On the Global Optimization Properties of Finite-Difference Local Descent Algorithms*

S. K. ZAVRIEV
Moscow State University, Faculty of Computational Mathematics and Cybernetics, S. U. 119899, Moscow, Russia.

(Received: 14 December 1990; accepted: 17 December 1991)

Abstract. The paper is devoted to the convergence properties of finite-difference local descent algorithms in global optimization problems with a special γ-convex structure. It is assumed that the objective function can be closely approximated by some smooth convex function. Stability properties of the perturbed gradient descent and coordinate descent methods are investigated. Basing on this results some global optimization properties of finite-difference local descent algorithms, in particular, coordinate descent method, are discovered. These properties are not inherent in methods using exact gradients.

Key words. Gradient method, coordinate descent method, global optimization, stability under perturbations.

1. Introduction

Consider the unconstrained multiextremal optimization problem

$$F(x) \rightarrow \inf_{x \in E_k} \tag{1}$$

where E_k is a k-dimensional Euclidean space.

DEFINITION 1.1. The function $\phi(\cdot)$ is called strongly convex on E_k with parameter $l > 0$ if

$$\phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2) - l\lambda(1-\lambda)\|x_1 - x_2\|^2/2$$

for any $x_1, x_2 \in E_k$ and λ, $0 \leq \lambda \leq 1$. □

Problems of the type (1) sometimes possess the following important structural property.

DEFINITION 1.2. The problem (1) is said to be γ-convex structured if there exists a strongly convex (with parameter $l > 0$) and differentiable function $\phi(\cdot)$, $\nabla\phi(\cdot) \in \mathscr{L}(E_k, L)$, such that

$$|\phi(x) - F(x)| \leq \gamma \quad \forall x \in E_k$$

*The paper was presented at the II. IIASA-Workshop on Global Optimization, Sopron (Hungary), December 9–14, 1990.

Journal of Global Optimization **3**: 67–78, 1993.

(see Figure 1). $\mathscr{L}(X, L)$ denotes the class of Lipschitz continuous functions on $X \subset E_k$ with constant $L > 0$. □

We may say that $\phi(\cdot)$ is a convex γ-skeleton of the objective function $F(\cdot)$. Note that $F(\cdot)$ may be multiextremal and nonsmooth and that a suitable skeleton function $\phi(\cdot)$ is usually unknown.

REMARK 1.3. Global optimization problems with the similar structural properties were considered in [1]. □

If γ is small, then the γ-convex problem (1) is well approximated by the smooth convex problem

$$\phi(x) \to \min_{x \in E_k}, \tag{2}$$

where $\phi(\cdot)$ is continuously differentiable on E_k, $\nabla\phi(\cdot) \in \mathscr{L}(E_k, L)$ and the sets

$$X_f = \{x \in E_k | \phi(x) \leq f\}, f \in R_1,$$

are compact.

Seemingly straightforward approach to solving γ-convex structured problems (1) is as follows: (i) to obtain a suitable unimodal differentiable approximation $\phi(\cdot)$ of the objective function $F(\cdot)$, for instance, using the formula

$$\phi(x) = \phi(x, \alpha) = \int_{-\alpha}^{\alpha} \dots \int_{-\alpha}^{\alpha} F(x + y)\, dy_1 \dots dy_k \tag{3}$$

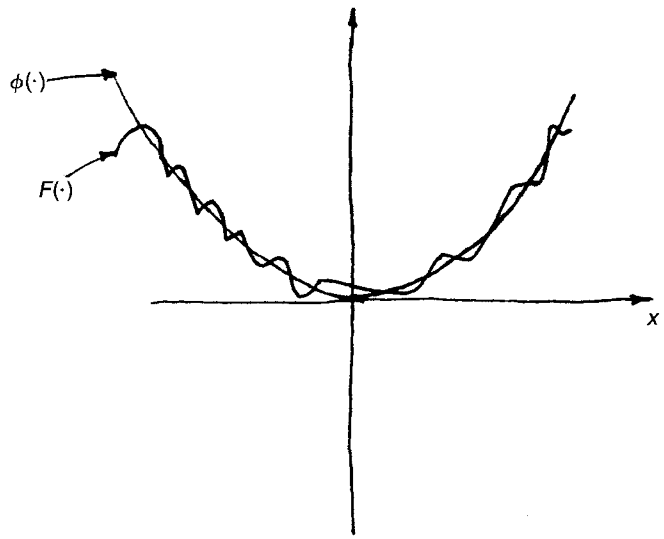

Fig. 1.

