UNIVERSITATIS IAGELLONICAE ACTA MATHEMATICA
doi: 10.4467/20843828AM.12.005.0457 FASCICULUS XLIX, 2011

SUFFICIENT CONDITIONS FOR THE CONVERGENCE OF NON-AUTONOMOUS STOCHASTIC SEARCH FOR A GLOBAL MINIMUM

by Dawid Tarłowski

Abstract. The majority of stochastic optimization algorithms can be written in the general form $x_{t+1} = T_t(x_t, y_t)$, where x_t is a sequence of points and parameters which are transformed by the algorithm, T_t are the methods of the algorithm and y_t represent the randomness of the algorithm. We extend the results of papers [**11**] and [**14**] to provide some new general conditions under which the algorithm finds a global minimum with probability one.

1. Introduction. Recent decades have been witnessing a great development of stochastic optimization techniques. Many methods are purely heuristic and their performance is experimentally confirmed. At the same time the corresponding mathematical background is underdeveloped. The global minimization problem concerns finding a solution of

$$\min_{x\in A} f(x),$$

where $f\colon A \to \mathbb{R}$ is the problem function given on a metric space (A, d) of all possible solutions. The most common mathematical tools of the stochastic convergence analysis are the probability theory and the Markov chains theory, see [**6, 3, 8**] for the general theory or [**1, 15, 13**] for some applications. This paper is a continuation of papers [**11**] and [**14**], where some concepts of the Lyapunov stability theory and the weak convergence of measures have been used. As it was discussed there, the majority of algorithms can be written in the general form $x_{t+1} = T_t(x_t, y_t)$, where x_t is a sequence of points and parameters which are successively transformed by the algorithm, y_t represents

Key words and phrases. Stochastic optimization, global optimization, Lyapunov function, weak convergence of measures.

the randomness of the algorithm and T_t are the methods of the algorithm. The algorithm was viewed as a non-autonomous dynamical system on Borel probability measures on the space A of admissible solutions; the proper Lyapunov function has been applied to it and some sufficient conditions for the global convergence have been established. As before, in theoretical analysis we assume that x_t belongs to A. This assumption does not prevent the applications of the theoretical results, even in the case of self-adaptive evolution strategies, like $(\mu + \lambda)$ and (μ, λ) algorithms, see [**2**] or Section 7 in [**14**]. In fact, if $x_t = (p_t^1, \cdots, p_t^k, c_t^1, \cdots, c_t^l) \in A^k \times C^l$, where C is a space of parameters, then we can consider the space $\widehat{A} = A^k \times C^l$ and the function $\widehat{f}(p^1, \cdots, p^k, c^1, \cdots, c^l) = \sum_{i=1}^{k} f(p^i)$. Roughly speaking, the basic convergence assumption, used in [**11, 14**], and in the previous papers [**9, 10, 12**], was

$$\int f(T(x,y))dy < f(x), \tag{1.1}$$

where T represents some methods of the algorithm and dy is an integration according to some probability distribution. The assumption means that the algorithm is capable of reaching from any position, in one step, the areas with lower function values. However, some algorithms, like Particle Swarm Optimization (PSO), [**4, 5**], gradually move through the search space and do not necessarily satisfy the condition, but remembering the best point found, they are capable of finding the global solution. In this paper we show that (1.1) can by replaced with a softer condition, which may be useful in further convergence analysis of some swarm intelligence algorithms, like PSO.

This paper is organized as follows. In Section 2 we define the algorithm and present the main results of the paper, Theorem 1 and Theorem 2. In Section 3 we recall one of the results of [**14**], where a Lyapunov function is applied to some non-autonomous dynamical system. Next, we use this result to provide a proof of Theorem 4 stated in Section 4. In Section 5 we show that Theorem 4 leads to Theorem 1, whilst Theorem 2 is a consequence of Theorem 1.

