ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria II: WIADOMOŚCI MATEMATYCZNE XXVII (1987)

wmptm 1897

Ministerstwo Nauki i Szkolnictwa Wyższego

DIETHARD PALLASCHKE (Karlsruhe)

Metoda elipsoidalna w programowaniu liniowym *

W referacie tym([1]) przedstawię metodę elipsoidalną, rozwiniętą w Kijowie w latach siedemdziesiątych przez N. Z. Shora [7] i wykorzystaną przez L. G. Khachiana [4] do programowania liniowego. Zacznę od przedstawienia metody Khachiana, a następnie przedstawię późniejsze jej modyfikacje.

Rozważmy typowe zadanie programowania liniowego. Będziemy maksymalizować funkcję liniową postaci $\langle c, x\rangle$, gdzie c jest stałym wektorem, określoną na przestrzeni $\boldsymbol{R}^n$ z iloczynem skalarnym $\langle , \rangle$, przy m ograniczeniach

$$\langle a_i, x\rangle \leqslant b_i, \qquad i \in \{1, \ldots, m\},$$

będącymi nierównościami liniowymi. Zakładamy, że $x \geqslant 0$, co nie ma wpływu na ogólność algorytmu.

Przy ujęciu Khachiana musimy założyć, że obszar dopuszczalny

$$P := \{x \in \boldsymbol{R}^n \mid \langle a_i, x\rangle \leqslant b_i,\ i \in \{1, \ldots, m\},\ x \geqslant 0\} \tag{1}$$

jest zwarty o niepustym wnętrzu.

Powyższe zadanie programowania liniowego rozwiązać można za pomocą algorytmu składającego się z dwóch kroków:

(i) Korzystając z metody elipsoidalnej, znajdujemy punkt $\tilde{x} \in P$ (metoda wyznaczania tego punktu zostanie przedstawiona poniżej; załóżmy na razie, że ten punkt jest już określony).

(ii) Obszar P zredukujemy przez dodanie następującej nierówności:

$$\langle c, x\rangle \geqslant \langle c, \tilde{x}\rangle.$$

---

* Jest to tekst referatu wygłoszonego (po polsku) dnia 13 września 1985 r. na Zjeździe Naukowym PTM w Kielcach (*przyp. red.*).

([1]) Autor serdecznie dziękuje Polskiemu Towarzystwu Matematycznemu za zaproszenie na zjazd w Kielcach oraz za możliwość wygłoszenia tego referatu. Szczególnie serdecznie autor dziękuje profesorowi S. Rolewiczowi za zainteresowanie go tym zjazdem.

Otrzymamy w ten sposób nowy obszar

$$\tilde{P} := P \cap \{x \in \boldsymbol{R}^n \mid \langle c, x\rangle \geqslant \langle c, \tilde{x}\rangle\}.$$

Do tego nowego obszaru stosujemy ponownie krok (i), tzn. znajdujemy punkt $\tilde{\tilde{x}} \in \tilde{P}$, itd. W ten sposób otrzymujemy ciąg punktów zbieżny do rozwiązania optymalnego.

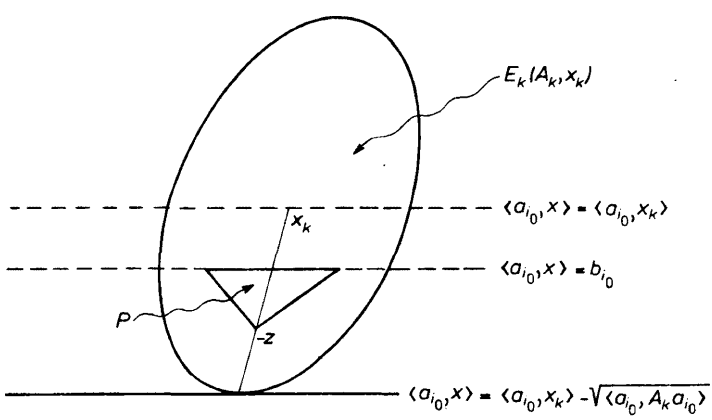
Khachian [4] udowodnił, że – przy pewnych założeniach – każdy pojedynczy krok w celu określenia punktu z obszaru dopuszczalnego jest wielomianowym([2]), tzn. istnieje pewien wielomian $p(n, m)$ dwu zmiennych (gdzie n oznacza wymiar przestrzeni, a m to ilość ograniczeń) taki, że liczba wszystkich działań algebraicznych (łącznie z pierwiastkowaniem) wykonanych w tym kroku jest nie większa niż $p(n, m)$.

