

Zastosowanie metod projekcyjnych w optymalizacji wypukłej nieróżniczkowalnej

Robert Dylewski

Streszczenie: W artykule przedstawiono metody projekcyjne stosowane w optymalizacji wypukłej nieróżniczkowalnej. Opisano problemy optymalizacji, w których pojawiają się funkcje nieróżniczkowalne. Przedstawiono wyniki uzyskane metodą projekcyjną z selekcją residualną dla problemu membrany.

Słowa kluczowe: metody projekcyjne, optymalizacja wypukła nieróżniczkowalna

1. WPROWADZENIE

W pracy rozważamy problem minimalizacji wypukłej bez ograniczeń postaci:

$$\begin{array}{ll} \text{minimalizować} & f(x) \\ \text{względem} & x \in R^n, \end{array} \quad (1)$$

gdzie funkcja $f: R^n \rightarrow R$ jest wypukła, niekoniecznie różniczkowalna.

Zakładamy, że zbiór rozwiązań optymalnych dla tego problemu $M^* = \{z \in R^n : f(z) \leq f(x) \text{ dla każdego } x \in R^n\}$ jest niepusty, czyli funkcja f osiąga minimum globalne f^* .

Jeżeli funkcja f jest różniczkowalna dla każdego $x \in R^n$, to: $x^* \in R^n$ jest rozwiązaniem optymalnym ($x^* \in M^*$) problemu (1) wtedy i tylko wtedy, gdy $\nabla f(x^*) = 0$. Funkcja wypukła jest różniczkowalna prawie wszędzie. Jeżeli natomiast funkcja wypukła f posiada w dziedzinie punkty, w których nie jest różniczkowalna (może ich być skończona liczba), to w punktach tych można wyznaczyć zamiast gradientu tzw. subróżniczkę $\partial f(x) = \{g \in R^n : f(y) - f(x) \geq \langle g, y - x \rangle, y \in R^n\}$, ($\langle x, y \rangle = x^T y$ oznacza standardowy iloczyn skalarny wektorów $x, y \in R^n$). Element $g_f(x) \in \partial f(x)$ nazywa się subgradientem funkcji f w punkcie x . Oczywiście, jeżeli f jest różniczkowalna w x , to subróżniczką w tym punkcie jest zbiorem jednoelementowym i $\partial f(x) = \{\nabla f(x)\}$.

Podsumowując, warunek konieczny i wystarczający optymalności dla problemu (1) ma postać [14]:

$$x^* \in M^* \Leftrightarrow 0 \in \partial f(x^*). \quad (2)$$

Dla zadań minimalizacji wypukłej (1) stawia się minimalne wymagania dotyczące informacji o funkcji celu f : dla każdego $x \in R^n$ można wyznaczyć wartość $f(x)$ funkcji celu f i jej (jeden) subgradient $g \in \partial f(x)$.

Klasyczne metody minimalizacji zastosowane bezpośrednio do problemu (1) nie są w ogólności zbieżne do rozwiązania optymalnego, nawet gdy funkcja celu jest różniczkowalna w każdym punkcie będącym kolejnym przybliżeniem rozwiązania. Może też się zdarzyć, że otrzymamy zbieżność do punktu, który nie jest rozwiązaniem optymalnym (przykład 3.3 w pracy [14]).

Dla funkcji różniczkowalnej gradient jest „mały” w pobliżu punktu realizującego jej minimum. Dla klasycznych metod minimalizacji różniczkowalnej można przyjąć kryterium zatrzymania postaci $\|\nabla f(x_k)\| < \varepsilon$ dla

zadanej tolerancji optymalności ε ($\|x\| = \sqrt{x^T x}$ oznacza normę euklidesową wektora $x \in R^n$). Kryterium tej postaci nie można jednak stosować w przypadku minimalizacji nieróżniczkowalnej, nawet jeśli gradient funkcji celu zastąpi się jej subgradientem. Subgradient bowiem nie musi być i najczęściej nie jest „mały” w pobliżu rozwiązania optymalnego. Wystarczy rozpatrzyć prosty przykład $f(x) = |x| = \max\{-x, x\}$. Dla $x \neq 0$ gradient jest równy $+1$ albo -1 , natomiast $\partial f(0) = [-1, 1]$.

