9 AN EXCHANGE ALGORITHM FOR MINIMIZING SUM-MIN FUNCTIONS

Alexei V. Demyanov

Faculty of Mathematics and Mechanics
St.Petersburg State University
Staryi Peterhof, St.Petersburg
198904 RUSSIA

Abstract: The problem of minimizing maxmin-type functions or the sum of minimum-type functions appears to be increasingly interesting from both theoretical and practical considerations. Such functions are essentially nonsmooth and, in general, they can successfully be treated by existing tools of Nonsmooth Analysis. In some cases the problem of finding a minimizer of such a function can be reduced to solving some mixed combinatorial–continuous problem.

In the present paper the problem of minimizing the sum of minima of a finite number of functions is discussed. It is shown that this problem is equivalent to solving a finite (though may be quite large) number of simpler (and sometimes quite trivial) optimization problems. Necessary conditions for global a minimum and sufficient conditions for a local minimum are stated. An algorithm for finding a local minimizer (the so-called exchange algorithm) is proposed. It converges to a local minimizer in a finite number of steps. A more general algorithm (called ε–exchange algorithm) is described which allows one to escape from a local minimum point. Numerical examples demonstrate the algorithm.

Key words: Sum-min function, necessary conditions for a minimum, sufficient conditions, exchange algorithm, ε–exchange algorithm.

1 INTRODUCTION

Let functions $\varphi_{ij}(x) : R^n \to R$ be given, where $i \in I := 1 : m, j \in J_i := 1 : N_i$. Construct the functions

$$\varphi_i(x) = \min_{j \in J_i} \varphi_{ij}(x) \quad \forall i \in 1 : m \tag{1.1}$$

and the function

$$F(x) = \sum_{i \in I} \varphi_i(x). \tag{1.2}$$

The function F defined by (1.2) is called a sum-min function. Many practical problems arising, e.g., in Mathematical Diagnostics, Data Mining, Clustering, Network Allocation etc. (see Mangasarian (1997), Rao (1971), Bagirov, Rubinov (1999), Bagirov, Rubinov, Yearwood (2001), Rubinov (2000)) can be described by mathematical models where it is required to find a minimizer of F. The problem of minimizing the function F is, first of all, generally speaking, nonsmooth (even if all the functions φ_{ij}'s are quite good) and "extremely" multiextremal. The nonsmoothness can be treated by the existing tools of Nonsmooth Analysis and Nondifferentiable Optimization (for example, local properties of F can be studied, under some conditions, imposed on φ_{ij}, by considering the directional derivatives of F). But the multiextremality remains the main unavoidable obstacle.

In the paper we consider the problem of minimizing one class of sum-min functions (the so-called separable-like sum-min functions). It is known that this problem is equivalent to solving a finite number of "simpler" problems. Namely, it is shown that

$$\inf_{x \in R^n} F(x) \;=\; \min_{j \in J} \inf_{x \in R^n} F_j(x) \tag{1.3}$$

where

$$J \;=\; \{j = (j_1, ..., j_m) \mid j_i \in J_i \quad \forall i \in J\},$$
$$F_j(x) \;=\; \sum_{i \in I} \varphi_{ij_i}(x).$$

A similar result for the best piece-wise polynomial approximation problem was proved in Vershik, Malozemov, Pevnyi (1975).

For every $j \in J$ the function $F_j(x)$ is called an elementary function, and it is assumed that one is able to find a minimizer of $F_j(x)$ (exactly or approximately). The problem of minimizing $F_j(x)$ will also be referred to as an elementary problem. In some cases the problem of minimizing F_j is quite simple (see Examples in Section 7).

Theoretically, it is possible to find a global minimizer of the function F by solving all elementary problems, however, the number of such elementary problems is too large, therefore it is important to be able to find a proper local minimizer at a reasonable computational price.

The paper is organized as follows. In Section 2 a special case of the above problem is stated (namely, the case $m = 2$ and $\varphi_{ij}(x) = \varphi_{ij}(x_i)$, where $x = (x_1, ..., x_m)$, $x_i \in R^{n_i}$, $\sum_{i \in I} n_i = n$). A result analogous to (1.3) is formulated in Section 3. Necessary conditions for a point to be a global minimizer and sufficient conditions for a point to be a local minimizer is proved in Section 4. An algorithm for finding a local minimizer is described in Section 5. This algorithm (called an exchange algorithm) is based on the necessary minimality conditions. In a finite number of steps a local minimizer is constructed. In Section 6 a modification of the exchange algorithm (ε–exchange algorithm) is proposed to "escape" from a local minimum point. Illustrative numerical examples are described in Section 7.

2 STATEMENT OF THE PROBLEM

Let a set of points $\Omega = \{t_1, \dots, t_N\} \subset Y$ (where Y is some space) and functions $\varphi_i : \Omega \times R^{n_i} \to R$, $i \in 1:m$, be given. Put $x = (x_1, \dots, x_m) \in S := R^{n_1} \times \ldots \times R^{n_m}$.

Let

$$\varphi(t, x) = \min_{i \in 1:m} \varphi_i(t, x_i)$$

and

$$F(x) = \sum_{t \in \Omega} \varphi(t, x) = \sum_{t \in \Omega} \min_{i \in 1:m} \varphi_i(t, x_i).$$

Problem P: Find a point $x^* = (x_1^*, \dots, x_m^*) \in S$, such that

$$F(x^*) = \min_{x \in S} F(x).$$

In the paper the case $m = 2$ is described in detail. Note that no specific requirements on φ_i are imposed (the functions φ_i's are not even assumed to be continuous).

Thus, we consider the case $m = 2$, $S = R^{n_1} \times R^{n_2}$. Then

$$F(x) = \sum_{t \in \Omega} \min \left\{ \varphi_1(t, x_1), \varphi_2(t, x_2) \right\}.$$

For every $\gamma_1 \subset \Omega$ and $\gamma_2 \subset \Omega$ put

$$c_1(x_1, \gamma_1) = \sum_{t \in \gamma_1} \varphi_1(t, x_1), \; c_2(x_2, \gamma_2) = \sum_{t \in \gamma_2} \varphi_2(t, x_2). \tag{2.1}$$

If $\gamma = \emptyset$, then by definition $c_1(x_1, \emptyset) = c_2(x_2, \emptyset) = 0$.

