), \underline{x}_i(1)] \\ \bar{u}_i(\alpha) &= \max[\underline{x}_i(\alpha), \bar{x}_i(\alpha), \underline{x}_i(1)] \end{aligned}$$



Rysunek 3.12: Przykład słabego rozwiązania. Dla układu równań z przykładu 3.4.2 i rozmytego wektora $\mathbf{y} = (\Delta(-2, 0, 2), \Delta(0, 1, 2))$ linią ciągłą przedstawiono funkcję $\underline{x}_2(\alpha)$, a przerywaną – funkcję $\bar{x}_2(\alpha)$. tak wyznaczone α -warstwy nie reprezentują liczby rozmytej. Jednakże $[\tilde{u}]_\alpha = [\bar{x}_2(\alpha), \underline{x}_2(\alpha)]$ jest dobrze określoną liczbą rozmytą, czyli słabym rozwiązaniem.

Uwzględnienie w powyższych równaniach wartości $\underline{x}_i(1)$ zapobiega konstrukcji liczb rozmytych, których funkcja przynależności przypomina kształtem trójkąt rozwartokątny.

Jeżeli tak zdefiniowane \tilde{u}_i jest liczbą rozmytą, mówimy o słabym rozwiązaniu układu (3.40). Oczywiście, jeżeli \tilde{x}_i jest liczbą rozmytą, to $\tilde{u}_i = \tilde{x}_i$.

Macierz \mathbf{S} ma szczególną postać, mianowicie

$$\mathbf{S} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C} & \mathbf{B} \end{bmatrix}$$

gdzie \mathbf{B} jest macierzą zawierającą dodatnie elementy oryginalnej macierzy \mathbf{A} , natomiast \mathbf{C} – macierzą zawierającą wartości bezwzględne ujemnych elementów macierzy \mathbf{A} . W [21] pokazano, że identyczną strukturę ma macierz \mathbf{S}^{-1} , tzn.

$$\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{E} & \mathbf{D} \end{bmatrix}$$

przy czym

$$\begin{aligned} \mathbf{D} &= \frac{1}{2}[(\mathbf{B} + \mathbf{C})^{-1} + (\mathbf{B} - \mathbf{C})^{-1}] \\ \mathbf{E} &= \frac{1}{2}[(\mathbf{B} + \mathbf{C})^{-1} - (\mathbf{B} - \mathbf{C})^{-1}] \end{aligned} \quad (3.46)$$

Własność tę wykorzystano w interesujący sposób w pracy [1]. Korzystając z formuły (3.35) i -te równanie układu (3.40) zapisujemy w postaci układu dwóch równań

$$\begin{cases} \sum_{a_{ij}>0} a_{ij}\underline{x}_j(\alpha) + \sum_{a_{ij}<0} a_{ij}\bar{x}_j(\alpha) = \underline{y}_i(\alpha) \\ \sum_{a_{ij}>0} a_{ij}\bar{x}_j(\alpha) + \sum_{a_{ij}<0} a_{ij}\underline{x}_j(\alpha) = \bar{y}_i(\alpha) \end{cases} \quad (3.47)$$

Odejmując pierwsze z tych równań od drugiego dostajemy jedno równanie postaci

$$\sum_{a_{ij}>0} a_{ij}[\bar{x}_j(\alpha) - \underline{x}_j(\alpha)] - \sum_{a_{ij}<0} a_{ij}[\bar{x}_j(\alpha) - \underline{x}_j(\alpha)] = \bar{y}_i(\alpha) - \underline{y}_i(\alpha), i = 1, \dots, n \quad (3.48)$$

Oznaczmy, jak poprzednio, przez \mathbf{B} macierz zawierającą dodatnie elementy macierzy \mathbf{A} , a przez \mathbf{C} – macierz zawierającą wartości bezwzględne ujemnych elementów macierzy \mathbf{A} . Innymi słowy $\mathbf{A} = \mathbf{B} - \mathbf{C}$. Oznaczając

$$w_j = \bar{x}_j - \underline{x}_j, \quad v_j = \bar{y}_j - \underline{y}_j$$

definiujemy wektory $\mathbf{w} = (w_1, \dots, w_n)^T$, $\mathbf{v} = (v_1, \dots, v_n)^T$. Przy tych oznaczeniach układ (3.48), który przyjmuje teraz postać

$$\sum_{a_{ij}>0} a_{ij}w_j - \sum_{a_{ij}<0} a_{ij}w_j = v_i, i = 1, \dots, n$$

możemy zapisać w postaci macierzowej (zauważmy, że współczynniki a_{ij} występujące w odjemniku powyższego równania są ujemne)

$$(\mathbf{B} + \mathbf{C})\mathbf{w} = \mathbf{v} \quad (3.49)$$

Niech dalej $x_i^c = (\bar{x}_i + \underline{x}_i)/2$ będzie punktem centralnym odcinka $[\underline{x}_i, \bar{x}_i]$ i niech $\mathbf{x}^c = (x_1^c, \dots, x_n^c)^T$. Jeżeli \mathbf{x} jest rozwiązaniem układu (3.40), to $\mathbf{A}\mathbf{x}^c = \mathbf{y}^c$. Można to sprawdzić sumując oba równania układu (3.47) i dzieląc je stronami przez 2. Innymi słowy, wektor \mathbf{x}^c wyznaczamy rozwiązując układ równań

$$(\mathbf{B} - \mathbf{C})\mathbf{x}^c = \mathbf{y}^c \quad (3.50)$$

Dysponując wektorami \mathbf{w} oraz \mathbf{x}^c , dolne $\underline{\mathbf{x}}$ i górne $\bar{\mathbf{x}}$ odgraniczenia rozwiązania \mathbf{x} znajdujemy w następujący sposób

$$\begin{aligned} \underline{\mathbf{x}} &= \mathbf{x}^c - \frac{1}{2}\mathbf{w} \\ \bar{\mathbf{x}} &= \mathbf{x}^c + \frac{1}{2}\mathbf{w} \end{aligned} \quad (3.51)$$

Oczywiście wektor \mathbf{w} wyznaczamy rozwiązując układ równań (3.49), a wektor \mathbf{x}^c – rozwiązując układ (3.50).

Twierdzenie 3.4.1. Układ równań (3.40) ma jednoznaczne rozmyte rozwiązanie zadane przez zbiór α -warstw $\mathbf{x}(\alpha)$, $\alpha \in [0, 1]$, jeżeli istnieją macierze $(\mathbf{B} + \mathbf{C})^{-1}$ i $(\mathbf{B} - \mathbf{C})^{-1}$, a ponadto $(\mathbf{B} + \mathbf{C})^{-1}$ jest macierzą nieujemną. Rozwiązanie \mathbf{x} nazywamy mocnym rozwiązaniem układu (3.40).

Dowód: wynika z obserwacji (3.46); por. [1]. □

Przykład 3.4.4. Rozważmy ponownie układ równań z przykładu ?? i załóżmy, że $\tilde{y}_1 \sim \Delta(0, 1, 2)$, $\tilde{y}_2 \sim \Delta(4, 6, 8)$. Skoro parametryczna reprezentacja obu tych liczb ma postać $[\tilde{y}_1]_\alpha = [\alpha, 2 - \alpha]$, $[\tilde{y}_2]_\alpha = [4 + 2\alpha, 8 - 2\alpha]$, to wektory \mathbf{v} i \mathbf{y}^c są postaci

$$\mathbf{v} = \begin{bmatrix} 2 - 2\alpha \\ 4 - 4\alpha \end{bmatrix}, \mathbf{y}^c = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

Z kolei macierze \mathbf{B} i \mathbf{C} wyglądają następująco

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Stąd

$$(\mathbf{B} - \mathbf{C})^{-1} = \begin{bmatrix} 0.75 & 0.25 \\ -0.25 & 0.25 \end{bmatrix}, (\mathbf{B} + \mathbf{C})^{-1} = \begin{bmatrix} 1.5 & -1.5 \\ -0.5 & 0.5 \end{bmatrix}$$

A więc rozwiązania układów równań (3.49) i (3.50) mają postać

$$\mathbf{x}^c = (\mathbf{B} - \mathbf{C})^{-1} \mathbf{y}^c = \begin{bmatrix} 2.25 \\ 1.25 \end{bmatrix}, \mathbf{w} = (\mathbf{B} + \mathbf{C})^{-1} \mathbf{v} = \begin{bmatrix} 1 - \alpha \\ 1 - \alpha \end{bmatrix}$$

Z równania (3.51) otrzymujemy, że parametryczną postacią rozwiązanie badanego tu równania jest

$$\mathbf{x} = \begin{bmatrix} 2.25 - \frac{1}{2}(1 - \alpha), 2.25 + \frac{1}{2}(1 - \alpha) \\ 1.25 - \frac{1}{2}(1 - \alpha), 1.25 + \frac{1}{2}(1 - \alpha) \end{bmatrix}$$

□

3.4.4.2 Metoda geometryczna

Przedstawmy prawą stronę \mathbf{y} układu równań (3.40) w postaci

$$[\mathbf{y}]_\alpha = \mathbf{m} + \tilde{\mathbf{y}}(\alpha) \quad (3.52)$$

gdzie

$$\mathbf{m} = (\underline{y}_1(1), \dots, \underline{y}_n(1))^T$$

jest wektorem średnich, a składowe $\tilde{y}_i(\alpha)$ wektora $\tilde{\mathbf{y}}(\alpha)$ mają postać

$$\tilde{y}_i(\alpha) = [\underline{z}_i(\alpha), \bar{z}_i(\alpha)] = [\underline{y}_i(\alpha) - \underline{y}_i(1), \bar{y}_i(\alpha) - \underline{y}_i(1)]$$

Zauważmy, że dla każdej wartości $\alpha \in [0, 1]$ dolne ograniczenie $\underline{z}_i(\alpha)$ składowej $\tilde{y}_i(\alpha)$ jest liczbą niedodatnią, a górne ograniczenie – liczbą nieujemną.

Tym samym rozwiązanie układu równań (3.40) przyjmuje teraz postać

$$\mathbf{x}(\alpha) = \mathbf{A}^{-1} \mathbf{y}(\alpha) = \mathbf{A}^{-1} \mathbf{m} + \mathbf{A}^{-1} \tilde{\mathbf{y}}(\alpha) = \mathbf{x}_d + \tilde{\mathbf{x}}(\alpha) \quad (3.53)$$

gdzie przez \mathbf{x}_d oznaczono składową deterministyczną rozwiązania, a przez $\tilde{\mathbf{x}}(\alpha)$ – składową rozmytą.

Aby wyznaczyć tę ostatnią składową, definiuje się zbiór

$$\Pi(\alpha) = \{\mathbf{v} = (v_1, \dots, v_n)^T \in \mathbb{R}^n : \underline{z}_i(\alpha) \leq v_i \leq \bar{z}_i(\alpha)\} \quad (3.54)$$

Wówczas zbiór $\mathbf{A}^{-1} \Pi(\alpha)$ definiujemy w następujący sposób

$$\mathbf{A}^{-1} \Pi(\alpha) = \{\mathbf{A}^{-1} \mathbf{v} : \mathbf{v} \in \Pi(\alpha)\} \quad (3.55)$$

Skoro każda składowa v_i należy do odpowiedniej liczby rozmytej y_i w stopniu $\alpha_i = \mu_{y_i}(v_i)$, to stopień przynależności wektora $\mathbf{v} \in \Pi$ jest równy $\alpha = \min_{i=1, \dots, n} \alpha_i$. Oznacza to, że we wprowadzonych w tej sekcji wzorach możemy pominąć symbol α wyznaczając go w razie potrzeby zgodnie z opisaną konwencją. W dalszym ciągu będziemy utożsamiać zbiór Π ze zbiorem $\Pi(0)$.

Stosując powyższą konwencję powiemy za [23], że wektor \mathbf{x} jest rozwiązaniem układu równań (3.40) w stopniu α jeżeli odpowiadający mu wektor $\mathbf{b} = \mathbf{A}\mathbf{x}$ należy do zbioru $\mathbf{m} + \Pi$ w stopniu α .

Twierdzenie 3.4.2. Jeżeli macierz współczynników układu równań (3.40) jest odwracalna, to zbiór rozwiązań S tego układu ma postać $S = \mathbf{x}_d + \mathbf{A}^{-1} \Pi$.

Dowód: (DOPRACOWAĆ) Niech \mathbf{x} należy do zbioru $\mathbf{x}_d + \mathbf{A}^{-1} \Pi$ w stopniu α . Na mocy przyjętej konwencji oznacza to, że istnieje taki wektor $\mathbf{z} \in \Pi$, że $\mathbf{x} = \mathbf{x}_d + \mathbf{A}^{-1} \mathbf{z} = \mathbf{A}^{-1}(\mathbf{m} + \mathbf{z})$. A więc stopnie przynależności wektorów \mathbf{x} i $\mathbf{m} + \mathbf{z}$ są identyczne i równe α .

Niech teraz \mathbf{x} będzie pewnym rozwiązaniem, którego stopień przynależności do zbioru S jest równy α . Wówczas musi istnieć taki wektor \mathbf{z} , że $\mathbf{z} \in (\mathbf{x}_d + \Pi)$ oraz $\mathbf{A}\mathbf{x} = \mathbf{z}$. □

Jeżeli składowymi wektora \mathbf{y} są trójkątne liczby rozmyte, $y_i \sim \Delta(\underline{y}_i, m_i, \bar{y}_i)$ to algorytm wyznaczania rozwiązania układu (3.40) jest następujący [23]:

- (1.) Wyznaczyć macierz odwrotną \mathbf{A}^{-1} .
- (2.) Wyznaczyć wektor $\mathbf{m} = (m_1, \dots, m_n)^T$ oraz wektor $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_n)^T$, gdzie $\tilde{y}_i = (\underline{y}_i - m_i, 0, \bar{y}_i - m_i) = (\underline{z}_i, 0, \bar{z}_i)$.
- (3.) Wyznaczyć deterministyczne rozwiązanie $\mathbf{x}_d = \mathbf{A}^{-1}\mathbf{m}$.
- (4.) Wyznaczyć wektory

$$\begin{array}{ll} \mathbf{v}_1 = (\underline{z}_1, 0, 0, \dots, 0)^T & \mathbf{u}_1 = (\bar{z}_1, 0, 0, \dots, 0)^T \\ \mathbf{v}_2 = (0, \underline{z}_2, 0, \dots, 0)^T & \mathbf{u}_2 = (0, \bar{z}_2, 0, \dots, 0)^T \\ \dots & \dots \\ \mathbf{v}_n = (0, 0, 0, \dots, \underline{z}_n)^T & \mathbf{u}_n = (0, 0, 0, \dots, \bar{z}_n)^T \end{array}$$

- (5.) Wyznaczyć zbiór rozwiązań postaci

$$S = \left\{ \mathbf{x}_d + \sum_{i=1}^n \alpha_i \mathbf{A}^{-1} \mathbf{w}_i : \mathbf{w}_i \in \{\mathbf{v}_i, \mathbf{u}_i\}, \alpha_i \in [0, 1] \right\}$$

Stopień przynależności takiego rozwiązania do zbioru S jest równy $\mu_S(\mathbf{x}) = 1 - \max_{i=1, \dots, n} \alpha_i$.

– aaa

Przykład 3.4.5. Rozważmy ponownie układ równań z przykładu 3.4.4. Macierz odwrotna współczynników tego układu ma postać

$$\mathbf{A}^{-1} = \frac{1}{4} \begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix}$$

a więc rozwiązaniem medialnym jest, jak poprzednio, wektor $\mathbf{x}_m = (2.25, 1.25)^T$.

Punktami wierzchołkowymi zbioru Π są natomiast

$$\mathbf{p}_1 = \mathbf{x}_m + \mathbf{A}^{-1} \begin{bmatrix} -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}, \mathbf{p}_2 = \mathbf{x}_m + \mathbf{A}^{-1} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.0 \\ 2.0 \end{bmatrix}$$

$$\mathbf{p}_3 = \mathbf{x}_m + \mathbf{A}^{-1} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix}, \mathbf{p}_4 = \mathbf{x}_m + \mathbf{A}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 1.5 \end{bmatrix}$$

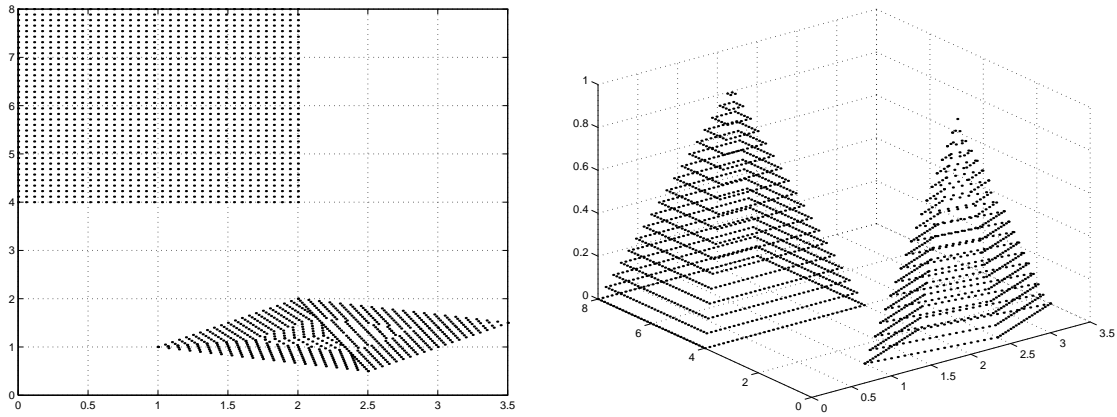
Podobnie możemy określić kształt zbioru $\Pi(\alpha)$ zastępując wektor $\tilde{\mathbf{y}}$ jego α -odcięciem $[\tilde{\mathbf{y}}]_\alpha$. Warstwie rozmytego rozwiązania przedstawiono na rysunku 3.13. \square

3.5 Zadania

Problem 3.5.1. Dla trójkątnej liczby rozmytej $\tilde{x} \sim \Delta(3, 5, 9)$ wyznaczyć:

- (a) analityczną postać funkcji przynależności $\mu_{\tilde{x}}$,
- (b) liczbę przeciwną $-\tilde{x}$ do liczby \tilde{x} , tzn. podać analityczną postać funkcji przynależności $\mu_{-\tilde{x}}$ i sporządzić wykres tej funkcji, oraz
- (c) narysować wykres funkcji przynależności dla liczby $\tilde{x} - \tilde{x}$.

Jak wyglądają α -warstwy liczby $-\tilde{x}$? Jak można skomentować wynik z punktu (c)?



Rysunek 3.13: Piramidalne rozwiązanie układu równań z przykładu 3.4.5. W lewej części rysunku przedstawiono nośnik zbioru $\tilde{y}_1 \times \tilde{y}_2$ oraz nośnik rozmytego rozwiązania, a w prawej stosowne zbiory piramidalne.

Problem 3.5.2. Niech $f(x) = x^2$ i niech \tilde{x} będzie trójkątną liczbą rozmytą. Stosując zasadę rozszerzania znaleźć $\mu_{f(\tilde{x})}(9)$ jeżeli:

- (a) $\tilde{x} \sim \Delta(2, 5, 10)$
- (b) $\tilde{x} \sim \Delta(-4, -1, 5)$

Problem 3.5.3. Wyznaczyć obraz liczby rozmytej \tilde{x} w odwzorowaniu $f(x) = x^2$ dla przypadków (a) i (b) z problemu 3.5.2.

Problem 3.5.4. Wyznaczyć odległość Hausdorffa między zbiorami $\tilde{A} \sim \Delta(-3, -1, 5)$ oraz $\tilde{B} \sim \Delta(2, 4, 6)$.

Problem 3.5.5. Oprócz odległości Hausdorffa definiuje się także odległość Hamminga

$$H(\tilde{A}, \tilde{B}) = \int_{-\infty}^{\infty} |\mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x)| dx$$

oraz odległość

$$C_{\infty}(\tilde{A}, \tilde{B}) = \sup_{x \in \mathbb{R}} |\mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x)|$$

Wyznaczyć obie te odległości między zbiorami $\tilde{A} \sim \Delta(-3, -1, 5)$ i $\tilde{B} \sim \Delta(2, 4, 6)$.

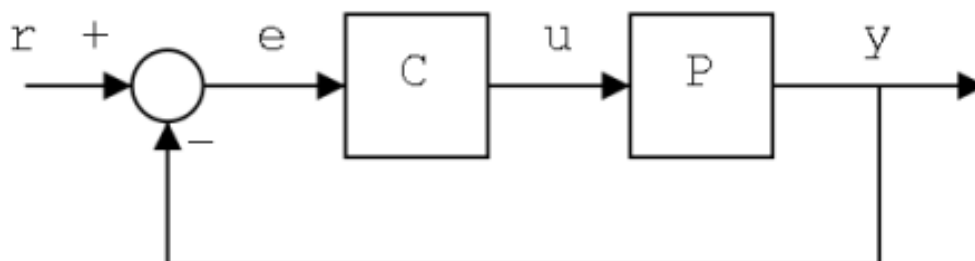
Problem 3.5.6. Stosując metodę wierzchołkową wyznaczyć obraz liczby rozmytej \tilde{x} w odwzorowaniu $f(x) = \sin x$ jeżeli:

- a) $\tilde{x} \sim \Delta(-\pi/4, 0, \pi/4)$,
- b) $\tilde{x} \sim \Delta(\pi/4, \pi/2, 3\pi/4)$

Rozdział 4

Sterowanie i sterowniki rozmyte

Regulatory to urządzenia generujące sygnał sterujący w oparciu o tzw. *uchyb regulacji*, czyli różnicę między wartością zadaną sygnału regulowanego a wartością aktualnie zmierzoną – por. rysunek 4.1. Za przykłady mogą służyć regulatory ciśnienia, temperatury, czy regulatory natężenia przepływu.



Rysunek 4.1: Podstawowy schemat sterowania. Sygnał wyjściowy y układu sterowanego P jest porównywany z zadaną wartością r , a różnica tych wielkości, czyli uchyb regulacji e jest podawana na wejście regulatora C . Dzięki temu regulator posiada informację o aktualnym błędzie sterowania i może tak zmieniać wartość wejściową u obiektu P , aby zmniejszyć uchyb do zera, gwarantując utrzymywanie wielkości wyjściowej układu możliwie blisko wartości zadanej.

W przypadku np. regulatorów temperatury można wyróżnić dwa rodzaje rozwiązań¹:

- Czujnik temperatury jest umieszczony wewnątrz zaworu i dzięki zmianie swych wymiarów powoduje przesuwanie grzybka względem gniazda.
- Czujnikiem jest termometr manometryczny dostarczający ciśnienie dla sprężystego mieszka, którego denko wprawia w ruch trzpień zaworu regulacyjnego.

Ogólnie, rolą regulatora jest doprowadzanie obiektu dożądanego stanu, ewentualnie poprawa niekorzystnych cech obiektu regulowanego lub też utrzymywanie obiektu w pożądanym stanie (jak to ma miejsce w przypadku regulatorów temperatury).

W automatyce przemysłowej² pod pojęciem regulatora (nazywanego też sterownikiem lub kontrolerem³) rozumie się samodzielny układ (komputerowy, elektryczny, elektroniczny,

¹Por. np. <http://www.sciaga.pl/tekst/34497-35-regulatory-bezposredniego-dzialania>.

²Rozumianej jako dziedzina wiedzy, która zajmuje się możliwościami wyeliminowania lub ograniczenia udziału człowieka w czynnościach związanych ze sterowaniem różnorodnymi procesami, głównie technologicznymi i przemysłowymi.

³Nazwę tę wprowadza się, aby odróżnić je od programowych sterowników urządzeń w systemie operacyjnym.

elektromechaniczny) zajmujący się nadzorowaniem pewnego urządzenia czy procesu technologicznego.

Podstawowe typy regulatorów to: regulator proporcjonalny P , proporcjonalno-różniczkujący, PD oraz proporcjonalno-całkujący, PI , a ich uogólnieniem jest regulator proporcjonalno-całkująco-różniczkujący, PID ; składa się on z członu proporcjonalnego P o wzmocnieniu K_p , całkującego I o czasie zdwojenia T_i oraz różniczkującego D o czasie wyprzedzenia T_d . Realizuje on algorytm

$$u(t) = K_p \left[e(t) + \frac{1}{T_i} \int e(t) dt + T_d \frac{de(t)}{dt} \right] \quad (4.1)$$

gdzie $u(t)$ oznacza sygnał wyjściowy, a $e(t)$ – uchyb regulacji (sygnał wejściowy).

W dalszym ciągu będziemy używać nazwy *sterownik* dla odróżnienia regulatora implementowanego jako program komputerowy w odróżnieniu od wspomnianych wcześniej regulatorów realizowanych w postaci urządzeń mechanicznych.

Projektowanie sterownika wymaga znajomości modelu procesu P poddawanego sterowaniu. Jednakże znalezienie odpowiedniego modelu jest problemem czasochłonnym i trudnym, wymuszającym zazwyczaj przyjęcie szeregu założeń upraszczających.

Zastosowanie teorii zbiorów rozmytych do sterowania procesami technologicznymi (ang. *Fuzzy Logic Control*) nie wymaga szczegółowej znajomości modeli tych procesów⁴. Formuluje się jedynie reguły postępowania w postaci rozmytych zdań warunkowych typu **if** A **then** B , gdzie przesłanka A określa warunki jakie muszą być spełnione aby zastosować akcję przewidzianą w konkluzji B . Jeżeli np. sygnałami wejściowymi sterownika układu klimatyzacji są wartość temperatury i wilgotności w pomieszczeniu, a sygnałem wyjściowym jest intensywność chłodzenia to przykładowa reguła może mieć postać:

if [(temperatura jest *wysoka*) **and** (wilgotność jest *duża*)]
then (intensywność chłodzenia jest **duża**)

Sterowniki rozmyte są przykładami urządzeń działających w oparciu o wiedzę pozyskiwaną na jeden z czterech sposobów:

- odwołanie się do wiedzy eksperta z wykorzystaniem pomocy inżyniera wiedzy,
- modelowanie zachowania operatora procesu,
- wykorzystanie rozmytego modelu procesu,
- zastosowanie algorytmów uczących się i samoorganizujących

Z tego powodu sterowniki rozmyte traktowane są jako reprezentanci sterowania eksperckiego (ang. *Expert Control*).

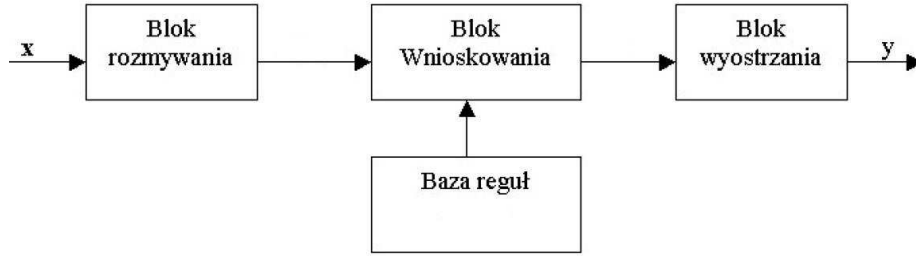
Sterowniki konstruowane z wykorzystaniem teorii zbiorów rozmytych znajdują obecnie szerokie zastosowanie ze względu na prostotę i taniość implementacji. Są wykorzystywane zarówno w prostych urządzeniach gospodarstwa domowego (pralki, lodówki, itp.) jak również w sterowaniu złożonymi procesami technologicznymi.

4.1 Architektura rozmytego sterownika

Schemat blokowy sterownika rozmytego przedstawiono na rysunku 4.2.

