Journal of Difference Equations and Applications

ISSN: 1023-6198 (Print) 1563-5120 (Online) Journal homepage: https://www.tandfonline.com/loi/gdea20

A proof of convergence of general stochastic search for global minimum

Jerzy Ombach &

To cite this article: Jerzy Ombach & (2007) A proof of convergence of general stochastic search for global minimum, Journal of Difference Equations and Applications, 13:8-9, 795-802, DOI: 10.1080/10236190701396560

To link to this article: https://doi.org/10.1080/10236190701396560

Copyright Taylor & Francis Group, LLC

Published online: 13 Aug 2007.

Submit your article to this journal

Article views: 458

View related articles

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gdea20

Journal of Difference Equations and Applications,
Vol. 13, Nos. 8–9, August–September 2007, 795–802

Taylor & Francis
Taylor & Francis Group

A proof of convergence of general stochastic search for global minimum†

JERZY OMBACH‡¶*

‡IM UJ, ul. Reymonta 4, 30 059 Kraków, Poland
¶IP PWSZ, ul. Chruslicka 6, 33 300 Nowy Sącz, Poland

(Received 22 September 2006; in final form 12 March 2007)

We present a general criterion guaranteeing stochastic convergence of a wide class of numerical methods used for finding global minimum of a continuous function.

Keywords: Global optimization; Evolutionary algorithm; Foias operator; Lyapunov function

The aim of this paper is to establish a general sufficient condition for stochastic convergence of a wide class of stochastic numerical methods used for solving global optimization problems. We consider methods of the form $X_t = T(X_{t-1}, \mathbf{Y}_{t-1})$ for $t = 1, 2, 3, \ldots$, where T is a given operator and $\mathbf{Y}_t$ are random variables. Such methods have been more and more developed in recent years and their most advantage is that they can be used in situations when an objective function is not differentiable. Several properties of such methods have been established, e.g. see Ref. [4] where Markov Chains techniques is extensively used. In this paper we apply a version of the classical Lyapunov Stability Theorem to a Foias operator on the space of probability measures to prove stochastic convergence of X_t to the set of the solutions of global minimization problem. Our main result is Theorem 1. In Section 5 we apply this result to get a criterion for stochastic convergence of a wide class of evolutionary algorithms which combine stochastic search with local deterministic methods.

1. Let A be a metric space and f: $A \to \mathbb{R}$ be a continuous function having its global minimum $\min f$ on A. Without loss of generality we can assume that $\min f = 0$. Let $A^* \subset A$ be the set of all the solutions of the global minimization problem, i.e.

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

A vast part of stochastic algorithms used for finding a solution of the global optimization problem yields the following form:

$$X_t = T(X_{t-1}, \mathbf{Y}_{t-1}) \text{ for } t = 1, 2, 3, \ldots \tag{1}$$

Here, X_0 is a fixed random variable having a known distribution on A, $\mathbf{Y}_t$ are random variables having a common distribution on B. T: $A \times B \to A$ is an operator identifying the algorithm. We are interested in convergence of the sequence of random variables X_t to a solution of the global optimization problem.

*Email: ombach@im.uj.edu.pl
†This paper is written in memory of Professor Bernd Aulbach.

Journal of Difference Equations and Applications
ISSN 1023-6198 print/ISSN 1563-5120 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10236190701396560

Let $\mathcal{B}(A)$ and $\mathcal{B}(B)$ denote the family of Borel subsets of the space A and B, respectively. Let $\mathcal{M}$ denote the set of all probability measures on $\mathcal{B}(A)$. Let ν be a probability measure on the σ-algebra $\mathcal{B}(B)$. Let $(\Omega, \Sigma, \text{Prob})$ be a probability space. Let X_0 be a random variable defined on Ω and assume that its distribution $\mu_0 \in \mathcal{M}$. Let $\mathbf{Y}_t$ be a sequence of independent random variables defined on Ω identically distributed on B with common distribution ν each. We assume that X_0 and $\mathbf{Y}_t$ are independent. Assume that T is a measurable function $A \times B \to A$. Thus X_t are random variables.

