Solving Deterministic Problems by
Stochastic Approximation

H. Sugiyama* and E. S. Lee

Department of Industrial Engineering, Kansas State University,
Manhattan, Kansas 66506

Received November 13, 1987

Stochastic approximation originally proposed by Robbins and Monro for stochastic problems is shown to be an effective computational tool for the numerical solution of deterministic problems. Some of the striking features of this approach are: Convergence can be proved for a general function and no differentiations or partial derivatives are needed. With some modifications of the required constant for deterministic problems, convergence can be accelerated considerably. To illustrate the approach, both algebraic equations and 2-point boundary value problems in ordinary differential equations are solved. Although these two classes of problems are quite different, they can be solved by essentially the same iterative approach. This shows the advantage and the generality of the method. Only 1-dimensional problems are investigated. Multidimensional problems and optimization problems with and without constraints will be discussed in other papers.

INTRODUCTION

Stochastic approximation was first proposed by Robbins and Monro in 1951 for locating a unique zero point of one dimensional monotone increasing regression function [1]. Various mathematical researchers have studied and applied this approach [2-12]. More recently Dr. Bellman [13] and Dr. Sugiyama [12] have emphasized the application of this approach in connection with the numerical solution of stochastic problems. However, this method has never been applied to obtain the numerical solution of deterministic problems where no noise is incorporated into the problem. The application has always been confined to stochastic problems for obtaining statistical estimates of unknown parameter values.

In this paper, we show that this technique also forms a very effective approach for obtaining the numerical solutions of problems which have no

* Dr. Sugiyama is a Professor Emeritus of Mathematical Engineering, School of Engineering, Osaka University, Japan.

Copyright © 1989 by Academic Press, Inc.
All rights of reproduction in any form reserved.

569
stochastic characteristics. Thus, it can be used as an effective numerical tool to solve iteratively various computational problems.

This approach has many advantages compared to currently available methods such as Newton's or the Newton–Raphson iteration method. Just like Newton’s method, the convergence rate of this approach is generally very fast and is quadratic under certain conditions. Unlike Newton’s method, no differentiation or partial derivative is needed. Furthermore, the ability of proving convergence for a given but unknown function gives this approach a much wider application. For example, 2-point boundary value problems in ordinary differential equations which are much more complicated in structure can also be solved by essentially the same approach. All the complications are taken care of by the conditional expression which is the realization of the random variable and is observable.

In this paper, both algebraic equations and 2-point boundary value problems in ordinary differential equations are solved by stochastic approximation. Only 1-dimensional problems are considered. It is shown that this approach converges to the solution reasonably fast. Furthermore, the convergence rate can be accelerated by Kesten's acceleration method [8].

In another paper, we shall show that Kiefer–Wolfowitz's approach [2] which uses stochastic approximation to obtain the optimum of a regression function can also be used to solve deterministic problems. Both multidimensional problems and optimization problems with and without constraints will be considered. It appears that the Kiefer–Wolfowitz approach is more powerful for multidimensional problems.

At first glance it seems strange to use stochastic methods to solve deterministic problems where very high accuracy or precision is required. This is probably the reason why this approach has never been used to solve numerical problems. As is shown in this work, this stochastic method is very efficient and also very versatile for solving deterministic problems. One explanation may lie in the fact that numerical computations are approximate with various round-off errors. It is well known that round-off errors can be considered as random variables with Gaussian distribution. This is because that computation must be carried out with a fixed and limited number of digits.

**STOCHASTIC APPROXIMATION**

Only the important features and equations which are essential for the proposed numerical computation will be summarized. The interested reader is referred to the original paper and the various research works accumulated after Robbins and Monro [1–10].
Let $Y$ be a random variable with finite variance $\sigma^2$ denoting the observed value of a regression function $M(x)$ at the level $x$. If we denote the conditional distribution function by $H(y \mid x)$, $M(x)$ is the expectation of random variable $Y$ and thus

$$M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x) \quad (1)$$

Stochastic approximation was initiated for estimating the root $\theta$ of the equation

$$M(x) = \alpha, \quad (2)$$

where $M(x)$ is assumed unknown to the experimenter. However, the random variable $Y(x)$ is assumed observable by the experimenter at the experimental level $x$. If $|M(x)|$ is bounded with probability one and if

$$M(x) \leq \alpha \quad \text{for} \quad x < \theta$$
$$\alpha \leq M(x) \quad \text{for} \quad \theta < x$$