W zastosowaniach wyróżnia się dwa typy sterowników rozmytych:

⁴Wspomnijmy tu o pionierskiej pracy Mamdani, E.H., Assilian, S. An experiment in linguistic synthesis with fuzzy controller. *Internat. J. of Man-Machine Studies*. **7**, 1975, 1-13.



Rysunek 4.2: Schemat blokowy sterownika rozmytego.

- (a) sterowniki Mamdaniego, w których zarówno przesłanki jak i konkluzje reguł mają charakter rozmyty,
- (b) sterowniki Takagi-Sugeno, w których przesłanki mają charakter rozmyty, a konkluzje – charakter funkcyjny, np.

if (x **is** A) **then** ($u = f(x)$)

Poniżej omawiamy dokładniej poszczególne bloki tworzące sterownik rozmyty.

4.1.1 Baza reguł

Baza reguł, nazywana też *modelem lingwistycznym* lub rozmytą pamięcią asocjacyjną (*FAM* – *Fuzzy Associative Memory*) składa się z rozmytych reguł $R^{(k)}$, $k = 1, \dots, N$, postaci

$$R^{(k)}: \text{if } [(x_1 \text{ is } A_1^{(k)}) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^{(k)})] \\ \text{then } [(u_1 \text{ is } B_1^{(k)})] \text{ and } \dots \text{ and } (u_m \text{ is } B_m^{(k)})] \quad (4.2)$$

gdzie $A_i^{(k)} \subset X_i \subset \mathbb{R}$, $i = 1, \dots, n$ są rozmytymi podzbiorami etykietującymi obserwowane wartości sygnałów wejściowych $x_i \in X_i$, a $B_j^{(k)} \subset Y_j \subset \mathbb{R}$, $j = 1, \dots, m$ są rozmytymi podzbiorami etykietującymi obserwowane wartości sygnałów wyjściowych $u_j \in Y_j$.

Zmienne wejściowe $(x_1, \dots, x_n)^T$ tworzą wektor obserwacji $\mathbf{x} \in \mathbf{X} = X_1 \times \dots \times X_n$; podobnie zmienne wyjściowe $(u_1, \dots, u_m)^T$ tworzą wektor sterowań $\mathbf{u} \in \mathbf{U} = U_1 \times \dots \times U_m$.

Tworząc zbiór reguł (4.2) należy pokrywać dziedziny zmiennych X_1, \dots, X_n w taki sposób, aby zbiory rozmyte charakteryzujące etykiety używane w regułach pokrywały każdą obserwację, tzn. aby

$$(\forall x_i \in X_i)(\exists j \in \{1, \dots, k_i\}): \mu_{A_j}(x_i) \geq \delta$$

gdzie k_i to liczba zbiorów rozmytych (nazywana też rozdzielczością) pokrywających dziedzinę X_i , δ to tzw. stopień kompletności pokrycia dziedziny. Uwzględniając uwagi z poprzedniego punktu, przyjmujemy $\delta = 0.5$. O j -tym zbiorze rozmytym pokrywającym dziedzinę X_i zakłada się ponadto, że jest on normalny i wypukły.

Podobnie jak w klasycznej bazie wiedzy, przyjmuje się, że reguły $R^{(k)}$, $k = 1, \dots, N$ są powiązane ze sobą za pomocą spójnika logicznego **lub**. Ponadto zakładamy, że wyjścia u_1, \dots, u_m są wzajemnie niezależne. Bez utraty ogólności można wówczas rozpatrywać reguły z atomową (elementarną) konkluzją postaci

$$R^{(k)}: \text{if } [(x_1 \text{ is } A_1^{(k)}) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^{(k)})] \\ \text{then } (u \text{ is } B^{(k)}) \quad (4.3)$$

W regule (4.3) część **if** nazywamy poprzednikiem, a część **then** następnikiem lub konkluzją. Poprzednik jest tzw. przesłanką *złożoną* stanowiącą połączenie za pomocą operatorów logicznych (w tym przypadku – operatora **and**) przesłanek *prostych* postaci $(x_i \text{ is } A_i^{(k)})$, $i = 1, \dots, n$.

Oznaczając przez $A^{(k)}$ iloczyn kartezjański zbiorów $A_1^{(k)}, \dots, A_n^{(k)}$ możemy regule (4.3) nadać postać rozmytej implikacji

$$R^{(k)}: A^{(k)} \rightarrow B^{(k)} \quad (4.4)$$

Regulę (4.3) można też interpretować jako relację rozmytą określoną na zbiorze $\mathbf{X} \times Y$ o funkcji przynależności

$$\mu_{R^{(k)}}(\mathbf{x}, u) = \mu_{A^{(k)} \rightarrow B^{(k)}}(\mathbf{x}, u) \quad (4.5)$$

Często blok zawierający definicje funkcji przynależności bywa nazywany bazą danych, a baza reguł wraz z bazą danych tworzą w takim kontekście bazę wiedzy.

4.1.2 Blok rozmywania

Skoro przesłanki reguły $R^{(k)}$ mają postać rozmytą, to konkretnej obserwacji $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)$ należy przydzielić stopnie przynależności do odpowiednich zbiorów rozmytych wymienionych w przesłance reguły.

Najczęściej zakłada się, że obserwowane są dokładne wartości zmiennych wejściowych. W takim przypadku wartość $\mu_{A^{(k)}}(\bar{\mathbf{x}})$ interpretujemy jako stopień zgodności wektora obserwacji z rozmytą przesłanką k -tej reguły. Np. jeżeli zbiór $A^{(k)}$ opisany jest trójkątną funkcją przynależności (2.14) postaci $(-10, 10, 30)$ to obserwacji $\bar{x}_i = 25$ przypisany zostanie stopień $\mu_{A^{(k)}}(25) = 0.25$ kwantyfikujący prawdziwość stwierdzenia " $x_i = 25$ is $A^{(k)}$ ".

Znając zgodność obserwacji \bar{x}_i z elementarną przesłanką, tzn. warunkiem zadany w postaci zbioru rozmytego $A_i^{(k)}$ trzeba jeszcze wyznaczyć stopień spełnialności złożonej przesłanki $A^{(k)}$ przez wektor obserwacji $\bar{\mathbf{x}}$. Proces wyznaczania stopnia zgodności dla przesłanki złożonej w oparciu o znajomość stopni zgodności przesłanek prostych nazywa się *agregacją*. Jeżeli przesłanka jest złożona z n przesłanek prostych połączonych spójnikiem **and**, jak w przypadku reguły (4.3), to dla danych wartości argumentów $x_i = \bar{x}_i$ stopień jej prawdziwości obliczamy jako stopień przynależności do relacji $A^{(k)}$:

$$\begin{aligned} \mu_{A^{(k)}}(\bar{x}_1, \dots, \bar{x}_n) &= \mu_{A_1^{(k)} \cap \dots \cap A_n^{(k)}}(\bar{x}_1, \dots, \bar{x}_n) \\ &= \mu_{A_1^{(k)}}(\bar{x}_1) * \dots * \mu_{A_n^{(k)}}(\bar{x}_n) \end{aligned} \quad (4.6)$$

gdzie $*$ jest symbolem normy trójkątnej (w skrócie T-normy).

Przypominamy, że funkcję $T: [0, 1] \times [0, 1] \rightarrow [0, 1]$ nazywamy T-normą, jeżeli dla dowolnych $a, b, c, d \in [0, 1]$ spełnia ona następujące warunki [75], [72], [64]:

- (a) $T(a, 0) = 0$, $T(a, 1) = a$ (ograniczoność)
- (b) $T(a, b) = T(b, a)$ (przemienność)
- (c) $T(T(a, b), c) = T(a, (b, c))$ (łączność)
- (d) $T(a, c) \leq T(b, d)$ gdy $a \leq b, c \leq d$ (monotoniczność)

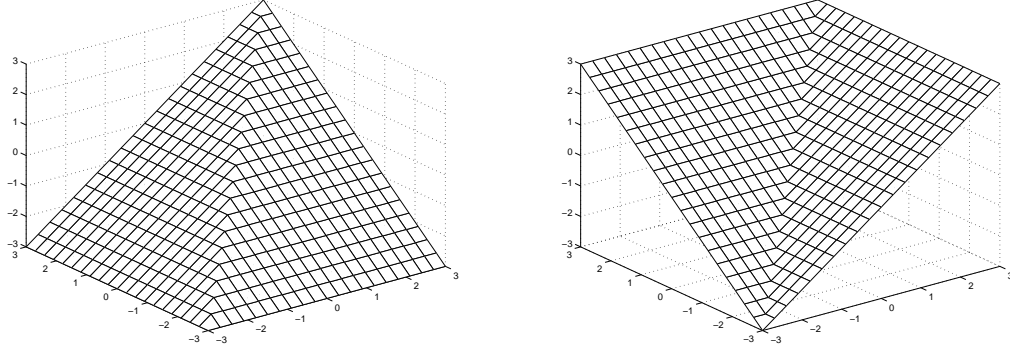
Dowolna T-norma jest ograniczona w sposób następujący

$$T_w(a, b) \leq T(a, b) \leq \min(a, b) \quad (4.7)$$

gdzie T_w jest T-normą postaci

$$T_w(a, b) = \begin{cases} a & \text{gdy } b = 1 \\ b & \text{gdy } a = 1 \\ 0 & \text{gdy } a, b \neq 1 \end{cases} \quad (4.8)$$

T-normę $\min(a, b)$ przedstawia lewa część rysunku 4.3. W teorii zbiorów rozmytych często zamiast $T(a, b)$ piszemy $a * b$.



Rysunek 4.3: Wykresy T-normy $\min(a, b)$ (lewy rysunek) oraz dualnej S-normy $\max(a, b)$ (prawy rysunek).

W przypadku (rzadziej stosowanych) przesłanek stanowiących alternatywę przesłanek prostych, stopień prawdziwości obliczamy jako stopień przynależności do relacji $A^{(k)} = A_1^{(k)}$ **or** \dots **or** $A_n^{(k)}$:

$$\begin{aligned} \mu_{A^{(k)}}(\bar{x}_1, \dots, \bar{x}_n) &= \mu_{A_1^{(k)} \cup \dots \cup A_n^{(k)}}(\bar{x}_1, \dots, \bar{x}_n) \\ &= \mu_{A_1^{(k)}}(\bar{x}_1) \diamond \dots \diamond \mu_{A_n^{(k)}}(\bar{x}_n) \end{aligned} \quad (4.9)$$

gdzie \diamond jest symbolem ko-normy trójkątnej, nazywanej w skrócie S-normą.

Funkcję dwóch zmiennych $S: [0, 1] \times [0, 1] \rightarrow [0, 1]$ nazywamy S-normą jeżeli jest ona nierosnąca względem obu argumentów, spełnia ona wymienione wcześniej warunki przemienności i łączności oraz następujące warunki brzegowe

$$S(a, 0) = a, S(a, 1) = 1 \quad (4.10)$$

Dowolna S-norma jest ograniczona w sposób następujący

$$\max(a, b) \leq S(a, b) \leq S_w(a, b) \quad (4.11)$$

gdzie S_w jest S-normą postaci

$$S_w(a, b) = \begin{cases} a & \text{gdy } b = 0 \\ b & \text{gdy } a = 0 \\ 1 & \text{gdy } a, b \neq 0 \end{cases} \quad (4.12)$$

a S-normę $\max(a, b)$ przedstawiono w prawej części rysunku 4.3.

Działanie S-normy na argumentach a i b będziemy oznaczać $S(a, b) = a \diamond b$.

Każdej T-normie odpowiada dualna do niej S-norma, zgodnie z poniższym równaniem

$$a * b = 1 - [(1 - a) \diamond (1 - b)] \quad (4.13)$$

Najczęściej stosowane T-normy i S-normy przedstawiono w tabelicy 4.1.

Nr	$T(a, b)$	$S(a, b)$	parametry
1	$\min(a, b)$	$\max(a, b)$	
2	ab	$a + b - ab$	
3	$\max(a + b - 1, 0)$	$\min(a + b, 1)$	
4	$T_w(a, b)$	$S_w(a, b)$	
5	$\frac{ab}{\gamma + (1-\gamma)(a+b-ab)}$	$\frac{a+b-(2-\gamma)ab}{1-(1-\gamma)ab}$	$\gamma > 0$
6	$\frac{ab}{\max(a, b, \alpha)}$	$\frac{a+b-ab-\min(a, b, 1-\alpha)}{\max(1-a, 1-b, \alpha)}$	$\alpha \in [0, 1]$
7	$\max\left(\frac{a+b-1+\lambda ab}{1+\lambda}, 0\right)$	$\min(a + b + \lambda ab, 1)$	$\lambda \geq -1$

Tablica 4.1: Normy i ko-normy trójkątne

4.1.3 Blok wnioskowania

Zakładamy, że znamy zbiór rozmyty $\tilde{A} \subset \mathbf{X}$ i chcemy wyznaczyć zbiór rozmyty \tilde{B} odpowiadający przesłance reguły $R^{(k)}$. Do tego celu stosuje się uogólniony schemat *Modus Ponens* postaci

$$\begin{array}{c|c} \text{fakt} & \bar{\mathbf{x}} \text{ is } \tilde{A} \\ \text{reguła} & \text{if } (\mathbf{x} \text{ is } A^{(k)}) \text{ then } (u \text{ is } B^{(k)}) \\ \hline \text{konkluzja} & (\bar{u} \text{ is } \tilde{B}) \end{array} \quad (4.14)$$

Warunkiem jego stosowalności jest posiadanie przez regułę $R^{(k)}$ *własności interpolacyjnych* umożliwiających jej stosowanie także w tych przypadkach gdy obserwacja x jest *mniej więcej* równa $A^{(k)}$; wówczas wyciągane wnioski są także mniej więcej zgodne z rzeczywistością [64].

Formalizacją schematu (4.14) jest równanie

$$\mu_{\tilde{B}}(u) = \sup_{\mathbf{x} \in \mathbf{X}} [\mu_{A^{(k)}}(\mathbf{x}) * \mu_{A^{(k)} \rightarrow B^{(k)}}(\mathbf{x}, u)] \quad (4.15)$$

Najczęściej spotykane definicje rozmytej implikacji przedstawiono w tablicy 4.2.

Nr	Nazwa	Definicja
1	reguła Mamdaniego	$\mu_{A \rightarrow B}(x, u) = \min[\mu_A(x), \mu_B(u)]$
2	reguła typu iloczyn (Larsena)	$\mu_{A \rightarrow B}(x, u) = \mu_A(x) \cdot \mu_B(u)$
3	reguła Łukasiewicza	$\mu_{A \rightarrow B}(x, u) = \min[1, 1 - \mu_A(x) + \mu_B(u)]$
4	reguła typu max-min	$\mu_{A \rightarrow B}(x, u) = \max\left\{\min[\mu_A(x), \mu_B(u)], 1 - \mu_A(x)\right\}$
5	reguła probabilistyczna	$\mu_{A \rightarrow B}(x, u) = \min[1, 1 - \mu_A(x) + \mu_A(x)\mu_B(u)]$
6	reguła ograniczonej sumy	$\mu_{A \rightarrow B}(x, u) = \min[1, \mu_A(x) + \mu_B(u)]$
7	reguła binarna	$\mu_{A \rightarrow B}(x, u) = \max[1 - \mu_A(x), \mu_B(u)]$
8	reguła Goguena	$\mu_{A \rightarrow B}(x, u) = \min\left[1, \frac{\mu_B(x)}{\mu_A(u)}\right]$
9	reguła Gödela	$\mu_{A \rightarrow B}(x, u) = \begin{cases} 1 & \text{gdy } \mu_A(x) \leq \mu_B(u) \\ \mu_B(u) & \text{gdy } \mu_A(x) > \mu_B(u) \end{cases}$
10	regułą Sharpa	$\mu_{A \rightarrow B}(x, u) = \begin{cases} 1 & \text{gdy } \mu_A(x) \leq \mu_B(u) \\ 0 & \text{gdy } \mu_A(x) > \mu_B(u) \end{cases}$

Tablica 4.2: Najczęściej spotykane definicje rozmytej implikacji

Podsumowując materiał tej sekcji można sformułować następującą procedurę wnioskowania w oparciu o zbiór N reguł i wektor obserwacji $\bar{\mathbf{x}}^T = (\bar{x}_1, \dots, \bar{x}_n)^T$.

1. Dla każdej obserwacji \bar{x}_i wyznaczyć stopień zgodności $\tau_i^{(k)} = \mu_{A_i^{(k)}}(\bar{x}_i)$ tej obserwacji ze zbiorem rozmytym występującym w i -tej elementarnej przesłance k -tej reguły.
2. (*Agregacja*) Dla każdej reguły wyznaczyć stopień zgodności obserwacji $\bar{\mathbf{x}}^T$ z przesłanką k -tej reguły, $\tau^{(k)} = \tau_1^{(k)} * \dots * \tau_n^{(k)}$.
3. (*Aktywacja*) Dla każdej reguły wyznaczyć, w oparciu o wybraną definicję implikacji, postać zbioru rozmytego opisującego k -tą konkluzję.
4. (*Akumulacja*) Zastosować wybraną S -normę (najczęściej \max) do utworzenia wynikowej funkcji przynależności opisującej rozmytą konkluzję.

Odnosimy, że wynikowa funkcja przynależności, o której mowa w punkcie (4) może być uzyskiwana na dwa sposoby [13]:

- Złożenie FITA (ang. *First Inference Then Aggregate*), gdzie najpierw każda reguła jest aktywowana wejściem rozmytym, a następnie poszczególne wyniki są łączone, celem stworzenia wynikowej funkcji przynależności opisującej końcową konkluzję.
- Złożenie FATI (ang. *First Aggregate Then Inference*), gdzie najpierw łączy się wszystkie reguły bazy wiedzy, a następnie aktywuje je wejściem rozmytym.

W [13] wskazano, że wyniki uzyskiwane metodą FITA są bardziej dokładne, a ponadto ta metoda akumulacji jest prostsza w implementacji. Operatorem umożliwiającym akumulację może być zarówno S -norma jak i uogólniony operator uśredniania:

$$\bigcup_{i=1}^N (\alpha, \xi_i) = \left(\frac{1}{N} \sum_{i=1}^N \xi_i^\alpha \right)^{\frac{1}{\alpha}} = \begin{cases} \min_{i=1, \dots, N} \xi_i & \text{gdy } \alpha \rightarrow -\infty \\ \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{\xi_i} \right)^{-1} & \text{gdy } \alpha \rightarrow -1 \\ \left(\prod_{i=1}^N \xi_i \right)^{1/N} & \text{gdy } \alpha \rightarrow 0 \\ \frac{1}{N} \sum_{i=1}^N \xi_i & \text{gdy } \alpha = 1 \\ \sqrt{\frac{1}{N} \sum_{i=1}^N \xi_i^2} & \text{gdy } \alpha = 2 \\ \max_{i=1, \dots, N} \xi_i & \text{gdy } \alpha \rightarrow \infty \end{cases} \quad (4.16)$$

Poniżej przedstawiamy konkretyzacje reguły (4.15) dla wybranych definicji implikacji. We wszystkich przykładach będziemy zakładać, że obserwowany wektor wejść ma postać $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2)$ a rozmyty sterownik opisują dwie reguły postaci

$$\begin{aligned} R^{(1)}: & \text{ if } [(x_1 \text{ is } A_1^{(1)}) \text{ and } (x_2 \text{ is } A_2^{(1)})] \text{ then } (u \text{ is } B^{(1)}) \\ R^{(2)}: & \text{ if } [(x_1 \text{ is } A_1^{(2)}) \text{ and } (x_2 \text{ is } A_2^{(2)})] \text{ then } (u \text{ is } B^{(2)}) \end{aligned}$$

W tym przypadku reguła (4.15) przyjmuje postać

$$\mu_{B^{(k)}}(u) = \sup_{X_1, x_2 \in X_1 \times X_2} [\mu_{A_1' \times A_2'}(x_1, x_2) * \mu_{R^{(k)}}(x_1, x_2, u)] \quad (4.17)$$

a funkcję przynależności wynikowego zbioru \tilde{B} wyznaczymy z równania

$$\mu_{\tilde{B}}(u) = \mu_{B^{(1)}}(u) \diamond \mu_{B^{(2)}}(u) \quad (4.18)$$

Przykład 4.1.1. Niech $*$ = min, \diamond = max i niech do zdefiniowania rozmytej implikacji posłuży reguła Mamdaniego. W takim przypadku reguła (4.6) przybiera postać (agregacja)

$$\mu_{A^{(k)}}(\bar{x}_1, \bar{x}_2) = \mu_{A_1^{(k)} \times A_2^{(k)}}(\bar{x}_1, \bar{x}_2) = \min[\mu_{A_1^{(k)}}(\bar{x}_1), \mu_{A_2^{(k)}}(\bar{x}_2)] \quad (4.19)$$

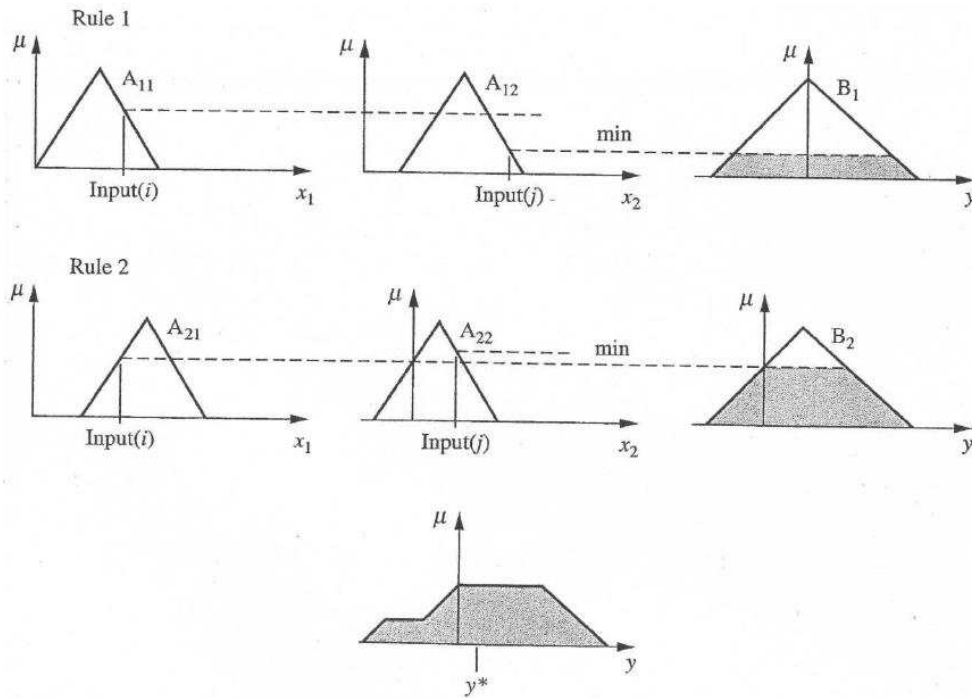
Stosując regułę Mamdaniego otrzymujemy (aktywacja)

$$\begin{aligned} \mu_{\tilde{B}^{(k)}}(u) &= \min \left\{ \min[\mu_{A_1^{(k)}}(\bar{x}_1), \mu_{A_2^{(k)}}(\bar{x}_2)], \mu_{B^{(k)}}(u) \right\} \\ &= \min[\mu_{A_1^{(k)}}(\bar{x}_1), \mu_{A_2^{(k)}}(\bar{x}_2), \mu_{B^{(k)}}(u)] \end{aligned} \quad (4.20)$$

i ostatecznie (akumulacja)

$$\mu_{\tilde{B}}(u) = \max_{k=1,2} \left\{ \min[\mu_{A_1^{(k)}}(\bar{x}_1), \mu_{A_2^{(k)}}(\bar{x}_2), \mu_{B^{(k)}}(u)] \right\} \quad (4.21)$$

Graficzną ilustrację tego procesu przedstawiono na rysunku 4.4 □



Rysunek 4.4: Graficzna reprezentacja procesu wnioskowania w oparciu o dwie reguły, których przesłanki złożone są z dwóch warunków elementarnych. Pierwsze dwie linie reprezentują agregację stopni spełnialności przesłanek oraz aktywację poszczególnych reguł, a w trzeciej linii przedstawiono akumulację uzyskanych wyników.

Przykład 4.1.2. Załóżmy teraz, że implikację definiujemy za pomocą iloczynu (por. regułę 2 w tablicy 4.2), a definicje T-normy i S-normy nie ulegają zmianie. Stosując rozumowanie z poprzedniego przykładu równanie (4.20) przyjmuje teraz postać

$$\mu_{\tilde{B}^{(k)}}(u) = \mu_{B^{(k)}}(u) \cdot \min[\mu_{A_1^{(k)}}(\bar{x}_1), \mu_{A_2^{(k)}}(\bar{x}_2)] \quad (4.22)$$

Wyznaczenie kształtu ostatecznego wyjściowego zbioru rozmytego odbywa się jak poprzednio. □

Przykład 4.1.3. Przyjmijmy, że zarówno T-norma jak i implikacja zdefiniowane są za pomocą iloczynu (definicja S-normy nie ulega zmianie). Stosując rozumowanie z poprzednich przykładów znajdujemy najpierw konkretyzację reguły (4.6):

$$\mu_{A^{(k)}}(\bar{x}_1, \bar{x}_2) = \mu_{A_1^{(k)} \times A_2^{(k)}}(\bar{x}_1, \bar{x}_2) = \mu_{A_1^{(k)}}(\bar{x}_1) \cdot \mu_{A_2^{(k)}}(\bar{x}_2) \quad (4.23)$$

Odnotujmy, że taki sposób agregacji jest najczęściej stosowany w praktyce. Równanie 2 z tablicy 4.2 przyjmuje teraz postać

$$\mu_{\tilde{B}^{(k)}}(u) = \mu_{B^{(k)}}(u) \cdot \mu_{A_1^{(k)}}(\bar{x}_1) \cdot \mu_{A_2^{(k)}}(\bar{x}_2) \quad (4.24)$$

Wyznaczenie kształtu ostatecznego wyjściowego zbioru rozmytego odbywa się jak w poprzednich przykładach. \square

Przykład 4.1.4. Dotychczas rozważaliśmy dokładne pomiary \bar{x}_1, \bar{x}_2 . W praktyce dopuszczalnie jest ich rozmywanie. Załóżmy więc, że obie obserwacje zadane są w postaci zbiorów rozmytych A_1, A_2 i powróćmy do sytuacji rozpatrywanej w przykładzie 4.1.1.

Tym razem kwantyfikacja stopnia prawdziwości faktu x_i **is** $A_i^{(k)}$ wyznaczana jest z równania

$$\tau_{A_i^{(k)}} = \sup_{x_i \in X_i} \min[\mu_{A_i^{(k)}}(x_i), \mu_{A_i}(x_i)] \quad (4.25)$$

a stopień prawdziwości przesłanki jest równy

$$\tau_{A^{(k)}} = \sup_{x_1 \in X_1} [\mu_{A_1^{(k)}}(x_1) \wedge \mu_{A_1}(x_1)] \wedge \sup_{x_2 \in X_2} [\mu_{A_2^{(k)}}(x_2) \wedge \mu_{A_2}(x_2)] \quad (4.26)$$

gdzie $a \wedge b = \min(a, b)$. Ostateczny kształt funkcji przynależności wynikowego zbioru rozmytego \tilde{B} wyznaczamy z równania

$$\mu_{\tilde{B}}(u) = \max_{k=1,2} \left\{ \tau_{A^{(k)}} \wedge \mu_{B^{(k)}}(u) \right\} \quad (4.27)$$

Definiując – jak w poprzednim przykładzie – T-normę za pomocą iloczynu, powyższe równanie przyjmuje postać

$$\mu_{\tilde{B}}(u) = \max_{k=1,2} \left\{ \sup_{x_1} [\mu_{A_1^{(k)}}(x_1) \mu_{A_1}(x_1)] \cdot \sup_{x_2} [\mu_{A_2^{(k)}}(x_2) \mu_{A_2}(x_2)] \cdot \mu_{B^{(k)}}(u) \right\} \quad (4.28)$$

\square

4.1.4 Blok wyostrzania

Pod pojęciem wyostrzania rozumiemy zastąpienie zbioru rozmytego o funkcji przynależności $\mu_{\tilde{B}}$ pojedynczą wartością $\bar{u} \in Y$, która będzie wyznaczonym sterowaniem. Istnieje wiele metod wyostrzania. Poniżej omawiamy najbardziej typowe. Inne metody omawiane są np. w pracy [64].

