

## ESTYMACJA BŁĘDÓW MODELU POWIERZCHNI OPISANYCH FUNKCJAMI KSZTAŁTU ZA POMOCĄ SIECI NEURONOWYCH

Maria Mrówczyńska  
Uniwersytet Zielonogórski

**Streszczenie.** W artykule przedstawiono zagadnienie estymacji błędów modeli powierzchni określonej na dyskretnym zbiorze punktów o danych wartościach współrzędnych przestrzennych  $(x, y, z)$ . Przyjęto, że obiekt opisują funkcje kształtu w postaci płaszczyzny, paraboloidy eliptycznej oraz paraboloidy hiperbolicznej. Realizacja numeryczna zadania polegała na wyznaczeniu błędów modeli określonych za pomocą sieci neuronowych oraz na podstawie rozwiązania zadań wyrównawczych. Modelowanie za pomocą sieci neuronowych zrealizowano za pomocą sieci jednokierunkowych wielowarstwowych z zastosowaniem gradientowych metod optymalizacji oraz algorytmu ResilientbackPropagation (*RPROP*). Wyniki porównano z wynikami aproksymacji wielomianem drugiego i trzeciego stopnia, funkcją sklejaną oraz metodą kriging.

**Słowa kluczowe:** sieci neuronowe jednokierunkowe, algorytmy gradientowe, aproksymacja powierzchni

### WSTĘP

Zagadnienie określania kształtu powierzchni rozpiętej na zadanym zbiorze punktów można rozpatrywać z geodezyjnego punktu widzenia w dwóch przypadkach, a mianowicie [Nowak 2001]:

- punkty  $P_i$  ( $i = 1, 2, \dots, m$ ), którym przyporządkowano współrzędne teoretyczne  $x_i, y_i, z_i$  ( $i = 1, 2, \dots, m$ ),
- punkty  $P_i$  ( $i = 1, 2, \dots, m$ ) o empirycznych współrzędnych  $x_i, y_i, z_i$  ( $i = 1, 2, \dots, m$ ).

Stawiając hipotezę, że funkcja kształtu o  $n$  parametrach przyjmuje postać

$$F(x, y, z) = 0$$

wówczas dla punktów o współrzędnych teoretycznych, estymacja błędów modelu sprowadza się do rozwiązania zagadnienia minimalizacji

$$\min \sum_{i=1}^m v_{f_i}^2 = \sum_{i=1}^m F^2(x_i, y_i, z_i)$$

gdzie  $v_{f_i}; i = 1, 2, \dots, m$  reprezentują różnice między wartością funkcji określoną na podstawie współrzędnych teoretycznych a wartością funkcji uzyskaną w procesie minimalizacji.

Biorąc pod uwagę współrzędne empiryczne punktów będziemy mieli do czynienia z zadaniem wyrównawczym o postaci ogólnej

$$\min \sum_{i=1}^m v_{x_i}^2 + v_{y_i}^2 + v_{z_i}^2$$

przy czym

$$F(x_i^{obs} + v_{x_i}, y_i^{obs} + v_{y_i}, z_i^{obs} + v_{z_i}) = 0 \quad \text{dla } P_1, \dots, P_m.$$

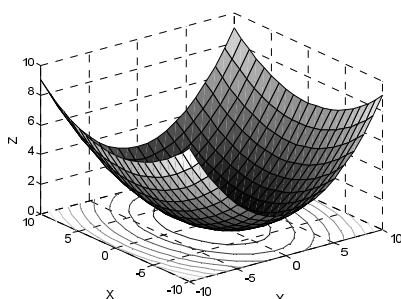
W tym przypadku błędy współrzędnych empirycznych nabierają szczególnego znaczenia w aspekcie dokładności modelowania obiektów.