Let us introduce the set

$$T(\Omega) = \left\{ \Gamma = (\gamma_1, \gamma_2) \middle| \gamma_1 \subset \Omega, \gamma_2 \subset \Omega, \; \gamma_1 \cap \gamma_2 = \emptyset, \; \gamma_1 \cup \gamma_2 = \Omega \right\} \subset 2^\Omega \times 2^\Omega.$$

For any $\Gamma = (\gamma_1, \gamma_2) \in T(\Omega)$ let us consider the function

$$F_\Gamma(x) = c_1(x_1, \gamma_1) + c_2(x_2, \gamma_2).$$

It is clear, that for every $\Gamma \in T(\Omega)$

$$F(x) \leq F_\Gamma(x) \quad \forall x \in S. \tag{2.2}$$

For each fixed point $x \in S$ let us introduce the sets

$$\widehat{\sigma}_1(x) = \left\{ t \in \Omega \middle| \varphi_1(t, x_1) < \varphi_2(t, x_2) \right\},$$

$$\widehat{\sigma}_2(x) = \left\{ t \in \Omega \middle| \varphi_1(t, x_1) > \varphi_2(t, x_2) \right\},$$

$$\Sigma(x) = \left\{ t \in \Omega \middle| \varphi_1(t, x_1) = \varphi_2(t, x_2) \right\}.$$

The set $\Sigma(x)$ is called the set of "common points".

Let sets $\widetilde{\sigma}_1, \widetilde{\sigma}_2$ be such that

$$\widetilde{\sigma}_1 \subset \Sigma(x), \; \widetilde{\sigma}_2 \subset \Sigma(x), \; \widetilde{\sigma}_1 \cap \widetilde{\sigma}_2 = \emptyset, \; \widetilde{\sigma}_1 \cup \widetilde{\sigma}_2 = \Sigma(x).$$

Put

$$\sigma_1 = \widetilde{\sigma}_1 \cup \widehat{\sigma}_1(x), \; \sigma_2 = \widetilde{\sigma}_2 \cup \widehat{\sigma}_2(x). \tag{2.3}$$

Clearly, $\sigma_1 \cup \sigma_2 = \Omega$, $\sigma_1 \cap \sigma_2 = \emptyset$. Hence, any distribution $(\widetilde{\sigma}_1, \widetilde{\sigma}_2)$ of the common points generates the corresponding disjoint partition (σ_1, σ_2) of the set Ω.

Definition 2.1 *Every such a partition (it depends upon x and, evidently, is not unique) is called an x-proper partition of the set Ω.*

Let $T(\Omega, x)$ denote the family of all x-proper partitions of the set Ω. Clearly,

$$T(\Omega, x) \subset T(\Omega). \tag{2.4}$$

Besides, it is easy to see that for any $\Gamma = (\gamma_1, \gamma_2) \in T(\Omega, x)$ the relation

$$F_\Gamma(x) = F(x) \tag{2.5}$$

holds.

For $\Gamma = (\gamma_1, \gamma_2) \in T(\Omega)$ let us introduce the function

$$\Phi(\Gamma) = \inf_{x \in S} F_\Gamma(x) = \inf_{x_1 \in R^{n_1}} c_1(x_1, \gamma_1) + \inf_{x_2 \in R^{n_2}} c_2(x_2, \gamma_2).$$

Now it is possible to formulate the following

Problem P1: Find $\Gamma^* \in T(\Omega)$ such that

$$\Phi(\Gamma^*) = \min_{\Gamma \in T(\Omega)} \Phi(\Gamma).$$

3 EQUIVALENCE OF THE TWO PROBLEMS

It will now be shown that the problem P is equivalent to the problem P1, that is the problem of minimizing the function $F(x)$ on S is equivalent to the problem of minimizing the function $\Phi(\Gamma)$ on the set $T(\Omega)$.

Let us suppose that for any $\sigma_1 \subset \Omega$ and $\sigma_2 \subset \Omega$ the functions $c_1(x_1, \sigma_1)$ and $c_2(x_2, \sigma_2)$ attain their minimal values on R^{n_1} and R^{n_2}, respectively.

Theorem 3.1 *The following equality holds:*

$$\inf_{x \in S} F(x) = \min_{\Gamma \in T(\Omega)} \Phi(\Gamma). \tag{3.1}$$

Proof: It is clear from (2.2) that for any $\Gamma = (\gamma_1, \gamma_2) \in T(\Omega)$

$$\inf_{x \in S} F(x) \leq \inf_{x \in S} F_\Gamma(x) = \inf_{x_1 \in R^{n_1}} c_1(x_1, \gamma_1) + \inf_{x_2 \in R^{n_2}} c_2(x_2, \gamma_2) = \Phi(\Gamma).$$

Hence,

$$\inf_{x \in S} F(x) \leq \min_{\Gamma \in T(\Omega)} \Phi(\Gamma). \tag{3.2}$$

Let us take an arbitrary $\overline{x} = (\overline{x}_1, \overline{x}_2) \in S$. For every $\Gamma_0 = (\sigma_1, \sigma_2) \in T(\Omega, \overline{x})$, the conditions (2.4) and (2.5) yield the relations

$$F(\overline{x}) = c_1(\overline{x}_1, \sigma_1) + c_2(\overline{x}_2, \sigma_2) \geq$$

$$\geq \inf_{x_1 \in R^{n_1}} c_1(x_1, \sigma_1) + \inf_{x_2 \in R^{n_2}} c_2(x_2, \sigma_2) = \Phi(\Gamma_0) \geq \min_{\Gamma \in T(\Omega)} \Phi(\Gamma).$$

Since an arbitrary $\overline{x} \in S$ was chosen,

$$\inf_{x \in S} F(x) \geq \min_{\Gamma \in T(\Omega)} \Phi(\Gamma). \tag{3.3}$$

The inequalities (3.2) and (3.3) imply (3.1). △

Take any $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega)$ and find $x_1(\sigma_1) \in R^{n_1}$, $x_2(\sigma_2) \in R^{n_2}$ such that

$$\min_{x_1 \in R^{n_1}} c_1(x_1, \sigma_1) = c_1(x_1(\sigma_1), \sigma_1), \tag{3.4}$$

$$\min_{x_2 \in R^{n_2}} c_2(x_2, \sigma_2) = c_2(x_2(\sigma_2), \sigma_2). \tag{3.5}$$

The point $x(\Gamma) = (x_1(\sigma_1), x_2(\sigma_2))$ is not unique, if the minima in (3.4) or (3.5) are attained at more than one point.