4.1.4.1 Sterownik Mamdaniego

Przypominamy, że w przypadku sterownika Mamdaniego konkluzja ma postać stwierdzenie "u **is** \tilde{B} ", a zadanie polega na znalezieniu pojedynczej wartości \bar{u} , która należy przyjąć za wartość sterowania.

Najczęściej stosowaną metodą wyostrzania jest metoda środka ciężkości (*COG* – *center of gravity*), w której za wartość \bar{u} przyjmuje się odciętą współrzędną środka ciężkości powierzchni zawartej pod wykresem funkcji przynależności $\mu_{\tilde{B}}$:

$$\bar{u} = \frac{\int \mu_{\tilde{B}}(u) \cdot u du}{\int \mu_{\tilde{B}}(u) du} \quad (4.29)$$

W przypadku dyskretnej funkcji przynależności powyższe równanie przyjmuje postać $\mu_{\tilde{B}}$:

$$\bar{u} = \frac{\sum_i \mu_{\tilde{B}}(u_i) \cdot u_i}{\sum_i \mu_{\tilde{B}}(u_i)} \quad (4.30)$$

Zaletą tej metody jest uwzględnienie wpływu wszystkich funkcji przynależności występujących w konkluzjach zaktywizowanych reguł. Jej wadą jest stosunkowo duży nakład obliczeniowy związany z całkowaniem powierzchni o nieregularnym kształcie. W przypadku trapezoidalnych lub trójkątnych funkcji przynależności można stosować proste metody całkowania numerycznego do wyznaczenia wartości \bar{u} .

Stosunkowo proste – z numerycznego punktu widzenia są metody polegające na wyborze takiej wartości \bar{u} , dla której funkcja $\mu_{\tilde{B}}$ osiąga swoje maksimum. Wyróżnia się tu trzy warianty:

- (a) metodę pierwszego maksimum (*FOM* – *first of maxima* lub *SOM* – *smallest of maxima*), w której za \bar{u} przyjmuje się najmniejszą wartość u odpowiadającą maksymalnemu stopniowi przynależności $\mu_{\tilde{B}}(u)$;
- (b) metodę ostatniego maksimum (*LOM* – *last*, lub *largest of maxima*), w której za \bar{u} przyjmuje się największą wartość u odpowiadającą maksymalnemu stopniowi przynależności $\mu_{\tilde{B}}(u)$;
- (c) metodę średniego maksimum (*MOM* – *middle of maxima* wybierającą centralny punkt zbioru tych wartości u , dla których funkcja przynależności $\mu_{\tilde{B}}(u)$ przyjmuje wartość maksymalną.

Wadą tych metod jest uwzględnianie wyłącznie konkluzji najbardziej zaktywizowanej reguły.

Graficzne porównanie wymienionych tu metod przedstawiono na rysunku 4.5. Inne metody wyostrzania, to:

1. Metoda środka sum (*CSD* – *center of sums defuzzification*). Wartość \bar{u} obliczana jest tu następująco

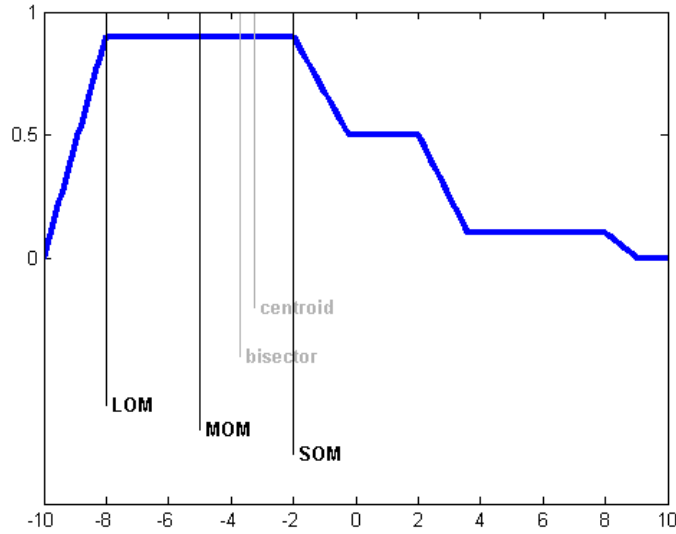
$$\bar{u} = \frac{\int u \sum_{k=1}^N \mu_{\tilde{B}^{(k)}}(u) du}{\int \sum_{k=1}^N \mu_{\tilde{B}^{(k)}}(u) du} \quad (4.31)$$

2. Metoda bisekcji (*BOA* – *bisector of area*), w której za \bar{u} wybiera się taką wartość odciętej, która dzieli powierzchnię pod krzywą $\mu_{\tilde{B}}$ na połowy, tzn.

$$\int_{u_{min}}^{\bar{u}} \mu_{\tilde{B}}(u) du = \int_{\bar{u}}^{u_{max}} \mu_{\tilde{B}}(u) du \quad (4.32)$$

3. Metoda *CAD* (*center average defuzzification*). Wartość \bar{u} wyznacza się ze wzoru

$$\bar{u} = \frac{\sum_{k=1}^N \mu_{\tilde{B}^{(k)}}(\bar{u}^k) \cdot \bar{u}^k}{\sum_{k=1}^N \mu_{\tilde{B}^{(k)}}(\bar{u}^k)} \quad (4.33)$$



Rysunek 4.5: Porównanie metod wyostrzania.

gdzie $\bar{u}^k = \operatorname{argmax}_u \mu_{\tilde{B}^{(k)}}(u)$. Punkt \bar{u}^k nazywany jest środkiem zbioru rozmytego $\tilde{B}^{(k)}$. Nie zależy on ani od kształtu ani od nośnika funkcji przynależności $\mu_{\tilde{B}^{(k)}}$.

4. Metoda *MFP* (*maksimum funkcji przynależności*). Jeżeli $\mu_{\tilde{B}}$ jest funkcją jedno-modalną, to wartość \bar{u} wyznacza się ze wzoru

$$\bar{u} = \sup_u \mu_{\tilde{B}}(u) \quad (4.34)$$

4.1.4.2 Sterownik Takagi-Sugeno

W przypadku tego sterownika konkluzja k -tej reguły ma postać $u = f^{(k)}(x_1, \dots, x_n)$. W celu wyznaczenia wartości sterowania \bar{u} postępujemy następująco:

1. Dla każdej reguły wyznaczamy stopień zgodności $w^{(k)}$ obserwacji $(\bar{x}_1, \dots, \bar{x}_n)$ z przesłankami k -tej reguły, $w^{(k)} = \mu_{A_1^{(k)}}(\bar{x}_1) * \dots * \mu_{A_n^{(k)}}(\bar{x}_n)$, gdzie $*$ jest symbolem T-normy (jest to zazwyczaj min lub iloczyn). Obliczamy także wartość $\bar{u}^{(k)} = f^{(k)}(x_1, \dots, x_n)$.
2. Poszukiwaną wartość \bar{u} wyznaczamy z równania

$$\bar{u} = \frac{\sum_{k=1}^N w^{(k)} \bar{u}^{(k)}}{\sum_{k=1}^N w^{(k)}} \quad (4.35)$$

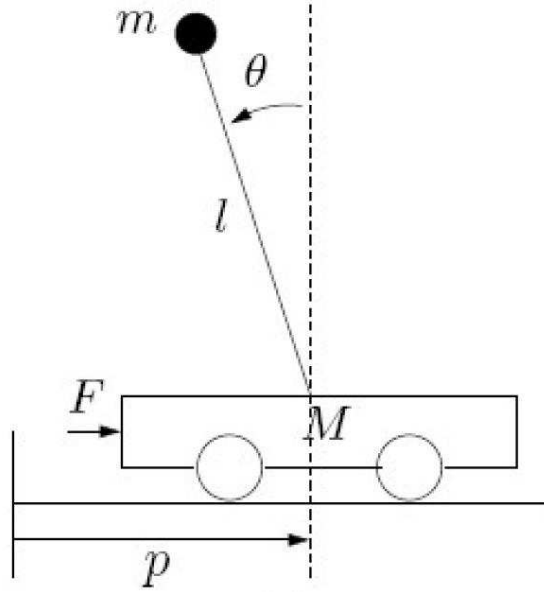
4.1.5 Przykłady

4.1.5.1 Odwrócone wahadło

Układ składa się z wózka i zamocowanego na elastycznym przegubie pionowego wahadła o masie m i długości l – por. rysunek 4.6.

Zachowanie systemu opisuje równanie różniczkowe

$$-ml^2 \frac{d^2\theta}{dt^2} + mlg \cdot \sin \theta = \tau = u(t)$$



Rysunek 4.6: Problem sterowania odwróconym wahadłem.

gdzie $\tau = u(t)$ jest sterowaniem (moment obrotowy), które należy przyłożyć w kierunku przeciwnym do wychylenia θ aby utrzymać wahadło w pozycji pionowej.

Oznaczając $x_1 = \theta$, $x_2 = d\theta/dt$ sprowadzamy powyższe równanie do układu równań opisującego zachowanie badanego układu w przestrzeni stanów

$$\begin{aligned} dx_1/dt &= x_2 \\ dx_2/dt &= (g/l) \sin x_1 - (1/ml^2)u(t) \end{aligned}$$

Dla małych wartości θ mamy $\sin \theta \approx \theta$, co pozwala zlinearyzować powyższy układ

$$\begin{aligned} dx_1/dt &= x_2 \\ dx_2/dt &= (g/l)x_1 - (1/ml^2)u(t) \end{aligned}$$

Jeżeli x_1 jest mierzone w stopniach, a x_2 w stopniach na sekundę (*sns*), to przyjmując $l = g$, $m = 180/(\pi \cdot g^2)$ uzyskujemy następujący różnicowy odpowiednik powyższego układu

$$\begin{aligned} x_1(t+1) &= x_1(t) + x_2(t) \\ x_2(t+1) &= x_1(t) + x_2(t) - u(t) \end{aligned}$$

Przyjmijmy, że $-2^\circ \leq x_1 \leq 2^\circ$, $-5sns \leq x_2 \leq 5sns$. Konstrukcja sterownika składa się z następujących kroków.

1. Dziedzinę zmiennej wejściowej x_1 pokrywamy trzema zbiorami rozmytymi: $\mu_{NEG} = \Delta(-\infty, -2, 0)$, $\mu_{Zero} = \Delta(-2, 0, 2)$, $\mu_{POS} = \Delta(0, 2, \infty)$. Podobnie, trzema zbiorami rozmytymi, lecz o innych nośnikach, pokrywamy dziedzinę zmiennej x_2 : $\mu_{NEG} = \Delta(-\infty, -5, 0)$, $\mu_{Zero} = \Delta(-5, 0, 5)$, $\mu_{POS} = \Delta(0, 5, \infty)$.
2. Pokrywamy dziedzinę zmiennej sterującej u pięcioma funkcjami przynależności: $\mu_{NB} = \Delta(-\infty, -24, -16)$, $\mu_N = \Delta(-16, -8, 0)$, $\mu_Z = \Delta(-8, 0, 8)$, $\mu_P = \Delta(0, 8, 16)$, $\mu_{PB} = \Delta(16, 24, \infty)$.

	NEG	Zero	POS
NEG	NB	N	Z
Zero	N	Z	P
POS	Z	P	PB

Tablica 4.3: Baza rozmytych reguł

3. Tworzymy reguły sterujące. Wygodną formą ich zapisu jest tzw. rozmyta pamięć asocjacyjna (*FAM*) – tablica 4.3: Wiersze tablicy odpowiadają rozmytym wartościom zmiennej x_1 , a kolumny – rozmytym wartościom zmiennej x_2 . Element znajdujący się na przecięciu wiersza i kolumny odpowiada konkluzji reguły, której przesłankami są odpowiednie wartości lingwistyczne zmiennych wejściowych. Przykładowo, przecięciu pierwszego wiersza i trzeciej kolumny odpowiada reguła **if x_1 is NEG and x_2 is Pos then u is Z**.

4. Właściwe sterowanie. Przypuśćmy, że w chwili $t = 0$ wahadło znajduje się w punkcie $x_1 = 1^\circ$, $x_2 = -4stn$. Punkt ten uaktywnia następujące reguły

$$\begin{aligned}
 \text{if } x_1 \text{ is Pos and } x_2 \text{ is Zero then } u \text{ is P} & \quad \min(0.5, 0.2) = 0.2 \text{ (P)} \\
 \text{if } x_1 \text{ is Pos and } x_2 \text{ is Neg then } u \text{ is Z} & \quad \min(0.5, 0.8) = 0.5 \text{ (Z)} \\
 \text{if } x_1 \text{ is Zero and } x_2 \text{ is Zero then } u \text{ is Z} & \quad \min(0.5, 0.2) = 0.2 \text{ (Z)} \\
 \text{if } x_1 \text{ is Zero and } x_2 \text{ is Neg then } u \text{ is N} & \quad \min(0.5, 0.8) = 0.5 \text{ (N)}
 \end{aligned}$$

Ostatecznie konkluzji N przypisujemy stopień 0.5, konkluzji Z stopień $\max(0.5, 0.2) = 0.5$ i konkluzji P stopień 0.2. Wyodrębiając wynikowy zbiór rozmyty stwierdzamy, że $u = -2$. Możemy więc wyznaczyć kolejne położenie wahadła w przestrzeni stanów

$$\begin{aligned}
 x_1(1) &= x_1(0) + x_2(0) = 1 - 4 = -3 \\
 x_2(1) &= x_1(0) + x_2(0) - u(0) = 1 - 4 - (-2) = -1
 \end{aligned}$$

Powtarzamy teraz procedurę wyznaczania kolejnego sterowania.

4.1.5.2 Automatyczny pilot

Rozważmy teraz symulację procesu lądowania samolotu. Interesują nas dwie zmienne: wysokość samolotu h i jego prędkość v . Równania opisujące dynamikę procesu wyprowadza się stosując następujące rozumowanie. Ciało o masie m poruszające się z prędkością v posiada pęd $p = mv$. Przy braku działania zewnętrznych sił, ciało to będzie poruszać się nadal ze stałą prędkością v . Dodatkowe oddziaływanie siłą f w przedziale czasu Δt skutkuje zmianą prędkości $\Delta v = f\Delta t/m$. Przyjmując $\Delta t = 1.0$ oraz $m = 1.0$ (i dobierając odpowiednio jednostki miary), otrzymujemy $\Delta v = f$, co oznacza, że zmiana prędkości jest proporcjonalna do użytej siły. Otrzymujemy więc następujący układ równań:

$$\begin{aligned}
 v(t+1) &= v(t) + f(t) \\
 h(t+1) &= h(t) + v(t)
 \end{aligned}$$

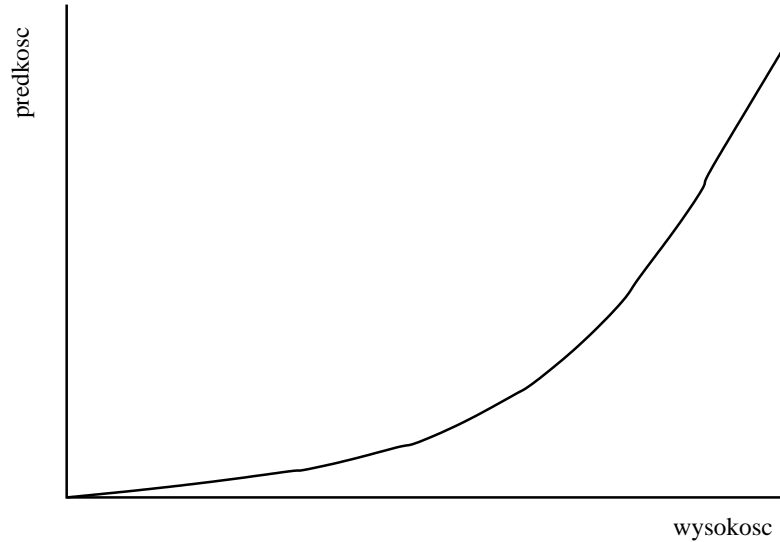
Przyjmijmy, że wysokość samolotu zmienia się w przedziale $[0, 1000]$. Pokrywamy dziedzinę zmiennej h czterema zbiorami rozmytymi: **Near Zero** o funkcji przynależności $\Delta(0, 0, 500)$, **Small** o funkcji przynależności $\Delta(-200, 300, 800) \cdot \mathbb{1}(x)$, gdzie $\mathbb{1}(x)$ jest funkcją przyjmującą wartość 0 dla $x \leq 0$ i 1 gdy $x > 0$, **Medium** o funkcji przynależności $\Delta(300, 800, 1300) \cdot (1 - \mathbb{1}(x - 1000))$ oraz **Large** o trójkątnej funkcji $\Delta(500, 1000, 1000)$.

Założmy dalej, że prędkość samolotu zmienia się w przedziale $[-30, 30]$ pokrytym pięcioma zbiorami rozmytymi, których definicje przedstawiono w Tablicy 4.4. Podobnie zmienia się wartości siły f (sterowania), tzn. $f \in [-30, 30]$. Dlatego dziedzinę tej funkcji pokryto również pięcioma zbiorami rozmytymi o identycznych, jak poprzednio, funkcjach przynależności.

etykieta	definicja
Down Large (DL)	$\Delta(-\infty, -20, -10)$
Down Small (DS)	$\Delta(-20, -10, 0)$
Zero (Z)	$\Delta(-10, 0, 10)$
Up Small (US)	$\Delta(0, 10, 20)$
Up Large (UL)	$\Delta(10, 20, \infty)$

Tablica 4.4: Pokrycie zbiorami rozmytymi dziedziny zmiennej v

Należy tak sterować samolotem, aby jego trajektoria ruchu wyglądała jak na poniższym rysunku.



Rysunek 4.7: Pożądana trajektoria lotu w przestrzeni fazowej (h, v)

Proponowana rozmyta pamięć asocjacyjna (FAM), reprezentująca reguły sterujące zapewniające uzyskanie powyższej trajektorii, ma następującą postać:

wysokość h	prędkość v				
	DL	DS	Zero	US	UL
L	Z	DS	DL	DL	DL
M	US	Z	DS	DL	DL
S	UL	US	Z	DS	DL
NZ	UL	UL	Z	DS	DS

Tablica 4.5: Rozmyta pamięć asocjacyjna

Dla przykładu założmy, że w chwili $t = 0$ mamy $h = 1000$, $v = -20$. Wyznaczając wartość siły f za pomocą metody z poprzedniego punktu i wyodrządzając wynikowy zbiór

rozmyty metodą środka ciężkości znajdujemy, że $f = 5.8$.

Problem: Przeprowadzić symulację lądowania samolotu rozpoczynając od podanych wyżej wartości wysokości i prędkości. Wykonać stosowny wykres. Porównać wyniki z tymi, uzyskanymi przez zastosowanie metody średniego maksimum (*MOM*).

4.2 Podstawowy algorytm generowania reguł rozmytych

Pochodzi on z pracy [79] (jego opis znajduje się także w książce [72]). Jej autorzy zauważyli, że sterowanie rozmyte jest znakomitą metodą pozwalającą korzystać z opisowej wiedzy o problemie (model lingwistyczny). Z drugiej strony podejścia bazujące na zastosowaniu sieci neuronowych w sterowaniu odwołują się do wiedzy reprezentowanej w postaci par (*wejście*, *wyjście*), gdzie *wejście* to zbiór obserwowanych wartości wejściowych, a *wyjście* – odpowiadające im wartości zmiennych wyjściowych. A więc sterowniki rozmyte korzystają z reguł lingwistycznych, a sterowniki neuronowe – z danych numerycznych.

Przedstawiane dalej podejście stawia sobie za cel połączenie obu paradygmatów, tzn. dysponując danymi numerycznymi konwertuje się je do postaci reguł rozmytych, które mogą być z kolei integrowane ze skonstruowanym wcześniej modelem lingwistycznym w celu utworzenia wspólnej bazy wiedzy pozwalającej generować wartości sterowań.

W dalszym ciągu będziemy zakładać, jak w obu cytowanych pracach, że chcemy utworzyć bazę reguł dla układu rozmytego o dwóch wejściach i jednym wyjściu. Dane uczące zadane są w postaci m trójek

$$\begin{aligned} & (x_1^{(1)}, x_2^{(1)}, u^{(1)}) \\ & (x_1^{(2)}, x_2^{(2)}, u^{(2)}) \\ & \dots \\ & (x_1^{(m)}, x_2^{(m)}, u^{(m)}) \end{aligned} \tag{4.36}$$

gdzie $x_1^{(i)}, x_2^{(i)}$ to wartości sygnałów wejściowych zarejestrowanych w i -tym przykładzie, a $u^{(i)}$ to odpowiadająca im wartość sygnału wyjściowego.

Poniżej omawiamy główne kroki algorytmu.

4.2.1 Podział wejściowej i wyjściowej przestrzeni na obszary

Dla każdego sygnału wejściowego i wyjściowego wyznaczamy jego dziedzinę $[x_{i,min}, x_{i,max}], [u_{min}, u_{max}], i = 1, 2$.

Każdy z wyznaczonych w ten sposób przedziałów dzielony jest na $2N + 1$ obszarów, oznaczających nośniki N zbiorów rozmytych. Liczba N może być różna dla każdego z sygnałów, a nośniki mogą być różnej długości. Z każdym zbiorem rozmytym kojarzymy etykietę oddającą intencję jego wprowadzenia. Przykładowe określenia to *NB* (*Negative Big*), *NM* (*Negative Medium*), *NS* (*Negative Small*), *Z*, *PS* (*Positive Small*), *PM* (*Positive Medium*), *PB* (*Positive Big*).

Kształt funkcji przynależności może być dowolny, jednak najchętniej stosowanymi są trójkątne funkcje przynależności opisane równaniem (2.14).

4.2.2 Tworzenie reguł rozmytych na podstawie przykładów

Przypuśćmy, że zakres zmienności sygnału x_1 pokryto zbiorami rozmytymi $A^j, j = 1, \dots, j_{max}$, sygnału x_2 – zbiorami $B^k, k = 1, \dots, k_{max}$, a dziedzinę sygnału wyjściowego pokryto zbiorami $C^l, l = 1, \dots, l_{max}$.

Dla każdego przykładu $(x_1^{(i)}, x_2^{(i)}, u^{(i)})$ wyznaczmy teraz stopnie przynależności do odpowiedniego obszaru: $\mu_{A^j}(x_1^{(i)}), \mu_{B^k}(x_2^{(i)}), \mu_{C^l}(u^{(i)})$. Ostatecznie konkretną wartość sygnału zaliczamy do obszaru o najwyższym stopniu przynależności. Jeżeli np. obserwacja $x_1^{(i)}$ należy do zbioru PB w stopniu 0.8, do zbioru PS w stopniu 0.3, a jej przynależność do pozostałych zbiorów rozmytych jest równa zeru, to zaobserwowaną wartość zaliczamy do zbioru PB .

Ogólnie, jeżeli obserwacja $x_1^{(i)}$ należy w maksymalnym stopniu do zbioru A^{j*} , obserwacja $x_2^{(i)}$ należy w maksymalnym stopniu do zbioru B^{k*} , a obserwacja $u^{(i)}$ należy w maksymalnym stopniu do zbioru C^{l*} , to przykładowi $(x_1^{(i)}, x_2^{(i)}, u^{(i)})$ odpowiada reguła

$$\text{if } (x_1 \text{ is } A^{j*}) \text{ and } (x_2 \text{ is } B^{k*}) \text{ then } (u \text{ is } C^{l*}) \quad (4.37)$$

4.2.3 Przyporządkowanie stopni prawdziwości regułom

W przypadku, gdy reguły o identycznych przesłankach mają różne konkluzje – mamy do czynienia z sytuacją konfliktową, a reguły o takiej własności tworzą zbiór sprzecznych reguł.

Najprostsza metoda rozwiązywania konfliktów polega na przypisaniu każdej regule (ze zbioru sprzecznych reguł) tzw. stopnia prawdziwości i wybrania tylko tej reguły, która ma najwyższy taki stopień. W ten sposób nie tylko rozwiązujemy konflikt, ale i redukujemy liczbę reguł. Dla reguły R postaci (4.37) stopień prawdziwości τ_R obliczamy następująco:

$$\tau_R = \mu_{A^{j*}}(x_1) \cdot \mu_{B^{k*}}(x_2) \cdot \mu_{C^{l*}}(u) \quad (4.38)$$

Problem: Zbadać, co będzie gdy występujący w powyższym równaniu iloczyn zastąpi się operatorem \wedge (min).

W praktyce posiadamy pewną *aprioryczną* wiedzę o rozwiązywanym problemie. Np. parze danych $(x_1^{(i)}, x_2^{(i)}, u^{(i)})$ można przypisać subiektywny stopień $\beta^{(i)}$ charakteryzujący jej wiarygodność, typowość czy też użyteczność w procesie formowania bazy wiedzy. W takim przypadku stopień prawdziwości przyznany tworzonej regule będzie miał postać

$$\tau_R = \mu_{A^{j*}}(x_1^{(i)}) \cdot \mu_{B^{k*}}(x_2^{(i)}) \cdot \mu_{C^{l*}}(u^{(i)}) \beta^{(i)} \quad (4.39)$$

4.2.4 Utworzenie bazy rozmytych reguł

Bazę reguł (wiedzy) stanowi tablica tworzona w sposób następujący: tworzymy tablicę, której boki odpowiadają wartościom rozmytym sygnałów wejściowych. Dla reguły postaci (4.37) rozmytą wartość sygnału wyjściowego umieszczamy na przecięciu wiersza odpowiadającego rozmytej wartości A^{i*} sygnału x_1 i kolumny odpowiadającej rozmytej wartości B^{j*} sygnału x_2 . W ten sposób otrzymujemy tablicę 4.6. Reprezentuje ona regułę **if** (x_1 is NS) **and** (x_2 is PS) **then** (u is PB).

4.2.5 Wyostrzenie

W oparciu o uzyskaną tablicę reguł poszukujemy teraz odwzorowania $g: (x_1, x_2) \rightarrow \bar{u}$, gdzie \bar{u} jest wartością wyjściową zwracaną przez sterownik.

Niech

$$\alpha^{(k)} = \mu_{A^{(k)}}(x_1) \cdot \mu_{B^{(k)}}(x_2) \quad (4.40)$$

oznacza stopień aktywności k -tej reguły.

<i>PB</i>					
<i>PM</i>					
<i>PS</i>		<i>PB</i>			
<i>Z</i>					
<i>NS</i>					
<i>NM</i>					
<i>NB</i>					
	<i>NM</i>	<i>NS</i>	<i>Z</i>	<i>PM</i>	<i>PB</i>

Tablica 4.6: Postać bazy rozmytych reguł

Poszukiwana wartość \bar{u} wyznaczamy korzystając z jednej z metod wyostrzania, np. stosując metodę *center average defuzzification* otrzymujemy

$$\bar{u} = \frac{\sum_{k=1}^N \alpha^{(k)} \bar{u}^{(k)}}{\sum_{k=1}^N \alpha^{(k)}} \quad (4.41)$$

gdzie $\bar{u}^{(k)}$ jest punktem, w którym funkcja przynależności występująca w konkluzji k -tej reguły osiąga swoje maksimum.

