Minimum Tracking with SPSA and Applications to Image Registration

Oleg Granichin, Lev Gurevich and Alexander Vakhitov

Department of Mathematics and Mechanics, St. Petersburg State University
St. Petersburg, Russia

Abstract. An application of simultaneous perturbation stochastic approximation (SPSA) algorithm with two measurements per iteration to the problem of object tracking on video is discussed. The upper bound of mean square estimation error is determined in case of once differentiable functional and almost arbitrary noises. Weak restrictions on uncertainty allow to use random sampling instead of full pixelwise difference calculation. The experiments show significant increase in performance of object tracking comparing to the classical Lucas-Kanade algorithm. The results can be generalized to improve more recent kernel-based tracking methods.

1 Introduction

Object tracking on video is an important problem for security surveillance systems. One of possible approaches to this problem is to look for a disparity vector h between object's locations in consequent frames I_n and I_{n+1} of the video stream. Disparity vector h can include object size and other properties as well. Well known algorithm of B. Lucas and T. Kanade [1] uses the functional of pixelwise differences in consequent frames. $F_{n+1}(x) = \sum_{p \in O_n} \frac{1}{w(p)}((I_{n+1}(x+p) - I_n(x_n+p))^2$, where O_n is a set of object's pixels p in the n-th frame, $\sum_{p \in O_n} w(p) = 1$, x_n is a center of previous object's location, p represents the relative position of a pixel woth respect to the object's center. The functional needs to be minimized to find real disparity $x^*_{n+1} = argmin F_{n+1}(x)$. If the object's locations on consequent frames have significant intersection, then the algorithm of minimum tracking based on convex optimization can be applied to this problem. Stochastic optimization allows to use not all the pixels of the object's image O_n, but some random sample $R_n \subset O_n$ to compute the functional each iteration, because the convergence of the algorithm can be proved in this case. This property allows to descrease the amount of pixelwise difference measurements per algorithm iteration. The experiments' results in the end of the paper show significant increase in the performance of the method. The same approach of randomization of the pixel sample combined with srtochastic optimization can be used in more recent kernel-based object tracking algorithms, described in [2].

More specifically, in this article the convergence of the SPSA (simultaneous perturbation stochastic optimization) algorithm of stochastic optimization class will be proved in case of minimum tracking. The functional measurement $y_n = F(x_n, w_n)$ depends

Granichin O., Gurevich L. and Vakhitov A. (2009).
Minimum Tracking with SPSA and Applications to Image Registration.
In *Proceedings of the International Workshop on Networked embedded and control system technologies: European and Russian R&D cooperation*, pages 66-74

on two arguments, one of which is the actual point of measurement (in case described above, $x_n = h_n$) and the second one is the uncertainty presented by random variable w, $\int P(dw) = 1$. The minimization is understood in average risk setting, such that $f_n(x) = E_w F(x, w, n) \to \min$.

In case of object tracking on video, we consider $F(x, w, n+1) = \sum_{p \in R_n} (I_{n+1}(x + p) - I_n(x_n + p))^2$, $f(x, n) = \sum_{p \in O_n} \frac{1}{w(p)} F(x, w, n)$. The average risk setting is perfectly suitable here, since we would like to evaluate a pixelwise difference on a randomly chosen subset of pixels. We demonstrate this approach in the example provided below, where the pixelwise difference is computed only 2 times per iteration.

Problem of functional optimization arises in many practical cases. While in some cases extreme points could be found analytically, many engineering applications deal with unknown functional, which can only be measured in selected points with possible noise. In some cases functional itself could vary over time and its extreme points could drift. In this case problem setting could be different, depending on goals of optimization and possible measurements. In general, there are two different variants of a function behavior over time - it has a limit function, to which it tends when time goes to infinity, or there is no such function [3]. In this paper we consider the second variant.

Non-stationary optimization problems can be described in discrete or continuous time. In our paper we consider only discrete time model. Let $f(x, n)$ be a functional we are optimizing at the moment of time n ($n \in \mathbf{N}$). In book [4] Newton method and gradient method are applied to problems like that, but they are applicable only in case of two times differentiable functional and $l < \nabla^2 f_k(x) < L$. Both methods require possibility of direct measurement of gradient in arbitrary point.

