On global unconstrained minimization of the difference of polyhedral functions

09.02.2018

Polyhedral functions

A function

$$f(x) = \max_{i \in I} \{\langle a_i, x \rangle + b_i\}, \quad a_i \in \mathbb{R}^n, \quad b_i \in \mathbb{R}, \quad i \in I = 1, \ldots, m,$$

is called a polyhedral function.

The subdifferential of a polyhedral function is a convex polyhedron, namely,

$$\partial f(x) = \text{co} \left\{ \bigcup_{i \in R(x)} a_i \right\},$$

where

$$R(x) = \{i \in I \mid f_i(x) = f(x)\}, \quad f_i(x) = \langle a_i, x \rangle + b_i, \quad i \in I.$$

For the function f at each point $x \in \mathbb{R}^n$ for any $\varepsilon \geq 0$ there exists the ε-subdifferential, and the ε-subdifferential mapping

$$\partial_\varepsilon f : \ \mathbb{R}^n \times (0, +\infty) \longrightarrow 2^{\mathbb{R}^n}$$

is already continuous in the Hausdorff metric.
For the polyhedral function, the formula of the ε-subdifferential at each point $x \in \mathbb{R}^n$ is

$$\partial_\varepsilon f(x) = \left\{ v = \sum_{i=1}^{m} \lambda_i a_i \in \mathbb{R}^n \,\middle|\, \begin{array}{c} \sum\limits_{i=1}^{m} \lambda_i (f(x) - \langle a_i, x \rangle - b_i) \leq \varepsilon, \\ \sum\limits_{i=1}^{m} \lambda_i = 1, \ \lambda_i \geq 0, \ i \in I \end{array} \right\}.$$

Thus, the ε-subdifferential of a polyhedral function f at each point $x \in \mathbb{R}^n$ is a convex polyhedron.

Remark. It is necessary to note that the points $a_i, \ \ i \in I$, belong to the set $\partial_\varepsilon f(x)$ for all $\varepsilon \geq f(x) - \langle a_i, x \rangle - b_i$.

Using the conjugate function, it is possible to give another definition of the ε-subdifferential of a convex closed function f at a point $x \in \operatorname{dom} f$:

$$\partial_\varepsilon f(x) = \{v \in \mathbb{R}^n \mid f(x) + f^*(v) - \langle x, v \rangle \leq \varepsilon\}.$$

The effective domain of the conjugate function f^* is the convex hull of the vectors a_i, $i \in I$, i.e.,

$$\operatorname{dom} f^* = \operatorname{co} \left\{ \bigcup_{i \in I} a_i \right\}.$$

Thus, the conjugate function is finite only at points of this polyhedron. Outside of it the function f^* takes the value $+\infty$.

Hypodifferentiable functions

V.F. Demyanov introduced the notions of hypodifferentiable functions and hypodifferentials

A function f is called a hypodifferentable function at a point $x \in \mathbb{R}^n$ if there exists a convex compact set $df(x) \subset \mathbb{R}^{n+1}$ such that

$$f(x+\Delta) = f(x) + \max_{[a,v]\in df(x)} [a + \langle v, \Delta\rangle] + o(x,\Delta), \;\; a \in \mathbb{R}, \;\; v \in \mathbb{R}^n,$$

where

$$\frac{o(x,\alpha\Delta)}{\alpha} \longrightarrow 0 \quad \text{if} \quad \alpha \to 0 \quad \forall \Delta \in \mathbb{R}^n.$$

The set $df(x)$ is called a hypodifferential of the function f at a point $x \in \mathbb{R}^n$.

The hypodifferential of a functions f at a point $x \in \mathbb{R}^n$ is not uniquely defined.

A function f is called continuously hypodifferentiable at a point $x \in \mathbb{R}^n$, if it is hypodifferentiable at x and in some neighborhood of this point there exists a continuous (in the Hausdorff metric) hypodifferentiable mapping $df(x)$. A polyhedral function is continuously hypodifferentiable in $\mathbb{R}^n$.

For example, the set

$$df(x) = \mathrm{co}\left\{\bigcup_{i\in I}\begin{pmatrix} a_i \\ \langle a_i, x\rangle + b_i - f(x)\end{pmatrix}\right\} \subset \mathbb{R}^n \times \mathbb{R}. \qquad (1)$$

can be used as a hypodifferential of the polyhedral function f.

The given hypodifferentiable mapping

$$df : \mathbb{R}^n \longrightarrow 2^{\mathbb{R}^{n+1}}$$

is continuous in the Hausdorff metric.
The set $df(x) \subset \mathbb{R}^{n+1}$ is also a convex polyhedron contained in the half-space

$$H = \{z = (z_1 \ldots, z_n, z_{n+1})^T \in \mathbb{R}^n \times \mathbb{R} \mid \; z_{n+1} \leq 0\}.$$

where T denotes transposition.
For a polyhedral function f at a point $x \in \mathbb{R}^n$ we shall define the number $\varepsilon^*(x) \geq 0$ by the formula

$$\varepsilon^*(x) = \max_{i \in I}\{f(x) - f_i(x)\}. \tag{2}$$

Fix any ε, satisfying the condition $0 \le \varepsilon \le \varepsilon^*$. Put

$$d_\varepsilon f(x) = \left\{ z \in \mathbb{R}^{n+1} \middle| \; z \in df(x), z = \begin{pmatrix} v \\ t \end{pmatrix}, v \in \mathbb{R}^n, \;\; t \in \mathbb{R}, \;\; -\varepsilon \le t \le 0 \right. \quad (3)$$

The set $d_\varepsilon f(x)$ is closed and convex for any $\varepsilon : \;\; 0 \le \varepsilon \le \varepsilon^*$. It is not difficult to see, that

$$d_{\varepsilon_1} f(x) \subset d_{\varepsilon_2} f(x), \quad 0 \le \varepsilon_1 \le \varepsilon_2 \le \varepsilon^*.$$

Lemma.

