Subdiferenciály a jejich role v optimalizaci

Milan Hladík

Katedra aplikované matematiky,
Matematicko-fyzikální fakulta,
Univerzita Karlova, Praha,
https://kam.mff.cuni.cz/~hladik

4. listopadu 2020

Motivace

Konvexní funkce hladká a nehladká.

Subgradient a subdiferenciál

Definice

Buď $f : \mathbb{R}^n \to \mathbb{R}$ konvexní. *Subgradientem* funkce $f(x)$ v bodě $x^0 \in \mathbb{R}^n$ je vektor $g \in \mathbb{R}^n$ splňující

$$f(x) \geq f(x^0) + g^T(x - x^0), \quad \forall x \in \mathbb{R}^n.$$

Subdiferenciál $\partial f(x^0)$ je množina všech subgradientů.

- subgradient je normálou tečné nadroviny ke grafu $f(x)$ v x^0
- pro hladkou funkci pojem gradientu a subgradientu splývají a subdiferenciál je jednobodová množina
- Pro řadu metod nehladké konvexní optimalizace stačí najít subgradient, což lze většinou snadno. Pokud chceme ale vyjádřit podmínky optimality, potřebujeme celý subdiferenciál, a ten bývá obecně obtížné spočítat.

Příklad

Pro $f(x) = |x|$ je $\partial f(0) = [-1, 1]$.

Vlastnosti

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Pozorování

Subdiferenciál konvexní funkce $f : \mathbb{R}^n \to \mathbb{R}$ v bodě $x^0 \in \mathbb{R}^n$ je uzavřená konvexní množina.

Důkaz.

Je to průnik poloprostorů. $\square$

Subdiferenciál splňuje základní řetízková pravidla:

$$\begin{aligned} \partial(\alpha f(x)) &= \alpha \partial f(x), \quad \alpha > 0, \\ \partial(f(x) + g(x)) &= \partial f(x) + \partial g(x), \\ \partial(f(Ax + b)) &= A^T(\partial f)(Ax + b). \end{aligned}$$

Vlastnosti

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Tvrzení

Pro konvexní funkce $f_1, \ldots, f_m \colon \mathbb{R}^n \to \mathbb{R}$ v bodě $x^0 \in \mathbb{R}^n$ je

$$\partial \max\{f_1(x^0), \ldots, f_m(x^0)\} = conv\ \cup_{k \in I(x^0)} \partial f_k(x^0),$$

kde množina aktivních podmínek $I(x^0) = \arg\max_{k=1,\ldots,m}\{f_k(x^0)\}$.

Důkaz (jen "$\supseteq$").

Buď $\lambda_k \geq 0$, $\sum_{k \in I(x^0)} \lambda_k = 1$, $g_k \in \partial f_k(x^0)$. Pak

$$f_k(x) \geq f_k(x^0) + g_k^T(x - x^0), \quad \forall x \in \mathbb{R}^n,$$

z čehož váženým součtem

$$\max_k\{f_k(x)\} \geq \sum_{k \in I(x^0)} \lambda_k f_k(x) \geq \max_k\{f_k(x^0)\} + \sum_{k \in I(x^0)} \lambda_k g_k^T(x - x^0),\ \forall x \quad \square$$

Příklad

Funkce $f(x) = \max_{k=1,\ldots,m}(a_k^T x + b_k)$ v bodě x^0 má subdiferenciál $\partial f(x^0) = conv\,\{a_k; \in I(x^0)\}$.

Vlastnosti

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Tvrzení

Buď $h\colon \mathbb{R}^m \to \mathbb{R}$ konvexní a v každé složce neklesající funkce a $g_1, \ldots, g_m\colon \mathbb{R}^n \to \mathbb{R}$ konvexní funkce. Pak pro subdiferenciál složené funkce $f(x) = h(g(x))$ bodě $x^0 \in \mathbb{R}^n$ platí

$$\partial h(g(x^0))^T \supseteq \partial h(y^0)^T \partial g(x^0)^T, \quad \textit{kde } y^0 := g(x^0).$$

Důkaz.

