Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 342 (2008) 326–333
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

# Stability of evolutionary algorithms

Jerzy Ombach [a,b,*]

[a] Instytut Matematyki, Uniwersytet Jagielloński, ul. Reymonta 4, 30 059 Kraków, Poland
[b] Państwowa Wyższa Szkoła Zawodowa, ul. Chruślicka 6, 33 300 Nowy Sącz, Poland

Received 7 July 2005

Available online 8 December 2007

Submitted by C. Simó

**Abstract**

We prove under mild conditions the convergence of some evolutionary algorithm to the solution of the global optimization problem. In the proof, the Lyapunov function's techniques is applied to some semi-dynamical system generated by a Foias operator on the space of the probability measures defined on the set of admissible solutions.


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

One of the most important problems in applied mathematics is to find a *global* minimum (or maximum) of a given real valued function $f$ considered on a given set $A \subset \mathbb{R}^n$. This problem was successfully solved under some restrictive assumptions like linearity, convexity, etc. on the objective function $f$ and the set of admissible solutions $A$. On the other hand, when $f$ or $A$ are not regular enough, the classical methods fail. One of the most important problems while applying such methods is to be not trapped by a local minimum/maximum. Yet, recent years have been witnessing a significant progress. Taking advantage of the increasing computational power new methods have been recently developed, implemented and successfully used, proving in practice its flexibility. Some of them are known as genetic, evolutionary and hybrid algorithms. They are relatively slow but numerical experiments exhibit their convergence to the solutions of global optimization problems. In particular, the problem of being trapped by a local optimum has been settled, to some extent, by the stochastic factor of those algorithms. There are some mathematical justifications of this phenomena based, generally, on the theory of classical Markov Chains and Markov Processes, see for example [6,8] and for some specific situation [3].

The aim of this paper is to demonstrate an alternative method of proving the convergence of some evolutionary and hybrid algorithms to the solution of the global optimization problems. In the framework of Markov operators on the space of measures we will use some concepts and results of the classical stability theory of dynamical systems. We concentrate on an algorithm which can be classified as a two-phase multistart, see [4]. The algorithm presented below is a kind of compromise between deterministic local search and pure stochastic search algorithms. It appears that it can be thought of as a Foias operator $P : \mathcal{M} \to \mathcal{M}$, where $\mathcal{M}$ is the set of all probability measures on $A$, and

* Address for correspondence: Instytut Matematyki, Uniwersytet Jagielloński, ul. Reymonta 4, 30 059 Kraków, Poland.
*E-mail address:* ombach@im.uj.edu.pl.

0022-247X/$ – see front matter 
doi:10.1016/j.jmaa.2007.12.006

$P$ is indeed a Feller operator, which means that it generates a discrete time semi-dynamical system on $\mathcal{M}$. Let $A^\star$ be the set of all the solutions of the global minimization problem. We will show that the set $\mathcal{M}^\star \subset \mathcal{M}$ of measures supported on $A^\star$ is globally asymptotically stable in this system, Theorem 6. Its proof relies on showing that the function $V : \mathcal{M} \ni \mu \to \int_A f\, d\mu \in \mathbb{R}$ is a strict Lyapunov function. As a consequence, the random variables resulting in the evolutionary algorithm are stochastically convergent to the set of solutions of the problem, $A^\star$, Corollaries 9 and 10. Our assumptions on the objective function $f$ and the set of admissible solutions $A$ are rather mild. In particular, neither uniqueness of the optimal solution nor differentiability of $f$ is required.

Section 1 is devoted to a general setting of the algorithm on a compact subset $A$ of $\mathbb{R}^n$ and to its reformulation in the language of a semi-dynamical system on $\mathcal{M}$. In Section 2 we state and prove our main result of global asymptotic stability of the set $\mathcal{M}^\star$ and then we will interpret this result in more classical terminology of stochastic convergence. The aim of Section 3 is to recall and adapt to our needs some known results on asymptotic stability and Lyapunov functions. In Section 4 we discuss some particular cases and discuss the meaning of our assumptions and possible extensions of the results of Section 2.