Teraz pokażemy, w jaki sposób wykorzystujemy metodę elipsoidalną do wyznaczenia punktu $\tilde{x}$. Obszar dopuszczalny (1) jest zwarty, zatem P jest zawarte w pewnej kuli. Przy danym iloczynie skalarnym kule są elipsoidami. Będziemy konstruować pewien ciąg elipsoid, których objętości maleją do zera (co więcej, stosunek kolejnych objętości jest ograniczony z góry przez pewną liczbę mniejszą od 1), nadto każda z nich zawiera obszar dopuszczalny. Ponieważ obszar dopuszczalny ma niepuste wnętrze, jego objętość jest dodatnia. Wynika stąd, że po pewnej liczbie kroków środki elipsoid będą należeć do obszaru dopuszczalnego.

W k-tym kroku rozważmy elipsoidę

$$E_k = E_k(A_k, x_k) := \{x \in R^n \mid \langle (x - x_k),\ A_k^{-1}(x - x_k)\rangle \leqslant 1\}, \tag{2}$$

gdzie A_k jest macierzą symetryczną dodatnio określoną, przedstawiającą osie elipsoidy, a x_k jest środkiem elipsoidy ($x_k \in \boldsymbol{R}^n$). Zakładamy, że



Rys. 1

([2]) Było to rozwiązanie problemu nurtującego przez kilkanaście lat środowisko numeryczne; nie wiadomo było, czy taki algorytm istnieje. Używany powszechnie algorytm symplex w niektórych złośliwych przypadkach może wymagać liczby obliczeń rosnącej eksponencjalnie wraz z wymiarem przestrzeni, nawet przy pewnych założeniach o współczynnikach (*przyp. red.*).

$P \subset E_k(A_k, x_k)$. Wzajemne położenie E_k i P pokazuje rysunek 1 (w przypadku dwóch zmiennych i trzech ograniczeń, tzn. $n = 2$, $m = 3$).

Teraz rozpatrzmy dwa przypadki: Jeżeli środek elipsoidy jest już punktem obszaru dopuszczalnego, tzn. $x_k \in P$, to algorytm jest zakończony. W przeciwnym przypadku istnieje co najmniej jedna hiperpłaszczyzna niedopuszczalna dla punktu x_k o równaniu

$$\langle a_{i_0}, x\rangle = b_{i_0}, \tag{3}$$

tzn. taka, że $\langle a_{i_0}, x_k\rangle > b_{i_0}$.

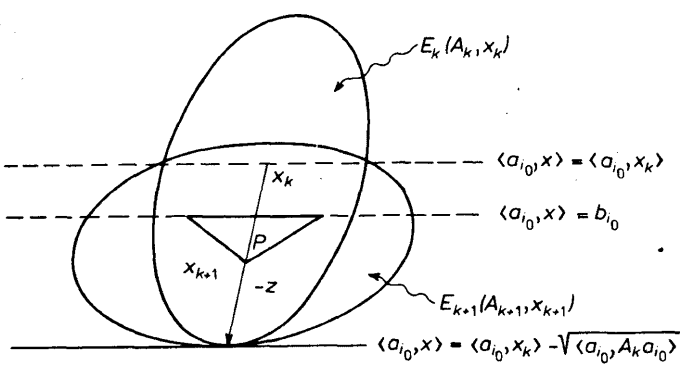
Konstrukcja przebiega dalej następująco: Hiperpłaszczyznę (3) przesuwamy równolegle do punktu środkowego x_k. Tę samą hiperpłaszczyznę przesuwamy również równolegle tak, aby była styczna do elipsoidy (rys. 1). Otrzymujemy więc dwie hiperpłaszczyzny. Pierwsza przechodzi przez punkt x_k i ma równanie

$$\langle a_{i_0}, x\rangle = \langle a_{i_0}, x_k\rangle. \tag{4}$$

Druga hiperpłaszczyzna ma równanie

$$\langle a_{i_0}, x\rangle = \langle a_{i_0}, x_k\rangle - \langle a_{i_0}, z\rangle, \tag{5}$$

gdzie z jest wektorem o początku w punkcie styczności i końcu x_k. Nowa elipsoida $E_{k+1}(A_{k+1}, x_{k+1})$ jest teraz wyznaczona w sposób jednoznaczny przez punkty przecięcia hiperpłaszczyzny (4) z elipsoidą $E_k(A_k, x_k)$ i przez punkt styczności hiperpłaszczyzny (5) z tą elipsoidą (rys. 2). Zauważmy, że ta nowa elipsoida ma tę samą hiperpłaszczyznę podparcia co elipsoida E_k.