W celu dokonania analizy estymacji błędów modelu zostały podjęte kroki poszukiwana minimum funkcji rozpiętej na zadanym zbiorze punktów o współrzędnych teoretycznych. Rozpatrywano trzy przypadki estymacji błędów modelu uzyskanych na podstawie rozwiązania zadania wyrównawczego z uwzględnieniem funkcji kształtu w postaci:

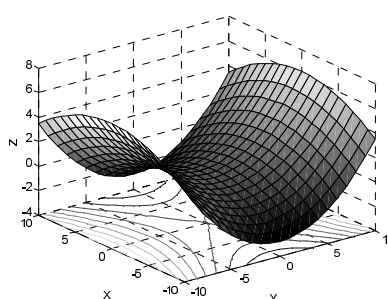
– równania paraboloidy eliptycznej  $z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$  (rys. 1a),

– równania paraboloidy hiperbolicznej  $z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$  (rys. 1b).

Rozwiązanie tego zagadnienia pozwala na przeprowadzenie krytycznej analizy dokładności odwzorowywania rzeźby terenu za pomocą sieci neuronowych, które stanowią uniwersalny układ aproksymacyjny, odwzorowujący wielowymiarowe zbiory danych oraz nie wymagają formułowania funkcji kształtu. Stosowanie sieci neuronowych wymaga jednak odpowiedniej struktury sieci, odpowiedniej liczby wzorców uczących w odniesieniu do ogólnej liczby punktów oraz doboru właściwej funkcji aktywacji (może być inna w warstwie ukrytej i w warstwie wyjściowej).



Rys. 1a.



Rys. 1b.

Rys. 1. Paraboloida eliptyczna oraz paraboloida hiperboliczna dla  $a=4$  i  $b=6$   
 Fig. 1. The elliptic paraboloid and the hyperbolic paraboloid for  $a=4$  and  $b=6$

## GRADIENTOWE METODY OPTIMALIZACJI

Realizacja zadań za pomocą sieci neuronowych wymaga nauczenia sieci prawidłowego przetwarzania informacji. Uczenie sieci polega na automatycznym wyznaczeniu wag neuronów w taki sposób, żeby na podstawie danych wprowadzonych do obliczeń (wektor wejściowy  $x$ ) uzyskać taki wektor wyjściowy  $y$ , którego współrzędne  $y_i$  byłyby pod względem wartości najbliższe współrzędnym  $d_i; i=1,2,\dots,M$  wektora zadanego  $d$ . Minimalizacja różnic między wartościami sygnałów wyjściowych  $y_i$  a wartościami zadanymi  $d_i$  odpowiada minimalizacji funkcji błędu  $E$  (funkcji celu, funkcji energetycznej), definiowanej z wykorzystaniem metryki euklidesowej jako:

$$E = \frac{1}{2} \sum_{i=1}^N \|y_i - d_i\|^2 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M (y_i^j - d_i^j)^2 \quad (1)$$

gdzie  $N$  jest liczbą wektorów wejściowych,  $M$  liczbą współrzędnych wektora wyjściowego.

Adaptacja wektora wag (uczenie sieci) przebiega zgodnie z regułą:

$$w^{(k+1)} = w^{(k)} + \eta p(w^{(k)}) \quad (2)$$

gdzie  $k$  oznacza numer kolejnej iteracji,  $\eta$  – współczynnik uczenia,  $p(w^{(k)})$  jest kierunkiem minimalizacji w przestrzeni wielowymiarowej  $w$ .

Metody gradientowe wykorzystują informacje dotyczące gradientu funkcji celu, który wykazuje kierunek największego wzrostu funkcji. Określenie gradientu wymaga rozwinięcia funkcji  $E(w)$  w dowolnie bliskim otoczeniu aktualnego rozwiązania

$w = [w_1, w_2, \dots, w_n]^T$  w szereg Taylora

$$E(\mathbf{w} + \mathbf{p}) = E(\mathbf{w}) + [\mathbf{g}(\mathbf{w})]^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{H}(\mathbf{w}) \mathbf{p} + \dots \quad (3)$$

gdzie

$$\mathbf{g}(\mathbf{w}) = \nabla E(\mathbf{w}) = \left[ \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \dots, \frac{\partial E}{\partial w_n} \right]^T \quad \text{– gradient } \mathbf{w} \text{ kierunku } \mathbf{p}$$

$$\mathbf{H}(\mathbf{w}) = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1 \partial w_1} & \dots & \frac{\partial^2 E}{\partial w_1 \partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial w_n \partial w_1} & \dots & \frac{\partial^2 E}{\partial w_n \partial w_n} \end{bmatrix} \quad \text{– hesjan.}$$