## 1. Evolutionary algorithms and Foias operators

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

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

*1.* One of the numerical methods for finding an approximation of the set $A^\star$ is an evolutionary algorithm which can be described as follows.

Let $\nu$ be a probability distribution on $A$, let $k$ and $m$ be natural numbers. Let $\varphi : A \to A$ be a measurable map satisfying:

$$\forall x \in A \quad f(\varphi(x)) \leqslant f(x). \tag{1}$$

We call such $\varphi$ a *local method*. Note that $A^\star$ is invariant under $\varphi$, i.e., $\varphi(A^\star) \subset A^\star$.

**Algorithm.**

(1) Choose an initial population, i.e., a sequence of points from $A$,

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

(2) Draw a simple sample $\mathbf{y} = (y_1, \dots, y_k) \in A^k$ according to the distribution $\nu$.

(3) Apply $\varphi$ to each $x_i$ and $y_j$ to get

$$(\varphi(x_1), \dots, \varphi(x_m), \varphi(y_1), \dots, \varphi(y_k)).$$

(4) Sort this sequence using $f$ as a criterion to get

$$(\bar{x}_1, \dots, \bar{x}_{m+k}) \quad \text{with } f(\bar{x}_1) \leqslant \cdots \leqslant f(\bar{x}_{m+k}).$$

(5) Form the next population with the first $m$ points

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

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

Repeat according to a stopping rule.

It is believed that under some natural assumptions points $\bar{x}_1, \dots, \bar{x}_m$, and in particular the first one, $\bar{x}_1$, from subsequent populations approximate the global minimum of $f$.

*2.* There are a number of local methods 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 is at the boundary of $A$, then more sophisticated method can be used, see [8] and survey paper [10]. Most of them satisfy condition (1) above. Let us note, that the identity function also satisfies (1) and hence it is a local method.

*3.* In the sequel we describe the Algorithm in terms of a semi-dynamical system with random perturbations. To simplify notations we will assume in the sequel that $m = 1$. The results and their proofs can been easily repeated with $m > 1$.

Fix a positive integer $k$ and define the map $T : A \times A^k \to A$ as follows. Let $(x, \mathbf{y}) \in A \times A^k$, $\mathbf{y} = (y_1, \dots, y_k)$. Now we put

$$T(x, \mathbf{y}) = \begin{cases} \varphi(x), & \text{if for all } i = 1, \dots, k \ f(\varphi(x)) < f(\varphi(y_i)), \\ \varphi(y_{i_0}), & \text{otherwise} \end{cases} \tag{2}$$

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

Let $\mathcal{B}(A)$ denote the family of Borel subsets of the set $A$. We will assume that the measure $\nu$ is defined on this $\sigma$-algebra. Denote by $\nu^k$ the product measure on the $\sigma$-algebra $\mathcal{B}(A^k)$. Let $\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \dots$ be the sequence of independent random vectors defined on a probability space $(\Omega, \Sigma, \mathrm{Prob})$, $\mathbf{Y}_i : \Omega \to A^k$, identically distributed with distribution $\nu^k$ each. In particular this means, that coordinates of each random vector $\mathbf{Y}_i$: $Y_{i1}, \dots, Y_{ik}$ are independent random variables with distribution $\nu$ each. In other words, their realization $y_{i1}, \dots, y_{ik}$ is a simple sample from the distribution $\nu$.

Fix $x_0 \in A$ and define the variables $X_t$:

$$X_0 \equiv x_0 \quad \text{and} \quad X_t = T(X_{t-1}, \mathbf{Y}_{t-1}) \quad \text{for } t = 1, 2, 3, \dots. \tag{3}$$

The following is a simple observation:

**Lemma 1.** *The map $T$ is measurable and then $X_t$ defined above are random variables on the probability space $(\Omega, \Sigma, \mathrm{Prob})$.*

We are interested in the distributions of $X_t$, i.e., in the measures $\mu_t^T$ defined as

$$\mu_t^T(C) = \mathrm{Prob}(X_t \in C) = \mathrm{Prob}\big(X_t^{-1}(C)\big), \quad C \in \mathcal{B}(A), \tag{4}$$

and in their convergence to a measure or to a set of measures supported on the set of the solutions of the global minimization problem. In the sequel we will be able to consider this sequence as a positive orbit of some semi-dynamical system and using concepts and terminology from the theory of dynamical system we will obtain appropriate results about the convergence of sequences $\{\mu_t^T\}$ and more generally, about behavior of such sequences.