Rys. 2

Łatwy rachunek pokazuje, że wektor z może być określony wzorem

$$z := \frac{A_k a_{i_0}}{\sqrt{\langle a_{i_0}, A_k a_{i_0}\rangle}};$$

ponadto

$$x_{k+1} = x_k - \frac{1}{n+1} z$$

oraz

$$A_{k+1} = \frac{n^2}{n^2-1}\left(A_k - \frac{2}{n+1} z \otimes z\right),$$

gdzie produkt tensorowy $z \otimes z$ jest macierzą $n \times n$ otrzymaną jako iloczyn $z^T z$ kolumny będącej transpozycją z^T wektora $z := (z_1, \ldots, z_n)$ przez wektor z.

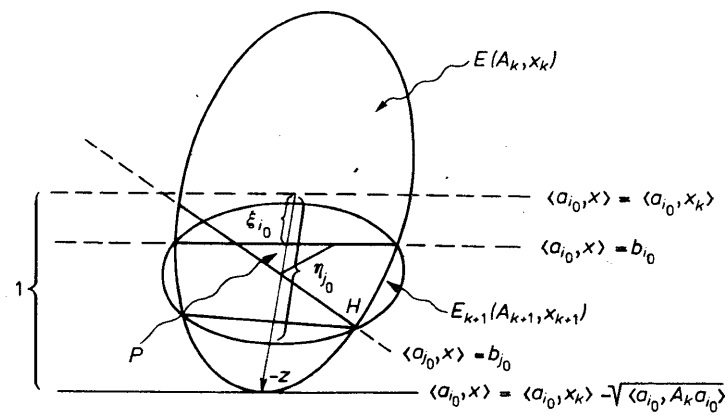
Powyższe wzory pochodzą od Khachiana. Czytając pracę Shora, można zauważyć, że wzory te są dokładnie wzorami na elipsoidę o najmniejszej objętości, zawierającą część wspólną elipsoidy E_k i obszaru między płaszczyznami (4) i (5) (rys. 2).

Teraz przedstawimy modyfikację tej metody. Została ona opisana w mojej pracy wspólnej z Hermannem Königiem [5].

Weźmy dwie hiperpłaszczyzny, z których jedna jest niedopuszczalna i ma równanie (3), a druga jest dopuszczalna, o równaniu

$$\langle a_{j_0}, x\rangle = b_{j_0}. \tag{6}$$

Niech teraz H będzie przecięciem elipsoidy z hiperpłaszczyzną (6). Rozważmy teraz hiperpłaszczyznę podpierającą zbiór H, równoległą do hiperpłaszczyzny (3). Nowa elipsoida jest teraz elipsoidą o najmniejszej objętości zawierającą obszar przedstawiony jako czworokąt na rysunku 3.



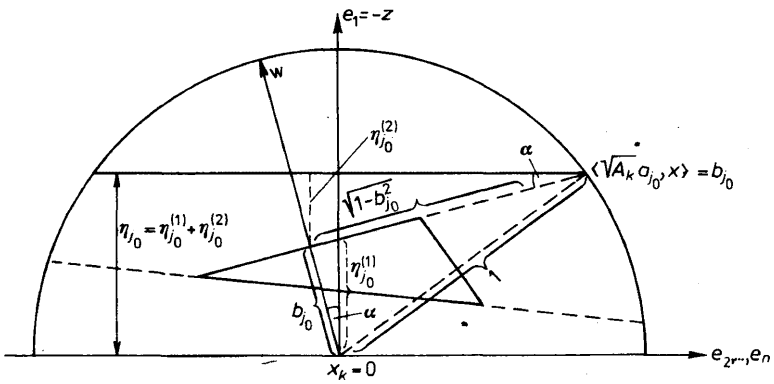
Rys. 3

Aby przedstawić wzory opisujące tę elipsoidę, wprowadzamy nowe współrzędne ξ_i i η_j. Wielkości te zaznaczone są klamrami na rysunku 3. Zauważmy, że na mocy twierdzenia Talesa mamy