For any $0 \le \varepsilon \le \varepsilon^*(x)$ *the equality*

$$\partial_\varepsilon f(x) = \left\{ v \in \mathbb{R}^n \mid \begin{pmatrix} v \\ t \end{pmatrix} \in d_\varepsilon f(x) \right\} \tag{4}$$

is valid.

Thus, the projection the set $d_\varepsilon f(x)$ onto $\mathbb{R}^n \times 0$ is the set $\partial_\varepsilon f(x) \times 0$.

Corollary.

For each $\varepsilon \geq \varepsilon^*(x)$ the equality

$$\partial_\varepsilon f(x) = \partial_{\varepsilon^*(x)} f(x) = \operatorname{co}\left\{\bigcup_{i \in I} a_i\right\} = \operatorname{dom} f^*.$$

holds.

Corollary .

If $v \notin \partial_\varepsilon f(x)$, then the point $z_t = \begin{pmatrix} v \\ t \end{pmatrix} \notin d_\varepsilon f(x)$ for every

Example 4

Let

$$f(x) = |x| = \max\{\ x, -x\ \},\ \ x \in \mathbb{R}.$$

Then

$$\text{dom } f^* = \text{co } \{-1, 1\} \subset \mathbb{R}.$$

Calculate a hypodifferential of f at $x \in \mathbb{R}$

$$df(x) = \text{co} \left\{ \begin{pmatrix} 1 \\ x - |x| \end{pmatrix}, \begin{pmatrix} -1 \\ -x - |x| \end{pmatrix} \right\} \subset \mathbb{R}^2.$$

Example 4

If $x = 0$ then

$$df(0) = \text{co}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^2.$$

In this case we have from (2) $\varepsilon^*(0) = 0$. Hence,

$$\partial_\varepsilon f(0) = \partial_{\varepsilon^*(0)} f(0) = \partial f(0) \quad \forall \varepsilon \geq 0.$$

Example 4

If $x = 1$, then

$$df(1) = \text{co}\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}-1\\-2\end{pmatrix}\right\} \subset \mathbb{R}^2,$$

and $\varepsilon^*(1) = 2$. Thus,

$$\partial_\varepsilon f(1) = \text{co}\{1-\varepsilon, 1\} \subset \mathbb{R}, \quad 0 \leq \varepsilon < \varepsilon^*(1),$$

$$\partial_\varepsilon f(1) = \partial_{\varepsilon^*(1)} f(1) = \partial f(0) = \text{dom } f^* \qquad \forall \varepsilon \geq \varepsilon^*(1).$$

Example 4.

If $x = -1$, then

$$df(-1) = \text{co}\left\{ \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^2,$$

and $\varepsilon^*(-1) = 2$. Thus,

$$\partial_\varepsilon f(-1) = \text{co}\{-1, -1 + \varepsilon\} \subset \mathbb{R}, \quad 0 \le \varepsilon < \varepsilon^*(-1),$$

$$\partial_\varepsilon f(-1) = \partial_{\varepsilon^*(-1)} f(-1) = \partial f(0) = \text{dom } f^* \qquad \forall \varepsilon \ge \varepsilon^*(-1).$$

Example 5.

Let

$$f(x) = \max\{x+1, 2x\}, \ x \in \mathbb{R}.$$

For the given function $\operatorname{dom} f^* = \operatorname{co}\{1,2\} \subset \mathbb{R}$. We have

$$df(x) = \operatorname{co}\left\{ \begin{pmatrix} 1 \\ x+1-f(x) \end{pmatrix}, \begin{pmatrix} 2 \\ 2x - f(x) \end{pmatrix} \right\} \subset \mathbb{R}^2.$$

If $x = 1$, then

$$df(1) = \operatorname{co}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^2,$$

and $\varepsilon^*(1) = 0$. Hence,

$$\partial_\varepsilon f(1) = \partial_{\varepsilon^*(1)} f(1) = \partial f(1) = \operatorname{co}\{1,2\} \quad \forall \varepsilon > 0.$$

Example 5.

If $x = 2$, then

$$df(x) = \operatorname{co}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^2,$$

and $\varepsilon^*(2) = 1$. Thus,

$$\partial_\varepsilon f(2) = \operatorname{co}\{2 - \varepsilon, 2\} \subset \mathbb{R}, \quad 0 \leq \varepsilon < \varepsilon^*(2),$$

$$\partial_\varepsilon f(2) = \partial_{\varepsilon^*(2)} f(2) = \partial f(1) = \operatorname{dom}\ f^* \quad \forall \varepsilon \geq \varepsilon^*(2).$$

Example 5.

If $x = 0$ then

$$df(0) = \mathrm{co}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\} \subset \mathbb{R}^2.$$

Then we have $\varepsilon^*(0) = 1$. Therefore

$$\partial_\varepsilon f(0) = \mathrm{co}\{1, 1+\varepsilon\} \subset \mathbb{R}, \quad 0 \le \varepsilon < \varepsilon^*(0),$$

$$\partial_\varepsilon f(0) = \partial_{\varepsilon^*(0)} f(0) = \partial f(1) = \mathrm{dom}\ f^* \quad \forall \varepsilon \ge \varepsilon^*(0).$$