The following theorem to be proved in Sections 2–4 provides a general sufficient condition for the stochastic convergence of X_t to the set A^*

Theorem 1. *Assume that A is compact and:*

(**A**) *For any $x_0 \in A$ and any sequence $x_n \to x_0$:*

$$T(x_n, \mathbf{y}) \to T(x_0, \mathbf{y}) \quad a.e. \ \ \nu, \ \ as \ \ n \to \infty.$$

(**B**) *For any $x \in A^*$ and $\mathbf{y} \in B$, $T(x, \mathbf{y}) \in A^*$.*

(**C**) *For any $x \in A \backslash A^*$:*

$$\int_B f(T(x, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) < f(x). \tag{2}$$

Then, for every $\varepsilon > 0$:

$$\lim_{t\to\infty} \text{Prob}(\text{dist}(X_t, A^*) < \varepsilon) = 1. \tag{3}$$

Corollary 1. *One can release the assumption of compactness of the set A assuming the following conditions instead:*

(**D**) *For every $x \in A$ and $\mathbf{y} \in B$: $f(T(x, \mathbf{y})) \le f(x)$.*

(**E**) *There exists $r > 0$ such that set $A_r := \{x \in A : f(x) \le r\}$ is compact and supp $\mu_0 \subset A_r$.*

In fact, by (**D**) $T(A_r \times B) \subset A_r$. Clearly, μ_0 is a probability measure on A_r and $A^* \subset A_r$. Hence, we may apply Theorem 1 to set A_r.

2. As we will see, the above Theorem is a simple consequence of Theorem 3 below on the asymptotic stability of the Foias operator $P{:}\mathcal{M} \to \mathcal{M}$ defined as follows:

$$P\mu(C) = \int_A \left(\int_B I_C(T(x, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) \right) \mu(\mathrm{d}x), \ \text{ for } \ \mu \in \mathcal{M}, \ \ C \in \mathcal{B}(A), \tag{4}$$

where I_C is the indicator function of the set C. We are interested in the successive iterations P^t of the Foias operator, i.e. maps defined as: $P^0 = id_{\mathcal{M}}$, $P^{t+1} = P \circ P^t$, for $t = 0, 1, 2, \ldots$.

First we recall some basic properties. It is known, see [6], that $\mathcal{M}$ with the Fortet–Mourier metric is a metric space and its topology is determined by the weak convergence of the sequences of measures as follows. The sequences $\mu_n \in \mathcal{M}$ converges to $\mu \in \mathcal{M}$ if and only if for any continuous (bounded as A is compact) function h,

$$\int_A h\mathrm{d}\mu_n \to \int_A h\mathrm{d}\mu, \tag{5}$$

as $n \to \infty$. Another equivalent conditions for (weak) convergence is:

$$\mu_n(C) \to \mu(C), \tag{6}$$

as $n \to \infty$, for any $C \in \mathcal{B}(A)$ such that $\mu(\delta C) = 0$, where δC is the boundary of C.

Since A is compact so is $\mathcal{M}$.

We will use two basic properties of the Foias operator summarized in two Lemmas below, for more details and proofs see for example Chapter 12 in Ref. [3].

LEMMA 1. *Let the initial random variable X_0 be distributed according to a measure $\mu_0 \in \mathcal{M}$. Then, for every $t = 1, 2, 3, \ldots$:*

$$P^t \mu_0 = \mu_t^T, \tag{7}$$

where the measures μ_t^T are defined by

$$\mu_t^T(C) = \mathrm{Prob}(X_t \in C) = \mathrm{Prob}(X_t^{-1}(C)), C \in \mathcal{B}(A). \tag{8}$$

For a measurable function $h{:}A \to \mathbb{R}$ we define the function Uh as:

$$Uh(x) = \int_B h(T(x, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) \tag{9}$$

LEMMA 2. *Let $\mu \in \mathcal{M}$.* IF *h is continuous, then:*

$$\int_A h\mathrm{d}(P\mu) = \int_A Uh\mathrm{d}\mu. \tag{10}$$

PROPOSITION 1. *Assumption* (**A**) *implies that $P{:}\mathcal{M} \to \mathcal{M}$ is a continuous map. In other words P is* (*weak*) *Feller, see Refs.* [3,4] *or* [8]. *for more details on Feller operators.*

