On the Convergence Rate of the Markov Symmetric Random Search

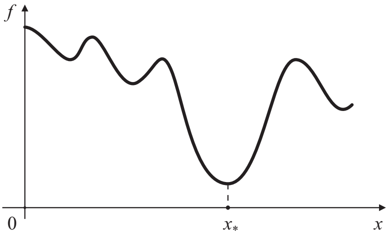
A.S. Tikhomirov

Novgorod State University

Velikiy Novgorod
2010

Introduction

Suppose that the objective function $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ takes its minimal value at a single point x_*.



Consider the problem of finding the global minimum of an objective function f. One possible approach to this problem is to apply random search optimization methods.

This paper is devoted to the theoretical study of the convergence rate of the Markov symmetric random search. We measure the convergence rate of such algorithms by the number of evaluations of the objective function required to attain the desired accuracy ε of the solution. It is shown that for the Markov symmetric random search it is not possible to obtain a rate better than $|\ln \varepsilon|$ as $\varepsilon \to 0$.

Optimization space

We consider the case of the Euclidean space $\mathbb{R}^d$ and the Euclidean metric ρ:

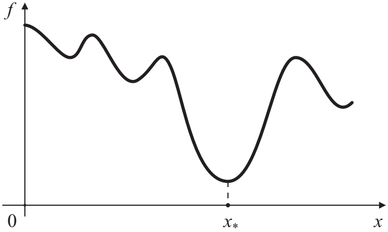
$$\rho(x,y) = \left(\sum_{n=1}^{d}(x_n - y_n)^2\right)^{1/2},$$

where $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$.
The closed ball of radius r centered at x will be denoted by

$$B_r(x) = \{y \in \mathbb{R}^d,\ \text{such that } \rho(x,y) \le r\}.$$

Objective function

We assume that the objective function $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ is measurable, bounded from below and takes its minimal value at a single point $x_* = \arg\min\{f(x),$ such that $x \in \mathbb{R}^d\}$.



Markov random search

Following the book by Zhigljavsky A. and Zilinskas A.

Zhigljavsky A., Zilinskas A. Stochastic Global Optimization. Berlin: Springer-Verlag. 2008.

we give a general scheme of Markov algorithms for global random search. The following algorithm simulates the Markov random search $\{\xi_n\}_{n\geq 0}$ in $\mathbb{R}^d$ with the initial point $\xi_0 = x \in \mathbb{R}^d$.

Algorithm 1 (A general scheme of Markov algorithms)

Step 1. Set $\xi_0 = x$ and the iteration number $n = 1$.

Step 2. Obtain a point η_n in $\mathbb{R}^d$ by sampling from the distribution $P_n(\xi_{n-1}, \cdot\,)$. Here $P_n(\xi_{n-1}, \cdot\,)$ is the transition probability; this probability may depend on n and ξ_{n-1}.

Step 3. Set

$$\xi_n = \begin{cases} \eta_n & \text{with probability } Q_n, \\ \xi_{n-1} & \text{with probability } 1 - Q_n. \end{cases}$$

Here Q_n is the acceptance probability; this probability may depend on η_n, ξ_{n-1}, $f(\eta_n)$ and $f(\xi_{n-1})$.

Step 4. Check a stopping criterion. If the algorithm does not stop, substitute $n + 1$ for n and return to Step 2.

Simulated annealing

Particular choices of transition probabilities $P_n(x,\cdot)$ and acceptance probabilities Q_n lead to specific Markov global random search algorithms. The most well-known among them is the celebrated 'simulated annealing' algorithm. A general simulated annealing algorithm is algorithm 1 with acceptance probabilities

$$Q_n = \begin{cases} 1 & \text{if } \Delta_n \le 0, \\ \exp(-\beta_n \Delta_n) & \text{if } \Delta_n > 0, \end{cases} \tag{1}$$

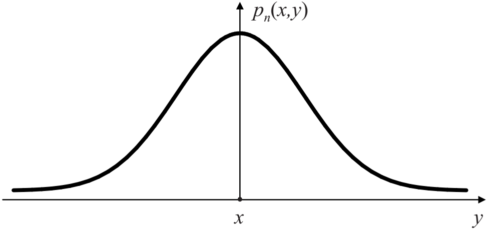
where $\Delta_n = f(\eta_n) - f(\xi_{n-1})$ and $\beta_n \ge 0$ $(n = 1, 2, \ldots)$.
The choice (1) for the acceptance probability Q_n means that any 'promising' new point η_n (for which $f(\eta_n) \le f(\xi_{n-1})$) is accepted unconditionally; a 'non-promising' point (for which $f(\eta_n) > f(\xi_{n-1})$) is accepted with probability $Q_n = \exp(-\beta_n \Delta_n)$.