A geometrical interpretation

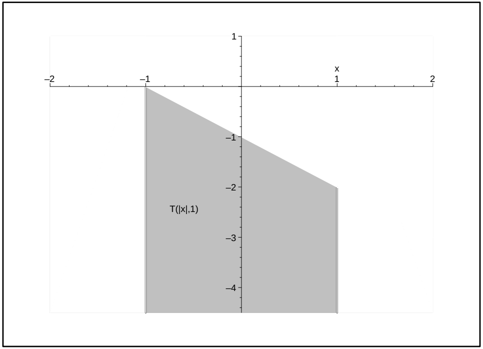
Let's denote

$$T(f,x) = df(x) + K, \quad T_\varepsilon(f,x) = T(f,x) \cap H(\varepsilon),$$

where

$$K = \{g \in \mathbb{R}^{n+1} \mid g = \lambda e, \; e = (\underbrace{0 \ldots 0}_{n}, -1)^T, \; \lambda \geq 0\},$$

$$H(\varepsilon) = \{z = (z_1 \ldots, z_n, z_{n+1})^T \in \mathbb{R}^{n+1} \mid z_{n+1} = -\varepsilon\}.$$


x
T(|x|,1)

Lemma

For any fixed $\varepsilon \geq 0$ at each point $x \in \mathbb{R}^n$ the equality

$$\partial_\varepsilon f(x) = \left\{ v \in \mathbb{R}^n \mid \begin{pmatrix} v \\ t \end{pmatrix} \in T_\varepsilon(f, x) \right\} \tag{5}$$

holds.

Let's consider the optimization problem: find

$$\inf_{x \in \mathbb{R}^n} f(x).$$

Necessary and sufficient conditions

Let f_1 and f_2 be the polyhedral functions defined on $\mathbb{R}^n$, i.e.

$$f_1(x) = \max_{i \in I} f_{1i}(x), \; f_{1i} = \{\langle a_i, x\rangle + b_i\}, \; I = \{1, \ldots, m\},$$

$$f_2(x) = \max_{j \in J} f_{2j}(x), \; f_{2j}(x) = \{\langle c_j, x\rangle + d_j\}, \; J = \{1, \ldots, p\},$$

Where $a_i, \; c_j \in \mathbb{R}^n, \; b_i, \; d_j \in \mathbb{R}, \; i \in I, \; j \in J$.

Consider the function $f(x) = f_1(x) - f_2(x)$. Then

$$f(x) = \max_{i \in I}\{\langle a_i, x\rangle + b_i\} - \max_{j \in J}\{\langle c_j, x\rangle + d_j\}.$$

Consider some optimization properties of the function f.
Reduce conditions of the unboundedness of the function f on $\mathbb{R}^n$.

Theorem.

For the function f to be unbounded from below in $\mathbb{R}^n$, it is necessary and sufficient that there exist $j^* \in J$ and a vector c_{j^*}, such that the condition

$$c_{j^*} \notin \operatorname{co}\left\{\bigcup_{i \in I} a_i\right\} \tag{6}$$

holds.

Corollary

For the function f to be unbounded from below in $\mathbb{R}^n$*, it is necessary and sufficient that the condition*

$$\text{dom } f_2^* \not\subset \text{dom } f_1^*$$

hold.

Let the function f be bounded from below in $\mathbb{R}^n$.

Theorem

For the point $x^* \in \mathbb{R}^n$ be a global minimizer of the function f on $\mathbb{R}^n$, it is necessary and sufficient, that the condition

$$df_1(x^*) \bigcap \operatorname{co} \left\{ \begin{pmatrix} c_j \\ f_{2j}(x^*) - f_2(x^*) \end{pmatrix}, \begin{pmatrix} c_j \\ 0 \end{pmatrix} \right\} \neq \emptyset \quad \forall j \in J, \tag{7}$$

hold.

Corollary

The condition (7) is equivalent to the following condition

$$0_{n+1} \in \left[df_1(x^*) - \text{co} \left\{ \begin{pmatrix} c_{j^*} \\ f_{2j^*}(x^*) - f_2(x^*) \end{pmatrix}, \begin{pmatrix} c_{j^*} \\ 0 \end{pmatrix} \right\} \right] \quad \forall j \in J.$$

Corollary.

The condition (7) is equivalent to the condition

$$0_{n+1} \in \bigcap_{j \in J} \left[df_1(x^*) - \text{co} \left\{ \begin{pmatrix} c_{j^*} \\ f_{2j^*}(x^*) - f_2(x^*) \end{pmatrix}, \begin{pmatrix} c_{j^*} \\ 0 \end{pmatrix} \right\} \right].$$

Corollary

(The sufficient condition for a global minimum of the function f on $\mathbb{R}^n$.) *If at a point $x^* \in \mathbb{R}^n$ the inclusion*

$$df_2(x^*) \subset df_1(x^*)$$

holds then the point x^* is a global minimizer of the function f on $\mathbb{R}^n$.

Example 6.

Consider the function

$$f(x) = f_1(x) - f_2(x),$$

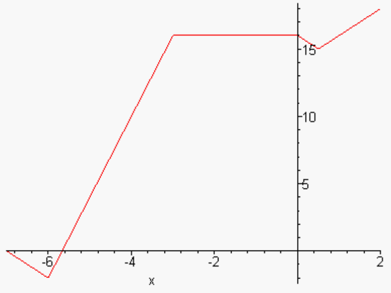
where

$$f_1(x) = \max\left\{|6x+23|, |2x+25|\right\}, \qquad f_2(x) = \max\left\{|4x+9|, |2x+9|\right\}$$

We have

$$f(x) = \begin{cases} -14-2x, & \text{if } -\infty < x \le -6, \\ 34-6x, & \text{if } -6 < x \le -3, \\ 16, & \text{if } -3 < x \le 0, \\ 16-2x, & \text{if } \quad 0 < x \le \frac{1}{2}, \end{cases}$$

On fig. 1 the function f is represented.



