J Optim Theory Appl
DOI 10.1007/s10957-015-0796-7

Second-Order Minimization Method for Nonsmooth Functions Allowing Convex Quadratic Approximations of the Augment

M. E. Abbasov[1]

Received: 25 October 2014 / Accepted: 10 August 2015

Abstract Second-order methods play an important role in the theory of optimization. Due to the usage of more information about considered function, they give an opportunity to find the stationary point faster than first-order methods. Well-known and sufficiently studied Newton's method is widely used to optimize smooth functions. The aim of this work is to obtain a second-order method for unconstrained minimization of nonsmooth functions allowing convex quadratic approximation of the augment. This method is based on the notion of coexhausters—new objects in nonsmooth analysis, introduced by V. F. Demyanov. First, we describe and prove the second-order necessary condition for a minimum. Then, we build an algorithm based on that condition and prove its convergence. At the end of the paper, a numerical example illustrating implementation of the algorithm is given.

Keywords Nonsmooth analysis · Nondifferentiable optimization · Coexhausters

Mathematics Subject Classification 49J52 · 90C30 · 65K05

1 Introduction

To study the local behavior of the function, the main part of the augment of a simple nature is always allocated. Thus, in a smooth case it is a linear or quadratic function, which is defined by means of the gradient and Hessian. Researchers use this approach for constructing optimization algorithms. Newton method can be mentioned as a rep-

✉ M. E. Abbasov
abbasov.majid@gmail.com; m.abbasov@spbu.ru

1 Saint-Petersburg University, St. Petersburg State University, SPbSU, SPbU,
7/9 Universitetskaya nab., St. Petersburg 199034, Russia

Published online: 21 August 2015

resentative of second-order methods. There have been several attempts to generalize the algorithm [1–3].

Various specific tools are employed in nonsmooth analysis to consider different classes of functions. For example, one can use exhausters to study an arbitrary directionally differentiable function. The concept of exhausters arose from works of Pshenichny [4], Rubinov [5,6] and Demyanov [7,8]. Exhausters are families of convex compact sets by means of which the main part of the augment can be represented in the form of min–max or max–min of linear functions. Optimality conditions were described in terms of exhausters [9,10]. This promoted the construction of optimization algorithms. Unfortunately, the fact that exhauster mapping is not continuous in Hausdorff metric produced computational problems [11]. This is why a new tool, free from this shortcoming, was introduced. Coexhausters are families of convex compact sets by means of which the main part of the augment can be represented in the form of min–max or max–min of affine functions. This notion was first introduced in [8]. It is possible to specify the class of functions which allows the expansion by means of continuous coexhausters in Hausdorff metric [5]. These objects were studied in [12–15], and the optimality conditions in terms of coexhausters were obtained. This gave an opportunity to describe new effective optimization methods. Further development of this fruitful idea led to the appearance of second-order coexhausters [5]. So the problem of constructing optimization algorithms, that are based on this new tool, arose.

2 Preliminary Information

Let a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ be given. The function f is said to have an upper second-order coexhauster at a point $x \in \mathbb{R}^n$ iff the representation

$$f(x+\Delta) = f(x) + \min_{C \in \widehat{E}^*(x)} \max_{[a,v,A] \in C} \left(a + \langle v, \Delta \rangle + \frac{1}{2} \langle A\Delta, \Delta \rangle \right) + o_x(\Delta), \quad (1)$$

is valid, where

$$\lim_{\alpha \downarrow 0} \frac{o_x(\alpha\Delta)}{\alpha^2} = 0 \quad \forall \Delta \in \mathbb{R}^n, \quad (2)$$

and $\widehat{E}^*(x)$—is a family of compacts in $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n\times n}$, called the upper second-order coexhauster.

The function f is said to have a lower second-order coexhauster at a point $x \in \mathbb{R}^n$ iff the representation

$$f(x+\Delta) = f(x) + \max_{C \in \widehat{E}_*(x)} \min_{[a,v,A] \in C} \left(a + \langle v, \Delta \rangle + \frac{1}{2} \langle A\Delta, \Delta \rangle \right) + o_x(\Delta), \quad (3)$$

is valid, where $o_x(\Delta)$ satisfies (2) and $\widehat{E}_*(x)$—is a family of compacts in $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n\times n}$, called the lower second-order coexhauster.