Proof. Let $\mu_n \to \mu$, $\mu_n,\mu \in \mathcal{M}$. *Let* $h{:}A \to \mathbb{R}$ be continuous function. We have to show that $\int_A h\mathrm{d}(P\mu_n) \to \int_A h\mathrm{d}(P\mu)$. Let $x_0 \in A$ and $x_m \to x_0$ be fixed. From (**A**) $T(x_m, \cdot) \to T(x_0, \cdot)$ a.e. ν and so $h(T(x_m, \cdot)) \to h(T(x_0, \cdot))$ on the set of full measure ν and the Dominated Convergence Theorem may be applied. We then have:

$$Uh(x_m) = \int_B h(T(x_m, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) \to \int_B h(T(x_0, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) = Uh(x_0),$$

so the function Uh is continuous. By Lemma 2 and condition (5):

$$\int_A h\mathrm{d}P\mu_n = \int_A Uh\mathrm{d}\mu_n \to \int_A Uh\mathrm{d}\mu = \int_A h\mathrm{d}P\mu,$$

and the Proposition follows. □

3. The above Proposition means that P induces semi-dynamical system on $\mathcal{M}$. Recall some definitions from the theory of dynamical systems. For any measure $\mu \in \mathcal{M}$, $\omega(\mu)$ denotes the ω-limit set of μ: $\omega(\mu) = \{\lambda \in \mathcal{M} : \exists t_i \to \infty, P^{t_i}\mu \to \lambda\}$. A compact set $\emptyset \neq \mathcal{K} \subset \mathcal{M}$ is invariant if $P(\mathcal{K}) \subset \mathcal{K}$. Let ϱ be a metric on $\mathcal{M}$ compatible with the topology. It is known and easy to see, that for any invariant set $\mathcal{K} \subset \mathcal{M}$: $\varrho(P^t\mu, \mathcal{K}) \to 0$ for $t \to \infty$, if and only if, $\omega(\mu) \neq \emptyset$ and $\omega(\mu) \subset \mathcal{K}$. As $\mathcal{M}$ is compact then, $\omega(\mu) \neq \emptyset$ for any $\mu \in \mathcal{M}$.

We recall a version of the famous Lyapunov Theorem on asymptotic stability, see [5] for a proof or [2] for a proof of its “continuous” counterpart.

THEOREM 2. *Let* (M, ϱ) *be a compact metric space,* $\emptyset \neq K \subset M$ *a compact and invariant set,* $P : M \to M$ *a continuous map. Let* $V : M \to \mathbb{R}$ *be a Lyapunov function, i.e.:*

1. *V is continuous,*
2. $V(x) = 0$*, for* $x \in K$*,*
3. $V(x) > 0$*, for* $x \in M \setminus K$*.*
4. *For every* $x \in M \setminus K$

$$V(P(x)) < V(x). \tag{11}$$

Then, for every $x \in M$*,*

$$\varrho(P^t x, K) \to 0, \text{ as } t \to \infty. \tag{12}$$

This theorem can be applied to our Foias operator as follows. Define:

$$\mathcal{M}^* = \{\mu \in \mathcal{M} : \operatorname{supp}\mu \subset A^*\}.$$

It is easy to see that $\mathcal{M}^*$ is a compact subset of $\mathcal{M}$ as A^* a compact subset of A. Also, $\mu \in \mathcal{M}^*$, if and only if, $\mu(A^*) = 1$. (If A^* is a singleton, so is $\mathcal{M}^*$. Otherwise, $\mathcal{M}^*$ is uncountable. In fact, if $a, b \in A^*$ are different points and $0 < p < 1$ then the measure μ defined by $\mu(\{a\}) = p$, $\mu(\{b\}) = 1 - p$, belongs to $\mathcal{M}^*$.)