2. The algorithm and the global convergence. Let (A, d) and (B, d_B) be separable metric spaces and let $f: A \to \mathbb{R}$ be a continuous function which attains its global minimum $f_{\min}$. Without loss of generality we assume that $f_{\min} = 0$. Let

$$A^\star = \{x \in A: \ f(x) = 0\}$$

be a set of global minimums. Let (Ω, Σ, P) be a probability space. We will provide some sufficient conditions for the convergence of a vast class of stochastic optimization methods, which can be modeled as the sequence of random variables $X_t: \Omega \to A$, $t = 0, 1, 2, \cdots$, defined by the non-autonomous equation

$$X_{t+1} = T_t(X_t, Y_t), \tag{2.1}$$

where

- $Y_t \colon \Omega \to B$ are measurable
- $T_t \colon A \times B \longrightarrow A$ are measurable
- the sequence $X_0, Y_0, Y_1, \cdots$ is independent.

X_t is a sequence, successively transformed by the algorithm, which approximates a global solution, Y_t represent the randomness of the algorithm and T_t represent the methods, by which the algorithm transforms the points and the parameters.

Let $(\mathcal{T}, d_{\mathcal{T}})$ denote a metric space of all measurable operators $T \colon A \times B \longrightarrow A$ with a uniform convergence metric and let $(\mathcal{N}, \tau_{\mathcal{N}})$ denote the space of all Borel probability measures on B equipped with a weak convergence topology. Let $X_t \colon \Omega \to A$ be a sequence defined by equation (2.1) and let ν_t denotes the distribution of Y_t, $t = 0, 1, \cdots$. It is clear that the sequence $\{(T_t, \nu_t)\}_{t=0}^{\infty}$ and the initial distribution μ_0 of X_0 determine the distributions of X_t.

For any $l \in \mathbb{N}$ define the sequence

$$T^{(t,l)} \colon A \times B^t \longrightarrow A, \ t = 1, 2, \cdots$$

as $T^{(1,l)} = T_l$ and

$$T^{(t+1,l)}(x, y_l, \cdots, y_{l+t}) = T_{t+l+1}\left(T^{(t,l)}(x, y_l, \cdots, y_{l+t-1}), y_{l+t}\right). \tag{2.2}$$

We will write $T^t := T^{(t,0)}$, $t = 1, 2, \cdots$. Clearly

$$X_{t+1} = T^{t+1}(X_0, Y_0, \cdots, Y_t)$$

and, for any $l \in \mathbb{N}$,

$$X_{l+t+1} = T^{(t+1,l)}(X_l, Y_l, Y_{l+1} \cdots, Y_{l+t}).$$

In Theorem 1 and Theorem 2 we present the conditions under which the algorithm, defined by (2.1), converges to the set of global solutions with probability 1.

Theorem 1. *Assume that A is compact. Let $U \subset \mathcal{T} \times \mathcal{N}$ and let $U_0 \subset U$ be such that U_0 is compact and*

- **(A)** *for any $(T, \nu) \in U_0$ and $x \in A$, T is continuous in (x, y) for any y from some set of full measure ν,*
- **(B)** *for any $(T, \nu) \in U$ and $x \in A$*

$$\int_B f(T(x, y)) v(dy) \le f(x), \tag{2.3}$$

(C) *there is $s \geq 0$ such that for any $\{(T_i, \nu_i)\colon\ i = 0, \cdots, s\} \subset U_0$ and $x \in A \setminus A^*$*

$$\int_{B^{s+1}} f\left(T^{s+1}(x, y_0, \cdots, y_s)\right)\ \nu_s \times \cdots \times \nu_0\ (dy_s, \cdots, dy_0) < f(x), \tag{2.4}$$

where $T^{s+1} = T^{(s+1,0)}$ is defined by (2.2). If $u_t = (T_t, \nu_t) \in U$ is such that for any $t \in \mathbb{N}$ there is $t_0 \geq t$ such that for $i \leq s$ we have $u_{t_0+i} \in U_0$, then

$$\forall \epsilon > 0\ P(d(X_t, A^*) < \varepsilon) \overset{t\to\infty}{\longrightarrow} 1$$

and

$$Ef(X_t) \searrow 0,\ t \to \infty.$$

$Ef(X_t)$ denotes the expected value of the random variable $f(X_t)\colon \Omega \to \mathbb{R}$, i.e. $E(f(X_t)) = \int\limits_\Omega f(X_t)dP$. If we express condition **(B)** in terms of the conditional probability, then we have

$$E(f(X_{t+1})|X_t = x) \leq f(x),$$

where $(T_t, \nu_t) \in U$. Similarly, condition **(C)** takes a form

$$E(f(X_{t+s+1})|X_t = x) < f(x),$$

where $x \in A \setminus A^\star$ and $(T_{t+i}, \nu_{t+1}) \in U_0$, $i = 0, 1, \cdots, s$. It gives the intuition behind the condition.