Remark 3.1 *Theorem 3.1 implies that the problem of minimizing the function F on S is reduced to the problem of solving a finite number (precisely, $|T(\Omega)|$) of problems of minimizing the functions of the form*

$$F_\Gamma(x) = c_1(x_1, \sigma_1) + c_2(x_2, \sigma_2),$$

where $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega)$. However, since $|T(\Omega)| = 2^{|\Omega|}$, then a quite large number of points in the set Ω can annihilate the practical value of this fact.

Here, as usual, $|A|$ stands for the number of points in a set A.

Thus, the problem of minimizing F on Ω becomes a combinatorial one (if the problem of minimizing $F_\Gamma(x)$ is assumed to be an elementary one).

4 MINIMALITY CONDITIONS

Theorem 4.1 *For a point $x^* \in S$ to be a global minimizer of the function F it is necessary that for any x^*-proper partition $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega, x^*)$ the following conditions hold:*

$$i)\ c_1(x_1^*, \sigma_1) = \min_{x_1 \in R^{n_1}} c_1(x_1, \sigma_1), \tag{4.1}$$

$$ii)\ c_2(x_2^*, \sigma_2) = \min_{x_2 \in R^{n_2}} c_2(x_2, \sigma_2). \tag{4.2}$$

If, in addition, the functions φ_i's, $i \in 1:2$, are continuous (respectively, in x_i), then the above conditions are sufficient conditions for a point x^* to be a local minimizer.

Proof: Necessity Let x^* be a global minimizer of the function F. Assume that the theorem is not valid. Let, for example, the condition (4.1) don't hold. Then there exists a point $\overline{x}_1 \in R^{n_1}$ such that

$$c_1(x_1^*, \sigma_1) > c_1(\overline{x}_1, \sigma_1). \tag{4.3}$$

Then for the point $\overline{x} = (\overline{x}_1, x_2^*) \in S$ by (4.3) and (2.2) one gets

$$F(x^*) = c_1(x_1^*, \sigma_1) + c_2(x_2^*, \sigma_2) > c_1(\overline{x}_1, \sigma_1) + c_2(x_2^*, \sigma_2) \geq F(\overline{x}),$$

which contradicts the fact, that the point x^* is a global minimizer of F.
Sufficiency Let the functions φ_i's be continuous in x_i and the conditions (4.1), (4.2) hold. Then the function

$$\varphi(t, x) = \min\left\{\varphi_1(t, x_1), \varphi_2(t, x_2)\right\}$$

is also continuous. The continuity of the function φ with respect to $x \in S$ implies that

- for every $t_i \in \widehat{\sigma}_1(x^*)$ there exists $\delta_i > 0$ such that $t_i \in \widehat{\sigma}_1(x)$ for any $x \in B(x^*, \delta_i)$,
- for every $t_j \in \widehat{\sigma}_2(x^*)$ there exists $\varepsilon_j > 0$ such that $t_j \in \widehat{\sigma}_2(x)$ for any $x \in B(x^*, \varepsilon_j)$.

Here $B(x^*, \delta) = \{x \in S \mid \|x - x^*\| \leq \delta\}$.

Put

$$\delta = \min\{\delta_i, \varepsilon_j | t_i \in \widehat{\sigma}_1(x^*),\ t_j \in \widehat{\sigma}_2(x^*)\}.$$

Thus, for all $\overline{x} \in B(x^*, \delta)$ we have

$$\widehat{\sigma}_1(x^*) \subset \widehat{\sigma}_1(\overline{x}), \quad \widehat{\sigma}_2(x^*) \subset \widehat{\sigma}_2(\overline{x}). \tag{4.4}$$

Let us consider the sets

$$A_1 = \widehat{\sigma}_1(\overline{x}) \setminus \widehat{\sigma}_1(x^*), \quad A_2 = \widehat{\sigma}_2(\overline{x}) \setminus \widehat{\sigma}_2(x^*).$$

Now we will prove the inclusions $A_1 \subset \Sigma(x^*)$, $A_2 \subset \Sigma(x^*)$.

Let $t \in A_1$, then, since

$$\Omega = \widehat{\sigma}_1(x^*) \cup \widehat{\sigma}_2(x^*) \cup \Sigma(x^*),$$

either $t \in \widehat{\sigma}_2(x^*)$, or $t \in \Sigma(x^*)$. However, if $t \in \widehat{\sigma}_2(x^*) \subset \widehat{\sigma}_2(\overline{x})$, then we have

$$t \in \widehat{\sigma}_1(\overline{x}) \cap \widehat{\sigma}_2(\overline{x}),$$

which contradicts the fact that $\widehat{\sigma}_1(\overline{x}) \cap \widehat{\sigma}_2(\overline{x}) = \emptyset$. So, $t \in \Sigma(x^*)$, and, hence, $A_1 \subset \Sigma(x^*)$. Similarly we get $A_2 \subset \Sigma(x^*)$.

Now let us take an arbitrary disjoint partition $(\widetilde{\sigma}_1, \widetilde{\sigma}_2)$ of the set $\Sigma(\overline{x})$. Since for the corresponding $\overline{x}$-proper partition $(\sigma_1, \sigma_2) \in T(\Omega, \overline{x})$ we obtain

$$\sigma_1 = \widehat{\sigma}_1(\overline{x}) \cup \widetilde{\sigma}_1 = \widehat{\sigma}_1(x^*) \cup A_1 \cup \widetilde{\sigma}_1, \tag{4.5}$$

$$\sigma_2 = \widehat{\sigma}_2(\overline{x}) \cup \widetilde{\sigma}_2 = \widehat{\sigma}_2(x^*) \cup A_2 \cup \widetilde{\sigma}_2. \tag{4.6}$$

It is easy to see, that from (4.4) it follows, that $\widetilde{\sigma}_1 \subset \Sigma(\overline{x}) \subset \Sigma(x^*)$, and $\widetilde{\sigma}_2 \subset \Sigma(\overline{x}) \subset \Sigma(x^*)$, therefore

$$A_1 \cup \widetilde{\sigma}_1 \subset \Sigma(x^*), \quad A_2 \cup \widetilde{\sigma}_2 \subset \Sigma(x^*),$$

and $(A_1 \cup \widetilde{\sigma}_1, A_2 \cup \widetilde{\sigma}_2)$ is a disjoint partition of the set $\Sigma(x^*)$.