It is easy to see, that

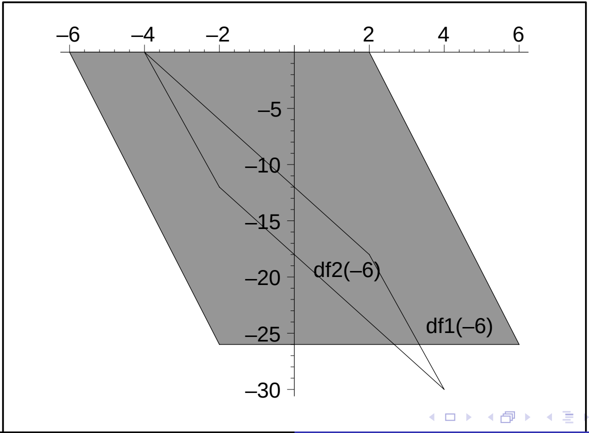
$$\mathrm{dom} f_1^* = \mathrm{co}\{-6, 6\}, \quad \mathrm{dom} f_2^* = \mathrm{co}\{-4, 4\}.$$

Thus, this function is bounded from below ($\mathrm{dom} f_2^* \subset \mathrm{dom} f_1^*$) and unbounded from above. The point $x^* = -6$ is a global minimizer of the function f on $\mathbb{R}$. At this point $f(-6) = -2$ and

$$\partial f_1(-6) = \mathrm{co}\left\{-6, 2\right\}, \; df_1(-6) = \mathrm{co}\left\{\begin{pmatrix} 6 \\ -26 \end{pmatrix}, \begin{pmatrix} -2 \\ -26 \end{pmatrix}, \begin{pmatrix} -6 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}\right\},$$

$$\partial f_2(-6) = -4, \quad df_2(-6) = \mathrm{co}\left\{\begin{pmatrix} 4 \\ -30 \end{pmatrix}, \begin{pmatrix} 2 \\ -18 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 12 \end{pmatrix}\right\}.$$

It is obvious, that the condition (??) holds. See fig.2.



Any point from the interval $(-3, 0)$ is a stationary point of the function f. The functions f_1 and f_2 are differentiable on this interval and $f_1'(x) = 2,\ f_2'(x) = 2$ for any $x \in (-3, 0)$.
Consider the point $x_1 = -2$. We have

$$df_1(-2) = \text{co}\left\{ \begin{pmatrix} 6 \\ -10 \end{pmatrix}, \begin{pmatrix} -2 \\ -42 \end{pmatrix}, \begin{pmatrix} -6 \\ -32 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right\},$$

$$df_2(-2) = \text{co}\left\{ \begin{pmatrix} 4 \\ -4 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ -6 \end{pmatrix}, \begin{pmatrix} -2 \\ -10 \end{pmatrix} \right\}.$$

The condition (??) holds, but the condition (7) does not. See fig 3.

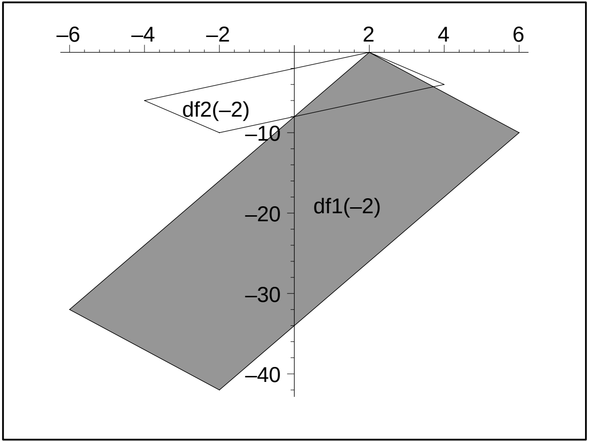


Fig. 3.

Take the point $x_2 = \frac{1}{2}$. This is a strict local minimizer of the function f. Then $f\left(\frac{1}{2}\right) = 15$ and

$$\partial f_1\left(\frac{1}{2}\right) = \text{co}\,\{2,6\}\,,\ df_1(-6) = \text{co}\left\{\begin{pmatrix}6\\0\end{pmatrix},\begin{pmatrix}-2\\-52\end{pmatrix},\begin{pmatrix}-6\\-52\end{pmatrix},\begin{pmatrix}2\\0\end{pmatrix}\right\},$$

$$\partial f_2\left(\frac{1}{2}\right) = 4, \quad df_2(-6) = \text{co}\left\{\begin{pmatrix}4\\0\end{pmatrix},\begin{pmatrix}2\\-1\end{pmatrix},\begin{pmatrix}-4\\-22\end{pmatrix},\begin{pmatrix}-2\\-21\end{pmatrix}\right\}.$$

Observe, that $\partial f_2\left(\frac{1}{2}\right) \subset \text{int}\ \partial f_2\left(\frac{1}{2}\right)$, i.e., the sufficient condition for a strict local minimum is satisfied. The condition (7) does not hold.

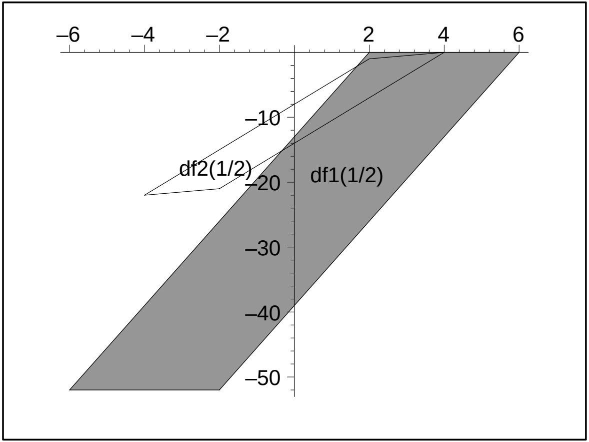


Fig. 4.