*4.* Let $\mathcal{M}$ denote the set of all probability measures on $\mathcal{B}(A)$. It is known, see [7], that $\mathcal{M}$ with the Fortet–Mourier metric is a compact metric space and its topology is determined by the weak convergence of the sequences of measures as follows. The sequence $\mu_t \in \mathcal{M}$ converges to $\mu \in \mathcal{M}$ if and only if for any continuous (so bounded by compactness of $A$) function $h$

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

as $t \to \infty$. We will use the following equivalent condition for weak convergence, see for example [1, p. 200]:

$$\text{for every } C \in \mathcal{B}(A) \text{ such that } \mu(\delta C) = 0, \quad \mu_t(C) \to \mu(C), \tag{6}$$

as $t \to \infty$, where $\delta C$ denotes the boundary of $C$.

We will use the concept of the Foias operator and its simple consequences summarized in the two lemmas below, for more details see for example Chapter 12 in [5].

The map $T$ defined above induces the Foias operator $P : \mathcal{M} \to \mathcal{M}$, as follows:

$$P\mu(C) = \int_A \left( \int_{A^k} \mathrm{I}_C\big(T(x, \mathbf{y})\big) \nu^k(d\mathbf{y}) \right) \mu(dx), \quad \text{for } \mu \in \mathcal{M},\ C \in \mathcal{B}(A), \tag{7}$$

where $\mathrm{I}_C$ is the indicator function of the set $C$.

We will be interested in the successive iterations $P^t$ of the Foias operator, i.e., maps defined as $P^0 = \mathrm{id}_{\mathcal{M}}$, $P^{t+1} = P \circ P^t$, for $t = 0, 1, 2, \ldots$.

**Lemma 2.** *Let $\delta_{x_0}$ denote the $\delta$-Dirac measure concentrated at $x_0 \in A$. Then, for every $t = 1, 2, 3, \ldots$,*

$$\mu_t^T = P^t \delta_{x_0}, \tag{8}$$

*where the measures $\mu_t^T$ are defined by* (4).

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

$$Uh(x) = \int_{A^k} h\bigl(T(x, \mathbf{y})\bigr) \nu^k(d\mathbf{y}). \tag{9}$$

**Lemma 3.** *Let $\mu \in \mathcal{M}$. If $h$ is continuous, then*

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

*5.* In the sequel we assume that:

(A1) The set $A$ is compact and the map $f$ is continuous.
(A2) $\nu(l_c) = 0$ for any level curve of $f$, $l_c := \{x \in A\colon f(x) = c\}$.
(A3) The local method $\varphi$ is continuous.
(A4) The local method $\varphi$ is $\nu$-nonsingular i.e.,

$$\nu(C) = 0 \quad \Rightarrow \quad \nu\bigl(\varphi^{-1}(C)\bigr) = 0 \quad \text{for any } C \in \mathcal{B}(A). \tag{11}$$