Many algorithm are monotonous, i.e. they satisfy $f(X_{t+1}) \leq f(X_t)$. If we strengthen condition **(B)** assuming the algorithm monotonous, then we will obtain Theorem 2. For any $\delta > 0$ let

$$A_\delta = \{x \in A\colon f(x) \leq \delta\} \text{ and } T_\delta = T|_{A_\delta}\ :\ A_\delta \times B \longrightarrow A.$$

For any $U \subset \mathcal{T} \times \mathcal{N}$ let

$$U(\delta) = \mathcal{T}_\delta \times \mathcal{N},\ \text{ where } \mathcal{T}_\delta = \{T_\delta\colon\ T \in \mathcal{T}\}.$$

It is simple that if A and U_0 are compact, then A_δ and $(U_0)_\delta$ are compact for any $\delta > 0$. In the case $A = \mathbb{R}^n$, by the continuity of f, for the compactness of A_δ, $\delta > 0$, it is enough to assume that the sets A_δ are bounded.

Theorem 2. *Assume that A_δ is compact, $\delta > 0$. Let $U \subset \mathcal{T} \times \mathcal{N}$ and $U_0 \subset U$ be such that $U_0(\delta)$ is compact for any $\delta > 0$ and conditions* **(A)** *and* **(C)** *are satisfied. Assume that*

(B') *for any $(T, \nu) \in U$ and $x \in A$, $y \in B$*

$$f(T(x, y)) \leq f(x).$$

Let $u_t = (T_t, \nu_t) \in U$. *If for any* $t \in \mathbb{N}$ *there is* $t_0 \geq t$ *such that for* $i \leq s$ *we have* $u_{t_0+i} \in U_0$*, then*

$$P(d(X_t, A^*) \to 0,\ t \to \infty) = 1$$

and

$$f(X_t) \searrow 0,\ t \to \infty \quad a.s.$$

REMARK 1. The case $s = 0$ was analyzed in [**11, 14**]. If $s = 0$, condition (**A**) of the theorems can be weakened, see Theorems 1 and 2 stated in [**14**].

3. Some concepts of the Lyapunov stability theory. Let $\mathcal{U}$ be a metric space and let M be a compact metric space. Let $\theta\colon \mathcal{U} \ni u \to \theta u \in \mathcal{U}$ and $\Pi\colon \mathcal{U} \times M\colon (u, m) \to \Pi_u m \in M$ be given continuous maps. For $t \geq 0$ define $\Pi^t\colon \mathcal{U} \times M \ni (u, m) \to \Pi_u^t m \in M$ as

$$\Pi^0(u, m) = m \text{ and } \Pi_u^{t+1} m = \Pi_{\theta^t u} \circ \Pi_u^t m, \text{ where } \theta^0 u = u. \tag{3.1}$$

In other words, $\Pi_u^t m = (\Pi_{\theta^{t-1}u} \circ \Pi_{\theta^{t-2}u} \circ \cdots \circ \Pi_u)(m)$, $t \geq 1$.

For any $u \in U$, the sequence Π_u^t determines a non-autonomous dynamical semi-system on M. For any $m \in M$, its orbits are given by $\{\Pi_u^t m\colon\ t = 0, 1, 2, \cdots\}$. At the same time $\Pi_u\colon M \to M$ is a continuous function which induces an autonomous dynamical system on M with orbits $\{(\Pi_u)^t m\colon\ t = 0, 1, \cdots\}$, where $(\Pi_u)^0 m = m$ and $(\Pi_u)^{t+1} m = \Pi_u(\Pi_u)^t m$. We will say that a closed set $K \subset M$ is invariant for Π_u, where $u \in U$, iff $\Pi_u(K) \subset K$.

THEOREM 3. *Let* $\varnothing \neq M^\star \subset M$ *be closed and invariant for any* Π_u, $u \in \mathcal{U}$. *Let* $V\colon M \to \mathbb{R}$ *be a Lyapunov function for any* Π_u, $u \in U$, *i.e:*

1. V *is continuous,*
2. $V(m) = 0$ *for* $m \in M^*$,
3. $V(m) > 0$ *for* $m \in M \setminus M^*$,
4. $V(\Pi_u m) \leq V(m)$ *for any* $u \in \mathcal{U}$ *and* $m \in M$.