Our main result, Theorem 1, is a consequence of the following:

THEOREM 3. *Assume that the set A is compact and conditions* (***A***), (***B***) *and* (***C***) *hold true. Then,* $\mathcal{M}^*$ *is invariant and for any measure* $\mu \in \mathcal{M}$:

$$\varrho(P^t \mu, \mathcal{M}^\star) \to 0, \text{ as } t \to \infty.$$

Proof. As mentioned above compactness of A implies compactness of $\mathcal{M}$. *It is easy to see that condition* (**B**) yields invariance of the set $\mathcal{M}^*$.

Define function $V: \mathcal{M} \to \mathbb{R}$:

$$V(\mu) = \int_A f \mathrm{d}\mu$$

to be a Lyapunov function. We are going to show, that the assumptions of Theorem 2 are satisfied.

Continuity of V is an immediate consequence of the definition of the topology on $\mathcal{M}$. Let $\mu_n \to \mu$. We put $h = f$ in (5) to get:

$$V(\mu_n) = \int_A f \mathrm{d}\mu_t \to \int_A f \mathrm{d}\mu = V(\mu).$$

Clearly $V(\mu) \geq 0$ for all $\mu \in \mathcal{M}$ and $V(\mu) = 0$ for all $\mu \in \mathcal{M}^*$. Let $V(\mu) = 0$ for some $\mu \in \mathcal{M}$. Then, we have: $0 = V(\mu) = \int_A f \mathrm{d}\mu = \int_{A^*} f \mathrm{d}\mu + \int_{A \setminus A^*} f \mathrm{d}\mu = \int_{A \setminus A^*} f \mathrm{d}\mu$. As f is strictly positive on $A \setminus A^*$, supp $\mu \subset A^*$, and hence $\mu \in \mathcal{M}^*$.

Let $\mu \in \mathcal{M} \backslash \mathcal{M}^*$. Condition (**C**) says that for any $x \in A \backslash A^*$:

$$\int_B f(T(x, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) < f(x) \tag{13}$$

and then by (9) and (10) and the choice of the measure μ.

$$V(P\mu) = \int_A f \mathrm{d}P\mu = \int_A Uf \mathrm{d}\mu = \int_A \left(\int_B f(T(x, \mathbf{y}))\mathrm{d}\nu(\mathbf{y}) \right) \mathrm{d}\mu(x) < \int_A f \mathrm{d}\mu = V(\mu)$$

Theorem 2 completes the proof. □

4. *Proof of Theorem* 1. We will interpret the above Theorem 3 in terms of random variables X_t. Note first that for any measure $\mu^* \in \mathcal{M}^*$ and any set $C \in \mathcal{B}(A)$ such that $A^* \subset \operatorname{int} C$ we have $\mu^*(\delta C) = 0$ and $\mu^*(C) = 1$, and then condition (6) implies that for any sequence of measures $\mu_n \in \mathcal{M}$ such thar $\mu_n \to \mu^*$ we have

$$\mu_n(C) \to 1, \text{ as } n \to \infty$$

In terms of random variables it can be expressed by Lemma 1 as follows. Let $B(A^*, \varepsilon) = \{a \in A : \operatorname{dist}(a, A^*) < \varepsilon\}$. Fix any measure $\mu_0 \in \mathcal{M}$. Theorem 3 guarantees that the ω-limit set, $\omega(\mu_0)$, is nonempty and is contained in $\mathcal{M}^*$. Hence, for any sequence $t_n \to \infty$, there exists a subsequences $t_{n_i} \to \infty$ and a measure $\mu^* \in \mathcal{M}^*$ such that $P^{t_{n_i}}\mu_0 \to \mu^*$ and hence $P^{t_{n_i}}\mu_0(B(A^*, \varepsilon)) \to \mu^*(B(A^*, \varepsilon)) = 1$. But this means that $P^t\mu_0(B(A^*, \varepsilon)) \to 1$, as $t \to \infty$. So by Lemma 1 we have: for every $\varepsilon > 0$:

$$\lim_{t \to \infty} \operatorname{Prob}(\operatorname{dist}(X_t, A^*) < \varepsilon) = 1, \tag{14}$$

what completes the proof of Theorem 1. □

5. We show an application of Theorem 1. One of the numerical methods for finding an approximation of the set A^* is an evolutionary algorithm which can be described as follows. As before we assume that A is a metric space, $f : A \to \mathbb{R}$ is a continuous function.