W przypadku, gdy przesłanki reguł zawierają n przesłanek elementarnych, w których występują zbiory $A_1^{(k)}, \dots, A_n^{(k)}$, stopień aktywności k -tej reguły jest równy

$$\alpha^{(k)} = \prod_{i=1}^n \mu_{A_i^{(k)}}(\bar{x}_i) \quad (4.42)$$

Niech \mathfrak{F} będzie zbiorem funkcji $f: [a_1, b_1] \times \dots \times [a_n, b_n] \rightarrow \mathbb{R}$, gdzie $|a_i b_i| < \infty$, $i = 1, \dots, n$. Wang i Mendel pokazali w [79], że odwzorowanie (4.41) jest uniwersalnym aproksymatorem, tzn. może aproksymować z dowolną dokładnością każdą funkcję $f \in \mathfrak{F}$.

4.3 Alternatywne metody generowania reguł

4.3.1 Generowanie uproszczonych reguł

Pomysł ten pochodzi z pracy [59], w której autorzy zajmują się regułami postaci

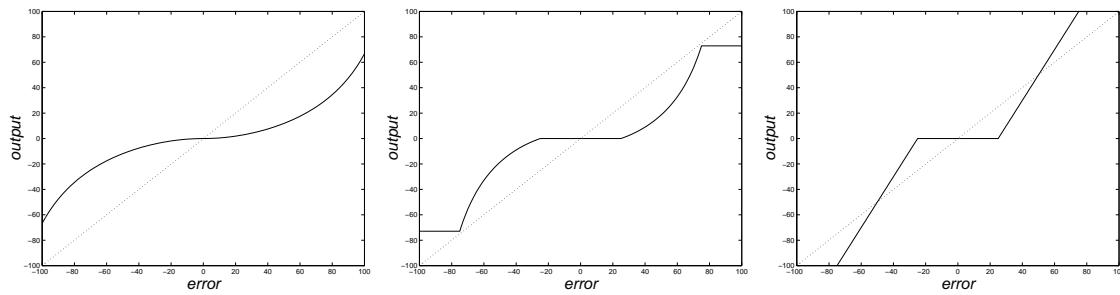
$$R^{(k)}: \text{if } [(x_1 \text{ is } A_1^{(k)}) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^{(k)})] \text{ then } (u \text{ is } b^{(k)}) \quad (4.43)$$

gdzie $b^{(k)} \in \mathbb{R}$ jest pewną liczbą. W dalszym ciągu reguły postaci (4.43) będziemy nazywać *prostymi* regułami rozmytymi w odróżnieniu od reguł postaci (4.2). Zaletą omawianych tu reguł jest to, że w procesie wnioskowania pomija się etap wyostrzania. W marę potrzeb można je także konwertować do typowego modelu lingwistycznego, co pokazujemy w sekcji 4.3.1.3. Ponadto – podobnie jak w przypadku ogólnym – reguły te posiadają dobre własności aproksymacyjne.

Różnicę między prostymi a "normalnymi" regułami sterowania przedstawiono na rysunkach 4.8. Zakładamy, że zachowanie serownika proporcjonalnego opisują trzy reguły:

1. **if error is *NEG* then output is *NEG***
2. **if error is *Zero* then output is *Zero***
3. **if error is *POS* then output is *POS***

Przyjęto też, że zbiory rozmyte *NEG*, *Zero* i *POS* opisane są za pomocą



Rysunek 4.8: Zależność sygnału wyjściowego od wejściowego w proporcjonalnym sterowniku rozmytym. Na lewym i środkowym rysunku przesłanki i konkluzje są zbiorami rozmytymi o: (a) trójkątnych, (b) trapezoidalnych funkcjach przynależności. Liniaą przerywaną zaznaczono rezultat uzyskany za pomocą prostych reguł sterowania z przesłankami zdefiniowanymi przez trójkątne funkcje przynależności. Prawy rysunek przedstawia wynik generowany przez proste reguły z przesłankami zdefiniowanymi przez trapezoidalne funkcje przynależności.

- trójkątnych funkcji przynależności, odpowiednio: $(-\infty, -100, 0)$, $(-100, 0, 100)$, $(0, 100, \infty)$,
- trapezoidalnych funkcji przynależności, odpowiednio: $(-\infty, -100, -75, -25)$, $(-75, -25, 25, 75)$, $(25, 75, 100, \infty)$.

W przypadku gdy zarówno sygnał wejściowy jak i wyjściowy charakteryzowany jest przez powyższe funkcje przynależności, kształt wyostrzonego sygnału wyjściowego ma postać jak na rysunkach 4.8a oraz 4.8b. W obu przypadkach linią przerywaną zaznaczono sygnał wyjściowy uzyskany za pomocą prostych reguł sterowania. W tym ostatnim przypadku wejście etykietowane jest zbiorami z trójkątnymi funkcjami przynależności, a wyjście charakteryzują trzy liczby: -100, 0 oraz 100. Dla porównania, rysunek 4.8c odpowiada sytuacji, gdy przesłanki prostych reguł opisane są trapezoidalnymi funkcjami przynależności.

W przypadku sterownika Mamdaniego⁵ (pierwszy wiersz na rysunku 4.8) wynikowa relacja między sygnałem wejściowym a wyjściowym pokazana jest w ostatniej kolumnie pierwszego wiersza. Jak widać, nie jest ona liniowa, a co istotniejsze – nie cały zakres sterowania jest tu wykorzystywany.

Ten ostatni problem znika w przypadku zastosowania prostych reguł rozmytych. Tutaj etykietcie *NEG* odpowiada wartość -100, etykietcie *Zero* – wartość 0, a etykietcie *POS* – wartość 100. W takim przypadku otrzymujemy liniową relację wejście-wyjście.

Można sprawdzić, że np. w przypadku trapezoidalnych funkcji przynależności, relacja wejście-wyjście ma także charakter liniowy, ale z *plateau* w okolicy zera. Eksperymenty wskazują, że przy projektowaniu sterownika najlepiej jest stosować się do następujących reguł:

1. używać trójkątnych funkcji przynależności przecinających się w punkcie $\mu = 0.5$
2. modelować spójnik **and** operatorem mnożenia
3. konkluzje reguł powinny być numeryczne, tzn. należy stosować proste reguły sterowania wyznaczania sygnału wyjściowego stosować wariant wyostrażania metodą środków ciężkości opisany równaniem (4.44).

⁵W obliczeniach przyjęto, że T-normę operator mnożenia, że S-normę – operator max, a wyostrażanie przeprowadzono metodą środka ciężkości.

4.3.1.1 Wnioskowanie z użyciem prostych reguł rozmytych

Niech $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)^T$ będzie wektorem obserwacji. Odpowiadającą mu wartość $u(\bar{\mathbf{x}})$ wyznaczamy z równania

$$u(\bar{\mathbf{x}}) = \frac{\sum_{k=1}^N \prod_{j=1}^n \mu_{A_j^{(k)}}(\bar{x}_j) b^{(k)}}{\sum_{k=1}^N \prod_{j=1}^n \mu_{A_j^{(k)}}(\bar{x}_j)} \quad (4.44)$$

Tutaj, jak poprzednio, iloczyn $\prod_{j=1}^n \mu_{A_j^{(k)}}(\bar{x}_j)$ określa stopień zgodności obserwacji $\bar{\mathbf{x}}$ z przesłankami k -tej reguły. Jest to wariant wyostrzania metodą środków ciężkości.

4.3.1.2 Identyfikacja prostych reguł rozmytych

Zakładamy, że dane mają postać identyczną z (4.36). Zakładamy też, że dokonano odpowiedniego pokrycia dziedzin zmiennych X_1, \dots, X_n zbiorami rozmytymi. Jeżeli A_i^k oznacza k_i -ty rozmyty podzbiór dziedziny zmiennej X_i to przez

$$w_{k_1 \dots k_n}(\bar{\mathbf{x}}) = \left(\prod_{j=1}^n \mu_{A_j^{k_j}}(\bar{x}_j) \right)^\alpha \quad (4.45)$$

oznacza stopień zgodności obserwacji $\bar{\mathbf{x}}$ z rozmytym obszarem wyznaczonym przez wybraną rodzinę podzbiorów. Zauważmy, że z równania tego wynika, że jeżeli dziedzinę zmiennej X_i pokryto K_i zbiorami rozmytymi, to dane uczące pozwalają wygenerować co najwyżej $K = K_1 \dots K_n$ reguł rozmytych. Jeżeli istotnie wygenerowano K reguł, mówimy wówczas o *kompletnej* bazie wiedzy.

Występujący w tym równaniu wykładnik $\alpha \in (0, 1)$ steruje stopniem dopasowania wyników reguł do danych trenujących.

Dysponując wagami (4.45), regule o indeksie $k = k_1 \dots k_n$ przypisuje się następująca numeryczną konkluzję:

$$b_{k_1 \dots k_n} = \frac{\sum_{j=1}^m w_{k_1 \dots k_n}(\bar{\mathbf{x}}_j) \cdot u_j}{\sum_{j=1}^m w_{k_1 \dots k_n}(\bar{\mathbf{x}}_j)} \quad (4.46)$$

tzn. $b_{k_1 \dots k_n}$ jest ważoną sumą m obserwacji zmiennej zależnej.

Oczywiście, jeżeli $w_{k_1 \dots k_n} = 0$ oznacza to niemożność wygenerowania reguły dla wybranej kombinacji zbiorów rozmytych etykietujących elementarne przesłanki. Jest to sygnał wskazujący na konieczność redefinicji rozmytych pokryć.

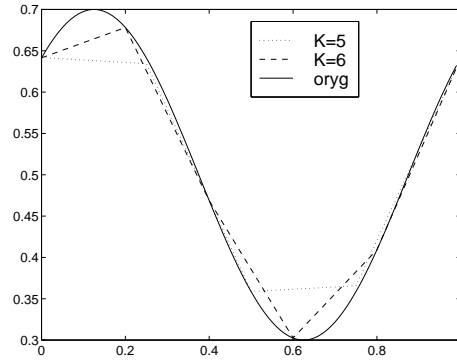
Decydując się na wybór konkretnego pokrycia, jak również na wybór wartości α można kierować się wartościami wskaźnika dopasowania

$$e = \sqrt{\sum_{j=1}^m \left(u(\bar{\mathbf{x}}_j) - u_j \right)^2} \quad (4.47)$$

Przykład 4.3.1. Celem ilustracji omówionej procedury rozważmy problem aproksymacji statycznego nieliniowego systemu opisanego równaniem

$$u = 0.2 \sin(2\pi x + \pi/4) + 0.5, 0 \leq x \leq 1$$

Zbiór trenujący utworzono wybierając 11 punktów o odciętych $x_{i+1} = x_i + 0.1$, $x_0 = 0.0$. Rozważono też dwa alternatywne pokrycia zbioru $[0, 1]$. Pierwsze to pięć trójkątnych funkcji przynależności postaci $\Delta_1(-\infty, 0.0, 0.25)$, $\Delta_2(0.0, 0.25, 0.50)$, $\Delta_3(0.25, 0.50, 0.75)$,



Rysunek 4.9: Porównanie jakości aproksymacji zadanej funkcji (przedstawionej linią ciągłą) dla dwóch różnych pokryć dziedzin $[0, 1]$.

$\Delta_4(0.50, 0.75, 1.0)$, $\Delta_5(0.75, 1.0, \infty)$. Drugie pokrycie obejmuje sześć trójkątnych funkcji przynależności postaci $\Delta_1(-\infty, 0.0, 0.2)$, $\Delta_2(0.0, 0.2, 0.4)$, $\Delta_3(0.2, 0.4, 0.6)$, $\Delta_4(0.4, 0.6, 0.8)$, $\Delta_5(0.6, 0.8, 1.0)$, $\Delta_6(0.8, 1.0, \infty)$.

W pierwszym przypadku jakość rekonstrukcji oryginalnej funkcji mierzona błędem średniokwadratowym jest równa $e = 0.003268$, a w drugim $e = 0.001972$. Zrekonstruowane krzywe przedstawiono na rysunku 4.3.1.2. Jeżeli w danych występują zakłócenia, np. *outliery*, to sensowny wybór wartości α pozwala zredukować ich wpływ na jakość rekonstrukcji. \square

4.3.1.3 Konwersja prostych reguł rozmytych

Przypuśćmy, że dany jest zestaw prostych reguł, a na dziedzinie zmiennej u określono rodzinę rozmytych podzbiorów $\mathbb{B} = \{B_1, \dots, B_N\}$. Niech $B_{j*}^{(k)} \in \mathbb{B}$ będzie zbiorem rozmytym o najwyższym stopniu zgodności z konkluzją $b^{(k)}$ tej reguły. Podobnie, niech $B_{j**}^{(k)} \in \mathbb{B} \setminus \{B_{j*}^{(k)}\}$ będzie zbiorem rozmytym o najwyższym (w zbiorze $\mathbb{B} \setminus \{B_{j*}^{(k)}\}$) stopniu zgodności z konkluzją $b^{(k)}$ tejże reguły. Wówczas regułę (4.43) można zastąpić dwiema regułami lingwistycznymi

$$\begin{aligned} R^{(k*)}: & \text{ if } [(x_1 \text{ is } A_1^{(k)}) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^{(k)})] \text{ then } (u \text{ is } B_{j*}^{(k)}) \\ R^{(k**)}: & \text{ if } [(x_1 \text{ is } A_1^{(k)}) \text{ and } \dots \text{ and } (x_n \text{ is } A_n^{(k)})] \text{ then } (u \text{ is } B_{j**}^{(k)}) \end{aligned} \quad (4.48)$$

W szczególnym przypadku, gdy stosujemy zbiory rozmyte o trójkątnych funkcjach przynależności przecinających się parami na wysokości $\mu = 0.5$ zauważamy co następuje:

- dowolna konkluzja $b^{(k)}$ jest pokrywana przez dwa sąsiadujące ze sobą zbiory $B_{j*}^{(k)}$ oraz $B_{j**}^{(k)}$;
- $\mu_{B_{j*}^{(k)}}(u) + \mu_{B_{j**}^{(k)}}(u) = 1$, $u \in [\hat{B}_{j*}^{(k)}, \hat{B}_{j**}^{(k)}]$, przy czym $\hat{B}_{j*}^{(k)}$ oznacza środek zbioru $B_{j*}^{(k)}$, a ponadto $\hat{B}_{j*}^{(k)} < \hat{B}_{j**}^{(k)}$;
- dowolną wartość $u \in [\hat{B}_{j*}^{(k)}, \hat{B}_{j**}^{(k)}]$ można przedstawić w postaci sumy ważonej $u = \mu_{B_{j*}^{(k)}}(u) \cdot \hat{B}_{j*}^{(k)} + \mu_{B_{j**}^{(k)}}(u) \cdot \hat{B}_{j**}^{(k)}$, tzn. dowolną wartość $u \in [\hat{B}_{j*}^{(k)}, \hat{B}_{j**}^{(k)}]$ można przedstawić jako kombinację wypukłą zbiorów B_{j*} oraz B_{j**} .

Ta ostatnia własność oznacza, że istotnie liczbę $b^{(k)}$ można przedstawić lingwistycznie za pomocą dwóch zbiorów rozmytych.

4.3.2 Zastosowanie algorytmów genetycznych

4.3.2.1 Genetyczna regresja

Przypuśćmy, że zbiór danych postaci (4.36) chcemy opisać równaniem liniowym $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + b$. Klasyczną metodą jest w tym wypadku metoda najmniejszych kwadratów polegająca na wyznaczeniu wektora $\hat{\mathbf{a}}$ oraz wyrazu wolnego \hat{b} minimalizujących wskaźnik jakości

$$q(\mathbf{a}, b) = \frac{1}{m} \sum_{i=1}^m [\mathbf{a}^T \mathbf{x}_i + b - u^{(i)}]^2 \quad (4.49)$$

Jest to zadanie regresji liniowej. Do jego rozwiązania można też wyznaczyć algorytm genetyczny. W tym przypadku chromosomem jest wektor o składowych reprezentujących poszczególne elementy wektora \mathbf{a} , natomiast (4.49) jest naturalną funkcją jakości. Ideę tę wykorzystano m.in. do wyboru "optymalnych" parametrów funkcji przynależności.

Pomysł ten wykorzystali Karr i Gentry [38] zakładając, że dany jest zbiór obserwacji (4.36) oraz zbiór reguł sterowania postaci (4.2).

Przykład 4.3.2. [69] Załóżmy, że zachowanie układu opisuje FAM pokazana w lewej części tablicy 4.7, a dane empiryczne przedstawiono w jej prawej części.

x	S	L	1	2	3	4	5
y	S	VL	1	4	9	16	25

Tablica 4.7: Lingwistyczny opis systemu wraz z danymi empirycznymi

Przypuśćmy, że terminy lingwistyczne S modelowane są przez trójkątne funkcje przynależności postaci $(-\infty, 0, c_i^1)$ a terminy L oraz VL przez funkcje postaci (a_i^2, ξ_i^2, ∞) , $i = 1, 2$, $\xi_1^2 = 5, \xi_2^2 = 25$. Załóżmy, że dziedziną zmiennej x jest przedział $[0, 5]$, a $y \in [0, 25]$ i przyjmijmy, że pojedynczy chromosom $\mathbf{v} = (\delta_1^1, \delta_2^1, \delta_1^2, \delta_2^2)$ reprezentuje rozpiętości odpowiednich zbiorów rozmytych, tzn. $\delta_1^i = c_i^1$, $\delta_2^i = \xi_i^2 - a_i^2$. Dla każdej wartości x_j z lewej części tablicy 4.7 wyznaczamy odpowiadającą jej wartość $\hat{y}_j = \mu_{R_k}^{-1}(\mu_{L_k}(x_j))$, gdzie R_k (L_k) oznacza prawą (lewą) stronę reguły uaktywnionej przez obserwację x_j ; wyznaczamy też kwadrat błędu $e_j = (\hat{y}_j - y_j)^2$. Funkcja dostosowania ma więc postać (4.49), tzn. $f(\mathbf{v}) = \sum_j e_j$.

Przykładowo, jeżeli chromosom ma postać (0.56, 1, 59, 8.73, 20.24) to p. wartość $x_4 = 4$ aktywuje regułę z drugiej kolumny lewej części tablicy 4.7 i $\mu_L(x) = 0.37$. Wartości tej odpowiada $\hat{y} = \mu_{VL}^{-1}(0.37) = 12.25$, a zatem błąd aproksymacji jest równy w tym wypadku $(16 - 12.25)^2 = 14.06$. \square

4.3.2.2 Uczenie reguł

Przedstawiony powyżej pomysł modyfikowano na wiele sposobów. Por. np. punkt 6.2.2 w [64], a także następujące prace: [34], [40], [63], [66]. Poniżej przedstawiono algorytm zaproponowany w [29].

Załóżmy, że do modelowania zbiorów rozmytych wykorzystujemy trapezoidalne funkcje przynależności, tzn. i -ta przesłanka k -tej reguły wykorzystuje funkcję $(a_i^{(k)}, b_i^{(k)}, c_i^{(k)}, d_i^{(k)})$. Oczywiście, gdy $b_i^{(k)} = c_i^{(k)}$, otrzymujemy trójkątną funkcję przynależności.

Skoro reguła składa się z n przesłanek i pojedynczej konkluzji, to do jej reprezentacji potrzebujemy wektora \mathbf{v} o wymiarze $4(n+1)$ zbudowanego z $(n+1)$ czwórek postaci (a_i, b_i, c_i, d_i) , $i = 1, \dots, n+1$.

Do modyfikacji chromosomów można stosować np. mutację gaussowską, lub mutację niejednostajną postaci

$$v'_k = \begin{cases} v_k + \Delta(t, v_{k,max} - v_k) & \text{gdy } r < 0.5 \\ v_k - \Delta(t, v_k - v_{k,min}) & \text{wpp} \end{cases} \quad (4.50)$$

gdzie r to liczba losowa, $[v_{k,min}, v_{k,max}]$ to przedział zmienności k -tego parametru, t jest numerem generacji, natomiast

$$\Delta(t, \xi) = \xi(1 - r_1^{(1-t/T)^b})$$

przy czym r_1 jest (inną) liczbą losową, T to maksymalna liczba generacji, a b jest parametrem (np. $b = 5$).

Do wymiany informacji między osobnikami \mathbf{v} i \mathbf{w} zastosowano wariant krzyżowania arytmetycznego polegającego na wygenerowaniu czterech osobników

$$\begin{aligned} \mathbf{z}_1 &= \alpha \mathbf{v} + (1 - \alpha) \mathbf{w} \\ \mathbf{z}_2 &= \alpha \mathbf{w} + (1 - \alpha) \mathbf{v} \\ \mathbf{z}_3 &\text{ gdzie } z_i = \min\{v_i, w_i\} \\ \mathbf{z}_4 &\text{ gdzie } z_i = \max\{v_i, w_i\} \end{aligned}$$

i wybraniu dwóch najlepszych. Oczywiście $\alpha \in (0, 1)$.

Do oceny osobników wykorzystano funkcję uwzględniającą pięć aspektów

- (a) (*Intensywność*) Jeżeli przez $\tau_k(e_j)$ oznaczmy zgodność j -tego przykładu z k -tą regułą (por. równanie 4.38), to intensywność reguły opisuje funkcja

$$f_1(R^{(k)}) = \frac{1}{m} \sum_{j=1}^m \tau_k(e_j)$$

- (b) (*Dodatnie pokrycie*) Mówimy, że j -ty element zbioru uczącego jest przykładem pozytywnym, jeżeli $\tau_k(e_j) \geq p$, gdzie p jest zadana i ustalona wartością progową. Niech $E^+(R^{(k)}) = \{e_j \in E : \tau_k(e_j) \geq p\}$ będzie zbiorem przykładów pozytywnych, a $n_k^+ = |E^+(R^{(k)})|$ jego liczebnością. Wówczas uśrednione pokrycie zbioru dodatnich przykładów określa funkcja

$$f_2(R^{(k)}) = \frac{1}{n_k^+} \sum_{e_j \in E^+(R^{(k)})} \tau_k(e_j)$$

- (c) (*Negatywne pokrycie*) Przykład e_j nazywamy negatywnym, jeżeli jego zgodność z przesłanką jest dodatnia, ale zgodność z konkluzją jest zerowa. A zatem istnieje zapewne reguła z inną konkluzją pokrywająca ten przykład. Niech n_k^- będzie liczebnością zbioru przykładów negatywnych. Definiujemy kolejną funkcję

$$f_3(R^{(k)}) = \begin{cases} 1 & \text{gdy } n_k^- \leq 5 \\ (n_k^- - 5 + \exp(1))^{-1} & \text{wpp} \end{cases}$$

- (d) (*Rozpiętość reguły*) Niech $s_i = d_i - a_i$, $i = 1, \dots, n+1$, będzie rozpiętością zbioru opisującego i -tą elementarną przesłankę reguły, $m_i = c_i - b_i$ określa długość jej mediany (rozmiar jądra) tejże przesłanki, a dom_i będzie rozpiętością jej dziedziny. Okreśmy następujące wartości:

$$\rho_1 = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{s_i}{dom_i}, \quad \rho_2 = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{m_i}{dom_i}$$

Rozpiętość reguły wyznaczamy ze wzoru

$$f_4(R^{(k)}) = g_1(\rho_1)g_2(\rho_2)$$

gdzie przyjmujemy $g_i(\xi) = \exp(1 - a\xi)$ gdy preferowane są niewielkie wartości ilorazów oraz $g_i(\xi) = \exp(-|1 - a\xi|)$ gdy ważna jest ustalona relacja między odpowiednimi wielkościami.

- (d) (*Symetria*) Niech $\delta_{i,1} = b_i - a_i$, $\delta_{i,2} = d_i - c_i$, $d_i = \max(\delta_{i,1}/\delta_{i,2}, \delta_{i,2}/\delta_{i,1})$ dla $i = 1 \dots, n+1$. Wówczas

$$f_5(R^{(k)}) = (\max_{i=1, \dots, n+1} d_i)^{-1}$$

Ostatecznie jakość reguły $R^{(k)}$ wyznaczamy mnożąc wartości powyższych pięciu funkcji. Zadanie polega na maksymalizacji globalnej funkcji jakości $f(R^{(k)})$.

Algorytm generowania "optymalnych" reguł przyjmuje następującą postać:

0. (*Inicjalizacja*). Wczytać zbiór przykładów E o liczebności m . Ustalić wartości progowe p oraz ϵ . Utworzyć tablicę $CV[j] = 0$, $j = 1, \dots, m$. Podstawić $\mathfrak{R} = \mathbf{nil}$ (\mathfrak{R} jest listą reguł).
1. Zastosować algorytm genetyczny do zbioru przykładów.
2. Wybrać najlepszy chromosom, a reprezentowaną przez niego regułę R^* dodać do zbioru reguł \mathfrak{R} .
3. Wyznaczyć zbiór pozytywnych przykładów $E^+(R^*)$. Dla każdego $e_j \in E^+(R^*)$: $CV[j] = CV[j] + \tau_*(e_j)$. Jeżeli $CV[j] \geq \epsilon$ usunąć e_j ze zbioru przykładów.
4. Jeżeli $E = \mathbf{nil}$ to STOP, w przeciwnym razie powrót do kroku (1).

Parametry algorytmu genetycznego: populacja złożona z 60 osobników, prawdopodobieństwo krzyżowania $p_c = 0.6$, parametr $\alpha = 0.35$, prawdopodobieństwo modyfikacji chromosomu $p_m = 0.6$, a prawdopodobieństwo modyfikacji genu $p_g = p_m/[4 * (n+1)]$. Warunek zatrzymania to zadana z góry liczba iteracji. **uzupełnić** ...

4.4 Przykładowe problemy

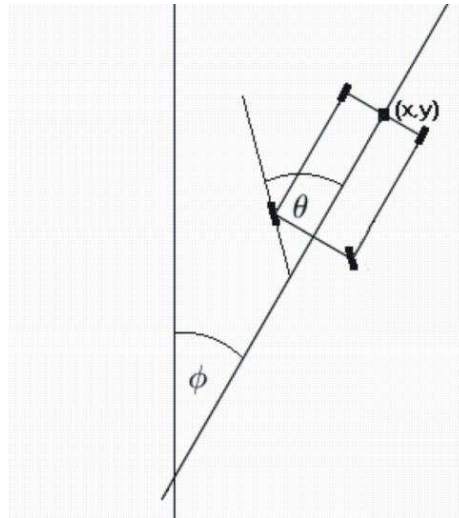
Jednym z popularniejszych zadań testowych z zakresu sterowania jest zadanie parkowania ciężarówki; omówiono je w sekcji 4.4.1. Popularnym zadaniem testowym z zakresu prognozowania, tzn. przewidywania przyszłych wartości szeregu czasowego jest prognozowanie dla szeregu generowanego przez równanie Mackey-Glassa, które omówiono w sekcji ??.

4.4.1 Zadanie parkowania ciężarówki

Ciężarówka porusza się wewnątrz prostokąta $[-150, 150] \times [0, 300]$. Należy sterować kątem skręcenia przednich kół θ w taki sposób aby zaparkować ciężarówkę tyłem do rampy znajdującej się w punkcie $(0, 0)$. Położenie ciężarówki jest określone przez trzy współrzędne: x , u (określają one środek tylnej części ciężarówki) oraz kąt ϕ pod jakim ciężarówka (wektor równoległy do niej) znajduje się względem osi u – por. rysunek 4.10. Jeżeli $\phi = 0$ – ciężarówka ulokowana jest "tyłem" do rampy. Przyjmujemy, że $\theta \in [-45^\circ, 45^\circ]$, a $\phi \in [-180^\circ, 180^\circ]$.