Rozwinięcie w szereg Taylora funkcji  $E(\mathbf{w})$  z uwzględnieniem składnika drugiego rzędu dostarcza informacji na temat krzywizny funkcji zawartej w hesjanie. Pominięcie tego składnika, jak to ma miejsce w przypadku stosowania metody największego spadku, pozwala jedynie uzyskać zbieżność liniową szczególnie w pobliżu punktu optymalnego. Jeżeli  $\mathbf{g}(\mathbf{w})^{(k)} = 0$  oraz hesjan  $\mathbf{H}(\mathbf{w})^{(k)}$  jest dodatnio określony, wówczas wartość funkcji celu  $E(\mathbf{w})$  w dowolnym punkcie należącym do bliskiego otoczenia punktu  $\mathbf{w}^{(k)}$  ma wartość większą niż w punkcie  $\mathbf{w}^{(k)}$ . Zatem punkt  $\mathbf{w}^{(k)}$  jest rozwiązaniem odpowiadającym minimalnej wartości funkcji celu.

Powszechnie stosowanym algorytmem wyboru kierunku minimalizacji jest algorytm największego spadku, który w procesie aktualizacji wag przyjmuje postać

$$w_{ij}^{(k+1)} = w_{ij}^{(k)} - \eta \frac{\partial E}{\partial w_{ij}}. \quad (4)$$

Metoda największego spadku zyskuje uznanie głównie ze względu na prostą realizację numeryczną oraz nie wymagającą zbyt dużej pamięci operacyjnej. W tym miejscu rozpatrywane zagadnienie należy uzupełnić uwagą, że korekta wag następuje w kierunku ujemnego wektora gradientu funkcji celu.

Metoda zmiennej metryki należy do metod quasi-newtonowskich, wykorzystujących w algorytmie informacje na temat krzywizny minimalizowanej funkcji bez konieczności kłopotliwego w realizacji jawnego wyznaczenia macierzy hesjanu. W metodach quasi-newtonowskich macierz hesjanu jest aproksymowana za pomocą różnicy pochodnych pierwszego rzędu. Ten tok postępowania nie wymaga trudnego do spełnienia (w ogólnym przypadku) warunku dodatniej określoności hesjanu w każdej iteracji, co ułatwia praktyczną implementację algorytmu. Zatem, stacjonarny punkt  $\mathbf{w}^*$  minimum funkcji celu  $E(\mathbf{w})$  będziemy poszukiwać na kierunku

$$\mathbf{p}^{(k)} = -[\mathbf{G}(\mathbf{w})^{(k)}]^{-1} \mathbf{g}(\mathbf{w})^{(k)}, \quad (5)$$

przy czym macierz odwrotna przybliżonego hesjanu  $V^{(k)} = [G(w)^{(k)}]^{-1}$  jako macierz modyfikowana z poprzedniej iteracji (wartość startowa  $V^0 = 1$ ) została opisana rekurencyjną zależnością (algorytm Davidona – Fletchera – Powella [Osowski 1996])

$$V^{(k)} = V^{(k-1)} + \frac{s^{(k)}(s^{(k)})^T}{(s^{(k)})^T r^{(k)}} - \frac{V^{(k-1)} r^{(k)}(r^{(k)})^T V^{(k-1)}}{(r^{(k)})^T V^{(k-1)} r^{(k)}} \quad (6)$$

gdzie  $s^{(k)}$  oraz  $r^{(k)}$  oznaczają odpowiednio przyrost wektora wag  $w$  oraz gradientu  $g(w)$  w dwóch kolejnych iteracjach, tj.  $s^{(k)} = w^{(k)} - w^{(k-1)}$ ,  $r^{(k)} = g(w)^{(k)} - g(w)^{(k-1)}$ .

