J. Math. Anal. Appl. 412 (2014) 631–645

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Nonautonomous stochastic search for global minimum in continuous optimization

Dawid Tarłowski *

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

ARTICLE INFO

Article history:
Received 1 April 2012
Available online 31 October 2013
Submitted by H. Frankowska

Keywords:
Stochastic optimization
Global optimization
Lyapunov function
Foias operator
Grenade explosion method
Evolution strategy

ABSTRACT

Various iterative stochastic optimization schemes can be represented as discrete-time Markov processes defined by the nonautonomous equation $X_{t+1} = T_t(X_t, Y_t)$, where Y_t is an independent sequence and T_t is a sequence of mappings. This paper presents a general framework for the study of the stability and convergence of such optimization processes. Some applications are given: the mathematical convergence analysis of two optimization methods, the elitist evolution strategy $(\mu + \lambda)$ and the grenade explosion method, is presented.

1. Introduction

Global optimization is a fast growing area of great practical importance. Let (A, d) be a separable metric space and let $f : A \to \mathbb{R}$ be a continuous function having its global minimum $\min f$. Without loss of generality we assume that $\min f = 0$. Let

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

The function f can be called a problem function and the elements of $A^\star$ can be called the solutions of the global minimization problem. Under some assumptions on the function f, like, for example, the differentiability, deterministic optimization techniques [21] can be used for solving optimization problems. However, global minima are usually hard to locate. Stochastic optimization techniques [39,38,27] are usually not dependent on the smoothness of the function. At the same time, if properly configured, these methods can be very effective in finding global solutions. There is great variety of available techniques, among them we have genetic and evolutionary algorithms [32,31,6,30], Simulated Annealing (SA) [5,37,18] or swarm intelligence algorithms like Particle Swarm Optimization (PSO) [11,10], Artificial Bee Colony (ABC) [16] or Ant Colony Optimization (ACO) [14]. Grenade Explosion Method (GEM) [1,2,36] is a technique proposed quite recently. Many algorithms are new variants or combinations of other methods. Accelerated Random Search (ARS) [4] can be viewed as the modification of Pure Random Search (PRS), while Random Multistart algorithms [39] combine global random search and local deterministic techniques. All the above mentioned iterative optimization techniques (except for some specific modifications,

* Fax: +48 12 664 66 74.
E-mail address: dawid.tarlowski@im.uj.edu.pl.

0022-247X/$ – see front matter
http://dx.doi.org/10.1016/j.jmaa.2013.10.070

for instance non-Markovian versions of Random Multistart are presented in [19]), and many other optimization methods, can be represented as discrete-time inhomogeneous Markov processes of the form

$$x_{t+1} = T_t(x_t, y_t), \quad \text{for } t = 0, 1, 2, \ldots, \tag{1.1}$$

where x_t is the sequence of states successively transformed by the algorithm, y_t is the independent sequence which represents the randomness of the algorithm and T_t is the sequence of the methods of the algorithm. The aim of this paper is to provide a general theoretical framework for the study of the stability and convergence of such optimization processes under conditions that can be verified in practice. Some applications are presented.

We refer to [9] for the general study of the processes of the form (1.1). As stated there, every Markov Chain given on a separable metric space has such a representation. Recursions of the form (1.1) have been studied for various purposes, including optimization, iterated function systems (IFS), fractals, control theory and other applications. Many examples, which correspond to the time-homogeneous situation, are given in [20,12,15,17]. These references focus mainly on the classical problem regarding the convergence of processes (1.1), which is how to prove the convergence to the unique stationary distribution.