Hypodifferentiable functions

Class of hypodifferentiable functions has been allocated by V.F. Demyanov among the nonsmooth functions.
Let $X \subset R^n$ be an open set, $x \in X$ and a function f be defined on X. We say that f is hypodifferentiable at the point x if there exist a convex compact set $df(x) \subset R^{n+1}$ such that

$$f(x+\Delta) = f(x) + \max_{(v,a)^T \in df(x)} [\langle v, \Delta \rangle + a] + o(x, \Delta),$$

where

$$\frac{o(x, \alpha\Delta)}{\alpha} \to 0 \quad \text{if } \alpha \downarrow 0 \quad \forall \Delta \in R^n,$$

$$a \in R,\ \ v \in R^n, \quad \text{co}\,\{x, x+\Delta\} \in X,$$

$$\max_{(v,a)^T \in df(x)} a = 0.$$

The set $df(x)$ is called a hypodifferential at x.

A function f is called continuously hypodifferentiable at a point x if it is hypodifferentiable in some neighbourhood of the point x and there exists a hypodifferential mapping

$$df : R^n \to 2^{R^n},$$

which is Hausdorff continuous at x. A hypodifferential is not uniquely defined. Using continuous hypodifferentials allows to construct numerical optimization methods with continuous descent directions, similar to gradient methods in the smooth case.

1) Let f be continuously differentiable function on R^n. Then f is continuously hypodifferentiable. As a continuous hypo differential can be chosen the set $df(x) = (f'(x), 0)^T \in R^n \times R$, where $f'(x)$ is the gradient of f at x.

2) Let f be convex on R^n. Then f is continuously hypodifferentiable. As a continuous hypodifferential can be chosen the set

$$df(x) = \text{co}\left\{ \bigcup_{\substack{v(z) \in \partial f(z),\ z \in R^n, \\ a = f(z) - f(x) + \langle v(z), x - z \rangle,}} (v(z), a)^T \right\},$$

where $\partial f(z)$ is the subdifferential of the convex function f at $z \in R^n$.

3) Let

$$f(x) = \max_{i \in I} f_i(x), \;\; I = 1, \dots, m,$$

where functions $f_i(x)$, $i \in I$, are continuously differentiable at x on R^n. Then f is continuously hypodifferentiable. As a continuous hypodifferential can be chosen the set

$$df(x) = \text{co} \left\{ \bigcup_{i \in I} (f_i'(x), f_i(x) - f(x))^T \right\} \subset R^n \times R.$$

Examples of continuously hypodifferential functions.

Example 1.

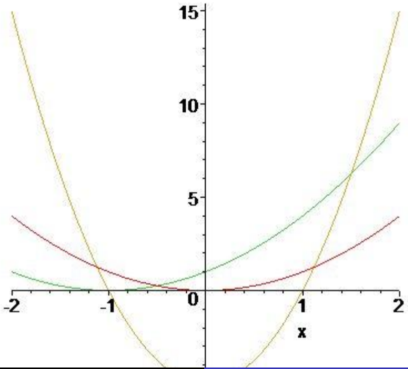
Consider function

$$f(x) = \max_{i \in I} f_i(x), \ \ I = 1, 2, 3, \ \ x \in R,$$

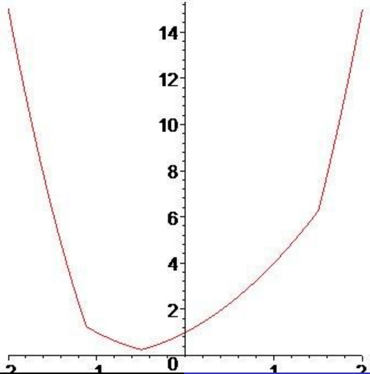
where

$$f_1(x) = x^2, \ \ f_2(x) = (x+1)^2, \ \ x_3(x) = 5x^2 - 5.$$

Examples of continuously hypodifferential functions.



Examples of continuously hypodifferential functions.



Consider points

$$x_1 = -\frac{\sqrt{5}}{2},\ x_2 = -\frac{1}{2},\ x_3 = \frac{3}{2}.$$

At these points the function f is not differentiable. We have

$$f_1(x_1) = \frac{5}{4},\ f_2(x_1) = \frac{9}{4} - \sqrt{5},\ f_3(x_1) = \frac{5}{4},\ f(x_1) = \frac{5}{4},$$

$$f_1(x_2) = \frac{1}{4},\ f_2(x_2) = \frac{1}{4},\ f_3(x_2) = -\frac{15}{4},\ f(x_2) = \frac{1}{4},$$

$$f_1(x_3) = \frac{9}{4},\ f_2(x_3) = \frac{25}{4},\ f_3(x_3) = \frac{25}{4},\ f(x_3) = \frac{25}{4}.$$

The sets

$$\partial f(x_1) = \text{co } \{-\sqrt{5}; 2-\sqrt{5}\} \subset R,$$

$$\partial f(x_2) = \text{co } \{-1; 1\} \subset R, \qquad \partial f(x_3) = \text{co } \{5; 15\} \subset R$$

are the subdifferentials of f at each points.

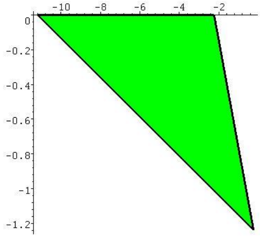
The set

$$df(x) = \text{co}\left\{ \begin{pmatrix} f_1'(x) \\ f_1(x) - f(x) \end{pmatrix}, \begin{pmatrix} f_2'(x) \\ f_2(x) - f(x) \end{pmatrix}, \begin{pmatrix} f_3'(x) \\ f_3(x) - f(x) \end{pmatrix} \right\}$$

is a hypodifferential of f at x.

