J Glob Optim (2010) 48:29–40
DOI 10.1007/s10898-009-9509-9

One-dimensional identification problem and ranking parameters

V. V. Demyanova · V. F. Demyanov

Received: 11 December 2009 / Accepted: 14 December 2009 / Published online: 29 December 2009

Abstract The problem of identifying points of two sets in $\mathbb{R}^n$ is considered. This problem is of interest by itself and has numerous practical applications. One of such applications—namely, to the ranking of parameters—is also mentioned in the paper. Two identification procedures are discussed: by means of a linear identifier (the separation method) and by means of isolation of one of the sets by an interval.

Keywords Mathematical diagnostics · One-dimensional identification problem · Ranking parameters · Separation method · Isolation method

1 Introduction

Problems of separating two sets of points in $\mathbb{R}$ arise, for example, in the theory of identification and mathematical diagnostics. In the present paper we treat the problem of identifying points of two finite sets located on the real line. Let $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ be given finite sets of points.

In Sect. 2, the identification problem is solved by means of a linear identifier. The points belonging to one semi-line are attributed to one set (say, A), while the points from the other semi-line—to the set B. By such an identification rule, some points can be misclassified in each of the two sets. It is required to find a linear identifier such that the number of misclassified points be minimal (in some proper sense). As a functional, the maximum of the number of misclassified points in the both of the sets is chosen. This functional is discontinuous and integer-valued. However, using the quasiconvexity property of this functional, it is possible to derive necessary and sufficient optimality conditions and to construct a numerical

The work was supported by the Russian Foundation for Fundamental Studies (RFFI) under Grant No. 06-01-00276.

V. V. Demyanova · V. F. Demyanov (✉)
Applied Mathematics Department, St.-Petersburg State University, Staryi Peterhof,
St.-Petersburg 198504, Russia
e-mail: vfd@ad9503.spb.edu

method. The mentioned conditions were obtained in [3]. Here their geometric interpretation is presented.

In Sect. 3 the identification is performed by isolating one of the sets by means of an interval. The points belonging to this interval are attributed to one set (set A), while the points not belonging to the interval are attributed to the set B. It is required to find an interval such that the number of misclassified points be minimal. As a functional, the maximum of the number of misclassified points in the both of the sets is chosen again. This functional (as a function of the ends of the interval, that is, a function of two variables) is discontinuous and takes integer values. In every of its coordinates this functional is quasiconvex, therefore, it is possible to use the coordinate-wise descent method (applying the algorithm described in Sect. 2 with respect to every coordinate for a fixed other coordinate). Necessary and sufficient optimality conditions are stated.

A ranking parameters method based on the one-dimensional identification is discussed in Sect. 3.3.

As a criterion functional, the maximum of the number of misclassified (in the sets A B) points is chosen. It is the so-called *natural functional*. This functional is discontinuous. In the one-dimensional case we have managed to overcome the difficulties caused by the discontinuity and to solve the stated identification problem. Another approach consists in replacing the natural functional by some other functional (called *surrogate*) possessing proper analytical properties allowing to employ the existing optimization methods. The relationship between natural and surrogate functionals is discussed in Sect. 4.

2 One-dimensional identification problem: identification by means of separation

2.1 Statement of the problem

Let sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}: \ A = \{a_i \in \mathbb{R} \mid i \in I\},\ B = \{b_j \in \mathbb{R} \mid j \in J\}$ be given where $I = 1 : N_1,\ J = 1 : N_2$. Assume that the points a_i and b_j are ordered: $a_1 < a_2 < \cdots < a_{N_1},\quad b_1 < b_2 < \cdots < b_{N_2}$ and they are distinct: $a_i \neq b_j \quad \forall i \in I,\ j \in J$.

It is required to find a rather simple algorithm for "separation" of the sets A and B. If

$$\max\{a_i | i \in I\} < \min\{b_j | j \in J\}$$

or

$$\min\{a_i | i \in I\} > \max\{b_j | j \in J\}$$

then the separation is not difficult. Consider the case where the points of the sets are "mixed".

In this section the following identification procedure is used.

Let a function $F(x, c) : \mathbb{R}^2 \to \mathbb{R}$ be given. The function $F(x, c)$ will be referred to as an identifier (or classifier). Take an arbitrary $c \in A \cup B$. If $F(x, c) > 0$ then the point c is attributed to the set A. If $F(x, c) < 0$ then c is attributed to the set B. In the case $F(x, c) = 0$ the point c will be referred to as unidentified by the classifier $F(x, c)$ (and, hence, misclassified). In the sequel only the case $F(x, c) = c - x$ is considered. (The function $F(x, c) = x - c$ is employed if the sets A and B are interchanged, in practice, one should use both classifiers).