Równania opisujące (przybliżoną) dynamikę ciężarówki mają postać

$$\begin{aligned} x(t+1) &= x(t) + \sin[\theta(t) + \phi(t)] - [\sin \theta(t)][\cos \phi(t)] \\ u(t+1) &= u(t) - \cos[\theta(t) + \phi(t)] - [\sin \theta(t)][\sin \phi(t)] \\ \phi(t+1) &= \phi(t) - \arcsin\left[\frac{2 \sin \theta(t)}{b}\right] \end{aligned} \quad (4.51)$$



Rysunek 4.10: Parametry wejściowe, (x, u, ϕ) oraz parametr sterujący θ w zadaniu parkowania ciężarówki.

gdzie $b = 20$ jest długością ciężarówki.

Dla uproszczenia zakładamy, że współrzędna u nie będzie uwzględniana jako wejście sterownika, tzn. dane uczące mają postać $(x(t), \phi(t), \theta(t))$. Jeżeli ciężarówka znajdzie się na osi OY i będzie umiejscowiona w stosunku do niej równolegle – zadanie można uznać za zakończone. Symulację należy też przerwać jeżeli ciężarówka znajdzie się poza "placem" (porażka).

Dane uczące przygotowujemy w następujący sposób: dla każdego położenia ciężarówki $(x(t), \phi(t))$ określa się sterowanie, tzn. kąt skręcenia kół ciężarówki, $\theta(t)$ opierając się na własnej intuicji. Po kilku próbach wybiera się dane odpowiadające najbardziej płynnej trajektorii jazdy (oczywiście zakończonej powodzeniem). Aby wygenerować dane uczące wybrano szesnaście położeń początkowych ciężarówki: $(-100, -90^\circ)$, $(-100, 0^\circ)$, $(-100, 90^\circ)$, $(-100, 180^\circ)$, $(-50, -90^\circ)$, $(-50, 0^\circ)$, $(-50, 90^\circ)$, $(-50, 180^\circ)$, $(50, -90^\circ)$, $(50, 0^\circ)$, $(50, 90^\circ)$, $(50, 180^\circ)$, $(100, -90^\circ)$, $(100, 0^\circ)$, $(100, 90^\circ)$, $(100, 180^\circ)$.

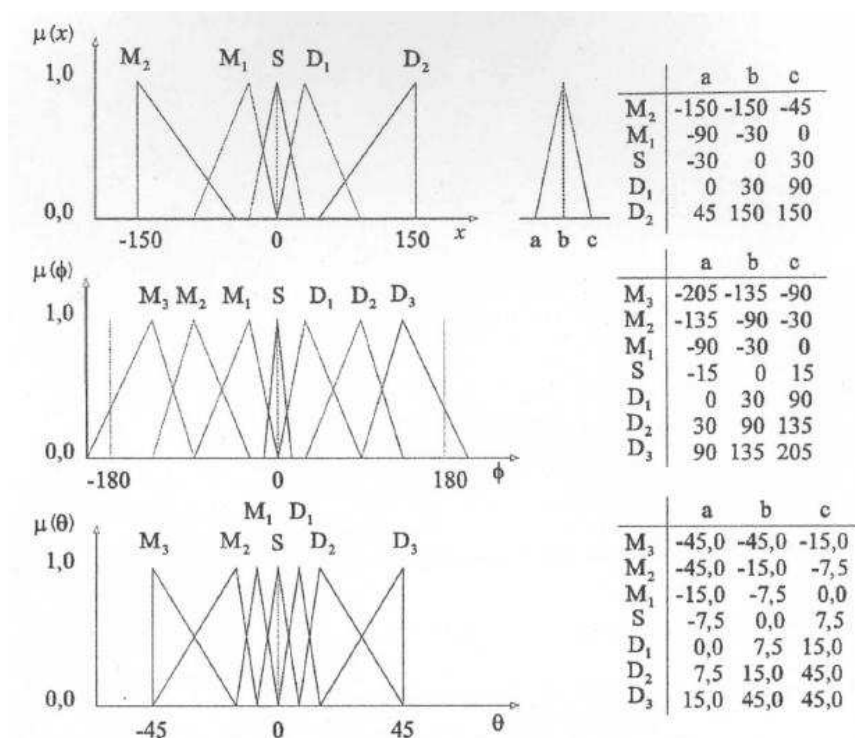
W książce [72] zastosowano do opisu badanego obiektu zbiory rozmyte, których definicje przedstawiono na rys. 4.11. W oparciu o nie oraz o zebrane dane empiryczne wygenerowano reguły przedstawione na rysunku 4.12

4.5 Równania różniczkowe

Wiedza o układach dynamicznych modelowanych za pomocą równań różniczkowych jest często niekompletna, niedokładna, czy wreszcie – niezbyt precyzyjna. Ów brak precyzji może dotyczyć wartości parametrów, zależności funkcyjnych, czy też warunków początkowych.

Pojęcie pochodnej funkcji rozmytej wprowadzili Chang i Zadeh [10], a rozwinęli Dubois i Prade [?] oraz Puri i Ralescu [?]. W tej ostatniej pracy uogólniono pojęcie H-pochodnej, czyli różniczkowalności w sensie Hukuhary odnoszącej się do odwzorowań, których wartościami są zbiory.

Prace nad równaniami różniczkowymi zainicjował pochodzący z 1978 roku artykuł [?]. Bardziej formalne podejście do tego zagadnienia, odwołujące się do pojęcia H-pochodnej, zaproponował Kaleva [?, ?] formułując twierdzenie o istnieniu jednoznaczności rozwiązania zagadnienia początkowego (o ile spełniony jest warunek Lipschitza). Spośród innych prac



Rysunek 4.11: Przykładowe definicje zbiorów rozmytych wykorzystane do sterowania procesem parkowania ciężarówki (źródło: [72]).

można wymienić m.in. [5], [?], [?], [?] czy też prace Chalco-Cano [?, ?].

4.5.1 Pojęcia wstępne

4.5.1.1 Pochodne rozmytych funkcji

W tym rozdziale będziemy korzystali z przedziałowej reprezentacji (2.22). Przykładowo, jeżeli x jest trójkątną liczbą rozmytą $\Delta(a, b, c)$ o funkcji przynależności (2.14), to przedstawiamy ją w postaci

$$(x)_\alpha = [\underline{x}(\alpha), \bar{x}(\alpha)] = [a + \alpha(b - a), c - \alpha(c - b)] \quad (4.52)$$

Niech $I \subset \mathbb{R}$ będzie pewnym przedziałem. Przez rozmyty proces będziemy rozumieli odwzorowanie $x: I \rightarrow \mathbb{R}_{\mathcal{F}}$. Jego α -przekroje mają postać, por. równanie (2.22)

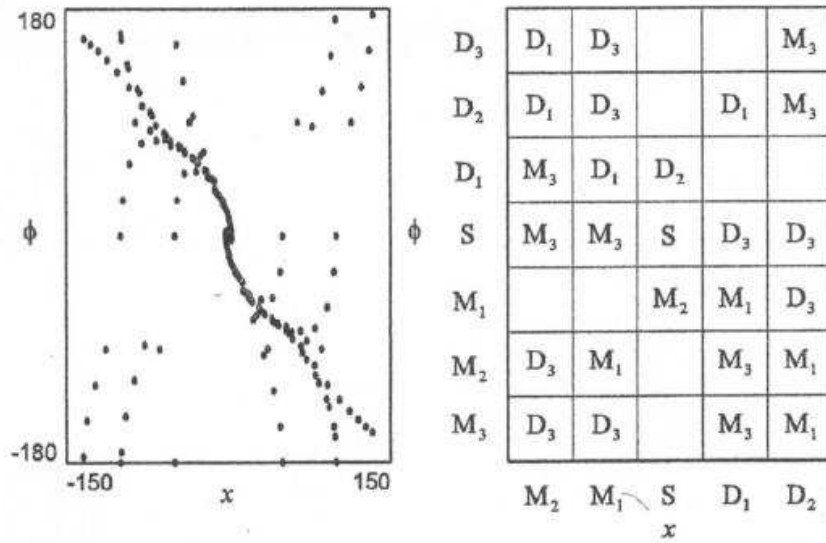
$$[x(t)]_\alpha = [\underline{x}(t; \alpha), \bar{x}(t; \alpha)], t \in I, \alpha \in (0, 1] \quad (4.53)$$

Podobnie definiujemy pochodną $x'(t)$ rozmytego procesu x

$$[x'(t)]_\alpha = [\underline{x}'(t; \alpha), \bar{x}'(t; \alpha)] \quad (4.54)$$

Rozważmy teraz funkcję $f(t, x(t))$, gdzie $x(t)$ jest liczbą rozmytą. Zgodnie z zasadą rozszerzania, stopień przynależności $f(t, x(t))(s)$ elementu $s \in \mathbb{R}$ do zbioru wartości tejże funkcji definiujemy jako

$$f(t, x(t))(s) = \sup_{s=f(t, \tau)} x(t)(\tau)$$



Rysunek 4.12: Dane empiryczne wraz z wygenerowanym zestawem reguł (źródło: [72]).

gdzie $x(t)(\tau)$ to skrótowy zapis stopnia przynależności $\mu_{x(t)}(\tau)$.

Ostatecznie, α -przekroje funkcji $f(t, x(t))$ mają postać

$$[f(t, x(t))]_{\alpha} = [f_1(t, x; \alpha), f_2(t, x; \alpha)]$$

gdzie

$$f_1(t, x; \alpha) = \min_{u \in [\underline{x}(t; \alpha), \bar{x}(t; \alpha)]} f(t, u)$$

$$f_2(t, x; \alpha) = \max_{u \in [\underline{x}(t; \alpha), \bar{x}(t; \alpha)]} f(t, u)$$

4.5.2 Rozmyte zagadnienie Cauchy'ego

Dany jest układ n równań różniczkowych postaci

$$x'_i(t) = f_i(t, x_1(t), \dots, x_n(t)) \quad (4.55)$$

z warunkiem początkowym $\mathbf{x}(0) = \mathbf{x}_0$, gdzie \mathbf{x}_0 jest wektorem, którego składowe są liczbami rozmytymi.

4.5.2.1 Liniowe równania różniczkowe z rozmytym warunkiem początkowym

Rozważmy liniowe równanie różniczkowe postaci

$$\begin{aligned} \mathbf{x}'(t) &= \mathbf{A}\mathbf{x}'(t) + \mathbf{B}\mathbf{x}(t) \\ \mathbf{x}(0) &= \tilde{\mathbf{x}}_0 \end{aligned} \quad (4.56)$$

przy założeniu, że składowe wektora $\tilde{\mathbf{x}}_0$ są liczbami rozmytymi [?].

Rozwiązanie tego równania przedstawimy w postaci α -przekrojów $[\mathbf{x}'(t)]_{\alpha} = [\underline{\mathbf{x}}'(t; \alpha), \bar{\mathbf{x}}'(t; \alpha)]$. Zgodnie z uwagami z punktu 3.4.4, układowi (4.56) odpowiada w takim wypadku układ równań wyznaczonych przez α -przekroje

$$(\mathbf{I} - \mathbf{S})\mathbf{X}'(t; \alpha) = \mathbf{TX}(t; \alpha) \quad (4.57)$$

gdzie \mathbf{I} to macierz jednostkowa, \mathbf{S} jest macierzą o elementach określonych równaniem (??), natomiast macierz \mathbf{T} wyznaczana jest podobnie jak \mathbf{S} , ale przez elementy macierzy \mathbf{B} .

Jeżeli $\mathbf{I} - \mathbf{S}$ nie jest macierzą osobiwą, to ostatnie równanie można zapisać w postaci

$$\mathbf{X}'(t; \alpha) = \mathbf{M}\mathbf{X}(t; \alpha) \quad (4.58)$$

gdzie $\mathbf{M} = (\mathbf{I} - \mathbf{S})^{-1}\mathbf{T}$. Jego rozwiązanie, przy uwzględnieniu warunku początkowego $\mathbf{X}_0 = [\underline{x}_0, \bar{x}_0]$, ma postać

$$\mathbf{X}(t; \alpha) = \exp(t\mathbf{M}) \cdot \mathbf{X}_0 \approx (\mathbf{I} + t\mathbf{M} + \frac{t^2}{2!}\mathbf{M}^2 + \frac{t^3}{3!}\mathbf{M}^3 + \dots)\mathbf{X}_0 \quad (4.59)$$

Przykład 4.5.1. Załóżmy, że

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 2 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1/3 & -1 \\ -2/3 & 0 \end{bmatrix}$$

natomiast wektor wartości początkowych ma składowe $x_{0,1} = \text{około } 2$, oraz $x_{0,2} = \text{około } 1$, co modelujemy przez α -przekroje $(x_{0,1})_\alpha = (1 + \alpha, 3 - \alpha)$, $(x_{0,2})_\alpha = (\alpha, 2 - \alpha)$.

Postępując zgodnie z opisaną wyżej procedurą, układ równań różniczkowych przyjmuje postać

$$\begin{bmatrix} -1 & 0 & 0 & 1 \\ -2 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} \underline{x}'_1 \\ \underline{x}'_2 \\ \bar{x}'_1 \\ \bar{x}'_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 & 0 & -1 \\ 0 & 0 & -\frac{2}{3} & 0 \\ 0 & -1 & \frac{1}{3} & 0 \\ -\frac{2}{3} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

Ostatecznie, rozwiązanie równania ma postać

$$\begin{bmatrix} \underline{x}'_1 \\ \underline{x}'_2 \\ \bar{x}'_1 \\ \bar{x}'_2 \end{bmatrix} = \mathbf{K}_m \begin{bmatrix} 1 + \alpha \\ \alpha \\ 3 - \alpha \\ 2 - \alpha \end{bmatrix}$$

gdzie $\mathbf{K}_m \approx \exp(t\mathbf{M}) = \mathbf{I} + \sum_{k=1}^m \frac{t^k}{k!}\mathbf{M}^k$, a macierz \mathbf{M} ma postać

$$\mathbf{M} = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 & -1 \\ 2 & -4 & 2 & -2 \\ 2 & -1 & 1 & -2 \\ 2 & -2 & 2 & -4 \end{bmatrix}$$

□

Rozdział 5

Sieci jednowarstwowe

5.1 Podstawowe określenia

Formalnie rzecz biorąc sieć neuronowa to ważony, skierowany graf, którego węzły reprezentują neurony, natomiast wagi wskazują na udział konkretnego sygnału w ostatecznej odpowiedzi.

W większości swoich zastosowań sieci neuronowe realizują zadanie uczenia nadzorowanego. Formułujemy je w następujący sposób. Dana jest przestrzeń wejściowa X , oraz zbiór T zawierający etykiety klas¹. Zazwyczaj $X \subset \mathbf{R}^n$, natomiast T jest k -elementowym zbiorem etykiet (numerów) klas, a k jest z góry ustaloną liczbą klas. Jeżeli rozważamy zadanie klasyfikacji binarnej (tzn. $k = 2$), przyjmuje się zazwyczaj, że $T = \{-1, 1\}$; klasę o etykiecie -1 nazywa się klasą negatywną, a klasę o etykiecie 1 – klasą pozytywną.

Starannie wyselekcjonowany, na ogół dyskretny, podzbiór $\mathcal{L} \subset X \times T$ tworzy tzw. zbiór trenujący, nazywany też próbą uczącą. Stanowi on kolekcję par postaci

$$\mathcal{L} = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_p, t_p)\} \quad (5.1)$$

Pierwszy element pary (\mathbf{x}_i, t_i) , $i = 1, \dots, p$ nazywamy przykładem (lub instancją), a drugi element – etykietą tego przykładu. Indeks p oznacza liczbę przykładów. Współrzędne wektora \mathbf{x}_i nazywamy cechami lub atrybutami, a często – po prostu zmiennymi.

Problem uczenia nadzorowanego formułuje się teraz jako zadanie znalezienia odwzorowania $f: X \rightarrow T$ takiego, że

$$f(\mathbf{x}_i) = t_i, \quad i = 1, \dots, m \quad (5.2)$$

Funkcję f nazywamy funkcją (regułą) decyzyjną, albo funkcją dyskryminacyjną, albo w skrócie: klasyfikatorem. Z definicji przydziela ona każdemu elementowi $\mathbf{x} \in X$ etykietę $f(\mathbf{x}) \in T$. Warto odnotować w tym momencie trzy uwagi:

- (a) Funkcja decyzyjna tworzona jest na podstawie próby uczącej \mathcal{L} . Dlatego uczenie nadzorowane bywa też nazywane klasyfikacją z próbą uczącą, lub klasyfikacją z nauczycielem, albo klasyfikacją pod nadzorem.
- (b) Próba ucząca służy do skonstruowania reguły klasyfikacyjnej. Samo zadanie klasyfikacji pod nadzorem jest zadaniem predykcji, czyli *przewidzenia* klasy do której należą nowe obserwacje. Typową procedurą oceny jakości klasyfikatora jest zbadanie trafności predykcji na starannie przygotowanej próbie walidacyjnej \mathcal{W} , która zawiera

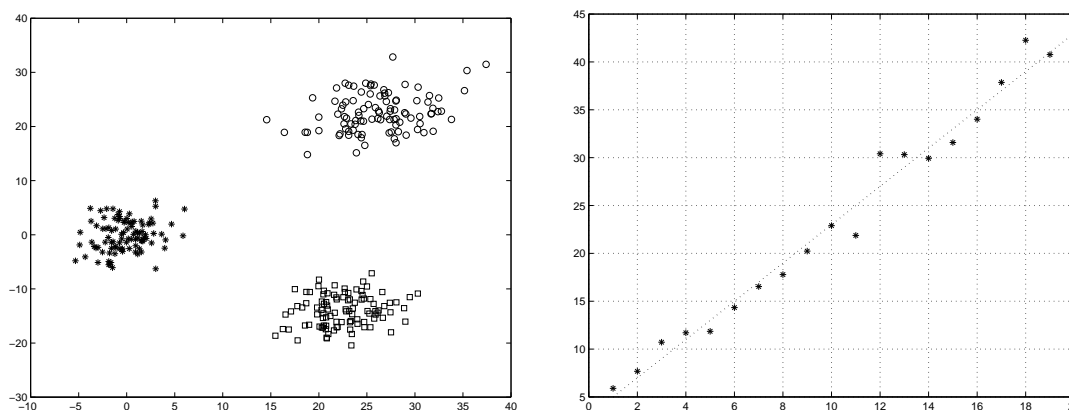
¹W literaturze poświęconej sieciom neuronowym symbol t jest interpretowany jako wartość pożądana (*target*).

$m_1 < m$ przykładów różnych od przykładów z próby uczącej. Szczegółowy opis takiej procedury można znaleźć w punkcie 2.3.2 monografii [46].

- (c) Jeżeli t_i są wartościami rzeczywistymi, tzn. $T \subset \mathbb{R}$, mamy do czynienia z zadaniem regresji polegającym na znalezieniu funkcji $r: X \rightarrow T$, nazywanej funkcją regresji, która minimalizuje błąd

$$\sum_{i=1}^m (t_i - r(\mathbf{x}_i))^2 \quad (5.3)$$

W najprostszym przypadku, gdy r jest funkcją liniową, tzn. $r = \mathbf{w}^T \mathbf{x} + b$, mówimy o zadaniu regresji liniowej. Sprowadza się ono wówczas do zadania wyboru takiego wektora współczynników \mathbf{w} i stałej b , które minimalizują wyrażenie (5.3). Jeżeli r może być dowolną funkcją rzeczywistą zależną od n zmiennych, to mówimy o zadaniu regresji nieliniowej. Różnicą między klasyfikacją a regresją zilustrowano symbolicznie na rysunku 5.1.



Rysunek 5.1: Różnica między klasyfikacją a regresją. W pierwszym przypadku (lewy rysunek) zbiór trenujący składa się z przykładów poszczególnych klas i należy znaleźć regułę pozwalającą określić przynależność dowolnego obiektu do jednej z wyróżnionych klas. W drugim przypadku (prawy rysunek) należy znaleźć taką funkcję (zaznaczona linią przerywaną), która w najlepszy sposób opisuje relacje między zmiennymi objaśniającymi x_1, \dots, x_n , a zmienną objaśnianą t .

W przypadku statystyki zadanie uczenia nadzorowanego rozwiązuje się metodami analizy dyskryminacyjnej. Warto jednak odnotować pewną subtelność różnicę między tymi zadaniami a zadaniem klasyfikacji z nauczycielem – por. [46]. W pierwszym przypadku (zadanie analizy dyskryminacyjnej) chodzi o jak najlepsze, w wybranym sensie, opisanie różnic między klasami, a w drugim – o skonstruowanie możliwie dobrej reguły przydziału obserwacji o nieznanym przynależności do jednej z klas wymienionych w zbiorze T .

Badania nad sieciami neuronowymi sięgają lat 50-tych XX wieku. Inicjują je prace McCullocha i Pittsa² oraz Hebba³, którzy stali na stanowisku, że inteligentne systemy powinny stanowić odwzorowanie pewnych funkcji układu nerwowego. Minsky i Pappert [53] przeciwstawili się poglądom lansowanym przez owych autorów twierdząc, że inteligencja opiera się na przetwarzaniu symboli w sposób przypominający działanie komputera von

²Por. W.S. McCulloch, W.A. Pitts. A logical calculus of the ideas imminent in nervous activity. *Bull. Math. Biophys.* 5(1943) 115-133.

³Por. D.O. Hebb. *Organization of Behavior*, Wiley, New York, 1949.

Neumanna. Nowego impulsu dla rozwoju sieci neuronowych dostarczają, ukazujące się w odstępie dwóch lat poczynając od 1982 r., prace Hopfielda [31], Kohonena [45] i Rummelharta [70]. Stworzyli oni nie tylko podstawy matematyczne dla zrozumienia dynamiki pewnych typów sieci (Hopfield), ale zajęli się także zagadnieniami uczenia nienadzorowanego (Kohonen), a ponadto wprowadzili struktury wielowarstwowe (Rummelhart). Wyczerpujący przegląd zastosowań sieci neuronowych można znaleźć np. w [71], jak również na wielu stronach internetowych, np. <http://www.calsci.com/Applications.html>.

5.2 Klasyfikacja liniowa

Prace nad liniowymi metody klasyfikacji zostały zapoczątkowane na gruncie statystyki przez sir Ronalda Fishera w 1936 roku, a na gruncie sztucznej inteligencji przez Franka Rosenblatta w 1957 roku⁴. O ile w pierwszym przypadku mówi się o liniowej dyskryminacji Fishera, o tyle w drugim – o perceptronie.

W formułowanych w teorii sieci neuronowych zadaniach klasyfikacji binarnej przyjmuje się, że obiekt $\mathbf{x} \in X$ należy do klasy pozytywnej jeżeli jednostka wyjściowa zwraca sygnał $y(\mathbf{x}) \geq 0$, a w przeciwnym razie – do klasy negatywnej.

Aby podkreślić fakt, że neuron wyjściowy kolekcjonuje sygnały wejściowe, a następnie emituje sygnał wyjściowy, wprowadzamy następujące oznaczenia. Pobudzenie neuronu określa się jako

$$\begin{aligned} y_{in} &= b + \langle \mathbf{w}, \mathbf{x} \rangle \\ &= b + \sum_{i=1}^n w_i x_i \end{aligned} \quad (5.4)$$

gdzie $\langle \mathbf{w}, \mathbf{x} \rangle$ oznacza iloczyn skalarny wektorów \mathbf{w} i \mathbf{x} , a $(b, \mathbf{w}) \in \mathbb{R} \times \mathbb{R}^n$ są parametrami sterującymi, które podlegają procesowi uczenia w oparciu o zbiór uczący. Często przyjmuje się, że przykłady są obiektami $(n+1)$ -wymiarowymi postaci $\mathbf{x} = (1, \mathbf{x}^T)^T$, czyli obiektami o współrzędnych $\mathbf{x} = (x_0, x_1, \dots, x_n)^T$, przy czym $x_0 = 1$. W takim wypadku równanie (5.4) można zapisać w zwartej postaci

$$\begin{aligned} y_{in} &= \langle \mathbf{w}, \mathbf{x} \rangle \\ &= \sum_{i=0}^n w_i x_i \end{aligned} \quad (5.5)$$

Sygnał y_{in} przekształcany jest następnie przez neuron do postaci

$$y = \alpha(y_{in}) \quad (5.6)$$

rejestrowanej na jego wyjściu. Powyższe przekształcenie to właśnie reguła klasyfikacyjna. Można zapisać ją w zwartej postaci jako $y = \text{sgn}(y_{in})$, gdzie

$$\alpha(y_{in}) = \text{sgn}(y_{in}) = \begin{cases} +1 & \text{gdy } y_{in} \geq 0 \\ -1 & \text{gdy } y_{in} < 0 \end{cases} \quad (5.7)$$

Wektor \mathbf{w} nosi nazwę wektora wagowego, albo w skrócie – wag, natomiast b to tzw. odchylenie lub obciążenie (ang. *bias*). Czasem $-b$ zastępuje się przez θ , tzn. przez wielkość

⁴Por. Rosenblatt, F. Perceptron: A perceiving and recognizing automaton (Project PARA). New York: Cornell Aeronautical Laboratory, January, 1957.

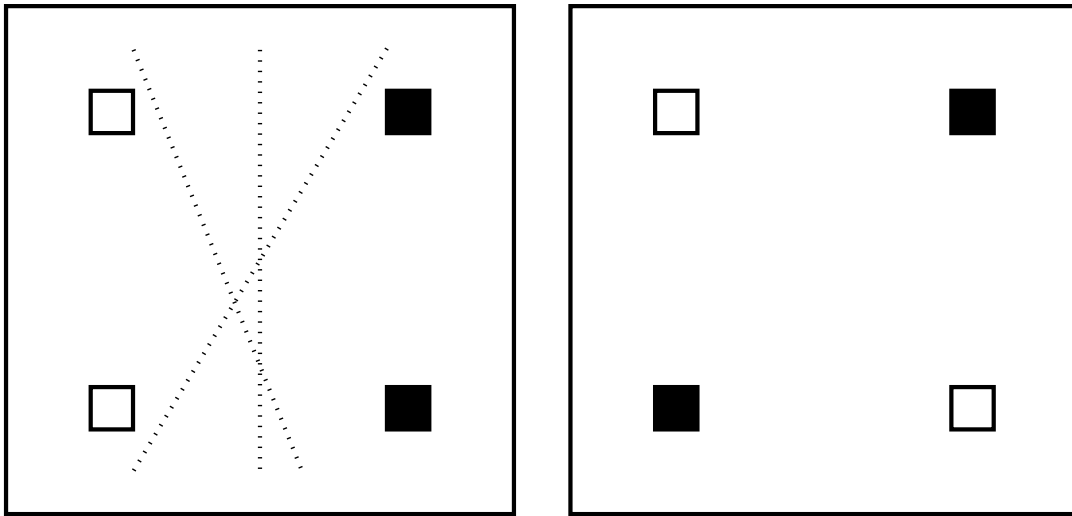
nazywaną progiem aktywacji, lub po prostu progiem. W tym ostatnim przypadku reguła decyzyjna ma postać

$$\alpha(y_{in}) = \begin{cases} +1 & \text{gdy } \langle \mathbf{w}, \mathbf{x} \rangle \geq \theta \\ -1 & \text{gdy } \langle \mathbf{w}, \mathbf{x} \rangle < \theta \end{cases} \quad (5.8)$$

Geometrycznie reguła decyzyjna tworzy podział zbioru obiektów X na k rozłącznych podzbiorów. Obserwacje z danego podzbioru są zaliczane do jednej i tylko jednej z k możliwych klas. Brzegi takich podzbiorów określają hiperpowierzchnie rozdzielające klasy. W przypadku binarnego klasyfikatora liniowego ową hiperpowierzchnią jest hiperpłaszczyzna \mathcal{H} opisana równaniem (5.4). Wektor \mathbf{w} wskazuje kierunek prostopadły do hiperpłaszczyzny, natomiast zmianom parametru b odpowiadają równoległe przesunięcia hiperpłaszczyzny zgodnie z kierunkiem \mathbf{w} .

