Automation and Remote Control, Vol. 55, No. 11, 1994

ANALYSIS OF SEARCH METHODS OF OPTIMIZATION BASED ON POTENTIAL THEORY III. CONVERGENCE OF METHODS

A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi

UDC 517.977.5

The convergence of trajectories of a random process used for the optimization of a nonsmooth multiextremal function is studied. The potential of the Newtonian vector field generated by the objective function is used as a Liapunov function of this random process. The convergence of the process is proved in the sense of the strong convergence, with preliminary investigation of other types of convergence.

1. INTRODUCTION

We consider the problem of minimizing a nonsmooth function $f : R^n \to R$. This problem is replaced by a randomized problem

$$F(X) = E[f(X)] \to \min,$$

where X is a random vector with values in R^n, and E is the symbol of mathematical expectation. To solve the problem we use a random process

$$X^{N+1} = X^N + \alpha_N Y^{N+1}, \quad E[\,\|X^0\|\,] < \infty, \tag{1}$$

where $\alpha_N > 0$, $\|X\|$ is the Euclidean norm of X, $N = 0, 1, 2, \ldots$.

The random process (1) is intended for minimizing nonsmooth and multiextremal functions. Therefore, it should have both the property of leaving of domains of (nonessential) extrema of the function f by the trajectories of (1) and the property of convergence of these trajectories to a domain of a (local) minimum of f.

In [1, 2] we studied the problems of leaving of domains of local extrema of the objective function $f(x)$. In this work we consider questions of the convergence of the random process (1) to a domain of a local extremum of $f(x)$.

2. THE DOMAIN OF CONVERGENCE

For convergence analysis of the trajectories of the random process (1) to a domain of a local extremum of $f(x)$ we use a Liapunov function of the following type:

$$\varphi(N, x) = \int\limits_{D(\varepsilon)} \widehat{f}(s) p_{U^N}(s) G(x, s)\, ds, \tag{2}$$

where $p_{U^N}(s)$ is a function that has the meaning of the density function of the random vector U^N, $\widehat{f}(s) = f(s) - c$, and c is an arbitrary constant. The integral in (2) is taken over the domain $D(\varepsilon) = \{x \in R^n |\; K > \widehat{f}(x) > \varepsilon\} (\infty > K > \varepsilon > 0)$.

Voronezh State University. Institute for System Analysis, Russian Academy of Sciences, Moscow. Translated from Avtomatika i Telemekhanika, No. 11, pp. 66–73, November, 1994. Original article submitted September 23, 1993.

 0005-1179/94/5511-1604$12.50

We also define the set

$$\overline{D} = \{x \in R^n | \; \|\nabla_x \varphi(x)\| = 0\},$$

where the existence of the limit of functions (2) is assumed if $N = 0, 1 \ldots$:

$$\varphi(x) = \lim \varphi(N, x), \quad N \to \infty. \tag{3}$$

For the existence of (3) it is sufficient that

$$\lim p_{U^N}(x) = p_U(x), \quad \lim c_N = c, \quad N \to \infty, \tag{4}$$

i.e., the limit of the density functions $p_{U^N}(x)$ as well as the constants c_N must exist.

LEMMA. Let $p_U(x) > 0$ in the domain $\overline{D}$; then the inclusion

$$\overline{D} \subset D, \quad D = \{x \in R^N | \; f(x) = c\} \tag{5}$$

holds.

The proof of the lemma is given in the Appendix.

To satisfy conditions (3), (4), and the condition

$$\lim \varepsilon_N = 0, \quad N \to \infty, \tag{6}$$

we can use the limit of the domains D_N as the domain D:

$$D = \lim D_N, \quad N \to \infty, \quad D_N = \{x \in R^n | \; \infty > K > \hat{f}_N(x) \geqslant \varepsilon_n > 0\}. \tag{7}$$

Because the lemma states that the set D includes the set $\overline{D}$, it is sufficient to prove the convergence of the trajectories of the random process (1) to the domain $\overline{D}$ in order to establish the convergence of the trajectories to the domain D.

