Journal of Global Optimization **23:** 63–80, 2002.

A method of truncated codifferential with application to some problems of cluster analysis

V.F. DEMYANOV[1], A.M. BAGIROV[2] and A.M. RUBINOV[2]
[1]*Department of Applied Mathematics, Saint-Petersburg State University, Russia;* [2]*School of Information Technology and Mathematical Sciences, University of Ballarat, Australia (e-mail: amr@ballarat.edu.au)*

Abstract. A method of truncated codifferential descent for minimizing continuously codifferentiable functions is suggested. The convergence of the method is studied. Results of numerical experiments are presented. Application of the suggested method for the solution of some problems of cluster analysis are discussed. In numerical experiments Wisconsin Diagnostic Breast Cancer database was used.

Key words: subdifferential, quasidifferential, codifferential, truncated codifferential, cluster analysis.

1. Introduction

A mathematical formalization of one of the main problems of cluster analysis leads to the following global optimization problem: for a given set of points $a_i \in \mathbb{R}^n,\ i = 1, \dots, m$ find a collection $\bar{x} = (\bar{x}_1, \dots, \bar{x}_p)$ of p n-dimensional vectors, which is a solution of the following problem:

$$f(x_1, \dots, x_p) = \sum_{i=1}^{m} \min_{l=1,\dots,p} \|x_l - a_i\| \longrightarrow \min \quad \text{subject to}$$
$$x_l \in S \subset \mathbb{R}^n,\ l = 1, \dots, p, \tag{1}$$

where S is a compact convex set in $\mathbb{R}^n$. As a rule m is a large number and p is substantially less than m.

The objective function in (1) is a DC function (see, for example, Tuy, 1998), that is, f can be represented as the difference of two convex functions. In fact

$$f(x) = \sum_{i=1}^{m} \sum_{l=1}^{p} \|x_l - a_i\| - \sum_{i=1}^{m} \max_r \sum_{l \neq r} \|x_l - a_i\|. \tag{2}$$

Since f is a DC function it follows that f is continuously codifferentiable. (For definition and properties of codifferentiable functions see Demyanov and Rubinov (1995) and also Section 2 below.) There are many continuous codifferentials for the function f at a point x. Starting from the functions $x_l \mapsto \|x_l - a_i\|$ and using method of codifferential calculus (Demyanov and Rubinov, 1995), we can easily give an explicit formula for the calculation of one of them. This codifferential can be called the *complete codifferential*. The construction of the complete codifferential requires some operations with polytopes in the n-dimensional space, the number of these polytopes depends on m and p. When the number m is large enough the calculation of the complete codifferential is too complicated. We show that different kinds of codifferentials can be constructed for the function (2). These codifferentials appear as truncations of the complete codifferential and we shall call them *truncated codifferentials*. It can be shown that a truncated codifferential of the function f at a point x is continuous only in some neighbourhood of x, so the known methods of minimization using continuous codifferentials cannot be directly applied in the case under consideration.

The codifferential descent method for the search for a local minimum was introduced and studied in Demyanov and Rubinov (1990) (see, also Demyanov and Rubinov, 1995). The results of Bagirov (2000) show that the version of this method, based on a continuous codifferential, is efficient for the minimization of DC functions and it has advantage over methods based on the quasidifferential mapping (Demyanov et al., 1986, 1996; Hiriart-Urruty, 1989; Kiwiel, 1986).

It should be noted that proofs of the convergence of the majority of methods of nonsmooth optimization are based only on some kinds of upper semicontinuity of the approximate set-valued mappings which are used for the construction of corresponding methods (see, for example, Polak et al., 1983; Polak and Mayne, 1985). However, computational experience shows that as a rule numerical methods based on the Hausdorff continuous mappings work better than methods employing the mappings which are only upper semicontinuous. For instance, we can refer to various versions of the bundle method for convex functions, which exploit the ε-subdifferential mapping (see Hiriart-Urruty and Lemarechal, 1993a,b).

The continuity of the truncated codifferential in some neighbourhood of a current point is also useful for the application of numerical methods based on this codifferential. In this paper we study a modification of the codifferential descent method, by assuming only this kind of continuity. In particular, we suggest a truncated codifferential descent method. The convergence of the method is studied. A number of numerical tests have been carried out. The results of these experiments are presented. We give an example which demonstrates that the codifferential descent method sometimes allows 'to jump over' some local minima points.

We shall discuss a possible application of the proposed method to the solution of the problem (1). We need to find a global minimizer of this problem. Proposed truncated codifferential descent, which is based on a local approximation, allows one to find only inf-stationary points. However a combination of this method with some known methods of global optimization (in particular with cutting angle method (Andramonov et al., 1999; Bagirov and Rubinov, 2000)) essentially reduces the computational time and is very useful for solving problem (1). We do not discuss such type of combinations in this paper. We use the well-known Wisconsin Diagnostic Breast Cancer database for numerical experiments. These experiments show that even the proposed local method, without any global search, leads to very good results, which are very close to the results obtained by the global cutting angle method (Bagirov et al., 1999). However, the computational time for the local method is substantially less than for the global one.

The paper has the following structure. Section 2 provides some preliminaries. In Section 3 a method of truncated codifferential descent is described and its convergence is studied. Results of numerical experiments are presented in Section 4. In Section 5 we give results of application of the method for the solution of cluster analysis problems. Section 6 concludes the paper.

