Stochastic Processes and their Applications 27 (1988) 217-231
North-Holland

CONVERGENCE AND ROBUSTNESS OF THE ROBBINS-MONRO ALGORITHM TRUNCATED AT RANDOMLY VARYING BOUNDS

Han-Fu CHEN, Lei GUO and Ai-Jun GAO

Institute of Systems Science, Academia Sinica, Beijing, People's Republic of China

Received 24 April 1987
Revised 14 September 1987

In this paper the Robbins-Monro (RM) algorithm with step-size $a_n = 1/n$ and truncated at randomly varying bounds is considered under mild conditions imposed on the regression function. It is proved that for its a.s. convergence to the zero of a regression function the necessary and sufficient condition is

$$(1/n) \sum_{i=1}^{n} \xi_i \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.}$$

where ξ_i denotes the measurement error. It is also shown that the algorithm is robust with respect to the measurement error in the sense that the estimation error for the sought-for zero is bounded by a function $g(\varepsilon)$ such that

$$g(\varepsilon) \xrightarrow[\varepsilon\to 0]{} 0 \quad \text{if} \quad \limsup_{n\to\infty} (1/n) \left\| \sum_{i=1}^{n} \xi_i \right\| \equiv \varepsilon > 0.$$

stochastic approximation * randomly varying truncation * robustness to noise * necessary and sufficient condition for convergence

1. Introduction

Since the pioneer work [1] of Robbins and Monro, stochastic approximation has drawn much attention from those interested in both its theory and applications. The earlier work [2] concentrated mainly on the case with independent measurement error. Later, the effort was devoted to weakening conditions imposed on the regression function and on the measurement noise [3-7]. By using the ordinary differential equation (ODE) method proposed by Ljung [3, 4] and further developed in [5], a large class of measurement errors can be treated in the convergence analysis. The a priori boundedness of the algorithm in this method is no longer assumed in [7] where the ODE method combined with martingale theory is used to analyse the case when the measurement error is of an ARMA process.

This project was supported by the National Natural Science Foundation of China and by TWAS Research Grant No. 87-43.

In a recent paper [8] Chen and Zhu consider a new algorithm consisting of truncating the RM algorithm at randomly varying bounds, and have proved the convergence of the algorithm to the zero set of the regression function under weak conditions imposed on both the regression function and measurement errors. Later, such algorithms are applied to the optimization problem in [9], to an optimization problem with constraints in [10], and to simultaneous estimations of zeros of the regression function and of unknown parameters contained in the measurement errors in [11, 12].

For the case when the regression function is dominated by a linear function and the Liapunov function is a quadratic function, Clark [13] has shown that the necessary and sufficient condition for convergence of the RM algorithm to the zero of the regression function is

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i \xrightarrow[n\to\infty]{} 0,$$

where ξ_i denotes the measurement error. In this paper, by use of an elementary approach completely different from that used in [13], we prove that the above mentioned conclusion remains valid for the RM algorithm truncated at randomly varying bounds in a more general case, namely, the restrictive conditions imposed (in [13]) on the regression function and Liapunov function can be removed. It is also shown that the algorithm considered in this paper is robust with respect to the measurement error in the sense that the estimation error for the zero of the regression function is bounded by an increasing and left-continuous function $g(\varepsilon)$ with

$$g(\varepsilon) \xrightarrow[\varepsilon\to 0]{} 0 \quad \text{if} \quad \limsup_{n\to\infty}(1/n)\left\|\sum_{i=1}^{n} \xi_i\right\| \equiv \varepsilon > 0.$$

We use $\|x\|$ to denote the Euclidean norm of a vector $x \in R^m$.

2. Main results

Let the regression function $h(\cdot): R^m \to R^m$ be an unknown continuous function with zero x^0:

$$h(x^0) = 0, \quad x^0 \in R^m, \tag{1}$$

and let x_i be the i-th approximation to x^0 based on the measurements $\{y_j : 0 \leqslant j \leqslant 1\}$. At time $i+1$ the regression function $h(x)$ is observed at x_i with random error ξ_{i+1}:

$$y_{i+1} = h(x_i) + \xi_{i+1}, \quad i \geqslant 0. \tag{2}$$

The assumptions made on $h(x)$ and ξ_i are as follows:

A. $\operatorname*{Lim\,sup}_{n\to\infty}(1/n)\left\|\sum_{i=1}^{n} \xi_{i+1}\right\| \equiv \varepsilon < \infty$ a.s. $\quad (3)$

B. There is a twice continuously differentiable function $v(x): R^m \to R$, such that

$$\begin{aligned} &v(x^0)=0, \qquad \lim_{\|x\|\to\infty} v(x)=\infty, \\ &v(x)>0, \qquad h^\tau(x)v(x)<0, \quad \forall x \neq x^0, \end{aligned} \tag{4}$$

where $v_x(x)$ denotes the gradient of $v(x)$.