(see [2, 3]); (ii) to solve the problem (2), (3) applying some effective descent algorithm.

Though performing this approach we face great difficulties, i.e., if $F(\cdot)$ is locally Lipschitz continuous, then $\phi(\cdot, \alpha)$ is actually continuously differentiable and for any compact X

$$\lim_{\alpha \to 0} \sup_{x \in X} |\phi(x, \alpha) - F(x)| = 0$$

but unimodality (and moreover convexity) of $\phi(\cdot, \alpha)$ in γ-convex structured problems is not studied and seems to be doubtful. Furthermore, Problems (2), (3) are complicated even when $\phi(\cdot, \alpha)$ is convex, because values of $\phi(\cdot, \alpha)$ are difficult to compute. Certainly, stochastic programming algorithms (see [3]) can be used for solving (2), (3), but the efficiency of these algorithms is not high.

In the present paper another approach to approximate solving of γ-convex structured problems (1) is proposed. Its mainpoints are as follows:

(i) Different local descent methods, for instance, gradient descent method and coordinate descent method can be used for solving (2).

(ii) Some of these algorithms have good properties of stability under perturbations in values of the objective function $\phi(\cdot)$, i.e., using in these algorithms perturbed values $\tilde{\phi}(x)$ instead of exact values $\phi(x)$ we shall obtain approximate solutions of the problem (2).

(iii) Thus, using in these algorithms values $F(x)$ instead of values $\phi(x)$ we shall obtain approximate solutions of the problem (2) and, therefore, approximate solutions of the initial problem (1).

In Section 2 stability properties of the gradient descent method using inexact gradients are studied, sharp estimations of attractors are obtained. Stability under perturbations of the gradient descent method has been studied before in [3–8]. A finite-difference analogue of the gradient method is considered in Section 3. In γ-convex structured problems this algorithm possesses some global optimization properties, which are not inherent in the gradient method using exact gradients. Some special abilities of descent algorithms, using finite-differences, were considered in [3, 9].

In Section 4 stability under perturbations is explored for the coordinate descent method. Due to its stability properties this method may be used in γ-convex problems for the approximate search of the global optimum (Section 5). Moreover it is shown that the coordinate descent method has some advantages in comparison with the simplest finite-difference version of the gradient descent method.

2. Stability of the Perturbed Gradient Descent Method

Consider the unconstrained optimization problem:

$$\phi(x) \to \min_{x \in E_k} , \tag{2}$$

where $\phi(\cdot)$ is continuously differentiable on E_k, $\nabla\phi(\cdot) \in \mathscr{L}(X', L)$, X' is an open set,

$$X' \supset X = \{x \in E_k | \phi(x) \leq f\} ,$$

$X \neq \varnothing$ is compact, $f \in R_1$.

The simplest algorithm of the perturbed gradient descent method is as follows:

$$\begin{aligned} &x^{n+1} = x^n + a_n(\nabla\phi(x^n) + \Delta_1(x^n)) , \\ &n = 1, 2, \dots, x^1 \in X^1 , \end{aligned} \tag{4}$$

here $\Delta_1(x) \in E_k$ is a perturbation of the gradient $\nabla\phi(x)$; the stepsizes $\{a_n\}$ satisfy

$$0 < a_0 \leq a_n \leq \bar{a}_0 < 1/L , \quad n = 1, 2, \dots ; \tag{5}$$

x^1 is an element of the set of starting points X^1.

ASSUMPTION 2.1. *Let the perturbation $\Delta_1(\cdot)$ satisfy the estimation*

$$\|\Delta_i(x)\| \leq \varepsilon_1 \quad \forall x \in \chi$$

with some constant level $\varepsilon_1 \geq 0$. □

Set

$$\phi^* = \min_{x \in E_k} \phi(x) = \min_{x \in X} \phi(x);$$

$$X_0(\varepsilon) = \{x \in E_k | \phi(x) \leq \phi^* + \varepsilon\} ,$$

$$X_S(\varepsilon) = \{x \in X | \|\nabla\phi(x)\| \leq \varepsilon\} ,$$

$$l_S(\varepsilon) = \sup_{x \in X_S(\varepsilon)} \phi(x) , \quad \varepsilon \geq 0 .$$