$$\frac{\xi_{i_0}}{1} = \frac{\langle a_{i_0}, x_k\rangle - b_{i_0}}{\langle a_{i_0}, x_k\rangle - (\langle a_{i_0}, x_k\rangle - \langle a_{i_0}, z\rangle)} = \frac{-b_{i_0} + \langle a_{i_0}, x_k\rangle}{\sqrt{\langle a_{i_0}, A_k a_{i_0}\rangle}},$$

co daje ξ_{i_0}. Wyznaczenie η_j nie jest tak proste. Zauważmy, że stosunek η_{j_0} do długości wektora $-z$ nie zmienia się przy przekształceniach afinicznych. Wobec tego przekształcamy elipsoidę w kulę jednostkową o środku w

punkcie zero. Oznaczamy przez e_k k-ty wektor jednostkowy w R^n. Bez zmiany ogólności możemy założyć, że obraz punktu środkowego x_k leży w początku układu współrzędnych, a obraz wektora $-z$ leży na pierwszej osi. Teraz część południowa elipsoidy leży w północnej części kuli (rys. 4).



Rys. 4

Po dokonaniu transformacji elipsoidy w kulę zachowujemy poprzednie oznaczenia. Przypuśćmy więc, że (3) jest równaniem hiperpłaszczyzny dopuszczalnej, a (6) — niedopuszczalnej. Wektor prostopadły do hiperpłaszczyzny dopuszczalnej oznaczamy przez w. Niech α oznacza kąt między wektorem w a wektorem jednostkowym e_1. Zauważmy, że α jest również kątem między hiperpłaszczyzną dopuszczalną a hiperpłaszczyzną prostopadłą do pierwszej osi oraz że $\cos\alpha = \langle w, e_1\rangle$.

Szukana wielkość jest sumą $\eta_{j_0} = \eta_{j_0}^{(1)} + \eta_{j_0}^{(2)}$, gdzie

$$\eta_{j_0}^{(1)} = b_{j_0}\cos\alpha,$$

a b_{j_0} jest odległością hiperpłaszczyzny dopuszczalnej od początku układu współrzędnych. Z twierdzenia Pitagorasa wynika, że

$$\eta_{j_0}^{(2)} = \sqrt{1-b_{j_0}^2}\sin\alpha = \sqrt{1-b_{j_0}^2}\sqrt{1-\cos^2\alpha}.$$

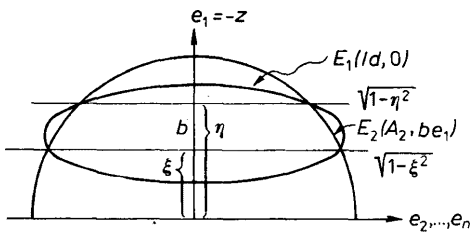
Wracając z tymi równaniami do wyjściowej k-tej elipsoidy, możemy określić:

$$\eta_{j_0} := \frac{(b_{j_0} - \langle a_{j_0}, x_k\rangle)\langle a_{j_0}, z\rangle}{\langle a_{j_0}, A_k a_{j_0}\rangle} + \left(1 - \frac{(b_{j_0} - \langle a_{j_0}, x_k\rangle)^2}{\langle a_{j_0}, A_k a_{j_0}\rangle}\right)^{\frac{1}{2}} \left(1 - \frac{\langle a_{j_0}, z\rangle^2}{\langle a_{j_0}, A_k a_{j_0}\rangle}\right)^{\frac{1}{2}}.$$

Niech teraz

$$\xi := \max_i \xi_i, \qquad \eta := \min_j \eta_j,$$

gdzie i jest wskaźnikiem hiperpłaszczyzny niedopuszczalnej, a j wskaźnikiem hiperpłaszczyzny dopuszczalnej. Biorąc liczby ξ i η, otrzymamy część K o najmniejszej grubości zawierającą obszar dopuszczalny. Teraz nasze zadanie sprowadza się do znalezienia wzorów opisujących elipsoidę o najmniejszej objętości zawierającą K. W tym celu zauważmy, że przekształcenie afiniczne zachowuje stosunki objętości. Transformujemy więc ponownie elipsoidę na kulę jednostkową (rys. 5).