Buď $v_{g_k} \in \partial g_k(x^0)\ \forall k$ a $v_h \in \partial h(y^0)$. Pak

$$\begin{aligned}
f(x) &= h(g_1(x), \ldots, g_m(x)) \\
&\geq h\big(g_1(x^0) + v_{g_1}^T(x - x^0), \ldots, g_m(x^0) + v_{g_m}^T(x - x^0)\big) \\
&\geq h(g_1(x^0), \ldots, g_m(x^0)) + v_h^T \left(v_{g_1}^T(x - x^0), \ldots, v_{g_m}^T(x - x^0)\right)^T \\
&= f(x^0) + \sum_{k=1}^{m} (v_h)_k v_{g_k}^T (x - x^0)
\end{aligned}$$

Tudíž $(v_{g_1} \mid \cdots \mid v_{g_k}) v_h \in \partial f(x^0)$. $\square$

Vlastnosti

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Příklady funkcí $f\colon \mathbb{R}_+ \to \mathbb{R}$, které nemají subdiferenciál v bodě $x^0 = 0$:

- $f(x) = 1$ pro $x = 0$ a $f(x) = 0$ pro $x > 0$,
- $f(x) = -\sqrt{x}$.

Na druhou stranu, pokud x^0 náleží do topologického vnitřku definičního oboru funkce $f(x)$, pak subdiferenciál v bodě x^0 je neprázdný.

Subdiferenciál normy

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Definice

Buď $\|\cdot\|\colon \mathbb{R}^n \to \mathbb{R}$ norma. Pak k ní *duální norma* $\|\cdot\|_*$ je definovaná předpisem

$$\|x\|_* = \max\{x^T y;\ \|y\| \leq 1\}.$$

- $\|x\|_p$ je duální k $\|x\|_q$ pro $\frac{1}{p} + \frac{1}{q} = 1$.
- Speciálně, eukleidovská norma je samo-duální.
- Jako limitní případ, součtová norma $\|x\|_1$ je duální k maximové $\|x\|_\infty$ a naopak.

Tvrzení (Subdiferenciál normy)

$$\partial\|x^0\| = \{y \in \mathbb{R}^n;\ y^T x^0 = \|x^0\|,\ \|y\|_* \leq 1\}.$$

Subdiferenciál normy

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Tvrzení (Subdiferenciál normy)

$$\partial\|x^0\| = \{y \in \mathbb{R}^n;\ y^T x^0 = \|x^0\|,\ \|y\|_* \leq 1\}.$$

Důkaz.

Podle definice, $g \in \mathbb{R}^n$ je subgradientem normy v bodě x^0, pokud

$$\|x\| \geq \|x^0\| + g^T(x - x^0), \quad \forall x \in \mathbb{R}^n.$$

Dosazením $x = 0$ a $x = 2x^0$ dostaneme $\|x^0\| = g^T x^0$, tudíž

$$\|x\| \geq g^T x, \quad \forall x \in \mathbb{R}^n$$

a z positivní homogenity se stačí omezit na $\|x\| = 1$, tedy

$$1 \geq g^T x, \quad \forall x : \|x\| = 1.$$

To je ale ekvivalentní s $\|g\|_* \leq 1$. $\square$

Subdiferenciál normy

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Tvrzení (Subdiferenciál normy)

$$\partial\|x^0\| = \{y \in \mathbb{R}^n;\ y^T x^0 = \|x^0\|,\ \|y\|_* \leq 1\}.$$

Například:

- v počátku $x^0 = 0$ je subgradientem jednotková koule v duální normě
$$\partial\|x^0\| = \{y \in \mathbb{R}^n;\ \|y\|_* \leq 1\}.$$

- Eukleidovská norma je v každém nenulovém bodě hladká, a tedy subdiferenciál splývá s gradientem $\partial\|x^0\| = \frac{1}{\|x^0\|_2} x^0$.
- Pro součtovou normu podmínka $y^T x^0 = \|x^0\|_1,\ \|y\|_\infty = 1$ platí jen pro taková y, pro která $y_i = \mathrm{sgn}(x_i^0)$, pokud $x_i^0 \neq 0$, a $y_i = [-1, 1]$, pokud $x_i^0 = 0$.

Subdiferenciál vlastního čísla

$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$

Tvrzení (Subdiferenciál maximálního vlastního čísla)

Uvažujme funkci $f(X) = \lambda_{\max}(X)$ *na prostoru symetrických matic. Buď* $X \in \mathbb{R}^{n\times n}$ *symetrická a* v, $\|v\|_2 = 1$, *příslušný vlastní vektor k* $\lambda_{\max}(X)$. *Pak*

$$v\,v^T \in \partial f(X).$$

Důkaz.

Pro každou symetrickou matici $Y \in \mathbb{R}^{n\times n}$ platí

$$\begin{aligned}
\lambda_{\max}(Y) &= \max\{u^T Y u;\ \|u\|_2 = 1\} \\
&\geq v^T Y v \\
&= v^T X v + v^T (Y - X) v \\
&= \lambda_{\max}(X) + \operatorname{tr}(v^T (Y - X) v) \\
&= \lambda_{\max}(X) + \operatorname{tr}(v v^T (Y - X)) \\
&= \lambda_{\max}(X) + \langle v v^T, Y - X \rangle
\end{aligned}$$

□

Subdiferenciál v optimalizaci

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Věta

Bod $x^0 \in \mathbb{R}^n$ je minimem konvexní funkce $f : \mathbb{R}^n \to \mathbb{R}$ právě tehdy, když $0 \in \partial f(x^0)$.