Consider measures $\mu_0, \nu_0 \in \mathcal{M}$ and let k, m be natural numbers. Let $\varphi : A \to A$ be a map such that A^* is invariant under φ, i.e. $\varphi(A^*) \subset A^*$. We call such φ a *local method*.

The Algorithm.

1. Choose an initial population, i.e. a simple sample of points from A distributed according to μ_0:

$$\mathbf{x} = (x_1, \ldots, x_m) \in A^m$$

2. Draw a simple sample $\mathbf{y} = (y_1, \ldots, y_k) \in A^k$ according to the distribution ν_0.
3. Apply φ to each x_i and y_i to get

$$(\varphi(x_1), \ldots, \varphi(x_m), \varphi(y_1), \ldots, \varphi(y_k))$$

4. Sort this sequence using f as a criterion to get

$$(\bar{x}_1, \ldots, \bar{x}_{m+k}) \text{ with } f(\bar{x}_1) \leq \ldots \leq f(\bar{x}_{m+k}).$$

5. Form the next population with the first m points

$$\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_m)$$

and go to point (2) with $\mathbf{x} = \bar{\mathbf{x}}$.

Repeat according to a stopping rule.

There are a number of local methods, φ, available. For example, a classical one is the gradient method. It requires differentiability of the objective function f still it is quite effective in finding local minima attained at interior points of the set A. If f is not a smooth function or its local minimum point are at the boundary of A, then more sophisticated method can be used, see Ref. [7] and survey paper [9]. Obviously, the identity map is a local method.

It is easy to describe the Algorithm in form (1). To simplify notations we will assume in the sequel $m = 1$. The results can be easily repeated with $m > 1$. Define the map $T : A \times A^k \to A$ as follows. Let $(x, \mathbf{y}) \in A \times A^k$, $\mathbf{y} = (y_1, \ldots, y_k)$. Now we put

$$T(x, \mathbf{y}) = \begin{cases} \varphi(x), & \text{if for all } i = 1, \ldots, k \ f(\varphi(x)) < f(\varphi(y_i)) \\ \varphi(y_{i_0}), & \text{otherwise} \end{cases} \tag{15}$$

where i_0 is the smallest number such that for all $i = 1, \ldots, k \ f(\varphi(y_{i_0})) \leq f(\varphi(y_i))$.

So we see that the Algorithm yields form (1) with $B = A^k$ and $\nu = \nu_0^k$.

The following theorem gives sufficient conditions for stochastic convergence of the above evolutionary algorithm to the solution of the global minimization problem. A similar result has been established in [5], where more direct proof was presented under assumption that A was compact and φ fulfilled an extra condition.

THEOREM 4. *Assume that:*

(**a**1) $\nu_0(l_c) = 0$ *for any level curve of* f, $l_c := \{x \in A : f(x) = c\}$.
(**a**2) *The local method* φ *is continuous.*
(**a**3) *The local method* φ *is* ν_0*-nonsingular, i.e.*

$$\nu_0(C) = 0 \Rightarrow \nu_0(\varphi^{-1}(C)) = 0 \quad \text{for any } C \in \mathcal{B}(A), \tag{16}$$

(**a**4) *If* $G \subset A$ *is a neighourhood of* A^*, *then* $\nu_0(G) > 0$.
(**a**5) *There exists* $r > 0$ *such that set* $A_r := \{x \in A : f(x) \leq r\}$ *is compact and supp* $\mu_0 \subset A_r$.*Then, for every* $\varepsilon > 0$:

$$\lim_{t \to \infty} \text{Prob}(\text{dist}(X_t, A^*) < \varepsilon) = 1. \tag{17}$$

Let A^* be a singleton a^*, i.e. the global optimization problem has a unique solution a^*. Under the above assumptions we have:

COROLLARY 2. *For every* $\varepsilon > 0$ *and any norm on* $\mathbb{R}^n$:

$$\lim_{t \to \infty} \text{Prob}(\|X_t - a^\star\| < \varepsilon) = 1. \tag{18}$$

In other words the sequence X_t *stochastically converges to the solution* a^*.