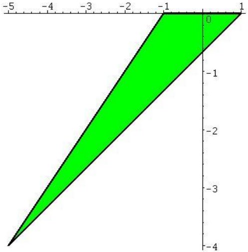
We have

$$df(x_1) = \text{co}\left\{\begin{pmatrix} -\sqrt{5} \\ 0 \end{pmatrix}, \begin{pmatrix} 2-\sqrt{5} \\ 1-\sqrt{5} \end{pmatrix}, \begin{pmatrix} -5\sqrt{5} \\ 0 \end{pmatrix}\right\} \subset R \times R.$$



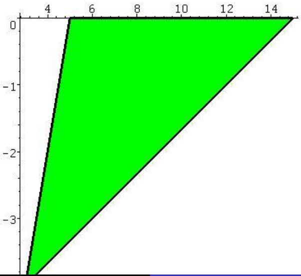
Examples of continuously hypodifferential functions.

$$df(x_2) = \text{co}\left\{\begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ -4 \end{pmatrix}\right\} \subset R \times R.$$



Examples of continuously hypodifferential functions.

$$df(x_3) = \mathrm{co}\left\{ \begin{pmatrix} 3 \\ -4 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \end{pmatrix}, \begin{pmatrix} 15 \\ 0 \end{pmatrix} \right\} \subset R \times R.$$



Necessary condition for a minimum of hypodifferential functions

Let a function f be continuously hypodifferential on $\mathbb{R}^n$ and $df(x)$ be a continuously hypodifferential of f at a point $x \in \mathbb{R}^n$. As the class of hypodifferential functions coincides with the class of subdifferential functions then at every point $x \in \mathbb{R}^n$ then

$$\partial f(x) = \left\{ v \in \mathbb{R}^n \;\middle|\; (v, 0)^T \in df(x) \subset \mathbb{R}^n \times \mathbb{R} \right\}, \qquad (8)$$

where $\partial f(x) \subset \mathbb{R}^n$ is the subdifferential and $df(x) \subset \mathbb{R}^{n+1}$ is a hypodifferential of f at $x \in \mathbb{R}^n$ and the directional derivative $f'(x,g)$ of f at $x \in \mathbb{R}^n$ along a given vector $g \in \mathbb{R}^n$ can be represented in the form

$$f'(x,g) = \max_{v \in \partial f(x)} \langle v, g \rangle \quad \forall g \in \mathbb{R}^n.$$

Theorem 1.

For the point $x^* \in \mathbb{R}^n$ to be a minimum point of f on R^n it is necessary that

$$0_{n+1} \in df(x^*). \tag{9}$$

A point $x^* \in \mathbb{R}^n$ is called a stationary point of f on $\mathbb{R}^n$, if (9) holds. If condition (9) does not hold at x then we project the point 0_{n+1} onto $df(x)$, i.e. solve the optimization problem

$$\min_{z \in df(x)} \|z\| = \|z(x)\|, \; z(x) = (w(x), t(x))^T \in \mathbb{R}^n \times \mathbb{R}.$$

Note that if $0_{n+1} \notin df(x)$, then $w(x) \neq 0_n$.
A direction $-w(x)$ is called a direction of hypodifferentiable descent of the function f on $\mathbb{R}^n$ at the point x. It is continuous and unique.

Lemma 1.

If a point $x \in \mathbb{R}^n$ is not a stationary point for f on R^n then the following inequality

$$f'(x, -w(x)) \leq -\|z(x)\|^2 \tag{10}$$

holds.

Corollary 1.

Let $-w(x)$ be a direction of hypodifferential descent of f at $x \in \mathbb{R}^n$ $(w(x) \neq 0_n)$ and $g(x) = -\frac{w(x)}{||w(x)||}$, then

$$f'(x, g(x)) \leq -||z(x)|| \leq -||w(x)||. \tag{11}$$

Continuous methods

Let a function f be defined, locally Lipschitz and continuously hypodifferentiable on $\mathbb{R}^n$.
Assume that a point $x \in R^n$ is not a stationary point of f on $\mathbb{R}^n$, i.e. condition (9) does not hold.
Since $df(x)$ is continuous then there exists a direction $-w(x)$ which is also a continuous descent direction of f at x as follows from (10).

Most iterative methods generate a minimizing sequence $\{x_k\}$ according to the rule

$$x_{k+1} = x_k + \alpha_k g(x_k),$$

where $g(x_k)$ is a descent direction (if $g(x_k) \neq 0_n$) at x_k and $\alpha_k, \alpha_k > 0$, is a step size along this direction.
As in smooth cases we consider two variants of choosing of a step size on each iteration.

1. One dimensional minimization.

A step size α_k is chosen from the condition

$$\alpha_k = \arg \min_{\alpha > 0} f(x_k - \alpha w(x_k)). \tag{12}$$

2. The Armijo rule

Fix any parameter $\theta \in (0, 0.5]$ and find the first value of $i_k = 0, 1, \ldots$ for which will be performed the following inequality

$$f(x_k - (0.5)^{i_k} w(x_k)) \leq f(x_k) - (0.5)^{i_k} \theta \|w(x_k)\|^2 \tag{13}$$

and put $\alpha_k = (0.5)^{i_k}$.

Choose an arbitrary point $x_0 \in \mathbb{R}^n$. If $0_{n+1} \in df(x_0)$, then x_0 is a stationary point of f on $\mathbb{R}^n$.
Let $x_k \in \mathbb{R}^n$ have already been found. If $0_{n+1} \in df(x_k)$, then x_k is a stationary point of f on $\mathbb{R}^n$.
Otherwise, put

$$x_{k+1} = x_k - \alpha_k w(x_k) = x_k - \alpha_k w_k,$$

where $-w(x_k) = -w_k$ is a hypodifferentiable descent direction at x_k, and step size α_k is chosen by using the one dimensional minimization or the Armijo rule (13). If the sequence $\{x_k\}$ is finite, then the last obtained point will be a stationary point.