So, the equalities (4.5) and (4.6) ensure us, that (σ_1, σ_2) is an x^*-proper partition of the set Ω, and the hypotheses of the theorem are satisfied. Hence, the condition (4.1) implies that

$$c_1(x_1^*, \sigma_1) \leq c_1(\overline{x}_1, \sigma_1), \tag{4.7}$$

and the condition (4.2) implies that

$$c_2(x_2^*, \sigma_2) \leq c_2(\overline{x}_2, \sigma_2). \tag{4.8}$$

It follows from (4.7) and (4.8) that x^* is a local minimizer of the function F.

$\triangle$

Definition 4.1 *A point $x^* \in S$ satisfying the conditions (4.1) and (4.2) is called a stationary point.*

If φ_i are continuous, a stationary point is a local minimizer. The opposite is not true: not every local minimizer is a stationary point.

Remark 4.1 *The notion of stationary point is closely related to the necessary condition used. Since the conditions (4.1) and (4.2) are necessary conditions for a global minimum then it is natural to expect that not every local minimizer is a stationary point. Note, that conditions (4.1) and (4.2) are of nonlocal nature.*

5 AN EXCHANGE ALGORITHM

Let us suppose that for every $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega)$ infima

$$\inf_{x_i \in R^{n_i}} c_i(x_i, \sigma_i), \ \ i = 1, 2,$$

are attained, that is there exists a point $x(\Gamma) = (x_1(\sigma_1), x_2(\sigma_2)) \in S$, such that

$$c_i(x_i(\sigma_i), \sigma_i) = \min_{x_i \in R^{n_i}} c_i(x_i, \sigma_i), \ \ i = 1, 2. \tag{5.1}$$

The following algorithm allows one to find a stationary point of the function F (and if φ_i, $i = 1, 2$, are continuous, then the resulting point will be a local minimizer of F).

1. Take an arbitrary $x^0 = (x_1^0, x_2^0) \in S$.
2. Let $x^k = (x_1^k, x_2^k) \in S$ have already been found. Construct the sets $\widehat{\sigma}_1(x^k)$, $\widehat{\sigma}_2(x^k)$ and $\Sigma(x^k)$.
3. Check the conditions (4.1) and (4.2) for all $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega, x^k)$.
4. If the conditions (4.1) and (4.2) are satisfied for all $\Gamma \in T(\Omega, x^k)$ then the point x^k is stationary, and the process terminates.
5. Otherwise, find any $\Gamma_k = (\sigma_1^k, \sigma_2^k) \in T(\Omega, x^k)$ for which one of the conditions (4.1), (4.2) is violated.
6. ▪ If the condition (4.1) holds, then put $\overline{x}_1^k = x_1^k$.
   ▪ If not, then find $\overline{x}_1^k \in R^{n_1}$ such that

$$\min_{x_1 \in R^{n_1}} c_1(x_1, \sigma_1^k) = c_1(\overline{x}_1^k, \sigma_1^k) < c_1(x_1^k, \sigma_1^k).$$

- If the condition (4.2) holds, then put $\overline{x}_2^k = x_2^k$.
- If not, then find $\overline{x}_2^k \in R^{n_2}$ such that

$$\min_{x_2 \in R^{n_2}} c_2(x_2, \sigma_2^k) = c_2(\overline{x}_2^k, \sigma_2^k) < c_2(x_2^k, \sigma_2^k).$$

7. Put $x^{k+1} = (\overline{x}_1^k, \overline{x}_2^k)$. Go to step 2.

Clearly,

$$F(x^{k+1}) < F(x^k). \tag{5.2}$$

As a result, a finite sequence $\{x^k\}$ is constructed such that condition (5.2) holds. Since every x-proper partition (σ_1^k, σ_2^k) may occur only once (due to (5.2)), then, taking into account the fact that $|T(\Omega)|$ is finite, one concludes that the algorithm converges to a stationary point in a finite number of steps.

Remark 5.1 *The algorithm described above may require at some steps complete enumeration of the set $|T(\Omega, x^k)| = 2^{|\Sigma(x^k)|}$. So, in practice the algorithm is effective if $|\Sigma(x^k)|$ is not very large. Theoretically, the case of complete enumeration of $T(\Omega)$ is possible (as for every algorithm of discrete mathematics).*

6 AN ε-EXCHANGE ALGORITHM

In this section it is assumed that φ_i's are continuous. For every fixed point $x \in S$ and $\varepsilon > 0$ let us introduce the sets

$$\widehat{\sigma}_{\varepsilon 1}(x) = \left\{ t \in \Omega \,\middle|\, \varphi_1(t, x_1) < \varphi_2(t, x_2) - \varepsilon \right\},$$

$$\widehat{\sigma}_{\varepsilon 2}(x) = \left\{ t \in \Omega \,\middle|\, \varphi_1(t, x_1) - \varepsilon > \varphi_2(t, x_2) \right\},$$

$$\Sigma_\varepsilon(x) = \left\{ t \in \Omega \,\middle|\, |\varphi_1(t, x_1) - \varphi_2(t, x_2)| \le \varepsilon \right\}.$$

The set $\Sigma_\varepsilon(x)$ is called the set of "ε-common points".

