

APPLICATION OF STOCHASTIC APPROXIMATION IN TECHNICAL DESIGN

VAIDA BARTKUTĖ
*Institute of Mathematics and Informatics
Akademijos st. 4
Vilnius 08663, Lithuania*

LEONIDAS SAKALAUSKAS
*Institute of Mathematics and Informatics
Akademijos st. 4
Vilnius 08663, Lithuania*

In this paper, we consider problems related to the implementation of Stochastic Approximation (SA) in technical design, namely, estimation of a stochastic gradient, improvement of convergence, stopping criteria of the algorithm, etc. The accuracy of solution and the termination of the algorithm are considered in a statistical way. We build a method for estimation of confidence interval of the objective function extremum and stopping of the algorithm according to order statistics of objective function values provided during optimization. We give some illustration examples of application of developed approach of SA to the optimal engineering design problems, too.

1. Introduction

In many practical problems of technical design some of the data may be subject to significant uncertainty which is reduced to stochastic-statistical models. Models applied in such cases are appropriate when data evolve over time and decisions need to be made prior to observing the entire data stream. Consequently, the performance of such problems can also be viewed like constrained stochastic optimization programming tasks. SA can be considered as alternative to traditional optimization methods, especially when objective functions are no differentiable or computed with noise.

Application of SA to nonsmooth optimization is both a theoretical and practical problem. Computational properties of SA algorithms are mainly determined by the approximation approach to the stochastic gradient, [4], [6], [9], [10], [13], [14]. At last time Simultaneous Perturbation Stochastic Approximation methods (SPSA) become rather popular in literature devoted to stochastic search. It is of interest to consider SPSA methods because in these methods values of the function for estimating the stochastic gradient are required only at one or several points. The SPSA algorithms were considered by several

authors who used various smoothing operators. SPSA methods, uniformly smoothing the function in an n -dimensional hypercube, are described in [10]. SPSA algorithm with the Bernoulli perturbation was proposed and the computational efficiency of SPSA as compared with the standard finite difference approximation was indicated in [13]. The convergence and asymptotic behaviour of this algorithm were established in the class of differentiable functions.

Application of the SPSA algorithms to non-differentiable optimization is of particular theoretical and practical interest. In this paper, we focus on the objective functions from the Lipschitz class. We consider and compare SPSA algorithms with various perturbation operators and SA finite difference algorithms. In this paper, we consider problems related to the implementation of SA algorithms, for example, estimation of a stochastic gradient, improvement of convergence, stopping criteria of the algorithm, etc. We build a method for estimation of confidence interval of the objective function extremum and stopping of the algorithm according to order statistics of objective function values provided during optimization. We give some illustration examples of applications of this method to the optimal engineering design problems, too.

2. Formulation of the optimization problem

Let optimization problem is

$$f(x) \rightarrow \min, \quad (1)$$

where the objective function $f: \mathcal{R}^n \rightarrow \mathcal{R}$ is Lipschitzian. Let $\partial f(x)$ be the generalized gradient (GG), i.e., the Clarke subdifferential [3] of this function. For the solving of the problem (1) we consider and compare three stochastic approximation methods: SPSA with Lipschitz perturbation operator, SPSA with Uniform perturbation operator and Standard Finite Difference Approximation (SFDA) method. General scheme of SA approach is as follows

$$x^{k+1} = x^k - \rho_k \cdot g^k, \quad k = 1, 2, \dots, \quad (2)$$

where g^k is the value of the stochastic gradient estimator at the point x^k , ρ_k is a scalar multiplier in iteration k . This scheme is the same for different SA algorithms whose distinguish only by approach for stochastic gradient estimation:

A) gradient estimator of the SPSA with Lipschitz perturbation operator is expressed as:

$$g(x, \sigma, \xi) = \frac{(f(x + \sigma \xi) - f(x)) \cdot \xi}{\sigma \cdot \|\xi\|} \quad (3)$$

where σ is the value of the perturbation parameter, vector ξ is uniformly distributed in the unit ball $\psi(y) = \begin{cases} \frac{1}{V_n}, & \text{if } \|y\| \leq 1, \\ 0, & \text{if } \|y\| > 1. \end{cases}$, V_n is the volume of the n -dimensional ball [1];

B) the same estimator for the SPSA with Uniform perturbation operator is expressed as:

$$g(x, \sigma, \xi) = \frac{(f(x + \sigma \cdot \xi) - f(x - \sigma \cdot \xi)) \cdot \xi}{2\sigma}, \quad (4)$$

where σ is the value of the perturbation parameter, $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ is a vector consisting of variables uniformly distributed from the interval $[-1; 1]$, [10];

C) gradient estimator $g = (g_1, g_2, \dots, g_n)$ for SFDA has following components:

$$g_i(x, \sigma, \xi, v) = \frac{f(x + \sigma \cdot \xi + v \cdot \varepsilon_i) - f(x + \sigma \cdot \xi)}{v}, \quad (5)$$

where ξ is the same like in (3), $\varepsilon_i = (0, 0, 0, \dots, 1, \dots, 0)$ is the vector with zero components except i^{th} one, which is equal to 1, v and σ are the values of the perturbation parameters, [10].

