Algorytmy analizy praktycznej
Podstawy analizy wypukłej

Szymon Majewski

15 marca 2019

1 Twierdzenia o oddzielaniu

Poniższe twierdzenia można znaleźć wraz z dowodami w skrpypcie [1], wykład 3:

Twierdzenie 1: Słabe twierdzenie o oddzielaniu
Niech $U, V \subseteq R^n$ będą niepustymi zbiorami wypukłymi, takimi że $U \cap V = \emptyset$. Wówczas istnieje hiperpłaszczyzna rozdzielająca U i V, czyli wektor $\xi \in \mathbb{R}^n$, taki że:

$$\sup_{u \in U} \xi^T u \leq \inf_{v \in V} \xi^T v$$

Twierdzenie 2: Słabe twierdzenie o oddzielaniu v2
Niech $U, V \subseteq R^n$ będą niepustymi zbiorami wypukłymi, takimi że $\mathrm{int}(U) \neq \emptyset$ oraz $\mathrm{int}(U) \cap V = \emptyset$. Wówczas istnieje hiperpłaszczyzna rozdzielająca U i V, czyli niezerowy wektor $\xi \in \mathbb{R}^n$, taki że:

$$\sup_{u \in U} \xi^T u \leq \inf_{v \in V} \xi^T v$$

Dowód. Powyższe twierdzenie wynika z poprzedniego, wystarczy zauważyć że wnętrze zbioru wypukłego jest zbiorem wypukłym. □

Twierdzenie 3: Silne twierdzenie o oddzielaniu
Niech U, V będą niepustymi zbiorami wypukłymi domkniętymi, U zwarty, $U \cap V = \emptyset$. Wówczas istnieje hiperpłaszczyzna ściśle rozdzielająca, to znaczy $\xi \in \mathbb{R}^n$ takie że:

$$\sup_{v \in V} \xi^T v < \inf_{u \in U} \xi^T u$$

2 Subgraidenty i subróżniczki

Definicja 1: Dla funkcji wypukłej $f : \mathbb{R}^n \supseteq W \to \mathbb{R}$ oraz punktu $x \in W$, subgradientem f w x nazywamy dowolny wektor $\xi \in \mathbb{R}^n$, taki że dla dowolnego $y \in W$ zachodzi:

$$f(y) \geq f(x) + \xi^T (y - x)$$

Definicja 2: Dla f, x jak w poprzedniej definicji, subróżniczką f w x nazywamy zbiór wszystkich subgradientów f w x. Subróżniczkę f w x oznaczamy $\partial f(x)$.

Własność 1: Dla $f, g : W \to \mathbb{R}^n$ oraz $x \in W$ mamy:

a) $\partial f(x) + \partial g(x) \subseteq \partial(f+g)(x)$

b) jeśli $x \in \mathrm{int}(W)$ mamy $\partial f(x) + \partial g(x) = \partial(f+g)(x)$

c) dla $\lambda > 0$ mamy $\partial(\lambda f)(x) = \lambda \partial f(x)$

d) zbiór $\partial f(x)$ jest domknięty i wypukły

e) jeśli $x \in \mathrm{int}(W)$ to $\partial f(x)$ jest ograniczony

f) jeśli $x \in \mathrm{int}(W)$ oraz f jest różniczkowalna w x, to $\partial f(x) = \{\nabla f(x)\}$

Własności $(a), (c), (d)$ wynikają wprost z defnicji i polecam spróbować udowodnić je samemu jeśli ktoś nie pamięta dowodów z wykładu. Dowód własności e, f też nie jest trudny i można spróbować wykazać tą własności samodzielnie, ale są one odrobinkę trudniejsze od trzech poprzednio wymienionych. Dowód (b) wymaga silnego twierdzenia o oddzielaniu. Sformułowanie i dowód tego twierdzenia, dowód (b) można znaleźć w skrypcie [1].
W dalszej części, do dowodu warunków optymalności, bedziemy potrzebowali jeszcze jednej własności subróżniczki. Najpierw przypominimy definicję pochodnej kierunkowej:

Definicja 3: Dla funkcji $f : \mathbb{R}^n \supseteq W \to \mathbb{R}$ punktu $x \in \mathrm{int}(W)$ oraz kierunku $d \in \mathbb{R}^n$, pochodną kierunkową f w punkcie x w kierunku d nazywamy:

$$f'(x, d) = \lim_{\lambda \to 0_+} \frac{f(x + \lambda d) - f(x)}{\lambda}$$

o ile ta granica istnieje

Okazuje się, że dla funkcji wypukłych pochodne kierunkowe zawsze istnieją we wnętrzu dziedziny:

Lemat 1: Dla $f : W \to \mathbb{R}$ wypukłej, oraz $x \in \mathrm{int}(W), d \in \mathbb{R}^n$, pochodna kierunkowa $f'(x, d)$ istnieje i jest skończona.