Důkaz.

Pokud $0 \in \partial f(x^0)$, pak podmínka z definice subgradientu má tvar

$$f(x) \geq f(x^0), \quad \forall x \in \mathbb{R}^n,$$

čili x^0 je minimem. Naopak, pokud je x^0 minimem, pak $0 \in \partial f(x^0)$. $\square$

Subdiferenciál v optimalizaci

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Příklad

Uvažujme úlohu

$$\min_{x \in \mathbb{R}^n} \max_{k=1,\dots,m} (a_k^T x + b_k).$$

Podle věty je bod $x^0 \in \mathbb{R}^n$ optimem právě tehdy, když existuje $y \geq 0$ takové, že $e^T y = 1$, $\sum_{k=1}^m y_k a_k = 0$ a $y_k = 0$ pro $k \notin I(x^0)$.

Pro srovnání, přepíšeme úlohu jako lineární program

$$\min\ z \quad \text{subject to} \quad a_k^T x + b_k \leq z \quad \forall k.$$

Duální úloha má tvar

$$\max\ b^T y \quad \text{subject to} \quad \sum_{k=1}^m y_k a_k = 0,\ y \geq 0,\ e^T y = 1.$$

Podmínky optimality dají stejné vyjádření, tj. existuje duální přípustné řešení y takové, že $y_k = 0$ pokud $a_k^T x^0 + b_k < z^0$, kde $z^0 = \max_k (a_k^T x^0 + b_k)$.

Subdiferenciál v optimalizaci

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Příklad (LASSO)

Uvažujme úlohu LASSO

$$\min\ \frac{1}{2}\|Ax - b\|_2 + \lambda\|x\|_1 \quad \text{subject to} \quad x \in \mathbb{R}^n,$$

kde $\lambda > 0$ je parametr. Podmínky optimality pro bod $x^0 \in \mathbb{R}^n$

$$0 \in \partial\tfrac{1}{2}\|Ax^0 - b\|_2 + \lambda\|x^0\|_1 = \partial\tfrac{1}{2}\|Ax^0 - b\|_2 + \partial\lambda\|x^0\|_1$$
$$0 \in A^T(Ax^0 - b) + \lambda\partial\|x^0\|_1.$$

Víme, že je $v \in \partial\|x^0\|_1$ právě tehdy, když $v_i = \operatorname{sgn}(x_i^0)$ pro $x_i^0 \neq 0$ a $v_i \in [-1, 1]$ pro $x_i^0 = 0$. Tedy podmínky optimality jsou

$$\begin{cases} A_{i*}^T(b - Ax^0) = \lambda \operatorname{sgn}(x_i^0) & \text{pokud } x_i^0 \neq 0, \\ |A_{i*}^T(b - Ax^0)| \leq \lambda & \text{pokud } x_i^0 = 0. \end{cases}$$

- Umožňují ověřit, zda dané x^0 je optimem (ale ne ho najít).
- Vlastnosti optima, jako například $|A^T(b - Ax^0)| \leq \lambda e$.

Subdiferenciál v optimalizaci

$$\boxed{f(x) \geq f(x^0) + g^T(x - x^0)\ \forall x}$$

Tvrzení

Platí pro směrovou derivaci ve směru v konvexní funkce $f : \mathbb{R}^n \to \mathbb{R}$:

$$f'_v(x^0) = \sup\{g^T v;\ g \in \partial f(x^0)\}.$$

Důkaz.

"$\geq$" Buď $g \in \partial f(x^0)$. Podle definice

$$f'_v(x^0) = \min_{\alpha \to 0^+} \frac{f(x^0 + \alpha v) - f(x^0)}{\alpha} \geq \min_{\alpha \to 0^+} \frac{f(x^0) + \alpha g^T v - f(x^0)}{\alpha} = g^T v.$$

"$\leq$" Idea: Ukaž, že $f'_v(x)$ je konvexní a její subgradient je subgradientem $f(x)$. Pak uvaž $v \to \alpha v$ pro $\alpha \to 0^+$. $\square$

- Optimalizační metody jsou často iterační ve směru klesání.
- Subgradient přímočaře použít nelze, např. funkce $f(x) = |x|$ má v bodě $x^0 = 0$ subgradient $g = 1$, ale funkce ve směru -1 neklesá.