Występuje jeszcze inny problem. Dla funkcji różniczkowalnej kierunek $t = -\nabla f(x)$ jest kierunkiem najszybszego spadku w punkcie x , natomiast dla funkcji wypukłej, która nie jest różniczkowalna w punkcie x , kierunek $t = -g$, gdzie $g \in \partial f(x)$ nie musi być nawet kierunkiem spadku (wektor t jest kierunkiem spadku funkcji f w punkcie x w kierunku t , jeśli istnieje $\tau > 0$ takie, że dla każdego $\lambda \in (0, \tau)$ zachodzi $f(x + \lambda t) < f(x)$).

2. METODY PROJEKCYJNE DLA PROBLEMÓW MINIMALIZACJI WYPUKŁEJ

W punkcie tym przedstawiamy metody projekcyjne, które zastosowane do problemów postaci (1) gwarantują zbieżność do rozwiązania optymalnego. Metody te mają postać procedury iteracyjnej generującej ciąg (x_k) kolejnych przybliżeń postaci:

$$\begin{array}{l} x_1 \in R^n - \text{dowolny} \\ x_{k+1} = x_k + \lambda_k t_k, \end{array} \quad (3)$$

gdzie:

- $\lambda_k \in (0,2)$ jest tzw. współczynnikiem relaksacyjnym;
- Wektor rzutowy t_k jest kierunkiem poszukiwań postaci

$$t_k = P_{S_k}(x_k) - x_k \quad (4)$$

($P_S(x) = \arg \min_{y \in S} \|y - x\|$ oznacza projekcję metryczną punktu x na zbiór S);

- zbiór S_k jest postaci

$$S_k = \bigcap_{i \in L_k} S(f_i, \alpha_k), \quad (5)$$

gdzie: $S(f_i, \alpha_k) = \{x \in R^n : f_i(x) \leq \alpha_k\}$ jest podpoziomicą funkcji f_i dla poziomu α_k , a $f_i(x) = \langle g_f(x_i), x - x_i \rangle + f(x_i)$ jest linearyzacją funkcji f w punkcie x_i ;

- $L_k \subset \{1, 2, \dots, k\}$ jest zbiorem wybranych linearyzacji takim, że $k \in L_k$;
- poziom

$$\alpha_k = (1 - \mu)\bar{\alpha}_k + \mu\underline{\alpha}_k \quad (6)$$

jest przybliżeniem wartości minimalnej f^* funkcji celu f ;

- $\mu \in (0, 1]$ jest parametrem poziomu;
- $\bar{\alpha}_k$ jest ograniczeniem górnym wartości f^*

$$\bar{\alpha}_k = \min_{1 \leq i \leq k} f(x_i); \quad (7)$$

- $\underline{\alpha}_k$ jest ograniczeniem dolnym wartości f^* aktualizowanym w trakcie procesu iteracyjnego

$$\underline{\alpha}_k \leq f^*. \quad (8)$$

Dodatkowo zakłada się, że znane jest ograniczenie górne R odległości punktu startowego x_1 od zbioru rozwiązań M^* , ($R \geq d(x_1, M^*)$) i początkowe ograniczenie dolne $\underline{\alpha}_1$ wartości minimalnej f^* funkcji f .

Jeżeli nie znane jest lepsze początkowe ograniczenie dolne $\underline{\alpha}_1$ wartości f^* , to można przyjąć $\underline{\alpha}_1 = f(x_1) - \|g_f(x_1)\|R$ (uwaga 2c) w pracy [2]).

Poszczególne metody różnią się przede wszystkim wyborem zbioru L_k i sposobem aktualizacji ograniczenia dolnego $\underline{\alpha}_k$ wartości f^* . Jeżeli znana jest wartość f^* , to w (6) można przyjąć $\mu = 1$, $\underline{\alpha}_1 = f^*$ i poziom $\alpha_k = f^*$. Jeżeli dodatkowo $L_k = \{k\}$, to otrzymujemy metodę Polyaka [13], gdzie wektor rzutowy $t_k = -(f(x_k) - f^*)g_f(x_k) / \|g_f(x_k)\|^2$. W przypadku, gdy f^* nie jest znane ($\mu \in (0, 1)$), sposób aktualizacji $\underline{\alpha}_k$ wykorzystujący informację o R podano w [9]. W pracy [10] pokazano, że jeżeli L_k jest pewnym podzbiorem zbioru $\{1, 2, \dots, k\}$ takim, że $k \in L_k$, to przy odpowiedniej aktualizacji ograniczenia dolnego $\underline{\alpha}_k$ (wykorzystującej informację o R) ciąg x_k jest zbieżny do rozwiązania. W [11] zaproponowano metodę, w której do zbioru L_k brane są wszystkie otrzymane linearyzacje $L_k = \{1, 2, \dots, k\}$. W pracy [2] wprowadzono model selekcji