3. CONVERGENCE OF TRAJECTORIES

THEOREM 1. Let conditions (3)–(7) be fulfilled and, in addition, the following conditions hold: the random process (1) has the Markov property

$$E[X^{N+1} | \; X^N, X^{N-1}, \ldots, X^0] = E[X^{N+1} | \; X^N]; \tag{8}$$

for any initial random vectors X^0 and Y^1 the inequalities

$$\begin{aligned} &(E[Y^{N+1} | \; X^N = x], -\nabla_x \varphi(N, x)) \geqslant 0, \\ &|(E[Y^{N+1} | \; X^N = x], \; \nabla_x \varphi(N, x))| \geqslant \Phi(\|\nabla_x \varphi(N, x)\|) \end{aligned} \tag{9}$$

are satisfied for all $N = 0, 1, 2, \ldots$, where

$$\varphi(N, x) = \int_{D_N} \hat{f}_N(s) p_{U^N}(s) G(x, s)\, ds,$$

and $\Phi(z)$ is a function of the scalar argument z with the properties $\Phi(0) = 0, \Phi(z) > 0, z \neq 0$;

the moments $M_{N,\gamma}$ of the order $1 + \gamma$ of the random vectors $Y^{N+1} | \; X^N = x$ with the density functions $p_{Y^{N+1}|X^N=x}(y)$ are bounded

$$\infty > M_{N,\gamma} = \int_{D_N} \|y\|^{(1+\gamma)} p_{Y^{N+1}|X^N=x}(y) dy, \; 0 < \gamma < 1; \tag{10}$$

the function $p_U(x) f(x)$ satisfies a Lipschitz condition of the form

$$p_{U^N}(x)\widehat{f}_N(x) \in \mathrm{Lip}_\gamma(D_N), \tag{11}$$

or

$$\left|p_{U^N}(x)\widehat{f}_N(x) - p_{U^N}(y)\widehat{f}_N(y)\right| \leqslant \Gamma_N\|x-y\|^\gamma, \quad 0<\Gamma_N \leqslant \Gamma < \infty, \quad 0<\gamma<1;$$

the parameters α_N and β_N satisfy the relations

$$\alpha_N > 0, \quad \sum_{N=0}^{\infty} \alpha_N = \infty, \quad \sum_{N=0}^{\infty} \beta_N < \infty, \quad \beta_N = (1+\gamma)^{-1}(\alpha_N)^{(1+\gamma)} c_N \Gamma_N M_{N,\gamma} > 0; \tag{12}$$

the limit of the Liapunov functions exists

$$\inf_{N \geqslant 0} \varphi(N, x) \to \infty \quad \text{if } \|x\| \to \infty; \tag{13}$$

the initial density function p_{U^0} is bounded in a domain D_0

$$p_{U^0} \leqslant p_0 < \infty. \tag{14}$$

Then

$$P\left\{\sup_{N \geqslant 0} \|X^N\| = R(\omega) < \infty\right\} = 1,$$
$$P\left\{\sum_{N=0}^{\infty} \alpha_N \Phi(\|\nabla_x \varphi(N, X^N)\|) < \infty\right\} = 1,$$
$$P\left\{\underline{\lim}\, \rho(X^N, \overline{D}) = 0\right\} = 1.$$

The proof of Theorem 1 is given in the Appendix.

Discussion. The Markov property of the random process (1) (condition (8)) is a traditional assumption for the recurrent procedures of stochastic approximation type [3]. Inequalities (9) represent the hypothesis of the pseudogradient condition [4] for the process (1), i.e., the motion along the direction given by the random vector Y^{N+1} constitutes on the average an acute angle with the antigradient of the Liapunov function (2).

Condition (10) is imposed on the moments of the order $1+\gamma$ of the random vectors $Y^{N+1} | \ X^N = x$ and means the finiteness of these moments.

The Lipschitz condition (11) is a kind of a priori information on the objective function $f(x)$.