Let $m_1(x) = |A^-|$ where $A^- = \{a_i \in A \mid a_i \leq x\}, m_2(x) = |B^-|$ with $B^- = \{b_j \in B \mid b_j \geq x\}$. Here $|C|$ is the number of points in the set C (the cardinality of C). Thus, $m_1(x)(m_2(x))$ represents the number of points of the set A (respectively, B) which are misclassified by the identifier $F(x, c)$. Now let us construct a criterion function putting

$$m(x) = \max\{m_1(x), m_2(x)\}. \tag{1}$$

It is required to find $\min_{x \in \mathbb{R}} m(x) = m^*$. The function $m(x), m_1(x), m_2(x)$ takes only integer values, they are discontinuous and piecewise constant. The set A is the set of discontinuity points of the function m_1, and the set B is the set of discontinuity points of the function m_2. The function $m_1(x)$ is nondecreasing and, hence, quasiconvex, while the function $m_2(x)$ is nonincreasing and also quasiconvex (see [6]). The function $m(x)$, as the maximum of quasiconvex functions, is also quasiconvex (see [6]).

Remind the definition of a quasiconvex function. Let a finite-valued function $f(x)$ be defined on a convex set $\Omega \subset \mathbb{R}^n$. The function f is called *quasiconvex* on Ω if

$$f(\alpha x + (1-\alpha)y) \leq \max\{f(x), f(y)\} \quad \forall \alpha \in [0,1], \quad \forall x, y \in \Omega. \tag{2}$$

The following properties hold (see [6]):

If functions $f_i(x), i \in 1:N$, are quasiconvex on Ω then the function

$$f(x) = \max_{i \in 1:N} f_i(x)$$

is also quasiconvex on Ω.

For a function f to be quasiconvex on Ω it is necessary and sufficient that, for any $c \in \mathbb{R}$, the set $\mathcal{D}_c = \{x \in \Omega \mid f(x) \leq c\}$ be convex.

This implies that *the set of global minimizers of a quasiconvex function on a convex set is convex.* Note that, unlike the convex case, a quasiconvex function may have local minimizers which are not global ones.

Any convex function is quasiconvex, the opposite is not true.

Remark 1 As a criterion function, one can choose different functions. For example, $f_1(x) = m_1(x) + m_2(x)$ is of interest (it represents the total number of misclassified points). However, the set of global minimizers of the function m is convex while the function f_1 may have local minimizers, and the set of global minimizers can be nonconvex.

Example 1 In the example shown in Figs. 1 and 2, the sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ contain 8 points each: $A = \{a_1, \ldots, a_8\}$, $B = \{b_1, \ldots, b_8\}$. They are located on the real line in the increasing order as follows:

$$b_1, a_1, b_2, b_3, a_2, b_4, b_5, a_3, a_4, b_6, b_7, a_5, b_8, a_6, a_7, a_8.$$

The set M^* of minimizers of the function $m(x) = \max\{m_1(x), m_2(x)\}$ is the interval (b_5, a_4) (and this is a convex set, and $m^* = 3$), while the set of minimizers of the function $f_1(x) = m_1(x) + m_2(x)$ is the union of three intervals: $M^* = (b_5, a_3) \bigcup (b_7, a_5) \bigcup (b_8, a_6)$, each of them being not convex.

This was the reason for us to use the functional $m(x)$.

2.2 Necessary and sufficient optimality conditions

Let M^* be the set of minimizers of the function m defined by the relation (1). The function m is quasiconvex, hence, the set M^* is convex (see [6]). Since we consider the one-dimensional case, then the convex set M^*, which is not a singleton, takes one of the forms: $M^* = [p,q]$, $M^* = (p,q]$, $M^* = [p,q)$, $M^* = (p,q)$. The function m is piecewise continuous on $\mathbb{R} \backslash [A \cup B]$, therefore, the points p and q belong to $A \cup B$.

By $m_1\left(x^+\right)$ we denote the value $m_1(z)$ for $z \in (x, x')$ where x' is such that $(x, x') \cap [A \cup B] = \emptyset$ (i.e., in the interval (x, x') there are no points of the sets A and B). By $m_1(x^-)$ let