then the unique root $\theta$ of Eq. (2) can be obtained by the sequence $\{x_n\}$, where $\{x_n\}$ is obtained by the simple iterative algorithm

$$x_{n+1} = x_n + a_n(\alpha - y(x_n)) \quad (3)$$

starting from an arbitrary value $x_1$, $y(x_n)$ is the realization of the random variable $Y(x_n)$, and the coefficient $a_n$ satisfies

$$a_n > 0, \quad (4a)$$
$$\sum_{n=1}^{\infty} a_n = \infty \quad (4b)$$

and

$$\sum_{n=1}^{\infty} a_n^2 < \infty. \quad (4c)$$

Under the above conditions, Robbins and Monro proved the mean square convergence of $\{x_n\}$ to the unique root $\theta$ with certain regularity conditions imposed upon $M(x)$, i.e.,

$$\lim_{n \to \infty} E\{(x_n - \theta)^2\} = 0 \quad (5)$$
Later, Blum [3] proved the convergence with probability one, i.e.,
\[ \Pr \left[ \lim_{n \to \infty} x_n = 0 \right] = 1 \] (6)
for the same problem under the condition
\[ |M(x)| \leq A + B |x| \] (7)
instead of
\[ \Pr[|M(x)| \leq c] = 1. \] (8)
The form of the coefficient assumed by the original investigators is
\[ a_n = c/n, \] (9)
where \( c \) is a constant.

One of the striking features of the above approach is that knowledge of the function \( M(x) \) itself is not needed as long as we can observe the outcome of the value of \( M(x) \) at a given level \( x \). This striking feature of being completely free from the form of the regression function \( M(x) \) makes the applicability of this approach very general. Furthermore, since no derivatives of \( M(x) \) are needed, this approximation scheme is much easier to use for complicated functions and can be applied to more general types of functions such as functions in tabular form.

**DETERMINISTIC COMPUTATION AND ACCELERATED CONVERGENCE**

Oviously, there are many differences between the stochastic problems originally proposed for the stochastic approximation approach and the deterministic problems to be solved in this paper. Since stochastic problems involve observation errors, high accuracy is not required. Furthermore, because of observation errors and the constant fluctuations of errors, the stochastic convergence scheme must be self-correcting. For deterministic problems, high accuracy is generally required.

Another difference lies in the unique nature of the accumulative probability density function. Due to this unique nature, the requirement of unique root and monotone increasing function for \( M(x) \) can frequently be satisfied. However, for the general deterministic problems, multiple roots are a common phenomenon.

Because of the high accuracy required for solving numerically deterministic problems, we discovered that good results cannot be obtained if we faithfully follow the original work and use Eq. (9) as the coefficient \( a_n \).
In fact, much better results can be obtained with a constant $a_n$ throughout all the iterations for a given problem.

Kesten [8] proposed a scheme to accelerate the rate of convergence of the stochastic approximation approach and gave a rigorous mathematical proof for the convergence of $\{x_n\}$ to $\theta$ with probability one based upon his modification. The acceleration scheme is based on the frequency of change of the sign of $(x_n - \theta) - (x_{n-1} - \theta) = x_n - x_{n-1}$. If the value of $|x_n - \theta|$ is small, sign fluctuation would be more frequent than the case of larger values of $|x_n - \theta|$. Thus, instead of Equation (9) for $a_n$, Kesten used the sequence

$$b_1 = a_1, \quad b_n = a_{t(n)},$$

where

$$t(n) = 1 + \sum_{i=2}^{n} l[(x_i - x_{i-1})(x_{i-1} - x_{i-2})]$$

and

$$l[u] = \begin{cases} 0 & \text{if } u > 0 \\ 1 & \text{if } u \leq 0 \end{cases}$$

Kesten also added the following condition to Eq. (4) in his proof,

$$a_{n+1} \leq a_n.$$  (13)

According to Eq. (9), the value $a_n$ changes after each iteration and becomes smaller as the number of iterations increase. But, in Kesten's approach, the value of $a_n$ decreases slowly according to Eqs. (10)–(12). Thus, instead of using Eq. (9) for $a_n$, the value of $a_n$ is kept constant as long as $x_n$ keeps monotone increasing or monotone decreasing. In fact, if we choose $a_n$ in Eq. (3) appropriately, we can preserve constant $a_n$ until we obtain a sufficiently accurate solution for $\theta$.