**Proposition 4.** *Under the above assumptions the Foias operator $P : \mathcal{M} \to \mathcal{M}$ is continuous.*

**Proof.** Let $\mu_t \to \mu$, $\mu_t, \mu \in \mathcal{M}$. Let $h : A \to \mathbb{R}$ be a continuous function. We have to show that $\int_A h\, d(P\mu_t) \to \int_A h\, d(P\mu)$. Let $x_0 \in A$ and $x_s \to x_0$ be fixed. Consider the level curve $l = l_{f(\varphi(x_0))}$. By assumption (A2) $\nu(l) = 0$ and as $\varphi$ is $\nu$-nonsingular, $\nu(\varphi^{-1}(l)) = 0$. Fix $\mathbf{y} = (y_1, \ldots, y_k) \in A^k$ such that for all $i = 1, \ldots, k$ $y_i \notin \varphi^{-1}(l)$. We claim that $T(x_s, \mathbf{y}) \to T(x_0, \mathbf{y})$. In fact, if $f(\varphi(x_0)) < \min(f(y_1), \ldots, f(y_k))$, then for large $s$ also $f(\varphi(x_s)) < \min(f(y_1), \ldots, f(y_k))$ and then $T(x_s, \mathbf{y}) = \varphi(x_s) \to \varphi(x_0) = T(x_0, \mathbf{y})$. 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 $s$ $f(y_{i_0}) < f(\varphi(x_s))$ and hence $T(x_s, \mathbf{y}) = \varphi(y_{i_0}) = \varphi(x_0) = T(x_0, \mathbf{y})$, which proves the claim. Now, $T(x_s, \cdot) \to T(x_0, \cdot)$ and so $h(T(x_s, \cdot)) \to h(T(x_0, \cdot))$ on the set of full measure $\nu^k$ and the Dominated Convergence Theorem may be applied. We then have

$$Uh(x_s) = \int_{A^k} h\bigl(T(x_s, \mathbf{y})\bigr) \nu^k(d\mathbf{y}) \to \int_{A^k} h\bigl(T(x_0, \mathbf{y})\bigr) \nu^k(d\mathbf{y}) = Uh(x_0),$$

so the function $Uh$ is continuous. By (5),

$$\int_A h\, dP\mu_t = \int_A Uh\, d\mu_t \to \int_A Uh\, d\mu = \int_A h\, dP\mu,$$

and the proposition follows. $\square$

**Remark 5.** Proposition 4 we have just proved actually says that under the above assumptions the operator $P$ is a Feller operator, see [5,6] or [9].

We complete this section noting that the set

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

is a compact subset of $\mathcal{M}$ as $A^\star$ a compact subset of $A$. Also, if $\mu \in \mathcal{M}^\star$, then $P\mu \in \mathcal{M}^\star$. In fact, by (1) $\mathrm{I}_{A^\star}(T(x, \mathbf{y})) \geqslant \mathrm{I}_{A^\star}(T(x))$, for all $x \in A^\star$, $\mathbf{y} \in A^k$, hence by (7) $P\mu(A^\star) \geqslant \mu(A^\star) = 1$.

If $A^\star$ is a singleton, so is $\mathcal{M}^\star$. Otherwise, $\mathcal{M}^\star$ is uncountable. In fact, if $a, b \in A^\star$ are different points and $0 < p < 1$ then the measure $\mu$ defined by $\mu(\{a\}) = p$, $\mu(\{b\}) = 1 - p$ belongs to $\mathcal{M}^\star$.

## 2. Global asymptotic stability of $\mathcal{M}^\star$

*1.* Our main result stated in the following theorem is expressed in terms of stability theory of dynamical systems, see Section 3 for the appropriate definitions.

**Theorem 6.** *Assume conditions* (A1) *to* (A4) *and*

(A5) *If $G \subset A$ is a neighbourhood of $A^\star$, then $\nu(G) > 0$.*

*Then, $\mathcal{M}^\star$ is globally asymptotically stable in the semi-dynamical system induced on $\mathcal{M}$ by the Foias operator $P$.*

**Proof.** In the proof of our theorem we will use the Lyapunov Stability Theorem adjusted to our situation, see Section 3 for its statement and a proof. Define function $V : \mathcal{M} \to \mathbb{R}$:

$$V(\mu) = \int_A f\, d\mu$$

to be a Lyapunov function. We are going to show, that all assumptions of point (iv) of Theorem 11 are satisfied.

