J. Math. Anal. Appl. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Global convergence of discrete-time inhomogeneous Markov processes from dynamical systems perspective

Dawid Tarłowski *

Institute of Mathematics, Faculty of Mathematics and Computer Science, Jagiellonian University, Łojasiewicza 6, 30 348 Kraków, Poland

ARTICLE INFO

Article history:
Received 15 September 2016
Available online xxxx
Submitted by U. Stadtmueller

Keywords:
Global optimization
Global convergence
Inhomogeneous Markov process
Foias operator
Lyapunov function

ABSTRACT

Given the continuous real-valued objective function f and the discrete time inhomogeneous Markov process X_t defined by the recursive equation of the form $X_{t+1} = T_t(X_t, Y_t)$, where Y_t is an independent sequence, we target the problem of finding conditions under which the X_t converges towards the set of global minimums of f. Our methodology is based on the Lyapunov function technique and extends the previous results to cover the case in which the sequence $f(X_t)$ is not assumed to be a supermartingale. We provide a general convergence theorem. An application example is presented: the general result is applied to the Simulated Annealing algorithm.

1. Introduction

Let (A, d) be a compact metric space. Assume that $f : A \to \mathbb{R}$ is the continuous problem function with global minimum $\min f = 0$ and $A^\star = \{x \in A : f(x) = 0\}$ is the set of the solutions of the global minimization problem. The last decades have witnessed the great development of iterative numerical techniques designed for finding an element from $A^\star$. The most popular methods are: genetic and evolutionary algorithms [33,32,5,31], inspired by the mechanisms of biological evolution, Simulated Annealing algorithm (SA) [4,39,19–21,1], which is based on analogy with the physical process of annealing, and methods based on the swarm intelligence of individuals [15] like Particle Swarm Optimization (PSO) [10,9] or Ant Colony Optimization (ACO) [13]. Those methods, and many other iterative heuristics, [3,16,23,30,40,41] are used in practice for solving difficult real world problems for which analytical methods fail. The corresponding literature puts great attention to the numerical aspect of the subject. From the theoretical perspective, majority of such optimization techniques can be represented as discrete-time inhomogeneous Markov processes of the form

$$x_{t+1} = T_t(x_t, y_t), \text{ for } t \in \mathbb{N}, \tag{1.1}$$

* Fax: +48 (12) 664 66 74.
E-mail addresses: dawid.tarlowski@im.uj.edu.pl, dawid.tarlowski@gmail.com.

http://dx.doi.org/10.1016/j.jmaa.2016.11.076
0022-247X/© 2016 Elsevier Inc. All rights reserved.

where the sequence x_t represents the successive states of the algorithm, y_t represents the probability distributions of the algorithm and T_t stands for the deterministic "mechanisms" of the algorithm. Recursions of the form (1.1) have been studied in many contexts, including control theory, iterated function systems (IFS), fractals, and other applications. Various examples can be found, for instance, in [24,11,14,18]. Generally speaking, the standard analysis of processes (1.1) concerns the problem of the existence and the convergence to the unique stationary distribution. This paper continues the research taken in the series of papers [26,25,27,38,37] which aim at the problem how to prove that the process given by equation (1.1) converges towards $A^\star$ under conditions that can be verified in practical cases.

General results on global convergence are often based on the classical probability theory [36,29]. Markov chains theory is used to prove the convergence towards $A^\star$ in some cases, see for example [33] or [1]. An important class of global optimization methods is methods with the supermartingale property – we shortly say that an optimization method X_t is a supermartingale if the corresponding sequence of record values $f(X_t)$ is a supermartingale. In this case stochastic Lyapunov functions arise quite naturally as a tool ensuring stability of the process X_t, see Chapter VIII in [2] for general framework or [34] for an example from evolutionary optimization. Previous papers [26,25,27,38,37] work under assumption $E(f(T_t(x,Y_t))| X_t = x) \leq f(x)$, $x \in A$, which implies that they also aim at the supermartingale class. The general methodology used there was to consider the nonautonomous dynamical system on the set $\mathcal{M}(A)$ of Borel probability measures on A (the system is induced by equation (1.1)) and next to prove the asymptotic stability of the set $M^\star = \{\mu \in \mathcal{M}(A)\colon \mu(A^\star) = 1\}$. One of the basic tools used in the proof was the Lyapunov function given by $V\colon \mathcal{M}(A) \ni \mu \to \int_A f d\mu \in [0,\infty)$. This paper extends this methodology to cover the case of non-supermartingales. For instance, Simulated Annealing algorithm and Evolution Algorithms with non-elitist selection strategies belong to the class of non-supermartingales. The main result of this paper is Theorem 2. Theorem 3, which is the conclusion of Theorem 2, is less general but easier to use and still covers some important practical cases like Simulated Annealing and many non-ellitist evolutionary methods. To present how the general results work in practice Theorem 2 is applied to the SA algorithm. The SA convergence result is expressed in Theorem 4.

This paper is organized as follows. Section 2 presents some general equivalences between basic modes of stochastic global convergence. These general results are rather easy to prove but according to the author's knowledge such general statements are not formulated in literature (special cases are proved separately in various papers). Section 3 presents and discusses the main results of this paper, Theorem 2, and its conclusion, Theorem 3. Section 4 applies Theorem 3 to the Simulated Annealing algorithm. Section 5 presents some facts on the weak convergence of Borel probability measures and Section 6 presents some ideas expressed in the language of dynamical systems and necessary for the proof of the main result. Finally, Section 7 uses the results of Sections 5 and 6 to prove Theorem 2. Appendix presents the proofs of results from Section 2.

2. Some equivalences for stochastic global convergence

This section presents some general equivalences for stochastic global convergence and introduces corresponding notation which will be used in further sections. Although this paper targets the compact case situation, the general results of this section are presented under assumption that the metric space A is separable. The results are rather simple but they generalize many partial observations stated in the literature and will be used further in this paper. The corresponding proofs can be found in Appendix.

We denote:

(1) $A^\star = \{x \in A\colon f(x) = 0\}$,
(2) $A_\delta = \{x \in A : f(x) \leq \delta\}$, where $\delta > 0$,

Please cite this article in press as: D. Tarłowski, Global convergence of discrete-time inhomogeneous Markov processes from dynamical systems perspective, J. Math. Anal. Appl. (2017), http://dx.doi.org/10.1016/j.jmaa.2016.11.076