**ALGEBRAIC EQUATIONS**

To illustrate the approach, several deterministic algebraic equations are solved by stochastic approximation with Kesten's acceleration procedure. Table I illustrates the convergence rate for the equation

$$M(x) \equiv x^3 - 3 = 0.$$  (14)
According to Eq. (2), \( x = 0 \) for this problem. The initial value for \( x \) and the initial coefficients are

\[
x_1 = 1.0, \quad a_1 = \frac{1}{3}.
\] (15)

In the original paper, \( Y(x) \) is assumed to be a random variable and \( y(x_n) \) is the realization of the random variable \( Y(x_n) \). This realization is assumed observable. In the present deterministic case, this realization is simply

\[
y(x_n) = x_n^3 - 3.
\] (16)

The correct solution, accurate up to nine decimal places is

\[
x = 1.442249570
\]

which was obtained in twelve iterations. Notice that accuracy up to four decimal places is obtained after seven iterations. Also notice that the value of \( a_n \) was chosen according to Kesten's approach. Let

\[
a_n = c/m,
\] (17)

where \( m \) is a positive integer. Whenever the signs of \( y(x_n) \) between adjacent iterations are changed, the value of \( m \) is increased by 1. On the other hand, if the signs of \( y(x) \) between adjacent iterations do not change then \( a_n \) remains constant.

On the other hand, if the appropriate initial value for \( a_1 \) is used, then the same \( a_1 \) value can be used for all the iterations until the desired accuracy is obtained. For example, with the initial values

\[
x_1 = 1.2, \quad a_1 = 0.1,
\]
the correct solution accurate up to nine decimal places for Eq. (14) was obtained in 22 iterations. This result was obtained by using the same constant \( a_n = 0.1 \) value throughout all the iterations.

In fact, the cubic root of almost any number can be obtained by this same simple algorithm. For example, the cubic root of 3.5 was obtained as 1.518294486 in 17 iterations with

\[ x_1 = 1.4 \quad \text{and} \quad a_n = 0.1, \quad n = 1, 2, \ldots, 17. \]

The cubic root of 110 was obtained as 4.791419857 in 18 iterations with

\[ x_1 = 4.5 \quad \text{and} \quad a_n = 0.01, \quad n = 1, 2, \ldots, 18. \]

Several other equations also solved by this approach. For example, the square root of 3 or

\[ M(x) = x^2 - 3 = 0 \quad (18) \]

is obtained with nine decimal places accuracy in eleven iterations with \( x_1 = 1.0 \) and \( a_1 = \frac{1}{3}. \) If we start with \( x_1 = 2.0 \), thirteen iterations are needed to obtain the same accuracy.

Consider the equation

\[ M(x) = x^5 + x^4 - 10x^3 - 10x^2 + 9x + 9 = 0. \quad (19) \]

The five roots of this equation are

\[ x = 3, \quad 1, \quad 1, \quad 1, \quad 3; \]

with \( x_1 = 2.5 \) and \( a_1 = 1.0 \), or \( x_1 = 1.5 \) and \( a_1 = 1.0 \), the root \( x = 3 \) was obtained with four decimal places accuracy in five iterations by using Kesten's procedure.

The root of \( x = 1 \) for Eq. (19) can also be obtained quickly starting with \( x_1 = 0 \) and \( a_1 = -10. \)

**Stochastic Approach**

It is interesting to see the effectiveness of the original approach without modification for deterministic problems. Thus, the same problem represented by Eq. (14) was also solved by using the original Robbins–Monro approach. In other words, the \( a_n \) in Eq. (3) is represented by Eq. (9) and Gaussian noise is added to the observed value \( y(x_n) \). The mean of the Gaussian noise is zero and the variance is \( 1/n \). The other values used are

\[ x_1 = 1.0, \quad c = 0.1. \]
The convergence rate is extremely slow under these conditions. After 10,000 iterations, only two digit accuracy was obtained, or

\[ x_{10,000} = 1.4. \]

With the same Gaussian noise and other values but with Kesten's acceleration method for \(a_n\), we obtained

\[
\begin{align*}
    x_{3813} &= 1.442253614, & a_{3813} &= 0.000048309 \\
    x_{5034} &= 1.442251985, & a_{5034} &= 0.000035958 \\
    x_{6043} &= 1.442251598, & a_{6043} &= 0.000029976
\end{align*}
\]

where the subscripts to \(x\) and \(a\) are the number of iterations. Thus, the addition of Gaussian noise and the use of Eq. (9) for \(a_n\) makes the approach very ineffective for deterministic problems.