Conditions (12) on the parameters of the random process (1) can be fulfilled simultaneously, e.g., if $\alpha_N = (1+N)^{-\alpha}$, $(1+\gamma)^{-1} < \alpha < 1$. This means that conditions (12) are consistent. In fact, we have the inequality $M_{N,\gamma} c_N \Gamma_N \leqslant M c \Gamma$ if $0 < M_{N,\gamma} \leqslant M < \infty$. From this the estimate

$$\sum_{N=0}^{\infty} \beta_N \leqslant M c \Gamma (1+\gamma)^{-1} \sum_{N=0}^{\infty} (1+N)^{-\alpha(1+\gamma)} < \infty$$

follows, which proves the consistency of conditions (12).

We also note the following. If the set D is bounded in R^n, then the set $\overline{D}$ is also bounded (this follows from the lemma). If a finite solution of the problem of optimizing the function $f(x)$ exists, then nothing hinders us from assuming the boundness of the set D. This fact can simplify essentially the development of numerical methods of minimizing the function f that satisfy the conditions of Theorem 1.

The condition $p_{U^0} \leqslant p_0 < \infty$ is sufficient for fulfilling the inequality $|E(\varphi(0, X^0))| < \infty$.

This inequality is necessary for the proof of the condition

$$P\left\{\sum_{N=0}^{\infty} \alpha_N \Phi(\|\nabla_x \varphi(N, X^N)\|) < \infty\right\} = 1.$$

We now establish the fact of the strong convergence (i.e., almost surely) of the trajectories of the process (1) to a domain of a (local) extremum of the function $f(x)$.

THEOREM 2. Let conditions (3)–(8) and (10)–(14) be fulfilled, and, in addition, the following conditions hold:

$$(E[Y^{N+1}|\ X^N=x], -\nabla_x\varphi(N,x)) \geqslant 0,$$
$$|(E[Y^{N+1}|\ X^N=x], \nabla_x\varphi(N,x))| \geqslant \Phi(\varphi(N,x)), \quad \Phi(z)>0, \quad z\neq 0, \quad \Phi(0)=0. \tag{15}$$

Then the trajectories of the process (1) converge almost surely to the set Ω, where

$$\Omega = \{x\in R^n|\ \varphi(x)=0\}.$$

In addition, the convergence of the moments

$$\lim E[\varphi(N,X^N)]^\delta = 0, \quad N\to\infty,$$

holds for any value of the parameter $\delta: 0<\delta<1$.

The proof of Theorem 2 is given in the Appendix.

Discussion. Theorem 2 states the fact of the strong convergence to the set Ω. The set D includes the set Ω; therefore, Theorem 2 establishes the fact of the strong convergence to the set D. The inclusion follows from the definition of the Liapunov function (2). In fact, according to the maximum principle for subharmonic functions [5], the inequality $\Delta\varphi \leqslant 0$ holds, i.e., the function $-\varphi(x)$ is subharmonic in R^n and cannot achieve its maximum inside the set D (if only $\varphi(x)\not\equiv$ const). Because maximizing the function $-\varphi(x)$ is equivalent to making it zero, we should concentrate the density function $p_U(x)$ at points s, for which the equality $G(x,s)(f(s)-c)=0$ holds. This means that we have to find a solution $s=x^*$ of the above equation: $G(x,x^*)(f(x^*)-c)=0$. In view of the structure of the kernel $G(x,s)$, we can achieve the coincidence of the sets of solutions of the equations $f(s)=c$ and $G(x,s)(f(s)-c)=0$ by the choice of the density function $p_U(s)$. This means that we can achieve the coincidence (not only inclusion) of the sets: $\Omega=D$.

4. ON THE IMPLEMENTATION OF THE SEARCH PROCEDURE OF OPTIMIZATION

In Theorems 1 and 2 the finiteness of the moments of the random vectors $Y^{N+1}|X^N=x$ with the conditional densities $p_{Y^{N+1}|X^N=x}(y)$ is assumed. Therefore, we can use these random vectors as the directions of motion in (1). The deterministic component of these random vectors is the vector $E[Y^{N+1}|X^N=x]$. Conditions (9) and (15) allow us to set $E[Y^{N+1}|\ X^N=x]=-\nabla_x\varphi(N,x)$. In this case the process (1) is deterministic. Therefore, the random process (1) can be used for developing deterministic methods of the nonsmooth nonlocal optimization.