2. Quasidifferentiable and codifferentiable functions

Consider a locally Lipschitz function f defined on an open set $X \subset \mathbb{R}^n$. Assume that f is directionally differentiable at each point $x \in X$, that is there exists the finite limit

$$f'(x, g) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha g) - f(x)}{\alpha}. \tag{3}$$

The value $f'(x, g)$ is called the directional derivative of f at the point x in the direction g. The function $f : X \to \mathbb{R}$ is called quasidifferentiable at a point $x \in X$ if it is directionally differentiable at x and if its directional derivative is of the form

$$f'(x, g) = \max_{v \in \underline{\partial} f(x)} (v, g) + \min_{w \in \overline{\partial} f(x)} (w, g)$$

where $\underline{\partial} f(x)$ and $\overline{\partial} f(x) \subset \mathbb{R}^n$ are compact convex sets. The pair $\mathcal{D} f(x) = [\underline{\partial} f(x), \overline{\partial} f(x)]$ is called the quasidifferential of f at x (the set $\underline{\partial} f$ being called the subdifferential and the set $\overline{\partial} f$ the superdifferential).

The quasidifferential mapping $\mathcal{D} f$ is essentially discontinuous (in the Hausdorff metric) at a point of the nondifferentiability of f, therefore special precautions are to be taken to guarantee the convergence.

To overcome the mentioned drawback we shall employ the notion of codifferentiable function.

A function $f : X \to \mathbb{R}$ is called codifferentiable at a point $x \in X$ if there exist compact convex sets $\underline{d} f(x) \subset \mathbb{R}^{n+1}$ and $\overline{d} f(x) \subset \mathbb{R}^{n+1}$ such that the following expansion holds

$$\begin{aligned} f(x + \Delta) = {} & f(x) + \max_{[a,v] \in \underline{d} f(x)} [a + (v, \Delta)] \\ & + \min_{[b,w] \in \overline{d} f(x)} [b + (w, \Delta)] + o(x, \Delta) \end{aligned} \tag{4}$$

where $\frac{o(x,\Delta)}{\|\Delta\|} \longrightarrow 0$ as $\|\Delta\| \to 0$.

The pair $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ is called the codifferential of f at x, $\underline{d}f(x)$ being called the hypodifferential and $\overline{d}f(x)$ the hyperdifferential. The properties of codifferentiable functions are discussed, e.g., in Demyanov and Rubinov (1990, 1995). The codifferential mapping Df (as well as the quasidifferential one) is not uniquely defined. A function f is called continuously codifferentiable at a point $x \in X$ if it is codifferentiable in a neighbourhood of x and if there exists a codifferential mapping Df which is Hausdorff continuous at the point x.

The class of codifferentiable functions is quite rich and enjoys a full-scale calculus which is a generalization of the classical differential calculus.

Without loss of generality we may assume that

$$\max_{[a,v]\in \underline{d}f(x)} a = \min_{[b,w]\in \overline{d}f(x)} b = 0.$$

It is easy to see that the classes of codifferentiable and quasidifferentiable functions coincide. In particular, if $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ is a codifferential of f at x, then the pair

$$\mathcal{D}f(x) = [\underline{\partial}f(x), \overline{\partial}f(x)],$$

where

$$\underline{\partial}f(x) = \{v \in \mathbb{R}^n \mid [0, v] \in \underline{d}f(x)\}, \quad \overline{\partial}f(x) = \{w \in \mathbb{R}^n \mid [0, w] \in \overline{d}f(x)\},$$

is a quasidifferential of f at x.

Necessary optimality conditions can be formulated in terms of the codifferential:
for a point $x^* \in X$ to be a (local) minimizer of f on X it is necessary that

$$0_{n+1} \in [\overline{w} + \underline{d}f(x^*)] \ \ \forall \overline{w} = [0, w] \in \overline{d}f(x^*). \tag{5}$$

A point $x^* \in X$ satisfying (5) is called an *inf-stationary* point of f.

If a point $x^0 \in X$ is not inf-stationary, then it is possible to find descent directions at this point in terms of hypo- and hyperdifferentials.

Though the classes of codifferentiable and quasidifferentiable functions coinside, the usage of the codifferentiability instead of the quasidifferentiability makes it possible to single out the subclass of continuously codifferentiable functions (which doesn't coincide with the class of arbitrary codifferentiable functions). It turns on that most known nonsmooth functions are continuously codifferentiable.

In Demyanov and Rubinov (1990, 1995) the method of codifferential descent (CDD-method) was described for finding inf-stationary points of a continuously codifferentiable function. The efficiency of this method and its applicability to solving optimization problems heavily depends on the continuous codifferential mapping which is being employed. In the following section we describe a modification of the CDD-method where a locally continuous codifferential mapping is used to construct a descent direction at each step.

3. A method of truncated codifferential descent

Let a function f be defined, locally Lipschitz and continuously codifferentiable on an open set $X \subset \mathbb{R}^n$. We shall consider only functions f such that for every $x \in X$ there exists a mapping $D_x f = [\underline{d}_x f, \overline{d}_x f]$ with the following property:

$$f(y + \Delta) = f(y) + \max_{[a,v] \in \underline{d}_x f(y)} [a + (v, \Delta)] + \min_{[b,w] \in \overline{d}_x f(y)} [b + (w, \Delta)] + o_x(y, \Delta) \quad (6)$$

where $\underline{d}_x f(y)$, $\overline{d}_x f(y) \subset \mathbb{R}^{n+1}$ are convex compact sets, the mapping $D_x f$ is Hausdorff continuous in y on $S_\delta(x) = \{x' \in \mathbb{R}^n \mid ||x - x'|| < \delta\}$ (with $\delta > 0$ the same for all $x \in X$),

$$\frac{o_x(y, \Delta)}{\|\Delta\|} \longrightarrow 0 \text{ as } \|\Delta\| \to 0 \quad (7)$$

uniformly with respect to $y \in S_\delta(x)$ and $x \in X$.

