ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria II: WIADOMOŚCI MATEMATYCZNE XXIX (1990)

wmptm 1897

Ministerstwo Nauki i Szkolnictwa Wyższego

DIETHARD PALLASCHKE (Karlsruhe)
RYSZARD URBAŃSKI (Poznań)

Funkcje quasi-różniczkowalne w nieróżniczkowej teorii optymalizacji

Wstęp. Praca zawiera przegląd wyników z teorii funkcji quasi-różniczkowalnych, które zostały wprowadzone przez W. F. Demianowa, A. N. Rubinowa oraz L. W. Wasiliewa w pracach [3] i [4]. Funkcje quasi-różniczkowalne na ogół nie są różniczkowalne, lecz pomimo to zachodzą dla nich reguły będące naturalnym rozszerzeniem klasycznego rachunku różniczkowego. Rozszerzenie to bazuje na pojęciu subgradientu funkcji wypukłej. W szczególności każda funkcja wypukła lub różniczkowalna jest quasi-różniczkowalna. Tak więc funkcje typu skończonego mini-maksu funkcji różniczkowalnych są quasi-różniczkowalne. Ponadto, wymierne kombinacje funkcji quasi-różniczkowalnych są również quasi-różniczkowalne.

Pojęcie funkcji quasi-różniczkowalnej jest jednym z wielu znanych uogólnień pojęcia pochodnej. Autorzy uważają, że to właśnie uogólnienie jest szczególnie wygodne w teorii optymalizacji, i dlatego je tu przedstawiają.

1. Funkcje quasi-różniczkowalne. Niech $U \subset \boldsymbol{R}^n$ będzie otwartym podzbiorem przestrzeni n-wymiarowej, $x_0 \in U$, $f: U \to \boldsymbol{R}$ jest ciągłą funkcją rzeczywistą. Mówimy, że f jest *quasi-różniczkowalna w punkcie* x_0, gdy spełnia następujące dwa warunki:

(i) dla dowolnego $g \in \boldsymbol{R}^n$ istnieje pochodna kierunkowa

$$\left.\frac{df}{dg}\right|_{x_0} := \lim_{\substack{\alpha \to 0 \\ \alpha > 0}} \frac{f(x_0+\alpha g)-f(x_0)}{\alpha},$$

(ii) istnieją dwa zwarte zbiory wypukłe

$$\underline{\partial} f|_{x_0},\ \overline{\partial} f|_{x_0} \subset \boldsymbol{R}^n$$

takie, że pochodna kierunkowa rozważana jako funkcja kierunku może być przedstawiona w postaci:

$$g \mapsto \left.\frac{df}{dg}\right|_{x_0} = \sup_{v \in \underline{\partial} f|_{x_0}} \langle v, g\rangle + \inf_{w \in \overline{\partial} f|_{x_0}} \langle w, g\rangle,$$

gdzie $\langle\ ,\ \rangle$ oznacza iloczyn skalarny w przestrzeni $\boldsymbol{R}^n$.

Jeśli funkcja f jest quasi-różniczkowalna w dowolnym punkcie zbioru U, to mówimy, że jest *quasi-różniczkowalna*. Zbiór $\underline{\partial} f|_{x_0}$ nazywamy *subróżniczką* funkcji f w punkcie x_0, a $\overline{\partial} f|_{x_0}$ nazywamy *superróżniczką* f w punkcie x_0. Parę

$$Df|_{x_0} := (\underline{\partial} f|_{x_0}, \overline{\partial} f|_{x_0})$$

nazywamy *quasi-różniczką* f w punkcie x_0.

Podamy teraz przykłady funkcji quasi-różniczkowalnych.

(a) Niech $U \subset \boldsymbol{R}^n$ będzie zbiorem otwartym oraz $f\colon U \to \boldsymbol{R}$ będzie funkcją różniczkowalną w U. Wówczas z reguły łańcuchowej

$$\frac{d}{d\alpha} f(x+\alpha g) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} g_i$$

wynika, że dla dowolnego wektora $g \in \boldsymbol{R}^n$ i $x_0 \in U$ pochodna kierunkowa jest równa

$$\left.\frac{df}{dg}\right|_{x_0} = \langle \nabla f|_{x_0}, g\rangle,$$

gdzie

$$\nabla f|_{x_0} = \left(\left.\frac{\partial f}{\partial x_1}\right|_{x_0}, \ldots, \left.\frac{\partial f}{\partial x_n}\right|_{x_0}\right)$$

oznacza gradient funkcji f w punkcie x_0. Weźmy teraz

$$\underline{\partial} f|_{x_0} = \{\nabla f|_{x_0}\}, \quad \overline{\partial} f|_{x_0} = \{0\};$$

wówczas drugi warunek quasi-różniczkowalności jest spełniony.