Let $\mathcal{U}_0 \subset \mathcal{U}$ *and* $\mathcal{U}_0' \subset \mathcal{U}$ *be such that* $\mathcal{U}_0'$ *is compact and*

(a) *for any* $u \in \mathcal{U}$ *there is* $k \geq 0$ *with* $\theta^k u \in \mathcal{U}_0$,
(b) *for any* $u \in \mathcal{U}_0$ *and* $m \in M \setminus M^\star$, $V(\Pi_u m) < V(m)$,
(c) *there is a surjection* $\zeta\colon \mathcal{U}_0 \to \mathcal{U}_0'$ *such that for* $u \in \mathcal{U}_0$ *and* $m \in M$

$$\Pi_u m = \Pi_{\zeta(u)} m. \tag{3.2}$$

Then, for any $u \in \mathcal{U}$ *and* $m \in M$,

$$V(\Pi_u^t m) \searrow 0, \ \textit{as } t \to \infty.$$

PROOF. The theorem is a direct consequence of Theorem 4 stated in [**14**]. □

4. Some concepts of the theory of weak convergence of measures. First recall some useful facts about the weak convergence of Borel probability measures. For more details, see for example [**6**] or [**3**].

Let $M(S)$ be a space of Borel probability measures on a separable metric space (S, d_S). We say that a sequence $\mu_t \in M(S)$ converges to some $\mu \in M(S)$ if for any bounded and continuous function $h\colon S \to \mathbb{R}$ we have

$$\int\limits_S h \; d\mu_t \to \mu, \text{ as } t \to \infty.$$

As S is separable, the topology of weak convergence is metrizable and one of accessible metrics is the Prohorov metric, defined by

$$d_M(\nu_1, \nu_2) = \inf\{\varepsilon > 0\colon \nu_1(D) \le \nu_2(D^\varepsilon) + \varepsilon \text{ for any Borel set } D \subset S\},$$

where $D^\varepsilon = \{y \in S\colon\ d_S(y, D) < \varepsilon\}$. Furthermore, if S is compact, then $M(S)$ is compact.

From now on,(M, d_M) will denote the metric space of Borel probability measures on A with the Prohorov metric d_M. Fix $(T, \nu) \in \mathcal{T} \times \mathcal{N}$. The function $P_{(T,\nu)}\colon M \ni \mu \to P_{(T,\nu)}\mu \in M$, defined by

$$P_{(T,\nu)}\mu(C) = (\mu \times \nu)(T^{-1}(C)), \text{ for any Borel set } C \subset M,$$

is a Foias operator, see [**7**]. We will also write $P_{(T,\nu)}\mu = (\mu \times \mu)T^{-1}$.

Proposition 1. *If $U_0 \subset \mathcal{T} \times M$ satisfies assumption* **(A)** *of Theorem 1, then the function $P : U_0 \times M \ni (u, \mu) \to P_u\mu \in M$ is continuous.*

Proof. For the proof see Proposition 1 established in [**11**]. □

Let

$$M^\star = \{\mu \in M\colon\ supp\ \mu \subset A^\star\}.$$

The following theorem is a basic tool for proving Theorem 1 stated in Section 5.

Theorem 4. *Assume that $U \subset \mathcal{T} \times \mathcal{N}$ and $U_0 \subset U$ are such that U_0 is compact and conditions **(A)**, **(B)** and **(C)** of Theorem 1 are satisfied. Let $(T_t, \nu_t) \in U$, $t \in \mathbb{N}$. If for any t there is $t_0 \ge t$ such that $\{(T_{t_0+i}, \nu_{t_0+i})\colon\ i = 0, 1, \cdots, s\} \subset U_0$, then for any $\mu_0 \in M$, the sequence $\mu_t \in M$, defined by $\mu_{t+1} = P_{(T_t,\nu_t)}\mu_t$, satisfies*

$$d_M(\mu_t, M^*) \to 0, \text{ as } t \to \infty$$

and

$$\int_A f d\mu_t \searrow 0, \ \ as\ t \to \infty.$$

Proof. We will take advantage of Theorem 3. Let $(\mathbb{N}, d_{\mathbb{N}})$ be a discrete metric space and let

$$\mathcal{U} = \mathbb{N} \times \{u = (u_0, u_1, \cdots) \in U^{\mathbb{N}} \colon \ \forall t\ (T_t, \nu_t) \in U_0 \Rightarrow u_t \in U_0\}$$

be a metric space with the product metric $d_{\mathcal{U}}$, which is defined by

$$d_{\mathcal{U}}((m,u),(n,v)) = d_{\mathbb{N}}(m,n) + \sum_{i=1}^{\infty} 2^{-i} d_U(u_i, v_i).$$