**TWO-POINT BOUNDARY VALUE PROBLEM**

Consider the 2-point boundary value problem in ordinary differential equations

\[ y'' = f(y', y, x) \quad (20a) \]

with either boundary conditions

\[ y(x = 0) = c \quad \text{and} \quad y(x = 1) = b \quad (20b) \]

or the boundary conditions

\[ y(x = 0) = c \quad \text{and} \quad y'(x = 1) = d. \quad (20c) \]

Since not all the boundary conditions are given at one point, this problem cannot be solved directly. Trial-and-error or shooting methods are frequently used. The generalized Newton–Raphson method has also been used [14, 15].

We shall show that the stochastic approximation approach forms a versatile and powerful technique to solve this problem. The reason this approach can be used is because the fact that the function \(M(x)\) can be any form and the observed results are given in conditional form. Let

\[ \theta = y'(0) \quad (21) \]
or $\theta$ represents the missing initial condition. Then for the problem represented by Eqs. (20a) and (20b), Eq. (3) becomes

$$\theta_{n+1} = \theta_n + a_n (b - y(1 | \theta_n)).$$  \hspace{1cm} (22)

where $y(x_n)$ in Eq. (3) is replaced by the conditional expression $y(1 | \theta_n)$. Notice that the value of $y$ at $x = 1$ is $b$. Similarly, for the problem represented by Eqs. (20a) and (20c), we have

$$\theta_{n+1} = \theta_n + a_n (d - y'(1 | \theta_n)).$$  \hspace{1cm} (23)

In order to obtain $y(1 | \theta_n)$ or $y'(a | \theta_n)$, Eq. (20a) must be integrated with initial conditions

$$y(0) = c \quad \text{and} \quad y'(0) = \theta_n.$$  \hspace{1cm} (24)

The conditional expression

$$y(1 | \theta_n) = \text{the value of } y \text{ at } x = 1 \text{ given that } y'(0) = \theta_n.$$  \hspace{1cm} (25)

In this work, the fourth-order Runge–Kutta technique is used to obtain the integration. To illustrate the approach, consider the problem

$$y'' = -0.49 (y')^2 - 1$$  \hspace{1cm} (26a)

with boundary conditions

$$y(0) = 0 \quad \text{and} \quad y(1) = 0$$  \hspace{1cm} (26b)

This example has been solved by quasilinearization [14].

This problem can be solved analytically as

$$y(x) = \frac{1}{0.49} \ln \left[ \frac{\cos(0.7)(x - 1/2)}{\cos(0.7/2)} \right].$$  \hspace{1cm} (27)

Thus, the missing initial condition for the problem represented by Eq. (26) can be calculated from Eq. (27) with $x = 0$ and the exact value is

$$y'(0) = 0.521469278.$$  \hspace{1cm} (28)

Equation (26) is solved by our stochastic approximation approach. The results are shown in Table II, where $h$ represents the integration step size. The value of $y(1 | \theta_n)$ was obtained by the fourth-order Runge–Kutta method. Compared to the value in Eq. (28), seven digits accuracy is obtained in six iterations.
TABLE II

\[ y'' = -0.49(y')^2 - 1 \]

| \( n \) | \( \theta_n \) | \( y(1 | \theta_n) \) | \( a_n \) | \( h \) |
|-------|--------|----------------|--------|--------|
| 1     | 0.2    | -0.32          | 1      | 0.2    |
| 2     | 0.52   | -0.0014        | 1      | 0.2    |
| 3     | 0.5214 | -0.000064      | 1      | 0.1    |
| 4     | 0.521464 | -0.000005      | 1      | 0.1    |
| 5     | 0.521469 | -0.000000253   | 1      | 0.1    |
| 6     | 0.521469253 |                 |        |        |

Notice that because the sign of \( y(1 | \theta_n) \) never changed in Table II, the value of \( a_n \) remained constant throughout the iteration.

