Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

SPSA for non-smooth optimization with application in ECG analysis

László Gerencsér
Computer and Automation Institute of
the Hungarian Academy of Sciences,
Budapest, Kende 13-17,
H-1111 Hungary,
gerencser@sztaki.hu

György Kozmann
Institute for Materials Science of
the Hungarian Academy of Sciences,
Budapest, Konkoly Thege út 29/33,
H-1121 Hungary,
kozmann@r1.atki.kfki.hu

Zsuzsanna Vágó
Computer and Automation Institute of
the Hungarian Academy of Sciences,
Budapest, Kende 13-17, Hungary,
H-1111 Hungary
vago@oplab.sztaki.hu

Abstract

It is shown that non-smooth optimization problems can be solved by a suitable extension of the simultaneous perturbation stochastic approximation or SPSA method (cf. [6]). The new optimization method has been tested in a min-max classification problem using both simulated and real data. The latter are ECG signals which were collected for the detection of so-called late potentials.

1 Introduction

The motivation of this research is to develop a fast, real-time classification method using a min-max class-distortion measure. The intended application area is real-time ECG analysis. For this purpose we are going to generate typical signals, which is performed by solving a min-max problem.

The solution of the min-max problem is non-trivial for large data-sets (cf. [5]). The specification for the present application is high speed and relatively low accuracy. This purpose is met by some randomization algorithm. We propose to use simultaneous perturbation stochastic approximation or SPSA method due to J.C. Spall (cf. [6]). The convergence properties of this function minimization algorithm is well-understood, (cf. also [3]), The standard SPSA procedure is applicable only to functions which are three-times continuously differentiable. The major result of the paper is the extension of SPSA to non-smooth problems. The new method gives a suboptimal solution. We have carried out numerical experiments for the solution of the min-max problem on both simulated and real data.

2 Non-smooth optimization via SPSA

Let $L(\theta), \theta \in \mathbb{R}^p$ be a cost-function, which is Lipschitz-continuous with Lipschitz-constant, say L, but not differentiable. On the contrary, we assume that $L(\theta)$ has a local, *strictly non-smooth minimum* at θ^* in the sense that for some $\delta, c > 0$, $|\theta - \theta^*| \leq \delta$ implies

$$L(\theta) - L(\theta^*) \geq c|\theta - \theta^*|. \tag{1}$$

A typical example is generated by defining

$$L(\theta) = \max_{i=1,\ldots,m} g_i(\theta),$$

where $g_i(\theta)$ themselves are Lipschitz-continuous, convex functions (cf. (2.2) of [5]).

For the minimization of $L(\theta)$ a number of methods are available in the theory of non-smooth optimization (cf. [5]). However their efficiency depends considerably on the complexity of the function $L(\theta)$ itself.

Our method is based on a randomization technique. First we generate a sufficiently smooth approximation $\overline{L}(\theta)$ of $L(\theta)$ by locally averaging it. Let $h(\theta)$ be a sufficiently smooth density function with bounded support, say $h(\theta) = 0$ for $|\theta_i| \geq \delta_m$ for all $i = 1, \ldots, p$, with some $\delta_m > 0$. Then the approximation of $L(\theta)$ will be defined as

$$\overline{L}(\theta) = \int L(\theta + \delta\theta) h(\delta\theta) d\delta\theta. \tag{2}$$

0-7803-4394-8/98 $10.00 © 1998 IEEE

Since the mass distribution defined by $h(\theta)$ is concentrated around zero, $\overline{L}(\theta)$ will be a good approximation of $L(\theta)$. On the other hand, since $L(\theta)$ has a strictly nonsmooth minimum at θ^*, it follows that the approximating function $\overline{L}(\theta)$ has a local minimum inside the sphere of radius δ_0 around θ^* for any δ_0, if δ_m is sufficiently small.