Let sets $\widetilde{\sigma}_1, \widetilde{\sigma}_2$ be such that

$$\widetilde{\sigma}_1 \subset \Sigma_\varepsilon(x), \ \widetilde{\sigma}_2 \subset \Sigma_\varepsilon(x), \ \widetilde{\sigma}_1 \cap \widetilde{\sigma}_2 = \emptyset, \ \widetilde{\sigma}_1 \cup \widetilde{\sigma}_2 = \Sigma_\varepsilon(x).$$

Put

$$\sigma_1 = \widetilde{\sigma}_1 \cup \widehat{\sigma}_{\varepsilon 1}(x), \ \sigma_2 = \widetilde{\sigma}_2 \cup \widehat{\sigma}_{\varepsilon 2}(x).$$

It is easy to see, that $\sigma_1 \cup \sigma_2 = \Omega$, $\sigma_1 \cap \sigma_2 = \emptyset$. Hence, any distribution $(\widetilde{\sigma}_1, \widetilde{\sigma}_2)$ of the set of ε-common points generates the corresponding disjoint partition (σ_1, σ_2) of the set Ω.

Definition 6.1 *Every such a partition (it depends upon x, ε and, evidently, it is not unique) is called an (x, ε)-proper partition of the set Ω.*

Let us denote by $T_\varepsilon(\Omega, x)$ the family of all (x, ε)-proper partitions of the set Ω. Clearly,

$$T(\Omega, x^*) \subset T_\varepsilon(\Omega, x) \subset T(\Omega).$$

Take $\Gamma = (\gamma_1, \gamma_2) \in T(\Omega)$ and construct the point $x(\Gamma)$ (see (5.1)). By applying the exchange algorithm (described in Section 5), and taking $x(\Gamma)$ as the initial point, in a finite number of steps we get a point $x^*(\Gamma)$ which is a stationary point of the function F.

Definition 6.2 *Let all φ_i be continuous and $\varepsilon > 0$. A point $x^* \in S$ is called an ε-local minimizer of the function F, if*

- *x^* is a stationary point of the function F,*
- *for every $(x^*, \varepsilon)-$proper partition $\Gamma = (\sigma_1, \sigma_2) \in T_\varepsilon(\Omega, x^*)$ of the set Ω for the point $x^*(\Gamma)$ (that is a local minimizer of F, delivered by the exchange algorithm with the initial point $x(\Gamma)$) the inequality*

$$F(x^*(\Gamma)) \geq F(x^*) \qquad (6.1)$$

holds.

Remark 6.1 *Since φ_i's are continuous, every stationary point is a local minimizer, the converse is not true: a local minimizer is not necessarily a stationary point.*

Remark 6.2 *The stationarity of a point is based on the necessary condition (see Rem. 4.1) while the property of being an ε-local minimizer depends also on the exchange algorithm. We will use this notion for the sake of convenience.*

Definition 6.3 *A point $x^* \in S$ is called a strict ε-local minimizer of the function F, if it is an ε-local minimizer of the function F, and, in addition, the inequality (6.1) is strict, i.e.*

$$F(x^*(\Gamma)) > F(x^*) \quad \forall \Gamma \in T_\varepsilon(\Omega, x^*) \setminus T(\Omega, x^*).$$

Note that for $\Gamma \in T(\Omega, x^*)$ the inequality (6.1) becomes the equality.

Let us describe an algorithm for finding ε-local minimizers. Let $\varepsilon > 0$ be fixed.

1. Choose an arbitrary stationary point $x^0 \in S$ (constructed, for example, by the exchange algorithm).

2. Let $x^k = (x_1^k, x_2^k) \in S$ have already been found.

3. Construct the sets $\widehat{\sigma}_{\varepsilon 1}(x^k), \widehat{\sigma}_{\varepsilon 2}(x^k)$ and $\Sigma_\varepsilon(x^k)$.

4. Check the ε-local minimality of x^k (since x^k is a stationary point, it remains to verify the condition (6.1) for all $\Gamma \in T_\varepsilon(\Omega, x^k)$).

   If the point x^k is an ε-local minimizer, then the process terminates. Otherwise, go to step 5.

5. Find any $\Gamma_k \in T_\varepsilon(\Omega, x^k)$ such that

   $$F(x^*(\Gamma_k)) < F(x^k).$$

   Such a Γ_k exists, since x^k is not an ε-local minimizer.

6. Put $x^{k+1} = x^*(\Gamma_k)$ and go to step 3.

As a result, in a finite number of steps an ε-local minimizer is constructed.

Remark 6.3 *If ε is sufficiently large then an ε-local minimizer is a global minimizer of the function F. In this case $T_\varepsilon(\Omega, x) = T(\Omega)$, and hence if ε is large, from the computational point of view, the ε-exchange algorithm is not effective (it is equivalent to the complete enumeration of the elements of the set $T(\Omega)$).*

If ε is fairly small, the above described algorithm allows one to escape from a local minimum point (if the point itself is not yet an ε-local minimizer).

7 AN APPLICATION TO ONE CLUSTERING PROBLEM

Let a set of points $\Omega = \{t_1, \ldots, t_N\} \subset R^n$ be given. Introduce the functions

$$\varphi_1(t, x_1) = \|t - x_1\|^2, \ \varphi_2(t, x_2) = \|t - x_2\|^2, \ x_i \in R^n, \ i = 1, 2,$$

where $\|x\|^2 = \langle x, x \rangle$. Put $x = (x_1, x_2)$,

$$\varphi(t, x) = \min\{\|t - x_1\|^2, \|t - x_2\|^2\}.$$

Note that φ_i's are continuously differentiable and convex. Consider the following clustering problem.

Problem CP: Find $x^* = (x_1^*, x_2^*) \in R^n \times R^n$ such that

$$F(x^*) = \min_{x \in R^n \times R^n} F(x),$$

where

$$F(x) = \sum_{t \in \Omega} \varphi(t, x).$$

Take any $\Gamma = (\sigma_1, \sigma_2) \in T(\Omega, x)$, the functions $c_i(x_i, \sigma_i)$ defined in (2.1) take the form

$$c_i(x_i, \sigma_i) = \sum_{t \in \sigma_i} \|t - x_i\|^2, \ i = 1, 2. \tag{7.1}$$

Clearly,

$$\min_{x_i \in R^n} c_i(x_i, \sigma_i) = c_i(x_i(\sigma_i), \sigma_i), \tag{7.2}$$

where

$$x_i(\sigma_i) = \frac{1}{|\sigma_i|} \sum_{t \in \sigma_i} t. \tag{7.3}$$

Many allocation problems can be described by the above model, maybe with slightly different performance functionals (for example, in **?** the functions $\varphi_i' s$ have the form $\varphi_i(t, x) = \|t - x_i\|$). We have chosen the quadratic functions φ_i (see (7.1)) for the simplicity reasons (since then the auxiliary problems (see (7.2), (7.3)) have the explicit solutions), because our main intention here is to demonstrate the Algorithm.