First we consider a broad class of functions f such that the mapping $D_x f$ with the required properties does exist. Let

$$f_1(x) = \max_{i \in I} \varphi_i(x), \qquad f_2(x) = \min_{j \in J} \psi_j(x)$$

and

$$f(x) = f_1(x) + f_2(x) = \max_{i \in I} \varphi_i(x) + \min_{j \in J} \psi_j(x). \quad (8)$$

Here $\varphi_i (i \in I)$ and $\psi_j (j \in J)$ are continuously differentiable on some open set $X \subset \mathbb{R}^n$, I and J are finite index sets. Since

$$\begin{aligned} f(x + \Delta) &= f(x) + \max_{i \in I}(\varphi_i(x + \Delta) - f_1(x)) + \min_{j \in J}(\psi_j(x + \Delta) - f_2(x)) \\ &= f(x) + \max_{i \in I}[\varphi_i(x) - f_1(x) + (\varphi_i'(x), \Delta)] \\ &\quad + \min_{j \in J}[\psi_j(x) - f_2(x) + (\psi_j'(x), \Delta)] + o(x, \Delta) \end{aligned}$$

then the mapping

$$Df(x) = [\underline{d} f(x), \overline{d} f(x)] \quad (9)$$

with

$$\underline{d} f(x) = \text{co}\{[a_i, v_i] \mid a_i = \varphi_i(x) - f_1(x),\ v_i = \varphi_i'(x),\ i \in I\}, \quad (10)$$

$$\overline{d} f(x) = \text{co}\{[b_j, w_j] \mid b_j = \psi_j(x) - f_2(x),\ w_j = \psi_j'(x),\ j \in J\} \quad (11)$$

is a codifferential mapping for the function f defined by (8). This mapping is Hausdorff continuous on X. We shall call the mapping Df the *complete codifferential mapping of* f.

Fix $\epsilon > 0$, $\mu > 0$ and put

$$R_\epsilon(x) = \{i \in I \mid \exists y \in S_\delta(x) : \varphi_i(y) \geqslant f_1(y) - \epsilon\},$$

$$Q_\mu(x) = \{j \in J \mid \exists y \in S_\delta(x) : \varphi_j(y) \leqslant f_2(y) + \mu\}.$$

The mapping

$$D_x f(y) = [\underline{d}_x f(y), \overline{d}_x f(y)] \tag{12}$$

with

$$\underline{d}_x f(y) = \mathrm{co}\{[a_i, v_i] \mid a_i = \varphi_i(y) - f_1(y),\ v_i = \varphi_i'(y),\ i \in R_\epsilon(x)\}, \tag{13}$$

$$\overline{d}_x f(y) = \mathrm{co}\{[b_j, w_j] \mid b_j = \psi_j(y) - f_2(y),\ w_j = \psi_j'(y),\ j \in Q_\mu(x)\} \tag{14}$$

is a codifferential mapping for f.

PROPOSITION 3.1. *The mapping $D_x f$ defined by (12) - (14) is Hausdorff continuous in y in a neighbourhood of the point x and possesses the property (6).*

Proof. The Hausdorff continuity of the mapping $D_x f$ follows immediately from its definition and continuously differentiability of the functions $\varphi_i (i \in I)$ and $\psi_j (j \in J)$. Let us prove that this mapping possesses the property (6). Let $y, y + \Delta \in S_\delta(x)$. Then we have

$$\begin{aligned} f(y+\Delta) =& f(y) + \max_{i \in I}(\varphi_i(y+\Delta) - f_1(y)) + \min_{j \in J}(\psi_j(y+\Delta) - f_2(y)) \\ =& f(y) + \max_{i \in R_\varepsilon(x)} (\varphi_i(y+\Delta) - f_1(y)) \\ & + \min_{j \in Q_\mu(x)} (\psi_j(y+\Delta) - f_2(y)) \\ =& f(y) + \max_{i \in R_\varepsilon(x)} [\varphi_i(y) - f_1(y) + (\varphi_i'(y), \Delta)] \\ & + \min_{j \in Q_\mu(x)} [\psi_j(y) - f_2(y) + (\psi_j'(y), \Delta)] + o_x(y, \Delta) \\ =& f(y) + \max_{[a,v] \in \underline{d}_x f(y)} [a + (v, \Delta)] \\ & + \min_{[b,w] \in \overline{d}_x f(y)} [b + (w, \Delta)] + o_x(y, \Delta). \end{aligned}$$

□

We shall call the mapping $D_x f(y)$ defined by (12)–(14) a truncated codifferential mapping or an (ϵ, μ)-truncated codifferential mapping.

REMARK 3.1. It can be shown that the function

$$f(x) = \max_{i \in I} \min_{j \in J_i} f_{ij}(x), \tag{15}$$

where I and $J_i(i \in I)$ are finite index sets, and functions $f_{ij}(i \in I, j \in J_i)$ are continuously differentiable on some open set $X \subset \mathbb{R}^n$, can also be presented in the form (8) (with different, of course, I and J) and therefore for this function we can construct a continuous codifferential mapping and a truncated codifferential mapping which is locally Hausdorff continuous.

REMARK 3.2. A truncated codifferential can also be considered for the function (2). For its computation we can use the same technique as for (8). Computational efforts depend on the chosen norm. For the 2-norm the truncated codifferential is simpler than for 1-norm. The computation of truncated codifferential for (2) with the 1-norm is possible only for small p.

Consider a locally Lipschitz continuously codifferentiable function f defined on an open set $X \subset \mathbb{R}^n$. Assume that for each $x \in X$ there exists a mapping $D_x f = [\underline{d}_x f, \overline{d}_x f]$, which possesses the property (6). The pair $\mathcal{D}_x f(y) = [\underline{\partial}_x f(y), \overline{\partial}_x f(y)]$ where

$$\underline{\partial}_x f(y) = \{v \mid [0, v] \in \underline{d}_x f(y)\}, \ \overline{\partial}_x f(y) = \{w \mid [0, w] \in \overline{d}_x f(y)\}$$

is a quasidifferential of f at y. We shall assume that

$$\mathcal{D}_x f(y) = \mathcal{D} f(y) \ \ \forall y \in S_\delta(x). \tag{16}$$

We also assume that

$$\max_{[a,v] \in \underline{d}_x f(y)} a = \min_{[b,w] \in \overline{d}_x f(y)} b = 0, \ \forall x \in X, \ \forall y \in S_\delta(x). \tag{17}$$

REMARK 3.3. The assumptions (16) and (17) hold for a broad class of functions. For example, it follows from the definition of the truncated codifferential that they hold for the functions defined by (8).