We now describe a method of reduction of the value of the variance of random vectors in (1) in order to show its usefulness for developing methods of the random search.

Let the dimension of the space R^n be $n>2$, and let the density function p_{U^N} in (2) be concentrated in the vicinity of the point x given by the ball $B_R=\{s\in R^n|\ \|x-s\|\leqslant R\}$ $(0<R<\infty)$.

We denote $p_U=p_{U^N}$ for all $N=0,1,2,\dots$ and write the gradient of the Liapunov function (2) in the form

$$\nabla_x\varphi(x) = \int\limits_{\|x-s\|\leqslant R} h(x,s)\|x-s\|^{-(n-1)}ds, \tag{16}$$

where

$$h(x,s)=\widehat{f}(s)p_U(s)\omega_n^{-1}\theta(x,s), \quad \theta=(x-s)\|x-s\|^{-1}.$$

We select the random vector θ uniformly distributed in a ball B_R: $p_\theta(s)=V_R^{-1}$, where the ball B_R has the finite volume V_R. We set $z=V_R h(x,\theta)\|x-\theta\|^{(1-n)}$. Then the relation

$$E[z]=V_R\int\limits_{B_R} h(x,s)\|x-s\|^{(1-n)}p_\theta(s)\,ds = I(x)=\nabla_x\varphi(x)$$

holds.

However, the variance

$$D(z)=[E[z]^2-(E[z])^2]=V_R^2\int\limits_{B_R} h^2(x,s)\|x-s\|^{2(1-n)}p_\theta(s)\,ds - I^2(x)$$

is not limited in this case because the integral diverges here.

To reduce the level of the variance $D(z)$ we will use the method of essential sampling.

In this case we specify the density function $p_\theta(s)$ with the same singularity that the integrand function has. For this we transfer the origin of coordinates to the point x and pass to the spherical coordinates $(r, \theta_1, \theta_2, \dots, \theta_{n-1})$ with center at this point x. The direction from the point x is given by the vector $\theta = \frac{x-s}{\|s-x\|}$, and the distance along this direction from the point x to the boundary of the ball equals R.

Let

$$p_{\theta|x}(s) = (n-\alpha)\omega_n^{-1} R^{(\alpha-n)} r^{-\alpha}, \quad \alpha = (n-1), \quad r = \|x-s\|; \tag{17}$$

then the relation

$$\int\limits_{B_R} p_{\theta|x}(s) ds = \int\limits_{S_R} \int\limits_0^R (n-\alpha)\omega_n^{-1} R^{(\alpha-n)} (n-\alpha) r^{-\alpha}\, dr\, dS_r =$$

$$\int\limits_{S_1} (n-\alpha)\omega_n^{-1} R^{(\alpha-n)} \int\limits_0^R r^{(n-1)} r^{-\alpha}\, dr\, dS_1 = 1,$$

$$S_R = \{z \in R^n \,|\, \|x-z\| = R\}$$

holds, i.e., expression (17) determines a density function of the random vector $\theta|x$. Here, S_R is a sphere of radius R and with center at the point x.

We consider now the random value

$$z(\theta|x) = \omega_n h(x, \theta) R^{(n-\alpha)} (n-\alpha)^{-1}, \tag{18}$$

and compute its mathematical expectation

$$E[z] = \int\limits_{B_R} z(s) p_{\theta|x}(s)\, ds =$$

$$\int\limits_{\|x-s\| \leqslant R} \omega_n h(x,s) R^{(n-\alpha)} (n-\alpha)^{-1} (n-\alpha) \omega_n^{-1} R^{\alpha-n} r^{-\alpha}\, ds =$$

$$\int\limits_{\|x-s\| \leqslant R} h(x,s) \|x-s\|^{-\alpha}\, ds = I(x).$$

The variance of this random value is bounded

$$D(z) = \omega_n^2 (n-\alpha)^{-2} \int\limits_{B_R} h^2(x,s) R^{2(n-\alpha)} p_{\theta|x}(s)\, ds - I^2(x) =$$

$$\omega_n (n-\alpha)^{-1} \int\limits_{B_R} h^2(x,s) R^{(n-\alpha)} \|x-s\|^{-\alpha}\, ds - I^2(x),$$

i.e., the integral converges in this case.

