

**PARAMETER ESTIMATION IN A HIGHLY NON-LINEAR MODEL
USING SIMULTANEOUS PERTURBATION STOCHASTIC
APPROXIMATION**

James E. Whitney, II Kerron Duncan, Maria Richardson
Isaac Bankman²

Morgan State University
Department of Electrical and Computer Engineering
5200 Perring Parkway
Baltimore, Maryland 21239

²Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Key Words: Optimization; SPSA.

ABSTRACT

Often it is necessary to estimate the parameters of a model or unknown system. Various techniques exist to accomplish this task, including Kalman and Wiener filtering, least-mean-square (LMS) algorithms, and the Levenberg-Marquardt(L-M) algorithm. These techniques require an analytic form of the gradient of the function of the parameters to be estimated. A key feature of the simultaneous perturbation stochastic approximation (SPSA) method is that it is a gradient-free optimization technique (Spall; 1992,1998a,b, 1999). In the current problem, the function of parameters to be identified is highly non-linear and of sufficient difficulty that obtaining an

analytic form of the gradient is impractical. Therefore, in this paper the performance of the SPSA algorithm will be examined in terms of parameter selection, data requirements, and convergence performance on this non-linear problem. Results will be reported on both a first-order "standard" implementation of SPSA and on a second-order version of SPSA that tends to enhance convergence.

1. INTRODUCTION

In estimating optimum parameters of a model or time series, there are several factors which must be considered when deciding on the appropriate optimization technique. Among these factors are convergence speed, accuracy, algorithm suitability, complexity, and computational costs in terms of time (coding, run-time, output) and power. In the current problem it is necessary to estimate the parameters of a geometrical object in real time. The model describing these parameter relationships is highly non-linear thus requiring the use of an iterative numerical technique. In addition, the complexity of the resulting loss² function suggests the use of a gradient-free technique. Initially, a batch L-M algorithm was tried using an algorithm which approximated the gradient from the input data. For the current model, the L-M had problems with convergence to local minima. Next, a simultaneous perturbation stochastic approximation (SPSA) algorithm was programmed and evaluated.

2. SPSA ALGORITHM

For the general optimization problem, the optimum set of parameters is obtained when

$$\frac{\partial L(\theta)}{\partial \theta} = 0$$

where $L(\theta)$ is the function to be optimized, and θ is the vector of parameters to be optimized. In this work $L(\theta)$ is a loss function describing the current best fit between physical measurement data and the model output. Thus, the SPSA algorithm is used to recursively optimize the parameter vector, θ . Unlike the Keifer-Wolfowitz stochastic approximation (SA) algorithm (Keifer, J. and Wolfowitz, J., 1952; Blum, J., 1954), which perturbs and optimizes over each parameter in turn, the SPSA algorithm simultaneously perturbs and optimizes over the entire parameter space. This increases algorithm efficiency and decreases the number of iterations necessary for a given problem.

The general steps in implementing the SPSA algorithm are: 1) initialization and coefficient selection, 2) generation of the simultaneous perturbation, 3) loss function evaluation, 4) gradient approximation, 5) updating of parameter vector $\hat{\theta}$, and 6) iteration (or, termination).

The recursive update form for the parameter vector is given by

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}_k(\hat{\theta}_k) \quad (1)$$

where,

a_k - weight, or gain constant for the current iteration

\hat{g}_k - gradient estimate for the current iteration

The i^{th} element of the gradient estimate, $\hat{g}_k(\hat{\theta})$ is given by

$$\hat{g}_k(\hat{\theta}) = \frac{y(\hat{\theta}_k + c_k \Delta_k) - y(\hat{\theta}_k - c_k \Delta_k)}{2c_k \Delta_k} \quad (2)$$

The term $\hat{\theta}_k \pm c_k \Delta_k$ represents a perturbation to the optimization parameters about the current estimate. Similar to a standard SA form, c_k is a small, positive weighting value. The vector Δ_k is a vector of zero-mean random variables which must have bounded inverse moments. One valid choice for Δ_k is a vector of Bernoulli-distributed, i.e., ± 1 , random perturbation terms.