If $x^* \in X$ is a minimizer of f then the necessary condition for a minimum (5) takes the form

$$0_{n+1} \in \{\underline{d}_x f(x^*) + [0, w]\} \ \forall [0, w] \in \overline{d}_x f(x^*), \ \forall x \in S_\delta(x^*). \tag{18}$$

Assume that condition (18) does not hold at a point $x \in X$. Then there exists a $\overline{w}_x = [0, w_x] \in \overline{d}_x f(x)$ such that

$$0_{n+1} \notin \{\underline{d}_x f(x) + \overline{w}_x\} \equiv L_{\overline{w}_x}(x). \tag{19}$$

Find

$$\min_{\overline{z} \in L_{\overline{w}_x}(x)} \|\overline{z}\| = \|\overline{z}_{\overline{w}_x}(x)\|.$$

It follows from (19) that

$$\overline{z}_{\overline{w}_x}(x) = [\eta_{\overline{w}_x}(x), z_{\overline{w}_x}(x)] \neq 0_{n+1}.$$

As in proof of Theorem V.4.1 Demyanov and Rubinov (1995) we conclude that $z_{\overline{w}_x}(x) \neq 0_n$ and that for the direction

$$g_{\overline{w}_x}(x) = -z_{\overline{w}_x}(x)/\|z_{\overline{w}_x}(x)\|$$

the following inequality holds:

$$f'(x, g_{\overline{w}_x}(x)) \leqslant -\|z_{\overline{w}_x}(x)\|.$$

Now let us describe the following algorithm:

Fix any $\mu > 0$. Choose an arbitrary $x_0 \in \mathbb{R}^n$. Let x_k have already been found. If condition (18) holds at the point x_k, i.e.

$$0_{n+1} \in \{\underline{d}_{x_k} f(x_k) + [0, w]\} \;\; \forall [0, w] \in \overline{d}_{x_k} f(x_k), \tag{20}$$

then x_k is an inf-stationary point and the process terminates. Otherwise, for every $\overline{w} \in \overline{d}_{\mu x_k} f(x_k)$ where

$$\overline{d}_{\mu x} f(x) = \{\overline{w} \in \overline{d}_x f(x) \mid \overline{w} = (\omega, w), 0 \leqslant \omega \leqslant \mu\} \tag{21}$$

we find

$$\min_{\overline{z} \in L_{\overline{w}}(x_k)} \|\overline{z}\| = \|\overline{z}_{k\overline{w}}\|, \tag{22}$$

where

$$\overline{z}_{k\overline{w}} = [\eta_{k\overline{w}}, z_{k\overline{w}}], \;\; L_{\overline{w}}(x_k) = \underline{d}_{x_k} f(x_k) + \overline{w}.$$

It follows from (6) that

$$f(x_k - \alpha z_{k\overline{w}}) \leqslant f(x_k) + \max_{[a,v] \in \underline{d}_{x_k} f(x_k)} [(a + \omega) - \alpha(v + w, z_{k\overline{w}})] + o(\alpha). \tag{23}$$

We have from (22)

$$(\overline{z}, -\overline{z}_{k\overline{w}}) \leqslant -\|\overline{z}_{k\overline{w}}\|^2 \;\; \forall \overline{z} \in L_{\overline{w}}(x_k), \tag{24}$$

i.e.

$$(a + \omega)(-\eta_{k\overline{w}}) - (v + w, z_{k\overline{w}}) \leqslant -\|\overline{z}_{k\overline{w}}\|^2 \tag{25}$$

or

$$-(v + w, z_{k\overline{w}}) \leqslant -\|\overline{z}_{k\overline{w}}\|^2 + (a + \omega)\eta_{k\overline{w}} \;\; \forall [a, v] \in \underline{d}_{x_k} f(x_k).$$

The inequalities (23) and (25) imply

$$\begin{aligned} f(x_k - \alpha z_{k\overline{w}}) &\leqslant f(x_k) + \max_{[a,v]\in \underline{d}_{x_k} f(x_k)} [(a+\omega) - \alpha\|\overline{z}_{k\overline{w}}\|^2 \\ &\quad + \alpha(a+\omega)\eta_{k\overline{w}}] + o_k(\alpha) \qquad (26)\\ &= f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + \max_{[a,v]\in \underline{d}_{x_k} f(x_k)} [(a+\omega)(1+\alpha\eta_{k\overline{w}})] + o_k(\alpha). \end{aligned}$$

For sufficiently small $\alpha \geqslant 0$ we have $(1 + \alpha\eta_{k\overline{w}}) > 0$, thus (26) yields

$$\begin{aligned} f(x_k - \alpha z_{k\overline{w}}) &\leqslant f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + (1+\alpha\eta_{k\overline{w}}) \max_{[a,v]\in \underline{d}_{x_k} f(x_k)} (a+\omega) + o_k(\alpha) \\ &= f(x_k) - \alpha\|\overline{z}_{k\overline{w}}\|^2 + (1+\alpha\eta_{k\overline{w}})\omega + o_k(\alpha). \qquad (27) \end{aligned}$$

Note that here $\omega \in [0, \mu]$, therefore it may happen that the direction $-z_{k\overline{w}}$ is not a descent direction (even if $\|z_{k\overline{w}}\| > 0$). However, since condition (18) does not hold at the point x_k, then there exists at least one $w_0 \in \overline{\partial}_{x_k} f(x_k)$ such that $\|\overline{z}_{k\overline{w}_0}\| > 0$, hence (in this case $\omega_0 = 0$) the direction $-z_{k\overline{w}_0}$ is (see (27)) a descent direction.