Consider the case when the sequence $\{x_k\}$ is infinite. Then the sequence $\{f(x_k)\}$ is monotonically decreasing, therefore, this method will be relaxation.
Let the level set

$$\mathcal{L} = \mathcal{L}(x_0) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\} \tag{14}$$

be bounded.

Theorem 2.

Every limit point of the sequence $\{x_k\}$ is a stationary point of the function f on $\mathbb{R}^n$.

Remark 1.

If the function f is continuously differentiable then the described methods coincide with respective gradient methods. Consequently, these hypodifferentiable descent methods also as gradient methods badly converge near a stationary point.

Finding of hypodifferentiable descent directions

The problem of finding of a hypodifferentiable descent direction of a continuous hypodifferentiable function f at the point x reduces to the solution of the following quadratic programming problem

$$\min_{z \in df(x)} \langle z, z \rangle = \min_{z \in df(x)} \|z\|^2 = \|z(x)\|^2,$$

where

$$z = (w, t)^T \in R^n \times \mathbb{R}, \; z(x) = (w(x), t(x)) \in \mathbb{R}^n \times R.$$

Consider a variant of this algorithm in which continuous hypodifferential $df(x)$ at the point x is a polyhedron in $\mathbb{R}^{n+1}$.

Let

$$f(x) = \max_{i \in I} f_i(x), \;\; I = 1, \dots, m,$$

where $f_i, i \in I$, are continuously differentiable functions on $\mathbb{R}^n$. Then the set

$$df(x) = \text{co} \; \left\{ \bigcup_{i \in I} \begin{pmatrix} f_i'(x) \\ f_i(x) - f(x) \end{pmatrix} \right\} \subset R^n \times R$$

is a continuous hypodifferential of f at $x \in \mathbb{R}^n$, because the mapping $df : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is continuous in the Hausdorff metric.

Let

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; \frac{1}{2}\langle A_1 x, x\rangle + \langle b_1, x\rangle + c_1 \le 0 \right\},$$

$$Y = \left\{ y \in \mathbb{R}^n \;\middle|\; \frac{1}{2}\langle A_2 y, y\rangle + \langle b_2, y\rangle + c_2 \le 0 \right\},$$

where matrices A_1, A_2 of size $n \times n$ are positive definite, $b_1, b_2 \in \mathbb{R}^n$, $c_1, c_2 \in \mathbb{R}$.

Suppose that X and Y are two nonempty sets. It is necessary to solve the optimization problem

$$||x - y|| \to \min, x \in X, y \in Y,$$

where $|| * ||$ is the Euclidean norm.

Consider this problem in R^{2n}.
It is necessary to solve

$$\frac{1}{2}||x-y||^2 \to \min,\ x \in X, y \in Y. \tag{15}$$

Denote by

$$f(z) = \frac{1}{2}\langle E_1 z, z\rangle = \frac{1}{2}\langle x-y, x-y\rangle, z=(x,y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

where

$E_1 = \begin{pmatrix} E_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & -E_{n\times n} \end{pmatrix}$ is the matrix of size $2n \times 2n$, $E_{n\times n}$ is the identity matrix of size $n \times n$, $0_{n\times n}$ is the zero matrix of size $n \times n$,

$$\varphi_1(z) = \frac{1}{2}\langle A_1 x, x\rangle + \langle b_1, x\rangle + c_1,$$

$$\varphi_2(z) = \frac{1}{2}\langle A_2 y, y\rangle + \langle b_2, y\rangle + c_2,$$

$$\varphi(z) = \max\{0, \varphi_1(z), \varphi_2(z)\},$$

$$Z = \{z = (x, y) \in R^n \times R^n \mid \varphi(z) = 0\}.$$

$$f(z) \to \min, z \in Z.$$

$$F(z, c) = f(z) + c\varphi(z), \ \ c \geq 0.$$

In our case the function $F(z, c)$ is an exact penalty function.

$$F(z) = \frac{1}{2}\langle x-y, x-y\rangle + c\max\{0, \varphi_1(z), \varphi_2(z)\} =$$

$$= \max\left\{\frac{1}{2}\langle x-y, x-y\rangle, \frac{1}{2}\langle x-y, x-y\rangle + c\left[\frac{1}{2}\langle A_1x, x\rangle + \langle b_1, x\rangle + c_1\right],\right.$$

$$\left.\frac{1}{2}\langle x-y, x-y\rangle + c\left[\frac{1}{2}\langle A_2y, y\rangle + \langle b_2, y\rangle + c_2\right]\right\},$$

For minimizing $F(z, c)$ it is possible to apply the method of hypodifferential descent.

Using the formula of hypodifferential calculus, we have

$$dF(z, c) = \text{co}\ \{t_0(z, c), t_1(z, c), t_2(z, c)\},$$

$$t_0(z, c) = \begin{pmatrix} x - y \\ -(x - y) \\ f(z) - F(z, c) \end{pmatrix},$$

$$t_1(z,c) = \begin{pmatrix} x - y + c\,(A_1 x + b_1) \\ y - x \\ c\varphi_1(z) - F(z,c) \end{pmatrix},$$

$$t_2(z,c) = \begin{pmatrix} x - y \\ y - x + c\,(A_2 y + b_2) \\ c\varphi_2(z) - F(z,c) \end{pmatrix},$$

$$t_0(z,c), t_1(z,c), t_2(z,c) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}.$$

Thus, in our problem a continuous hypodifferential is a triangle in space $\mathbb{R}^n \times R^n \times \mathbb{R}$. Therefore, to find the direction of a steepest descent it is necessary to project the zero point onto the triangle. Consider this procedure.