Dowód. Oznaczmy:

$$f_{x,d}(\lambda) = \frac{f(x + \lambda d) - f(x)}{\lambda}$$

i niech L będzie takie że $x + Ld \in W$. Wówczas z definicji wypukłości można wykazać, że funkcja $f_{x,d}(\lambda)$ jest niemalejącą funkcją λ na odcinku $(0, L]$ (polecam jako ćwiczenie). Ponadto, funkcja ta jest ograniczona z dołu, bo dla dowolnego $\xi \in \partial f(x)$ mamy $f_{x,d}(\lambda) \geq \xi^T d$. Zatem granica

$$f'(x, d) = \lim_{\lambda \to 0_+} f_{x,d}(\lambda)$$

istnieje i jest skończona. $\square$

Istnieje następujący ważny związek między pochodną kierunkową a subróżniczką:

Twierdzenie 4: Dla $f : W \to \mathbb{R}$, $x \in \mathrm{int}(W)$ oraz $d \in \mathbb{R}^n$, zachodzi:

$$f'(x, d) = \sup_{\xi \in \partial f(x)} \xi^T d$$

Dowód. Dowód tego stwierdzenia można znaleźć w rozdziale trzecim skryptu [1]. □

Na zakończenie jeszcze jedno zadanie na przećwiczenie definicji:

Zadanie 1: Niech $f : W \to \mathbb{R}$ będzie funkcją wypukłą, niech $V \subseteq W$ będzie zbiorem wypukłym zawartym w W i niech f_V będzie obcięciem f do V. Udowodnij, że dla każdego $x \in V$ mamy $\partial f(x) \subseteq \partial f_V(x)$.

Warto zauważyć, że subróżniczka obcięcia może być istotnie większa od subróżniczki przed obcięciem!

3 Stożki styczne i normalne

Definicja 4: Stożkiem w dowolnej przestrzeni liniowej L nad $\mathbb{R}$, nazywamy dowolny zbiór $S \subseteq L$, taki że dla dowolnych $x \in S$ oraz $\lambda \in \mathbb{R}_+ \cup \{0\}$ zachodzi $(\lambda x) \in S$.

Stożek jest więc sumą promieni wychodzących z punktu 0. Zauważmy, że stożek niepusty zawsze zawiera punkt 0.

Dla dowolnego zbioru $X \subseteq \mathbb{R}^n$ oraz $x \in X$ można zdefiniować stożek styczny do X w punkcie x w następujący sposób:

Definicja 5: Dla zbioru $X \subseteq \mathbb{R}^n$ wektor $d \in \mathbb{R}^n$ nazywamy kierunkiem stycznym do X w x, jeśli istnieje ciąg elementów $x_n \in X$ oraz liczb $t_n \in \mathbb{R}, t_n > 0$, takie że $\lim_{n\to\infty} x_n = x$ oraz $\lim_{n\to\infty} \frac{x_n - x}{t_n} = d$.
Zbiór wszystkich kierunków stycznych do X w x nazywamy stożkiem stycznym

Korzystając bezpośrednio z definicji, można wykazać:

Zadanie 2: Udowodnij, że dla dowolnego niepustego zbioru X oraz $x \in X$ zbiór $T_X(x)$:

a) zawiera punkt 0

b) jest stożkiem

c) jest domknięty

Dla zbioru wypukłego powyższa definicja sprowadza sie do dużo prostszej:

Twierdzenie 5: Dla zbioru wypukłego W i punktu $x \in W$ mamy:

$$T_W(x) = cl\left(\{\lambda(y - x) : y \in W, \lambda \geq 0\}\right)$$

Zadanie 3: Udowodnij twierdzenie 5.

Dla dowolnego stożka, można zdefiniować stożek sprzężony (czasem nazywany polarnym):

Definicja 6: Dla danego stożka K stożekiem sprzężonym do K nazywamy zbiór:

$$K^\circ = \{y : y^t x \leq 0 \text{ dla wszystkich } x \in K\}$$

Zadanie 4: Udowdnij, że:

a) dla dowolnego stożka K zbiór K° jest stożkiem

b) stożek K° jest zawsze domknięty i wypukły, nawet jeśli K nie jest wypukły

c) mamy $K \subseteq (K^\circ)^\circ$

Powyższe trzy własności są proste do udowodnienia wprost z definicji i polecam je jako ćwiczenie. Poniższe twierdzenie jest odwróceniem trzeciego punktu, i przyda nam się w dowodzie warunku optymalności:

Twierdzenie 6: Dla wypukłego i domkniętego zbioru K zachodzi $(K^\circ)^\circ = K$.