Now, for every $\overline{w} \in \overline{d}_{\mu x_k} f(x_k)$ we find

$$\min_{\alpha>0} f(x_k - \alpha z_{k\overline{w}}) = f(x_k - \alpha_{k\overline{w}} z_{k\overline{w}}) \qquad (28)$$

and then

$$\min_{\overline{w}\in \overline{d}_{\mu x_k} f(x_k)} f(x_k - \alpha_{k\overline{w}} z_{k\overline{w}}) = f(x_k - \alpha_{k\overline{w}_k} z_{k\overline{w}_k}). \qquad (29)$$

Put $x_{k+1} = x_k - \alpha_{k\overline{w}_k} z_{k\overline{w}_k}$. Continuing in the same manner we construct a sequence $\{x_k\}$ such that

$$f(x_{k+1}) < f(x_k). \qquad (30)$$

THEOREM 3.1. *Let the set*

$$P \equiv \{x \in X \mid f(x) \leqslant f(x_0)\} \qquad (31)$$

be bounded and x^* *be a limit point of the sequence* $\{x_k\}$. *Then the point* x^* *is an inf-stationary point of the function* f *on* X *(i.e. condition (18) holds).*

Proof. The existence of limit points of the sequence $\{x_k\}$ follows from the compactness of the set (31) and from (30). Let $x_{k_s} \to x^*$. Assume that condition (18) does not hold at x^*, i.e. x^* is not a stationary point of f. At the point x^* we have the family of codifferentials $\{D_x f(x^*) \mid x \in S_\delta(x^*)\}$, however, all quasidifferentials

$$\mathcal{D}_x f(x^*) = [\underline{\partial}_x f(x^*), \overline{\partial}_x f(x^*)]$$

where

$$\underline{\partial}_x f(x^*) = \{v \in \mathbb{R}^n \mid [0, v] \in \underline{d}_x f(x^*)\},$$

$$\overline{\partial}_x f(x^*) = \{w \in \mathbb{R}^n \mid [0, w] \in \overline{d}_x f(x^*)\}$$

are equivalent (see Demyanov and Rubinov (1999, 1995)). Since x^* is not a stationary point then there exists a $c > 0$ such that

$$\max_{w\in\overline{\partial}_x f(x^*)} \min_{v\in\underline{\partial}_x f(x^*)} \|v + w\| = \|v^* + w^*\| = c > 0 \ \ \forall x \in S_\delta(x^*). \tag{32}$$

Taking into account the uniform (with respect to $x \in S_\delta(x^*)$) continuity of $D_x f(y)$ we can claim that there exists a neighbourhood of x^* (without loss of generality we again denote this neighbourhood by $S_\delta(x^*)$) and a vector-function $\overline{w}_x$ defined on $S_\delta(x^*)$ such that

$$\min_{\overline{z}\in L(x)} \|\overline{z}\| = \|\overline{z}_{\overline{w}_x}(x)\| \geqslant c/2 \ \ \forall x \in S_\delta(x^*)$$

where

$$L(x) = \overline{w}_x + \underline{d}_x f(x), \quad \overline{w}_x = [\omega_x, w_x] \in \overline{d}_x f(x),$$

$$\omega_x \to 0,\ w_x \to w^* \ \text{ as } \ x \to x^*.$$

Due to the continuity of the codifferential mapping there must exist a sequence $\{\overline{w}_{k_s}\}$ such that

$$\overline{w}_{k_s} = [\omega_{k_s}, w_{k_s}] \in \overline{d}_{\mu x_{k_s}} f(x_{k_s}), \quad \omega_{k_s} \to 0,\ \overline{w}_{k_s} \to \overline{w} \ \text{ as } \ k_s \to \infty.$$

We have (see (22)) $\overline{z}_{k_s} = [\eta_{k_s}, z_{k_s}]$, and we may assume $\|\overline{z}_{k_s}\| \geqslant c/4 > 0$.

It follows from the properties of the function $o(x, \Delta)$ (see (7)) and (27) that there exists an $\alpha_0 > 0$ such that for sufficiently large k_s

$$f(x_{k_s} - \alpha_0 z_{k_s}) \leqslant f(x_{k_s}) - \frac{\alpha_0 c}{8} + 2\omega_{k_s}.$$

Since $\omega_{k_s} \to 0$, then, for sufficiently large k_s,

$$f(x_{k_s} - \alpha_0 z_{k_s}) \leqslant f(x_{k_s}) - \frac{\alpha_0 c}{16}.$$

Moreover,

$$f(x_{k_s+1}) \leqslant f(x_{k_s}) - \frac{\alpha_0 c}{16}.$$

Therefore from (30) we have $f(x_k) \to -\infty$, which contradicts the boundedness of the continuous function f on the bounded closed set P (see (31)). □

REMARK 3.4. It is assumed that the one-dimensional minimization problem in (28) is solved exactly. To find the exact solution of (28) is a very complicated problem in many instances. Therefore below we use the well-known Armijo rule (see Armijo, 1966) for the estimation of the stepsize in numerical experiments. Let $c,\ \sigma \in (0, 1)$ be a given numbers. The stepsize $\alpha_{k\overline{w}}$ is defined as follows:

$$\alpha_{k\overline{w}} = \text{argmax}\{\sigma^i \mid f(x_k - \sigma^i z_{k\overline{w}}) - f(x_k) \leqslant -c\sigma^i \|z_{k\overline{w}}\|^2\}. \tag{33}$$

REMARK 3.5. The problems (22) and (29) are the most time-consuming parts of the suggested method. For the solution of the problem (22) one can use Wolfe's terminating algorithm (see Wolfe, 1976). The problem (29) is solved effectively when the set $d_\mu f(x)$ is a polyhedron described by its vertices. In this case it is sufficiently to solve this problem for vertices of the polyhedron.