Powyższy prosty przykład pokazuje już charakterystyczną trudność występującą w teorii funkcji quasi-różniczkowalnych. Mianowicie, sub- oraz superróżniczka nie są określone w sposób jednoznaczny: dla

$$\underline{\partial} f|_{x_0} = \{0\} \quad \text{oraz} \quad \overline{\partial} f|_{x_0} = \{\nabla f|_{x_0}\}$$

druga część definicji quasi-różniczkowalności jest również spełniona. Widać stąd, że w teorii funkcji quasi-różniczkowalnych rozważa się tylko klasy równoważności par zbiorów. Mówiąc bardziej precyzyjnie, para (A, B) jest równoważna z (C, D), jeśli $A+D = B+C$ (por. [10]).

Rozważmy teraz następujący przykład:

(b) Jeśli $U \subset \boldsymbol{R}^n$ jest otwartym podzbiorem wypukłym przestrzeni $\boldsymbol{R}^n$ oraz $f\colon U \to \boldsymbol{R}$ jest funkcją wypukłą, wówczas subróżniczka w sensie analizy wypukłej [5] jest równa

$$\partial f|_{x_0} = \{v \in \boldsymbol{R}^n |\ \langle v, (z-x_0)\rangle \leqslant f(z)-f(x_0) \text{ dla wszystkich } z \in U\}.$$

Ponieważ dla każdego $g \in \boldsymbol{R}^n$ (zob. [5])

$$\left.\frac{df}{dg}\right|_{x_0} = \sup_{v \in \partial f|_{x_0}} \langle v, g\rangle,$$

możemy więc przyjąć

$$\underline{\partial} f|_{x_0} = \partial f|_{x_0}, \qquad \overline{\partial} f|_{x_0} = \{0\}.$$

Zatem każda funkcja wypukła jest quasi-różniczkowalna.

(c) Z ostatniego przykładu bezpośrednio wynika, że różnica dwóch funkcji wypukłych jest również quasi-różniczkowalna. Niech $U \subset \boldsymbol{R}^n$ będzie wypukłym zbiorem otwartym oraz niech $f_1, f_2\colon U \to \boldsymbol{R}$ będą funkcjami wypukłymi. Jeśli $g \in \boldsymbol{R}^n$ i $x_0 \in U$, wówczas dla $f := f_1 - f_2$ mamy

$$\left.\frac{df}{dg}\right|_{x_0} = \sup_{v \in \partial f_1|_{x_0}} \langle v, g\rangle - \sup_{w \in \partial f_2|_{x_0}} \langle w, g\rangle.$$

Zatem dla

$$\underline{\partial} f|_{x_0} = \partial f_1|_{x_0}, \qquad \overline{\partial} f|_{x_0} = -\partial f_2|_{x_0}$$

warunek quasi-różniczkowalności jest spełniony.

2. Rachunek quasi-różniczkowy. Istotna przewaga określonych wyżej pojęć nad innymi rozszerzeniami rachunku różniczkowego dla pewnych klas funkcji nieróżniczkowalnych [1] polega na tym, że dla funkcji quasi-różniczkowalnych prawa rachunku różniczkowego są rzeczywiście spełnione. Mówiąc bardziej precyzyjnie, reguły różniczkowania są wyrażone za pomocą równości, a nie tylko przez inkluzje.

W celu przedstawienia tych reguł zdefiniujmy dodawanie par podzbiorów zwartych przestrzeni $\boldsymbol{R}^n$

$$(A, B)+(C, D) := (A+C, B+D),$$

oraz mnożenie pary zbiorów (A, B) przez skalar $\alpha \in \boldsymbol{R}$:

$$\alpha(A, B) = \begin{cases} (\alpha A, \alpha B), & \alpha \geqslant 0, \\ (-\alpha B, -\alpha A), & \alpha < 0. \end{cases}$$

Stosując powyższe działania, otrzymamy następujące reguły różniczkowania ([3], [4]).