We have already noted that $\mathcal{M}^\star$ is invariant under $P$. We prove continuity of $V$. Let $\mu_t \to \mu$. We put $h = f$ in (5) to get

$$V(\mu_t) = \int_A f\, d\mu_t \to \int_A f\, d\mu = V(\mu).$$

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

Let $\mu \in \mathcal{M} \setminus \mathcal{M}^\star$. We will show that $V(P\mu) < V(\mu)$. Let $x \in A \setminus A^\star$. Thus $f(\varphi(x)) \leqslant f(x)$. If $\varphi(x) \in A^\star$, then also $T(x, \mathbf{y}) \in A^\star$, hence $f(T(x, \mathbf{y})) < f(x)$ for all $\mathbf{y} \in A^k$. So assume now that $\varphi(x) \notin A^\star$. We then have $f(\varphi(x)) > 0$. As for any $y \in A^\star$, $f(\varphi(y)) \leqslant f(y) = 0$, then, by continuity of $f \circ \varphi$ and compactness of $A^\star$ there exists $G$, a neighborhood of $A^\star$, such that $f(\varphi(y)) < f(\varphi(x))$ for $y \in G$, and hence $f(T(x, \mathbf{y})) < f(x)$ for all $\mathbf{y} \in A^k$ having at least one coordinate $y_i \in G$, i.e., for $\mathbf{y} \in A^k \setminus (A \setminus G)^k$. By our assumption (A5) $\nu(G) > 0$, and hence $\nu(A^k \setminus (A \setminus G)^k) > 0$. In the both cases for any fixed $x \in A \setminus A^\star$ we have

$$f\bigl(T(x, \mathbf{y})\bigr) < f(x), \quad \text{for all } \mathbf{y} \text{ from a set of positive measure } \nu^k, \tag{12}$$

with $f(T(x, \mathbf{y})) \leqslant f(x)$ for all $x \in A$ and $\mathbf{y} \in A^k$.

It follows that for any $x \in A \setminus A^\star$:

$$\int_{A^k} f\bigl(T(x, \mathbf{y})\bigr)\nu^k(d\mathbf{y}) < \int_{A^k} f(x)\nu^k(d\mathbf{y}) = f(x) \tag{13}$$

and then by (10) and (9) and the choice of the measure $\mu$

$$V(P\mu) = \int_A f\, dP\mu = \int_A Uf\, d\mu = \int_A \biggl(\int_{A^k} f\bigl(T(x, \mathbf{y})\bigr)\nu^k(d\mathbf{y})\biggr)\mu(dx) < \int_A f\, d\mu = V(\mu).$$

Point (iv) of Theorem 11 completes the proof. □

*2.* Now we will interpret the above theorem in terms of random variables $X_t$ defined by (3) resulting from the Algorithm. Note first that for any measure $\mu^\star \in \mathcal{M}^\star$ and any set $C \in \mathcal{B}(A)$ such that $A^\star \subset \operatorname{int} C$ we have $\mu^\star(\delta C) = 0$ and $\mu^\star(C) = 1$, and then condition (6) implies that for any sequence of measures $\mu_t \in \mathcal{M}^\star$, $\mu_t \to \mu^\star$ we have

$$\mu_t(C) \to 1, \quad \text{as } t \to \infty.$$

If the optimization problem has an unique solution, say $a^\star$, then the set $\mathcal{M}^\star$ is a singleton $\delta_{a^\star}$ and Theorem 6 then says that $P^t\mu \to \delta_{a^\star}$, as $t \to \infty$, for any probability measure in $\mathcal{M}$. So we have

**Corollary 7.** *Let $a^\star$ be an unique solution of the minimization problem. Let $\mu$ be a measure in $\mathcal{M}$. Then for every $C \in \mathcal{B}(A)$ such that $a^\star \in \operatorname{int} C$,*

$$\lim_{t\to\infty} P^t\mu(C) = 1.$$

In terms of random variables it can be expressed by Lemma 2 as follows.