REMARK 3.6. It is clear from (27) that the direction $x_{k+1} - x_k$ may happen not to be a descent direction (in this direction the function may first increase and then decrease, i.e. the algorithm allows to 'jump over' some points of local minima).

4. Numerical experiments

In order to verify the practical efficiency of the proposed algorithm a number of numerical experiments have been carried out. In this section we describe results of these experiments. We consider only unconstrained problems with DC objective functions.

The following notation will be used for the description of the test problems:

- $f = f(x)$ is the objective function,
- n is the number of variables,
- x_0 is a starting point,
- x^* is a local minimizer, $f_* = f(x^*)$.

PROBLEM 4.1.

$$f(x) = \max\left\{\sum_{i=1}^{n} a_{ij}(b_i x_i - 1)^2 : j = 1, ...m\right\} + \min\left\{\sum_{i=1}^{n} a_{ij}(b_i x_i - 1)^2 : j = 1, ...m\right\},$$

$$x \in \mathbb{R}^n,\ a_{ij} = 1/(i + j - 1),\ i = 1, ...n,\ \ j = 1, ..., m,\ b_i = i,\ i = 1, ..., n,$$

$$x_0 = (5, ..., 5),\ x^* = (1, 1/2, ..., 1/n) \in \mathbb{R}^n,\ f_* = 0.$$

PROBLEM 4.2.

$$f(x) = \max\left\{\exp\left(\sum_{j=1}^{n} a_{ij} x_j (x_j + 1)\right) : i = 1, ..., 60\right\} + \min\left\{\exp\left(\sum_{j=1}^{n} b_{ij} x_j (x_j + 1)\right) : i = 1, ..., 60\right\},$$

$$x \in \mathbb{R}^n,\ a_{ij} = 1/2(i + j - 1),\ i = 1, ..., 30,\ j = 1, ..., n,$$
$$a_{ij} = -1/2(i + j - 1),$$

$$i = 31, ..., 60,\ j = 1, ..., n,\ b_{ij} = a_{ij}/2,\ i = 1, ..., 60,\ j = 1, ..., n,$$

$$x_0 = (1, \dots , 1),\ x^* = (0, ..., 0),\ f_* = 2.$$

PROBLEM 4.3.

$$f(x) = n \max\{|x_i| :\ i = 1, ..., n\} - \sum_{i=1}^{n} |x_i|,$$

$$x \in \mathbb{R}^n,\ x_0 = (i,\ i = 1, ...,]n/2[,\ \ -i,\ i =]n/2[+1, ..., n),\ x^* = (\alpha, ..., \alpha),$$

$$\alpha \in \mathbb{R}^1,\ f_* = 0.$$

PROBLEM 4.4.

$$f(x) = \sum_{j=1}^{100} \left| \sum_{i=1}^{n} (x_i - x_i^*) t_j^{i-1} \right|$$
$$- \max \left\{ \left| \sum_{i=1}^{n} (x_i - x_i^*) t_j^{i-1} \right| :\ j = 1, ..., 100 \right\},$$

$$x \in \mathbb{R}^n,\ t_j = 0.01j,\ j = 1, ..., 100,\ x_0 = (0, ..., 0),$$

$$x^* = (1/n, ..., 1/n),\ f_* = 0.$$

It should be noted that in Problem 4.1 $\underline{\partial} f(x^*) = \overline{\partial} f(x^*) = \{0\}$, whereas in Problems 4.2, 4.3 and 4.4 $\underline{\partial} f(x^*) \neq \{0\}, \overline{\partial} f(x^*) \neq \{0\}$.

In Problems 4.1 and 4.2 we compute the complete codifferentials, whereas in Problems 4.3 and 4.4 – only truncated codifferentials. In latter case for Problem 4.3 we took $\epsilon = 0.1$ and for Problem 4.4 $\epsilon = 0.001 \div 0.0001$. For all problems μ is chosen as $\mu = 0.0001$. For the compution of the stepsize we use the Armijo rule (33). We took the following values of the parameters c and σ: $c = 0.01,\ \sigma = 0.6$. It should be noted that for all problems and n we use the same values of μ, c and σ. Taking them different for different problems and n we can get more better results. For the solution of the problem (22) we use Wolfe's method (Wolfe, 1976). We solved all problems with the precision 10^{-4}, that is we computed a point $\overline{x}$ for which

$$f(\overline{x}) - f_* \leqslant 10^{-4}.$$

Numerical experiments have been carried out on an IBM Pentium-S with CPU 150 MHz. Their results are given in Table 1. For the description of these results we use the following notations:

Table 1.

N	n	m_1	m_2	t	N	n	m_1	m_2	t
1	5	78	87	0.17	3	5	64	339	0.05
1	10	138	293	0.49	3	10	94	557	0.28
1	20	239	856	1.60	3	15	163	1065	1.19
1	30	484	2200	5.11	3	20	162	1004	3.57
1	40	884	6725	15.11	3	30	198	556	18.23
1	50	2207	13445	41.47	3	40	255	619	39.33
1	70	1785	11570	47.35	3	50	283	811	104.47
2	5	31	50	0.11	4	5	188	548	1.65
2	10	70	84	0.33	4	10	126	607	6.48
2	20	122	125	1.04	4	15	191	784	19.56
2	30	59	88	0.83	4	20	460	2969	34.49
2	40	140	179	2.80	4	30	498	3125	109.36
2	50	66	115	1.43	4	40	464	2836	156.42
2	70	86	117	2.59	4	50	751	6709	281.66

- N is the number of problem,
- n is the number of variables,
- m_1 is number of iterations,
- m_2 is number of function evaluations,
- t is the computation time (in s).