Let $X \subset \mathbb{R}^n$ is a triangle with vertices $a_1, a_2, a_3 \in \mathbb{R}^n$, that is

$$X = \text{co}\,\{a_1, a_2, a_3\} \subset \mathbb{R}^n.$$

Consider a optimization problem

$$\min_{x \in X} ||x||^2 \tag{16}$$

This problem can be reduced to a quadratic programming problem. In fact, as any point $x \in X$ can be represented as:

$$x = \lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3, \ \lambda_1 + \lambda_2 + \lambda_3 = 1, \ \lambda_1, \lambda_2, \lambda_3 \geq 0.$$

Then problem (16) is equivalent to the following optimization problem

$$\min_{\lambda\in\Lambda} ||\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3||^2 = \min_{\lambda\in\Lambda} \langle A\lambda, \lambda\rangle, \qquad (17)$$

where

$$\Lambda = \left\{\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3 \;\middle|\; \lambda_1 + \lambda_2 + \lambda_3 = 1,\ \lambda_1, \lambda_2, \lambda_3 \geq 0\right\},$$

$$A = \begin{pmatrix} \langle a_1, a_1\rangle & \langle a_1, a_2\rangle & \langle a_1, a_3\rangle \\ \langle a_1, a_2\rangle & \langle a_2, a_2\rangle & \langle a_2, a_3\rangle \\ \langle a_1, a_3\rangle & \langle a_2, a_3\rangle & \langle a_3, a_3\rangle \end{pmatrix}.$$

But the solution of problem (16) can also be found in the following way. If the points $\{a_1, a_2, a_3\}$ are not on the line then the vectors $e_1 = a_2 - a_1,\ e_2 = a_3 - a_1$ are linearly independent. The set

$$M = a_1 + \lambda_1 e_1 + \lambda_2 e_2 \subset \mathbb{R}^n,\ \lambda_1, \lambda_2 \in \mathbb{R}$$

is a linear manifold and $X \subset M$.

Project the zero point onto the set of M. Introduce the function

$$F(\lambda) = \langle a_1 + \lambda_1 e_1 + \lambda_2 e_2, a_1 + \lambda_1 e_1 + \lambda_2 e_2 \rangle, \ \lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2.$$

We have

$$\frac{\partial F(\lambda)}{\partial \lambda_1} = 2(\langle a_1, e_1 \rangle + \lambda_1 \langle e_1, e_1 \rangle + \lambda_2 \langle e_1, e_2 \rangle),$$

$$\frac{\partial F(\lambda)}{\partial \lambda_2} = 2(\langle a_1, e_2 \rangle + \lambda_1 \langle e_1, e_2 \rangle + \lambda_2 \langle e_2, e_2 \rangle).$$

Denote by

$$\hat{A} = \begin{pmatrix} \langle e_1, e_1 \rangle & \langle e_1, e_2 \rangle \\ \langle e_1, e_2 \rangle & \langle e_2, e_2 \rangle \end{pmatrix}, \quad \hat{b} = \begin{pmatrix} \langle a_1, e_1 \rangle \\ \langle a_1, e_2 \rangle \end{pmatrix}.$$

Calculate the vector

$$\lambda^* = -\hat{A}\,\hat{b}, \quad \lambda^* = (\lambda_1^*, \lambda_2^*) \in \mathbb{R}^2.$$

Projecting of the zero point onto a segment.

Then the projection of the zero point onto the linear manifold M is calculated by the formula

$$x^* = a_1 + \lambda_1^* e_1 + \lambda_2^* e_2.$$

If $x^* \in X$, then we receive a solution of problem (16).
Otherwise, project the zero point onto three segments. Define

$$x_1^* = \arg\min_{x \in X_1} ||x||^2, \quad x_2^* = \arg\min_{x \in X_2} ||x||^2, \quad x_3^* = \arg\min_{x \in X_3} ||x||^2,$$

where

$$X_1 = \mathrm{co}\{a_1, a_2\}, \quad X_2 = \mathrm{co}\{a_1, a_3\}, \quad X_3 = \mathrm{co}\{a_2, a_3\}.$$

Obviously, the point with the smallest norm is the solution of problem (16).

Projecting of the zero point onto a segment.

Describe the procedure of projecting the zero point onto a segment. The problem is to find the vector of least length co $\{a, b\}$, $a, b \in R^n$, $a \neq b$. Any vector of this segment can be represented in the form

$$x = \mu a + (1 - \mu)b, \ \mu \in [0, 1].$$

Introduce a function

$$t(\mu) = (\mu a + (1 - \mu)b)^2 = (\mu(a - b) + b)^2 = \langle \mu(a - b) + b, \mu(a - b) + b \rangle.$$

Then it is necessary to solve an optimization problem

$$t(\mu) \to \min, \ \mu \in [0, 1].$$

Calculate

$$t'(\mu) = 2\mu\langle a-b, a-b\rangle + 2\langle a-b, b\rangle.$$

Obviously that $t'(\mu) = 0$ under

$$\mu^* = -\frac{\langle a-b, b\rangle}{\langle a-b, a-b\rangle}.$$

If $\mu^* > 1$, then put $\mu^* = 1$. If $\mu^* < 0$, then put $\mu^* = 0$.
Thus, the vector

$$x^* = \mu^*(a-b) + b = -\frac{\langle a-b, b\rangle}{\langle a-b, a-b\rangle}(a-b) + b =$$

$$\frac{\langle a-b, b\rangle}{\langle a-b, a-b\rangle}(b-a) + b.$$

is our solution.