stożka rozwartego do wyboru zbioru L_k i w konsekwencji konstrukcji wektora t_k , a w pracach [3] i [4] wprowadzono tzw. model selekcji residualnej, dla której selekcja stożka rozwartego jest szczególnym przypadkiem. Wykorzystując model selekcji residualnej albo selekcji stożka rozwartego można dodatkowo wykrywać kiedy poziom α_k jest za niski ($\alpha_k < f^*$) w przypadku kiedy wyznaczony zbiór $S_k = \emptyset$ (korzystając z rozkładu Cholesky'ego macierzy Grama $G_L^T G_L$ kolumn macierzy $G_L = [g_f(x_i) : i \in L_k]$). Jeżeli stwierdzi się, że $\alpha_k < f^*$, to w kolejnej iteracji można wykonać aktualizację ograniczenia dolnego $\underline{\alpha}_{k+1} = \alpha_k$.

W podanych metodach projekcyjnych wektor t_k niekoniecznie jest spadkowy, ale w przypadku, gdy $\alpha_k \geq f^*$ dla każdego $k \geq 1$, to ciąg (x_k) jest monotoniczny w sensie Fejéra względem zbioru M^* , tzn. $\|x_{k+1} - x\| \leq \|x_k - x\|$ dla dowolnego $x \in M^*$ i dla dowolnego $k \geq 1$.

W prezentowanych metodach projekcyjnych możemy stosować kryterium zatrzymania postaci $\bar{\alpha}_k - \underline{\alpha}_k \leq \varepsilon$ dla bezwzględnej tolerancji optymalności ε . Po spełnieniu takiego kryterium zatrzymania, dla punktu $\bar{x}_k = \arg \min_{1 \leq i \leq k} f(x_i)$ zachodzi $f(\bar{x}_k) \leq f^* + \varepsilon$, zatem $\bar{x}_k$ jest tzw. rozwiązaniem ε -optymalnym dla problemu (1).

3. PRZYKŁADY PROBLEMÓW OPTYMALIZACJI WYPUKŁEJ NIERÓŻNICZKOWALNEJ

Problemy postaci (1) pojawiają się w optymalizacji w sposób naturalny. W punkcie tym podajemy kilka przykładów problemów, które mają postać (1) i występują w nich funkcje nieróżniczkowalne.

3.1. PROBLEM DOPUSZCZALNOŚCI LINIOWEJ

Problem dopuszczalności liniowej ma postać:

$$\begin{aligned} &\text{znaleźć} && x \in R^n \\ &\text{spełniający} && A^T x \leq b, \end{aligned} \quad (9)$$

gdzie: A jest macierzą typu $n \times m$ i $b = (b_1, \dots, b_m)^T \in R^m$, o ile taki x istnieje.

Można też rozpatrywać problem postaci:

$$\begin{aligned} &\text{minimalizować} && f(x) = \max\{A_1^T x - b_1, \dots, A_m^T x - b_m, 0\} \\ &\text{względem} && x \in R^n, \end{aligned} \quad (10)$$

gdzie: A_i oznacza i -tą kolumnę macierzy A , $i \in \{1, \dots, m\}$. Otrzymana w (10) funkcja f jest wypukła, ale może być nieróżniczkowalna.

Jeżeli wiadomo, że zbiór rozwiązań $M_0 = \{x : A^T x \leq b\}$ problemu (9) jest niepusty, to problem (9) jest równoważny problemowi (10) ze znaną wartością minimalną $f^* = 0$. Jeżeli natomiast nie wiemy, czy

$M_0 \neq \emptyset$, to wyznaczając rozwiązanie problemu (10) metodą projekcyjną opisaną w punkcie 2, możemy przyjąć $\alpha_1 = 0 \leq f^*$. Rozwiązanie ε -optymalne $\bar{x}_k$ problemu (10) spełnia wtedy warunek

$$A_i^T \bar{x}_k - b_i \leq f^* + \varepsilon \quad (11)$$

dla każdego $i \in \{1, \dots, m\}$, gdzie f^* jest minimalną wartością, dla której spełniony jest ten warunek. Oczywiście, jeżeli $M_0 \neq \emptyset$, to powyższy warunek przyjmuje postać $A_i^T \bar{x}_k - b_i \leq \varepsilon$ dla każdego $i \in \{1, \dots, m\}$.