The number of computation of complete or truncated codifferential is the same as the number of iterations m_1 so we do not give it.

The following example demonstrates that the algorithm allows to 'jump over' some points of local minima. Let

$$\begin{aligned} f(x) = &\min\{\max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5\},\\ &\max\{-2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\}\}. \end{aligned}$$

This function can be represented as the difference of two convex functions:

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

where

$$\begin{aligned} f_1(x) = &\max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5\}\\ &+ \max\{-2x_1 - x_2 + 21, 6x_1 + 3x_2 - 15\},\\ f_2(x) = &\max\{-x_1 - 2x_2 + 4, 2x_1 + 4x_2 - 5, -2x_1\\ &- x_2 + 21, 6x_1 + 3x_2 - 15\}. \end{aligned}$$

Table 2.

k	x_1	x_2	$f(x)$	μ
0	3.100	3.100	12.900	
1	2.976	3.038	12.010	0.01
2	2.976	3.038	12.010	0.01
3	2.976	3.038	12.010	0.01
4	2.976	3.038	12.010	0.01
5	2.976	3.038	12.010	0.02
6	2.976	3.038	12.010	0.04
7	2.976	3.038	12.010	0.08
8	2.976	3.038	12.010	0.16
9	2.976	3.038	12.009	0.32
10	2.976	3.038	12.009	0.64
11	1.104	−0.707	4.310	1.28
12	1.907	0.900	2.414	0.01
15	1.734	0.553	1.160	0.01
20	1.750	0.586	1.077	0.01
89	1.764	0.613	1.010	0.01

The function f has two sets of minimum points:

$$X_1^* = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid x_1 + 2x_2 = 3\},$$
$$X_2^* = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 + x_2 = 9\}.$$

$f(x) = 1$ for all $x \in X_1^*$ and $f(x) = 12$ for all $x \in X_2^*$. Thus the set X_1^* is the set of global minimizers of the function f over $\mathbb{R}^2$.

In Table 2 we give the results of numerical experiments where the value $\mu = 1.28$ allowed to jump over a local minimum point.

5. Clustering via codifferentiability

One of the basic problems of data mining is the following clustering problem: for a given set of points $a^i,\ i = 1, \dots, m$ in $\mathbb{R}^n$, find cluster centers $x^l,\ l = 1, \dots, p$ in $\mathbb{R}^n$ such that the sum of the minima over $l \in \{1, \dots, p\}$ of the distance between each point a^i and the cluster centers $x^l,\ l = 1, \dots, p$ is minimized. The problem we are concerned with is

$$f(x) = \sum_{i=1}^{m} \min_{l=1,\dots,p} \|x^l - a^i\| \to \min \text{ s.t. } x^l \in S \subset \mathbb{R}^n,\ l = 1, \dots, p \qquad (34)$$

where S is a compact convex set in $\mathbb{R}^n$.

The objective function in (34) can be represented as the difference of two convex functions:

$$f(x) = \sum_{i=1}^{m} \sum_{l=1}^{p} \|x^l - a^i\| - \sum_{i=1}^{m} \max_r \sum_{l \neq r} \|x^l - a^i\|.$$

In the case $p = 2$ the objective function is the simplest:

$$f(x) = \sum_{i=1}^{m} (\|x^1 - a^i\| + \|x^2 - a^i\|) - \sum_{i=1}^{m} \max\{\|x^1 - a^i\|, \|x^2 - a^i\|\}.$$

We conclude that the objective function f in (34) is codifferentiable and in the latter case we can construct its truncated codifferential. Thus for solving clustering analysis problems we can use the method of truncated codifferential descent described above.

It has been applied to analyse the Wisconsin Diagnostic Breast Cancer database. This database consists of 569 vectors with known outcomes and it represents a training set with which a classifier can be constructed to diagnose future examples. It was created by W.H. Wolberg, General Surgery Dept., University of Wisconsin, Clinical Sciences Center, W. N. Street and O.L. Mangasarian, Computer Sciences Dept., University of Wisconsin.

Ten real-valued features are computed for each cell nucleus: radius (mean of distances from center to points on the perimeter), texture (standard deviation of gray-scale values), perimeter, area, smoothness (local variation in radius lengths), compactness, concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry, fractal dimension (of the boundary). The mean value, standard error and extreme value (i.e., largest or worst value: biggest size, most irregular shape) of each of these cellular features are computed for each image, resulting in a total of 30 real-valued features. Thus $n = 30$ in the case under consideration. The papers (Wolberg et al., 1994, 1995a–c) contain more detailed description of this database. Some approaches based on linear and bilinear programming techniques are considered in papers (Mangasarian, 1997; Mangasarian et al., 1995).

The first set consists of vectors related to benign cases and contains 357 vectors; the second set consists of vectors related to malignant cases and contains 212 vectors.

By applying the algorithm and using only first 10 parameters (mean values) we computed two clusters for both sets of vectors. We took 80% of all vectors as a traing set and remaining 20% of vectors were used for testing the obtained clusters. The numerical experiments show that the algorithm found two clusters for each set which describe all vectors with 96.5% accuracy (97.2% for benigns and with 92.5% for malignants accuracy.)

In the numerical experiments we considered the 1-norm, the 2-norm and the max-norm. Best result we obtained using the 1-norm. Results with the 2-norm were close enough to the results obtained by means of the 1-norm. Clusters obtained by means of the max-norm did not give a good description of both sets.

