Random search for global extremum of a function using Markov chains simulation

L V Vladimirova¹ and D A Ovsyannikov¹

¹St. Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia

E-mail: l.vladimirova@spbu.ru, d.a.ovsyannikov@spbu.ru

Abstract. The problem under study is global extremum search for multieextremal function. Random search method based on Markov chains simulation is used. A comparative analysis of this method with another method – the method of random search for the global extremum with "memory", using the normal distribution – done.

1. Introduction

The problem of global extremum search is relevant for various applied problems, including optimization of physical processes [1-7].

Among the methods of search for global extremum special place is taken by the algorithms of random search. As practice shows, when solving complex extreme problems, they are almost the only way to achieve the goal, while they have the simplicity of the algorithm and can be used for large dimensions of the search area.

To increase the efficiency of random search of global extremum in this paper we use uneven distribution with density associated with the target function. This allows us to find the point of global extremum as a mode of such distribution [2].

2. Problem statement

Assume there is a nonnegative function $f(x)$, set in the region $D \subset E^n$ -dimensional Euclidean space. The volume $D$ is expected to be limited. The objective is to find

$$f(x') = \sup_D f(x).$$

To solve problem (1), you can enter an uneven distribution in the area $D$

$$p(x) = f(x)(\int_D f(x)dx)^{-1}$$

and find the point $x'$ where the mode $p(x)$ will be reached. It is clear that the realization of a random vector distributed with density (2) will appear more often near the highest value $f(x)$. Moreover, the density $p_s(x) = f'(x)(\int_D f'(x)dx)^{-1}$ will be when the $s \to \infty$ converges to the $\delta$ -function concentrated at the point $x^*$ [2]. However, modeling such densities is difficult.
3. Random search for global extremum with "memory"

Let's consider a method that allows the use of uneven distribution and evaluation of the covariance matrix to model the density

$$p(x) = Cf(x)N(x^{(0)}, B),$$

(3)

instead of $p(x)$. In (3) $C$ is the normalization constant the density, function

$$N(x^{(0)}, B) = \frac{(2\pi)^{n/2}}{(\det B)^{1/2}} \exp\left(-(x-x^{(0)})^T B^{-1} (x-x^{(0)})\right)$$

(4)

is the density of the normal distribution with parameters: vector $x^{(0)} \in D$ and covariance matrix $B$. The covariance matrix $B$ to a certain extent reflects the structure of the function $f(x)$. In this sense, the described method is called the method with “memory”.

It is convenient to use the generalized Neumann method for density $p(x)$ modeling, and normally distributed random vectors with density (4) are modelled in the region $D$. This is a multi-iterative process. At the zero iteration, it is required to model random vectors uniformly distributed in $D$ (the usual Neumann method for density modeling (2)). Algorithm and convergence theorem for random search of global extremum of function based on modeling of normal distribution and estimation of covariance matrix are presented in [2, 8].

4. Metropolis Method

To model density (2), you can use a rejection sampling [9]. To do this, it’s needed to find a convenient simulated density $q(x)$, for which the following inequality is valid: $Cq(x) > f(x)$ for all $x \in D$, where $C$ is a constant. However, in the algorithm of sampling variance, there is an issue with the choice the constant $C$ (figure 1). If the constant $C$ is too large, the algorithm will give the correct results, but the ratio of the number of test points in the strikethrough area to the total number of test points will be small, which means that the effectiveness of the method is low [9]. If an error inclines towards the lower values of the part of the area $D$ where $f(x) > Cq(x)$ the rejection sampling cannot be applied. To resolve the issue with constant $C$ selection, you can use the metropolis method.

The algorithm of the metropolis method eliminates the problem with the choice of the constant $C$, using the transitive density of the Markov process as a function $q(x)$ depending on two variables $q(x, y)$. Instead of independent realizations of a random value with the density $q(x)$, the Markov chain with transition density $q(x, y)$ is modeled. Thus, the algorithm of the metropolis method serves as the basis of the Monte Carlo method on the Markov chains [10].