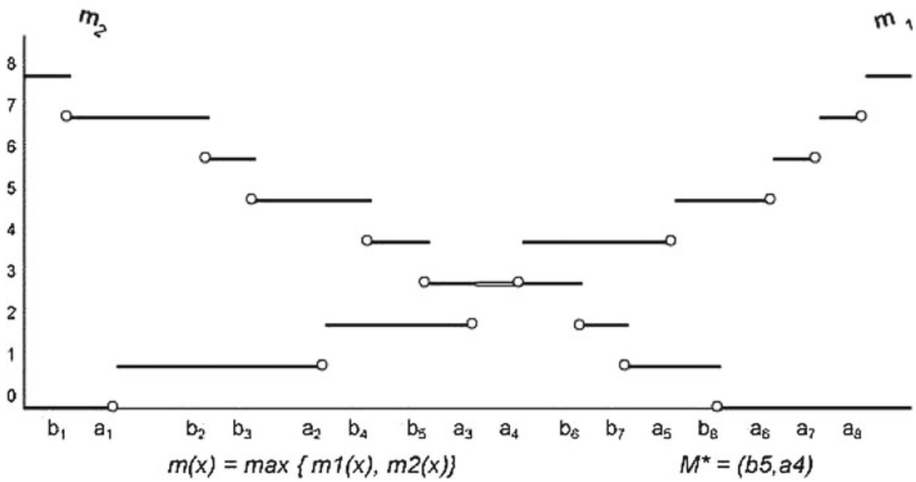

Fig. 1 Example 1. The functions $m_1, m_2, m = \max\{m_1, m_2\}$

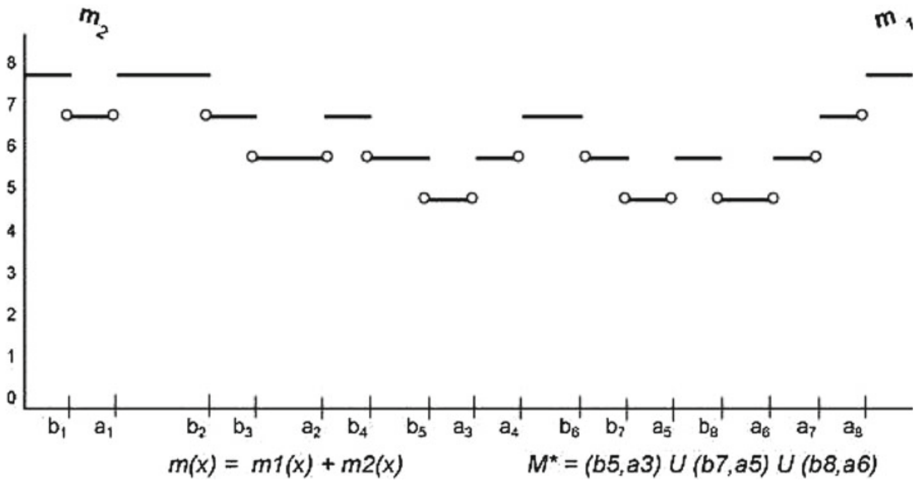

Fig. 2 Example 1. The function $f_1 = m_1 + m_2$

us denote the value $m_1(z)$ for $z \in (x', x)$ where x' is such that $(x', x) \cap [A \cup B] = \emptyset$ (i.e., in the interval (x', x) there are no points of the sets A and B).

The functions $m_2\left(x^+\right)$ and $m_2(x^-)$ are defined analogously. Put

$$m\left(x^+\right) = \max\left\{m_1\left(x^+\right), m_2\left(x^+\right)\right\}, m(x^-) = \max\left\{m_1(x^-), m_2(x^-)\right\}.$$

Theorem 1 *The set M^* is of the form $M^* = (b, a)$ where $a \in A, \quad b \in B$. Becides,*

$$m_2\left(b^+\right) = m_1(a^-) = m^*. \tag{3}$$

Proof Clearly,

$$m_1\left(x^+\right) \geq m_1(x), \quad m_2\left(x^+\right) \leq m_2(x),$$
$$m_1(x^-) \leq m_1(x), \quad m_2(x^-) \geq m_2(x).$$

If $a \in A, b \in B$ then

$$m_1\left(a^+\right) = m_1(a),\ m_1(a^-) = m_1(a) - 1,$$
$$m_1\left(b^+\right) = m_1(b^-) = m_1(b), m_2\left(a^+\right) = m_2(a^-) = m_2(a),$$
$$m_2\left(b^+\right) = m_2(b) - 1,\quad m_2(b^-) = m_2(b).$$

It is easy to observe that

$$m(a^-) \le m(a) = m\left(a^+\right), \tag{4}$$
$$m\left(b^+\right) \le m(b) = m(b^-). \tag{5}$$

Now, let us show that M^* is of the form $M^* = (p, q)$. Assuming the opposite, we get the following.

In the case $p \in M^*$: if $p = b \in B$, then $m_1(b^-) = m_1(b),\ m_2(b^-) = m_2(b)$, and the interval M^* can be expanded (to the left); if $p = a \in A$, then

$$m_1(a^-) = m_1(a) - 1,\quad m_2(a^-) = m_2(a),$$

and the interval M^* can be expanded (also to the left).

In the case $q \in M^*$: if $p = b \in B$, then

$$m_1\left(b^+\right) = m_1(b),\quad m_2\left(b^+\right) = m_2(b) - 1,$$

and the interval M^* can be expanded (to the right); if $p = a \in A$, then

$$m_1\left(a^+\right) = m_1(a),\quad m_2\left(a^+\right) = m_2(a),$$

and the interval M^* can be expanded (also to the right).