There is a wide class of functions admitting representations (1) and (3).

In this paper, we will consider the problem of minimizing functions whose upper second-order coexhausters consist of a single set

$$f(x+\Delta) = f(x) + \max_{[a,v,A]\in C(x)} \left(a + \langle v, \Delta\rangle + \frac{1}{2}\langle A\Delta, \Delta\rangle \right) + o_x(\Delta), \qquad (4)$$

where all the matrices $A \in C(x)$ are assumed to be positive definite for all x in the domain of f. Furthermore, we will study the case of finite set $C(x)$, which is the most common in practice.

Let us construct a bijection between the set $C(x)$ and some set of indexes I. That is $C(x) = \left\{[a_i(x), v_i(x), A_i(x)]:\ i \in I\right\}$. For the sake of convenience, instead of (4) we will use an equivalent notation

$$f(x+\Delta) = f(x) + \max_{i\in I} \left(a_i(x) + \langle v_i(x), \Delta\rangle + \frac{1}{2}\langle A_i(x)\Delta, \Delta\rangle \right) + o_x(\Delta).$$

3 Necessary Second-Order Extremum Condition

In this section, we describe the optimality condition for the studied function and then construct the optimization algorithm based on this condition.

Theorem 3.1 *Let a continuous function $f(x)\colon \mathbb{R}^n \to \mathbb{R}$ allowing the expansion*

$$f(x+\Delta) = f(x) + \max_{i\in I} \left(a_i(x) + \langle v_i(x), \Delta\rangle + \frac{1}{2}\langle A_i(x)\Delta, \Delta\rangle \right) + o_x(\Delta),$$

where I is a finite set of indexes, all the matrices $A_i(x) \in I$ are positive definite for all $x \in \mathbb{R}^n$, and $o_x(\Delta)$ satisfy

$$\lim_{\alpha\downarrow 0} \frac{o_x(\alpha\Delta)}{\alpha^2} = 0, \quad \forall \Delta \in \mathbb{R}^n.$$

If $x_* = \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x)$, then

$$0_n = \operatorname*{argmin}_{\Delta\in\mathbb{R}^n} \max_{i\in I} \left(a_i(x_*) + \langle v_i(x_*), \Delta\rangle + \frac{1}{2}\langle A_i(x_*)\Delta, \Delta\rangle \right). \qquad (5)$$

Proof First note that the main part of the augment of f

$$\phi_{x_*}(\Delta) = \max_{i\in I} \left(a_i(x_*) + \langle v_i(x_*), \Delta\rangle + \frac{1}{2}\langle A_i(x_*)\Delta, \Delta\rangle \right)$$

is the convex function. Moreover, since f is continuous $\phi_{x_*}(0_n) = 0$ holds.

Prove the theorem by contradiction. Assume that conditions of the theorem are valid, but (5) is not satisfied at x_*, that is

$$0_n \neq \widetilde{g} = \operatorname*{argmin}_{\Delta \in \mathbb{R}^n} \max_{i \in I} \left(a_i(x_*) + \langle v_i(x_*), \Delta \rangle + \frac{1}{2} \langle A_i(x_*)\Delta, \Delta \rangle \right).$$

If so, then function ϕ has a direction of descent $\widetilde{\Delta} = \dfrac{\widetilde{g}}{\|\widetilde{g}\|}$ at $\Delta = 0_n$.

The behavior of the main part of the augment f, while moving from the point x_* along the direction $\widetilde{\Delta}$ with some positive step α, is defined by a function

$$\phi_{x_*}(\alpha \widetilde{\Delta}) = \max_{i \in I} \phi_{i,x_*}(\alpha \widetilde{\Delta}),$$

where

$$\phi_{i,x_*}(\alpha \widetilde{\Delta}) = a_i(x_*) + \alpha \langle v_i(x_*), \widetilde{\Delta} \rangle + \frac{\alpha^2}{2} \langle A_i(x_*)\widetilde{\Delta}, \widetilde{\Delta} \rangle.$$