In real world measurement always contains noise. Sometimes the algorithms that perfectly solve the problem on paper do not provide good estimates in practical cases. Robustness is important in engineering applications. For problems with noise the Robbins-Monro and Kiefer-Wolfowitz stochastic approximation algorithms were developed in 1950s. The history of development of such algorithms is described in [5, 8]. Common approach used in these algorithms can be formalized in a following way:

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \alpha_n \hat{g}_n(\hat{\theta}_n), \tag{1}$$

where $\{\hat{\theta}_n\}$ —is the sequence of extreme points estimates generated by algorithm, g_n — pseudo-gradient (replacing the gradient from Newton method). Pseudo-gradient has to approximate the true gradient. The important properties of algorithms described in this form are simplicity and recurrence. Because of these properties they are often applied in different areas.

Kiefer-Wolfowitz algorithm with randomized differences is also known as SPSA (Simultaneous perturbation stochastic approximation). Algorithms of this type with one or two measurements on each iteration appeared in papers of different researchers in the end of the 1980s [9–12]. Later in the text we will refer to this class of algorithms as SPSA for simplicity. These algorithms are known for their applicability to problems with almost arbitrary noise [8]. The measurement noise should be bounded and only slightly correlated with perturbation on each iteration. Moreover, the number of measurements made on each iteration is only one or two and is independent from the number of dimensions of the state space d. This property sufficiently increases the rate of convergence of the algorithm in multidimensional case ($d >> 1$), comparing to algorithms, that use direct estimation of gradient, that requires $2d$ measurements of function in case if direct measurement of function gradient is impossible. Detailed review of development of such methods is provided in [8, 13].

Stochastic approximation algorithms were initially proven in case of the stationary functional. The gradient algorithm for the case of minimum tracking is provided in [4], however the stochastic setting is not discussed there. Further development of these ideas could be found in paper [3], where conditions of drift pace were relaxed. The book [5] uses the ordinary differential equations (ODE) approach to describe stochastic approximation. It addresses the issue of applications of stochastic approximation to tracking and time-varying systems in a following way: it is proven there that when the step size goes to zero in the same time as the number of the algorithm's iterates over a finite time interval tends to infinity, then the minimum estimates tend to true minimum values. This is not the case here, since we consider the number of iterates per unit of time to be fixed. In this paper we consider application of simultaneous perturbation stochastic approximation algorithm to the problem of tracking of the functional minimum. SPSA algorithm does not rely on direct gradient measurement and is more robust to non-random noise than gradient-based methods mentioned earlier. The most closely case was studied in [6], but we do not use the ODE approach and we establish more wide conditions for the estimates stabilization. In the following section we will give the problem statement that is more general than in [3, 4], in the third section we provide the algorithm and prove its estimates mean squared stabilization. In the last section we illustrate the algorithm, applying it to minimum tracking in a particular system.

2 Problem Statement

Consider the problem of minimum tracking for average risk functional:

$$f(x, n) = E_w\{F(x, w, n)\} \to \min_x, \tag{2}$$

where $x \in \mathbf{R}^d$, $w \in \mathbf{R}^p$, $n \in \mathbf{N}$, $E_w\{\cdot\}$ — mean value conditioned on the minimal σ-algebra in which w is measurable.

The goal is to estimate θ_n — minimum point of functional $f(x, n)$, changing over time:

$$\theta_n = argmin_x f(x, n).$$

Let us assume that on the iteration we can do a following measurement:

$$y_n = F(x_n, w_n, n) + v_n, \tag{3}$$

where x_n — arbitrary measurement point chosen by algorithm, w_n — random values, that are non-controlled uncertainty and v_n — observation noise.

Time in our model is discrete and implemented in number of iteration n.

To define the quality of estimates we will use the following definition:

Definition 1. *[7] A random matrix (or vector) sequence* $\{A_k, k \geq 0\}$ *defined on the basic probability space* $\{\Omega, \mathcal{F}, P\}$ *is called* L_P*-stable* ($p > 0$) *if*

$$\sup_{k \geq 0} E\|A_k\|^p < \infty.$$

We will use the definition 1 in case of $p = 2$.

Further we will consider generation of sequence of estimate $\{\hat{\theta}_n\}$ for problem (2), satisfying the definition 1 for p=2, in following conditions.