Dowód. Zawieranie w jedna stronę jest proste i jest częścią poprzedniego zadania. W drugą stronę: założmy że istnije $x \in (K^\circ)^\circ \setminus K$. Ponieważ K jest domknięty i wypukły, a zbiór $\{x\}$ jest zwarty, możemy zastosować silne twierdzenie o oddzielaniu (3). Istnieje zatem niezerowy wektor ξ, taki że $\sup_{y \in K} \xi^T y < \xi^T x$. Ponieważ K jest stożkiem, musi zachodzić $\sup_{y \in K} \xi^T y \leq 0$. Gdyby bowiem istniało $y \in K$ takie że $\xi^T y > 0$, to ponieważ dla dowolnego $\lambda > 0$ punkt λy również należy do K, wiec byłoby $\sup_{y \in K} \xi^T y = \infty$, co byłoby sprzeczne z $\sup_{y \in K} \xi^T y < \xi^T x$. Zatem mamy $\sup_{y \in K} \xi^T y = 0$, bo z definicji stożka $0 \in K$. Stąd $\xi \in K^\circ$, ale $\xi^T x > 0$. Sprzeczność, zatem musi być $(K^\circ)^\circ \subseteq K$. $\square$

Możemy teraz zdefiniować stożek normalny do zbioru wypukłego:

Definicja 7: Dla niepustego zbioru wypukłego W oraz punktu $x \in W$ stożkiem normalnym do W w x nazywamy zbiór:

$$N_W(x) = (T_W(x))^\circ$$

Zadanie 5: Przy oznaczeniach jak w definicji powyżej, udowodnij że $\xi \in N_W(x)$ wtedy i tylko wtedy gdy $\xi^T x \leq 0$ dla wszystkich $x \in W$.

Jeszcze kilka prostych ćwiczeń na oswojenie się ze stożkami:

Zadanie 6: Niech W będzie zbiorem wypukłym i $x \in \operatorname{int}(W)$. Znajdź $T_W(x)$ oraz $N_W(x)$.

Zadanie 7: Dla zbioru wypukłego $W \subseteq \mathbb{R}^n$, niech I_W będzie funkcją daną wzorem:

$$I_W(x) = \begin{cases} 0 & \text{dla } x \in W \\ +\infty & \text{dla } x \notin W \end{cases}$$

Wykaż że funkcja I_W jest wypukła na $\mathbb{R}^n$ oraz że $\partial I_W(x) = N_W(x)$ dla wszystkich $x \in W$ (robiliśmy to na wykładzie).

Zadanie 8: Dla zbioru wypukłego W oraz punktu x na brzegu W (czyli $x \in W \setminus \operatorname{int}(W)$) istnieje $\xi \neq 0$ ktore należy do $N_W(x)$ (może przydać się któreś twierdzenie o oddzielaniu).

4 Warunki optymalności

Prosty warunek opymalności dla funkcji wypukłych na zbiorze wypukłym był następujący:

Twierdzenie 7: Dla funkcji wypukłej $f : W \to \mathbb{R}$ mamy $x \in \arg\min_{y \in W} f(y)$ wtedy i tylko wtedy gdy $0 \in \partial f(x)$

Dowód. Jak na wykładzie. Z definicji subróżniczki i subgradientu mamy $0 \in \partial f(x)$ wtedy i tylko wtedy gdy dla wszystkich $y \in W$ mamy $f(y) \geq f(x) + 0^T(y - x)$, czyli $f(y) \geq f(x)$. □

Poniżej wraz z dowodem naprawiony warunek optymalności z wykładu:

Twierdzenie 8: Niech $W, V \subseteq \mathbb{R}^n$ będą zbiorami wypukłymi, $V \subseteq \mathrm{int}(W)$. Niech $f : W \to \mathbb{R}$ będzie funkcją wypukłą. Wówczas $x \in \arg\min_{y \in V} f(y)$, wtedy i tylko wtedy gdy:

$$0 \in \partial f(x) + N_V(x)$$

W dowodzie przyda się następujący lemat:

Lemat 2: Niech U, V będą zbiorami domkniętymi i wypukłymi. Załóżmy że U jest zwarty. Wówczas zbiór $S = U + V$ jest wypukły i domknięty.