Let $R_{x_*}(0_n) = \{i \in I \colon a_i(x_*) = \phi_{x_*}(0_n) = 0\}$ be a set of indexes of functions active at zero. Then, due to the fact that ϕ_{x_*} is decreasing at zero along the direction $\widetilde{\Delta}$, its directional derivative along $\widetilde{\Delta}$ should be strictly negative:

$$\exists \mu > 0 \colon \phi'_{x_*}(0_n, \widetilde{\Delta}) = \max_{i \in R_{x_*}(0_n)} (v_i, \widetilde{\Delta}) = -\mu < 0.$$

Finally, taking into account the continuity of f, conditions imposing on o_x by the theorem, boundedness of coexhauster and finiteness of the set I, it can be stated that there is $\widetilde{\alpha}$ small enough that for every positive $\alpha < \widetilde{\alpha}$ the chain

$$\begin{aligned} & f(x_* + \alpha \widetilde{\Delta}) - f(x_*) \\ & \quad = \max_{i \in I} \left(a_i(x_*) + \alpha \langle v_i(x_*), \widetilde{\Delta} \rangle + \frac{\alpha^2}{2} \langle A_i(x_*)\widetilde{\Delta}, \widetilde{\Delta} \rangle \right) + o_{x_*}(\alpha \widetilde{\Delta}) \\ & \quad = \max_{i \in R_{x_*}(0_n)} \left(a_i(x_*) + \alpha \langle v_i(x_*), \widetilde{\Delta} \rangle + \frac{\alpha^2}{2} \langle A_i(x_*)\widetilde{\Delta}, \widetilde{\Delta} \rangle \right) + o_{x_*}(\alpha \widetilde{\Delta}) \\ & \quad \leq -\alpha\mu + \frac{\alpha^2}{2} \max_{i \in I} \|A_i(x_*)\| + \frac{\alpha^2}{2} \max_{i \in I} \|A_i(x_*)\| < 0. \end{aligned}$$

is valid. This contradicts the fact that x_* is the local minimum of f. □

Now we can proceed to the method itself.

4 Minimization Method

Let x_0 be an initial point. Suppose that we have obtained a point x_k. Define

$$\begin{aligned} y_k &= \operatorname*{argmin}_{y\in\mathbb{R}^n} \max_{i\in I} \left(a_i(x_k) + \alpha\langle v_i(x_k), y\rangle + \frac{\alpha^2}{2}\langle A_i(x_k)y, y\rangle \right), \\ \Delta_k &= \frac{y_k}{\|y_k\|}, \\ \alpha_k &= \operatorname*{argmin}_{\alpha} f\left(x_k + \alpha\Delta_k\right), \\ x_{k+1} &= x_k + \alpha_k\Delta_k. \end{aligned}$$

As a result, we construct the sequence $\{x_k\}$ such that

$$f(x_{k+1}) < f(x_k). \tag{6}$$

Theorem 4.1 *Let a set*

$$P = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$$

be bounded, mapping $C(x) = \big\{[a_i(x), v_i(x), A_i(x)] : i \in I\big\}$ *is continuous in Hausdorff metric,* x_* *is the limit point of the sequence* $\{x_k\}$*, and function* $o_x(\Delta)$ *satisfies the condition* $\frac{o_x(\alpha\Delta)}{\alpha^2} \xrightarrow[\alpha\downarrow 0]{} 0$ *uniformly with respect to* x *in some neighborhood of* x_* *and with respect to* Δ *in* $\mathbb{S} = \{\Delta \in \mathbb{R}^n : \|\Delta\| = 1\}$. *Then, the point* x_* *is a stationary point of* f *on* $\mathbb{R}^n$*; that is, condition* (5) *satisfies at* x_*.

Proof The existence of the limit point of the sequence $\{x_k\}$ follows from the boundedness of the set P and inequality (6). Thus, there is a subsequence $\{x_{k_m}\}$ converging to x_*.