Of course, better initial approximation frequently can be obtained by some approximation methods. Consider the problem

\[ y'' + y + 1 = 0 \]  \hspace{1cm} (29a)

with boundary conditions.

\[ y(0) = 0 \quad \text{and} \quad y(1) = 0. \]  \hspace{1cm} (29b)

An initial approximation of the missing initial condition can be obtained by using difference equations with \( h = 0.25 \). Solving the simultaneous difference equations, we obtain

\[ y'(0) = 0.5460. \]

Using this value as the initial approximation, the results are tabulated in Table III. Seven digits accuracy is obtained in six iterations.

Another differential equation considered is of the form

\[ y'' + \left( 1 - \frac{2}{x^2} \right) y = 0 \]  \hspace{1cm} (30a)

TABLE III

\[ y'' + y + 1 = 0 \text{ with } h = 0.1 \]

| \( n \) | \( \theta_n \) | \( y(1 | \theta_n) \) | \( a_n \) |
|-------|--------|----------------|--------|
| 1     | 0.5460 | 0.000255       | 1      |
| 2     | 0.546255 | 0.000040     | 1      |
| 3     | 0.546295 | 0.0000064   | 1      |
| 4     | 0.5463014 | 0.0000010  | 1      |
| 5     | 0.5463024 | 0.0000002  | 1      |
| 6     | 0.5463026 | 5.91 \times 10^{-9} | 1      |
SOLVING DETERMINISTIC PROBLEMS

TABLE IV

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\theta_n$</th>
<th>$y(1.5 \mid \theta_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.0452</td>
</tr>
<tr>
<td>2</td>
<td>0.9548</td>
<td>1.0224</td>
</tr>
<tr>
<td>3</td>
<td>0.9324</td>
<td>1.0110</td>
</tr>
<tr>
<td>4</td>
<td>0.9214</td>
<td>1.0054</td>
</tr>
<tr>
<td>5</td>
<td>0.9160</td>
<td>1.0027</td>
</tr>
<tr>
<td>6</td>
<td>0.9133</td>
<td>1.001334</td>
</tr>
<tr>
<td>7</td>
<td>0.911966</td>
<td>1.000658</td>
</tr>
<tr>
<td>8</td>
<td>0.911308</td>
<td>1.000325</td>
</tr>
<tr>
<td>9</td>
<td>0.910983</td>
<td>1.000160</td>
</tr>
<tr>
<td>10</td>
<td>0.910823</td>
<td>1.000079</td>
</tr>
<tr>
<td>11</td>
<td>0.910744</td>
<td>1.000039</td>
</tr>
<tr>
<td>12</td>
<td>0.910705</td>
<td>1.000019</td>
</tr>
<tr>
<td>13</td>
<td>0.910686</td>
<td>1.000010</td>
</tr>
<tr>
<td>14</td>
<td>0.910676</td>
<td>1.000005</td>
</tr>
<tr>
<td>15</td>
<td>0.910671</td>
<td>1.000002</td>
</tr>
<tr>
<td>16</td>
<td>0.910669</td>
<td>1.000001</td>
</tr>
<tr>
<td>17</td>
<td>0.910668</td>
<td>1.000001</td>
</tr>
<tr>
<td>18</td>
<td>0.910666</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

with boundary conditions

\[ y(1) = 0.5061 \quad \text{and} \quad y(1,.50) = 1.0. \]  \hspace{1cm} (30b)

Equation (22) for this problem becomes

\[ \theta_{n+1} = \theta_n + a_n(1 - y(1.50 \mid \theta_n)). \]  \hspace{1cm} (31)

In order to test the convergence rate without Kesten's acceleration, this problem is solved with a constant $a_n \equiv 1.0$. The results are listed in Table IV. It needs 18 iterations in order to obtain six digits accuracy. With Kesten's modification, 9 digits accuracy can be obtained with 16 iterations. The missing initial condition obtained is

\[ y^1(1.0) = 0.910666. \]

CONCLUSION

The numerical examples show clearly the effectiveness of this modified approach for deterministic problems. Because of the very general nature of the function $M(x)$ and also because the use of the conditional expression for the observable realization $y(x_n)$, this approach is very versatile and can be applied to solve many different types of problems.
It should be noted that our proposed procedure is much more general than the original approach. Robbins and Monro were concerned with stochastic problems and thus their proof is restricted to a monotone increasing function with a single root. Our procedure is applicable for problems with multiple roots.

In the experiment to solve Eq. (14), the addition of observation noise only slows down the convergence rate considerably. It should be emphasized that even without the addition of Gaussian noise, there is still a noise but with a very small variance. This noise comes from the truncation error.

REFERENCES