Thus, $M^* = (p, q)$. In the sequel it will be shown that $p \in B,\ q \in A$. First let us prove that $p \in B$. Assume the opposite, let $p = a \in A$. Then

$$m_1(a) = m_1\left(a^+\right),\quad m_1(a^-) = m_1(a) - 1 = m_1\left(a^+\right) - 1,$$
$$m_2(a^-) = m_2\left(a^+\right) = m_2(a),$$

therefore, $m(a^-) \le m\left(a^+\right)$ (i.e., the points which are close to a from the left, belong to M^*). It is a contradiction. Hence, $p \in B$.

Now let us show that $q \in A$. Assume the opposite, let $q = b \in B$. Then

$$m_2\left(b^+\right) = m_2(b) - 1 = m_2(b^-) - 1,\quad m_1\left(b^+\right) = m_1(b^-) = m_1(b),$$

therefore, $m\left(b^+\right) \le m(b^-)$ (i.e., the points which are close to a from the right, belong to M^*). It is again a contradiction. Hence, $q \in A$.

Thus, it is proved that $M^* = (b, a)$. Now let us show that $m_2\left(b^+\right) = m_1(a^-) = m^*$. Indeed, assume, e.g., that $m_2\left(b^+\right) < m^*$. Since

$$m_2(b^-) = m_2(b) = m_2\left(b^+\right) + 1 \le m^*,\quad m_1(b^-) = m_1\left(b^+\right) \le m^*,$$

i.e., $m(b^-) \le m^*$, and since $m(z) \ge m^*\quad \forall z$, then $m(b^-) = m^*$. Therefore, the set M^* can be expanded (to the left), which is impossible.

Analogously, assuming that $m_1(a^-) < m^*$, we have

$$m_1\left(a^+\right) = m_1(a^-) + 1 \le m^*,\quad m_2\left(a^+\right) = m_2(a^-) \le m^*,$$

i.e., $m\left(a^+\right) \le m^*$. Since $m(z) \ge m^*\quad \forall z$, then this yields $m\left(a_i^+\right) = m^*$. Therefore, one concludes that the set M^* can be expanded (this time to the right), which is again impossible.

Thus, the theorem is proved. □

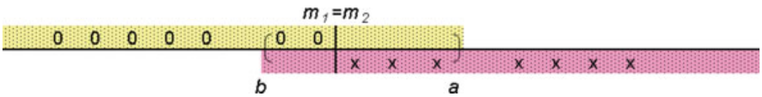

Fig. 3 Illustation to Theorem 1

Corollary 1 *In the immediate right-hand side neighborhood of the point b, the number of misclassified points of the set B is equal to the number of misclassified points of the set A in the immediate left-hand side neighborhood of the point a.*

Corollary 2 *There exist an subinterval* $\overline{M} = (\alpha^*, \beta^*) \subset M^*$ *(possibly coinciding with* M^**) such that*

$$m_1(x) = m_2(x) = m(x) \forall x \in (\alpha, \beta).$$

This implies that in the interval (b, a) *may be points from the set A and points from the set B, and the points from A should be located to the left of the points from the set B (see Fig. 3).*

In the above Example 1 (see Fig. 1) $M^* = (b_5, a_4)$, $\overline{M} = (\alpha^*, \beta^*) = (a_3, a_4)$. *It has already been observed that* $m^* = 3$.

Corollary 3 *For a point* x^* *to be a minimizer of the function* $m(x)$*, it is sufficient that*

$$m_2\left(x^*\right) = m_1\left(x^*\right). \tag{6}$$

Corollary 4 *The condition (3) is necessary and sufficient for every point of the interval* (b, a) *to be a minimizer of the functional* $m(x)$*. Such an interval is unique.*

Remark 2 It may happen that $M^* = (-\infty, a^*)$ or $M^* = (b^*, +\infty)$.

2.3 A numerical method for minimizing $m(x)$.

The above theorem makes it possible to describe a numerical method for constructing the set M^*. Choose an arbitrary set $M_0 = (p_0, q_0)$, where

$$p_0 \le \min\left\{\min_{i\in I} a_i,\ \min_{j\in J} b_j\right\}, \quad q_0 > \max\left\{\max_{i\in I} a_i, \max_{j\in J} b_j\right\}.$$

Then

$$m(p_0) = m_2(p_0) = N_1, \quad m(q_0) = m_1(q_0) = N_2, \quad m_1(p_0) = m_2(q_0) = 0.$$

Let the set $M_k = (p_k, q_k)$ have already been constructed such that $m_1(p_k) < m_2(p_k)$, $m_2(q_k) < m_1(q_k)$. Put $c_k = \frac{1}{2}(p_k + q_k)$.

If $m_1(c_k) = m_2(c_k)$ then $m(c_k) = m^*$ (i.e., c_k is a minimizer since m_2 is a decreasing function while m_1 is an increasing one). Now let us find the point $b_{j_k} \in B$, which is the closest to c_k from the left, and the point $a_{i_k} \in A$, which is closest to c_k from the right. Then $M^* = (b_{j_k}, a_{i_k})$.