Definicja 5.2.1. Jeżeli hiperpłaszczyzna zadana równaniem (5.4) pozwala poprawnie przydzielić elementy zbioru trenującego \mathcal{L} do odpowiednich klas, to \mathcal{L} nazywamy zbiorem liniowo separowalnym. W przypadku przeciwnym mówimy, że \mathcal{L} jest zbiorem nieseparowalnym. \square

Na rysunku 5.2 podano przykłady zbiorów obu typów. Istotne jest to, że w przypadku zbiorów separowalnych istnieje wiele hiperpłaszczyzn rozdzielających elementy należące do poszczególnych klas.



Rysunek 5.2: Przykład zbioru separowalnego (lewy rysunek) i nieseparowalnego (prawy rysunek). W przypadku zbiorów separowalnych istnieje na ogół wiele hiperpłaszczyzn oddzielających poszczególne klasy. W przypadku 2-wymiarowym są to linie; kilka z nich zaznaczono liniami przerywanymi

5.3 Algorytm perceptronu

W latach 60-tych zaproponowano kilka iteracyjnych algorytmów optymalizujących różne funkcje kosztu stosowane do rozdzielania hiperpłaszczyzną dwóch zbiorów obiektów. Oryginalny algorytm uczenia wektora wag i wartości progu aktywacji w oparciu o próbę uczącą sformułował Rosenblatt w pracy [68]. Jest to sekwencyjna procedura korekcji błędów, której istotą jest iteracyjne „poprawianie” wektora wag i wartości progu jeżeli aktualnie

prezentowany przykład został źle sklasyfikowany. Modyfikacja wag odbywa się zgodnie z pseudokodem 5.1.

Algorytm 5.1 Prymalny algorytm perceptronu, [11]

WE: liniowo separowalny zbiór trenujący \mathcal{L} oraz współczynnik uczenia $\eta \in \mathbb{R}$.

WY: wektor wagowy \mathbf{w} oraz próg aktywacji $\theta = -b$.

(a) (Inicjalizacja) $\mathbf{w}_0 = \mathbf{0}$, $b_0 = 0$, t .

$$R = \max_{1 \leq i \leq p} \|\mathbf{x}_i\|.$$

(b) powtarzaj

dla $i = 1, \dots, p$

jeżeli i -ty obiekt został błędnie sklasyfikowany, to

$$(i) \quad \mathbf{w}_{t+1} = \mathbf{w}_t + \eta t_i \mathbf{x}_i$$

$$(ii) \quad b_{t+1} = b_t + \eta t_i R^2$$

$$(iii) \quad t = t + 1$$

dopóki perceptron nie popełni żadnej pomyłki w klasyfikacji zbioru trenującego.

(c) zwróć \mathbf{w}_t , b_t , oraz t – liczbę popełnionych błędów.

Uwaga 5.3.1. W notacji poszerzonej równania (i) oraz (ii) można zapisać w postaci pojedynczego równania

$$\mathbf{v}_{t+1} = \mathbf{v}_t + \eta t_i \mathbf{x}_i \quad (5.9)$$

Taka modyfikacja skutkuje jednak wolniejszą zbieżnością algorytmu. Ponadto, w opisie nieliniowych klasyfikatorów, takich jak np. SVM, wyraźnie odróżnia się wagi \mathbf{w} i współczynnik b . Taką też notację będziemy stosowali w tym punkcie.

Równanie (5.9) pozostaje zwartym zapisem równań (i) oraz (ii) z algorytmu 5.1 jeżeli tylko założymy, że poszerzona reprezentacja obiektu ma postać $\mathbf{x} = (R^2, \mathbf{x}^T)^T$. Nie jest to postać uniwersalna, gdyż najpierw należy obliczyć wartość R^2 , ale znakomicie upraszcza ona zapis równań perceptronu. \square

Dowodzi się, por. np. [83], że jeżeli zbiór \mathcal{L} jest liniowo separowalny i $\eta > 0$, to algorytm perceptronu znajduje hiperpłaszczyznę rozdzielającą w skończonej liczbie kroków. Liczba ta zależy od stałej R oraz wielkości nazywanej marginesem.

Definicja 5.3.1. Niech (\mathbf{w}, b) będzie pewną hiperpłaszczyzną, a (\mathbf{x}_i, t_i) pewnym przykładem i jego etykietą. Marginesem funkcyjnym tego przykładu względem hiperpłaszczyzny rozdzielającej jest wielkość

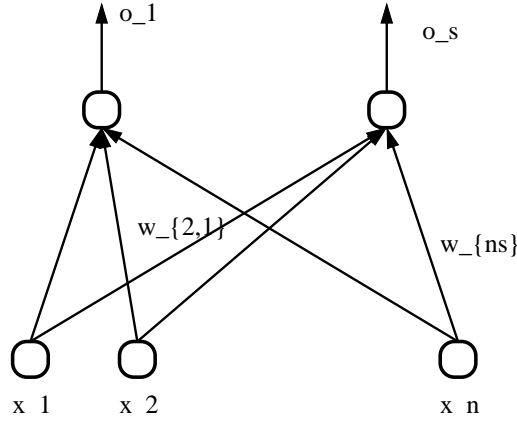
$$\gamma_i = t_i \cdot (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \quad (5.10)$$

\square

Zauważmy, że $\gamma_i > 0$ oznacza prawidłową klasyfikację danego obiektu.

W przypadku, gdy sieć posiada $s > 1$ wyjść, por. rysunek 5.3, rolę wektora wag pełni macierz wag $\mathbf{W} = [w_{lj}]_{n \times s}$, gdzie w_{lj} to waga połączenia l -tego wejścia (pobierającego l -tą cechę obiektu) z j -tym wyjściem. Podobnie wektor \mathbf{t} zastępujemy przez macierz $\mathbf{T} = [t_{ij}]_{m \times s}$, której elementy reprezentują wartość j -tej etykiety dla i -tego przykładu. W tym wypadku reguła uczenia perceptronu przyjmuje postać

$$w_{lj} \leftarrow w_{lj} + \frac{\eta}{2} \left[t_{ij} - \text{sgn}(b_j + \sum_{l=1}^n w_{lj} x_{il}) \right] \mathbf{x}_i \quad (5.11)$$



Rysunek 5.3: Jednowarstwowa sieć neuronowa złożona z n jednostek wejściowych i s jednostek wyjściowych.

$$b_j \leftarrow b_j + \frac{\eta}{2} \left[t_{ij} - \text{sgn}(b_j + \sum_{l=1}^n w_{lj} x_{il}) \right] R^2 \quad (5.12)$$

t.b.c.

5.4 Reguła *delta*.

Jej autorami są Widrow i Hoff [80]. O ile w regule perceptronu adjustacja wag dokonywana jest w przypadku, gdy jednostka wyjściowa zwraca wartość różną od wartości etykiety przypisanej prezentowanemu obiektowi, to teraz istotą algorytmu jest minimalizacja błędu będącego kwadratem różnicy między wartością etykiety t oraz $f(\mathbf{x})$. Poprawia to znacząco zdolności generalizacyjne wynikowej sieci jednowarstwowej. Sieć ta znana jest w literaturze pod nazwą adaptacyjny sumator liniowy, albo ADALINE. Regułę delta, nazywaną też regułą Widrowa-Hoffa, można traktować jako prekursora algorytmu propagacji wstecznej.

5.4.1 Reguła Widrowa-Hoffa

Rozpoczynamy od klasycznego sformułowania Widrowa i Hoffa, w którym zakłada się, że α jest liniową funkcją aktywacji, tzn. $\alpha(\xi) = \xi$, natomiast błąd wynikający z prezentacji obiektu \mathbf{x} ma postać

$$E(\mathbf{w}) = \frac{1}{2} (t - \alpha[y_{in}])^2 = (t - y_{in})^2 \quad (5.13)$$

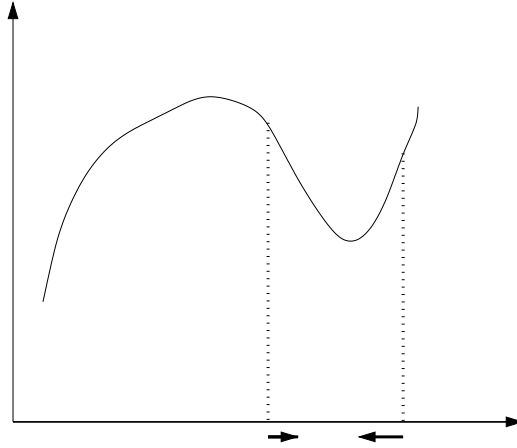
gdzie jak poprzednio $y_{in} = b + \mathbf{w}^T \mathbf{x}$ oznacza aktywację (pobudzenie) neuronu wyjściowego. Minimalizacja błędu wymaga rozwiązania równania $\partial E / \partial \mathbf{w} = 0$. Pochodna błędu względem wektora wagowego ma postać

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{w}} &= \frac{\partial (t - y_{in})^2}{\partial \mathbf{w}} \\ &= -(t - y_{in}) \frac{\partial y_{in}}{\partial \mathbf{w}} = -(t - y_{in}) \frac{\partial (\mathbf{w}^T \cdot \mathbf{x})}{\partial \mathbf{w}} \\ &= -(t - y_{in}) \mathbf{x} \end{aligned} \quad (5.14)$$

Zauważmy, że pochodna funkcji $g: \mathbb{R} \rightarrow \mathbb{R}$, czyli

$$g'(x) = \lim_{x_n \rightarrow 0} \frac{g(x) - g(x_n)}{x - x_n} \quad (5.15)$$

wskazuje kierunek wzrostu wartości funkcji. Jeżeli więc szukamy minimum funkcji metodą iteracyjną, to kolejne punkty jego przybliżenia x_0, x_1, \dots wybiera się w kierunku przeciwnym do kierunku pochodnej. Tzn. jeżeli $g'(x_0) < 0$, tzn. w punkcie x_0 funkcja g jest malejąca, to punkt x_1 powinien leżeć na prawo od x_0 , natomiast jeżeli $g'(x_0) > 0$, czyli g jest funkcją rosnącą w tym punkcie, to x_1 powinien leżeć na lewo od x_0 – por. rysunek 5.4.



Rysunek 5.4: Reguła najszybszego spadku. Wybór kolejnego punktu odbywa się w kierunku przeciwnym do kierunku pochodnej.

Korzystając z tej obserwacji, rozwiązanie równania $\partial E / \partial \mathbf{w} = 0$ prowadzi do równań postaci

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta(t - y_{in})\mathbf{x} \quad (5.16)$$

$$b_{t+1} = b_t + \eta(t - y_{in})R^2 \quad (5.17)$$

Uwaga 5.4.1. Często różnica $(t - y_{in})$ oznaczana jest symbolem δ , tzn.

$$\delta = (t - y_{in}) \quad (5.18)$$

Stąd bierze się określenie „reguła delta”. Ponadto wprowadzając kolejne oznaczenie $\Delta \mathbf{w} = \mathbf{w}_{t+1} - \mathbf{w}_t$ powyższe równania zapisuje się w zwartej postaci

$$\begin{aligned} \Delta \mathbf{w} &= \eta \delta \mathbf{x} \\ \Delta b &= \eta \delta R^2 \end{aligned} \quad (5.19)$$

□

Pseudokod 5.2 opisuje implementację algorytmu ADALINE. Przyjęto tu prostszy warunek zatrzymania postaci

$$\max_{i=1, \dots, p} (t_i - y_i)^2 < \epsilon \quad (5.20)$$

gdzie ϵ to dokładność aproksymacji zbioru etykiet przez prostą $\mathbf{w}^T \mathbf{x} + b$. Inną często stosowaną postacią warunku zakończenia jest

$$\sum_{i=1}^m (t_i - y_i)^2 < \epsilon \quad (5.21)$$

Algorytm 5.2 Algorytm ADALINE

Require: zbiór trenujący, współczynnik uczenia $\eta \in \mathbb{R}$, oraz dokładność ϵ .

Ensure: wektor wagowy \mathbf{w} oraz próg aktywacji $\theta = -b$.

```

1: (Inicjalizacja)  $\mathbf{w}_0 = \mathbf{0}$ ,  $b = 0$ ,  $t = 0$ .
    $e = +\infty$ 
   done = false;
2: while (not done) do
3:   for  $i = 1, \dots, p$  do
4:     wyznacz wartość  $y_{in_i} = \mathbf{w}^T \mathbf{x}_i + b$ 
5:     wyznacz  $\delta_i = t_i - y_i$ 
6:     if  $|e - |\delta_i|| < \epsilon$  then
7:       done = true
8:     else
9:        $e = |\delta_i|$ 
10:    end if
11:    modyfikuj wagi zgodnie z równaniami (5.19)
12:  end for
13: end while

```

Przykład 5.4.1. Rozważmy problem uczenia dla zaszumionych danych generowanych zgodnie z równaniem

$$(3x_1 + 7x_2 - 4)(1 + 0.05\epsilon) \quad (5.22)$$

gdzie ϵ jest liczbą losową z przedziału $[-1, 1]$. Innymi słowy pomiar danych obarczony jest błędem $\pm 5\%$.

Dla celów testowych wygenerowano $p = 2000$ punktów, przy czym x_1, x_2 są każdorazowo losowane ze zbioru $[-5, 10]$.

Trzykrotne uruchomienie algorytmu (z nowym losowo wygenerowanym zbiorem danych trenujących) daje, przy $\eta = 0.001$ i losowej inicjalizacji wektora \mathbf{w} i współczynnika b , następujące wyniki

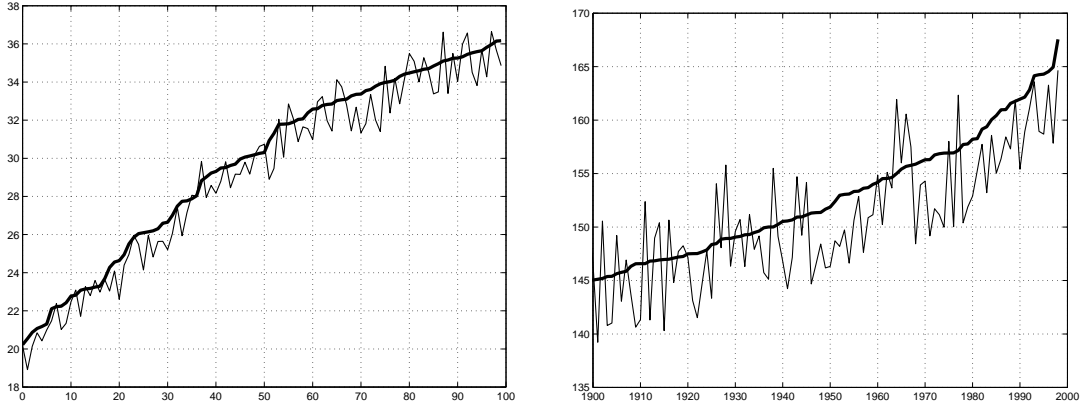
lp	b	w_1	w_2	odl
1	-4.079547237	2.993985367	6.840148085	0.1787
2	-3.932731443	2.93184988	7.06128253	0.1137
3	-3.857211862	3.012499594	7.022436701	0.1451

Tablica 5.1: Wartości wektora wagowego uzyskane w trzech losowych eksperymentach. W ostatniej kolumnie podano odległość estymowanego wektora od wektora docelowego $w^* = (-4, 3, 7)^T$.

Z kolei na rysunku 5.6 pokazano relację między mierzonymi wartościami t_i (zaznaczono je pogrubioną linią), a wartościami generowanymi przez sieć.

W przypadku równań liniowych nawet niewielka liczba danych trenujących wystarcza do prawidłowego oszacowania składowych wektora wagowego. Wzrasta co najwyżej liczba

epok niezbędnych do jego stabilnego oszacowania. W rozważanym tu przypadku $p = 10$ losowych punktów należących do płaszczyzny $3x_1 + 7x_2 - 4 = 0$ pozwala znaleźć następujące wagi: $b = -3.929956789941455$, $w_1 = 2.999679031635004$, $w_2 = 6.995109367575962$. Ich nauczanie wymaga jednak 8512 epok, gdy $\eta = 0.001$. \square



Rysunek 5.5: Wartości rzeczywiste (linia pogrubiona), a wartości wyznaczone przez sieć neuronową dla funkcji liniowej postaci (5.22). Na lewym rysunku podano 100 pierwszych pomiarów, a na prawym – ostatnie 100 pomiarów.

Uwaga 5.4.2. Podobnie jak w przypadku perceptronu można rozważać sieć z $s > 1$ wyjściami. W tym wypadku $\delta_i = t_i - f_i(\mathbf{x})$, $i = 1, \dots, s$. \square

5.4.2 Wybór parametrów

Widrow, Winter i Baxter sugerują w [81] aby współczynnik uczenia był stosunkowo niewielką liczbą spełniającą warunek

$$0.1 \leq \eta < 1.0 \quad (5.23)$$

przy czym proponują oni aby równania (5.19) zastąpić przez

$$\begin{aligned} \Delta \mathbf{w} &= \frac{\eta}{\|\mathbf{x}\|^2} \delta \mathbf{x} \\ \Delta b &= \frac{\eta}{\|\mathbf{x}\|^2} \delta R^2 \end{aligned} \quad (5.24)$$

Oznacza to adaptację rzeczywistej wartości współczynnika uczenia do aktualnie prezentowanego obiektu.

Baum i Haussler badali w pracy [3] jak liczny powinien być zbiór uczący aby uzyskać zadowalającą dokładność uczenia. Ich wyniki dotyczą wprowadzie wielowarstwowych sieci uczonych metodą propagacji wstecznej, ale mogą one stanowić także pewną wskazówkę w przypadku sieci jednowarstwowych. Mianowicie, jeżeli przez Ω oznaczymy liczbę wag, a przez ϵ dokładność aproksymacji, to

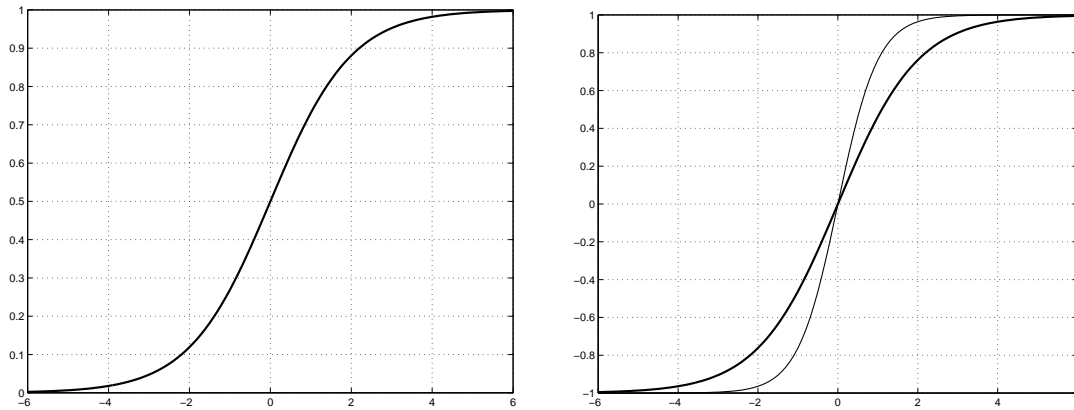
$$m \approx \frac{\Omega}{\epsilon} \quad (5.25)$$

5.4.3 Reguła delta z nieliniową funkcją aktywacji

W praktycznych zastosowaniach oprócz liniowej funkcji aktywacji stosuje się także i inne funkcje. Najbardziej popularne wymieniono w tabelicy 5.2. Dla prostoty zastosowano tu „poszerzoną” reprezentację obiektów i wag, tzn. $\mathbf{x} = (1, x_1, \dots, x_n)^T$, $\mathbf{w} = (b, w_1, \dots, w_n)^T$.

Nazwa	Postać
Liniowa	$\alpha_1(y_{in}) = \mathbf{w}^T \mathbf{x}$
Kawałkami liniowa	$\alpha_2(y_{in}) = \begin{cases} 1 & \text{gdy } \mathbf{w}^T \mathbf{x} > 1 \\ \langle \mathbf{w}, \mathbf{x} \rangle & \text{gdy } \mathbf{w}^T \mathbf{x} \leq 1 \\ -1 & \text{gdy } \mathbf{w}^T \mathbf{x} < -1 \end{cases}$
Unipolarna sigmoidalna	$\alpha_3(y_{in}) = \left(1 + \exp(-\mathbf{w}^T \mathbf{x})\right)^{-1}$
Bipolarna sigmoidalna (1)	$\alpha_4(y_{in}) = \tanh(\mathbf{w}^T \mathbf{x})$
Bipolarna sigmoidalna (2)	$\alpha_5(y_{in}) = 2/(1 + \exp(-\mathbf{w}^T \mathbf{x})) - 1$

Tablica 5.2: Często stosowane postacie funkcji aktywacji. Uwaga: przyjęto tu „poszerzoną” reprezentację obiektów i wag, tzn. $\mathbf{x} = (x_1, \dots, x_n, R^2)^T$, $\mathbf{w} = (w_1, \dots, w_n, b)^T$.



Rysunek 5.6: Unipolarna (lewy rysunek) i bipolarne funkcje aktywacji (prawy rysunek). Linia cieńszą zaznaczono funkcję $\tanh(\cdot)$, a pogrubioną – funkcję α_5 . Zauważmy, że o ile wartości unipolarnej funkcji aktywacji należą do zbioru $[0, 1]$, o tyle wartości bipolarnych funkcji należą do zbioru $[-1, 1]$.

Załóżmy, że wybrano unipolarną funkcję aktywacji. W takim wypadku błąd E wyraża się formułą

$$E(\mathbf{w}) = \frac{1}{2}(t - y)^2 = \frac{1}{2} \left(t - \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x})} \right)^2 \quad (5.26)$$

Biorąc pod uwagę, że (por. problem 5.5.2)

$$\frac{\partial y}{\partial \mathbf{w}} = y(1 - y)\mathbf{x}$$

można łatwo sprawdzić, że modyfikacja wag odbywa się zgodnie z regułą

$$\Delta \mathbf{w} = \eta \delta y(1 - y)\mathbf{x} \quad (5.27)$$

5.5 Zadania

Problem 5.5.1. Sprawdź prawdziwość ostatniego akapitu w uwadze 5.3.1.

Problem 5.5.2. Sprawdź, że jeżeli $f(\xi) = 1/(1 + e^{-\xi})$, to

$$f'(\xi) = f(\xi)[1 - f(\xi)]$$

Problem 5.5.3. Sprawdź, że jeżeli $f(\xi) = 2/(1 + e^{-\xi}) - 1$, to

$$f'(\xi) = \frac{2e^{-\xi}}{(1 + e^{-\xi})^2} = \frac{1}{2}[1 - f^2(\xi)]$$

Problem 5.5.4. Jeżeli wartości aproksymowanej funkcji należą do przedziału $[\underline{f}, \bar{f}]$, to często stosuje się sigmoidalną funkcję aktywacji $\beta(\xi) = a\alpha(\xi) + b$, gdzie $a = \bar{f} - \underline{f}$, $b = -\underline{f}$, natomiast $\alpha(\xi) = 1/(1 + e^{-\xi})$. Sprawdź, że

$$\beta'(\xi) = \frac{1}{a}[b + \beta(\xi)][a - b - \beta(\xi)]$$

W szczególności, gdy $a = 2$, $b = 1$, otrzymujemy bipolarną funkcję z przykładu 5.5.5.

Problem 5.5.5. Wprowadzając dodatkowy parametr c można sterować szybkością wzrostu funkcji aktywacji α , a w konsekwencji funkcji β z poprzedniego przykładu. Sprawdź, że jeżeli $\alpha(\xi) = 1/(1 + ce^{-\xi})$, to

$$\beta'(\xi) = \frac{c}{a}[b + \beta(\xi)][a - b - \beta(\xi)]$$

Problem 5.5.6. Podaj postać delta reguły dla bipolarnej funkcji sigmoidalnej (2), tzn. funkcji postaci $f(\xi) = 2/(1 + e^{-\xi}) - 1$.

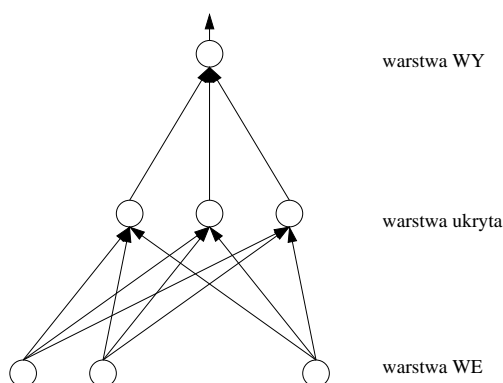
Rozdział 6

Sieci wielowarstwowe

6.1 Uogólniona reguła delta

Rozważmy prostą sieć trójwarstwową, por. rysunek 6.2. Zawiera ona n jednostek w warstwie wejściowej, h jednostek w warstwie ukrytej oraz m jednostek wyjściowych. Jednostki tych warstw będziemy oznaczali odpowiednio symbolami X_i , Y_j , Z_k , gdzie dolny indeks wskazuje na konkretny neuron w danej warstwie. Niech

- v_{ij} oznacza wagę połączenia i -tej jednostki warstwy wejściowej z j -tą jednostką warstwy ukrytej i niech
- w_{jk} będzie wagą połączenia j -tej jednostki warstwy ukrytej z k -tą jednostką wyjściową.



Rysunek 6.1: Struktura sieci wielowarstwowej. Warstwa wejściowa składa się z n neuronów, warstwa ukryta – z u neuronów, a w warstwie wyjściowej jest pojedynczy neuron.

Proces uczenia (trenowania sieci) składa się z, powtarzanych do uzyskania zadowalających wyników, faz przesyłania sygnału wejściowego *w przód* do warstwy wyjściowej. Tam obliczana jest różnica między sygnałem zwracającym przez sieć, a sygnałem docelowym. Obliczony w ten sposób błąd propagowany jest *wstecz* do warstwy wejściowej.