Dowód. Wypukłość sumy zbiorów wypukłych wynika wprost z definicji, pozostawiam wykazanie tej implikacji czytelnikowi. Weźmy ciąg $\{s_i\}_{i=1}^{\infty}$ elmentów S, taki że $\lim_{n\to\infty} s_n = s$. Chcemy wykazać $s \in S$. Ponieważ $S = U + V$, istnieją ciągi $\{u_i\}_{i=1}^{\infty}, \{v_i\}_{i=1}^{\infty}$ takie że dla wszystkich $n \geq 1$ mamy $u_n \in U, v_n \in V, s_n = u_n + v_n$. Ze zwartości U, istnieje podciąg $\{u_{n_k}\}_{k=1}^{\infty}$ ciągu $\{u_n\}_{n=1}^{\infty}$ zbieżny do pewnej granicy $u \in U$. Wówczas ponieważ $u_{n_k} + v_{n_k}$ zbiega, to v_{n_k} zbiega do pewnej granicy v. Z domkniętości V mamy $v \in V$. Ponieważ $u_{n_k} + v_{n_k}$ zbiega do s, stąd $s = u + v$. Co kończy dowód domkniętości $U + V$. □

Dowód. **twierdzenia** 8
Dowód w jedną stronę jest prosty. Załóżmy, że $0 \in \partial f(x) + N_V(x)$, czyli że istnieje $\xi \in \partial f(x)$, takie że $(-\xi) \in N_V(x)$. Wówczas:

$$f(y) \geq f(x) + \xi^T(y - x)$$

dla wszystkich $y \in W$, oraz:

$$0 \geq (-\xi)^T(y - x)$$

dla wszystkich $y \in V$. Dodając te nierówności stronami, otrzymujemy $f(y) \geq f(x)$ dla wszystkich $y \in V$.

W drugą stronę, załóżmy że $f(y) \geq f(x)$ dla wszystkich $y \in V$. Niech $S = \partial f(x) + N_V(x)$. Wykażemy, że dla wszystkich $d \in \mathbb{R}^n$ zachodzi $\sup_{s \in S} d^T s \geq 0$. Jeśli $d \notin T_V(x)$, to dla pewnego $n \in N_V(x)$ mamy $d^T n > 0$ (wynika to z tego, że z twierdzenia 6 mamy $T_V(x) = (N_V(x))^\circ$). Ponieważ dla dowolnego $\lambda > 0$ wektor λn również należy do stożka $N_V(x)$, to mamy $\sup_{s \in S} d^T s = \infty$.

Jeśli natomiast $d \in T_V(x)$, to mamy $\sup_{s \in S} d^T s \geq \sup_{\xi \in \partial f(x)} \xi^T d$. Wystarczy zatem wykazać $\sup_{\xi \in \partial f(x)} \xi^T d \geq 0$. Załóżmy, że $d \in T_V(x)$ jest takie, że dla pewnego $L > 0$ mamy $x + Ld \in V$. Wówczas korzystając z twierdzenia 4 mamy:

$$\sup_{\xi \in \partial f(x)} \xi^T d = \lim_{\lambda \to 0_+} \frac{f(x + \lambda d) - f(x)}{\lambda} \geq 0$$

gdzie do uzyskania nierówności wykorzystujemy założenie, że $f(y) \geq f(x)$ dla wszystkich $y \in V$. Zatem $\sup_{\xi \in \partial f(x)} \xi^T d \geq 0$ dla d takich że $\lambda d \in V$ dla pewnego $\lambda > 0$. Ponieważ $g(z) = \sup_{\xi \in \partial f(x)} \xi^T z$ jest ciągłą funkcją z (polecam udowodnić jako ćwiczenie), wynika z tego, że $\sup_{\xi \in \partial f(x)} \xi^T d \geq 0$ dla wszystkich $d \in T_V(x)$, a zatem $\sup_{s \in S} d^T s \geq 0$ dla wszystkich $d \in \mathbb{R}^n$.