W problemach postaci (9) często $M_0 = \emptyset$ albo nie wiemy, czy $M_0 = \emptyset$. W większości metod stosowanych bezpośrednio do problemu (9) mamy zagwarantowaną zbieżność, tylko gdy $M_0 \neq \emptyset$ (patrz np. [1]), ewentualnie możemy wykryć, że $M_0 = \emptyset$ w trakcie procesu iteracyjnego (patrz np. [5]). Stosując metodę projekcyjną do problemu (10) nie musimy wiedzieć, czy $M_0 \neq \emptyset$, a otrzymane rozwiązanie ma pożądaną własność (11).

Problem postaci (9) pojawia się np. w radioterapii. W zagadnieniu planowania radioterapii z użyciem wiązek o modulowanej intensywności (intensity modulated radiation therapy – IMRT) trzeba ułożyć plan naświetleń promieniami rentgenowskimi mający na celu zniszczenie guza nie niszcząc jednocześnie zdrowych organów. Model matematyczny otrzymany w wyniku dyskretyzacji problemu IMRT ma właśnie postać (9), gdzie układ ograniczeń w postaci nierówności liniowych najczęściej jest sprzeczny. Należy określić wektor intensywności radiacji „najbliższy” spełnieniu tych ograniczeń (szczegóły można znaleźć w [8]). Podobnie, w zagadnieniu tomografii komputerowej (wyznaczanie przekroju pewnego obiektu) otrzymuje się model matematyczny postaci (9) z macierzą rzadką A dużych wymiarów (patrz np. [7]).

3.2. PROBLEM MINIMALIZACJI WYPUKŁEJ Z OGRANICZENIAM

Problem minimalizacji wypukłej z ograniczeniami ma postać:

$$\begin{aligned} &\text{minimalizować} && f(x) \\ &\text{względem} && x \in R^n \\ &\text{przy ograniczeniach} && c_i(x) \leq 0, i \in I, \end{aligned} \quad (12)$$

gdzie: $f, c_i : R^n \rightarrow R, i \in I = \{1, \dots, m\}$ są funkcjami wypukłymi. Problem (12) można sprowadzić do problemu bez ograniczeń (1), wprowadzając tzw. dokładną funkcję kary $\Phi(x) = \alpha f(x) + \|c^+(x)\|$, gdzie:

$$\begin{aligned} c^+(x) &= [c_1^+(x), \dots, c_m^+(x)]^T, \quad c_i^+(x) = \max\{c_i(x), 0\}, i \in I, \\ \alpha &> 0, \text{ zaś } \|\cdot\| \text{ oznacza jakąkolwiek normę w } R^n, \text{ na} \\ &\text{przykład normę } l_1 \text{ lub } l_\infty. \text{ W przypadku normy } l_\infty \text{ problem} \\ &\text{minimalizacji dokładnej funkcji kary przyjmuje postać:} \\ &\text{minimalizować} && \Phi(x) = \alpha f(x) + \max\{0, c_i(x) : i \in I\} \\ &\text{względem} && x \in R^n. \end{aligned} \quad (13)$$

Dla dostatecznie małego $\alpha > 0$ przy pewnych założeniach dotyczących funkcji ograniczeń $c_i(x), i = \{1, \dots, m\}$ problemy (13) i (12) są sobie

równoważne. Szczegóły można znaleźć w [6] (ustęp 14.3). Zaletą powyższego podejścia jest sprowadzenie trudniejszego problemu minimalizacji z ograniczeniami, do prostszego problemu minimalizacji bez ograniczeń. Ceną za to jest utrata różniczkowalności. Dokładna funkcja kary jest najczęściej nieróżniczkowalna, nawet jeśli funkcje $f, c_i, i \in I$ są różniczkowalne. Natomiast dokładna funkcja kary jest wypukła, jeśli funkcje $f, c_i, i \in I$ są wypukłe.

3.3. PROBLEM MEMBRANY

Problem membrany można przedstawić, przy pomocy równania Laplace’a z warunkiem brzegowym, w postaci:

$$\begin{aligned} &\text{wyznaczyć } u \text{ takie, że} \\ &-\Delta u = f \text{ na } \Omega \\ &u = 0 \text{ na } \partial\Omega, \end{aligned} \quad (13)$$

gdzie: $\Omega = [0,1] \times [0,1]$, $\partial\Omega$ oznacza brzeg obszaru Ω ,

$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$. W równaniu tym f oznacza siłę działającą

prostopadle do elastycznej membrany, zaś u ugięcie membrany pod wpływem działającej siły.