Oba etapy wyglądają następująco (por. np. [20, pkt. 6.3]):

(a) faza *w przód*

(a1) Sygnał $\mathbf{x} = (x_1, \dots, x_n)$ przesyłany jest przez warstwę wejściową do warstwy

ukrytej. Każda jednostka z warstwy ukrytej oblicza wartość aktywacji

$$z_in_j = \sum_{i=0}^n x_i v_{ij} \quad (6.1)$$

a następnie wyznacza swój sygnał wyjściowy

$$z_j = \alpha_Z(z_in_j) \quad (6.2)$$

(a2) Jednostki warstwy wyjściowej zbierają sygnały od jednostek warstwy ukrytej

$$y_in_k = \sum_{j=0}^h z_j w_{jk} \quad (6.3)$$

i wyznaczają odpowiedź sieci

$$y_k = \alpha_Y(y_in_k) \quad (6.4)$$

(a3) Oblicza się błąd

$$E = \sum_{k=0}^m (t_k - y_k)^2 \quad (6.5)$$

(b) faza *wstecz*

(b1) Stosując twierdzenie o pochodnej funkcji złożonej,

$$\frac{d}{dx} f[\xi(x)] = \frac{df(\xi)}{d\xi} \frac{d\xi}{dx} = f'(\xi) \frac{d\xi}{dx}$$

wyznacza się pochodną cząstkową

$$\begin{aligned} \frac{\partial E}{\partial w_{jk}} &= -(t_k - y_k) \frac{\partial y_k}{\partial w_{jk}} = -(t_k - y_k) \frac{\partial}{\partial w_{jk}} \alpha_Y(y_in_k) \\ &= -(t_k - y_k) \alpha'_Y(y_in_k) \frac{\partial y_in_k}{\partial w_{jk}} \\ &= -(t_k - y_k) \alpha'_Y(y_in_k) z_j = -\delta_k z_j \end{aligned} \quad (6.6)$$

Tutaj symbolem δ_k oznaczono iloczyn

$$\delta_k = -(t_k - y_k) \alpha'_Y(y_in_k) \quad (6.7)$$

(b2) Wyznacza się pochodną cząstkową

$$\begin{aligned} \frac{\partial E}{\partial v_{ij}} &= - \sum_{k=1}^m (t_k - y_k) \frac{\partial y_k}{\partial v_{ij}} = - \sum_{k=1}^m (t_k - y_k) \frac{\partial}{\partial v_{ij}} \alpha_Y(y_in_k) \\ &= - \sum_{k=1}^m (t_k - y_k) \alpha'_Y(y_in_k) \frac{\partial y_in_k}{\partial v_{ij}} \\ &= - \sum_{k=1}^m \delta_k \frac{\partial y_in_k}{\partial v_{ij}} = - \sum_{k=1}^m \delta_k w_{jk} \frac{\partial z_j}{\partial v_{ij}} \\ &= - \sum_{k=1}^m \delta_k w_{jk} \alpha'_Z(z_in_j) x_i \end{aligned} \quad (6.8)$$

(b3) Dokonuje się synchronicznej (tzn. jednocześnie) modyfikacji wszystkich wag, tzn.

$$\begin{aligned}\Delta w_{jk} &= \eta \delta_k z_j \\ \Delta v_{ij} &= \eta \sum_{k=1}^m \delta_k w_{jk} \alpha'_Z(z_{in_j}) x_i\end{aligned}\tag{6.9}$$

UWAGA Analogicznie do sytuacji propagowania sygnału można myśleć o propagowaniu błędu od jednostek warstwy wyjściowej do neuronów leżących w coraz niższych warstwach. W przypadku rozważanej tu sieci trój-warstwowej błędy δ_k zbierane są przez jednostki warstwy ukrytej, tzn.

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk}\tag{6.10}$$

natomiast sygnał korygujący wagi v_{ij} ma postać

$$\delta_j = \delta_{in_j} \alpha'_Z(z_{in_j})\tag{6.11}$$

Ostatecznie drugie równanie (??) przyjmuje prostą postać

$$\Delta v_{ij} = \eta \delta_j x_i\tag{6.12}$$

Jest ono wprowadzie równokształtne z pierwszym równaniem, ale należy pamiętać, że δ_j jest zdecydowanie inne niż δ_k . \square

Przykład 6.1.1. Załóżmy, że warstwa ukryta sieci neuronowej zawiera tylko jeden neuron, tzn. $m = 1$, oraz, że α_Z i α_Y są bipolarnymi funkcjami typu 2, tzn.

$$\alpha_Z(\xi) = \alpha_Y(\xi) = \alpha(\xi) = \frac{2}{1 + \exp(-\xi)} + 1\tag{6.13}$$

Pamiętając, że (por. ćwiczenie 5.5.5)

$$\alpha'(\xi) = \frac{1}{2} [1 - \alpha^2(\xi)]$$

sprawdzamy, że (skoro $m = 1$, to opuszczamy indeks k)

$$\delta = \frac{1}{2} (t - y)(1 - y^2)$$

$$\delta_{in_j} = \delta w_j$$

$$\delta_j = \frac{1}{2} \delta_{in_j} (1 - y^2)$$

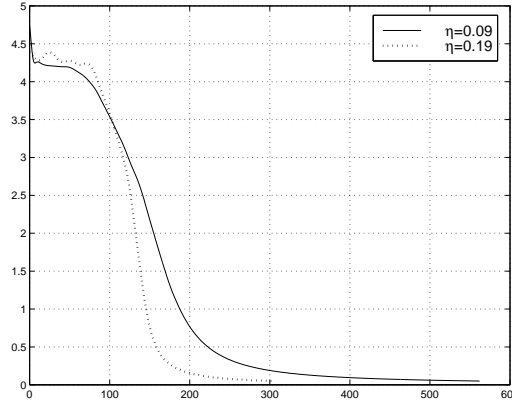
Zatem modyfikacja wag realizowana jest zgodnie z równaniami

$$\Delta w_j = \eta \delta z_j, \quad j = 0, 1, \dots, h$$

$$\Delta v_{ij} = \eta \delta_j x_i, \quad i = 0, 1, \dots, n$$

\square

Na zakończenie wspomnijmy o różnorodnych udoskonaleniach podstawowego algorytmu propagacji wstecznej: stochastycznym algorytmie BackProp [50] oraz algorytmie optymalizacyjnym 2-go rzędu QuickProp [19].



Rysunek 6.2: Uczenie funkcji XOR. Na osi odciętych zaznaczono epoki, a na osi rzędnych wartość błędu

6.2 Radialne sieci neuronowe

W przypadku radialnych sieci neuronowych zasadniczy proces obliczeniowy realizowany jest przez jednostki warstwy ukrytej. Innymi słowy zakłada się, że wagi $v_{ij} = 1$ dla wszystkich $i = 1, \dots, p$ oraz $j = 1, \dots, h$. Działanie sieci opiera się na założeniu, że dane tworzące zbiór trenujący posiadają wewnętrzną strukturę manifestującą się występowaniem skupień. Jeżeli znane są środki owych skupień, można lokować w nich tzw. jednostki radialne. Odległości przykładów od owych środków poddawane są nieliniowemu przekształceniu, w którym obiektom znajdującym się w pobliżu centrum¹ przypisuje się wartość (podobieństwa) bliską 1, natomiast obiektom oddalonym od centrum wartości bliskie 0.

Zakłada się dalej, że obszar otaczający środek skupienia ma charakter izotropowy, a funkcja realizująca wspomniane wyżej nieliniowe odwzorowanie nazywana jest radialną funkcją bazową. Najbardziej popularnym przykładem funkcji radialnej jest tzw. gaussowska funkcja radialna postaci

$$g_j(\mathbf{x}) = g(\mathbf{x}; \mathbf{c}_j, \sigma_j) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{2\sigma_j^2}\right) \quad (6.14)$$

gdzie g to tzw. funkcja bazowa, a \mathbf{c}_j i σ_j to jej parametry: środek i dyspersja.

W opisie funkcji radialnych posługujemy się zazwyczaj normowanym promieniem

$$r = \frac{\|x - c_j\|}{\sigma_j} \quad (6.15)$$

Wówczas gaussowską funkcję radialną zapisuje się krótko jako $g(r) = \exp(-r^2/2)$. Inne, równie często stosowane funkcje radialne to $g(r) = \sqrt{r^2 + 1}$ lub $g(r) = 1/\sqrt{r^2 + 1}$.

W przypadku sieci z pojedynczą jednostką wyjściową, odpowiedź na impuls \mathbf{x} ma postać

$$y(\mathbf{x}) = \sum_{j=1}^h w_j g_j(\mathbf{x}) = \mathbf{w}^T \mathbf{g}(\mathbf{x}) \quad (6.16)$$

gdzie $\mathbf{w}^T = (w_1, \dots, w_h)$, $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_h(\mathbf{x}))^T$.

Problem uczenia sieci radialnej polega na wyborze parametrów (\mathbf{c}_j, σ_j) oraz na wyborze parametrów w_j zapewniających minimalizację wskaźnika

¹A więc obiektom „typowym” dla danego skupienia.

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^p \left(t_i - y(\mathbf{x}_i) \right)^2 \quad (6.17)$$

Poniżej omawiamy te problemy. Czytelnika zainteresowanego szczegółami odsyłamy do raportów [60] i [61].

6.2.1 Uczenie wag

Niech

$$\mathbf{G} = \begin{bmatrix} g_1(\mathbf{x}_1) & g_2(\mathbf{x}_1) & \dots & g_h(\mathbf{x}_1) \\ g_1(\mathbf{x}_2) & g_2(\mathbf{x}_2) & \dots & g_h(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\mathbf{x}_p) & g_2(\mathbf{x}_p) & \dots & g_h(\mathbf{x}_p) \end{bmatrix} \quad (6.18)$$

oznacza macierz wyjść jednostek warstwy ukrytej. Jej i -ty wiersz reprezentuje zbiór sygnałów przesyłanych przez warstwę ukrytą do jednostki wyjściowej po prezentacji i -tego przykładu, natomiast j -ta kolumna zawiera wartości sygnałów emitowanych przez j -tą jednostkę tej warstwy gdy prezentowane są kolejne sygnały $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$. Niech dalej

$$\mathbf{t} = (t_1, \dots, t_p)^T \quad (6.19)$$

będzie wektorem wartości docelowych ze zbioru \mathcal{L} .

Przy powyższych oznaczeniach wskaźnik jakości (6.17) można zapisać w równoważnej postaci

$$E(\mathbf{w}) = \frac{1}{2} \|\mathbf{G}\mathbf{w} - \mathbf{t}\|^2 \quad (6.20)$$

gdzie $\|\cdot\|$ oznacza normę euklidesową, tzn. $\|\boldsymbol{\xi}\|^2 = \sum_{i=1}^p \xi_i^2$.

Jeżeli $h = p$, tzn. za centra funkcji radialnych przyjmiemy wszystkie elementy zbioru trenującego, to \mathbf{G} staje się macierzą kwadratową. Jeżeli ponadto \mathbf{G} jest macierzą nieosobliwą², to wektor wagowy minimalizujący powyższy wskaźnik wyznaczamy z równania

$$\mathbf{w} = \mathbf{G}^{-1}\mathbf{t} \quad (6.21)$$

Jednak przyjęcie za $h = p$ niesie za sobą szereg niedogodności. Przede wszystkim nie zawsze \mathbf{G} jest macierzą nieosobliwą. Po drugie, wektor wag jest zbyt silnie dopasowany do danych, co osłabia zdolności generalizacyjne sieci. Ponadto ze wzrostem liczby jednostek ukrytych rośnie prawdopodobieństwo złego uwarunkowania macierzy \mathbf{G} z uwagi na rosnące niebezpieczeństwo redundancji danych. Na ogół dąży się do tego, aby $h < p$. W takim przypadku \mathbf{G} jest macierzą prostokątną i do wyznaczenia wektora \mathbf{w} stosujemy równanie

$$\mathbf{w}^* = \mathbf{G}^\dagger \mathbf{t} \quad (6.22)$$

gdzie symbol \mathbf{G}^\dagger oznacza tzw. pseudoinwersję, albo odwrotność Moore-Penrose'a². Spełnia ona cztery równania Penrose'a [52, s. 44]

²Formalnie nieosobliwość macierzy gwarantuje umieszczenie w zbiorze trenującym p różnych punktów – por. C.A. Micchelli, "Interpolation of scattered data: Distance matrices and conditionally positive definite functions". *Const. Approx.*, **2**, 1986, 11-22. Jednak w praktyce współczynnik uwarunkowania takiej macierzy może być bardzo duży, co wynika z „wyglądających” własności sigmoidalnych funkcji aktywacji.

$$\begin{aligned} \mathbf{G}\mathbf{G}^\dagger\mathbf{G} &= \mathbf{G}, & \mathbf{G}^\dagger\mathbf{G}\mathbf{G}^\dagger &= \mathbf{G}^\dagger \\ (\mathbf{G}\mathbf{G}^\dagger)^\mathrm{T} &= \mathbf{G}\mathbf{G}^\dagger, & (\mathbf{G}^\dagger\mathbf{G})^\mathrm{T} &= \mathbf{G}^\dagger\mathbf{G} \end{aligned} \quad (6.23)$$

Macierz \mathbf{G}^\dagger wyznacza się następująco

$$\mathbf{G}^\dagger = \begin{cases} (\mathbf{G}^\mathrm{T}\mathbf{G})^{-1}\mathbf{G}^\mathrm{T} & \text{gdy } \mathbf{G}^\mathrm{T}\mathbf{G} \text{ jest macierzą nieosobliwą} \\ \mathbf{G}^\mathrm{T}(\mathbf{G}\mathbf{G}^\mathrm{T})^{-1} & \text{gdy } \mathbf{G}\mathbf{G}^\mathrm{T} \text{ jest macierzą nieosobliwą} \end{cases} \quad (6.24)$$

W praktyce okazuje się, że nawet jeżeli $\mathbf{A} = \mathbf{G}^\mathrm{T}\mathbf{G}$ (czy też $\mathbf{A} = \mathbf{G}\mathbf{G}^\mathrm{T}$) jest macierzą nieosobliwą, to jej wyznacznik jest bardzo bliski zera lub równoważnie, jej współczynnik uwarunkowania $\text{cond}(\mathbf{A})$ jest bardzo duży³. A to dramatycznie redukuje stabilność procesu obliczeniowego, co skutkuje wektorem \mathbf{w} o składowych odbiegających od wartości teoretycznych. Dlatego wprowadza się tzw. regularyzację polegającą na zastąpieniu macierzy \mathbf{A} przez macierz regularyzowaną⁴

$$\mathbf{A}_r = \mathbf{A} + \lambda\mathbf{I} \quad (6.25)$$

gdzie \mathbf{I} jest macierzą jednostkową, a $\lambda > 0$ to tzw. współczynnik regularyzacji.

Należy też odnotować⁵, że jeżeli $p < h$, tzn. liczba wierszy macierzy \mathbf{G} jest mniejsza od liczby kolumn, to macierz $\mathbf{G}^\mathrm{T}\mathbf{G}$ jest macierzą osobliwą. Jeżeli natomiast $p > h$ to osobliwą jest macierz $\mathbf{G}\mathbf{G}^\mathrm{T}$. Ostatecznie można przyjąć, że

$$\mathbf{w}^* = \mathbf{G}^\dagger \mathbf{t} = \begin{cases} (\mathbf{G}^\mathrm{T}\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^\mathrm{T}\mathbf{t} & \text{gdy } p > h \\ \mathbf{G}^\mathrm{T}(\mathbf{G}\mathbf{G}^\mathrm{T} + \lambda\mathbf{I})^{-1}\mathbf{t} & \text{gdy } p < h \end{cases} \quad (6.26)$$

W teorii równań liniowych udowadnia się, że wyznaczony w powyższy sposób wektor \mathbf{w}^* spełnia warunek

$$\|\mathbf{G}\mathbf{w}^* - \mathbf{t}\| = \min_{\mathbf{w} \in \mathbb{R}^h} \|\mathbf{G}\mathbf{w} - \mathbf{t}\| \quad (6.27)$$

W szczególności, gdy stosuje się regularyzację, to wektor \mathbf{w}^* zapewnia minimalizację wyrażenia (por. np. Dodatek 4 raportu [60])

$$\|\mathbf{G}\mathbf{w} - \mathbf{t}\|^2 + \lambda\|\mathbf{w}\|^2$$

Co więcej jest to rozwiązanie o minimalnej normie, tzn. dla dowolnego innego wektora \mathbf{w}^{**} minimalizującego wskaźnik E spełniony jest warunek $\|\mathbf{w}^*\| \leq \|\mathbf{w}^{**}\|$.

6.2.2 Wybór centrów

Jeżeli p jest liczbą najbliższych sąsiadów centrum c_j , to parametr σ_j wybieramy jako

$$\sigma_j = \frac{1}{p} \sqrt{\sum_{i=1}^h \|\mathbf{c}_j - \mathbf{c}_i\|^2}$$

³Współczynnik uwarunkowania informuje nas o tym jak zaburzenie względne danych przenosi się na błąd względny wyniku rozwiązania x^* układu równań liniowych $Ax = b$. Szybko informację na ten temat można znaleźć na stronie <http://wazniak.mimuw.edu.pl/index.php?title=MN07>, ewentualnie w [52] lub książkach poświęconych analizie numerycznej.

⁴Składnik $\lambda\mathbf{I}$ poprawia współczynnik uwarunkowania $\text{cond}(\mathbf{A})$. Na ogół λ jest liczba bliska zeru.

⁵Por. np. rozdział 5 w książce C. Moler. *Numerical Computing with MATLAB*, SIAM, 2008. Jej starsza wersja elektroniczna dostępna pod adresem <http://www.mathworks.com/moler/chapters.html>.

Algorytm 6.1 Algorytm k -średnich dla lokalizacji centrów sieci radialnej

Require: zbiór trenujący, współczynnik uczenia $\eta \in [0, 1]$.**Ensure:** centra funkcji radialnych.

- 1: (Inicjalizacja) Inicjuj k centrów.
 `done = false;`
 - 2: **while** (not done) **do**
 - 3: wybierz losowo element $\mathbf{x} \in \mathcal{L}$
 - 4: znajdź centrum $\mathbf{c}_{\mathbf{x}}$ najbliższe obiektowi \mathbf{x}
 - 5: aktualizuj lokalizację wybranego centrum zgodnie z równaniem $\mathbf{c}_{\mathbf{x}} = \mathbf{c}_{\mathbf{x}} + \eta(\mathbf{x} - \mathbf{c}_{\mathbf{x}})$
 - 6: jeżeli lokalizacje zmieniają się nieznacznie, podstaw `done = true`
 - 7: **end while**
-

Według innej propozycji współczynnik dyspersji powinien być identyczny dla wszystkich funkcji radialnych i równy

$$\sigma = \frac{d_{max}}{\sqrt{2h}}$$

gdzie d_{max} to maksymalna odległość między centrami. Taki wybór wartości σ jest szczególnie uzasadniony gdy jednostki ukryte pokrywają równomiernie dziedzinę aproksymowanej funkcji.

6.3 Sieci wielowarstwowe jako uniwersalne aproksymatory

Problem aproksymacji formułuje się zazwyczaj następująco. Niech \mathcal{K} będzie zwartym podzbiorem \mathbb{R}^n . Typowym przykładem zbioru \mathcal{K} jest n -wymiarowa kostka $I = [0, 1]^n$ lub ogólniej iloczyn kartezjański domkniętych przedziałów o skończonej długości

$$\mathcal{K} = [a_1, b_1] \times \cdots \times [a_n, b_n]$$

Niech dalej $C(\mathcal{K}, \mathbb{R})$ oznacza zbiór ciągłych funkcji $f: \mathcal{K} \rightarrow \mathbb{R}$. Poszukuje się takiego podzbioru $\mathcal{A} \subset C(\mathcal{K}, \mathbb{R})$, że dla dowolnej funkcji $f \in \mathcal{A}$ można znaleźć funkcję $\tilde{f} \in \mathcal{A}$ taką, że

$$\|f - \tilde{f}\| < \epsilon \tag{6.28}$$

gdzie $\|\cdot\|$ jest pewną normą, zazwyczaj

$$\|f\| = \sup_{\mathbf{x} \in \mathcal{A}} |f(\mathbf{x})| \tag{6.29}$$

Jako źródło tak postawionego problemu wskazuje się trzynaste, spośród 26 zagadnień przedstawionych przez Davida Hilberta na Międzynarodowym Kongresie Matematyków w Paryżu w 1900 w jego słynnym odczycie „Mathematische Probleme”⁶. Mianowicie Hilbert stwierdza tam, rozwiązania równania wielomianowego siódmego stopnia, $x^7 + ax^3 + bx^2 + cx + 1 = 0$, które można potraktować jako funkcję trzech zmiennych $x^*(a, b, c)$ nie mogą być

⁶Ich oryginalne sformułowanie można znaleźć np. na stronie <http://www.mathematik.uni-bielefeld.de/~kersten/hilbert/rede.html>. Ciekawe omówienie tych zagadnień podaje Ivor Grattan-Guinness w swoim artykule „A Sideways Look at Hilbert’s Twenty-three Problems of 1900”, który ukazał się w *Notices of the AMS*, vol. 47, no. 7, 2000, pp. 752-757. Artykuł ten jest dostępny pod adresem <http://www.ams.org/notices/200007/fea-grattan.pdf>.

przedstawione w postaci sumy i superpozycji skończonej liczby funkcji dwóch zmiennych. Problem ten rozwiązał w 1957 roku Kołmogorow. Jego twierdzenie o superpozycji głosi, że dowolną ciągłą funkcję n zmiennych $f: [0, 1]^n \rightarrow \mathbb{R}$ można przedstawić w postaci sumy [43]

$$f(x_1, \dots, x_n) = \sum_{j=1}^{2n+1} \chi_j \left(\sum_{i=1}^n \varphi_{ij}(x_i) \right) \quad (6.30)$$

gdzie χ_j oraz φ_{ij} są ciągłymi funkcjami jednej zmiennej, a ponadto φ_{ij} są monotonicznymi funkcjami, które nie zależą od f .

W terminologii sieci neuronowych powyższe równanie można interpretować następująco: każda spośród $2n+1$ jednostek warstwy ukrytej zasilana jest przez sumę n nieliniowych sygnałów emitowanych przez jednostki warstwy wejściowej (każdy sygnał jest nieliniową transformacją innej cechy X_i). Każda jednostka ukryta emituje sygnał χ_j , a jednostka wyjściowa akumuluje owe sygnały. Na fakt ten, jako pierwszy, zwrócił uwagę Hecht-Nielsen w artykule [28]. Mianowicie, jeżeli przez $y(\mathbf{x})$ oznaczmy wartość zwracaną przez wielowarstwową sieć neuronową po prezentacji sygnału $\mathbf{x} = (x_1, \dots, x_n)$, to z twierdzenia Kołmogorowa wynika, że dla dowolnej ciągłej funkcji f będzie spełniony warunek

$$\sum_{\mathbf{x} \in \mathcal{K}} |f(\mathbf{x}) - y(\mathbf{x})| < \epsilon \quad (6.31)$$

Jak odnotowano w [18] bezpośrednie zastosowanie twierdzenia Kołmogorowa jest w praktyce nieco skomplikowane. Przede wszystkim funkcje χ_j oraz φ_{ij} znacząco różnią się od ważonych sum stosowanych w opisie sieci neuronowych. Mają one zdecydowanie bardziej skomplikowaną postać. Ponadto twierdzenie Kołmogorowa jest twierdzeniem egzystencjalnym, co oznacza, że z faktu istnienia aproksymacji f nie wynikają żadne wskazówki pozwalające zidentyfikować funkcje χ_j oraz φ_{ij} dla zadanego zbioru uczącego.

Przegląd różnorodnych podejść do efektywnego wykorzystania twierdzenia Kołmogorowa zawiera praca [48]. Odnotujmy jako ciekawostkę, że o ile druga część tytułu pracy [24] z 1989 roku brzmi *Kolmogorov's theorem is irrelevant*⁷, to w dwa lata później w tym samym czasopiśmie opublikowano pracę pod tytułem *Kolmogorov's theorem is relevant*, [47]. Neruda, Stedry i Drkosová przedstawili w [57] algorytm pozwalający rekonstruować ciągłą funkcję n zmiennych drogą superpozycji ciągłych funkcji jednej zmiennej. Korzystali oni z wyników Sprechera [77] oraz [78].

Ważny z praktycznego punktu widzenia wynik podał Hornik [32] stwierdzając, że jeżeli α jest ciągłą, ograniczoną i różną od stałej funkcją aktywacji, to dowolna ciągła funkcja o zwartej dziedzinie może być aproksymowana przez sieć neuronową z pojedynczą warstwą ukrytą. Leshno *i in.* [51] wykazali dodatkowo, że taka aproksymacja jest możliwa o ile α jest funkcją różną od wielomianu. Podobny rezultat dla funkcji radialnych uzyskali Park i Sandberg [62]. Pokazali oni, że jeżeli jednostki radialne o identycznej wartości dyspersji σ zostaną właściwie rozmieszczone wewnątrz dziedziny aproksymowanej funkcji, to funkcje te można aproksymować z dowolną dokładnością.

Zakończmy ten punkt precyzyjnym sformułowaniem twierdzenia, którego autorem jest Funahashi [22].

Twierdzenie 6.3.1. Niech $\alpha(\xi)$ będzie ograniczoną i monotonicznie rosnącą ciągłą funkcją. Niech $\mathcal{K} \subset \mathbb{R}^n$ będzie zwartym zbiorem i niech $f: \mathcal{K} \rightarrow \mathbb{R}$ będzie ciągłą funkcją. Wówczas,

⁷Autorzy tej pracy powołują się m.in. na artykuł G.G. Lorentza p.t. "Metric entropy, widths, and superposition of functions", który ukazał się w *Am. Math. Monthly*, **69**, 1962, 469-485. Lorentz pyta w nim (mając na myśli twierdzenie Kołmogorowa) "Will it have useful applications? ... One wonders whether Kolmogorov's theorem can be used to obtain positive results of greater [than trivial] depth."

dla dowolnego $\epsilon > 0$ i rzeczywistych stałych ω_i , w_{ij} takich, że

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^u \omega_i \alpha \left(\sum_{j=1}^n w_{ij} x_j \right)$$

spełniony jest warunek

$$\|f - \tilde{f}\|_{\infty} = \sup_{x \in \mathcal{K}} |f(x) - \tilde{f}(x)| \leq \epsilon$$

□

Innymi słowy dowolna ciągła funkcja może być aproksymowana dostatecznie dokładnie przez trój-warstwową sieć neuronową, w której jednostki warstwy ukrytej emitują sygnały postaci $\alpha[f(x)]$, funkcja aktywacji jednostki wyjściowej jest liniowa.

6.4 Zadania

Problem 6.4.1. Sprawdź, że jeżeli \mathbf{G} jest nieosobliwą macierzą kwadratową, to oba warianty równania (6.24) prowadzą do wyniku $\mathbf{G}^{\dagger} = \mathbf{G}^{-1}$.

Problem 6.4.2. Sprawdź, że jeżeli do wyznaczenia uogólnionej macierzy odwrotnej \mathbf{G}^{\dagger} wykorzystano pierwsze z równań (6.24), to spełniony jest warunek $\mathbf{G}^{\dagger} \mathbf{G} = \mathbf{I}$, czyli \mathbf{G}^{\dagger} jest lewostronną odwrotnością macierzy \mathbf{G} . Jeżeli natomiast \mathbf{G}^{\dagger} wyznaczono za pomocą drugiego z równań (6.24), to jest to prawostronna odwrotność macierzy \mathbf{G} , tzn. $\mathbf{G} \mathbf{G}^{\dagger} = \mathbf{I}$. Symbolem \mathbf{I} oznaczono macierz jednostkową.

Problem 6.4.3. Rozwiąż równanie

$$\mathbf{G} = \begin{bmatrix} 3 & 7 & 4 \\ 11 & 2 & 9 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 25 \\ 51 \end{bmatrix} \quad (6.32)$$

Porównaj wynik uzyskany dla obu wariantów macierzy \mathbf{G}^{\dagger} zdefiniowanych w równaniu (6.24). Jak zmieniają się wyniki jeżeli macierz $\mathbf{G}^T \mathbf{G}$ zastąpimy macierzą $\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I}$, gdzie $\lambda = 0.0001$?

Zauważ, że innym rozwiązaniem tego równania jest wektor $\mathbf{w} = (2, 1, 3)^T$. Porównaj normę tego wektora z normą rozwiązania \mathbf{w}^* wyznaczonego przez zastosowanie procedury (6.26).