It is obvious that

$$l_S(\varepsilon') \leq l_S(\varepsilon'') \quad \forall \varepsilon', \varepsilon'', \quad 0 \leq \varepsilon' \leq \varepsilon'' .$$

LEMMA 2.2. (1). *If $\phi(\cdot)$ is convex on E_k, then*

$$l_S(0) = \phi^*$$

(2). *If $\phi(\cdot)$ is strongly convex on E_k with parameter $l > 0$, then*

$$l_S(\varepsilon) \leq \phi^* + \varepsilon^2/2l ,$$

$$X_S(\varepsilon) \subset X_0(\varepsilon^2/2l) \quad \forall \varepsilon \geq 0 .$$

Proof. Fix arbitrary $\varepsilon \geq 0$and $x \in X_S(\varepsilon)$.
Applying the following estimation

$$2l(\phi(x) - \phi^*) \leq \|\nabla\phi(x)\|^2$$

(see Lemma 1.4.3 [5]), we obtain

$$\phi(x) \leq \phi^* + \varepsilon^2/2l ;$$

$$x \in X_0(\varepsilon^2/2l).$$

The lemma is proved. □

ASSUMPTION 2.3. *Let*

$$X^1 = X \supset X_S(\varepsilon_1) . \qquad \square$$

We shall observe the behaviour of trajectories of the algorithm (4), (5) with a fixed level ε_1 of perturbations.

LEMMA 2.4. *Under Assumptions* 2.1, 2.3 *every trajectory* $\{x^n\}$ *of* (4), (5) *satisfies*

$$x^n \in X , \quad n = 1, 2, \ldots \qquad \square$$

Applying Lemma 2.4 and the convergence analysis technique developed in [6], it is easy to prove the following

THEOREM 2.5. *Let Assumptions* 2.1, 2.3 *be fulfilled.*

(1) *Then every trajectory* $\{x^n\}$ *of the perturbed gradient descent algorithm* (4), (5) *satisfies*

$$\limsup_{n\to\infty} \phi(x^n) \leq l_S(\varepsilon_1) . \tag{6}$$

(2) *If* $\phi(\cdot)$ *is strongly convex on* E_k *with parameter* $l > 0$ *then there exist* C, $q > 0$, $0 < q < 1$, *such that every trajectory* $\{x^n\}$ *of* (4), (5) *satisfies*

$$\overline{lt}\{x^n\} \subset X_0\ (\varepsilon_1^2/2l) ,$$

$$\phi(x^n) \leq \phi^* + \varepsilon_1^2/2l + Cq^n , \quad n = 1, 2, \ldots$$

($\overline{lt}\{x^n\}$ *denotes the set of all limit points due to subsequences* $\{x^n\}$). □

REMARK 2.6. It is easy to see that the estimation (6) is sharp for any problem (2) on the considered class of perturbations, i.e., for any $\phi(\cdot)$ there exist a starting point x^1 and a perturbation $\Delta_1(\cdot)$, $\|\Delta_1(x)\| \leq \varepsilon_1$ $\forall x \in X$, such that the corresponding trajectory $\{\bar{x}^n\}$, $\bar{x}^1 = x^1$, of (4), (5) satisfies

$$\phi(\bar{x}^n) = l_S(\varepsilon_1) , \quad n = 1, 2, \ldots \qquad \square$$

REMARK 2.7. Theorem 2.5 generalizes the corresponding results in [4, 5].

3. Global Optimization Properties of the Simplest Finite-Difference Version of the Gradient Descent Method

Let us consider the γ-convex structured problem (1). Set

$$g(x, h) = (g_1(x, h), \ldots, g_k(x, h)),$$
$$g_i(x, h) = (F(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_k) - F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k))/h,$$
$$i = 1, 2, \ldots, k,\ h > 0,\ x \in E_k.$$

The simplest finite-difference analogue of the gradient descent method is as follows:

FDG algorithm (FDGA)

$$x^{n+1} = x^n - a_n g(x^n, h), \quad n = 1, 2, \ldots,$$
$$x^1 \in E_k,$$

where for the parameter h it holds $h > 0$, the stepsizes $\{a_n\}$ satisfy (4) and $x^1 \in E_k$ is a starting point.