This paper extends the results of previous papers [23,22,28,24] which deal with the following problem: Eq. (1.1) induces a dynamical system determined by the family of Foias operators, which are given on the metric space $M(A)$ of Borel probability measures on the space A, and the goal is to prove the global attractiveness of the set $M^\star = \{\mu \in M(A): \mu(A^\star) = 1\}$. In the present paper we work under weaker assumptions and hence the convergence results have significantly more applications compared to the previous results. Sections 6 and 7 present some applications: the general convergence results from Section 2 are applied to the GEM algorithm (grenade explosion method) and to the evolution strategy $(\mu/\rho + \lambda)$. In the present paper we do not assume that the Foias operators corresponding to the algorithm are continuous (the continuity is equivalent to the Feller property, see [20,17]) and the assumption (**A**) of Theorem 1 is used instead. As the continuity is no longer assumed, the dynamical system corresponding to Eq. (1.1) is in fact a pseudo-dynamical system, see [25]. Additionally, the compactness assumption of the set U of the algorithm's parameters and distributions is released and the t_0 steps contraction from assumption (U_0) of [24] is replaced with a softer condition. For monotonic methods the important case when the space A is not compact is covered for the class of functions with compact lower level sets. Still, the basic convergence assumption is that the inequality $\int f(T_t(x, y)\nu_t(dy)) < f(x)$ is satisfied for some pairs (T_t, ν_t), where ν_t is the distribution of Y_t, and the Lyapunov function technique is used. In paper [35], under the continuity assumption of Foias operators, the above strong inequality is replaced with a weaker, t_0 steps inequality.

The methodology behind the results of this paper is based on purely topological approach to the stochastic optimization. The weak convergence topology on M is considered and the probability theory is used only to interpret the stability of $M^\star$ in terms of the algorithm convergence (it may be also useful in applications). The basic convergence condition $\int f(T_t(x, y)\nu_t(dy)) < f(x)$ expresses that the algorithm is capable of reaching regions with smaller function values at one step starting from x at step t. It connects the algorithm's parameters values with the topology of the function f. The classical general convergence results for monotonic search methods which are based on the probability theory either use the assumption that the algorithm can reach arbitrarily small vicinity of the set of global solutions $A^\star$ from any position [34] or are based on conditions hardly verifiable in many practical situations [26]. Many results on the convergence of monotonic methods are based on this $A^\star$ accessibility-type assumption, see for instance [36,29,30] for many examples from evolutionary optimization. The case of non-monotonic methods is also often analyzed with the use of the classical probability theory, see for example [5,18,37], the $A^\star$ accessibility assumption is often released and the analysis is usually more difficult. Markov chains theory is sometimes used for the convergence analysis, see for example [32] or [3]. However, monotonic methods do not satisfy proper ergodic-type assumption and many advanced tools of Markov chains are hardly applicable to this case.

This paper is organized as follows. Section 2 presents and discusses the main results of the paper, Theorem 1 and Theorem 2. In Section 3 and Section 4 we prepare some necessary tools from the stability theory and the weak convergence of measures. The global asymptotic stability of $M^\star$ is stated in Theorem 5. In Section 5 this theorem is interpreted in the terms of the algorithm convergence which proves Theorem 1 and Theorem 2. A reader interested in the applications of the results from Section 2, not in the theoretical details of the proofs, can omit Sections 3, 4, 5 with no loss of continuity. In Section 6 we provide sufficient conditions for the convergence of grenade explosion method. In Section 7 the functionality of Theorem 2 is presented on the $(\mu/\rho + \lambda)$ strategy.

2. Theorem 1 and Theorem 2

Let (Ω, Σ, P) be a probability space and let (B, d) be a separable metric space. Let $X_t : \Omega \to A$, $t = 0, 1, \ldots$, be a measurable sequence defined by the nonautonomous recursive equation:

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

where $Y_t : \Omega \to B$ are independent random variables, the $X_0 : \Omega \to A$ is independent of the sequence Y_t, and maps $T_t : A \times B \to A$ are measurable. Let ν_t denote the probability distribution of Y_t, $t = 0, 1, \ldots$. It is clear that the distributions μ_t of X_t are determined by the initial distribution μ_0 of X_0 and the sequence (T_t, ν_t).