We will assume that drift of the minimum point is limited in following sense:

$$(\mathbf{A}) \quad \|\theta_n - \theta_{n-1}\| \leq A.$$

Function $f(\cdot, n)$ is a strictly convex function for each n:

$$(\mathbf{B}) \quad \langle \nabla f(x, n), x - \theta_n \rangle \geq \mu \|x - \theta_n\|^2.$$

Gradient $\nabla F(\cdot, w, n)$ is Lipschitz with constant B, $\forall n, \forall w$:

$$(\mathbf{C}) \quad \|\nabla F(x, w, n) - \nabla F(y, w, n)\| \leq B\|x - y\|.$$

Average difference of function $F(x, \cdot, n)$ in any point x for moments n and $n + 1$ is limited in a following way:

$$(\mathbf{D}) \quad E_{w_1, w_2}|F(x, w_1, n + 1) - F(x, w_2, n)| \leq C\|x - \theta_n\| + D.$$

($\mathbf{E}$) Local Lebesgue property for the function $\nabla F(w, x)$: $\forall x \in \mathbb{R}^d \, \exists$ neigbourhood U_x such that $\forall x' \in U_x$ $\|\nabla F(w, x)\| < \Phi_x(w)$ where $\Phi_x(w) : \mathbb{R}^p \to \mathbb{R}$ is integrable by w: $\int_{\mathbb{R}^q} \Phi_x(w) dw < \infty$

The last condition is necessary for the commutation of differentiation and integration operations, that is used to change order of expectation and gradient in the proof of the theorem. For more discussions about such properties see [18].

($\mathbf{F}$) For the observation noise v_n the following conditions are satisfied:

$$|v_{2n} - v_{2n-1}| \leq \sigma_v,$$

or if it has statistical nature then:

$$E\{|v_{2n} - v_{2n-1}|^2\} \leq \sigma_v^2.$$

Here we should make several notes:

1). Sequence $\{v_n\}$ could be of non-statistical but unknown deterministic nature. 2). Constraint ($\mathbf{A}$) allows both random and deterministic drift. In certain cases Brownian motion could be described without tracking. Tracking is needed when there is both determined and non-determined aspects of drift. Similar condition is introduced in [4], it is slightly relaxed in [3]. For example it could be relaxed in a following way:

$$(\mathbf{A'}) \quad \theta_n \leq A_1 \theta_{n-1} + A_2 + \xi_n,$$

where ξ_n is random value.

In this paper we will only consider drift constraints in form **(A)**. Mean square stabilization of estimation under condition **(A)** implies its applicability to wide variety of problems.

3 Algorithm and Stabilization of Estimates

In this section we are introducing a modification of SPSA algorithm provided by Chen et al [16], which takes one perturbed and one non-perturbed measurement on each step.

Let perturbation sequence $\{\Delta_n\}$ be an independent sequence of Bernoulli random vectors, with component values $\pm 1/\sqrt{d}$ with probability $\frac{1}{2}$. Let vector $\hat{\theta}_0 \in \mathbb{R}^d$ be the initial estimation. We will estimate a sequence of minimum points $\{\theta_n\}$ with sequence $\{\hat{\theta}_n\}$ which is generated by the algorithm with fixed stepsize:

$$\begin{cases} x_{2n} = \hat{\theta}_{2n-2} + \beta\Delta_n,\ x_{2n-1} = \hat{\theta}_{2n-2}, \\ y_n = F(x_n, w_n, n) + v_n, \\ \hat{\theta}_{2n} = \hat{\theta}_{2n-2} - \frac{\alpha}{\beta}\Delta_n(y_{2n} - y_{2n-1}), \\ \hat{\theta}_{2n-1} = \hat{\theta}_{2n-2}. \end{cases} \tag{4}$$

(**G**) We will further assume that random values Δ_n generated by algorithm are not dependent on $\hat{\theta}_k$, w_k , $\hat{\theta}_0$ and on v_k (if they are assumed to have random nature) for $k = 1, 2, \ldots, 2n$.

Let us define $H = (\alpha^2 B + \frac{\alpha^2}{\beta^2} D + \frac{\alpha^2}{\beta^2}\sigma_v)(\beta B + C) + \alpha AB + \alpha\beta B + A$. Let K and $\delta > 0$ be constants satisfying following condition:

$$K = 1 - 2\alpha\mu + \frac{\alpha^2}{\beta} B + \frac{\alpha^2}{\beta^2} C + \delta < 1.$$

Denote $L = \frac{H^2}{\delta} + A^2 + 2\alpha\beta AB + \frac{\alpha^2}{\beta^2}((\beta^2 B + D)^2 + \sigma_v^2 + 2(\beta^2 B + D)\sigma_v)$.

Theorem 1. *[15, 19]* **.** *Assume that conditions* **(A)–(G)** *on functions f, F and ∇F and values θ_n, $\hat{\theta}_n$, v_n, w_n, y_n and Δ_n are satisfied. Let us further assume that constants $\alpha, \beta > 0$ satisfy the inequality:*

$$0 < \delta < 2\alpha\mu - \frac{\alpha^2}{\beta} B - \frac{\alpha^2}{\beta^2} C. \tag{5}$$

Then estimates provided by the algorithm (4) stabilize in mean squares and following inequality holds:

$$E\{\|\theta_n - \hat{\theta}_n\|^2\} \le K^n \|\theta_0 - \hat{\theta}_0\|^2 + \frac{L(1 - K^n)}{1 - K}. \tag{6}$$

Note that Theorem 1 provides asymptotically effective value for the estimates:$\bar{L} = L/(1 - K)$.