Proof. It is enough to show that conditions (**a**1)–(**a**5) imply conditions (**A**)–(**E**).

We prove (**A**). Fix $x_0 \in A$ and a sequence $x_n \to x_0$. Consider the level curve $l = l_{f(\varphi(x_0))}$. By Assumption (**a**1) $\nu_0(l) = 0$ and as φ is ν_0-nonsingular, $\nu_0(\varphi^{-1}(l)) = 0$. Fix $\mathbf{y} = (y_1, \ldots, y_k) \in B$ such that for all $i = 1, \ldots, k$ $y_i \notin \varphi^{-1}(l)$. We claim that $T(x_n, \mathbf{y}) \to T(x_0, \mathbf{y})$. In fact, if $f(\varphi(x_0)) < \min(f(y_1), \ldots, f(y_k))$, then by continuity of f and φ for large n's also $f(\varphi(x_n)) < \min(f(y_1), \ldots, f(y_k))$ and then $T(x_n, \mathbf{y}) = \varphi(x_n) \to \varphi(x_0) = T(x_0, \mathbf{y})$. Otherwise, $f(\varphi(x_0)) > f(y_{i_0})$, where i_0 is the smallest index such that $f(y_{i_0}) = \min(f(y_1), \ldots, f(y_k))$. Like above, for large n's $f(y_{i_0}) < f(\varphi(x_n))$ and hence $T(x_n, \mathbf{y}) = \varphi(y_{i_0}) = \varphi(x_0) = T(x_0, \mathbf{y})$, which proves the claim. Now, $T(x_n, \cdot) \to T(x_0, \cdot)$ on the set of full measure ν^k as required.

Conditions (**B**) follows from the definition of the local method.

We prove (**C**). Let $x \in A \backslash A^*$. If $\varphi(x) \in A^*$, then also $T(x, \mathbf{y}) \in A^*$, hence $f(T(x, \mathbf{y})) < f(x)$ for all $\mathbf{y} \in B$. So assume now that $\varphi(x) \notin A^*$. We then have $f(\varphi(x)) > 0$. As for any $y \in A^*, f(\varphi(y)) \leq f(y) = 0$, then, by continuity of $f \circ \varphi$ and compactness of A^* there exists G, a neighborhood of A^*, such that $f(\varphi(y)) < f(\varphi(x))$ for $y \in G$, and hence $f(T(x, \mathbf{y})) < f(x)$ for all $\mathbf{y} \in B$ having at least one coordinate $y_i \in G$, i.e. for $\mathbf{y} \in B \backslash (A \backslash G)^k$. By condition (**a**4) we have $\nu_0(G) > 0$, and hence $\nu(B \backslash (A \backslash G)^k) > 0$. In the both cases we have:

$$f(T(x, \mathbf{y})) < f(x), \quad \text{for all } \mathbf{y} \text{ from a set of positive measure } \nu, \tag{19}$$

By the definition of T we clearly see that $f(T(x, \mathbf{y})) \leq f(x)$ for all $\mathbf{y} \in B$. It implies that:

$$\int_B f(T(x, \mathbf{y}))\nu(\mathrm{d}\mathbf{y}) < \int_B f(x)\nu(\mathrm{d}\mathbf{y}) = f(x), \tag{20}$$

what is required. Conditions (**D**) follows from the description of the map T. Conditions (**E**) is just the same as (**a**5).