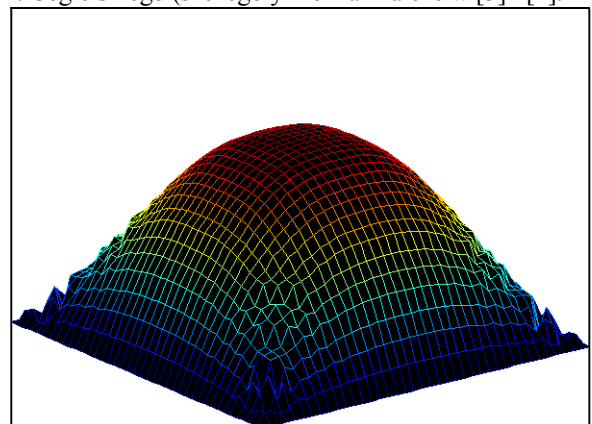
Problem (13) można sformułować równoważnie jako problem minimalizacji postaci:

$$\begin{aligned} &\text{minimalizować} && \frac{1}{2} a(u, u) - (f, u) \\ &\text{względem} && u \in K, \end{aligned} \quad (14)$$

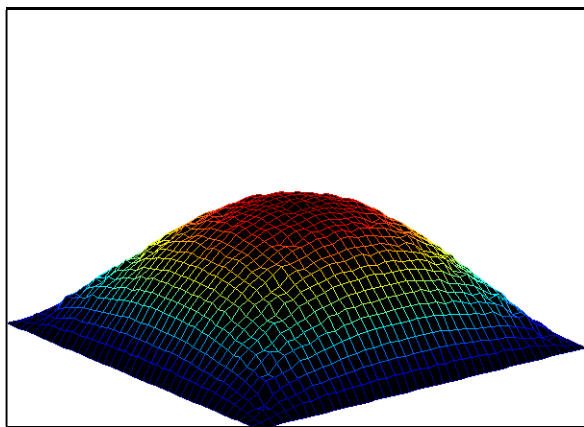
gdzie: $a(u, w) = \int_{\Omega} \nabla u^T \nabla w$ i $K = \{u : u = 0 \text{ na } \partial\Omega\}$.

Natomiast problem (14) można rozwiązać przez dyskretyzację, dzieląc obszar Ω na kwadraty i zastępując u i f przez ich kawałkami liniowe aproksymacje. Po dyskretyzacji otrzymujemy problem postaci (1) z kwadratową funkcją celu $p : R^{N+1} \times R^{N+1} \rightarrow R$, gdzie $N \times N$ jest liczbą kwadratów, na którą został podzielony obszar Ω . Szczegóły można znaleźć w [12].

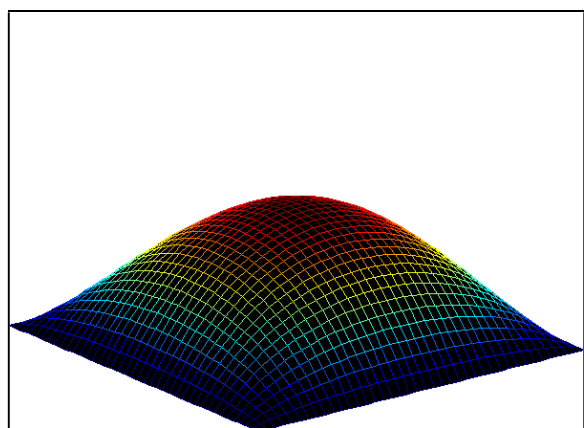
Na koniec prezentujemy wyniki dla problemu membrany, w którym przyjęto siłę $f = 1$ i $N = 39$, czyli wymiar zadania po dyskretyzacji wynosi $40 \times 40 = 1600$. Do wyznaczenia rozwiązania użyto metody projekcyjnej opisaną przez (3, 4, 5, 6, 7, 8), gdzie do wyboru zbioru L_k (i w konsekwencji konstrukcji wektora t_k) zastosowano metodę selekcji residualnej opracowaną przez autora i A. Cegielskiego (szczegóły można znaleźć w [3] i [4]).