REMARK 5.1. If centres of clusters are known, we can try to find clusters themselves, or, at least, a part of each cluster and then to compare distances to clusters (parts of clusters) instead of distances to centers of clusters. Thus, the following two step approach can be considered. First, to find clusters (part of clusters), using known centres of clusters. Second, to define a sort of the point by comparison of distances from this point to clusters (part of clusters). Of course this is only a heuristic idea. Indeed, since our method is based on calculation of distances to points, we should determine a sort of a point by comparison distances from this point to the centres of clusters. However, in some instances this idea demonstrates good results. For example, for the database under consideration usage this idea allows one to improve the description of the sets 1–2%.

6. Conclusions

In this paper we have studied a numerical algorithm for solving unconstrained problems of nonsmooth optimization with quasidifferentiable objective functions based on the truncated codifferential mapping. Some numerical experiments have been carried out using this method. In these experiments test problems with the objective functions represented as the difference of two convex functions were used. The results of numerical experiments show the effectiveness of the suggested method. The presented example shows that this method sometimes allows to jump over local minimum points and to find a global one. At the same time we cannot assert that it allows always to find a global solution of the optimization problem under consideration.

This method has been applied to the solution of one cluster analysis problem. The clusters obtained describe quite well the sets under consideration and we can conclude that this method solves such a problem effectively if the number of clusters is not large.

Acknowledgements

We would like to thank the anonymous referee for valuable comments. This research has been supported by the Australian Research Council under Grant No. A49906152. This work has been completed when V.F. Demyanov visited School of Information Technology and Mathematical Sciences, University of Ballarat, Australia. V.F. Demyanov was also partially supported by the Russian Foundation for Fundamental Studies under the grant RFFI No. 97-01-00499.

References

Andramonov, M. Yu., Rubinov, A.M. and Glover, B. M. (1999), Cutting angle methods in global optimization, *Applied Mathematics Letters* 12: 95–100.

Armijo, L. (1966), Minimization of functions having continuous partial derivatives, *Pacific J. Math.* 16: 1–13.

Bagirov, A.M. (2000), Numerical methods for minimizing quasidifferentiable functions: a survey and comparison, In: *Quasidifferentiability and Related Topics,* Demyanov, V.F. and Rubinov, A.M. (eds.), Ser.: Nonconvex Optimization and Its Applications, Vol. 43, 33–71, Kluwer Academic Publishers, Dordrecht.

Bagirov, A.M. and Rubinov, A.M. (2000), Global minimization of increasing positively homogeneous functions over the unit simplex, *Annals of Operations Research* 98: 171–187.

Bagirov, A.M., Rubinov, A. M., Stranieri, A. and Yearwood, J. (1999), The global optimization approach to the clustering analysis, Research Report 45/99, University of Ballarat, Australia.

Demyanov, V.F., Gamidov, S. and Sivelina, T.I. (1986), An algorithm for minimizing a certain class of quasidifferentiable functions, *Math. Program. Study* 29: 74–85.

Demyanov, V.F. and Rubinov, A.M. (1990), *Foundations of Nonsmooth Analysis and Quasidifferential Calculus,* Nauka, Moscow (in Russian)

Demyanov, V.F. and Rubinov, A.M. (1995), *Constructive Nonsmooth Analysis,* Springer, Frankfurt am Main.

Demyanov, V.F., Stavroulakis, G.E., Polyakova, L.N. and Panagiotopoulos, P.D. (1996), *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics,* Kluwer Academic Publishers, Dordrecht.

Hiriart-Urruty, J.-B. (1989), From convex optimization to nonconvex optimization. Necessary and sufficient conditions for global optimality, In: Clarke F.H., Demyanov V.F. and Gianessi F. (eds.), *Nonsmooth Optimization and Related Topics,* Plenum Publ., New York, 219-239.

Hiriart-Urruty, J.B. and Lemarechal, C. (1993a), *Convex Analysis and Minimization Algorithms*, Springer, Heidelberg, Vol. 1.

Hiriart-Urruty, J.B. and Lemarechal, C. (1993b), *Convex Analysis and Minimization Algorithms,* Springer, New York, Vol. 2.

Kiwiel, K.C. (1986), A linearization method for minimizing certain quasidifferentiable functions, *Math. Program. Study* 29: 86–94.

Mangasarian, O.L. (1997), Mathematical programming in data mining, *Data Mining and Knowledge Discovery* 1: 183–201.

Mangasarian, O.L., Street, W.N. and Wolberg, W.H. (1995), Breast cancer diagnosis and prognosis via linear programming, *Operations Research* 43(4): 570–577.

Polak, E., Mayne, D.Q. and Wardi, Y. (1983), On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems, *SIAM J. Control and Optimization* 21(2): 179–203.

Polak, E. and Mayne, D.Q. (1985), Algorithm models for nondifferentiable optimization, *SIAM J. Control and Optimization* 23(3): 477–491.

Tuy, H., (1998), *Convex Analysis and Global Optimization,* Kluwer Academic Publishers, Dordrecht.

Wolberg, W.H., Street, W.N. and Mangasarian, O.L. (1994), Machine learning techniques to diagnose breast cancer from fine-needle aspirates, *Cancer Letters* 77: 163–171.

Wolberg, W.H., Street, W.N. and Mangasarian, O.L. (1995a), Image analysis and machine learning applied to breast cancer diagnosis and prognosis, *Analytical and Quantitative Cytology and Histology* 17(2): 77–87.

Wolberg, W.H., Street, W.N., Heisey, D.M. and Mangasarian, O.L. (1995b), Computerized breast cancer diagnosis and prognosis from fine needle aspirates, *Archives of Surgery* 130: 511–516.

Wolberg, W.H., Street, W.N., Heisey, D.M. and Mangasarian, O.L. (1995c), Computer-derived nuclear features distinguish malignant from benign breast cytology, *Human Pathology* 26: 792–796.

Wolfe, P.H. (1976), Finding the nearest point in a polytope. *Math. Programm.* 11(2): 128–149.