The second-order SPSA algorithm (Spall, 1999) incorporates the first-order algorithm, usually at a reduced number of iterations, to do an initial estimate

of the optimum values for the parameters, θ . These values are then used as the starting point for the second-order algorithm. The second-order algorithm makes use of the Hessian matrix to increase the rate-of-convergence. Although the second-order algorithm is more costly in terms of computer resources, the decrease in the number of iterations needed to reach convergence should offset this higher computational cost.

3. APPLICATION

Next, first- and second-order SPSA algorithms were implemented to estimate the unknown parameters of the highly non-linear physical model. The model was generated to represent the complexity of the data expected from lidar returns and is given by

$$\rho(\beta, \phi; x, z) = (\cos \phi)(\cot^2 \beta)(z \sin \beta) \tan^{-1} \left(\frac{x}{z} \right) \quad (3)$$

where,

- β, ϕ - shape parameters to be estimated
- z - height along shape
- x - width along shape

The independent variables x and z are known. z is the vertical distance up the geometrical body through which a horizontal plane is extended, and is a constant for any particular data set. The returns, $\rho(\beta, \phi; x, z)$, are calculated at equal x intervals across this plane. The parameters β and ϕ are unknown and are initialized based on knowledge of their expected ranges. Under different scenarios it may be required to estimate one, or more, of the model parameters in various combinations.

The loss function, $L(\theta)$, describing the current best fit between the measurement data and the model output is given by

$$L(\theta) = [\rho(\beta, \phi; x, z) - \rho(\hat{\beta}, \hat{\phi}; x, z)]^2 \quad (4)$$

TABLE I

β , actual value	β , estimated value
10.0	9.998
15.0	15.000
20.0	20.002
25.0	24.998
30.0	30.000
33.0	33.002
38.0	37.999

4. RESULTS

Several scenarios were investigated, including; first-order algorithm, one-parameter; first-order algorithm, two-parameter; and second-order algorithm, two-parameter.

The results obtained from the first-order, one-parameter algorithm are illustrated in Table I. In this scenario, β was the parameter to be estimated. The parameter ϕ was known and had a value of 40.0. The measurement data density is 4000 samples-per-unit-length, with algorithm convergence parameters $\alpha = 0.702$, and $\gamma = 0.101$. The initial value of $\hat{\theta}_0$ was 20.0.

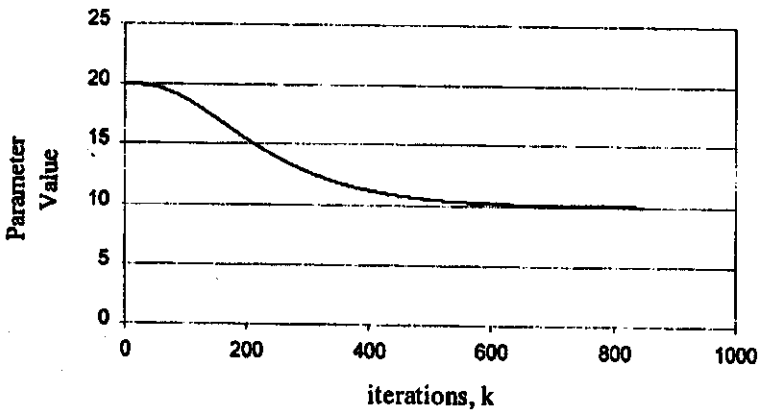
In a second trial, the SPSA algorithm was programmed to estimate the parameter ϕ . The parameter β was known and had a value of 10.0. Table II illustrates the results.

Figures 1 and 2 illustrate convergence curves for estimation of the model parameter ϕ . In these figures, the value of β was 10.0, and the actual values of ϕ are 10 and 45, respectively. In both cases the initial value of ϕ was 20.

In the first-order two-parameter example, the parameter vector has been modified so that both parameters β and ϕ are estimated by the first-order SPSA algorithm. The algorithm was run for various combinations of β and ϕ . These results are shown in Table III.