Markov symmetric random search

In accordance with the structure of algorithm 1, we call the distributions $P_n(\xi_{n-1}, \cdot)$ trial transition probabilities. We shall consider the case where the trial transition probabilities $P_n(x, \mathrm{d}y)$ have symmetric densities $p_n(x, y)$ of the form

$$p_n(x,y) = g_n\big(\rho(x,y)\big), \qquad (2)$$

where ρ is the Euclidean metric and g_n are non-increasing functions of a positive argument.



Markov search with transition densities of the form (2) will be called Markov symmetric random search.

Characteristics of random search

We use a random search to find the minimizer x_* with the prescribed accuracy ε (approximation with respect to the argument). In this case, we want the search to hit the ball $B_\varepsilon(x_*)$. Denote by

$$\tau_\varepsilon = \min\{n \geq 0, \text{ such that } \xi_n \in B_\varepsilon(x_*)\}$$

the instant when the search first hits the set $B_\varepsilon(x_*)$.

Usually, we assume that there is no need to evaluate the function f for the simulation of the distributions P_n. Therefore, in the process of performing τ_ε iterations of the algorithm, the function f is evaluated $\tau_\varepsilon + 1$ times.

We use two characteristics of the convergence rate of the random search. The computational effort of the random search is defined as $\mathsf{E}\,\tau_\varepsilon$. It is interpreted as the average number of steps needed to hit the set $B_\varepsilon(x_*)$.

The other characteristic of τ_ε considered in this paper is the guaranteeing number of steps. It is defined as the minimal number of steps $N(x, f, \varepsilon, \gamma)$ at which the hit of $B_\varepsilon(x_*)$ is ensured with the probability not less than γ; in other words, $N(x, f, \varepsilon, \gamma) = \min\{n \geq 0, \text{ such that } \mathsf{P}(\tau_\varepsilon \leq n) \geq \gamma\}$.

The rate of convergence

It is shown that for the Markov symmetric random search it is not possible to obtain a rate better than $|\ln \varepsilon|$ as $\varepsilon \to 0$.

Theorem 1

Assume that $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ takes its minimal value at a single point x_*. Consider any Markov symmetric random search in $\mathbb{R}^d$ with the initial point $x \in \mathbb{R}^d$. Let $0 < \varepsilon < \rho(x, x_*)$ and $0 < \gamma < 1$. Then we have

$$\mathsf{E}\,\tau_\varepsilon \geq c_d \ln\bigl(\rho(x, x_*)/\varepsilon\bigr), \quad N(x, f, \varepsilon, \gamma) \geq \gamma c_d \ln\bigl(\rho(x, x_*)/\varepsilon\bigr),$$

where $c_d = 2^{1-d} \sup\{(1-q)^d / |\ln q|$, such that $q \in (0,1)\}$.

This result gives us an opportunity to estimate potential capabilities of the Markov symmetric random search methods.

Fast optimization methods

Assume that the objective function f is 'non-degenerate' (see [Zhigljavsky, Zilinskas]).
Examples of homogeneous Markov algorithms with the computational effort and the guaranteeing number of steps behaving as $O(\ln^2 \varepsilon)$ were given here:

- Zhigljavsky A., Zilinskas A. Stochastic Global Optimization. Berlin: Springer-Verlag. 2008.
- A.S. Tikhomirov, "On the Markov Homogeneous Optimization Method," Computational Mathematics and Mathematical Physics. **46**, 361–375 (2006).
- A. Tikhomirov, T. Stojunina, and V. Nekrutkin, "Monotonous Random Search on a Torus: Integral Upper Bounds for the Complexity," Journal of Statistical Planning and Inference. **137**, 4031–4047 (2007).
- A.S. Tikhomirov, "On Fast Variants of the Simulated Annealing Algorithm," Stochastic Optimisation in Computer Science. **5**, pp. 65–90 (2009) [in Russian].
- A.S. Tikhomirov, "On the Convergence Rate of the Simulated Annealing Algorithm," Computational Mathematics and Mathematical Physics. **50**, 19–31 (2010).

Fast optimization methods

Nonhomogeneous Markov algorithms with the computational effort and the guaranteeing number of steps behaving as $O(|\ln\varepsilon|\ \ln|\ln\varepsilon|)$ were given here:

- V.V. Nekrutkin and A.S. Tikhomirov, "Speed of Convergence as a Function of Given Accuracy for Random Search Methods," Acta Applicandae Mathematicae. **33**, 89–108 (1993).
- A.S. Tikhomirov and V.V. Nekrutkin, "Markov Monotone Search for Extrema: Survey of Some Theoretic Results," in Mathematical Modeling: Theory and Applications, No. 4 (VVM, St. Petersburg, 2004), pp. 3–47 [in Russian].
- A.S. Tikhomirov, "On the Convergence Rate of the Simulated Annealing Algorithm," Computational Mathematics and Mathematical Physics. **50**, 19–31 (2010).

The inequalities of theorem 1 show that these algorithms are fast optimization methods (from an asymptotic viewpoint). Their asymptotic rate of convergence is just marginally worse than the lower bounds in theorem 1.