The function $\overline{L}(\theta)$ is not computable explicitly, however, for the purpose of iterative minimization we can easily generate a Monte-Carlo estimate $\hat{L}_n(\theta)$ of $\overline{L}(\theta)$. For any fixed θ at any time n generate a random perturbation $\delta\theta_n$ with density function $h(\theta)$ and set

$$\hat{L}_n(\theta) = L(\theta + \delta\theta_n). \qquad (3)$$

Thus $\hat{L}_n(\theta)$ is an unbiased estimation of $\overline{L}(\theta)$. Note, however, that the "measurement error" defined as

$$\varepsilon_n = \hat{L}_n(\theta) - \overline{L}(\theta) = L(\theta + \delta\theta_n) - \overline{L}(\theta) \qquad (4)$$

depends on θ, i.e. we have a *state-dependent noise*.

Since there is no direct way to estimate the gradient of $\overline{L}(\theta)$. Thus we have to resort to numerical differentiation using the noise-corrupted function values $\hat{L}_n(\theta)$. This can be done in a very effective manner, using a simultaneous random perturbation of the parameters (cf. [6]).

The basic ingredients of the SPSA method are given in [6] in which the almost sure convergence and asymptotic normality of the estimator sequence has been established. In [2] an almost sure convergence rate has been given for a modified, truncated version of the SPSA algorithm (Theorem 3 in [2]). A further modification of the SPSA algorithm was developed in [3], where the assumed boundedness condition of [6] is replaced by a resetting mechanism, enforcing the estimator sequence to stay in a bounded domain. In the cited papers a rate of convergence result is derived for the moments of the estimation error.

A major technical condition in all the cited works is that it is assumed that the measurement noise is *state-independent*. This condition is violated by the measurement noise ε_n defined by (4), and thus previous results are not directly applicable. However, it can be shown that the methods can be extended to cover the state-dependent case, and thus the application of SPSA for the present problem is justified (cf. [4]).

3 Computational experiments

We have tested the non-smooth SPSA-method on both simulated and real data. In the case of *simulated data* we considered the function $f(x) = \sum_{i=1}^{n} |x_i|$. In a typical experiment the dimension was 100, and the coordinates of the starting point were chosen as independent random variables, with standard Gaussian distribution. The smoothing parameter in the randomization step was $\delta = 0.001$. We have obtained a reasonably accurate result in 100 steps.

In the case of *real data* we considered three ECG records, taken from two healthy subjects and a patient with frequent ectopic beats. The length of the records were 2.5 min, and from all the technical aspects the measurements were conform with the recommendations formulated in [1] by an international task force. In all cases data were recorded on three leads for the same heart cycle, which is represented by 33 numbers; the middle point representing the fiducial point of the cycle. Thus our data sets consisted of points in 99 dimension. We identified one outlier that is obviously due to measurement error.

Acknowledgement

The authors express their thanks to J.C. Spall and J.L. Maryak for cooperating in this research, and to G. Bortolan for helpful discussions. This research was partially supported by by the National Research Foundation of Hungary (OTKA) under Grant T 020984.

References

[1] G. Breithardt, M.E. Cain, N. El-Sherif, N. Flowers, V. Hombach, M. Janse, M.B. Simpson, and G. Steinbeck. Standards for analysis of ventricular late potentials using high resolution or signal-averaged electrocardiography. *Eur. Heart J.*, 12:473–480, 1991.

[2] H.F. Chen, T.E. Duncan, and B. Pasik-Duncan. A stochastic approximation algorithm with random differences. In J.Gertler, J.B. Cruz, and M. Peshkin, editors, *Proceedings of the 13th Triennal IFAC World Congress, San Francisco, USA*, pages 493–496, 1996. Volume editors: R.Bitmead, J. Petersen, H.F. Chen and G. Picci.

[3] L. Gerencsér. Rate of convergence of moments of Spall's SPSA method. In *European Control Conference, ECC'97*, volume 3, pages WE–E–F1, 1997.

[4] L. Gerencsér. SPSA with state-dependent noise–a tool for direct adaptive control. In *Proceedings of the Conference on Decision and Control, CDC 37*, 1998.

[5] C. Lemarechal. Nondifferentiable optimization. In G.L. Nemhauser et al., editor, *Handbooks in Operations Research and Management Science*, pages 529–572. Elsevier Science Publishers B.V. (North-Holland, 1989.

[6] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Contr.*, 37:332–341, 1992.