Rys. 5

Łatwo zauważyć, że w przestrzeni $\boldsymbol{R}^2$ prawdziwe są następujące wzory:

$$A_{k+1} := \lambda(A_k - \mu z \otimes z), \qquad x_{k+1} := x_k - \gamma z,$$

opisujące osie i środek elipsy o najmniejszej objętości, gdzie $\lambda, \mu, \gamma \in \boldsymbol{R}$. Wzory te opisują również poszukiwaną elipsoidę o najmniejszej objętości w przestrzeni $\boldsymbol{R}^n$. Wynika to z twierdzenia F. Johna [3], mówiącego, że dla obszaru zwartego i wypukłego w $\boldsymbol{R}^n$, elipsoida o najmniejszej objętości zawierająca ten obszar jest jednoznacznie wyznaczona. Zatem przez obrót układu współrzędnych wokół pierwszej osi otrzymamy ogólną sytuację w $\boldsymbol{R}^n$.

Oznaczamy teraz kulę jednostkową przez $E_1(I, 0)$, gdzie I jest macierzą jednostkową, oraz poszukiwaną elipsoidę przez $E_2(A_2, be_1)$, gdzie $b \in \boldsymbol{R}$ (rys. 5). A więc A_2 jest postaci $\lambda(I - \mu e_1^T \otimes e_1)$, tzn.

$$A_2 = \begin{bmatrix} \lambda(1-\mu) & & & 0 \\ & \lambda & & \\ & & \lambda & \\ & & & \ddots \\ 0 & & & \lambda \end{bmatrix}.$$

Stąd wynika, że równanie szukanej elipsoidy jest postaci:

$$\sum_{i=2}^{n} \frac{x_i^2}{\lambda} + \frac{(x_1 - b)^2}{\lambda(1-\mu)} = 1.$$

Dla $x_1 = \xi$ otrzymujemy równanie

$$\frac{(1-\xi^2)}{\lambda} + \frac{(\xi - b)^2}{\lambda(1-\mu)} = 1, \tag{7}$$

a dla $x_1 = \eta$ otrzymamy drugie równanie postaci:

$$\frac{(1-\eta^2)}{\lambda} + \frac{(\eta-b)^2}{\lambda(1-\eta)} = 1. \tag{8}$$

Z powyższych równań wynika, że $b = \mu(\xi+\eta)/2$. Po podstawieniu b do równania (8) i zastąpieniu $(1-\mu)$ przez φ, otrzymamy równanie kwadratowe:

$$(\xi+\eta)^2\varphi^2 - \left(4(\lambda-1)+2(\eta^2+\xi^2)\right)\varphi + (\eta-\xi)^2 = 0.$$

Równanie to ma dwa pierwiastki:

$$\varphi_{1,2} = \frac{2(\lambda-1)+(\eta^2+\xi^2)+\left[(\lambda-1)^2+(\eta^2+\xi^2)(\lambda-1)+\frac{\xi^2\eta^2}{2}\right]^{\frac{1}{2}}}{(\xi+\eta)^2}.$$

Wystarczy rozpatrzyć tylko jeden z tych pierwiastków, np. φ_1, pomnożony przez stałą $(\xi+\eta)^2$. Objętość elipsoidy jest równa

$$\operatorname{vol}(E_2(A_2, be_1)) = c(\lambda^n\varphi_1)^{\frac{1}{2}},$$

gdzie $c > 0$ jest pewną stałą. Zatem objętość jako funkcję zmiennej λ piszemy w postaci:

$$V(\lambda) := \lambda^n\left[2(\lambda-1)+(\eta^2+\xi^2)-2\left((\lambda-1)^2+(\eta^2+\xi^2)(\lambda-1)+\frac{\eta^2\xi^2}{2}\right)^{\frac{1}{2}}\right].$$

Aby znaleźć ekstremum tej funkcji, zauważmy, że ma ona tylko ekstrema wewnętrzne. Pochodna jej jest postaci

$$V'(\lambda) = n\lambda^{n-1}\left[2(\lambda-1)+(\eta^2+\xi^2)-2\left((\lambda-1)^2+(\eta^2+\xi^2)(\lambda-1)+\frac{\eta^2\xi^2}{2}\right)^{\frac{1}{2}}\right]$$
$$+\lambda^n\left[2-\frac{2(\lambda-1)+(\xi^2+\eta^2)}{\left((\lambda-1)^2+(\eta^2+\xi^2)(\lambda-1)+\frac{1}{2}\eta^2\xi^2\right)^{\frac{1}{2}}}\right]$$