Let $\{t_k\}_{k=0}^{\infty} \subset \mathbb{N}$ be a sequence defined by

$$t_0 = \min\{t \in \mathbb{N} \colon \ (T_{t+i}, \nu_{t+i}) \in U_0 \colon \ i = 0, 1, \cdots, s\}$$

and

$$t_{k+1} = \min\{t \geq t_k + s + 1 \colon \ \ (T_{t+i}, \nu_{t+i}) \in U_0 \colon \ i = 0, 1, \cdots, s\}.$$

Let

$$t(\mathbb{N}) = \{t_k \colon \ k = 0, 1, \cdots\}.$$

Let $\alpha \colon \mathbb{N} \ni k \to \alpha_k \in \mathbb{N}$ satisfy

$$\alpha_0 = 0 \ \text{ and } \ \alpha_{k+1} = \begin{cases} \alpha_k + s + 1 & , \text{ if } \alpha_k \in t(\mathbb{N}) \\ \min\ \{k_1 > \alpha_k \colon \ k_1 \in t(\mathbb{N})\} & , \text{ if } \alpha_k \notin t(\mathbb{N}) \end{cases}$$

and let $\beta \colon \mathcal{U} \to \mathcal{U}$ be a shift map defined by

$$\beta(u_0, u_1, \cdots) = (u_1, u_2, \cdots).$$

Clearly, for $k \in \mathbb{N}$, $\beta^k(u_0, u_1, \cdots) = (u_k, u_{k+1}, \cdots)$. We will also write $(u)_k := \beta^k(u)$. Let

$$\theta \colon \mathbb{N} \times U \ni (k, u) \longrightarrow (k+1, (u)_{\alpha(k+1)}) \in \mathbb{N} \times U.$$

Clearly a shift map is continuous, thus θ is continuous. For any natural numbers $l < t$ and $u \in U^{\mathbb{N}}$ define $P_u^{(t,l)} \colon M \to M$ as

$$P_u^{(t,l)} = P_{u_{t-1}} \circ P_{u_{t-2}} \circ \cdots \circ P_{u_l}.$$

Let $\Pi \colon \mathbb{N} \times \mathcal{U} \times M \ni (k, u, \mu) \longrightarrow \Pi_{(k,u)}\mu \in M$ be as follows

$$\Pi_{(k,u)}\mu = P_u^{(\alpha_{k+1} - \alpha_k, 0)}\mu \in M.$$

By Proposition 1, Π is continuous. In fact, for any natural k the function $\Pi_{(k,\cdot)}(\cdot)$ is a composition of continuous functions P_{u_i}. Furthermore,

$$P_{(u)_{\alpha_k}}^{(\alpha_{k+1} - \alpha_k, 0)}\mu = P_u^{(\alpha_{k+1}, \alpha_k)}\mu.$$

Thus, for $u = \{(T_t, \nu_t)\}_{t=0}^{\infty}$ and $t \geq 1$, the Π^t defined by (3.1), satisfy

$$\begin{aligned}\Pi^t_{(0,u)}\mu_0 &= P^{(\alpha_t-\alpha_{t-1},0)}_{(u)_{\alpha_{t-1}}} \circ \cdots \circ P^{(\alpha_1,0)}_u \mu \\ &= P^{(\alpha_t,\alpha_{t-1})}_u \circ \cdots \circ P^{(\alpha_1,0)}_u \mu_0 = P^{(\alpha_t,0)}_u = \mu_{\alpha_t}.\end{aligned}$$