Thus, in numerical methods of optimization based on the process (1) the Liapunov function (2) can be replaced by the mathematical expectation $E[z]$ over the density function (17) with bounded variance.

APPENDIX

Proof of the lemma. We have $\overline{D} = \{x \in R^n \,|\, \|\nabla\varphi(x)\| = 0\}, D = \{x \in R^n \,|\, f(x) = c\}$. By virtue of the properties of the Newtonian field for the function (2), we have $\Delta\varphi(N, x) = \widehat{f}(x) p_{U^N}(x)$. Therefore, $\Delta\varphi = \operatorname{div}(\nabla_x \varphi) = \widehat{f}(x) p_U(x)$. If $x \in \overline{D}$, then $\|\nabla\varphi(x)\| = 0$. Consequently, $\nabla_x \varphi(x) = 0$. From this it follows that $\Delta\varphi(x) = 0$. Because $p_U(x) \neq 0$, we have $f(x) = c$. Therefore, $\overline{D} \subset D$. The lemma is proved.

Proof of Theorem 1. The proof of Theorem 1 consists of the verification of all conditions of Remark 5.1 in [4]. We shall also use the following auxiliary results:

the formula of finite increments for the Liapunov function (2), and

the Lipschitz condition for the gradient $\nabla_x \varphi(N,x)$ of the Liapunov function $\varphi(N,x)$ (this follows from the Lipschitz condition (11)).

To prove Theorem 1, we estimate the increment of the generator of the process (1):

$$
\begin{aligned}
&L[\varphi(N,x)] \equiv \int\limits_{D_N} p_{Y^{N+1}|\, X^N=x}(y)(\varphi(N+1,y)-\varphi(N,x))\,dy = \\
&\int\limits_{D_N} p_{Y^{N+1}|X^N=x}(y)\left[\alpha_N(\nabla_x\varphi(N,x),y)+\int\limits_0^{\alpha_N}(\nabla\varphi(N,x+\tau y)-\nabla\varphi(N,x),y)\,d\tau\right]dy \leqslant \\
&\alpha_N(\nabla\varphi(N,x),E[Y^{N+1}|\, X^N=x])+\int\limits_{D_N} p_{Y^{N+1}|\, X^N=x}(y)\int\limits_0^{\alpha_N}\Gamma_N c_N \tau^{\gamma}\|y\|^{(1+\gamma)}\,d\tau\,dy \leqslant \\
&-\alpha_N\Phi[\,\|\nabla_x\varphi(N,x)\|\,]+c_N\Gamma_N M_{N,\gamma}(1+\gamma)^{-1}\alpha^{(1+\gamma)} \leqslant \\
&-\alpha_n\Phi[\,\|\nabla_x\varphi(N,x)\|\,]+(1+\gamma)^{-1}\alpha_N^{(1+\gamma)}c_N\Gamma_N M_{N,\gamma}(1+\varphi(N,x)).
\end{aligned}
\tag{A.1}
$$

