JOURNAL OF APPROXIMATION THEORY 56, 118–120 (1989)

Note

Necessary and Sufficient Conditions for a Stochastic Approximation Method

DEAN S. CLARK

Department of Mathematics, University of Rhode Island, Kingston, Rhode Island 02881, U.S.A.

Communicated by Oved Shisha

Received December 24, 1986; revised June 9, 1987

The simplest interpretation of the stochastic approximation (SA) problem is to estimate a zero θ of an unknown function $f\colon \mathbb{R} \to \mathbb{R}$ via a sequence of iterates X_n which, rather than providing exact values $f(X_n)$, give only "noise corrupted" observations $f(X_n) + \xi_n$, where ξ_n denotes the random observation error. If f is thought to have enough monotonicity, say, were the graph of f to lie above that of $y = -\rho(x-\theta)$ for $x < \theta$ and below it for $x > \theta$, for some positive constant ρ, then

$$X_{n+1} = X_n + a_n(f(X_n) + \xi_n), \qquad a_n > 0, \tag{1}$$

supplies such a sequence (X_n). Equation (1) is the original recursive SA method: the Robbins–Monro method [4].

In [1] we gave necessary and sufficient conditions for the convergence of X_n to θ with probability 1 (wp 1) that were in the form of laws of large numbers

$$a_n \cdot \sum_{j=0}^{n-1} \xi_j \to 0 \qquad \text{as} \quad n \to \infty, \quad \text{wp } 1. \tag{2}$$

It turned out that the rate of decrease of the step-sizes a_n was critical in determining whether (2) was a necessary or a sufficient condition for $X_n \to \theta$. If a_n decreased at least as rapidly (slowly) as c/n, $c > 0$, then (2) was necessary (sufficient) for convergence.

In [1], as in almost all the SA literature, it was assumed that $a_n \to 0$ as $n \to \infty$, thus enabling the convergence $X_n \to \theta$. In this note we ask if this condition $a_n \to 0$ is strictly necessary for the approximation of θ in *some* useful probabilistic sense as $n \to \infty$ and answer that it is not.

0021-9045/89 $3.00

Consider the multidimensional version of (1),

$$X_{n+1} = X_n + E_n(F(X_n) + \xi_n), \qquad F\colon \mathbb{R}^m \to \mathbb{R}^m, \qquad F(0) = 0, \tag{3}$$

where (E_n) is a sequence of diagonal matrices with positive entries ε_n^j, and (ξ_n) is a sequence of m-dimensional random vectors. We assume that none of the ε_n^j go to zero as $n \to \infty$, ruling out the reasonable possibility of convergence of (X_n) wp 1. Nevertheles, if the ε_n^j become small, there is a reasonable possibility of the asymptotic approximation of $\theta = 0 \in \mathbb{R}^m$ in the mean-square (L_2) sense, hence in probability. To tie the ε_n^j to a small known parameter $\varepsilon > 0$ we assume that

$$|\varepsilon_n^j - \varepsilon| = O(\varepsilon^2) \qquad \text{for all } j, n, \tag{4}$$

where $O(h)$ denotes a numerical value satisfying $O(h) \leqslant M\,|h|$ for a constant $M > 0$ independent of h. Let $\|\xi\| = \sqrt{E(|\xi|^2}$, with E the usual expectation operator, denote the L_2 norm of a random vector ξ. We emphasize in the remarks that follow that this is not to be confused with the *Euclidean* norm of a random vector at a fixed sample point, denoted $\sqrt{|\xi(\omega)|^2}$.

THEOREM. *Let* (X_n) *be defined by* (3), *assuming* (4), *and suppose that* F *is Lipschitz continous on* $\mathbb{R}^m$. *If* $\lim_{\varepsilon \downarrow 0} \overline{\lim}_{n \to \infty} \|X_n\| = 0$, *then* (6), *below, holds. Conversely, if we also have*

$$\langle F(X), X \rangle \leqslant -\rho\,|X|^2, \qquad X \in \mathbb{R}^m, \tag{5}$$

for some positive constant ρ, *then* (6) *implies* $\lim_{\varepsilon \downarrow 0} \overline{\lim}_{n \to \infty} \|X_n\| = 0$.

$$\sup_{n \geqslant 0} \|\xi_n\| < \infty, \tag{6a}$$

$$\lim_{N \to \infty} \left\{ \sup_{n \geqslant 0} \frac{1}{N+1} \left\| \sum_{j=n}^{n+N} \xi_j \right\| \right\} = 0. \tag{6b}$$

Remarks. Condition (5), where $\langle \cdot, \cdot \rangle$ and $|\cdot|$ denote the Euclidean inner product and distance, respectively, assures that (X_n) will approximate θ in the absence of the random errors ξ_n as ε becomes small. It amounts to the existence of a Liapunov function (the classical one $V(X) = \frac{1}{2}|X|^2$) for the differential equation $\dot{X} = F(X)$, and it seems clear that a weaker condition, say, simply the existence of any suitable Liapunov function, would also suffice.

Condition (6b) would be very unnatural if $\|\xi\|$ were replaced by $|\xi(\omega)|$ and (6b) were required to hold in this latter sense wp 1. With this interpretation, (6b) holds wp 0 even for the sequence of classical coin-tossing random variables (ψ_n) despite the fact that, for any *fixed* n, $\lim_{N \to \infty} 1/(N+1) |\sum_{j=n}^{n+N} \psi_j(\omega)| = 0$ at almost every sample point ω (the strong law of large numbers).

However, (6) is a statement about the boundedness and ergodic behavior of (ξ_n) in the L_2-sense. In that case, (6) becomes quite a bit more reasonable. For instance, boundedness and orthogonality of (ξ_n) in L_2 is more than sufficient for (6). The kinds of stochastic processes that could satisfy (6) have been the subject of much study. See, e.g. [3].

The conditions on F are the same as those used in [1], and the proof of the theorem, although technically more complicated, follows along the lines of that given in [1]. See [2].

References

1. D. S. Clark, Necessary and sufficient conditions for the Robbins–Monro method, *Stochastic Process. Appl.* **17** (1984), 359–367.
2. D. S. Clark, Stability in a system of difference equations, *Quart. Appl. Math.* **45** (1987), 441–453.
3. I. A. Ibragimov and Yu. V. Linnik, "Independent and Stationary Sequences of Random Variables," Wolters-Noordhoff, Groningen, 1971.
4. H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Stat.* **22** (1951), 400–407.

Printed in Belgium