Define $V\colon M \to \mathbb{R}$ as

$$V(\mu) = \int\limits_A f d\mu.$$

We will show that V satisfies all assumptions (1), (2), (3) and (4) of Theorem 3. Since f is continuous (and bounded as A is compact), then the continuity of V follows directly from the definition of weak convergence. To see (2),(3) note that for any $\mu \in M$, *supp* $\mu \subset A^\star$ iff $\mu(A^\star) = 1$. Since f is positive without the set $A^\star$ and equal to 0 on $A^\star$, then it is clear that for any μ from M, $V(\mu) = \int\limits_A f d\mu \geq 0$ and $V(\mu) = 0 \Leftrightarrow \mu(A \setminus A^\star) = 0 \Leftrightarrow \mu \in M^\star$. To see (4), by the definition of Π, it will be enough to know that $V(P_u\mu) \leq V(\mu)$ for any $u = (T, \nu) \in U$ and $\mu \in M$.

By the definition of Foias operator, change of variable, Fubini's theorem and **(B)**,

$$\begin{aligned}V(P_u\mu) &= \int\limits_A f dP_u\mu = \int\limits_{A\times B} f \circ T d(\mu \times \nu) \\ &= \int\limits_A \left(\int\limits_B f(T(x,y))\nu(dy) \right) \mu(dx) \leq \int\limits_A f(x)\mu(dx) = V(\mu).\end{aligned}$$

From (2), (3), (4), there immediately follows that $M^\star$ is invariant under Π_u, for any $u \in \mathcal{U}$. Fix $k_0 \in \alpha^{-1}(t(\mathbb{N}))$, i.e. fix k_0 such that $\alpha_{k_0} \in t(\mathbb{N})$. Define

$$\mathcal{U}_0 = \left(\alpha^{-1}(t(\mathbb{N})) \times (U_0)^{s+1} \times U^{\mathbb{N}} \right) \cap \mathcal{U}$$

and

$$\mathcal{U}_0' = \{k_0\} \times \{u \in (U_0)^{\mathbb{N}} \colon\ u_i = u_{i+s+1},\ i = 0, 1, \cdots\}.$$

Clearly $\mathcal{U}_0'$ is compact as a closed subset of a compact set $\{k_0\} \times (U_0)^{\mathbb{N}}$. It remains to show that $\mathcal{U}_0, \mathcal{U}_0'$ satisfy assumptions (a),(b),(c) of Theorem 3. (a) is an immediate consequence of the definitions of $\mathcal{U}_0$, α and θ. To see (b), we need $V(\Pi_{(k,u)}\mu) < V(\mu)$ for any $(k,u) \in \mathcal{U}_0$ and $\mu \in M \setminus M^\star$. Since $\alpha_k \in t(\mathbb{N})$, then $\Pi_{(k,u)}\mu = P^{(s+1,0)}_u \mu$. Hence, we need

$$V\left(P_{u_s} \circ \cdots \circ P_{u_0})\mu\right) < V(\mu)$$

for any $(u_0, \cdots, u_s) \in (U_0)^{s+1}$ and $\mu \in M \setminus M^\star$. We have $\int\limits_A f dP_{(T,\nu)}\mu = \int\limits_A \left(\int\limits_B f(T(x,y))\nu(dy) \right) \mu(dx)$. Using the induction, by change of variable and

Fubini theorem, we obtain

$$(4.1)\qquad \int_A \left(fd(P_{u_s}\circ\cdots\circ P_{u_0})\right)\mu = \int_A fd\left(((\mu\times\nu_0)T_0^{-1})\times\cdots\times\nu_s\right)T_s^{-1}$$

$$= \int_A \left(\int_{B^{s+1}} f\left(T^{s+1}(x,y_0,\cdots,y_s)\right)\ \nu_s\times\cdots\times\nu_0\ (dy_s,\cdots,dy_0)\right)\mu(dx).$$

Note that the condition **(B)** implies that for any $(T_i,\nu_i)_{i=0}^s \in (U_0)^{s+1}$, $T^{s+1}=T^{(s+1,0)}$, defined by (2.2), satisfies
(4.2)

$$\forall\ x\in A^\star \int_{B^{s+1}} f\left(T^{s+1}(x,y_0,\cdots,y_s)\right)\ \nu_s\times\cdots\times\nu_0\ (dy_s,\cdots,dy_0)\le f(x)=0.$$