Metoda Levenberga-Marquardta jest bardzo zbliżona do metody zmiennej metryki. Jej zaletą jest to, że prawie zawsze jest zbieżna i nie „zwalnia zbieżności”, co często występuje w metodzie największego spadku. Analogicznie jak w metodzie zmiennej metryki wykorzystuje się kwadratowe przybliżenie funkcji celu  $E(w)$  oraz aproksymowaną wartość hesjanu  $G(w)$  z uwzględnieniem czynnika regularyzacji  $r^{(k)}I$ .

Metoda gradientów sprzężonych wykorzystuje model kwadratowy funkcji celu bez konieczności wykonania w każdej iteracji szeregu operacji macierzowych. Kierunki  $p_1, p_2, \dots, p_n$  nazywamy kierunkami sprzężenia, które generowane są na podstawie informacji o nowych pochodnych i poprzednich kierunkach. Wektor kierunkowy w  $k$ -tej iteracji przyjmuje postać

$$p^{(k)} = -g(w)^{(k)} + \beta^{(k-1)} p^{(k-1)} \quad (7)$$

gdzie:  $g(w)^{(k)}$  - wektor gradientu,  $\beta^{(k-1)}$  - współczynnik sprzężenia.

Pośród wielu reguł wyznaczenia współczynnika sprzężenia  $\beta^{(k-1)}$ , najczęściej stosowane są [Osowski 1996]:

$$\text{– metoda Polaka – Riebiere'a } \beta^{(k-1)} = \frac{g(w)^{(k)T} (g(w)^{(k)} - g(w)^{(k-1)})}{(g(w)^{(k-1)})^T (g(w)^{(k-1)})}, \quad (8)$$

$$\text{– metoda Fletchera – Reevesa } \beta^{(k-1)} = \frac{g(w)^{(k)T} (g(w)^{(k)} - g(w)^{(k-1)})}{-p^{(k)T} g(w)^{(k-1)}}. \quad (9)$$

## ALGORYTM RPROP

Algorytm *Resilient Backpropagation* (algorytm RPROP) umożliwia przyspieszenie zbieżności i zmniejsza ryzyko zatrzymania procesu w lokalnym minimum. Istota algorytmu polega na aktualizacji wag w zależności od znaku składowych gradientu, zaś

sama jego wartość jest pomijana [Duch i in. 2000, Osowski 1996, Riedmiller i Braun 1992].

Korekta wag przebiega według zależności

$$w^{(k+1)} = w^{(k)} - \eta^{(k)} \operatorname{sgn}(\nabla E(\mathbf{w})^{(k)}), \quad (10)$$

w której współczynnik uczenia  $\eta$  jest zależny od znaku gradientu. Współczynnik  $\eta$  jest dobierany w każdym cyklu dla każdej wagi  $w$  indywidualnie. Wzrost wartości tego współczynnika następuje wtedy, gdy znak gradientu w dwóch kolejnych iteracjach jest jednakowy, natomiast w przeciwnym wypadku następuje jego redukcja. W związku z tym

$$\eta^{(k)} = \begin{cases} \min(a\eta^{(k-1)}, \eta_{\min}) & \text{dla } \nabla E(\mathbf{w})^{(k)} \nabla E(\mathbf{w})^{(k-1)} > 0 \\ \max(b\eta^{(k-1)}, \eta_{\max}) & \text{dla } \nabla E(\mathbf{w})^{(k)} \nabla E(\mathbf{w})^{(k-1)} < 0 \\ \eta^{(k-1)} & \text{w innym przypadku} \end{cases} \quad (11)$$

Zawarte we wzorach symbole  $a$  i  $b$  są stałymi:  $a=1.2$ ,  $b=0.5$ , natomiast  $\eta_{\min}$  i  $\eta_{\max}$  oznaczają odpowiednio minimalną i maksymalną wartość współczynnika uczenia, równą w algorytmie RPROP odpowiednio  $10^6$  oraz 50 [Duch i in. 2000, Osowski 1996, Riedmiller i Braun 1992].