**Corollary 8.** *Let $a^\star$ be a unique solution of the minimization problem. Let $x_0 \in A$ and let $X_t$ be the sequence of random variables resulted from the Algorithm as in* (3). *Then for every $C \in \mathcal{B}(A)$ such that $a^\star \in \operatorname{int} C$,*

$$\lim_{t\to\infty} \operatorname{Prob}(X_t \in C) = 1.$$

Putting $C = B(a^\star, \varepsilon) = \{a \in A\colon\ \|a - a^\star\| < \varepsilon\}$ for $\varepsilon > 0$ we have

**Corollary 9.** *For every $\varepsilon > 0$ and any norm on $\mathbb{R}^n$*

$$\lim_{t\to\infty} \operatorname{Prob}\bigl(\|X_t - a^\star\| < \varepsilon\bigr) = 1. \tag{14}$$

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

If the optimization problem has more than one solution, then $\mathcal{M}^\star$ is an uncountable set of measures. Still, our theorem yields similar result to that one in the previous case. Let $B(A^\star, \varepsilon) = \{a \in A\colon\ \operatorname{dist}(a, A^\star) < \varepsilon\}$. Fix any measure $\mu \in \mathcal{M}$. Theorem 6 guarantees that the $\omega$-limit set, $\omega(\mu)$, is nonempty and is contained in $\mathcal{M}^\star$. Hence, for any sequence $t_n \to \infty$, there exist subsequences $t_{n_i} \to \infty$ and a measure $\mu^\star \in \mathcal{M}^\star$ such that $P^{t_{n_i}}\mu \to \mu^\star$ and hence $P^{t_{n_i}}\mu(B(A^\star, \varepsilon)) \to \mu^\star(B(A^\star, \varepsilon)) = 1$. But this means that $P^t\mu(B(A^\star, \varepsilon)) \to 1$, as $t \to \infty$. So we have

**Corollary 10.** *For every $\varepsilon > 0$ and every $x_0 \in A$*

$$\lim_{t\to\infty} \operatorname{Prob}\bigl(\operatorname{dist}(X_t, A^\star) < \varepsilon\bigr) = 1. \tag{15}$$

## 3. Lyapunov Stability Theorem in metric spaces

In this section we assume that $(X, \varrho)$ is a metric space and $P : X \to X$ is a continuous map. The map $P$ determines a semi-dynamical system on $X$ and the orbit $o(x)$ of a point $x \in X$ is the set $o(x) = \{P^t(x)\colon\ t = 0, 1, 2, \ldots\}$, where $P^t$ is the $t$th iteration of $P$, i.e., $P^0 = \mathrm{id}_X$, $P^{t+1} = P \circ P^t$, for $t = 0, 1, 2, \ldots$. Let $K \subset X$ be a compact set, invariant under $P$, i.e., $P(K) \subset K$. For any $\varepsilon > 0$ we consider an $\varepsilon$-neighborhood of $K$, $B(K, \varepsilon) = \{x \in X\colon\ \varrho(x, K) < \varepsilon\}$. Let $\omega(x)$ denote the set of limit points of $x$, i.e., $\omega(x) = \{y \in X\colon\ \exists t_i \to \infty,\ P^{t_i}(x) \to y\}$. It is known and easy to see, that $\varrho(P^t(x), K) \to 0$ for $t \to \infty$, if and only if, $\omega(x) \neq \emptyset$ and $\omega(x) \subset K$. If $X$ is compact then, $\omega(x) \neq \emptyset$ for any $x \in X$.

*1.* We say that a point $x$ is attracted to $K$, if $\varrho(P^t(x), K) \to 0$ for $t \to \infty$.

We say that $K$ is stable if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in B(K, \delta)$, then for all $t \geqslant 0$ $P^t(x) \in B(K, \varepsilon)$.

We say that $K$ is asymptotically stable, if $K$ is stable and there exists $\delta_0 > 0$ such that any $x \in B(K, \delta_0)$ is attracted to $K$.

We say that $K$ is globally asymptotically stable, if $K$ is stable and any $x \in X$ is attracted to $K$.

*2.* In the following theorem we summarize results known as the Lyapunov Stability Theorem. For reader convenience we provide a proof that summarizes classical proofs of the Lyapunov Stability Theorem. It seems, that the closest versions to the statements of the theorem below can be found in [2, Section V.4], where continuous-time dynamical systems on locally compact metric spaces were considered.