Prove the theorem by contradiction. Assume that the conditions of the theorem are valid, but (5) is not satisfied at x_*; that is, $\widetilde{g} \neq 0_n$, where

$$\widetilde{g} = \operatorname*{argmin}_{\Delta\in\mathbb{R}^n} \max_{i\in I} \left(a_i(x_*) + \langle v_i(x_*), \Delta\rangle + \frac{1}{2}\langle A_i(x_*)\Delta, \Delta\rangle \right) = \operatorname*{argmin}_{\Delta\in\mathbb{R}^n} \phi_{x_*}(\Delta).$$

This means that the convex function $\phi_{x_*}(\Delta)$ is strictly decreasing in point 0_n along the direction $\widetilde{\Delta} = \widetilde{g}/\|\widetilde{g}\|$; that is, there exists $\mu > 0$ such that

$$\phi'_{x_*}(0_n, \widetilde{\Delta}) = -\mu < 0.$$

Denote $\Lambda = \max_{i\in I} \|A_i(x_*)\|$. Due to the continuity of the coexhauster mapping, there exists ε_1-neighborhood of the point x_*, such that for all x in this neighborhood inequality

$$\max_{i\in I} \|A_i(x)\| \leq \frac{3}{2}\Lambda,$$

is valid.

Consider the function $\phi_{x_*}(\Delta)$. As I is the finite set, one can find a δ-neighborhood centered at zero in which not active functions are the same as in zero. That is for any Δ in such a neighborhood will hold an equality

$$\max_{i\in I}\left(a_i(x_*)+\langle v_i(x_*),\Delta\rangle+\frac{1}{2}\langle A_i(x_*)\Delta,\Delta\rangle\right)$$
$$=\max_{i\in R_{x_*}(0_n)}\left(a_i(x_*)+\langle v_i(x_*),\Delta\rangle+\frac{1}{2}\langle A_i(x_*)\Delta,\Delta\rangle\right),$$

where $R_{x_*}(0_n)=\{i\in I\colon a_i(x_*)=\phi_{x_*}(0_n)=0\}$ is the set of indexes of functions active at zero. Since the coexhauster mapping is continuous, there is ε_2-neighborhood of the point x_* such that for any x in this neighborhood and Δ in $\frac{\delta}{2}$-neighborhood centered at zero

$$\max_{i\in I}\left(a_i(x)+\langle v_i(x),\Delta\rangle+\frac{1}{2}\langle A_i(x)\Delta,\Delta\rangle\right)$$
$$=\max_{i\in R_{x_*}(0_n)}\left(a_i(x)+\langle v_i(x),\Delta\rangle+\frac{1}{2}\langle A_i(x)\Delta,\Delta\rangle\right),$$

is valid.

Since $\max\limits_{i\in R_{x_*}(0_n)}\langle v_i(x_*),\widetilde{\Delta}\rangle=\phi'_{x_*}(0_n,\widetilde{\Delta})<-\mu<0$, one can conclude that $\langle v_i(x_*),\widetilde{\Delta}\rangle<-\mu$ is valid for any $i\in R_{x_*}(0_n)$. Due to the fact that the coexhauster mapping is continuous, $\Delta_{k_m}\to\widetilde{\Delta}$ holds, and thus, there is ε_3-neighborhood of the point x_*, such that for any x in this neighborhood

$$\langle v_i(x_{k_m}),\Delta_{k_m}\rangle<-\frac{\mu}{2}<0,\quad \forall i\in R_{x_*}(0_n)$$

is true.

In virtue of the theorem condition, there is $B(x_*,\varepsilon_4)$—ε_4-neighborhood of the point x_* such that

$$\exists\widetilde{\alpha}>0\colon \forall x\in B(x_*,\varepsilon_4),\ \forall\Delta\in\mathbb{S},\ \forall\alpha\in[0,\widetilde{\alpha}]\quad |o_x(\alpha\Delta)|<\frac{\alpha^2}{2}.$$