4.1. Metropolis method algorithm

1. Assume $k = 0$. Choose arbitrarily $X^{(0)}$ and the transition density $q(x, y)$. Modeling density $q(X^{(0)}, y)$, we obtain the realization of a random variable $Y^{(k)}$. 

[Figure 1. The rejection sampling]
2. Modeling \( q(Y^{(k)}, y) \), we obtain the realization of a random variable \( Y' \).
3. The value is calculated
   \[
   a = \left( f(Y') f^{-1}(Y^{(k)}) \right) \left( q(Y^{(k)}, Y') q^{-1}(Y', Y^{(k)}) \right).
   \]
   (5)
4. If \( a \geq 1 \), we assume \( Y^{(k+1)} = Y' \).
5. If \( a < 1 \), we assume \( Y^{(k+1)} = Y' \) with probability \( a \) and \( Y^{(k+1)} = Y^{(k)} \) with probability \( 1 - a \).

The random process \( Y^{(k)} \) with discrete time \( k = 0, 1, 2, 3, \ldots \) is a Markov process, and the transition density of a random process 1-5 is equal to \( A(x, y) = q(x, y) \cdot \min(1, p(y)q(y, x)p^{-1}(x)q^{-1}(x, y)) \) and for all \( x, y \) the ratio \( A(x, y) p(y) = A(y, x) p(x) \) is valid. Then \( p(x) \) is the density of the stationary distribution for \( A(x, y) \) [9]. This means that with a sufficiently large number of \( N \) the sequence
   \[
   Y^{(1)}, Y^{(2)}, \ldots, Y^{(N)}
   \]
   (6)
is the sequence of independent realizations of a random vector distributed with density \( p(x) \). Thus, according to the sequence (6), the density mode can be approximately found. As the \( p(x) \) is proportional to the function \( f(x) \), then the maximum value of function \( f(x) \) can be approximately found as well.

4.2. Special case of the metropolis method

As a transitional density, it is convenient to choose the density of the normal probabilities distribution (Gaussian)
   \[
   q(x, y) = \frac{1}{(2\pi)^n/2} \exp \left( -\frac{1}{2} (x - y)^T B^{-1} (x - y) \right), \quad x, y \in D \subseteq E^n
   \]
   (7)
with covariance matrix \( B \). Function (7) is symmetric and value search (4) is simplified:
   \[
   a = f(Y') f^{-1}(Y^{(k)})
   \]

5. The problem of finding a locally optimal plan for the non-linear exponential regression

Let's apply the metropolis method to find the global extremum of a complex multiextremal function with 4 variables. We consider the problem of finding the locally optimal plan for the non-linear exponential regression
   \[
   \eta(x, \theta) = \beta_1 e^{-\theta_1 x} + \beta_2 e^{-\theta_2 x}, \quad x \in [0, B] \subseteq R^1, \quad \beta_1, \beta_2 \neq 0, \quad \theta_2 > \theta_1 > 0,
   \]
   (8)
with the vector of parameters \( \theta = (\beta_1, \theta_1, \beta_2, \theta_2) \). In the ratio (8) \( B \) is some given value.

As we can see, only the parameters \( \theta_1, \theta_2 \) enter nonlinearly. Functions of the type (8) form an important class of solutions of systems of differential equations and have a variety of practical applications, for example, in atomic physics.

For this particular problem, the plan \( x_1, x_2, x_3, x_4 \) consists of the minimum possible number of values \( x \in [0, B] \) that is equal to the number of parameters \( \beta_1, \theta_1, \beta_2, \theta_2 \). To find the locally optimal plan, we consider the optimality criterion related to the information matrix \( M(x, \theta) \) with fixed parameters. Such optimality criterion for regression (8) is the determinant of the following form
   \[
   \det(M(x, \theta)) = \beta_1^2 \beta_2^2 A(x, \theta_1, \theta_2) \] [11-13], where \( x = (x_1, x_2, x_3, x_4) \) and
The function (9) depends on the points of the planning set \( x_1, x_2, x_3, x_4 \) and on the parameters \( \theta_1, \theta_2 \) entering nonlinearly in (8).