If $m_1(c_k) < m_2(c_k)$ then put $M_{k+1} = (p_{k+1}, q_{k+1})$ where $p_{k+1} = c_k,\ q_{k+1} = q_k$. Note that

$$m_1(p_{k+1}) = m_1(c_k) < m_2(c_k) = m_2(p_{k+1}),$$
$$m_2(q_{k+1}) = m_2(q_k) < m_1(q_k) = m_1(q_{k+1}).$$

Finally, if $m_1(c_k) > m_2(c_k)$ then take $M_{k+1} = (p_{k+1}, q_{k+1})$, where $p_{k+1} = p_k,\ q_{k+1} = c_k$. Let us observe again that $m_1(p_{k+1}) = m_1(p_k) < m_2(p_k) = m_2(p_{k+1}), m_2(q_{k+1}) = m_2(c_k) < m_1(c_k) = m_1(q_{k+1})$. Continuing in an analogous way, let us construct a sequence of intervals $\{M_k\}$. It is not difficult to see that the process terminates in a finite number of steps, and the last obtained set is the desired set M^*. The number of steps will not exceed the value K where K is such that

$$\frac{q_0 - p_0}{2^K} < d = \min_{i \in I, j \in J} |a_i - b_j|.$$

Remark 3 We have described one possible algorithm. More sophisticated methods for minimizing quasiconvex functions can be used (see,e.g., [5]).

3 Identification via isolation

In this Section, the identification is performed by isolating one of the sets by means of an interval. The points belonging to this interval are attributed to one set (say, the set A), while the points not belonging to the interval are attributed to the set B. It is required to find an interval such that the number of misclassified points be minimal. As a functional, the maximum of the number of misclassified points in the both of the sets is chosen again. This functional (as a function of the ends of the interval, that is, a function of two variables) is discontinuous and takes integer values. In every of its coordinates this functional is quasiconvex, therefore, it is possible to use the coordinate-wise descent method (applying the algorithm described in Sect. 2 with respect to every coordinate for a fixed other coordinate). Necessary and sufficient optimality conditions are stated in Sect. 3.2.

3.1 Isolation method

Let $z = (x_1, x_2) \in \mathbb{R}^2, x_1 < x_2$.

As a classifier, let us take the function $F(z, c) = F(x_1, x_2, c) : \mathbb{R}^3 \to \mathbb{R}$, where $F(x_1, x_2, c) = \min\{c - x_1, x_2 - c\}$. The identification rule is as follows:

Let $c \in A \cup B$. If $F(z, c) > 0$ then the point c is attributed to the set A. If $F(z, c) < 0$ then c is attributed to B. In the case $F(z, c) = 0$ the point c is considered as misclassified by the classifier $F(z, c)$.

If $A \subset (x_1, x_2)$ and $B \bigcap [x_1, x_2] = \emptyset$, then the sets A and B are completely identified by the classifier F.

The function $F(z, c) = \min\{x_1 - c, c - x_2\}$ should be employed if the sets A and B are interchanged, in practice, both classifiers must be used.

Let $m_1(z) = |A^-|,\ m_2(z) = |B^-|$,

$$A^- = \{a_i \in A \mid a_i \notin (x_1, x_2)\},\ \ B^- = \{b_j \in B \mid b_j \in [x_1, x_2]\}.$$

Here $|C|$ is the cardinality of the set C. Thus, $m_1(z)$ (respectively, $m_2(z)$) is the number of points of the set A (respectively, B), which are misclassified by the classifier $F(z, c)$. As a criterion function let us choose the function

$$m(z) = \max\{m_1(z), m_2(z)\}. \tag{7}$$

It is required to find

$$\min_{z \in \mathbb{R}^2} m(z) = m^*.$$

The functions $m(z), m_1(z), m_2(z)$ are integer-valued, they are *discontinuous and peacewise constant*. A point $z = (x_1, x_2)$ is a discontinuity point of the function m_1 if at least one of the points x_1, x_2 belongs to the set A. A point $z = (x_1, x_2)$ is a discontinuity point of the function m_2 if at least one of the points x_1, x_2 belongs to the set B.

The function $m_1(z) = m_1(x_1, x_2)$ is nondecreasing in x_1 and nonincreasing in x_2, i.e., it is quasiconvex in every coordinate for a fixed other coordinate, and the function $m_2(z) = m_2(x_1, x_2)$ is nonincreasing in x_2 and nondecreasing in x_1, hence, it is quasiconvex in every coordinate for a fixed other coordinate. Therefore, the function $m(x_1, x_2)$, as the maximum of quasiconvex functions, is also quasiconvex in every coordinate for a fixed other coordinate. Note again that *the function $m(z)$ is discontinuous and integer-valued*.

The function $f_1(z) = m_1(z) + m_2(z)$ is of interest as well (it represents the total number of misclassified points in both sets), however, it lacks the coordinate-wise quasiconvexity.

3.2 Necessary optimality conditions

Let M^* be the set of intervals minimizing the functional $m(z)$.