TWIERDZENIE 1. *Niech* $U \subset \boldsymbol{R}^n$ *będzie zbiorem otwartym, niech* $f_1, f_2\colon U \to \boldsymbol{R}$ *będą ciągłymi funkcjami quasi-różniczkowalnymi w punkcie* $x_0 \in U$ *i niech* $\alpha, \beta \in \boldsymbol{R}$. *Wówczas zachodzą następujące reguły*:

(i) *Liniowość*:

$$D(\alpha f_1 + \beta f_2)|_{x_0} = \alpha Df_1|_{x_0} + \beta Df_2|_{x_0}.$$

(ii) *Reguła iloczynu*:

$$D(f_1 f_2)|_{x_0} = f_1(x_0) Df_2|_{x_0} + f_2(x_0) Df_1(x_0).$$

(iii) *Reguła ilorazu*:

$$D\left(\frac{f_1}{f_2}\right)\Big|_{x_0} = \frac{1}{(f_2(x_0))^2}\left(f_2(x_0) Df_2|_{x_0} - f_1(x_0) Df_2(x_0)\right),$$

przy założeniu, że $f_2(x_0) \neq 0$.

Zauważmy, że spełniony jest również wariant reguły łańcuchowej [3]. Ponieważ quasi-różniczka nie jest wyznaczona w sposób jednoznaczny, więc powyższe reguły opisują działania na klasach równoważności.

Reguły różniczkowania są interesującym faktem w teorii funkcji quasi-różniczkowalnych. Dla lepszego zrozumienia, dlaczego reguły te są spełnione, spójrzmy na definicję quasi-różniczkowalności z innego punktu widzenia.

Najbardziej istotnym warunkiem w definicji quasi-różniczkowalności jest istnienie zwartych zbiorów wypukłych $\underline{\partial} f|_{x_0}$, $\overline{\partial} f|_{x_0} \subset \boldsymbol{R}^n$ takich, że

$$g \mapsto \frac{df}{dg}\Big|_{x_0} = \sup_{v \in \underline{\partial} f|_{x_0}} \langle v, g\rangle + \inf_{w \in \overline{\partial} f|_{x_0}} \langle w, g\rangle.$$

Ponieważ każdej funkcji subliniowej $p\colon \boldsymbol{R}^n \to \boldsymbol{R}$, tj. takiej, że

(α) $p(\alpha x) = \alpha p(\lambda)$ dla $x \in \boldsymbol{R}^n$, $\alpha \geqslant 0$,

(β) $p(x+y) \leqslant p(x) + p(y)$ dla $x, y \in \boldsymbol{R}^n$,

odpowiada dokładnie jeden zwarty zbiór wypukły $K \subset \boldsymbol{R}^n$ taki, że

$$p(x) = \sup_{v \in K} \langle v, x\rangle \qquad ([9])$$

oraz

$$\inf_{w \in \overline{\partial} f|_{x_0}} \langle w, g\rangle = - \sup_{w \in -\underline{\partial} f|_{x_0}} \langle w, g\rangle,$$

widzimy, że w teorii quasi-różniczkowalności istotny jest fakt, iż pochodna kierunkowa może być reprezentowana przez różnicę dwóch funkcji subliniowych.

Niech więc

$$\mathscr{P} := \{p\colon \boldsymbol{R}^n \to \boldsymbol{R} \mid p \text{ jest subliniowe}\},$$

oznacza zbiór funkcji subliniowych określonych na $\boldsymbol{R}^n$ i niech $\mathscr{D}(\boldsymbol{R}^n) = \mathscr{P} - \mathscr{P}$. Wówczas warunek (ii) quasi-różniczkowalności można zapisać inaczej:

$$g \mapsto \frac{df}{dg}\Big|_{x_0} \in \mathscr{D}(\boldsymbol{R}^n).$$

Dla funkcji spełniających lokalnie warunek Lipschitza prawdziwe jest następujące twierdzenie:

TWIERDZENIE 2. *Niech $U \subset \boldsymbol{R}^n$ będzie zbiorem otwartym, a $f\colon U \to \boldsymbol{R}$ funkcją spełniającą lokalnie warunek Lipschitza. Niech $x_0 \in U$. Wówczas f jest quasi-różniczkowalna w x_0 wtedy i tylko wtedy, gdy istnieje element $T \in \mathscr{D}(\boldsymbol{R}^n)$ o następującej własności: dla każdego $\varepsilon > 0$ istnieje $\delta > 0$ takie, że dla wszystkich $g \in U$ z $\|g\| \leqslant \delta$ i $x_0 + g \in U$ zachodzi nierówność*

$$|f(x_0+g) - f(x_0) - T(g)| \leqslant \varepsilon \|g\|.$$

Dowód znajduje się w pracy [8].

Powyższe równoważne sformułowanie quasi-różniczkowalności pokrywa się z definicją różniczkowalności w sensie Frécheta, z wyjątkiem faktu, że T nie jest odwzorowaniem liniowym, lecz różnicą dwóch odwzorowań subliniowych. Łatwo wykazać, że odwzorowanie $T \in \mathscr{D}(\boldsymbol{R}^n)$ jest określone w sposób jednoznaczny. Będziemy je oznaczać przez $df|_{x_0}$, tzn. $df|_{x_0} = T$ jest quasi-różniczką funkcji f w punkcie x_0.

Bardzo znanym przykładem funkcji nieróżniczkowalnej w zerze jest funkcja $f\colon \boldsymbol{R}^2 \to \boldsymbol{R}$ zdefiniowana następująco:

$$f(x, y) = \begin{cases} \dfrac{x^2 y}{x^2 + y^2} & \text{dla } (x, y) \neq (0, 0), \\ 0 & \text{dla } x = y = 0. \end{cases}$$

Można udowodnić, że funkcja ta jest quasi-różniczkowalna.

Stożek złożony z klas abstrakcji par wypukłych, zwartych podzbiorów $\boldsymbol{R}^n$ jest izomorficzny ze stożkiem, którego elementami są różnice funkcji subliniowych określonych na $\boldsymbol{R}^n$ ([5]).

3. Kierunek największego spadku. Głównym problemem teorii optymalizacji jest wyznaczenie maksimum i minimum lokalnego funkcji. Nie będziemy omawiać szczegółowo tego problemu, lecz ograniczymy się do prostego przypadku znalezienia minimum lokalnego funkcji quasi-różniczkowalnej $f\colon \boldsymbol{R}^n \to \boldsymbol{R}$ określonej na całej przestrzeni. Metoda, którą stosuje się do wyznaczenia minimum lokalnego, nazywa się *metodą największego spadku.* W zarysie można ją przedstawić następująco: obieramy punkt $x_0 \in \boldsymbol{R}^n$ oraz konstruujemy ciąg $(x_k)_{k \in N}$ według wzoru

$$x_{k+1} = x_k + \alpha_k g_k,$$

gdzie g_k jest kierunkiem największego spadku funkcji f w punkcie x_k, a α_k jest rozwiązaniem równania

$$f(x_k+\alpha_k g) = \min_{\alpha>0} f(x_k+\alpha g).$$

Przy pewnych warunkach regularności podciąg ciągu $(x_k)_{k\in N}$ jest zbieżny do minimum lokalnego ([2], [7]).

Kierunek największego spadku jest to wektor unormowany g_0 taki, że

$$\left.\frac{df}{dg_0}\right|_{x_0} = \inf_{\|g\|=1} \left.\frac{df}{dg}\right|_{x_0}.$$

Rozpatrując przestrzeń $\boldsymbol{R}^n$ z normą euklidesową oraz zakładając, że f jest quasi-różniczkowalna, otrzymamy następujący problem optymalizacji:

$$\inf_{\|g\|=1} \left.\frac{df}{dg}\right|_{x_0} = \inf_{\|g\|=1} \left(\sup_{v\in\underline{\partial}f|_{x_0}} \langle v, g\rangle + \inf_{w\in\bar{\partial}f|_{x_0}} \langle w, g\rangle\right),$$