Fix $\mu\in M\setminus M^\star$. By (4.1), (4.2), **(C)** and $\mu(A\setminus A^\star)>0$, for any $(u_0,\cdots,u_s)\in(\mathcal{U}_0)^{s+1}$,

$$V\left((P_{u_s}\circ\cdots\circ P_{u_0})\mu\right)$$

$$=\int_A\left(\int_{B^{s+1}} f\left(T^{s+1}(x,y_0,\cdots,y_s)\right)\nu_s\times\cdots\times\nu_0\ (dy_s,\cdots,dy_0)\right)\mu(dx)$$

$$=\int_{A\setminus A^\star}\left(\int_{B^{s+1}} f\left(T^{s+1}(x,y_0,\cdots,y_s)\right)\nu_s\times\cdots\times\nu_0\ (dy_s,\cdots,dy_0)\right)\mu(dx)+\int_{A^\star}0d\mu$$

$$<\int_{A\setminus A^\star} f(x)\mu(dx)=\int_A f(x)\mu(dx)=V(\mu).$$

Let

$$\zeta\colon \mathcal{U}_0\ni(k,u)\longrightarrow\left(k_0,(u_{i\ mod(s+1)})_{i=0}^\infty\right)\in\mathcal{U}_0',$$

where $i\ mod(s+1)=k\in\{0,1,\cdots,s\}$ with $(i-k)=c\cdot(s+1)$ for some natural c. Clearly, ζ is a surjection. For any $\alpha_k\in t(\mathbb{N})$ we have $\alpha_{k+1}-\alpha_k=s+1$. Hence, for any $(k,u)\in\mathcal{U}_0$ and $\mu\in M$,

$$\Pi_{(k,u)}\mu=P_u^{(\alpha_{k+1}-\alpha_k,0)}\mu=P_u^{(s+1,0)}\mu=P_u^{(\alpha_{(k_0+1)}-\alpha_{k_0},0)}\mu=\Pi_{\zeta(k,u)}\mu,$$

which proves (c). We have shown that the defined objects V, $\mathcal{U}$, θ, Π, $\mathcal{U}_0$, $\mathcal{U}_0'$ and ζ satisfy all the assumptions of Theorem 3. Since $\mu_{\alpha_t}=\Pi_u^t\mu_0$, where $u=(T_t,\nu_t)_{t=0}^\infty\in\mathcal{U}$, then $V(\mu_{\alpha_t})\searrow 0$. As we have shown, $V(\mu_{t+1})\le V(\mu_t)$. Hence, $V(\mu_t)=\int_A fd\mu_t\searrow 0$. Now, note that the continuity of V and the compactness of M imply that V is separated from zero without any open set D with $D\supset M^\star$. Thus, since $V(\mu_t)\searrow 0$, then $d(\mu_t,M^\star)\to 0$. $\square$

5. Proofs of Theorem 1 and Theorem 2. First, recall a simple lemma.

Lemma 1. *Let $X_t\colon \Omega \to A$ be a sequence of random variables distributed according to $\mu_t \in M$. If $d_M(\mu_t, M^\star) \to 0$, then*

$$\forall \varepsilon > 0 \ P(d(X_t, A^\star) < \varepsilon) \to 1, \quad t \to \infty.$$

Proof. For the proof see Section 5.1 in [**14**]. □

The results of Section 4 lead to Theorem 1.

Proof of Theorem 1. We will make use of Theorem 4. Let μ_t denote the distribution of X_t, $t = 0, 1, \cdots$. Note that, by the definition of X_t, the random variables X_t and Y_t are independent. Thus, $X_{t+1} = T_t(X_t, Y_t)$ is distributed according to $(\mu_t \times \nu_t)T_t^{-1} = P_{(T_t,\nu_t)}\mu_t$. By Theorem 4, $d_M(\mu_t, M^\star) \to 0$ and $\int\limits_A f d\mu_t \searrow 0$. From Lemma 1,

$$\forall \varepsilon > 0 \ P(d(X_t, A^\star) < \varepsilon) \to 0, \quad t \to \infty.$$

Now, it is enough to note that by change of variables,

$$Ef(X_t) = \int\limits_\Omega f(X_t) dP = \int\limits_A f d\mu_t.$$

□

Proof of Theorem 2. Fix $x_0 \in A$. If $\mu_0 = \delta_{x_0}$ is a Dirac measure, then *supp* $\mu_0 = \{x_0\} \subset A_{f(x_0)}$. $A_{f(x_0)} = \{x \in A\colon \ f(x) \leq \delta\}$ is compact and, by **(B')**, $T_t(A_{f(x_0)} \times B) \subset A_{f(x_0)}$ for any $t \in \mathbb{N}$. Clearly $A^\star \subset A_{f(x_0)}$. Thus we may apply Theorem 1 to $A_{f(x_0)}$. Hence, under the assumption $\mu_0 = \delta_{x_0}$, we have $Ef(X_t) \searrow 0$.