Theorem 2 *If $z_0 \in M^*$ then there exists a set $M(z_0) \subset M^*$ of the form*

$$M(z_0) = \{(x_1, x_2) \mid x_1 \in (b_1, a_1),\ x_2 \in (a_2, b_2)\},$$

where $a_1, a_2 \in A,\ b_1, b_2 \in B$. In addition,

$$m_1\left(a_1^-, a_2^+\right) = m_2\left(b_1^+, b_2^-\right) = m^*. \tag{8}$$

Proof Let $z_0 = (x_{10}, x_{20}) \in M^*$. We construct the set $M(z_0)$ starting from the point z_0 and arguing for every coordinate as in the proof of Theorem 1. □

The condition (8) means that the number of points of the set A which are beyond the interval (a_1, a_2) (that is, misclassified) is equal to the number of points of the set B which are inside the interval (b_1, b_2) (that is, also misclassified). Besides, *there exist a subinterval* $(\beta_1, \alpha_1) \subset (b_1, a_1)$ (possibly coinciding with b́1 (b_1, a_1)) and *a subinterval* $(\alpha_2, \beta_2) \subset (a_2, b_2)$ (possibly coinciding with (a_2, b_2)) such that $m_1(x_1, x_2) = m_2(x_1, x_2)$ $\forall x_1 \in (\beta_1, \alpha_1), x_2 \in (\alpha_2, \beta_2)$.

This implies that it may happen that in the interval (b_1, a_1) there are points from the set A and points from B, and the points from A should be located to *the left* from the points of B, and in the interval (a_2, b_2) there are points from the set A and points from B, and the points from A should be located to *the right* from the points of B (see Fig. 4).

The condition (8) is necessary but not sufficient.

Since the functional $m(x_1, x_2)$ is *quasiconvex* in every variable, one can use the coordinate-wise descent method employing, e.g., the algorithm described in Sect. 2.3.

Put

$$\overline{M}(z_0) = \{(x_1, x_2) \mid x_1 \in (\beta_1, \alpha_1),\ x_2 \in (\alpha_2, \beta_2)\},$$

where $a_1, a_2 \in A,\ b_1, b_2 \in B$.

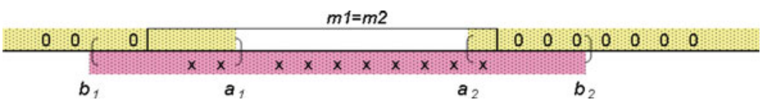

Fig. 4 Illustration to Theorem 2

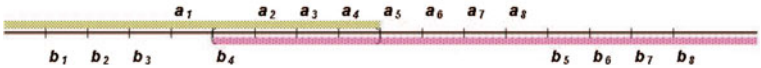

Fig. 5 Example 2. Separation method. $M^* = (b_4, a_5),\ m^* = 4$

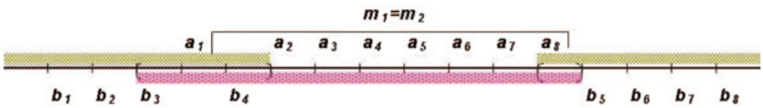

Fig. 6 Example 2. Isolation method. Set $M(z_0)$

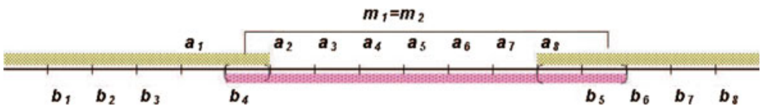

Fig. 7 Example 2. Isolation method. Set $M(z'_0)$

The aforesaid implies that

$$\overline{M}(z_0) \subset M(z_0),\ m_1(x_1, x_2) = m_2(x_1, x_2)\ \forall x_1 \in (\beta_1, \alpha_1), x_2 \in (\alpha_2, \beta_2).$$

Remark 4 Taking instead of $z_0 \in M^*$ another point $z'_0 \in M^*$, one can construct sets $M(z'_0)$ and $\overline{M}(z'_0) \subset M(z'_0)$, possessing the same properties as the sets $M(z_0)$ and $\overline{M}(z_0)$. There exist a finite number of such sets. Therefore, it is possible to formulate the following evident statement.

Theorem 3 *There exist a finite number of points $z_i \in M^*$ $(i \in 1 : r)$, such that*

$$M^* = \bigcup_{i \in 1:r} M(z_i).$$

Example 2 Let every of the sets A and B contain 8 points in $\mathbb{R}$ located in the increasing order as follows:

$$b_1, b_2, b_3, a_1, b_4, a_2, a_3, a_4, a_5, a_6, a_7, a_8, b_5, b_6, b_7, b_8.$$

First let us perform identification by the separation method. It is easy to see (see Fig. 5), that $M^* = (b_4, a_5)$, and $m^* = 4$.