LEMMA 3.1. *For every $h > 0$ and $x \in E_k$ $\|g(x, h) - \nabla\phi(x)\| \leq k^{1/2}(Lh/2 + 2\gamma/h)$.* □

THEOREM 3.2. *Let the problem* (1) *be γ-convex structured. Then every trajectory $\{x^n\}$ of FDGA satisfies*

$$\overline{lt}\{x^n\} \subset \{x \in E_k | F(x) \leq \inf_{x' \in E_k} F(x') + k(Lh/2 + 2\gamma/h)^2/2l + 2\gamma\}.$$

In particular, when

$$h = 2\sqrt{\gamma/L}$$

we get

$$\overline{lt}\{x^n\} \subset \{x \in E_k | F(x) \leq \inf_{x' \in E_k} F(x') + 2\gamma(k(L/l) + 1)\}.$$

Proof. Fix an arbitrary trajectory $\{x^n\}$ of FDGA. Applying Lemma 3.1 and Theorem 2.5 we obtain immediately

$$\overline{lt}\{x^n\} \subset \{x \in E_k | \phi(x) \leq \phi^* + k(Lh/2 + 2\gamma/h)^2/2l\}.$$

It is obvious that

$$|\inf_{x \in E_k} F(x) - \phi^*| \leq \gamma;$$

therefore

$$\{x \in E_k | \phi(x) \leq \phi^* + \varepsilon\} \subset \{x \in E_k | F(x) \leq \inf_{x' \in E_k} F(x') + \varepsilon + 2\gamma\}$$

for any $\varepsilon \geq 0$.

Thus,

$$\overline{lt}\{x^n\} \subset \{x \in E_k | F(x) \leq \inf_{x' \in E_k} F(x') + 2\gamma + k(Lh/2 + 2\gamma/h)^2/2l\}.$$

The theorem is established. □

Thus, we can see that when the parameter γ is sufficiently small and condition L/l of the problem (2) is not large it is possible to search approximate global solutions of the multiextremal problem (1) using the finite-difference algorithm FDGA with the appropriate parameter h,

$$h \approx 2\sqrt{\gamma/L}\,.$$

The search of the appropriate h may be performed by some adaptive rules.

4. Stability of the Perturbed Coordinate Descent Method

The coordinate descent algorithm [10] can be considered as some finite-difference analogon of the gradient method. Thus, we shall show that this algorithm when used in γ-convex problems possesses some global optimization properties.

Let us study convergence properties of the perturbed coordinate descent method.

Consider the optimization problem (2), where the objective function $\phi(\cdot)$ is assumed to be continuously differentiable on E_k, $\nabla\phi(\cdot) \in \mathscr{L}(X', L)$, $X' \supset X = \{x \in E_k | \phi(x) \leq f\}$ is an open set, $X \neq \varnothing$ is compact, $f \in R_1$.

The simplest algorithm of the perturbed coordinate descent method is as follows:

PCD ALGORITHM (PCDA)

$$x^{n,1} = x^n$$

$$x^{n,i+1} = \begin{cases} x^{n,i} + h_n e_i\,, \text{ if} \\ \phi(x^{n,i} + h_n e_i) + \Delta_0(x^{n,1} + h_n e_i) < \phi(x^{n,i}) + \Delta_0(x^{n,i})\,; \\ \text{else } x^{n,i} - h_n e_i\,, \text{ if} \\ \phi(x^{n,i} - h_n e_i) + \Delta_0(x^{n,i} - h_n e_i) < \phi(x^{n,i}) + \Delta_0(x^{n,i})\,; \\ \text{else } x^{n,i}\,; \end{cases}$$

$$i = 1, \ldots, k\,;$$

$$x^{n+1} = x^{n,k+1}\,;$$

$$h_{n+1} = \begin{cases} h_n\,, & x^{n+1} \neq x^n\,; \\ \theta h_n\,, & x^{n+1} = x^n\,; \end{cases}$$

$$n = 1, 2, \ldots, \quad x^1 \in X^1\,,$$

where $\Delta_0(x) \in R_1$ is a perturbation of $\phi(x)$; θ, $0 < \theta < 1$, is an adapting parameter; and $e_1, \ldots, e_k$ are coordinate vectors,

$$\langle e_i, e_j \rangle = \begin{cases} 0, & i \neq j; \\ 1, & i = j; \end{cases}$$

$$i, j = 1, \ldots, k; \quad h_1 > 0;$$

$x^1 \in X^1$ is an element of the set of starting points X^1.