TABLE II

ϕ , actual value	ϕ , estimated value
10.0	9.998
15.0	15.000
18.0	18.002
23.0	22.999
30.0	30.000
37.0	37.002
45.0	45.000

FIG. 1 Convergence Curve, $\phi = 10$

The convergence characteristics for this scenario are illustrated in Figure 3, where the true parameter values are $\beta = 12.0$, and $\phi = 25.0$

The results of the second-order two-parameter simulation were similar to those of the first-order, two-parameter example and are illustrated in Table IV. In this case, the number of iterations performed in the first-order section of the algorithm was restricted to 50. The remaining iterations were then performed in the second-order section of the algorithm.

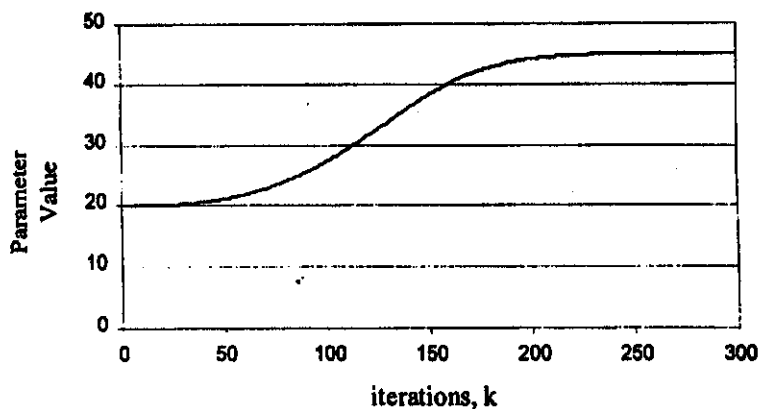
FIG. 2 Convergence Curve, $\phi = 45$

TABLE III

β , actual value	ϕ , actual value	β , estimated value	ϕ , estimated value
12.0	25.0	11.8	26.50
15.0	15.0	16.50	12.00
15.0	45.0	18.87	39.89

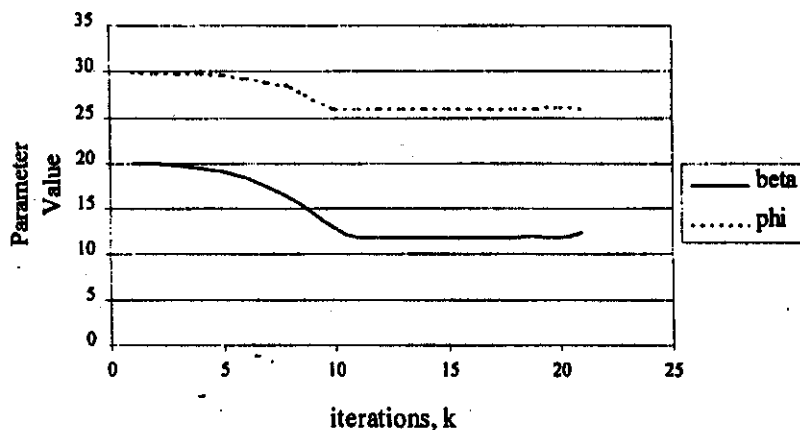
FIG. 3 Convergence Curves; $\beta = 12$, $\phi = 25$

TABLE IV

β , actual value	ϕ , actual value	β , estimated value	ϕ , estimated value
12.0	25.0	9.02	24.84
15.0	15.0	14.98	15.39
15.0	45.0	14.40	47.40