Remark 7.1 *Let*

$$x_1^* = \frac{1}{N} \sum_{t \in \Omega} t, \ M = \max_{t \in \Omega} \|t - x_1^*\|.$$

Then for every $x_2 \in R^n$ such that $\|x_2 - x_1^*\| > 2M$, the point (x_1^*, x_2) is stationary (that is, a local minimizer as well). Such local minimizers will be called trivial stationary point, and we shall ignore them, looking for better ones.

7.1 Example 1

Let $n = 2$, and Ω be as shown in Table 7.1. The set Ω contains 32 points in R^2, and we have to find two clusters $x_1^* \in R^2$ and $x_2^* \in R^2$, which minimize the functional

$$F(x) = F(x_1, x_2) = \sum_{i \in I} \min\{\|t_i - x_1\|^2, \|t_i - x_2\|^2\},$$

where $I = 1 : 32$. First, let us solve the 1-cluster problem:

Find

$$\min_{x_1 \in R^2} F_1(x_1) = F_1(x_1^*),$$

where

$$F_1(x_1) = \sum_{i \in I} \|t_i - x_1\|^2.$$

It follows from (7.3) that $x_1^* = (-0.5, 2.0656)$ and $F_1(x_1^*) = 782.2722$.

The first and quite natural idea is to take the point $x^0 = (x_1^*, x_1^*)$ as the initial point for our 2-cluster problem. In this case $\Sigma(x^0) = \Omega$ (and $F(x^0) = F_1(x_1^*) = 782.2722$), and at the first step of the exchange algorithm (only to check the conditions (4.1) and (4.2)!!!) it is required to solve 2^{32} elementary problems (which is absolutely unacceptable from practical considerations). However, it is sufficient to slightly perturb x^0 to overcome this difficulty: e.g., take $\overline{x}^0 = (x_1^* + z_1, x_1^*)$, where $z_1 = (0, 0.01)$.

Applying the exchange algorithm with $\overline{x}^0$ as the initial point, in two steps we get the local minimizer $\overline{x} = (\overline{x}_1, \overline{x}_2)$, where

$$\overline{x}_1 = (-0.53846, 5.5),\ \overline{x}_2 = (-0.47368, 0.28421)$$

and $F(\overline{x}) = 523.9929$.

Now take $\overline{\overline{x}}^0 = (x_1^* + z_2, x_1^*)$ where x_1^* is as above, $z_2 = (0.01, 0)$.

Applying the exchange algorithm with $\overline{\overline{x}}^0$ as the initial point, in four steps we get another local minimizer $\overline{\overline{x}} = (\overline{\overline{x}}_1, \overline{\overline{x}}_2)$, where

$$\overline{\overline{x}}_1 = (1.95, 2.98),\ \overline{\overline{x}}_2 = (-4.5833, 0.54167)$$

and $F(\overline{\overline{x}}) = 417.5478$.

The point $\overline{x}^0$ and the corresponding partition (σ_1, σ_2) of the set Ω at $\overline{x}^0$ are shown in Figure 7.2. The notations used in the Figures are explained in Figure 7.1. The points in $\widehat{\sigma}_1(\overline{x}^0)$ are referred to as the points in the first cluster, and the points in $\widehat{\sigma}_2(\overline{x}^0)$ are referred to as the points in the second cluster.

For the initial point $x^0 = (x_1^0, x_2^0) \in R^2 \times R^2$, where $x_1^0 = (-5, 10)$, $x_2^0 = (5, -5)$ (the function value $F(x^0) = 1707.81$), the exchange algorithm terminated in three steps, resulting in the local minimizer $x^* = (x_1^*, x_2^*)$, where

$$x_1^* = (-1.8421, 4.1316),\ x_2^* = (1.4615, -0.9538).$$

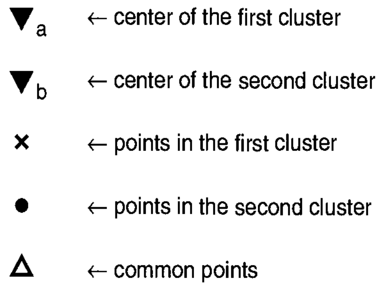

Figure 7.1 Legend.

The function value is $F(x^*) = 498.4104$. The point x^* and the partition (σ_1, σ_2) of the set Ω at x^* are shown in Figure 7.3.

For the initial point $x^0 = (x_1^0, x_2^0)$ with $x_1^0 = (10, 10)$, $x_2^0 = (-10, -10)$ in three steps the local minimizer $x^* = (x_1^*, x_2^*)$, where $x_1^* = (1.9500, 2.9800)$, $x_2^* = (-4.5833, 0.5417)$, is found with $F(x^*) = 417.5478$.

Table 7.1 The set of points $\Omega = \{t_1, \ldots, t_{32}\} \subset R^2$.

i	t_i	i	t_i	i	t_i	i	t_i
1	(3, 1)	*9*	(-5, -2)	*17*	(-1, 6)	*25*	(2, 5)
2	(3, -1)	*10*	(-6, 1)	*18*	(1, 6)	*26*	(2, 4)
3	(5, 1)	*11*	(-6, -1)	*19*	(0, 7)	*27*	(2, -3.4)
4	(-2, 1)	*12*	(2, 0)	*20*	(0, 5)	*28*	(-2, 7)
5	(-2, -1)	*13*	(6, 0)	*21*	(-1, 7)	*29*	(-5, 4)
6	(-3, 2)	*14*	(4, 2)	*22*	(1, 7)	*30*	(-5, 4)
7	(-3, -2)	*15*	(4, -2)	*23*	(2, 6)	*31*	(-6, 2)
8	(-5, 2)	*16*	(5, -1)	*24*	(1, 3)	*32*	(-7, 4.5)

7.2 Example 2

Consider again the problem discussed in Example 1, and apply the ε-exchange algorithm. As the initial point for the ε-exchange algorithm let us choose one of the local minimizers, obtained in Example 1.