ASSUMPTION 4.1. *Let a perturbation $\Delta_0(\cdot)$ satisfy the estimation*

$$|\Delta_0(x)| \leqslant \varepsilon_0 \forall x \in X,$$

with some constant level $\varepsilon_0 \geqslant 0$. □

ASSUMPTION 4.2. *Let*

$$X^1 \subset \{x \in E_k | \Phi(x) \leqslant f - 2\varepsilon_0\}.$$ □

It is obvious that $X^1 \subset X \subset X'$.

We shall observe the behaviour of trajectories of PCDA with a fixed level ε_0 of perturbation.

LEMMA 4.3 *Under Assumptions 4.1, 4.2 every trajectory of PCDA satisfies the estimations*

$$\limsup_{n\to\infty} \phi(x^n) \leqslant \phi(x^N) + 2\varepsilon_0,$$

$$\limsup_{n\to\infty} (\phi(x^n) + \Delta_0(x^n)) \leqslant \phi(x^N) + \varepsilon_0, \; N = 1, 2, \ldots$$ □

COROLLARY 4.4. *Under Assumptions 4.1, 4.2 every trajectory $\{x^n\}$ of PCDA satisfies*

$$x^n \in X, \quad n = 1, 2, \ldots$$ □

Let us fix an arbitrary trajectory $\{x^n\}$ of PCDA and set

$$\mathcal{N} = \{n \in \mathbb{N} | x^n = x^{n+1}\} = \{N_1, N_2, \ldots\},$$

where $N_i < N_{i+1}$, $i = 1, 2, \ldots$, $\mathbb{N} = \{1, 2, \ldots\}$.

It is easy to see that

$$h_{n+1} = \begin{cases} h_n, & n \notin \mathbb{N}; \\ \theta h_n, & n \in \mathbb{N}, \end{cases}$$

$$n = 1, 2, \ldots;$$

therefore $h_{N_i} = \theta^{i-1} h_1$, $i = 1, 2, \ldots$

LEMMA 4.5. *The set $\mathbb{N}$ is infinite, $|\mathbb{N}| = \infty$($|A|$ denotes the number of elements of the set A).* □

COROLLARY 4.6. $\lim_{n\to\infty} h_n = 0$. □

LEMMA 4.7. *For every* $i = 1, 2, \ldots$ *the inequality*

$$\|\nabla\phi(x^{N_i})\| \leqslant \sqrt{k}\left(\frac{2\varepsilon_0}{\theta^{i-1}h_1} + \frac{L}{2}\,\theta^{i-1}h_1\right)$$

is valid. □

LEMMA 4.8. *If* $h_1 > 2\sqrt{\varepsilon_0/L}$ *then*

$$\min_{i\in\mathbb{N}}\left(\frac{2\varepsilon_0}{\theta^{i-1}h_1} + \frac{L}{2}\,\theta^{i-1}h_1\right) \leqslant \sqrt{\varepsilon_0 L}\left(\sqrt{\theta} + \frac{1}{\sqrt{\theta}}\right).$$ □

Applying Lemmas 4.3, 4.5, 4.7, 4.8, we arrive at

THEOREM 4.9. *Let Assumptions 4.1, 4.2 hold and the parameter* h_1 *fulfil*

$$h_1 > 2\sqrt{\varepsilon_0/L}\,.$$

(1) *Then every trajectory* $\{x^n\}$ *of PCDS satisfies*

$$\limsup_{n\to\infty} \phi(x^n) \leqslant l_S\left(\sqrt{k\varepsilon_0 L}\left(\sqrt{\theta} + \frac{1}{\sqrt{\theta}}\right)\right) + 2\varepsilon_0\,.$$

(2) *If* $\phi(\cdot)$ *is strongly convex on* E_k *with parameter* $l > 0$ *then every trajectory* $\{x^n\}$ *of PCDA satisfies*

$$\overline{lt}\{x^n\} \subset X_0\left(k\varepsilon_0 L\left(\sqrt{\theta} + \frac{1}{\sqrt{\theta}}\right)^2/2l + 2\varepsilon_0\right).$$ □

5. Global Optimization Properties of the Coordinate Descent Method

Turning back to the γ-convex structured problem (1) let us consider the coordinate descent algorithm:

CD ALGORITHM (CDA)

$$x^{n,1} = x^n\,,$$

$$x^{n,i+1} = \begin{cases} x^{n,i} + h_n e_i\,, \text{ if} \\ F(x^{n,i} + h_n e_i) < F(x^{n,i})\,; \\ \text{else } x^{n,i} - h_n e_i\,, \text{ if} \\ F(x^{n,i} - h_n e_i) < F(x^{n,i})\,; \\ \text{else} \\ \quad x^{n,i}\,; \end{cases}$$

$$i = 1, \dots, k\,; \quad x^{n+1} = x^{n,k+1}\,;$$

$$h_{n+1} = \begin{cases} h_n\,, & x^{n+1} \neq x^n\,; \\ \theta h_n\,, & x^{n+1} = x^n\,; \end{cases}$$

$$n = 1, 2, \dots, \quad x^1 \in E_k\,,$$

where all denotations are the same as in Section 4.