## WYNIKI SYMULACJI NUMERYCZNYCH

Odwzorowanie powierzchni opisanych funkcjami kształtu paraboloidą eliptyczną oraz hiperboliczną przeprowadzono dla zbioru uczącego o liczebności 20 punktów, zbioru testowego o liczebności 440 punktów (rys. 2a i 2b) dla sieci o architekturze 2\_10\_1 (2 wejścia, 10 neuronów w warstwie ukrytej, 1 wyjście) z wykorzystaniem bipolarnej funkcji aktywacji postaci

$$y = f(\text{net}) = \frac{1 - \exp(-\lambda \text{net})}{1 + \exp(-\lambda \text{net})}, \quad \lambda > 0 \quad (12)$$

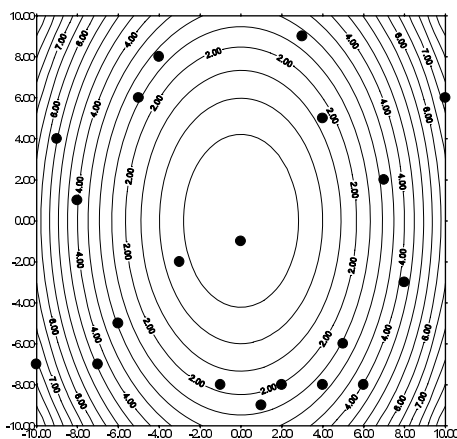
W procesie uczenia jako kryterium zakończenia procesu minimalizacji przyjęto zmianę błędu uczenia w dwóch kolejnych iteracjach wynoszącą  $1e-10$  oraz liczbę iteracji na poziomie 25000.

Zastosowana metoda krigingu daje najlepsze rezultaty dla wartości punktowych lub blokowych badanej zmiennej oraz pozwala na osiągnięcie minimalnej wariancji podczas procesu estymacji. Wartości wariancji krigingu zależą od położenia punktów względem szacowanej lokalizacji, od wag przypisanych punktom oraz parametrów modelu semiwariogramu. Wartości funkcji semiwariogramu zostały obliczone na podstawie klasycznej formuły Matherona [Matheron 1962]:

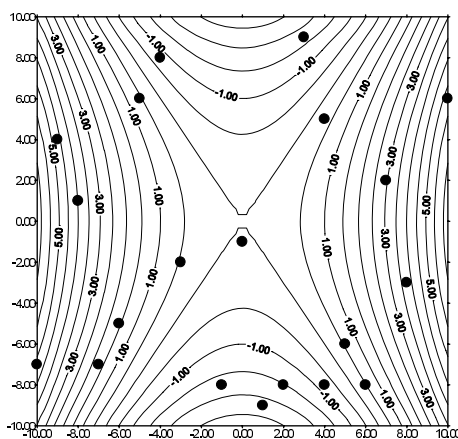
$$\gamma(h) = \frac{1}{2N_h} \sum_{i=1}^n (z_{i+h} - z_i)^2 \quad (13)$$

gdzie  $\gamma(h)$  – wartość funkcji semiwariogramu,  $z_i, z_{i+h}$  – wysokość w punktach  $i$  i  $i+h$  odległych o wartość  $h$ ,  $N_h$  – liczba par punktów pomiarowych odległych o wartość  $h$ .

W pracy została również zastosowana aproksymacja typu spline wykorzystująca funkcje sklejane złożone z wielomianów 3 stopnia.



Rys. 2a.



Rys. 2b.

Rys. 2. Rozmieszczenie punktów zbioru uczącego – paraboloida eliptyczna (2a) oraz paraboloida hiperboliczna (2b)

Fig. 2. The layout of training set points – the elliptic paraboloid (2a) and the hyperbolic paraboloid (2b)

Wyniki uczenia oraz testowania sieci w postaci błędu średniego

$$RMSE = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n (d_i - y_i)^2} \quad (14)$$

wraz z wynikami uzyskanymi za pomocą wielomianów, aproksymacji typu spline oraz krigingu zestawiono w tabeli 1.