Remark 1. Conditions A and B are the only restrictions on the measurement errors and on the regression function and are possibly the weakest in comparison with those used in preceding results. Condition A prescribes the boundedness of $\{\xi_i\}$ in average and it is satisfied by a large class of random vectors, for example, if $\{\xi_i\}$ is a stationary ARMA process with zero mean [2, 7], and also if $\sum_{i=1}^{\infty} \xi_i/i$ converges [5, 8]. Condition B assumes the existence of a Liapunov function $v(x)$ but does not require $v(x)$ to be known. This condition implies asymptotic stability of the solution x^0 for the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = h(x(t)).$$

These kinds of conditions are frequently used in the convergence analysis of algorithms [5, 7, 8–13].

For describing our algorithm we first take two arbitrary points $x_1^* \neq x_2^*$ in R^m and define $d_0 = \max(v(x_1^*), v(x_2^*))$.

Since $\lim_{\|x\|\to\infty} v(x)=\infty$, there is an $M>0$ such that

$$d_0 < \inf\{v(x): \|x\| > M\} \equiv d \tag{5}$$

$$\max(\|x_1^*\|, \|x_2^*\|) < M. \tag{6}$$

Assuming that M determined by (5)–(6) is now fixed, we apply the RM algorithm truncated at randomly varying bounds [8] to estimate x^0 as follows. Let $\{M_i\}$ be an arbitrary increasing sequence of positive numbers tending to infinity, with $M_0 > M+8$.

Define, for $n = 1, 2, \ldots,$

$$x_{n+1} = (x_n + (1/n)y_{n+1}) I_{[\|x_n+(1/n)y_{n+1}\| \le M_{\sigma(n)}]} + x_n^* I_{[\|x_n+(1/n)y_{n+1}\| > M_{\sigma(n)}]}, \tag{7}$$

$$\sigma(n) = \sum_{i=1}^{n-1} I_{[\|x_i+(1/i)y_{i+1}\| > M_{\sigma(i)}]}, \qquad \sigma(1)=0, \tag{8}$$

$$x_n^* = \begin{cases} x_1^* & \text{if } \sigma(n) \text{ is even,} \\ x_2^* & \text{if } \sigma(n) \text{ is odd.} \end{cases} \tag{9}$$

Since the truncation bound is time-varying and allowed to increase to infinity, there is no a priori boundedness assumption imposed on $\{x_n\}$.

Theorem 1. *If Conditions* A *and* B *hold for the algorithm defined by* (7)-(9), *then there is an* $\varepsilon^* > 0$ *such that, for any* $\varepsilon \in [0, \varepsilon^*)$,

$$\limsup_{n\to\infty} \|x_n - x^0\| \leq g(\varepsilon) \quad a.s.$$

where ε *is given in* (3) *and* $g(\varepsilon)$ *is an increasing, left-continuous function on* $[0, \infty)$ *tending to zero as* $\varepsilon \to 0$.

Theorem 2. *Under Condition* B, $\{x_n\}$ *defined by* (7)-(9) *converges to* x^0 *if and only if Condition* A *holds with* $\varepsilon = 0$.

Remark 2. Theorem 1 means that the less the influence of the noise is (or the smaller ε is), the better x_n approximates x^0 in the limit. In other words, the algorithm is robust to noise. Unlike [13] Theorem 2 requires neither growth rate for $h(\cdot)$ and $v(\cdot)$, nor [5] a priori boundedness of the iterates x_n.

3. Finiteness of the number of truncations

In the sequel we always assume that the sampling point ω is fixed.

For the convenience of the reader we collect in one place, (10)-(20), below, our basic parameter definitions and their permissible range of values.

Let δ_1, δ_2 be real numbers such that

$$\delta_2 - \delta_1 = (d - d_0)/2, \qquad [\delta_1, \delta_2] \subset (d_0, d) \tag{10}$$

and set

$$D \equiv \{x\colon \delta_1 \leq v(x) \leq \delta_2\} \cap \{x\colon \|x\| \leq M\}$$

where d_0, d and M are given in (5)-(6).