**Theorem 11.** *Let $X$ be a metric space, $\emptyset \neq K \subset X$ a compact and invariant set, $P: X \to X$ a continuous function. Let $V: X \to \mathbb{R}$ be a Lyapunov function, i.e.*:

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

$$V\big(P(x)\big) \leqslant V(x). \tag{16}$$

*Then*:

(i) *If $X$ is locally compact, then $K$ is stable.*
(ii) *If for every $x \in X \setminus K$*

$$V\big(P(x)\big) < V(x), \tag{17}$$

*then, for any point $x \in X$ either $\omega(x) = \emptyset$ or $x$ is attracted to $K$.*
(iii) *If $X$ is locally compact and condition* (17) *is satisfied, then $K$ is asymptotically stable.*
(iv) *If $X$ is compact and condition* (17) *is satisfied, then $K$ is globally asymptotically stable.*

**Proof.** (i) Let $\varepsilon > 0$ be such that the closure $\overline{B(K,\varepsilon)}$ is compact. Let $K_\varepsilon = \overline{B(K,\varepsilon)} \setminus P^{-1}(B(K,\varepsilon))$. We consider two cases. (1) $K_\varepsilon = \emptyset$. In this case we put $\delta = \varepsilon$. Clearly, by our assumption if $x \in B(K,\varepsilon)$, then also $P(x) \in B(K,\varepsilon)$ and inductively $P^t(x) \in B(K,\varepsilon)$, $t = 1, 2, 3, \ldots,$ which proves stability. (2) $K_\varepsilon \neq \emptyset$. Let $m = \min_{K_\varepsilon} V$. Note that by invariance of $K$, $K \cap K_\varepsilon = \emptyset$, and as $B(K,\varepsilon)$ is open and $P$ is continuous, then $K_\varepsilon$ is compact. So $m > 0$. Again, by continuity of $V$ and compactness of $K$, there exists $\delta > 0$, $\delta \leqslant \varepsilon$, such that $B(K,\delta) \subset \{x \in X\colon\ V(x) < m\}$. Now, the following implication is true:

$$x \in B(K,\varepsilon) \quad \text{and} \quad V(x) < m \quad \Rightarrow \quad P(x) \in B(K,\varepsilon). \tag{18}$$

In fact, if there existed $x \in B(K,\varepsilon)$ and $P(x) \notin B(K,\varepsilon)$, then $x \in K_\varepsilon$ and $V(x) \geqslant m$ which is a contradiction. Now, let $x \in B(K,\delta)$. Then, for each $t$, $V(P^t(x)) \leqslant V(x) < m$ by the choice of $\delta$. Simple induction and (18) show that the orbit $o(x) \subset B(K,\varepsilon)$, which proves the stability of $K$.

(ii) Let $x \in X$ and assume that $\omega(x) \neq \emptyset$. We will show that $V$ is constant on $\omega(x)$. In fact, choose two points $y, z \in \omega(x)$, and then, corresponding sequences $\{s_i\}$ and $\{t_i\}$ going to infinity such that $P^{s_i}(x) \to y$ and $P^{t_i}(x) \to z$. Taking subsequences, if necessary, one can assume that for all $i$'s: $s_i < t_i < s_{i+1} < t_{i+1}$. Hence

$$V\big(P^{s_i}(x)\big) \leqslant V\big(P^{t_i}(x)\big) \leqslant V\big(P^{s_{i+1}}(x)\big) \leqslant V\big(P^{t_{i+1}}(x)\big).$$

Letting $i \to \infty$:

$$V(y) \leqslant V(z) \leqslant V(y) \leqslant V(z),$$

and then $V(y) = V(z)$ as we have claimed. As $\omega(x)$ is invariant, then $P(y) \in \omega(x)$ for any $y \in \omega(x)$, hence for any such $y$, $V(P(y)) = V(y)$, hence consequently $y \in K$ by our assumption (17). So $\omega(x) \subset K$ and $x$ is attracted to $K$ as required.