Note, that only two function values have to be computed for the estimation of the stochastic gradient according to SPSA methods A) and B), and the SFDA method, in its turn, requires that $n+1$ function values be computed.

Method (2) converges a. s. for all considered gradient estimators under certain conditions, [1], [10]. The SPSA convergence rate

$$E\|x^k - x^*\|^2 = O\left(\frac{1}{k^\gamma}\right), 1 \leq \gamma < 2 \quad (6)$$

has been established in a presence of function computed without noise, [1].

3. Computer modelling

The convergence of the proposed method was studied by computer modelling,

as well. We considered a class of test function $f = \sum_{k=1}^n a_k |x_k|^\beta + M$,

where a_k were randomly and uniformly generated in the interval $[\mu, K]$,

$K > \mu > 0$. The samples of $T=500$ test functions were generated, when $\mu = 2$, $K=5$.

Table 1. Empirical and theoretical rates of convergence by SA methods.

| | $\beta = 0.5$ | $\beta = 0.75$ | $\beta = 0.9$ | |
|---|---------------|----------------|---------------|---|
| Theoretical | 1.5 | 1.75 | 1.9 | |
| <i>Empirical</i> | | | | |
| SPSA (Lipshitz perturbation) | | | | |
| $n = 2$ | 1.45509 | 1.72013 | 1.892668 | $\rho_i = n \cdot \min\left(0.005; \frac{0.25}{i}\right)$ $\sigma_i = \sqrt{\frac{(n+2)(n+3)}{n(n+1)}} \min\left(0.01; \frac{1}{i^\beta}\right)$ |
| $n = 4$ | 1.41801 | 1.74426 | 1.958998 | |
| SPSA (Uniform perturbation) | | | | |
| $n = 2$ | 1.605244 | 1.938319 | 1.988265 | $\rho_i = n \cdot \min\left(0.055; \frac{1}{i}\right)$ $\sigma_i = \sqrt{\frac{(n+2)(n+3)}{n(n+1)}} \min\left(0.05; \frac{0.5}{i^\beta}\right)$ |
| $n = 4$ | 1.551486 | 1.784519 | 1.998132 | |
| Stochastic Difference Approximation method | | | | |
| $n = 2$ | 1.52799 | 1.76399 | 1.90479 | $\rho_i = \min\left(0.5; \frac{0.25}{i}\right)$ $\sigma_i = \sqrt{\frac{(n+2)(n+3)}{n(n+1)}} \min\left(0.1; \frac{1}{i^\beta}\right); v_i = \frac{\sigma_i}{\sqrt{i}}$ |
| $n = 4$ | 1.50236 | 1.75057 | 1.90621 | |

The test functions were minimized by algorithms described in Section 2 and rate of convergence $\Delta_k = E\|x^k - x^*\|^2$ was investigated. The empirical least square estimates of the rate of convergence by the Monte-Carlo method and coefficients of optimizing sequence (2) are presented in Table 1. Theoretical rates of convergence are given for comparison, too. Hence, empirical rates of convergence look similarly for all methods and corroborate theoretical conclusions (6).

4. Application of order statistic to optimality testing

Application of extreme value theory in optimization algorithms was studied by several authors, however this interesting topic is not studied sufficiently well. Statistical inferences about the maximal (minimal) value of a function are described in [11], [15], [16]. An application of the extreme order statistics to the estimation of the location of the maximum of a regression function can be found in [2].

Let consider description of the inference of order statistics to optimality testing. Assume, the problem (1) be solved by the SA algorithms providing a sample

$$H = \{\eta_1, \dots, \eta_N\}, \quad (7)$$

whose elements are function values $\eta_k = f(x_k)$ provided during optimization. We build a method for the estimation of the minimum of the objective function by order statistics of sequence (7). We will apply for this purpose the extreme value theory of i. i. d. variables and examine our approach by experimental way, because theoretical distribution of extremes of sequence (7) is not studied yet.