Define the quantities

$$h_1 \equiv \max_{\|x\| < M+8} \|h(x)\|, \tag{11}$$

$$r_1 \equiv \max_{x\in U} \|v_x(x)\|, \qquad r_2 \equiv \max_{x\in U} \|v_{xx}(x)\|, \tag{12}$$

$$a \equiv -\max_{x\in D} h^{\tau}(x) v_x(x), \tag{13}$$

where

$$U \equiv \{x\colon \|x\| \leq M + 5 + h_1\}, \tag{14}$$

and $v_{xx}(x)$ denotes the matrix of second partial derivatives of $v(x)$. $\|X\|$ denotes the square root of the maximum eigenvalue of XX^{τ} for any matrix X.

It is clear $x^0 \notin D$ and $a > 0$ by Condition B.

Since $h(\cdot)$ and $v(\cdot)$ are continuous, there are constants $\varepsilon^* \in (0, 1]$, $T^* \in (0, 1]$, $\alpha_1 > 0$, $\beta > 0$, and $\delta^* > 0$ satisfying

$$\text{(i)} \quad T^* < 1/(1+h_1), \qquad \tfrac{1}{2} r_2 (h_1+2)^2 T^* - a < -\alpha_1, \tag{15}$$

$$\text{(ii)} \quad r_1 \cdot \max_{\substack{\|x-y\| < 3\varepsilon^* + T^*(h_1+2) \\ x,y \in U}} \|h(x) - h(y)\| - \alpha_1 < -\beta, \tag{16}$$

$$\text{(iii)} \quad \varepsilon^* < \beta T^* / [8r_1 + r_2(3h_1 + 11)], \tag{17}$$

$$\text{(iv)} \quad \max_{\substack{\|x-y\| < 4\varepsilon^* \\ x,y \in U}} |v(x) - v(y)| < \delta_2 - \delta_1, \tag{18}$$

$$\text{(v)} \quad \delta_1 + [7\varepsilon^* + T^*(h_1 + 2\varepsilon^*)] r_1 < \delta_2, \tag{19}$$

$$\text{(vi)} \quad \min_{\substack{\|x-x^0\| \geq \delta^* \\ \|x\| \leq M_\sigma}} (-h^\tau(x) v_x(x)) \geq \sqrt{\varepsilon^*} \cdot \max_{\|x\| \leq M\sigma + 2\varepsilon^*} v(x), \tag{20}$$

where σ is defined in Theorem 3, below.

By Condition A we can take an integer N sufficiently large so that

$$N > 2h_1/\varepsilon^*, \qquad \frac{1}{n} \left\| \sum_{i=1}^{n} \xi_{i+1} \right\| \leq \tfrac{11}{8} \varepsilon^*, \quad \forall n \geq N \tag{21}$$

if $0 < \varepsilon < \varepsilon^*$. The quantity $\varepsilon^* \in (0, 1]$ satisfying (17)-(20) is the one used in Theorem 1.

In this section we prove that the number of truncations $\sigma(n)$ is bounded as $n \to \infty$ (Theorem 3). For this we first prove two lemmas showing that x_m is close to x_n when $\|x_n\| \leq M$ and $\sum_{i=n}^{m} 1/i$ is small. These lemmas use only Condition A and the boundedness of $h(\cdot)$ on a bounded set.

Lemma 1. *Let h_1 be defined by* (11) *and $\Delta \equiv 1/(1+h_1)$. Suppose that Condition* A *holds for some $\varepsilon < \varepsilon^* \leq 1$ and that $\|x_n\| \leq M$ for some $n > N$, where M and N are defined in* (5)-(6) *and* (21), *respectively. Then, for any $T \in [0, \Delta]$,*

$$\left\| \sum_{i=n}^{m+1} \frac{1}{i} y_{i+1} \right\| \leq 6\varepsilon^* + 2, \quad \forall m\colon\ n \leq m \leq m(n, T),$$

where

$$m(n, T) = \max\left\{ m\colon \sum_{i=n}^{m} \frac{1}{i} \leq T \right\}. \tag{22}$$

Proof. If the set of integers defined by (22) is empty for all $T \in [0, \Delta]$ there is nothing to prove. Otherwise, since $m(n, T)$ is non-decreasing as T increases, we only need to show the lemma for $T = \Delta$.