Choose $\varepsilon=\min\{\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4\}$, $\alpha_0=\min\{\widetilde{\alpha},\frac{\delta}{2},\frac{\mu}{6},\frac{\mu}{6\Lambda}\}$. Starting with some sufficiently large $M>0$, all x_{k_m} will be included in ε-neighborhood of the x_*. Then, for $m\geq M$, we can write

$$f(x_{k_{m+1}})-f(x_{k_m})$$
$$=\max_{i\in I}\left(a_i(x_{k_m})+\alpha_{k_m}\langle v_i(x_{k_m}),\Delta_{k_m}\rangle+\frac{\alpha_{k_m}^2}{2}\langle A_i(x_{k_m})\Delta_{k_m},\Delta_{k_m}\rangle\right)$$
$$+o_{x_{k_m}}(\alpha_{k_m}\Delta_{k_m})$$

$$\leq \max_{i\in I}\left(a_i(x_{k_m})+\alpha_0\langle v_i(x_{k_m}),\Delta_{k_m}\rangle+\frac{\alpha_0^2}{2}\langle A_i(x_{k_m})\Delta_{k_m},\Delta_{k_m}\rangle\right)$$
$$+o_{x_{k_m}}(\alpha_0\Delta_{k_m})$$
$$=\max_{i\in R_{x_*}(0_n)}\left(a_i(x_{k_m})+\alpha_0\langle v_i(x_{k_m}),\Delta_{k_m}\rangle+\frac{\alpha_0^2}{2}\langle A_i(x_{k_m})\Delta_{k_m},\Delta_{k_m}\rangle\right)$$
$$+o_{x_{k_m}}(\alpha_0\Delta_{k_m})<-\alpha_0\frac{\mu}{2}+\frac{\alpha_0^2}{2}\Lambda+\frac{\alpha_0^2}{2}<-\alpha_0\frac{\mu}{3}<0.$$

Hence considering (6), we get $f(x_k)\to-\infty$, which contradicts to the fact that the continuous functions f on the bounded closed set P are bounded. □

Remark 4.1 Let us note that these results remain true also for functions, allowing positive semidefinite quadratic approximation of the augment.

Remark 4.2 Since the algorithm must converge to the point where condition (5) holds, the value of $\|y_k\|$ can be taken as a stopping criteria.

Example 4.1 Consider a function $f:\mathbb{R}^2\to\mathbb{R}$, $f(x)=\max\{f_1(x);f_2(x)\}$, where

$$f_1(x)=(x^1-1)^4+(x^2)^4,\quad f_2(x)=(x^1+1)^4+(x^2)^4.$$

Denote components of x as x^1 and x^2. Choose $x_0=(1,2)^{\mathrm{T}}$. It is obvious that

$$f(x+\Delta)=f(x)+\max_{i\in\{1;2\}}\left(a_i(x)+\langle v_i(x),\Delta\rangle+\frac{1}{2}\langle A_i(x)\Delta,\Delta\rangle\right)+o_x(\Delta),$$

where

$$a_1(x)=f_1(x)-f(x),\quad a_2(x)=f_2(x)-f(x),$$

$$v_1(x)=\begin{pmatrix}4(x^1-1)^3\\4(x^2)^3\end{pmatrix},\quad v_2(x)=\begin{pmatrix}4(x^1+1)^3\\4(x^2)^3\end{pmatrix},$$
$$A_1(x)=\begin{pmatrix}12(x^1-1)^2&0\\0&12(x^2)^2\end{pmatrix},\quad A_2(x)=\begin{pmatrix}12(x^1+1)^2&0\\0&12(x^2)^2\end{pmatrix},$$

$$\lim_{\alpha\downarrow 0}\frac{o_x(\alpha\Delta)}{\alpha^2}=0\quad\forall\Delta\in\mathbb{R}^n.$$

As

$$a_1(x_0)=-16,\quad a_2(x_0)=0,$$
$$v_1(x_0)=\begin{pmatrix}0\\32\end{pmatrix},\quad v_2(x_0)=\begin{pmatrix}32\\32\end{pmatrix},$$
$$A_1(x_0)=\begin{pmatrix}0&0\\0&48\end{pmatrix},\quad A_2(x_0)=\begin{pmatrix}48&0\\0&48\end{pmatrix},$$

the quadratic approximation of the main part of the augment at x_0 has the form

$$\max\{32y^1 + 32y^2 + 24(y^1)^2 + 24(y^2)^2; -16 + 32y^2 + 24(y^2)^2\}.$$

This function attains minimum at the point $y_0 = (-2/3, -2/3)^{\mathrm{T}}$. Therefore, doing a one-dimensional minimization along normalized direction Δ_0, where $\Delta_0 = -\left(\sqrt{2}/2, \sqrt{2}/2\right)^{\mathrm{T}}$, we get $\alpha_0 = \sqrt{2}$, $x_1 = (0, 1)^{\mathrm{T}}$.