Now, let us perform identification by the isolation method (in this case, let us try to isolate the set A). If we take the point $z_0 = (x_{10}, x_{20})$, with $x_{10} \in (b_3, a_1), x_{20} \in (a_8, b_5)$, then employing the above procedure, we get the sets (see Fig. 6)

$$M(z_0) = \{(x_1, x_2) \mid x_1 \in (b_3, a_2),\ x_2 \in (a_8, b_5)\},$$
$$\overline{M}(z_0) = \{(x_1, x_2) \mid x_1 \in (a_1, b_4),\ x_2 \in (a_8, b_5)\}.$$

If one takes the point $z'_0 = (x'_{10}, x'_{20})$, where $x'_{10} \in (b_4, a_2), x'_{20} \in (a_8, b_5)$, he gets the sets (see Fig. 7)

$$M(z'_0) = \{(x_1, x_2) \mid x_1 \in (b_4, a_2),\ x_2 \in (a_8, b_6)\},$$
$$\overline{M}(z'_0) = \{(x_1, x_2) \mid x_1 \in (b_4, a_2),\ x_2 \in (b_5, b_6)\}.$$

It is clear that

$$M^* = M(z_0) \bigcup M\left(z_0'\right).$$

Observe also that $m^* = 1$.

Remark 5 The discussed problems (identification by separation and isolation) can be solved just by considering all possible combinations and sorting out, however, if in the identification method only $N_1 + N_2$ cases should be considered, the isolation method requires to consider $C^2_{N_1+N_2}$ cases. The problem becomes much more complicated if the isolation is performed by means of two intervals.

Remark 6 In the case of isolation by two or more intervals, it is possible to get optimality conditions generalizing Theorem 2.

Remark 7 The choice of separation and isolation methods is justified by the following. Let two normally distributed one-dimensional random variables ξ_1 and ξ_2 be given As a classifier take the following one: let us partition $\mathbb{R}$ into two sets Ω_1 and $\Omega_2 : \mathbb{R} = \Omega_1 \bigcup \Omega_2$. If some $x \in \mathbb{R}$ is observed, we attribute x to ξ_1 if $x \in \Omega_1$, and to ξ_2, if $x \in \Omega_2$. It is known that the optimal (in the sense of minimizing the maximum of probability of misclassified points) is the partitioning of $\mathbb{R}$ into two semi-lines (if the variances of ξ_1 and ξ_2 are equal) or the partitioning into an interval and its complement (if the variances of ξ_1 and ξ_2 differ).

3.3 Ranking by means of one-dimentional identification

Let two sets $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^n$ be given:

$$A = \left\{a_i \in \mathbb{R}^n \mid i \in I\right\}, \quad B = \left\{b_j \in \mathbb{R}^n \mid j \in J\right\},$$

$I = 1 : N_1,\ J = 1 : N_2,\ a_i = (a_{i1}, \ldots, a_{in}),\ b_j = (b_{j1}, \ldots, b_{jn})$.

For the simplicity reason it is assumed that $N_1 = N_2 = N$. For the k-th coordinate we obtain the one-dimensional set $A_k = \{a_{ik} \mid i \in I\},\ B_k = \{b_{jk} \mid j \in J\}$. Applying the above described methods, let us identify the sets A_k and B_k and find the value m_k^*. The value $\frac{m_k^*}{N}$ can be treated as the part (fraction) of misclassified points of the sets A_k and B_k, while $\mu_k = 1 - \frac{m_k^*}{N}$ is the part (fraction) of correctly classified points. Arranging μ_k's in the decreasing order, it is possible to rank the coordinates according to their informative value.

Example 3 Let the sets A and B contain 8 points in the n-dimensional space. Assume that their first and second coordinates represent the sets described in Examples 1 and 2, respectively. To identify the sets A and B in the first coordinate (i.e., the sets A_1 and B_1, see Fig. 1) the separation method was used and it was found that $m_1^* = 3$. Applying the separation method for identifying the sets A and B in the second coordinate (i.e., the sets A_2 and B_2) we have obtained (see Fig. 5) $m_2^* = 4$. Then $\mu_1 = 1 - \frac{m_1^*}{8} = 5/8$, and $\mu_2 = 1 - \frac{m_2^*}{8} = 4/8 = 1/2$. Therefore, we conclude that the informative value of the first coordinate is higher than the informative value of the second coordinate (since $\mu_1 > \mu_2$). But if for identifying the sets A and B in the second coordinate (i.e., the sets A_2 and B_2) we use the isolation method (see Figs. 6 and 7), then one gets $m_2^* = 1$. Then $\mu_1 = 1 - \frac{m_1^*}{8} = 5/8$, and $\mu_2 = 1 - \frac{m_2^*}{8} = 7/8$, that is the informative value of the second coordinate is higher than the informative value of the first coordinate (since $\mu_2 > \mu_1$). Thus, the informative value of a coordinate depends on an identification method used.