Rys. 1. Uzyskane rozwiązanie dla $k = 34$, $\varepsilon = 0.1130$



Rys. 2. Uzyskane rozwiązanie dla $k = 305$, $\varepsilon = 0.0126$



Rys. 3. Uzyskane rozwiązanie dla $k = 1047$, $\varepsilon = 10^{-6}$

Implementację wykonano w Fortranie 90 (Lahey Fortran 90 v. 3.5), a obliczenia były przeprowadzone przy podwójnej precyzji. Do graficznego przedstawienia rozwiązania wykorzystano pakiet matematyczny Matlab.

Przykładowe wyniki przedstawiono na rysunkach 1, 2 i 3. Przez k oznaczono liczbę wyznaczeń wartości funkcji celu i jej subgradientu, a przez ε bezwzględną tolerancję optymalności, dla której spełnione było kryterium zatrzymania $\bar{\alpha}_k - \underline{\alpha}_k \leq \varepsilon$.

4. PODSUMOWANIE

W artykule przedstawiono metody projekcyjne służące rozwiązywaniu problemów minimalizacji wypukłej nieróżniczkowalnej. Opisano problemy z optymalizacji, w których występują funkcje nieróżniczkowalne. Dla przykładowego problemu membrany wyznaczono rozwiązanie tego problemu, wykorzystując tzw. dyskretyzację. Zastosowano metodę projekcyjną z modelem selekcji residualnej linearyzacji wprowadzonym w pracach [3, 4]. Dla metody projekcyjnej z selekcją residualną można zaobserwować, w problemach testowanych przez autora, liniową zbieżność do rozwiązania. Wiadomo też, że metody projekcyjne należą do klasy najefektywniejszych metod w minimalizacji wypukłej nieróżniczkowalnej [11].

LITERATURA

- [1] Cegielski A., *Metody relaksacyjne w problemach optymalizacji wypukłej*. Zielona Góra: Monografie 67, IM WSI, 1993.
- [2] Cegielski A., „A method of projection onto an acute cone with level control in convex minimization”, *Mathematical Programming*, 85, ss. 469-490, 1999.
- [3] Cegielski A., Dylewski R., „Selection strategies in projection methods for convex minimization problems”, *Discussiones Mathematicae. Differential Inclusions, Control and Optimization*, 22, ss. 97-123, 2002.
- [4] Cegielski A., Dylewski R., „Residual selection in a projection method for convex minimization problems”, *Optimization*, 52, ss. 211-220, 2003.
- [5] Dylewski R., „Projection method with residual selection for linear feasibility problems”, *Discussiones Mathematicae. Differential Inclusions, Control and Optimization*, 27, ss. 43-50, 2007.
- [6] Fletcher R., *Practical Methods of optimization*, vol. 2, Chichester, New York, Brisbane, Toronto: John Wiley & Sons, 1981.
- [7] Herman G.T., „A relaxation method for reconstructing objects from noisy X-rays”, *Mathematical Programming*, 8, ss. 359-369, 1975.
- [8] Hamacher H.W., Küfer K.-H., „Inverse raduation therapy planning – a multiple objective optimization approach”, *Discrete Applied Mathematics*, 118, ss. 145-161, 2002.
- [9] Kim S., Ahn H., Cho S.-C., „Variable target value subgradient method”, *Mathematical Programming*, 49, ss. 359-369, 1991.
- [10] Kiwiel K.C., „The efficiency of subgradient projection methods for convex optimization, part I: General level methods”, *SIAM J. Control and Optimization*, 34, ss. 660-676, 1996.
- [11] Lemaréchal C., Nemirovskii A.S., Nesterov Yu.E., „New variants of bundle methods”, *Mathematical Programming*, 69, ss. 111-147, 1995.
- [12] Outrata J., Kocvara M., Zowe J., *Nonsmooth approach to optimization problems with equilibrium constraints*, Dordrecht: Kluwer, 1998.
- [13] Polyak B.T., „Minimization of unsmooth functionals”, *Zh. Vychisl. Mat. Mat. Fiz.*, 9, ss. 509-521, 1969.
- [14] Schramm H., Zowe J., „A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results”, *SIAM J. Optimization*, 2, ss. 121-152, 1992.



dr Robert Dylewski

Uniwersytet Zielonogórski
Wydział Matematyki, Informatyki i
Ekonometrii

ul. prof. Z. Szafrana 4a
65-516 Zielona Góra

tel.: 68 328 28 75

e-mail: r.dylewski@wmie.uz.zgora.pl