The task is to find the maximum of function (9) via variables \( x_1, x_2, x_3, x_4 \) from a set of scheduling \([0,B]\) with fixed values of parameters \( \theta_1 = \theta_{1_u}, \theta_2 = \theta_{2_u} \). The plan resulting from optimization of the optimality criterion with fixed \( \theta_{1_u}, \theta_{2_u} \) is called a locally optimal plan.

At the figure 2 locally optimal plans \( x_1, x_2, x_3, x_4 \) for 9 variants of parameters \( \theta_{1_u}, \theta_{2_u} \) are presented. This result are obtained in two ways of optimization. Locally optimal plans found using the metropolis method are indicated by asterisks. Locally optimal plans found with the application of random search with "memory" are circle-marked [8,12,13].

![Figure 2](image)

**Figure 2.** Four points \( x_1, x_2, x_3, x_4 \) of locally optimal plans for 9 variants of nonlinear parameters \( \theta_{1_u}, \theta_{2_u} \) found by two optimization methods

As can be seen from figure 2 the applied simplified metropolis method gives adequate results. Due to its simplicity and convenience, it can be successfully used for solving various applied problems, including optimization problems of various accelerating structures.

6. References

[1] Ovsyannikov A D, Ovsyannikov D A, Altsybeyev V V, Durkin A P and Papkovich V G 2014 *Application of Optimization Techniques for RFQ Design* (Problems of Nuclear Science and Engineering vol 91 No 3) pp. 116-119

[2] Vladimirova L V 2014 *Global Extremum Search on the Basis of Density and Its Mode Estimation* (Proc. of 20th International Workshop: Beam Dynamics & Optimization (BDO), June-July 2014 St. Petersburg Russia IEEE) pp 186-188

[3] Vladimirova L V 2016 *Multicriterial Approach to Beam Dynamics Optimization Problem* (II Conference on Plasma&Laser Research and Technologies Journal of Physics: Conference Series vol 747 No 1) 012070 [http://iopscience.iop.org/1742-6596/747/1/012070](http://iopscience.iop.org/1742-6596/747/1/012070)

[4] Rubtsova I D 2015 *Optimization of iterative beam dynamic process* (Proc. III Int. Conf. “Stability and Control Processes” in Memory of V.I.Zubov (SCP) (5-9 October 2015 St. Petersburg) ed L A Petrosyan and A P Zhabko (St. Petersburg: IEEE)) DOI: 10.1109/SCP.2015.7342092 pp 198-200

[5] Ovsyannikov D A and Altsybeyev V V 2013 *Optimization of APF accelerators* (Problems of Atomic Science and Technology No 88(6)) pp 119-122

[6] Zavadsky S V, Ovsyannikov D A and Chung S-L 2009 *Parametric Optimization Methods for...
the Tokamak Plasma Control Problem Int. J. Mod. Phys. A 24 No 5 pp.1040–47

[7] Kotina E D 2006 Discrete optimization problem in beam dynamics Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 558(1), c. 292-294

[8] Vladimirova L V 2013 Using random search with "memory" in the estimation of nonlinear regression parameter (Bulletin SPbSUTD series 1 No 4) pp 30-34

[9] Ermakov S M and Sipin A S 2014 Monte-Carlo and Parametric Rasilimali Algorithms (SPb.: Publishing House of Saint Petersburg University) p 248

[10] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 Equations of state calculations by fast computing machines (Journal Chemical Physics vol 21) pp 1087-1091

[11] Melas V B 1978 Optimal designs for exponential regression (Math. Operations forsh. Statist. Ser. Statistics. vol. 9) pp 45-59

[12] Vladimirova L V and Fatyanova I A 2015 Construction of optimal regression experiment plan based on random search with "memory" using parallel calculations (Sustainability and management processes: Proceedings of the III international conference (SPb, 5-9 October 2015) Ed by A P Zhabko, L A Petrosyan: Fedorova G V Publishing House) pp 303-304.

[13] Vladimirova L V and Setina I A 2017 Locally Optimal Designs Search for Exponential Regression (48th international scientific conference of postgraduates and students “Management Processes and sustainability" April 3-6 2017 vol 4 No 1) pp 16-21