(iii) Choose $\varepsilon > 0$ such that $\overline{B(K,\varepsilon)}$ is compact. Let $\delta_0 > 0$ correspond to this $\varepsilon$ by (i) just proven. Then for $x \in B(K,\delta_0)$ the orbit $o(x) \subset \overline{B(K,\varepsilon)}$ and by compactness of the latter set $\omega(x) \neq \emptyset$. By (ii) $x$ is attracted to $K$.

(iv) Like in the proof of (iii) we can show that any point from $X$ is attracted to $K$, which completes the proof of the theorem. $\square$

## 4. Remarks and comments

*1.* Let us note that we have not imposed any condition on a local method $\varphi$ but quite natural (1), and (A3) and (A4). The assumption on continuity of the local method $\varphi$, (A3), is in some way restrictive, as some local methods used in practice are not continuous at some points and then Theorem 6 cannot be used. On the other hand purely statistical algorithms do not use local methods at all, in other words $\varphi(x) = x$ for each $x \in A$ and all the above assumptions are clearly satisfied.

*2.* We can use any measure $\nu$ if only assumptions (A2) and (A4) are satisfied and it seems that the measures absolutely continuous with respect to the Lebesgue measure meet these requirements with most functions being optimized in practice. In fact, if function $f$ is $C^1$ 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 (A2) is satisfied. Also, (A4) is satisfied if $\nu$ is absolutely continuous with respect to the Lebesgue measure and $\varphi$ is a local diffeomorphism.

*3.* J. Dupuis [3] has established so-called EAR algorithm for obtaining the Maximum Likelihood Estimator for some problems with incomplete data. In his method our measure $\nu$ is a distribution with density proportional to the target function, $k = 1$ and $\varphi(x) = x$ for each $x \in A$. Under some additional assumption and using Markov Chains techniques he has proven a result, Theorem 3, which is a counterpart of Corollary 9 above.

*4.* Instead of the product measure $\nu^k$ on the space $A^k$ one can use any measure $\lambda$ which is equivalent to $\nu^k$. In other words, $y_1, \ldots, y_k$ may be not a simple sample.

*5.* If assumption (A5) is not satisfied, then we may lose global asymptotic stability. Still, using Theorem 11(i) and modifying the proof of Theorem 6 one gets that the set $\mathcal{M}^\star$ is stable.

*6.* If the set $A$ is not bounded, then in general $\mathcal{M}$ it is neither compact nor locally compact and our theorem cannot be used. On the other hand, if $f$ is bounded and the optimization problem has a solution and all other assumptions of Theorem 6 are satisfied, then using Theorem 11(ii) one can easily prove that any partial limit of the sequence $\{P^t\mu\}$ is contained in $\mathcal{M}^\star$.

## References


[1] R. Ash, Real Analysis and Probability, Academic Press, 1972.
[2] N. Bhatia, G. Szegö, Stability Theory of Dynamical Systems, Springer-Verlag, 2002.
[3] J. Dupuis, A global optimization algorithm of the likelihood for a Markov model with randomly missing data, preprint.
[4] C. Guus, E. Boender, E. Romeijn, Stochastic methods, in: R. Horst, P. Pardalos (Eds.), Handbook of Global Optimization, Kluwer Academic Publishers, 1995.
[5] A. Lasota, M. Mackey, Chaos, Fractals and Noise, Springer-Verlag, 1994.
[6] S. Meyn, R. Twedie, Markov Chains and Stochastic Stability, Springer-Verlag, 1993.
[7] K. Parthasarathy, Probability Measures on Metric Spaces, Academic Press, 1967.
[8] Ch. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag, 2004.
[9] W. Słomczyński, Dynamical Entropy, Markov Operators and Iterated Functions Systems, Wydawnictwo Uniwersytetu Jagiellońskiego, 2003.
[10] M. Wright, The interior-point revolution in optimization: History, recent developments, and lasting consequences, Bull. Amer. Math. Soc. 42 (2005) 39–56.