Suppose that the lemma were not true. Then we would find m_1 such that $n \leq m_1 \leq m(n, T)$ and

$$\left\| \sum_{i=n}^{m_1+1} \frac{1}{i} y_{i+1} \right\| > 6\varepsilon^* + 2.$$

Without loss of generality let

$$m_1 = \min\left\{m:\ n \leqslant m \leqslant m(n, \Delta);\ \left\|\sum_{i=n}^{m+1} \frac{1}{i} y_{i+1}\right\| > 6\varepsilon^* + 2\right\}.$$

Hence, for any m: $n \leqslant m \leqslant m_1$,

$$\left\|x_n + \sum_{i=n}^{m} \frac{1}{i} y_{i+1}\right\| \leqslant \|x_n\| + \left\|\sum_{i=n}^{m} \frac{1}{i} y_{i+1}\right\|$$

$$\leqslant M + 6\varepsilon^* + 2 < M_0 < M_{\sigma(n)}, \tag{23}$$

and by (7) we obtain the untruncated iterations

$$x_{m+1} = x_m + \frac{1}{m} y_{m+1}, \quad \forall m:\ n \leqslant m \leqslant m_1. \tag{24}$$

Hence

$$\|x_{m_1+2} - x_n\| = \left\|x_{m_1+1} - x_n + \frac{1}{m_1+1} y_{m_1+2}\right\|$$

$$= \left\|\sum_{i=n}^{m_1+1} \frac{1}{i} y_{i+1}\right\| > 6\varepsilon^* + 2 \tag{25}$$

where the last inequality follows from the definition of m_1.

We now show the inequality converse to (25), and the contradiction will prove the lemma.

From (23), (24) we have

$$\|x_m\| \leqslant M + 6\varepsilon^* + 2, \quad \forall m:\ n \leqslant m \leqslant m_1 + 1$$

and by (11)

$$\|h(x_m)\| \leqslant h_1, \quad \forall m:\ n \leqslant m \leqslant m_1 + 1.$$

Let $S_i \equiv \sum_{j=1}^{i} \xi_{j+1}$. Using (21), (22) and noting $n > N$, we see that

$$\|x_{m_1+1} - x_n\| \leqslant \left\|\sum_{i=n}^{m_1} \frac{1}{i} h(x_i)\right\| + \left\|\sum_{i=n}^{m_1} \frac{1}{i} \xi_{i+1}\right\|$$

$$\leqslant \left(\sum_{i=n}^{m_1} \frac{1}{i}\right) h_1 + \left\|\frac{S_{m_1}}{m_1} - \frac{S_{n-1}}{n} + \sum_{i=n}^{m_1-1} \frac{S_i}{i} \frac{1}{i+1}\right\|$$

$$\leqslant h_1 \Delta + \tfrac{11}{4}\varepsilon^* + \tfrac{11}{8}\varepsilon^* \Delta.$$

Further, by (21) and the selection of ε^*,

$$m_1 > N > 2h_1/\varepsilon^*$$

and we have

$$\begin{aligned}
&\left\| x_{m_1+1} - x_n + \frac{1}{m_1+1} y_{m_1+2} \right\| \\
&\quad \leqslant \|x_{m_1+1} - x_n\| + \left\| \frac{1}{m_1+1} (h(x_{m_1+1}) + \xi_{m_1+2}) \right\| \\
&\quad \leqslant h_1 \Delta + \tfrac{11}{4} \varepsilon^* + \tfrac{11}{8} \varepsilon^* \Delta + \frac{h_1}{m_1+1} + \frac{1}{m_1+1} \|S_{m_1+1} - S_{m_1}\| \\
&\quad \leqslant (h_1 + \tfrac{11}{8} \varepsilon^*) \Delta + \tfrac{11}{4} \varepsilon^* + \tfrac{1}{2} \varepsilon^* + \tfrac{11}{4} \varepsilon^* \\
&\quad \leqslant (h_1 + \tfrac{11}{8} \varepsilon^*)/(h_1+1) + 6\varepsilon^* < 2 + 6\varepsilon^*
\end{aligned}$$

which contradicts (25).

Lemma 2. *Under the conditions and definitions of Lemma* 1,

$$\|x_m - x_n\| < 3\varepsilon^* + T(h_1 + 2\varepsilon^*)$$

for any $T \in [0, \Delta]$ *and any* m: $n \leqslant m \leqslant m(n, T) + 1$.

Proof. It is easy to see that

$$\begin{aligned}
\left\| x_n + \sum_{i=n}^{m} \frac{1}{i} y_{i+1} \right\| &\leqslant M + 6\varepsilon^* + 2 \\
&\leqslant M_0 \leqslant M_{\sigma(n)}, \quad \forall m: \ n \leqslant m \leqslant m(n, T) + 1
\end{aligned}$$

because of Lemma 1 and $\|x_n\| \leqslant M$.