Conditions **(A)–(C)**,**(E)–(G)** are standard for SPSA algorithms [8]. Earlier the proof of the similar theorem was given in [15] with more strict conditions. See the proof of Theorem 1 in appendix.

The condition (5) on α can be satisfied only when inequality $0 < \alpha < 2\mu\frac{\beta^2}{B\beta+C}$ is true. It follows from the result of the Theorem 1 that $E\{\|\theta_n - \hat{\theta}_n\|^2\} \leq O(\frac{A^2}{\alpha})$ $(\alpha \to 0)$ which leads to a simple decision rule: to presume the upper bound α should tend to zero with the same pace as A^2. α can be arbitrarily close to 0, which diminishes the effects of the gradient approximation bias and the noise.

To build the upper bound of the algorithm's estimates error using Theorem 1, it is needed to find α and β satisfying the condition (5), then to find δ which gives minimum value of the fraction $\frac{L}{1-K}$. Using the resulting bound obtained, it is possible to reduce it by applying the ideas concerning the relation of α and dreif parameters such as A and the function parameters such as B or μ. The resulting estimates behavior will be the trade-off between the tracking ability and noise sensitivity. The similar problem of algorithm parameters choosing was studied in [17] for the same algorithm but in the linear case.

4 Application to Object Tracking

Here we would like to present the derivation of the basic convergence conditions (A)-(F), which is done considering the application of the theorem to the object tracking problem. The problem itselkf was descibed in the beginning of the article. The conditions should be true in some neighbourhood of the object.

$$(\mathbf{A}) \quad \|x_n - x_{n-1}\| \leq A.$$

That is a constraint on the speed of object's movement.

Function $f(\cdot, n)$ is a strictly convex function for each n:

$$(\mathbf{B}) \quad \langle \sum_{p \in O_n} \frac{1}{w(p)} 2(I_n(x+p) - I_{n-1}(x_{n-1}+p)), x - x_n^* \rangle \geq \mu \|x - x_n^*\|^2.$$

This condition is about the difference between the object and the background.

Gradient $\nabla F(\cdot, w, n)$ is Lipschitz with constant B, $\forall n, \forall w$:

$$(\mathbf{C}) \quad \|2I_n(x+p) - 2I_n(y+p)\| \leq B\|x - y\|.$$

This describes the local proximity of pixels in the picture.

Average difference of function $F(x, \cdot, n)$ in any point x for moments n and $n+1$ is limited in a following way:

$$\begin{aligned}(\mathbf{D}) \quad \sum \frac{1}{w(p_1)w(p_2)} &|(I_n(x+p_1) - I_{n-1}(x+p_2) + I_{n-1}(x+p_1) - I_{n-2}(x+p_2)) \\ (I_n(x+p_1) - I_{n-1}(x+p_2) &- I_{n-1}(x+p_1) + I_{n-2}(x+p_2))| \leq C\|x - x_n^*\| + D.\end{aligned}$$

This condition means that the object's pixels are similar to each other, and in the same time different from the background.

$(\mathbf{E})$ Local Lebesgue property for the function $\nabla F(w, x)$: $\forall x \in \mathbb{R}^d$ $\exists$ neigbourhood U_x such that $\forall x' \in U_x$ $\|\nabla F(w, x)\| < \Phi_x(w)$ where $\Phi_x(w) : \mathbb{R}^p \to \mathbb{R}$ is integrable by w: $\int_{\mathbb{R}^q} \Phi_x(w) dw < \infty$

The last condition is necessary for the commutation of differentiation and integration operations, that is used to change order of expectation and gradient in the proof of the theorem. For more discussions about such properties see [18].

$(\mathbf{F})$ For the observation noise v_n the following conditions are satisfied:

$$|v_{2n} - v_{2n-1}| \le \sigma_v,$$

or if it has statistical nature then:

$$E\{|v_{2n} - v_{2n-1}|^2\} \le \sigma_v^2.$$

The noise boundary depends on the light conditions and camera properties.

5 Object Tracking Example

The application of the algorithm to the problem of object tracking on video is demonstrated by the following example. We took an image presented at the Fig. 5 as a second image, I_{n+1}. The object is a rectangle with a picture of butterfly, of size 208x120=24960 pixels. The image I_n contains the object in tyhe lower-right corner. Vector h is 2-dimensional.

