Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 128 (2018) 1073–1094
stochastic processes and their applications
www.elsevier.com/locate/spa

Convergence results for a class of time-varying simulated annealing algorithms

Mathieu Gerber[a,*], Luke Bornn[a,b]

[a] *Harvard University, Department of Statistics, United States*
[b] *Department of Statistics and Actuarial Science, Simon Fraser University, Canada*

Received 5 September 2016; received in revised form 22 March 2017; accepted 5 July 2017
Available online 19 July 2017

Abstract

We provide a set of conditions which ensure the almost sure convergence of a class of simulated annealing algorithms on a bounded set $\mathcal{X} \subset \mathbb{R}^d$ based on a time-varying Markov kernel. The class of algorithms considered in this work encompasses the one studied in Bélisle (1992) and Yang (2000) as well as its derandomized version recently proposed by Gerber and Bornn (2016). To the best of our knowledge, the results we derive are the first examples of almost sure convergence results for simulated annealing based on a time-varying kernel. In addition, the assumptions on the Markov kernel and on the cooling schedule have the advantage of being trivial to verify in practice.

Keywords: Digital sequences; Global optimization; Simulated annealing

1. Introduction

Simulated annealing (SA) algorithms are well known tools to evaluate the global optimum of a real-valued function φ defined on a measurable set $\mathcal{X} \subseteq \mathbb{R}^d$. Given a starting value $x_0 \in \mathcal{X}$, SA algorithms are determined by a sequence of Markov kernels $(K_n)_{n\geq 1}$, acting from $\mathcal{X}$ into itself, and a sequence of temperatures (also called cooling schedules) $(T_n)_{n\geq 1}$ in $\mathbb{R}_{>0}$. Simulated annealing algorithms have been extensively studied in the literature and it is now well established that, under mild assumptions on φ and on these two tuning sequences, the

* Correspondence to: University of Bristol, School of Mathematics, United Kingdom.
E-mail address: mathieu.gerber@bristol.ac.uk (M. Gerber).

http://dx.doi.org/10.1016/j.spa.2017.07.007

resulting time-inhomogeneous Markov chain $(X^n)_{n\geq 1}$ is such that the sequence of value functions $(\varphi(X^n))_{n\geq 1}$ converges (in some sense) to $\varphi^* := \sup_{x\in\mathcal{X}}\varphi(x)$. Most of these results are derived under the condition $K_n = K$ for all $n \geq 1$, see for instance see Bélisle [2] and Locatelli [7] for convergence results of SA on bounded spaces and Andrieu et al. [1] and Rubenthaler et al. [11] for results on unbounded spaces.

However, it is a common practice to use as input of SA a sequence of Markov kernels $(K_n)_{n\geq 1}$ whose variance reduces over time in order to improve local exploration as n increases. For instance, the simulated annealing functions in Matlab (function `simulannealbnd`) and in R (option "SANN" of the function `optim`) are both based on a Markov kernel whose scale factor is proportional to the current temperature. Some convergence results for such SA algorithms based on a time-varying Markov kernel can be found e.g. in [12].

Recently, Gerber and Bornn [4] proposed a modification of SA algorithms where extra dependence among the random variables generated in the course of the algorithm is introduced to improve the exploration of the state space. The idea behind this new optimization strategy is to replace in SA algorithms the underlying i.i.d. uniform random numbers in $[0, 1)$ by points taken from a random sequence with better equidistribution properties. More precisely, Gerber and Bornn [4] take for this latter a $(t, s)_R$-sequence, where the parameter $R \in \bar{\mathbb{N}} := \{0, 1, \ldots, \infty\}$ controls for the degree of randomness of the input point set, with the case $R = 0$ corresponding to i.i.d. uniform random numbers and the limiting case $R = \infty$ to a particular construction of quasi-Monte Carlo (QMC) sequences known as (t, s)-sequences; see Section 2.3 for more details. Convergence results and numerical analysis illustrating the good performance of the resulting algorithm are given in [4]. Their theoretical analysis only applies for the case where $K_n = K$ for all $n \geq 1$; in practice, as explained above, it is however desirable to allow the kernels to shrink over time to improve local exploration as the chain becomes more concentrated around the global optimum.

In this work we study SA type algorithms based on a time-varying kernel by making two important contributions. First, we provide under minimal assumptions an almost sure convergence result for Monte Carlo SA which constitute, to the best of our knowledge, the first almost sure convergence result for this class of algorithms. Second, we extend the analysis of Gerber and Bornn [4] to the time-varying set-up. As in [6] and [12], the conditions on the sequence $(K_n)_{n\geq 1}$ for our results to hold amount to imposing a bound on the rate at which the tails of K_n decrease as $n \to \infty$. Concerning the cooling schedules, all the results presented in this paper only require that, as in [4], the sequence $(T_n)_{n\geq 1}$ is such that the series $\sum_{n=1}^{\infty} T_n \log(n)$ converges.

The results presented below concern the limit of the sequence $(\varphi(X^n))_{n\geq 1}$ but, in practice, we are mostly interested in the sequence $\left(\max_{1\leq k\leq n}\varphi(X^k)\right)_{n\geq 1}$ to estimate φ^*. However, if $\varphi^* < +\infty$ (as assumed below) it is clear from the relation

$$\varphi(X^n) \leq \max_{1\leq k\leq n} \varphi(X^k) \leq \varphi^*, \quad \forall n \geq 1$$

that the convergence of the former sequence to φ^* implies the convergence of the latter sequence to that value.

The rest of the paper is organized as follows. Section 2 introduces the notation and the general class of SA algorithms studied in this work. The main results are provided in Section 3 and are illustrated for some classical choice of Markov kernels in Section 4. All the proofs are collected in Section 5.