Thus, to estimate confidence intervals for the minimum A of the objective function, it suffices to choose from sample H only $m+1$ order statistic:

$\eta_{(0)}, \dots, \eta_{(m)}$, where $m = m(N)$, $\frac{m^2}{N} \rightarrow 0$, $N \rightarrow +\infty$, [11], [15]. Then the linear estimators for A can be as follows:

$$A_{N,m} = \sum_{i=0}^m a_i \eta_{(i)}, \quad (8)$$

where m is much smaller than N , a_0, \dots, a_m are some coefficients satisfying the

condition $\sum_{i=1}^m a_i = 1$. Let us consider simple sets of coefficients proposed by [7]:

$$a = (1 + c_m, 0, \dots, 0, -c_m), \quad c_m = \frac{1}{\prod_{i=1}^m \left(1 + \frac{1}{i \cdot \alpha}\right) - 1}, \quad \text{where } \alpha \text{ is the}$$

parameter of extreme values distribution. We examine a choice of this parameter for continuous optimization

$$\alpha = \frac{n}{\varphi}, \quad (9)$$

where φ is the parameter of homogeneity of the function $f(x)$ in the neighbourhood of the point of minimum [15].

The one-side confidence interval of the minimum of the objective function is as follows:

$$[\eta_{(0)} - r_{m,\gamma} \cdot (\eta_{(m)} - \eta_{(0)}), \eta_{(0)}], \quad (10)$$

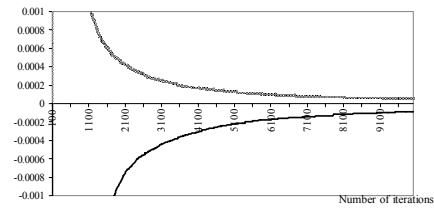
where $r_{m,\gamma}$ - certain constant, γ is the confidence level, [15].

We investigate the approach developed by computer modeling for SPSA method with Lipschitz perturbation, where parameters taken from the Table 1, $\gamma = 0.95$. The upper and lower bounds of the confidence interval for the

minimal value of the objective function are given in Table 2 and Figure 1. These bounds were estimated by the Monte-Carlo method with the number of iterations $N=10000$ and the number of trials $T=500$, when the confidence level was $\gamma = 0.95$. From Figure 1 we can see that the length of the confidence interval decreases when the number of iterations increases.

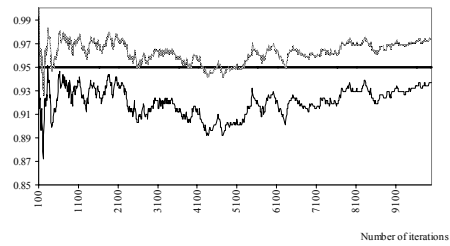
Table 2. Experimental results of the one-side confidence interval.

| N=10000, T=500, $\gamma = 0.95$ | Estimate of $A = 0$ | Confidence interval | | Empirical probability of hitting of A to the confidence interval |
|---------------------------------------|------------------------|---------------------|-------------|---|
| | | Lower bound | Upper bound | |
| n=2 | 0.000014 | -0.000036 | 0.0000275 | 0.954 |
| n=4 | 0.000289 | -0.000569 | 0.0005355 | 0.946 |
| n=6 | 0.001027 | -0.001879 | 0.001819 | 0.942 |
| n=10 | 0.004488 | -0.005359 | 0.0070606 | 0.946 |



— Lower bound of minimum
..... Upper bound of minimum

Figure 1. Confidence bounds of the minimum ($A=0$, $\gamma = 0.95$, $n = 2$, $T = 500$)



— Confidence probability (0,95)
— Lower bound of hitting probability
..... Upper bound of hitting probability

Figure 2. Confidence interval of the hitting probability $n = 2$, $\gamma = 0.95$, $T = 500$)

Thus, from the results of Table 2 and Figure 1 we can see that formula (10) approximates the confidence interval of objective function minimum rather well. Results of Table 2 and Figure 2 show also that empirical probability of minimal value hitting the confidence interval corroborates well to theoretical admissible confidence level γ . From the experimental results it follows that formula (10) can be used to create the stopping criteria for the algorithm, namely, the algorithm stops when the length of the confidence interval becomes smaller than the admissible value $\varepsilon > 0$.

5. Computer modelling

We will demonstrate the applicability of SA for two real-life problems.