Tabela 1. Wyniki odwzorowania modeli powierzchni  
Table 1. The results of the area models projection

Metoda odwzorowania The approximation method	Paraboloida eliptyczna błąd $RMSE$ [m] The elliptic paraboloid error $RMSE$ [m]		Paraboloida hiperboliczna błąd $RMSE$ [m] The hyperbolic paraboloid error $RMSE$ [m]	
	zbiór uczący training set	zbiór testowy testing set	zbiór uczący training set	zbiór testowy testing set
	Algorytm <i>RPROP</i>	0.02	0.03	0.05
Metoda Backpropagation	0.68	0.78	0.68	0.76
Gradientów sprzężonych	0.15	0.19	0.22	0.23
Zmiennej metryki	0.00	0.00	0.00	0.00
Levenberga – Marquardta	0.00	0.00	0.00	0.00
Aproksymacja typu spline	0.02	0.03	0.01	0.02
Aproksymacja typu kriging	0.08	0.09	0.07	0.05
Wielomian 2 stopnia	0.00	0.00	0.00	0.00
Wielomian 3 stopnia	0.00	0.00	0.00	0.00

## PODSUMOWANIE

Aproksymacja formy kwadratowej trzech liczb rzeczywistych  $x, y, z$  (paraboloida eliptyczna i hiperboliczna) za pomocą sieci wielomianów drugiego lub trzeciego stopnia oraz za pomocą sieci jednokierunkowych wielowarstwowych uczonych metodą zmiennej metryki i Levenberga – Marquardta daje jakościowo najlepszy wynik. Niewiele większe błędy uzyskamy stosując algorytm heurystyczny *RPROP* oraz aproksymację typu spline i kriging. W tym miejscu należy dodać, że algorytm *RPROP* wymaga około 10 razy więcej iteracji, aby osiągnąć zbieżność niż podstawowe metody gradientowe. Najmniej korzystne rezultaty osiągnięto stosując metodę największego spadku, ze względu na zbieżność liniową oraz niewielkie postępy minimalizacji w pobliżu punktu optymalnego.

Na podstawie uzyskanych wyników można stwierdzić, że aproksymacja modeli powierzchni opisanych podstawowymi funkcjami kształtu nie odbiega pod względem dokładności od metod tradycyjnych, takich jak aproksymacja typu spline i kriging. Należy jedynie sprostać wszelkim wymaganiom dotyczącym uzyskania minimum lokalnego, położonego w bliskiej odległości od minimum globalnego.

## PIŚMIENNICTWO

- Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R., 2000. Sieci neuronowe, Akademicka Oficyna Wydawnicza Exit, Warszawa.
- Matheron G., 1962. Traite de Geostatistique Appliquee, Memoire du Buremu de Recherch Geologique et Minierem, vol.14, Editions Technik, Paris.
- Nowak E., Estymacja i weryfikacja numerycznego modelu terenu, 2001. XI Konf. Naukowo-Techniczna „Systemy Informacji Przestrzennej”, Warszawa 28 – 30 maja 2001.
- Nowak E., Wyznaczanie kształtu poprzez estymację błędów pomiaru i modelu, 2001. V Konf. Naukowo-Techniczna „Problemy Automatyzacji w Geodezji Inżynierskiej”, Warszawa 29 – 30 marca 2001.



- Osowski S., 1996. Sieci neuronowe w ujęciu algorytmicznym, WNT, Warszawa.  
Riedmiller M., Braun H., 1992. A fast adaptive learning algorithm, Technical Report, University Karlsruhe, Germany.

**THE ESTIMATION OF ERRORS  
OF AREA MODELS DESCRIBED  
BY THE SHAPE FUNCTIONS BY THE MEANS  
OF NEURAL NETWORKS**

**Abstract.** The article deals with the issue of estimation of the area models errors determined on the basis of a discrete points set with the given values of space coordinates (x, y, z). The object was assumed to be described by shape functions in the form of the elliptic paraboloid and the hyperbolic paraboloid. The digital task accomplishment consisted in the statistic verification of errors of the models determined by neural networks and by the accomplishment of adjustment tasks. Modeling by the means of neural networks was carried out by the unidirectional multilayer networks with the application of gradient methods of optimalization and by ResilientbackPropagation algorithm (RPROP). The obtained results were compared with the following results of approximation of the second and the third degree of polynomial, the b-spline function and the kriging's method.

**Key words:** neural networks, gradient methods of optimalization, approximation method

Zaakceptowano do druku – Accepted for print: 29.03.2007