Proceeding similarly, we obtain $y_1 = -(0, 1/3)^{\mathrm{T}}$, $\Delta_1 = -(0, 1)^{\mathrm{T}}$, $\alpha_1 = 3$, $x_2 = (0, 0)^{\mathrm{T}}$. The quadratic approximation of the main part of the augment of f at x_2 is

$$\max\{4y^1 + 6(y^1)^2; -4y^1 + 6(y^1)^2\}$$

and hence $y_2 = (0, 0)^{\mathrm{T}}$. This means that the necessary condition of minimum (5) holds at the point x_2.

5 Conclusions

Let us note that, for an implementation of the proposed method, we have to find a minimum of the quadratic approximation at each iteration. This problem is not related to the searching of directions of descent, and therefore, it can be solved by using different smoothing techniques.

When the specified method is applied to a smooth convex function, we get a well-known modified Newton's method [16]. Thus, the proposed algorithm is a generalization of Newton's method for a class of nonsmooth functions, allowing convex quadratic approximation of the augment.

Acknowledgments The author dedicates the article to the blessed memory of Professor Vladimir Fedorovich Demyanov, whose ideas led to the appearance of this work. The work is supported by the Saint-Petersburg State University under Grant No. 9.38.205.2014.

References

1. Panin, V.M.: About second order method for descrete minimax problem. Comput. Math. Math. Phys. **19**(1), 88–98 (1979). (in Russian)
2. Qi, L., Sun, J.: A nonsmooth version of Newtons method. Math. Program. **58**(1–3), 353–367 (1993)
3. Gardasevic-Filipovic, M., Duranovic-Milicic, N.: An algorithm using Moreau–Yosida regularization for minimization of a nondifferentiable convex function. Filomat **27**(1), 15–22 (2013)
4. Pshenichny, B.N.: Convex analysis and extremal problems. Nauka, Moscow (1980). (in Russian)
5. Demyanov, V.F., Rubinov, A.M.: Constructive nonsmooth analysis. Approximation and optimization, vol. 7. Peter Lang, Frankfurt am Main (1995)
6. Demyanov, V.F., Rubinov, A.M.: Exhaustive families of approximations revisited. From convexity to nonconvexity. Nonconvex optimization and its applications, vol. 55, pp. 43–50. Kluwer Academic, Dordrecht (2001)
7. Demyanov, V.F.: Exhausters af a positively homogeneous function. Optimization **45**(1–4), 13–29 (1999)
8. Demyanov, V.F.: Exhausters and convexificators—new tools in nonsmooth analysis. In: Demyanov, V., Rubinov, A. (eds.) Quasidifferentiability and related topics, pp. 85–137. Kluwer Academic Publishers, Dordrecht (2000)

9. Demyanov, V.F., Roschina, V.A.: Optimality conditions in terms of upper and lower exhausters. Optimization **55**(5–6), 525–540 (2006)
10. Abbasov, M.E., Demyanov, V.F.: Proper and adjoint exhausters in nonsmooth analysis: optimality conditions. J. Glob. Optim. **56**(2), 569–585 (2013)
11. Demyanov, V.F., Malozemov, V.N.: Introduction to minimax. Halsted Press/Wiley, New York (1974)
12. Dolgopolik, M.V.: Inhomogeneous convex approximations of nonsmooth functions. Russ. Math. **56**(12), 28–42 (2012)
13. Abbasov, M.E., Demyanov, V.F.: Extremum conditions for a nonsmooth function in terms of exhausters and coexhausters. Proc. Steklov Inst. Math. **269**(1 Supplement), 6–15 (2010)
14. Demyanov, V.F.: Proper exhausters and coexhausters in nonsmooth analysis. Optimization **61**(11), 1347–1368 (2012)
15. Abbasov, M.E., Demyanov, V.F.: Adjoint coexhausters in nonsmooth analysis and extremality conditions. J. Optim. Theory Appl. **156**(3), 535–553 (2013)
16. Minoux, M.: Mathematical programming: theory and algorithms. Wiley, New York (1986)