Równanie $V'(\lambda) = 0$ ma rozwiązanie $\lambda = 0$ oraz

$$\lambda = \frac{n^2}{n^2-1}\left[1-\frac{\xi^2+\eta^2}{2}+\left(\frac{(\eta^2-\xi^2)^2}{4}+\frac{(1-\eta^2)(1-\xi^2)}{n^2}\right)^{\frac{1}{2}}\right].$$

Dla tej drugiej wartości funkcja V ma minimum globalne. Łatwo obliczyć, że

$$\mu = 1-\frac{(\eta-\xi)^2}{\eta^2+\xi^2+2\left(\frac{n+1}{n}\lambda-1\right)}, \qquad \gamma = \xi+[(\lambda-1+\xi^2)(1-\mu)]^{\frac{1}{2}}.$$

W przypadku specjalnym, gdy $\eta = 1$, otrzymujemy wzory Goldfarba–Todda [2], a mianowicie

$$\lambda = \frac{n^2}{n^2-1}(1-\xi^2), \quad \mu = \frac{2}{n+1}\frac{1+n\xi}{1+\xi}, \quad \gamma = \frac{1+n\xi}{n+1}.$$

Natomiast w przypadku, gdy $\eta = 1$ i $\xi = 0$, otrzymamy oryginalne wzory Khachiana.

Tę metodę można też zastosować do programowania wypukłego. Można to znaleźć w pracach Eckera–Kupferschmida [1] i P. Rechta [6].

Na Uniwersytecie w Karlsruhe został opracowany program komputerowy (w języku FORTRAN) dla zadania liniowego i kwadratowego wykorzystującego metodę elipsoidalną. W porównaniu algorytmu sympleks z metodą elipsoidalną (w przypadku liniowym), sympleks jest dużo szybszy. Wynika to z tego, że w metodzie elipsoidalnej w każdym kroku należy obliczyć bardzo wiele iloczynów skalarnych. Największymi przykładami, dla których zastosowano u nas metodę elipsoidalną, były problemy liniowe z 500 ograniczeniami i 100 zmiennymi. Na komputerze IBM 360/165 albo SIEMENS 7880 czas obliczeń dla tych przykładów jest dłuższy od jednej godziny CPU. Oznacza to, że ten algorytm nie nadaje się do użycia dla zadań o dużym wymiarze.

Z drugiej strony algorytm elipsoidalny jest bardziej stabilny. Istnieją mianowicie przykłady (pierwszy z nich podany przez S. Walukiewicza [8]), w których algorytm sympleks nie jest zbieżny do punktu optymalnego, ponieważ kroki w tym algorytmie nie są stabilne. Metoda elipsoidalna zastosowana do tego przykładu daje rozwiązanie optymalne.

Bibliografia

[1] J. G. Ecker, M. Kupferschmid, *An implementation and generalization of Shor's ellipsoid algorithm for convex programming*, Math. Prog. 27 (1983), str. 83–106.

[2] D. Goldfarb, M. J. Todd, *Modifications and implementations of the Shor–Khachian algorithm for linear programming*, Math. Prog. 23 (1982), str. 1–19.

[3] F. John, *Extremum problems with inequalities as subsidiary conditions*, Courant Anniv. Vol., Interscience, N. Y. (1948), str. 187–204.

[4] L. G. Chacijan, *Polinomijal'nyj algoritm v linejnom programmirovanii*, Dokl. A. N. SSSR 224 (1979), str. 1093–1096.

[5] H. König, D. Pallaschke, *On Khachian's algorithm and minimal ellipsoids*, Num. Math. 36 (1981), str. 211–223.

[6] P. Recht, *An Objective-Function Ellipsoid Algorithm for Convex*, Quadratical Programming, Optimization 17 (1986), str. 333–347.

[7] N. Z. Shor, *Generalized Gradient Methods of Nondifferentiable Optimization Employing Space Dilatation Operations*, Math. Programming – State of the Art, ed. A. Bachem, M. Grötschel, B. Korte, Springer-Verlag (1983), str. 501–529.

[8] B. Waluk, S. Walukiewicz, *Wyniki eksperymentu obliczeniowego z algorytmem elipsoidalnym*, Archiwum Automatyki i Telemechaniki 35 (1983), str. 239–257.