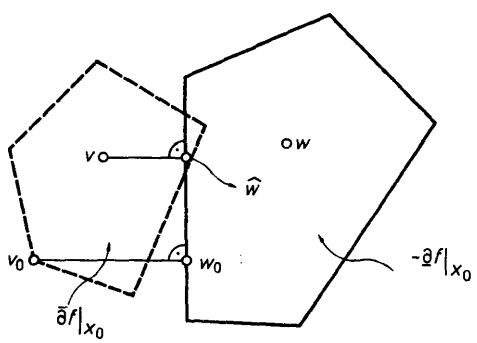
którego wszystkie rozwiązania są postaci

$$g_0 = -\frac{v_0+w_0}{\|v_0+w_0\|},$$

gdzie

$$\|v_0+w_0\| = \sup_{v\in\bar{\partial}f|_{x_0}} \inf_{w\in\underline{\partial}f|_{x_0}} \|v+w\|.$$

Powyższy problem dla optymalizacji kwadratowej można przedstawić geometrycznie jak na rys. 1. Mianowicie dla $v\in\bar{\partial}f|_{x_0}$ wyznaczamy najbliższy punkt $\hat{w}\in -\underline{\partial}f|_{x_0}$ i maksymalizujemy v.



Rys. 1

Jeśli funkcja różniczkowalna ma skończoną liczbę maksimów i minimów, to można przyjąć, że sub- i superróżniczka są reprezentowane za pomocą wielościanów. Wówczas powyższy problem optymalizacji może być rozwią-

zany za pomocą zmodyfikowanego algorytmu Franka–Wolfa [4]. Odpowiednie oprogramowanie zagadnienia przedstawiono w pracy [6].

Zakończenie. Funkcje quasi-różniczkowalne są naturalnym uogólnieniem funkcji różniczkowalnych. Zamiast przybliżania liniowego stosuje się przybliżanie za pomocą różnic funkcji subliniowych. W praktyce nie wykorzystuje się jednak powyższego wyniku do rozwiązywania problemów optymalizacji, ponieważ pary zbiorów mają lepszą reprezentację w komputerze od różnic funkcji subliniowych.

Bibliografia

[1] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York 1983.
[2] V. F. Demyanov, S. Gamidov, T. I. Sivelina, *An Algorithm for Minimizing a Certain Class of Quasidifferentiable Functions*, IIASA – Working Papers, WP-83-122 (Dec. 1983).
[3] V. F. Demyanov, A. M. Rubinov, *Quasidifferential Calculus*, Optimization Software, Springer-Verlag, Berlin 1986.
[4] V. F. Demyanov, L. V. Vasilev, *Nondifferentiable Optimization*, Optimization Software, Springer-Verlag, Berlin 1985.
[5] L. Hörmander, *Sur la fonction d'appui des ensembles convexes dans un espace localement convexe*, Ark. Mat. 3 (1954), 181–186.
[6] H. Jandl, *On the Implementation of Quasidifferentiable Optimization Problems under Constraints*, rozprawa doktorska, 1987.
[7] D. Pallaschke, P. Recht, *On the steepest descent method for a class of quasi-differentiable optimization problems*, w: *Nondifferentiable Optimization: Motivations and Applications*, Proceedings, Sopron, Hungary, 1984, Springer-Verlag, Berlin 1985, 252–263.
[8] D. Pallaschke, P. Recht, R. Urbański, *On locally-Lipschitz quasi-differentiable functions in Banach spaces*, Optimization 17 (3) (1986), 287–295.
[9] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton 1970.
[10] R. Urbański, *A generalization of the Minkowski–Rådström–Hörmander theorem*, Bull. Acad. Polon. Sci. Sér. Sci. Math. 24 (1976), 709–715.

Miscellanea

12. *Według nas nauczyciel matematyki ma zadanie:*

- *pozwolić każdemu uczniowi rozwinąć – w ramach jego uzdolnień – umiejętności i rozumienie matematyki niezbędne w dorosłym życiu, w pracy i w dalszej nauce; nauczyciel powinien być przy tym świadom trudności, jakich niektórzy uczniowie będą doznawać w dążeniu do takiego rozumienia;*
- *wyposażyć każdego ucznia w taką matematykę, jaka jest konieczna przy uczeniu się innych przedmiotów;*