The results of numerical experiments for the ε-exchange algorithm with different ε are presented in Table 7.2. The local minimizer $x^0 = (x_1^0, x_2^0)$, where

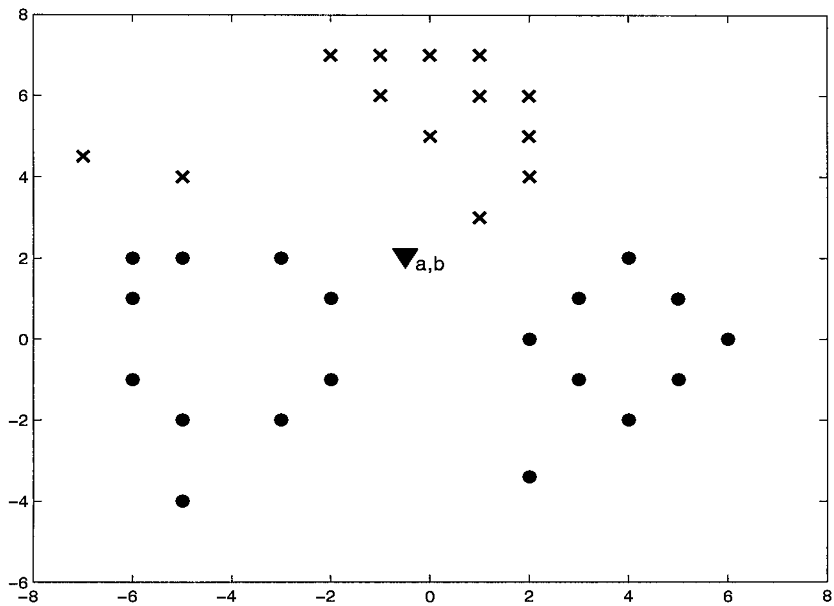

Figure 7.2 The partition of the set Ω at $\overline{x}^0 = (x_1^* + z_1, x_1^*)$.

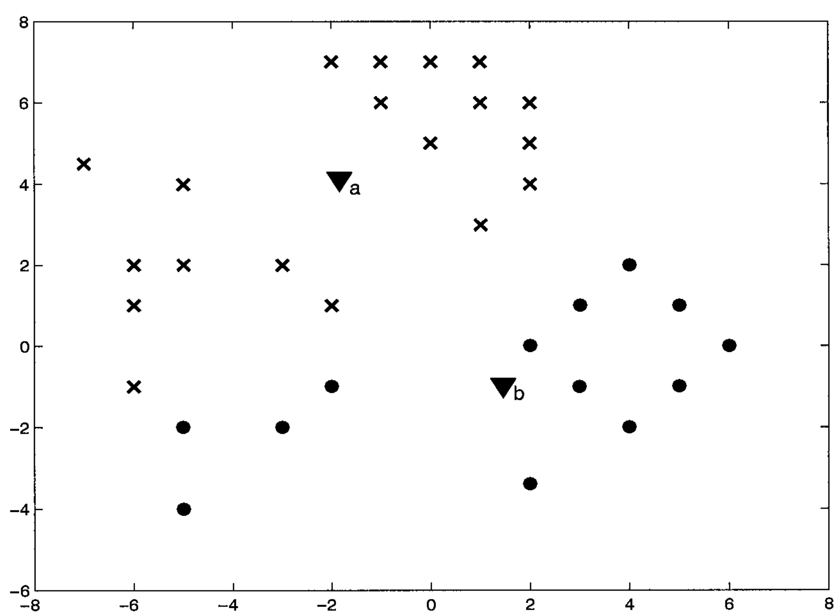

Figure 7.3 Results of the exchange algorithm.

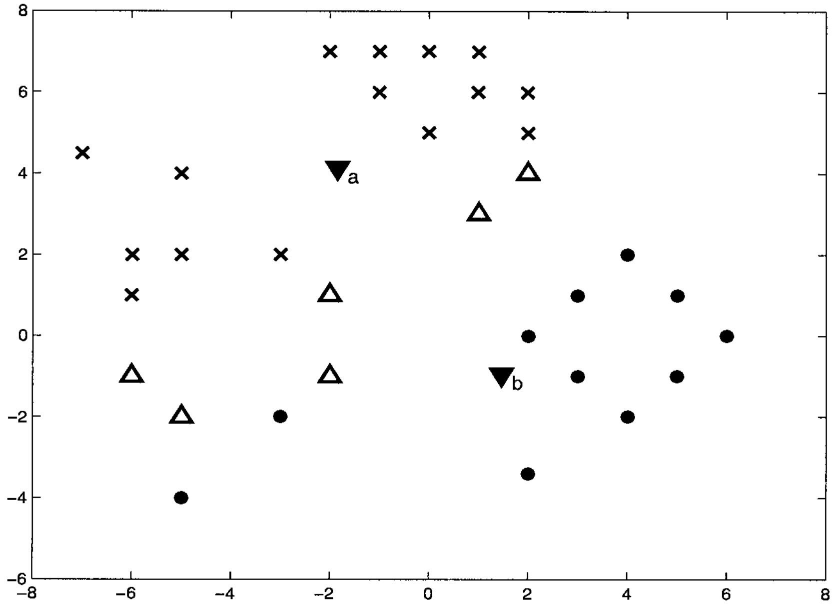

Figure 7.4 The first step of the ε-exchange algorithm with $\varepsilon = 15$ (six ε-common points).