5.1. Volatility estimation by Stochastic Approximation algorithm

Financial engineering, as well as risk analysis in the market research and management is often related to the implied and realized volatility. Let us consider the application of SA to the minimization of the mean absolute pricing error for the parameter calibration in the Heston stochastic volatility model, [8]. In this model option pricing biases can be compared to the observed market prices, based on the latter solution and pricing error. We consider the mean absolute pricing error (MAE) defined as:

$$\text{MAE}(\kappa, \sigma, \rho, \lambda, v, \theta) = \frac{1}{N} \sum_{i=1}^N |C_i^H(\kappa, \sigma, \rho, \lambda, v, \theta) - C_i|, \quad (11)$$

where N is the total number of options, C_i and C_i^H represent the realized market price and the implied theoretical model price, respectively, while $\kappa, \sigma, \rho, \lambda, v, \theta$ ($n=6$) are the parameters of the Heston model to be estimated.

To compute option prices by the Heston model, one needs input parameters that can hardly be found from the market data. We need to estimate the above parameters by an appropriate calibration procedure. The estimates of the Heston model parameters are obtained by minimizing MAE:

$$\text{MAE}(\kappa, \sigma, \rho, \lambda, v, \theta) \rightarrow \min. \quad (12)$$

Heston model was implemented for the Call option on SPX (29 May 2002). The SPSA algorithm with Lipschitz perturbation was applied to the calibration of the Heston model. Usually, SPSA requires that MAE be computed several hundred times that is reasonable for interactive Heston model calibration. Figure 3 below illustrates the applicability of the SPSA algorithm in practice, where we can see the dependence of MAE on the number of computations of function values compared with the same dependence obtained by the SFDA method

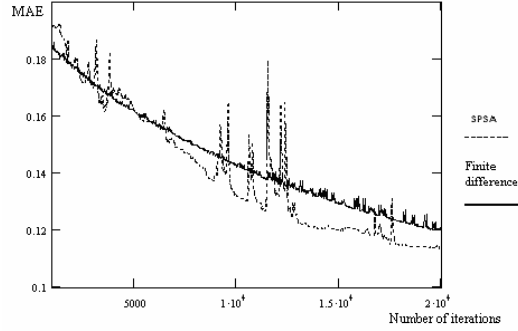


Figure 3. Minimization of MAE by SPSA and SFDA methods.

5.2. Optimal Design of Cargo Oil Tankers

In cargo oil tankers design, it is necessary to choose such sizes for bulkheads, that the weight of bulkheads would be minimal. After some details the minimization of weight of bulkheads for the cargo oil tank we can formulate like nonlinear programming task, [12]:

$$f(x) = \frac{5.885 \cdot x_4 (x_1 + x_3)}{x_1 + \sqrt{(x_3^2 - x_2^2)}} \rightarrow \min ,$$

$$\text{subject to } g_1(x) = x_2 x_4 \left(0.4 \cdot x_1 + \frac{1}{6} x_3 \right) - 8.94 \cdot \left(x_1 + \sqrt{(x_3^2 - x_2^2)} \right) \geq 0 ,$$

$$g_2(x) = x_2^2 x_4 \left(0.2 \cdot x_1 + \frac{1}{12} x_3 \right) - 2.2 \cdot \left(8.94 \cdot \left(x_1 + \sqrt{(x_3^2 - x_2^2)} \right) \right)^{\frac{4}{3}} \geq 0 ,$$

$$g_3(x) = x_4 - 0.0156 x_1 - 0.15 \geq 0 ,$$

$$g_4(x) = x_4 - 0.0156 x_3 - 0.15 \geq 0 ,$$

$$g_5(x) = x_4 - 1.05 \geq 0 ,$$

$g_6(x) = x_3 - x_2 \geq 0$, where x_1 - width, x_2 - debt, x_3 - lenght, x_4 - thickness.

Let us consider the application of SA to the minimization of the bulkheads weight by the penalty method. In Figure 4 the penalty function and the best feasible objective functions under the number of iterations minimized by SPSA with Uniform perturbation and SFDA method are depicted. In Figure 5 the averaged upper and lower bounds of the minimum are illustrated. For comparison the function minimum value is presented, taken from [12]. As we see, the linear estimators by order statistics make it possible to evaluate minimum with admissible accuracy and introduce rule for algorithm stopping when the confidence interval becomes less than certain small value.