Fig. 1. Experiment with tracking of an object on video.

The results are presented at the Table 1. We see significant increase in performance gained from the application of more advanced optimization technique.

6 Conclusions

In our work we apply the SPSA-type algorithm to the problem of object tracking on video. The novelty of the approach comes from the use of stochastic optimization instead of standard pseudogradient technique to find a location of an object. SPSA does

Table 1. The results of the experiment: amount of pixelwise difference calculation for Lucas-Kanade and SPSA-based algorithms.

Characteristic	Lucas-Kanade	SPSA-based
Number of measurements per iteration	24960	2
Number of iterations	3	1201
Number of measurements	74880	2402

not require possibility of direct gradient measurement, needs only 2 function measurement on each iteration and once differentiable function. Drift is only assumed to be limited, which includes random and directed drift. It was proven that the estimation error of this algorithm is limited with constant value. The modeling was performed on a multidimensional case.

The results show the potential performance gains of using the SPSA-type algorithms for tracking of objects on video. Probably, more sofisticated algorithms based on optimization such as kernel-based methods of tracking [2] can also be improved by the same technique. Authors want to try such application in future.

References

1. Lukas B., Kanade T., "An Iterative Registration Technique with an Application to Stereo Vision,". *In Proc. Imaging Understanding Workshop.*, 1981, pp. 121-130.
2. Tsin Y. *Kernel Correlation as an Affinity Measure in Point-Sampled Vision Problems*. PhD Thesis, Carnegie Mellon University, 2003.
3. Popkov A. Yu., "Gradient methods for nonstationary unconstrained optimization problems," *Automat. Remote Control*, No. 6, pp. 883–891, 2005.
4. Polyak B. T., *Introduction to Optimization.* New York: Optimization Software. 1987.
5. Kushner H., Yin G., *Stochastic Approximation and Recursive Algorithms and Applications.* Springer. 2003.
6. Borkar V. S., *Stochastic Approximation. A Dynamical Systems Viewpoint.* Cambridge University Press. 2008.
7. Guo L., "Stability of recursive stochastic tracking algorithms," *SIAM J. Control and Optimization*, vol. 32, No 5, pp. 1195–1225, 1994.
8. Granichin O. N., Polyak B. T., *Randomized Algorithms of an Estimation and Optimization Under Almost Arbitrary Noises.* Moscow: Nauka. 2003.
9. Granichin O. N., "Procedure of stochastic approximation with disturbances at the input," *Automat. Remote Control*, vol. 53, No 2, part 1, pp. 232–237, 1992.
10. Granichin O. N., "A stochastic recursive procedure with dependent noises in the observation that uses sample perturbations in the input," *Vestnik Leningrad Univ. Math.*, vol. 22, No 1(4), pp. 27–31, 1989.
11. Polyak B. T., Tsybakov A. B., "Optimal order of accuracy for search algorithms in stochastic optimization," *Problems of Information Transmission*, vol. 26, No 2, pp. 126–133, 1990.
12. Spall J. C., "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 332–341, 1992.
13. Spall J. C., "Developments in stochastic optimization algorithms with gradient approximations based on function measurements," *Proc. of the Winter Simulation Conf.*, pp. 207–214, 1994.

14. Katkovnik V. Ya., Khejsin V. E., "Dynamic stochastic approximation of polynomial drifts," *Automat. Remote Control*, vol. 40, No 5, pp. 700–708, 1979.
15. Gurevich L., Vakhitov A., "SPSA Algorithm for Tracking," *In Proc. 12th International Student Olympiad on Automatic Control (Baltic Olympiad)*, pp. 52–57, 2008.
16. Chen H.-F., Duncan T. E., Pasik-Duncan B. A "Kiefer-Wolfowitz algorithm with randomized differences," *IEEE Trans. Automat. Contr.*, vol. 44, No 3, pp. 442–453, 1999.
17. Granichin O. N., "Linear Regression and Filtering Under Nonstandard Assumptions (Arbitrary Noise)," *IEEE Trans. Automat. Contr*., vol. 44, No 3, pp. 442–453, 1999.
18. Van der Vaart H. R., Yen E. H., "Weak Sufficient Conditions for Fatou's Lemma and Lebesgue's Dominated Convergence Theorem," *Mathematics Magazine*, vol. 44, No 3, pp. 442–453, 1968.
19. Vakhitov A., Granichin O., Gurevich, L. SPSA Algorithm In Case of Tracking, *Automat. Remote Contr.*, to be printed. 2009.