Applying the results of Section 4 we obtain

THEOREM 5.1. *Let the problem* (1) *be* γ*-convex structured. Then every trajectory* $\{x^n\}$ *of CDA, with the parameter* $h_1 > 2\sqrt{\gamma/L}$, *satisfies*

$$\overline{lt}\{x^n\} \subset \Big\{ x \in E_k \,|\, F(x) \leq \inf_{x' \in E_k} F(x') + 2\gamma\Big(k(L/l)\Big(\Big(\sqrt{\theta} + \frac{1}{\sqrt{\theta}}\Big)\Big/2\Big)^2 + 1\Big)\Big\}\,. \qquad \square$$

Comparing the algorithms FDGA and CDA, we see that the advantage of CDA is that this algorithm does not need the special selection of the appropriate value of finite-difference parameter h. We may say that the selection of the appropriate h is an interior property of the coordinate descent method. On the other side the convergence rate of the coordinate descent method is less than the rate of the gradient method.

Analysing Theorem 5.1 it is easy to see that the more is the value of the adapting parameter θ, $0 < \theta < 1$, in the CDA – the more is a precision of solving (1) by the use of CDA. But it is obvious that as θ is tending to 1 the convergence rate of CD is decreasing.

6. Conclusion

A special γ-convex structure of the optimization problem (1) can be essentially used in the construction of global optimization procedures. In these problems on the first stage of global optimization we may apply finite-difference local descent algorithms, in particular, the coordinate descent method. Certainly, for the further improvement of the obtained approximate solution we have to use some other global optimization algorithm, for instance, of "covering" type (see [11, 12]), but the area searched (covered) by this algorithm may be much reduced in comparison with the initial one.

Finite-difference local descent algorithms can be applied in global optimization procedures for solving problems with more complicated structure than the γ-convex one. We give some definitions to discuss this abilities of finite-difference local descent algorithms.

DEFINITION 6.1. Let the function $\phi(\cdot)$ be continuously differentiable on E_k and $\nabla\phi(\cdot)$ be locally Lipschitz continuous, $\phi(\cdot)$ is said to be nonsingular if the stationary set

$$X_S = \{x \in E_k | \nabla\phi(x) = 0\}$$

is bounded and

$$J\nabla\phi(x) \text{ is nonsingular for every } x \in X_S ,$$

where $J\nabla\phi(x)$ denotes the generalized Jacobian of $\nabla\phi(\cdot): E_k \rightarrow E_k$ in $x \in E_k$. □

It is easily seen that the stationary set X_S of the nonsingular function $\phi(\cdot)$ consists of a finite number of elements, i.e.

$$|\phi(X_S)| \leq |X_S| < \infty ,$$

$$\phi(\chi_S) = \{f \in R_1 | f = \phi(x) , x \in X_S\} .$$

DEFINITION 6.2. The problem (1) is said to be (γ, M)-nonsingularly structured, $\gamma \geq 0$, $M \in \mathbb{N}$, if there exists a nonsingular $\phi(\cdot)$ such that

$$|\phi(x) - F(x)| \leq \gamma \quad \forall x \in E_k$$

and

$$|\phi(X_S)| = M$$

(see Figure 2). □

For solving global optimization problems of the type (1) with sufficiently smooth $F(\cdot)$ the multistart method is commonly used [5]. Analysing Theorems 3.3 and 5.1, it is easy to see that if in a (γ, M)-nonsingularly structured problem (1) the parameters γ and M are sufficiently small (the set $\phi(X_S)$ is rarefied) then the finite-difference local descent algorithms, for instance, CD algorithm, are more preferable for applying in multistart method than local descent algorithms using exact gradients. Actually, if M and γ are small, then, for instance, the trajectories $\{x^n\}$ of CDA have the property that

$$\overline{lt}\{F(x^n)\} \subset B_\varepsilon(\phi(X_S))$$

for some small $\varepsilon > 0$ ($B_\varepsilon(Y)$ denote $\{x | \exists y \in Y: \|x - y\| < \varepsilon\}$). And therefore, for most x^1 the difference

$$F(x^1) - \lim_{n\rightarrow\infty} F(x^n)$$

($F(\cdot)$-improvement of x^1 by CDA) is sufficiently large. The similar characteristics of the exact gradient algorithm can be essentially smaller because at a number of local minima in problem (1).