4 Surrogate functionals

As a criterion functional, we have chosen the maximum of the number of misclassified (in the sets A B) points. It is the so-called *natural functional*. This functional is discontinuous. In the one-dimensional case we have managed to overcome the difficulties caused by the discontinuity and to solve the stated identification problem. Another approach consists in replacing the natural functional by some other functional (called *surrogate*) possessing proper analytical properties allowing to employ the existing optimization methods.

As an example, consider the class of linear classifiers (i.e., as a classifier we take the function $F(x, c) = c - x$). This is identification by the separation method. As a surrogate functional, one can choose, for example, the sum of the distances of misclassified points to the point x, that is the functional

$$H(x) = \sum_{i \in I} \max\{0, x - a_i\} + \sum_{j \in J} \max\{0, b_j - x\}. \tag{9}$$

Put

$$\varphi_i(x) = \max\{0, x - a_i\} \quad \forall i \in I, \qquad \psi_j(x) = \max\{0, b_j - x\} \quad \forall j \in J.$$

The function $H(x)$ is continuous and even convex. Its subdifferential (in the sense of convex analysis—see [4]) is

$$\partial H(x) = \sum_{i \in I_-(x)} \partial\varphi_i(x) + \sum_{j \in J_+(x)} \partial\psi_j(x), \tag{10}$$

where

$$I_-(x) = \{i \in I \mid \varphi_i(x) \le 0\}, \qquad J_+(x) = \{j \in J \mid \psi_j(x) \ge 0\},$$

$$\partial\varphi_i(x) = \begin{cases} 1, \varphi_i(x) < 0, \\ co\{0, 1\}, \ \varphi_i(x) = 0, \end{cases} \qquad \partial\psi_j(x) = \begin{cases} -1, \ \psi_j(x) > 0, \\ co\{-1, 0\}, \ \varphi_j(x) = 0. \end{cases}$$

If a point x^* is a minimizer of the function H, then by the necessary and sufficient condition for a minimum $0 \in \partial H(x^*)$. First, consider the case where x^* belongs neither the set A nor the set B. Then (10) yields

$$\left|I_-\left(x^*\right)\right| = \left|J_+\left(x^*\right)\right|, \tag{11}$$

that is, the number of misclassified points of the set A is equal to the number of misclassified points of the set B. In Sect. 2.2, Corollary 2 it was shown that this is the set $\overline{M}$. It is an open interval. At the boundary points the necessary condition for a minimum holds as well since in this case either the subdifferential of the function φ_i will be added to the subdifferential of H (if this is a point of A and the subdifferetial of φ is $[0, 1]$), or the subdifferential of the function ψ_j (if this is a point of B and the subdifferetial of ψ is $[-1, 0]$).

The condition $0 \in \partial H(x^*)$ holds if as a representative of this set we take the point 0. At points not belonging to the closure of the set $\overline{M}$ the condition for a minimum (11) does not hold. Thus, the set of minimizers of the *surrogate* functional H coincides with the closure of the set $\overline{M}$ (and this is a subset of the set of minimizers of the *natural* functional $m(x)$).

Remark 8 For the problem of identification by means of the isolation method it is also possible to construct a similar surrogate functional. In this case as a classifier we take the function $F(z, c) = F(x_1, x_2, c) : \mathbb{R}^3 \to \mathbb{R}$, where $F(x_1, x_2, c) = \min\{c - x_1, x_2 - c\}$. Put

$$\varphi(z, c) = \max\{0, \min\{c - x_1, x_2 - c\}\}, \quad \psi(z, c) = \max\{0, x_1 - c, c - x_2\}.$$

Introduce the functional

$$H_1(z) = \sum_{i \in I} \psi(z, a_i) + \sum_{j \in J} \varphi(z, b_j). \tag{12}$$

The value of this functional at a point z is equal to the sum of distances of misclassified points to the nearest boundary point of the interval $[x_1, x_2]$. This functional is not convex anymore, however, it is quasidifferentiable (see [1,2]). By rules of quasidifferential calculus one can formulate a necessary condition for a minimum (which isn't, unlike in the convex case, sufficient anymore) and show that the set of inf-stationary points of this functional (i.e., points satisfying the necessary condition for a minimum) is a closure of a subset of the set M^* defined in Theorem 3.

References

1. Demyanov, V.F., Vasiliev, L.V.: Nondifferentiable Optimization. Springer-Optimization Software, New York (1986)
2. Demyanov, V.F., Rubinov, A.M.: Constructive Nonsmooth Analysis. Verlag Peter Lang, Frankfurt a/M (1995)
3. Demyanova, V.V.: One-dimensional identification by the separation method // Vestnik of St.-Petersburg University; Ser. 10. Applied mathematics, informatics, control processes. No. 3, pp. 28–31 (2006)
4. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
5. Wilde, D.J.: Optimum-Seeking Methods. Prentice-Hall, Englewood Cliffs (1963)
6. Zabotin, Y.I., Korablev, F.I., Khabibulin, R.F.: On minimizing quasiconvex functionals // Izvestiya Vuzov. Mathematics. No. 10, pp. 27–33 (1972)