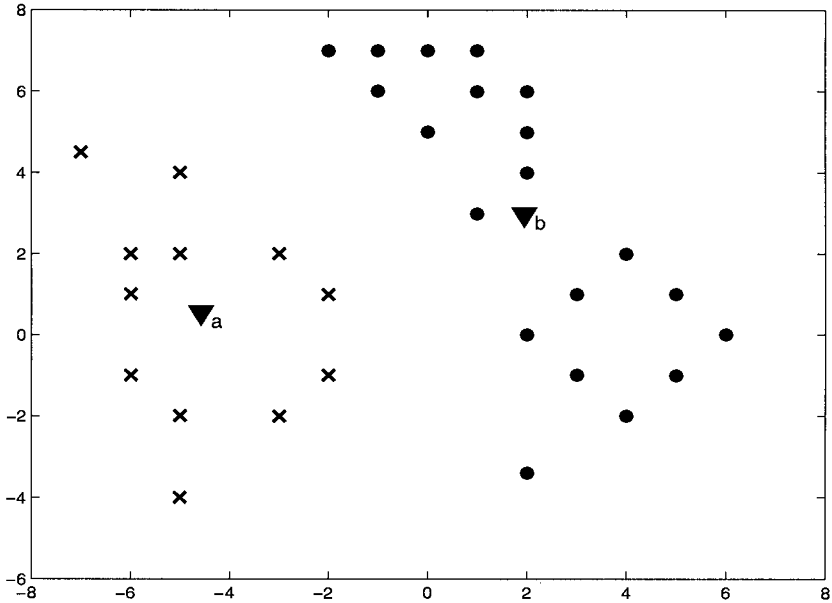

Figure 7.5 An ε-local minimizer with $\varepsilon = 15$ for the initial point $x^0 = (x_1^0, x_2^0)$, $x_1^0 = (-1.8421, 4.1316)$, $x_2^0 = (1.4615, -0.95385), F(x^*) = 417.5478$.

$x_1^0 = (-1.8421, 4.1316)$, $x_2^0 = (1.4615, -0.95385)$, was taken as the initial point for these computations. In Figure 7.4 the sets $\widehat{\sigma}_{\varepsilon 1}(x^0)$, $\widehat{\sigma}_{\varepsilon 2}(x^0)$, $\Sigma_\varepsilon(x^0)$ are depicted (they are denoted by $\times$, $\bullet$, $\triangle$, respectively) for $\varepsilon = 15$.

In Figure 7.5 the results of application of the ε-exchange algorithm are shown (for ε=15 and the initial point x^0). As a result of the computations we received four local minimizers:

- $x^{*1} = (x_1^{*1}, x_2^{*1})$, where

  $$x_1^{*1} = (-1.8421, 4.1316),\ x_2^{*1} = (1.4615, -0.95385),\ F(x^{*1}) = 498.4104.$$

  It is an ε-local minimizer for ε up to 4.

- $x^{*2} = (x_1^{*2}, x_2^{*2})$, where

  $$x_1^{*2} = (-2.0000, 3.825),\ x_2^{*2} = (2.0000, -0.8667),\ F(x^{*2}) = 497.1842.$$

  This point is an ε-local minimizer for ε up to 8.

- $x^{*3} = (x_1^{*3}, x_2^{*3})$, where

  $$x_1^{*3} = (-2.1739, 3.0217),\ x_2^{*3} = (3.7778, -0.3778),\ F(x^{*3}) = 478.3746.$$

  The point x^{*3} is an ε-local minimizer for ε up to 10.

- $x^{*4} = (x_1^{*4}, x_2^{*4})$, where

  $$x_1^{*4} = (-4.5833, 0.5417),\ x_2^{*4} = (1.9500, 2.9800),\ F(x^{*4}) = 417.5478.$$

  The point x^{*4} is an ε-local minimizer for ε up to 30.

Remark 7.2 *We have computed only integer values of ε up to 30. It seems that the point x^{*4} is a global minimizer of F. It is interesting to note that for $\varepsilon = 9$ at the first step the set of common points was the same as for $\varepsilon = 8$. However, at further steps the increase of ε affected the result, since due to deeper "ε-diving" we were able to pick up a better ε-minimizer.*

Remark 7.3 *Numerical experiments with several real databases demonstrated the effectiveness of the exchange algorithm.*

8 CONCLUSIONS

Thus, we have described two algorithms: the exchange algorithm for constructing a stationary point and the ε–exchange algorithm for finding, may be, a better minimizer. The ε–exchange algorithm allows one to "escape" from a local minimum. These algorithms are conceptual (in the terminology of E.Polak) (see Polak (1971)), though in some cases (as is demonstrated in section 7) they are directly applicable.

It may happen, that the number of common (or ε-common points) is large. In such a case it is useful to perform some preliminary aggregation of these points reducing their number to a reasonable quantity (as well as to reduce or increase the value of ε). The aggregation idea was proposed by Prof. M. Gaudioso.

Computationally implementable modifications of the above algorithms for specific classes of functions will be reported elsewhere.

At each step of the both algorithms an elementary problem of minimizing a function of the form $F_\Gamma(x)$ is to be solved. The algorithms converge in a finite number of steps to, at least, a local minimizer. If ε is sufficiently large, the ε–exchange algorithm will produce a global minimizer, however, theoretically it may require the complete enumeration (and solution) of all elementary problems. The hope and expectation are that if we take a reasonable ε, then (at least statistically) the price asked for a fairly good local minimizer will not be too high.

The case $m > 2$ can be studied in a similar way. Analogous results and algorithms can be formulated. The number of "elementary" problems becomes $m^{|\Omega|}$ (cf. Rem. 3.1) and, of course, all calculations are more complicated. However, the exchange and ε-exchange algorithms can be constructed.

Acknowledgments

The author is thankful to an anonymous referee for his careful reading of the manuscript, useful advice, remarks and suggestions.

References

Bagirov, A.M., Rubinov, A.M. (1999), Global Optimization of IPH functions over the unit simplex with applications to cluster analysis, In *Proceedings of* *the Third Australia-Japan Joint Workshop on Intelligent and Evolutionary Systems*, Canberra, Australia, pp. 191-192.

Bagirov, A.M., Rubinov, A.M. and Yearwood J. (2001), Using global optimization to improve classification for medical diagnosis, *Topics in Health Information Management*, Vol. 22, pp. 65-74.

Mangasarian, O.L. (1997), Mathematical Programming in Data Mining, *Data Mining and Knowledge Discovery*, Vol. 1, pp. 183-201.

Polak E. (1971), *Computational Methods in Optimization; A unified approach*, Academic Press, New York, NY.

Rao, M.R. (1971), Cluster analysis and mathematical programming, *Journal of the American Statistical Association*, Vol. 66, pp. 622-626.

Rubinov, A.M. (2000), *Abstract Convexity and Global Optimization*, Kluwer Academic Publishers, Dordrecht.

Vershik, A.M., Malozemov, V.N., Pevnyi, A.B. (1975), The best piecewise polynomial approximation, *Siberian Mathematical Journal*, Vol. XVI, Nr. 5, pp. 925-938.