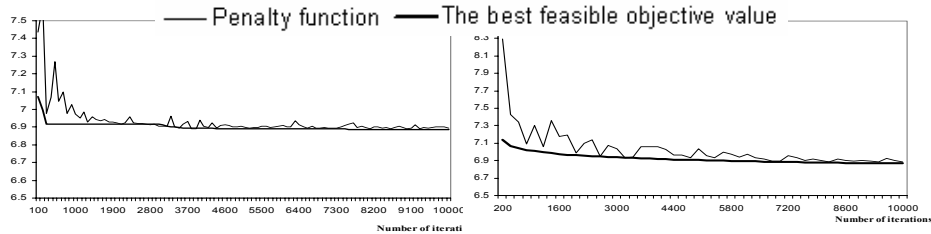
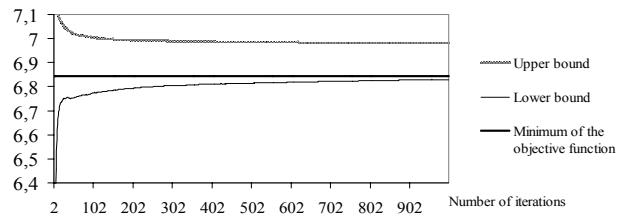


Figure 4. SPSA with Lipschitz perturbation and SFDA for the cargo oil target design.


 Figure 5. Confidence bounds of the minimum ($A=6.84241$, $T=100$, $N=1000$).

6. Conclusion

Application of SA to engineering problems has been considered comparing three algorithms. The rate of convergence of the developed approach was explored for the functions with a sharp minimum by computer simulation when there are no noises in computation of the objective function. Computer simulation by Monte-Carlo method has shown that the empirical estimates of the rate of convergence corroborate the theoretical estimation of the convergence order $O\left(\frac{1}{k^\gamma}\right)$, $1 \leq \gamma < 2$. The SPSA algorithms have appeared to be more efficient for small n than the SFDA approach. However, when the dimensionality of the task increases, the SFDA method becomes more efficient than SPSA algorithm according to the number of function value computed for optimization.

The linear estimator for the minimum value of optimized function has been proposed, using the theory of order statistics, and studied in experimental way. The estimator proposed are simple and depend only on the parameter of the extreme value distribution α . The parameter α is easily estimated, using the parameter of homogeneity of the objective function. Theoretical considerations and computer examples have shown that the confidence interval of the function minimum can be estimated with an admissible accuracy, when the number of iterations is increased. Finally, the developed algorithms were applied to the minimization of the mean absolute pricing error for parameter estimation in the

Heston stochastic volatility model and minimization of weight of bulkheads for cargo oil tanks demonstrate applicability for practical purposes.

References

1. V. Bartkutė, L. Sakalauskas. Convergence of Simultaneous Perturbation Stochastic Approximation in a Lipshitz class. *Lith. Mathematical Journal*, **44**, p. 603-608, ISSN 0132-2818 (in Lithuanian) (2004).
2. H. Chen. Estimation of the Location of the Maximum of a Regression Function Using Extreme Order Statistics. *Journal of multivariate analysis*, **57**, 191 – 214 (1996).
3. F. H. Clarke. Generalized gradients and applications. *Trans. Amer. Math. Soc*, **205**, 247 – 262 (1975).
4. Yu. M. Ermoliev. Methods of stochastic programming. *Nauka*, Moscow. (in Russian) (1976).
5. Yu. M. Ermoliev, V. I. Norkin, R. J-B Wets. The minimization of semicontinuous functions: mollifier subgradients. *Control and optimization*, **33**(1), 149-167 (1995).
6. O.N. Granichin, B.T. Poliak. Randomized algorithms for estimation and optimization with almost arbitrary errors. *Nauka*, Moscow. (in Russian) (2003).
7. P. Hall. On estimating the endpoint of a distribution. *Annals of Statistics*. **10**, 556-568 (1982).
8. S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, **6** (2), 327-343 (1993).
9. H.J. Kushner, G. G. Yin. Stochastic Approximation and Recursive Algorithms and Applications. *Springer*: N.Y., Heidelberg, Berlin (2003).
10. V.S. Mikhalevitch, A.M. Gupal, V.I. Norkin. Methods of Nonconvex Optimization. *Nauka*: Moscow (in Russian) (1987).
11. J. Mockus. Multiextremal problems in design. Moscow, Nauka (1967).
12. G.V. Reklaitis, A. Ravindran, K.M. Ragsdell. Engineering Optimization. Methods and Applications, Moscow. (In Russian) (1986).
13. J. C. Spall. Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, *Wiley, Hoboken*, NJ. (2003).
14. M.T. Vazan. Stochastic Approximation. Cambridge. *Transactions in Mathematics and Mathematical Physics*. Cambridge University Press, Cambridge (1969).
15. A. Žilinskas, A. Zhigljavsky. Methods of the global extreme searching. *Nauka*, Moscow. (In Russian) (1991).
16. A. Zhigljavsky. Branch and probability bound methods for global optimization. *Informatica*, Vilnius **1**(1), 125-140 (1990).