Now, let $\mu_0 \in M$. By Fubini's theorem,

$$Ef(X_t) = Ef(T^{t+1}(X_0, Y_0, \cdots, Y_t)) = \int\limits_A Ef(T^{t+1}(x_0, Y_0, \cdots, Y_t))\mu_0(dx_0).$$

Since $Ef(T^{t+1}(x_0, Y_0, \cdots, Y_t) \searrow 0$, for any $x_0 \in A$, then, by the Lebesgue Monotone Convergence Theorem, $Ef(X_t) \searrow 0$. Since, from **(B')**, $f(X_t) \leq f(X_{t+1})$, then $f(X_t) \searrow 0$ almost everywhere (on some set of full measure P). In fact, in the opposite case, again by the Monotone Convergence Theorem, we would have $Ef(X_t) \searrow \delta$ for some $\delta > 0$. Now, it is enough to know that $f(x_t) \searrow$ implies that $d(x_t, A^\star) \to 0$ for any sequence $x_t \in A$. It holds true, because there is $\delta > 0$ such that A_δ is compact. Hence, as a continuous function, f is separated from zero without any open set $D \subset A$ with $D \supset A^\star$. Therefore, $P\left(d(X_t, A^\star) \to 0\right) = 1$. □

References

1. Appel M. J., Labarre R., Radulovic D., *On Accelerated Random Search*, SIAM J. Optim., **14** (2003), 708–731.
2. Beyer H. G., Schwefel H. P., *Evolution Strategies – A Comprehensive Introduction*, Natural Computing, **1** (2002), 3–52.
3. Billingsley P., *Convergence of Probability Measures*, Second Edition, A Wiley-Interscience Publication, 1999.
4. Clerc M., Kennedy J., *The Particle Swarm – Explosion, Stability, and Convergence in a Multidimensional Complex Space*, IEEE Transactions on Evolutionary Computation, **6** (2002).
5. Clerc M., *Particle Swarm Optimization*, ISTE Ltd, London, 2006.
6. Dudley R. M., *Real Analysis and Probability*, Cambridge University Press, 2004.
7. Lasota A., Mackey M., *Chaos, Fractals and Noise*, Springer Verlag, 1994.
8. Meyn S., Twedie R., *Markov Chains and Stochastic Stability*, Springer Verlag, London, 1993.
9. Ombach J., *Stability of evolutionary algorithms*, J. Math. Anal. Appl., **342** (2008), 326–333.
10. Ombach J., *A Proof of Convergence of General Stochastic Search for Global Minimum*, Journal of Difference Equations and Applications, **13** (2007), 795–802.
11. Ombach J., Tarłowski D., *Nonautonomous Stochastic Search in Global Optimization*, Preprint.
12. Radwański M., *Convergence of nonautonomous evolutionary algorithm*, Univ. Iagel. Acta Math., **45** (2007), 197–206.
13. Rudolph G., *Convergence of Evolutionary Algorithms in General Search Spaces*, 50–54, Proceedings of the Third IEEE Conference on Evolutionary Computation, Piscataway, IEEE Press, NJ, 1996.
14. Tarłowski D., *Non-autonomous Stochastic Search for Global Minimum in Continuous Optimization*, Preprint.
15. Tarłowski D., *Global Convergence of Stochastic Optimization Algorithms*, Preprint.
16. Zhigljavsky A., Žilinskas A., *Stochastic Global Optimization*, Springer, New York, 2008.

Received September 3, 2011

Institute of Mathematics
Jagiellonian University
Łojasiewicza 6
30-348 Kraków, Poland
e-mail: `dawid.tarlowski@im.uj.edu.pl`
e-mail: `dawid.tarlowski@uj.edu.pl`