Thus, in (A.1) we established the estimate

$$L[W(N,x)] \leqslant -\alpha_N\Phi[\,\|\nabla_x\varphi(N,x)\|\,] \leqslant 0.$$

Therefore, the pair $[W(N,x),\mathcal{N}_N]$ is a nonnegative supermartingale, where

$$W(N,x)=(1+\varphi(N,x))\prod_{m=N}^{\infty}(1+\beta_m).$$

Consequently, the processes $W(N,X^N)$ and $\varphi(N,X^N)$ have almost surely finite limits if $N\to\infty$. This limit is nonnegative, because $\varphi(N,X)\geqslant 0$. Hence, taking into account Theorem 5.2 from [4], we proved that

$$P\left[\sum_{N=0}^{\infty}\alpha_N\Phi(\|\nabla_x\varphi(N,X)\|)<\infty\right]=1,$$

and, taking into account the conditions of Theorem 1, we proved that a subsequence $N_k\to\infty$ exists for which

$$\lim\|\nabla_x\varphi(N_k,X_{N_k})\|=0,\quad N_k\to\infty;$$

holds. Theorem 1 is proved.

Proof of Theorem 2. The proof of Theorem 2 consists of the verification of all conditions of theorem 7.1 from [4].

Similarily to (A.1), we obtain an estimate of the increment of the generator for the random process (1) with the Liapunov function (2)

$$
\begin{aligned}
&L[\varphi(N,x)] \equiv \int\limits_{D_N} p_{Y^{n+1}|\, X^N=x}(y)[\varphi(N+1,y)-\varphi(N,x)]\,dy \leqslant \\
&\alpha_n(\nabla_x\varphi(N,x),\ E(Y^{N+1}|\, X^N=x))+\int\limits_{D_N} p_{Y^{N+1}|\, X^N=x}(y)\int\limits_0^{\alpha_N}\Gamma_N c_N\tau^{\gamma}\|y\|^{(1+\gamma)}\,d\tau\,dy \leqslant \\
&-\alpha_N\Phi[\varphi(N,x)]+(1+\gamma)^{-1}c_N\Gamma_N M_{N,\gamma}\alpha_N^{(1+\gamma)}(1+\varphi(N,x)).
\end{aligned}
\tag{A.2}
$$

According to [4], we proved in (A.2) that the estimate

$$L[W(N,x)] \leqslant -\alpha_N\Phi(\varphi(N,x)) \leqslant 0$$

holds. Here, $W(N,x)=(1+\varphi(N,x))\prod_{m=N}^{\infty}(1+\beta_m)$.

Therefore, the random processes $W(N, X^N)$ and $\varphi(N, X^N)$ have almost surely finite limits if $N \to \infty$. According to Theorem 7.1 from [4], we proved in (A.2) that the following series converges almost surely:

$$P\left[\sum_{N=0}^{\infty} \alpha_N \Phi(\varphi(N, X^N)) < \infty\right] = 1.$$

Hence, taking into consideration the conditions of Theorem 2, a subsequence $N_k \to \infty$ exists for which the following limit exists almost surely:

$$\lim \varphi(N_k, X^{N_k}) = 0.$$

Taking into account condition (15), we obtain that $\lim \varphi(N_k, X^{N_k}) = 0, N_k \to \infty$. Because $\varphi(N, X) \geqslant 0$, we have $\lim \varphi(N, X^N) \geqslant 0, N \to \infty$.

Taking into account that these limits are equal to zero for the subsequence $N_k \to \infty$, we find that $\lim \varphi(N, X^N) = 0, N \to \infty$ almost surely. This completes the proof of Theorem 2.

REFERENCES

1. A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi, "Analysis of search methods of optimization based on potential theory. I. Nonlocal properties," *Avtomat. Telemekh.*, No. 8, 97–105 (1994).
2. A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi, "Analysis of search methods of optimization based on potential theory. II. On differentiability of the Liapunov function," *Avtomat. Telemekh.*, No. 9 (1994).
3. M. B. Nevel'son and R. Z. Has'minskii, *Stochastic Approximation and Recurrent Estimation* [in Russian], Nauka, Moscow (1972).
4. B. T. Polyak and Ya. Z. Tsypkin, "Optimal pseudogradient algorithms of adaptation," *Dokl. Akad. Nauk SSSR*, **250**, No. 5, 1084–1087 (1980).
5. W. K. Hayman and P. B. Kennedy, *Subharmonic Functions*, Academic Press, New York (1976).