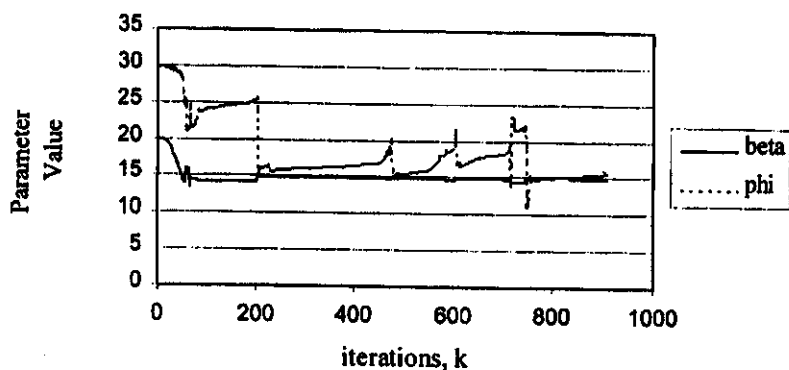
FIG. 4 Convergence Curves; $\beta = 15$, $\phi = 15$

Fig. 4 illustrates the convergence characteristics of the second-order SPSA algorithm for the parameter set $\beta = 15$, $\phi = 15$.

It should be noted that the values listed in the tables as 'actual value' were the values used to generate the data. However, the estimated optimum parameter values, $\hat{\theta}$, may not necessarily equal the actual values due to being estimated from the generated finite length data set.

5. DISCUSSION

It can be seen from Figures 1 and 2 that the first-order algorithm provided outstanding estimation performance on the one-parameter problem. Although estimation performance for the first-order one-parameter algorithm is not

shown for the parameter β , it was found that in general β was the easier of the two parameters to estimate - both in terms of accuracy, and rate of convergence. This fact can also be observed in Figure 4, which illustrates the convergence behavior for the second-order two-parameter algorithm. In this case, the true values are $\beta=15$, and $\phi=15$. It can be seen that the curve for β is relatively smooth, while the curve for ϕ is more discontinuous. Part of this effect is due to the number of iterations run in the first-order parameter estimation section of the algorithm; i.e., the estimated value, $\hat{\beta}$, upon entry to the second-order section of the program is closer to the true value of β , than $\hat{\phi}$ is to the true value of ϕ . This disparity in convergence rates might be overcome with more experimentation with the user-adjustable 'blocking factor' which constrains the range of parameter perturbation at each iteration.

One technicality which had to be accounted for in these simulations was that Eq. 3 contained several minima; i.e., over the span of viable parameter combinations of β and ϕ , there were several combinations where the loss function, $L(\theta)$, went to zero. The resulting optimum values, $\hat{\theta}$, were thus a function of the initial starting parameters, $\hat{\theta}_0$.

In conclusion, it was found that both the first-order and second-order SPSA algorithms performed well in the prescribed application. The performance of both algorithms was highly dependent on the shape of the loss function surface. Consequently, this places a higher burden on the selection of initial parameter values and user-selectable algorithm tuning variables, such as the two convergence gain parameters, a_k and c_k .

ACKNOWLEDGMENTS

This work was performed under a contract with the Johns Hopkins Applied Physics Laboratory. The authors would like to thank Dr. James Spall and Dr. John Maryak for their comments and support.

BIBLIOGRAPHY

- Blum, J. R. (1954) "Multidimensional Stochastic Approximation Methods", *Ann. Math. Stat.*, **25**, 737-744.
- Kiefer, J. and Wolfowitz, J. (1952) "Stochastic Estimation of the Maximum of a Regression Function", *Ann. Math. Stat.*, **23**, 462-466.
- Spall, J. C. (1992) "Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation", *IEEE Transactions on Automatic Control*, **37**(3), 332-341.
- Spall, J. C. (1998a) "Implementation of the Simultaneous Perturbation Algorithm for Stochastic Optimization", *IEEE Transactions on Aerospace and Electronic Systems*, **34**(3), 817-823.
- Spall, J. C. (1998b) "An Overview of the Simultaneous Perturbation Method for Efficient Optimization", *John Hopkins APL Technical Digest*, **19**(4), 482-492.
- Spall, J. C. (1999) "Adaptive Stochastic Approximation by the Simultaneous Perturbation Method", *IEEE Transactions on Automatic Control*, to appear (in condensed form in *Proceedings of the IEEE Conference on Decision and Control*, 1998, 3872-3879).

Received June, 1999; Revised January, 2000.