Problem 6.4.4. Sprawdź jakość aproksymacji podanych niżej funkcji dwóch zmiennych⁸ z użyciem metody propagacji wstecznej oraz z metodą uczenia wag radialnej sieci neuronowej

⁸Por. np. J.N. Hwang, S.R. Lay, M. Maechler, D. Martin, J. Schimert. "Regression modeling in backpropagation and pursuit learning", *IEEE Trans. Neural Networks*, **5**, 1994, 342-353.

$$\begin{aligned}
f_1(x_1, x_2) &= 10.391 \left((x_1 - 0.4)(x_2 - 0.6) + 0.36 \right) \\
f_2(x_1, x_2) &= 24.234 \left(r^2(0.75 - r^2) \right), \\
&\quad r^2 = (x_1 - 0.4)^2 + (x_2 - 0.5)^2 \\
f_3(x_1, x_2) &= 42.659 \left(0.1 + \xi_1(0.05 + \xi_1^4 - 10\xi_1^2\xi_2^2 + 5\xi_2^4) \right), \\
&\quad \xi_1 = x_1 - 0.5, \xi_2 = x_2 - 0.5 \\
f_4(x_1, x_2) &= 1.3356 \left[1.5(1 - x_1) + e^{2x_1-1} \sin 3\pi(x_1 - 0.6)^2 + e^{3x_2-0.5} \sin 4\pi(x_2 - 0.9)^2 \right. \\
f_5(x_1, x_2) &= 1.9 \left[1.35 + e^{x_1} \sin 13(x_1 - 0.6)^2 \right] e^{-x_2} \sin 7x_2
\end{aligned} \tag{6.33}$$

Wszystkie funkcje określone są na zbiorze $[0, 1] \times [0, 1]$. Sporządź wykresy wszystkich funkcji. Jako zbiór trenujący wygeneruj losowo (zgodnie z rozkładem jednostajnym) $p = 225$ punktów ze zbioru $[0, 1] \times [0, 1]$. Dla każdego z tak otrzymanych punktów (x_1, x_2) wyznacz wartość $t = f_i(x_1, x_2)$. Utwórz zbiór testowy złożony z 10000 punktów będących węzłami regularnej kraty rozpiętej na dziedzinie $[0, 1] \times [0, 1]$. Powtórz eksperymenty z zaszumionym zbiorem trenującym otrzymanym z \mathcal{L} przez dodanie do każdej wartości t liczby gaussowskiej pochodzącej z rozkładu normalnego o zerowej wartości średniej i odchyleniu standardowym $\sigma = 0.25$.

Rozdział 7

Sieci samoorganizujące się

Rozdział 8

Inne sieci neuronowe

Rozdział 9

Systemy neuro-rozmyte

Wiedza zawarta w bazie wiedzy sterownika rozmytego ma postać zdań warunkowych zapisanych w języku bliskim naturalnemu. Baza wiedzy nie stanowi jednak narzędzia umożliwiającego pozyskiwanie takiej wiedzy. Z drugiej strony, sieci neuronowe stanowią narzędzie pozwalające na uczenie się z przykładów. Jednakże wiedza reprezentowana przez zbiór wag przypisanych połączeniom jednostek tworzących daną sieć ma zdecydowanie nieintuicyjny charakter. Połączeniem obu formalizmów są systemy neuro-rozmyte posiadające zarówno zdolności uczenia się jak też reprezentujące wiedzę w sposób przyjazny dla użytkownika.

Rozdział 10

Projekty

Poniżej przedstawiamy trzy propozycje poważniejszych problemów (nazywanych dalej projektami), które Czytelnik powinien umieć rozwiązać. Sekcja ?? zawiera dwa zadania o zbliżonym stopniu złożoności ilustrujące zastosowanie zbiorów rozmytych w zagadnieniach prognozowania i aproksymacji funkcji. W sekcji ?? przedstawiono dwa tematy ilustrujące typowe zastosowania sztucznych sieci neuronowych, a w sekcji ?? – tematy ilustrujące cechy systemów neuro-rozmytych. Z każdej grupy tematycznej należy wybrać **jedno** zadanie.

Sposób dokumentowania prowadzonych prac opisano w poniższym podrozdziale.

10.1 Jak przygotowywać raporty dokumentujące projekty?

10.1.1 Część opisowa

Składa się ona z sześciu podstawowych bloków, które omawiamy poniżej.

10.1.1.1 Streszczenie

Obejmuje ono zazwyczaj jeden lub dwa akapity. Rolą streszczenia jest zainteresowanie potencjalnego czytelnika i zachęcenie go do przeczytania całej pracy. Powinno ono zawierać zwięzłe przedstawienie treści zawartych w raporcie. Pisząc je należy pamiętać, że jest adresowane do osób mających ogólną wiedzę o dziedzinie, ale też i do ekspertów. Należy więc krótko objaśnić problem, przedstawić metodę rozwiązania i scharakteryzować osiągnięte wyniki. Pisząc streszczenie, nie używamy skrótów, symboli specjalnych, itp.

10.1.1.2 Wprowadzenie

Kogoś zainteresowało przeczytane/przejrzane przed chwilą streszczenie. Aby zachęcić czytelnika do dalszej lektury należy uwzględnić następujące aspekty:

- (a.) **Motywacja.** Na czym polega rozważany w raporcie problem i dlaczego jest on interesujący (potencjalne zastosowania). Czy inni też uznają go za interesujący? Dlaczego w pracy zastosowano taką właśnie, a nie inną, metodę rozwiązania? („*Bo nie znam innej*” to raczej słaba odpowiedź)
- (b.) **Background.** Kto dotychczas zajmował się tym problemem? (odnośniki do literatury).
- (c.) Krótki opis zastosowanej metodologii. trzeba pamiętać, że czytelnikiem może być osoba nieznająca dokładnie technik stosowanych w pracy. Należy poświęcić chociaż jeden

akapit objaśniający proponowaną/stosowaną metodologię ze wskazaniem potencjalnie innych jej zastosowań.

- (d.) Krótkie przedstawienie uzyskanych wyników. (Nie pisz jak znakomite są one, ale po prostu omów je).
- (e.) Krótkie omówienie zawartości reszty pracy, tzn. co znajduje się w jej kolejnych częściach wraz z logicznym uzasadnieniem takiego właśnie układu pracy.

Celem *Wprowadzenia* jest wyjaśnienie czytelnikowi, w jaki sposób uzyskano raportowane wyniki. To powinno zachęcić go do czytania dalszych części pracy, a jeśli jest specjalistą – do przejrzenia wybranych jej fragmentów (z pominięciem fragmentów dla niego oczywistych; aby mógł on omijać takie fragmenty, musi jednak z góry wiedzieć, gdzie ulokowano je w pracy). Zazwyczaj *Wprowadzenie* składa się z kilku akapitów.

10.1.1.3 Metodologia

Sekcja ta przeznaczona jest głównie dla czytelników, którzy nie mają dostatecznej wiedzy w tej szczegółowej dziedzinie. Z tego powodu może być ona rozbita na kilka pod-sekcji.

Przyda się zapewne (pod-)sekcja dotycząca przyjętej reprezentacji problemu, ew. ogólnego schematu stosowanego algorytmu. W szczególności:

1. Jeżeli Twój algorytm jest specjalizacją pewnego ogólnego przepisu (algorytmu), należy ten przepis dokładnie objaśnić ze wskazaniem podstaw teoretycznych.
2. Należy dokładnie wyjaśnić istotę rozwiązywanego problemu i omówić sposób wykorzystania ogólnej metodologii z poprzedniego punktu.
3. Należy wskazać, na czym polega własne rozwiązanie, tzn. omówić zastosowany sposób kodowania (reprezentacja) zadania, wykorzystane operatory, ew. heurystyki czy też własne usprawnienia ze *starannym i przemyślanym* wskazaniem intencji przyświecających ich użyciu.

10.1.1.4 Wyniki i ich analiza

Ta część pracy powinna zawierać dużo wykresów i tabel. Wykresy powinny być estetyczne. Należy opisywać osie (a więc: co reprezentuje oś odciętych, a co oś rzędnych). Rekomendowane narzędzie do robienia wykresów to **gnuplot** dostępny na stronie <http://www.gnuplot.info/>¹.

Sekcja ta powinna składać się z jednej lub dwóch pod-sekcji. Na początku należy zaznaczyć:

- co zrobiono?,
- dlaczego tak właśnie postąpiono?,
- czy uzyskane wyniki są zgodne z oczekiwaniami?
- dlaczego ? (w szczególności, jeśli wyniki nie są zgodne z oczekiwaniami, należy starannie przeanalizować źródła porażki).

¹Przykłady jego użycia (wraz ze stosownymi skryptami) znajdują się na stronie <http://gnuplot.sourceforge.net/demo/>, a szereg tutoriali dostępnych jest na stronie <http://www.gnuplot.info/help.html>.

Typowy sposób pisania jest zgodny z formatem: „*Idea skłaniająca mnie do użycia opisanej wcześniej metodologii, była stosowana z nadzieją uzyskania wyniku X. Kiedy ją zaimplementowałem, uzyskałem efekty Y, a to z kolei zainspirowało mnie do modyfikacji Z, których skutkiem jest (...)*”.

Pożądane jest porównanie jakości uzyskanych wyników z wynikami uzyskiwanymi innymi, popularnymi, metodami.

10.1.1.5 Konkluzje i plany dalszych badań

Należy krótko podsumować uzyskane wyniki i wskazać ich konsekwencje, tzn. co z wykonanej pracy wynika dla danej dziedziny. Jeżeli coś wynika, opisać jak podejść do owych implikacji. Mogą to być zarówno sformalizowane pomysły jak również pewne idee, czy „intuicje”.

Pamiętaj jednak, że ktoś, kto dotarł do tego miejsca, przeczytał wcześniej przynajmniej streszczenie i sekcję z wynikami. Pisząc konkluzje dbaj o spójność logiczną pracy, aby przypadkiem nie obniżyć jej jakości nieprzemyślanymi wnioskami i konkluzjami mającymi niewiele wspólnego z wstępnymi deklaracjami. Wbrew pozorom jest to bardzo istotna część pracy! Pisz ją starannie i z rozmysłem!

10.1.1.6 Bibliografia

Powinna być także **przemyślana**! Musi zawierać zarówno pozycje o fundamentalnym znaczeniu dla danej dziedziny (przeznaczone np. dla osób, które chciałyby się czegoś dowiedzieć) jak i pozycje specjalistyczne (przeznaczone dla ekspertów), które stały się inspiracją dla Twojej pracy. Nie cytuj jednak pozycji, które znasz tylko ze słyszenia.

Bibliografię przygotowujemy wg. następującego schematu:

1. książki:

Rutkowska, D., Piliński, M. Rutkowski, L. *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. PWN, Warszawa 1997

Yan, J., Ryan, M., Power, J. *Using Fuzzy Logic*. Prentice Hall, London, 1994

2. artykuły w czasopismach: Wang, L.X., Mendel, J.M. Generating fuzzy rules by learning from examples. *IEEE Trans. on Systems, Man and Cybernetics*, **22**, 1992, 1414-1427

3. rozdziały w książkach, artykuły w materiałach konferencyjnych:

Altenberg, L. The evolution of evolvability in genetic programming. W: K. Kinnear, ed., *Advances in Genetic Programming*. MIT Press, Cambridge, MA, 1994, pp. 47-74

Angeline, P. Evolutionary optimization versus particle swarm optimization: Philosophy and performance difference. In: *Proc. 7-th Annual Conference on Evolutionary programming*, LNCS 1447, Springer-Verlag, 1998, pp. 601-610

4. prace doktorskie:

Guo, H. Algorithm selection for sorting and probabilistic inference: A machine learning-based approach. Ph. D. Thesis, Department of Computing and Information Sciences, Kansas State University, 2003

5. raporty:

Fraley, C., Raftery, A. MCLUST: Software for model-based cluster and discriminant analysis. Technical Report 342, Dept. Statistics, University of Washington, 1999

10.1.2 Część programowa

Program, przy użyciu którego uzyskano wyniki omawiane w części opisowej powinien posiadać krótką dokumentację uruchomieniową.

Na początku programu należy umieścić nazwiska jego autorów wraz z krótką informacją o jego przeznaczeniu. Ponadto zmienne globalne używane w programie powinny być opisane, a również każda metoda winna być poprzedzona komentarzem omawiającym jej rolę, parametry wejściowe oraz zwracany wynik. Pożądany język programowania: **Java**.

Bibliografia

- [1] Abbasbandy, S., Alavi, A. A method for solving fuzzy linear systems. *Iranian Journal of Fuzzy Systems*, **2**(2), 2005, 37-43
- [2] Abe, S., Lan, M.-S. Fuzzy rules extraction directly from numerical data for function approximation. *IEEE Trans. on Systems, Man and Cybernetics*, **25**(1), 1995, 119-129. URL: <http://www.lib.kobe-u.ac.jp/repository/90000209.pdf>
- [3] Baum, E.B., Haussler, D. What size net gives valid generalization? *Neural Computation*, **1**(1), 1989, 151-160
- [4] Buckley, J.J., Eslami, E. *Introduction to Fuzzy Logic and Fuzzy Sets*, Physica-Verlag, Heidelberg, Germany. 2001
- [5] Buckley, J.J., Feuring, T. Fuzzy differential equations. *Fuzzy Sets and Systems*, **110**, 2000, 43-54
- [6] Buckley, J.J., Feuring, T., Hayashi, Y. Solving fuzzy equations using evolutionary algorithms and neural nets. *Soft Computing*, **6**, 2002, 116-123
- [7] Castro, J. L. Delgado, M., Herrera, F. A learning method of fuzzy reasoning by genetic algorithms, pp. 804-809, URL: citeseer.ist.psu.edu/article/castro93learning.html
- [8] Chalco-Cano, Y., Román-Flores, H., Rojas-Medar, M.A., Saavedra, O.R., Jiménez-Gamero, M. The extension principle and a decomposition of fuzzy sets. *Information Sciences*, **177**, 2007, 5394-5403
- [9] Chalco-Cano, Y., Jiménez-Gamero, M., Román-Flores, H., Rojas-Medar, M.A. An approximation to the extension principle using decomposition of fuzzy intervals. *Fuzzy Sets and Systems*, **159**, 2008, 3245-3258
- [10] Chang, S.L., Zadeh, L.A. On fuzzy mapping and control. *IEEE Trans. Systems Man Cybernet.*, **2**, 1972, 30-34
- [11] Cristianini, N., Shwe-Taylor, J. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000
- [12] Crowder, R.S., Predicting the Mackey-Glass timeseries with cascade-correlation learning. In: D. Touretzky, G. Hinton, and T. Sejnowski (eds), *Proc. of the 1990 Connectionist Models summer School*. Carnegie Mellon University, 1990, pp. 117-123
- [13] Czogała, E., Łęski, J., *Fuzzy and Neuro-fuzzy Intelligent Systems*. Physica-Verlag, A Springer-Verlag Company, Heidelberg, New York, 1999

- [14] Diamond, P., Kloeden, P. Metric spaces of fuzzy sets. *Fuzzy Sets and Systems*, **100**, Supplement 1, 1999, 63-71 (także: Diamond, P., Kloeden, P. *Metric Spaces of Fuzzy Sets*, World Scientific, Singapore 1994)
- [15] Dong, W., Shah, H.C. Vertex method for computing functions of fuzzy variables. *Fuzzy Sets and Systems*, **24**, 1987, 65-78
- [16] Dubois, D., Prade, H. *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, 1980
- [17] Dubois, D., Prade, H. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988
- [18] Duda, R.O., Hart, P.E., Stork, D.G. *Pattern Classification* (2nd Edition). John Wiley & Sons Inc., New York, 2001
- [19] Fahlman, S. Faster-learning variations on back-propagation: An empirical study. W: Proc. of 1988 Connectionist Models Summer School, D. S. Touretzky, G. E. Hinton, and T. J. Sejnowski (eds.), Morgan Kaufmann Publishers, Los Altos CA, pp. 38-51
- [20] Fausett, L. *Fundamentals of Neural Networks. Architectures, Algorithms, and Applications*. Prentice Hall International Inc., Enlewood Cliffs, NJ 1994
- [21] Friedman, M., Ming, M., Kandel, A. Fuzzy linear systems. *Fuzzy Sets and Systems*, **96**, 1998, 201-209
- [22] Funahashi, K. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2(1989) 183-192
- [23] Gasilov, N., Amrahov, Sh.G., Fatullayev, A.G., Karakas, H.I., Akin, O. A geometric approach to solve fuzzy linear systems. arXiv:0910.4049v1 [cs.NA]
- [24] Girosi, F., Poggio, T. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Computation*, **1**, 1989, 465-469
- [25] Goodman, J.R. Some new results concerning random sets and fuzzy sets. *Information Science*, textbf34, 1984, 93-113
- [26] Hanss, M. The transformation method for the simulation and analysis of systems with uncertain parameters. *Fuzzy Sets and Systems*, **130**, 2002, 277-289
- [27] Hanss, M. *Applied Fuzzy Arithmetic. An Introduction with Engineering Applications*. Springer, 2005
- [28] Hecht-Nielsen, R. Kolmogorov's mapping neural network existence theorem. W: *Proc. of the International Conference on Neural Networks*, pp. 11-14, New York, 1987. IEEE Press
- [29] Herrera, F., Lozano, M., Verdegay, J.L. Generating fuzzy rules from examples using genetic algorithms. In: *Proc. of 5th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'94*, pp. 675-680, URL: <http://citeseer.ist.psu.edu/347754.html>
- [30] Hill, G., Horstkotte, E., Teichrow, J. *Fuzzy-C development system – user's manual*. Togai Infralogic, 30 Corporate Park, Irvine, CA 92714, USA

- [31] Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79(1982) 2554-2558
- [32] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4, 1991, 251-257
- [33] Huang, G.-B. Chen, L., Siew, Ch.-K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. on Neural Networks*, 17(4), 2006, 879-892
- [34] Ishibuchi, H. Nozaki, K., Yamamoto, N. Selecting fuzzy rules by genetic algorithm for classification problems. In: *Second IEEE International Conference on Fuzzy Systems*, pp. 1119-1124, San Francisco, CA., 1993
- [35] Jantzen, J. Design of fuzzy controllers. <http://www.iau.dtu.dk/~jj/pubs/design.pdf>
- [36] Jones, R.D. *et al.* Function approximation and time series prediction with neural networks. *1990 IJCNN International Joint Conference on Neural Networks*, 1990, pp. 649-665
- [37] Jowers, L.J., Buckley, J.J., Reilly, K.D. Simulating continuous systems. *Information Sciences*, 177, 2007, 436-448
- [38] Karr, C., Gentry, E. Fuzzy control of ph using genetic algorithms, *IEEE Transactions on Fuzzy Systems*, 1, 1993, 46-53
- [39] Khedkar, P.S., Keshav, S. Fuzzy prediction of timeseries. In: *Proc. IEEE Conference on Fuzzy Systems-92*, URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.30.2499>
- [40] Kim, H.S., Roschke, P.N. Design of fuzzy logic controller for smart base isolation system using genetic algorithm. *Engineering Structures*, 28, 2006, pp. 84-96
- [41] Klimke, A. An efficient implementation of the transformation method of fuzzy arithmetic. Technical Report 2003/009, Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, URL: <http://preprints.ians.uni-stuttgart.de>
- [42] Kloeden, P.E. Fuzzy dynamical systems. *Fuzzy Sets and Systems*, 7, 1982, 275-296
- [43] Kolmogorov, A.N. On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition. *Doklady Akademii Nauk SSSR*, 114(5), 1957, 953-956
- [44] Kołodziej, W. *Analiza matematyczna*. WNT, Warszawa 1978
- [45] Kohonen, T. *Self-Organization and Associative Memory*, Springer-Verlag, New York, 1984
- [46] Koronacki, J., Ćwik, J. *Statystyczne systemy uczące się*. WNT, Warszawa 2005
- [47] Kurková V. Kolmogorov's theorem is relevant, *Neural Computation*, 3(4), 1991, 617-622
- [48] Kurková V. Kolmogorov's theorem and multilayer neural networks, *Neural Networks*, 5, 1992, 501-506

- [49] Lapedes, A.S., Farber, R. Nonlinear signal processing using neural networks. Tech. Rep. No. LA-UR-87-2662). Los Alamos, NM: Los Alamos National Laboratory, 1987
- [50] LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R. Efficient BackProp. W: Orr, G. and Muller K. (Eds), *Neural Networks: Tricks of the trade*, LNCS 1524, pp. 9-50, Springer, 1998
- [51] M. Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, **6**, 1993, 861–867
- [52] Meyer, C. *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000
- [53] Minsky, M., Pappert, S. *Perceptrons*. MIT Press, Cambridge, Mass. 1969
- [54] Moody, J. Fast learning in multi-resolution hierarchies. *Advances in Neural Information Processing Systems*. Morgan Kaufmann Publishers Inc. 1989, pp. 29 - 39
- [55] Moody, J., Darken, C. Fast learning in networks of locally-tuned processing units. *Neural Computation*, **1**, 1989, pp. 289-303
- [56] Negoita, C.V., Ralescu, D.A. *Applications of Fuzzy Sets to System Analysis*. Birkhauser, 1975
- [57] Neruda, R., Stedry, A., Drkosová, J. Implementation of Kolmogorov's learning algorithm for feedforward neural networks. W: V.N. Alexandrov *et al.* *Computational Science – ICCS 2001 (Proc. Int. Conference, San Francisco, CA, USA, May 28-30, 2001)*, LNCS 2074, pp. 986-995, Springer 2001
- [58] Nguyen, H.T. A note on the extension principle for fuzzy sets. *J. Math. Anal. Appl.*, **64**, 1978, 369-380
- [59] Nozaki, K., Ishibuchi, H., Tanaka, H. A simple but powerful heuristic method for generating fuzzy rules from numerical data. *Fuzzy Sets and Systems*, **86**, 1997, 251-270
- [60] Orr, M.J.L. Introduction to radial basis function networks. Centre for Cognitive Science, University of Edinburgh, Edinburgh, Scotland 1996. URL: <http://anc.ed.ac.uk/rbf/papers/intro.ps.gz>
- [61] Orr, M.J.L. Recent advances in radial basis function networks. Centre for Cognitive Science, University of Edinburgh, Edinburgh, Scotland 1999. URL: <http://anc.ed.ac.uk/rbf/papers/recad.ps.gz>
- [62] Park, J., Sandberg, I.W. Universal approximation using radialbasis function networks. *Neural Comput*, **3**, 1991, 246–257
- [63] Park, Y.J., Lee, S.Y., Cho, H.S. A genetic algorithm-base fuzzy control of an electrohydraulic fin position servo system. In: *Proc. of 1999 IEEE Internat. Conference on Fuzzy Systems*, Seoul, Korea 1999, pp.III-1361-III-1366
- [64] Piegat, A. *Modelowanie i sterowanie rozmyte*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 1999
- [65] Plucińska, A., Pluciński, E. *Probabilistyka. Rachunek prawdopodobieństwa, statystyka matematyczna, procesy stochastyczne*. WNT, Warszawa, 2000

- [66] Porter L.L., Passino K.M., Genetic adaptive observers, *Engineering Applications of Artificial Intelligence*, **8**, 1995, pp. 261–269
- [67] Román-Flores, H., Barros, L.C., Bassanezi, C. A note on Zadeh's extensions. *Fuzzy Sets and Systems*, **117**, 2001, 327–331
- [68] Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 1959, 386–486
- [69] Ross, T.J. *Fuzzy Logic with Engineering Applications* J. Wiley, 2004
- [70] Rummelhard, D.E., McClelland, J.L., and the PDP Research Group. *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. MIT Press/Bradford Books, Cambridge, Mass, 1986
- [71] Rummelhard, D.E., Widrow, B., Lehr, M.A. Neural networks: applications in industry, business and science. *Communications of the ACM*, **37**(3), 1994, 93–105
- [72] Rutkowska, D., Piliński, M. Rutkowski, L. *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, PWN, Warszawa 1997
- [73] Rutkowski, L. *Metody i techniki sztucznej inteligencji. Inteligencja obliczeniowa*. PWN, Warszawa, 2005
- [74] Sanger, T.D. A tree-structured adaptive network for function approximation in high-dimensional spaces. *IEEE Transactions on Neural Networks*, **2**, 1991, pp. 285–293
- [75] Schweizer, B., Sklar, A. *Probabilistic Metric Spaces*. Elsevier & North-Holland, 1983
- [76] Shafer, G. *A Mathematical Theory of Evidence*, Princeton Univ. Press, N.J., 1976
- [77] D.A. Sprecher. A numerical construction of a universal function for Kolmogorov's superpositions. *Neural Network World*, **7**(4), 1996, 711–718
- [78] D.A. Sprecher. A numerical implementation of Kolmogorov's superpositions II. *Neural Networks*, **10**(3), 1997, 447–457
- [79] Wang, L.X., Mendel, J.M. Generating fuzzy rules by learning from examples. *IEEE Trans. on Systems, Man and Cybernetics*, **22**(6), 1992, 1414–1427, URL: <http://sipi.usc.edu/~mendel/publications/Wang&Mendel\%20SMC\%2011-12-92.pdf>
- [80] Widrow, B., Hoff, M.E. Adaptive switching circuits. *IRE WESCON Convention Record*, part 4, pp. 96–104, 1960
- [81] Widrow, B., Winter, R.G., Baxter, R.A. Learning phenomena in layered neural networks. *Proc. IEEE First Intl. Conference on Neural Networks*, 1987, vol. II, pp. 411–429
- [82] Yager, R.R. Level sets for membership evaluation of fuzzy subsets. In: R.R. Yager, ed. *Fuzzy Sets and Possibility Theory: Recent Developments*, Pergamon Press, Oxford, 1982, pp. 90–97
- [83] Tou, J., Gonzalez, R.C. *Pattern Recognition Principles*. Addison-Wesley, London 1974

- [84] Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. Parts 1, 2, and 3. *Infor. Sci.*, **8**, 1975, 199-249, **8**, 1975, 301-357, **9**, 1975, 43-80
- [85] Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1**: 3-28

Skorowidz

- aktywacji
 - próg, 80
- cechy (atrybuty), 77
- funkcja
 - decyzyjna, 77
 - regresji, 78
- klasa negatywna (pozytywna), 77
- klasyfikator, 77
- macierz wag, 81
- metoda transformacji
 - pełna, 32
 - rozszerzona, 32
 - zredukowana, 32
- obciążenie (*bias*), 79
- odległość Hausdorffa
 - definicja, 38
 - uogólniona, 38
- odwzorowanie rozmyte
 - ciągłe, 39
 - monotonicznie rosnące, 24
 - określenie, 23
- próba ucząca (zbiór trenujący), 77
- próg (aktywacji), 80
- przykład (instancja) klasy, 77
- pseudoinwersja (odwrotność Moore-Penrose'a), 93
- reguła delta
 - podstawowa, 83
- rozmyta liczba
 - trójkątna, 18
- rozmytej liczby
 - przedział niepewności (ufności), 18
 - wartość średnia, 18
 - wyostrzenie, 21
- rozmyty
 - proces, 73
 - przedział, 18
- rozwiązanie równania liniowego
 - słabe, 39
 - silne, 39
- sterownik (kontroler, regulator), 50
- twierdzenie
 - Funahashi'ego, 96
 - o reprezentacji, 20
- uczenie
 - nadzorowane, 77
- wektor wagowy (wagi), 79
- zadanie
 - regresji, 78
 - regresji
 - liniowej, 78
 - nieliniowej, 78
- zasada rozszerzania, 23, 73
- zbiór
 - liniowo separowalny, 80
 - nieseparowalny, 80
 - trenujący (próba ucząca), 77
- zbiór rozmyty
 - normalny, 14
 - subnormalny, 14
 - trójkątny, 17
 - trapezoidalny, 17
 - wypukły, 15
- zbioru rozmytego
 - α -przekrój (α -warstwa), 14
 - średnica, 14
 - jądro, 14
 - nośnik, 14
 - wysokość, 14