Certainly, finite-difference local descent algorithms do not pretend to supply some new general method of global optimization. But we suppose that these

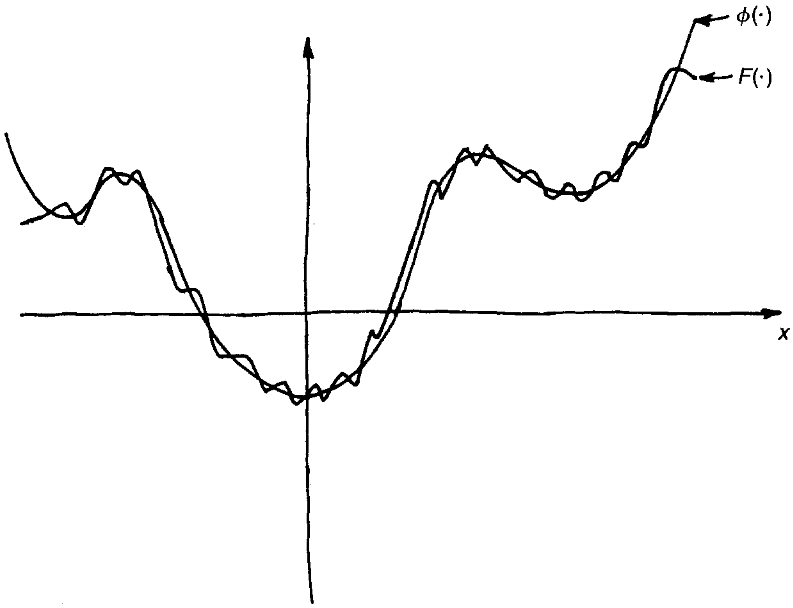

Fig. 2.

algorithms can be fruitfully used in global optimization procedures as some other supplementary instrument.

References

1. Hu, T. C., V. Klee and D. Larman (1989), Optimization of Globally Convex Functions, *SIAM J. Control and Optimization* **27** (5), 1026–1047.
2. Mayne, D. Q. and E. Polak (1984), Nondifferential Optimization via Adaptive Smoothing, *JOTA* **43** (4), 601–614.
3. Mikhalevitch, V. S., A. M. Gupal and V. I Norkin (1987), *Methods of Nonconvex Optimization*, Nauka, Moscow (in Russian).
4. Boggs, P. T. and J. E. Dennis, Jr. (1976), A Stability Analysis for Perturbed Nonlinear Iterative Methods, *Mathematics of Computation* **30** (134), 199–215.
5. Polyak, B. T. (1987), *Introduction to Optimization*. Optimization Softwear, Inc., NY.
6. Zavriev, S. K. (1984), *Stochastic Sub-Gradients Methods for Minmax Problems*. Izd. MGU, Moscow, (1984) (in Russian).
7. Dorofeev, P. A. (1985), On Some Properties of Quasigradient Method, *USSR Computational Mathematics and Mathematical Physics* **25**, 181–189.
8. Zavriev, S. K. (1990), Convergence Properties of the Gradient Method in Presence of Perturbations of a Variable Level, *USSR Computational Mathematics and Mathematical Physics* **30**, 997–1007.
9. Norkin, V. I. (1979), On the Convergence Conditions for Combined Nonlinear Programming Algorithms, *Kibernetica* **6**, 73–77.
10. Luenberger, D. G. (1984), *Linear and Nonlinear Programming*, Addison-Wesley, London.
11. Evtushenko, U. G. (1971), Numerical Method for the Search of Global Extremum of Function, *USSR Computational Mathematics and Mathematical Physics* **11**, 1390–1403.
12. Horst, R. (1986), A General Class of Branch-and-Bound Methods in Global Optimization with Some New Approaches for Concave Minimization, *Journal of Optimization Theory and Applications* **51**, 271–291.