Z drugiej strony, ponieważ $\partial f(x)$ oraz $N_V(x)$ są domknięte oraz wypukłe, S jest wypukły i domknięty (na mocy lematu2). Załóżmy że $0 \notin S$. Z silnego twierdzenia o oddzielaniu dla zbiorów $\{0\}, S$, wnioskujemy że istnieje taki wektor $\beta \in \mathbb{R}^n$, że $\sup_{s \in S} \beta^T s < \beta^T 0$, czyli $\sup_{s \in S} \beta^T s < 0$. Jest to sprzeczne z poprzednio wykazaną nierównością, zatem jeśli dla wszystkich $y \in V$ zachodzi $f(y) \geq f(x)$, to musi również zachodzić $0 \in \partial f(x) + N_V(x)$. $\square$

5 Rzutowanie

Definicja 8: Dla domkniętego, niepustego zbioru wypukłego $W \subseteq \mathbb{R}^n$, rzutowaniem na zbiór W nazywamy funkcję $P_W : \mathbb{R}^n \to W$ określoną wzorem:

$$P_W(x) = \arg\min_{y \in W} ||x - y||_2$$

Dla domkniętego wypukłego zbioru powyższa funkcja jest dobrze określona, to znaczy minimum jest przyjmowane w dokładnie jednym punkcie zbioru W. Dowód tego faktu polecam jako ćwiczenie, rozbite na dwa problemy.

Pierwszy z nich udowodniliśmy na wykładzie, ale polecam sobie powtórzyć (wskazówka: korzystaliśmy ze zwartości jakiegoś zbioru).

Zadanie 9: Dla domknietego, niepustego zbioru $W \subseteq \mathbb{R}^n$ oraz punktu $x \in \mathbb{R}^n$ istnieje przynajmniej jeden punkt $x_* \in W$, taki że:

$$||x - x_*||_2 = \inf_{y \in W} ||y - x||_2$$

Drugie zadanie było zostawione do przemyślenia w domu:

Zadanie 10: Jeśli W jest niepustym, domkniętym zbiorem wypukłym, to dla każdego x istnieje dokładnie jeden punkt x_* taki że:

$$||x_* - x||_2 = \inf_{y \in W} ||y - x||_2$$

6 Operator prox

Wprowadziliśmy następujący operator:

Definicja 9: Dla funkcji wypukłej $f : W \to \mathbb{R}$ oraz $x \in \mathbb{R}^n$ operator prox jest zdefiniowany następująco:

$$\mathrm{prox}_f(x) = \arg\min_{y \in W} f(y) + \frac{1}{2}||y - x||_2^2$$

Ogólnym warunkiem na to by operator prox był dobrze zdefiniowany dla wszystkich x, jest dolna półciągłość f połączona z domkniętością W. Poniższy nieco ogólniejszy fakt można potraktować jako zadanie na przećwiczenie definicji:

Zadanie 11: Dla wypukłej, ciągłej funkcji $f : W \to \mathbb{R}^n$ określonej na wypukłym, domkniętym zbiorze W, $\mathrm{prox}_f(x)$ jest dobrze określony dla każdego $x \in \mathbb{R}^n$ (trzeba wykazać że minimum istnieje, i jest dobrze określone).

Z warunków optymalności wynika następujący wniosek:

Wniosek 1: Niech W będzie wypukłym otwartym zbiorem, $f : W \to \mathbb{R}$ będzie wypukła oraz $\lambda > 0$. Mamy $\mathrm{prox}_{\lambda f}(x) = y$ wtedy i tylko wtedy, gdy $x - y \in \lambda \partial f(y)$.

Dowód. Ponieważ W jest otwarty to $\partial(\lambda f + \frac{1}{2}|| \cdot -x||_2^2)(y) = \lambda \partial f(y) + (y - x)$ z własności subróżniczki. Z warunku optymalności wynika teza. □

Operator prox jest użyteczny w algorytmach tylko wtedy, gdy dla dowolnego x potrfimy szybko policzyć $\mathrm{prox}_{\lambda f}(x)$ - na przykład dla f dla których $\mathrm{prox}_{\lambda f}$ można wyrazić jawnym wzorem. Poniżej przykłady:

Zadanie 12: Znajdź wzór jawny na $\mathrm{prox}_{\lambda f}$ dla $\lambda > 0$ oraz

a) $f(x) = |x|_1 = \sum_{i=1}^n |x_i|$

b) $f(x) = ||x - z||_2^2 = \sum_{i=1}^n (x_i - z_i)^2$ (dla pewnego ustalonego $z \in \mathbb{R}^n$)

c) $f(x) = Ax + b$ dla pewnej macierzy A

Literatura

[1] Jan Palczewski *Optymalizacja 2* `http://mst.mimuw.edu.pl/lecture.php?lecture=op2`

[2] Stephen Boyd, Lieven Vandenberghe *Convex optimization* `http://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf`

[3] Neal Parikh, Stephen Boyd *Proximal algorithms* `http://web.stanford.edu/~boyd/papers/pdf/prox_algs.pdf`