Theorem 1 and Corollary 1 complete the proof. □

Let us note that (**a1**)–(**a5**) are rather mild conditions. Assume that A is a compact set of finite dimensional space, say $A \subset \mathbb{R}^n$, and the objective function f is defined on some neighborhood of A. It seems that the measures ν_0 absolutely continuous with respect to the Lebesgue measure meet these requirements with most functions being optimized in practise. In fact, if function f is of class C^1 and $a \in \mathbb{R}^n$ is not a critical point of f, then the level curve passing through a, $l_{f(a)}$, is locally a submanifold of $\mathbb{R}^n$ of dimension $n - 1$ and as such has its Lebesgue measure zero. So, if f is a Morse function, then (**a1**) is satisfied. Also, (**a3**) is satisfied if ν_0 is absolutely continuous with respect to the Lebesgue measure and φ is a local diffeomorphism. Condition (**a4**) is satisfied if ν_0 has a density with respect to the Lebesgue measure which is positive on a neighborhood of A, e.g. it is non degenerate normal. Also, most real situation correspond to assumption (**a5**).

6. Theorem 1 and Corollary 1 have nice interpretation in a very special case when the map T does not depend on the second variable $\mathbf{y}$. Let (X, ϱ) be a metric space. Let $S : X \to X$ be a continuous function. A nonempty set $K \subset X$ is globally asymptotically stable iff: (i) It is stable, i.e. for every $\varepsilon > 0$ there exits $\delta > 0$ such that if $\varrho(x, K) \leq \delta$, then for all $n \geq 0$ $\varrho(S^n(x), K) \leq \varepsilon$, and (ii) For every $x \in X$ $S^n(x) \to K$ as $n \to \infty$. Let X_0 be a random variable on X distributed according to a probability measure μ_0. Then we can form sequence of random variables:

$$X_n = S(X_{n-1}), \quad n = 1, 2, 3, \ldots$$

We can prove the following quite natural folklore result:

THEOREM 5. *Let X be a locally compact metric space and $S : X \longrightarrow X$ a continuous function. Let nonempty compact set $K \subset X$ be a globally asymptotically stable set. Let supp μ_0 be a compact set. Then: for every $\varepsilon > 0$:*

$$\lim_{n \to \infty} \text{Prob}(\text{dist}(X_n, K) < \varepsilon) = 1. \tag{21}$$

Proof. As K is globally asymptotically stable there exists a Lyapunov function, say $f : X \longrightarrow [0, \infty)$, such that: (a) f is continuous, (b) $f(x) = 0$ iff $x \in K$, (c) $f(S(x)) < f(x)$ for $x \notin K$ (d) For every $c > 0$ there exists a compact set L such that $f(x) \geq c$ for $x \notin L$, see [1] and [2]. Now, we can apply Theorem 1 and Corollary 1 with $A = X$, an arbitrary set B and any measure μ, and $T(x, \mathbf{y}) := S(x)$. As supp μ_0 is compact and f is continuous, there exists $r > 0$ such that supp $\mu_0 \subset A_r = \{x \in A : f(x) \leq r\}$. Putting $c > r$ in (d) we have a compact set L such that $A_r \subset L$. So A_r is compact and condition (**E**) holds true. Since conditions (**A**)–(**D**) are also evidently satisfied the proof is complete. □

References

[1] Berger, A. and Siegmund, S., 2003, On the gap between randon dynamical systems and continuous skew products. *Journal of Dynamics and Differential Equations*, **15**, 237–279.
[2] Bhatia, N. and Szegö, G., 2002, *Stability Theory of Dynamical Systems* (New York: Springer).
[3] Lasota, A. and Mackey, M., 1994, *Chaos, Fractals and Noise* (New York: Springer-Verlag).
[4] Meyn, S. and Twedie, R., 1993, *Markov Chains and Stochastic Stability* (New York: Springer-Verlag).
[5] Ombach, J., Stability of evolutionary algorithms, preprint.
[6] Parthasarathy, K., 1967, *Probability Measures on Metric Spaces* (New York: Academic Press).
[7] Robert, Ch. and Casella, G., 2004, *Monte Carlo Statistical Methods* (New York: Springer).
[8] Słomczyński, W., 2003, *Dynamical Entropy, Markov Operators and Iterated Functions Systems* (Krakow: Wydawnictwo Uniwersytetu Jagillońskiego).
[9] Wright, M., 2005, The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bulletin of the American Mathematical Society*, **42**, 39–56.