Then by the definition of our algorithm we know that

$$x_{m+1} = x_m + \frac{1}{m} y_{m+1}, \quad \forall m: \ n \leqslant m \leqslant m(n, T) + 1 \tag{26}$$

and

$$\|x_m\| \leqslant M + 6\varepsilon^* + 2, \qquad \|h(x_m)\| \leqslant h_1.$$

Hence we have

$$\begin{aligned}
\|x_m - x_n\| &\leqslant \left\| \sum_{i=n}^{m-1} \frac{1}{i} (h(x_i) + \xi_{i+1}) \right\| \\
&\leqslant h_1 T + \left\| \frac{S_{m-1}}{m-1} - \frac{S_{n-1}}{n} + \sum_{i=n}^{m-2} \frac{S_i}{i} \frac{1}{i+1} \right\| \\
&\leqslant h_1 T + \tfrac{11}{4} \varepsilon^* + \tfrac{11}{8} \varepsilon^* T < 3\varepsilon^* + T(h_1 + 2\varepsilon^*)
\end{aligned}$$

for any m: $n \leqslant m \leqslant m(n, T) + 1$, and the lemma follows.

We now show that the number of truncations for the algorithm is finite. To emphasize the dependence on ε we write the $\sigma(n)$ of (8) as $\sigma_\varepsilon(n)$.

Theorem 3. *Under the conditions of Theorem* 1 *there is a constant σ independent of ε such that*

$$\sup_n \sigma_\varepsilon(n) \leqslant \sigma < \infty$$

for any $\varepsilon \in [0, \varepsilon^*)$.

Proof. Suppose the claim were not true. Let σ_0 be an integer greater than both N and $1/T^*$.

By the contradictory assumption, there would exist $\varepsilon \in [0, \varepsilon^*)$ and n such that

$$\sigma_\varepsilon(n) > \sigma_1 \quad \text{with } \sigma_1 = \sigma_0 + 2. \tag{27}$$

Let n_0 be the maximal time for which

$$\sigma_0 = \sum_{i=1}^{n_0-1} I\left[\left\| x_i + \frac{1}{i} y_{i+1} \right\| > M_{\sigma_\varepsilon(i)}\right]. \tag{28}$$

Then by (7), (8) we have

$$\sigma_\varepsilon(n_0) = \sigma_0, \qquad \sigma_\varepsilon(n_0+1) = \sigma_0 + 1 \quad \text{(because of the maximality)},$$
$$x_{n_0+1} = x^*_{n_0} \tag{29}$$

and

$$n_0 > \sigma_0 > \max(N, 1/T^*). \tag{30}$$

For an initial contradiction, suppose that $\|x_n\| \leqslant M$ for all $n \geqslant n_0 + 2$. Then by (21)

$$\begin{aligned} \left\| x_n + \frac{y_{n+1}}{n} \right\| &\leqslant M + \left\| \frac{h(x_n)}{n} \right\| + \left\| \frac{\xi_{n+1}}{n} \right\| \\ &\leqslant M + \frac{h_1}{n} + \left\| \frac{\sum_{i=1}^{n} \xi_{i+1}}{n} - \frac{\sum_{i=1}^{n-1} \xi_{i+1}}{n} \right\| \\ &\leqslant M + \frac{\varepsilon^*}{2} + 3\varepsilon^* \\ &\leqslant M + 4 < M_0 \leqslant M_{\sigma_\varepsilon(n)}, \quad \forall n \geqslant n_0 + 2, \end{aligned} \tag{31}$$

which in conjunction with (8) and (29) shows

$$\sigma_\varepsilon(n) \equiv \sigma_\varepsilon(n_0+2) \leqslant 1 + \sigma_\varepsilon(n_0+1) = \sigma_0 + 2$$

for all $n \geqslant n_0 + 2$, and this contradicts (27).

Hence there is an $m_0 \geqslant n_0 + 2$ such that

$$\|x_{m_0}\| > M. \tag{32}$$

By (29) and (32) from (5) and (10) we see that

$$v(x_{n_0+1}) = v(x^*_{n_0}) < \delta_1, \tag{33}$$

$$v(x_{m_0}) > \delta_2, \quad (m_0 \geqslant n_0 + 2). \tag{34}$$

It is important to know that $m_0 > n_0+2$, i.e., strict inequality. For this we only need to show that $v(x_{n-1}) < \delta_1$ for some $n \geq n_0+2$ leads to $v(x_n) < \delta_2$.

Indeed, if $v(x_{n-1}) < \delta_1$, by (5), (10) we have $\|x_{n-1}\| \leq M$, and by the argument of (31)

$$\left\| x_{n-1} + \frac{1}{n-1}(h(x_{n-1}) + \xi_n) \right\| \leq M + 4\varepsilon^* \leq M_{\sigma_\varepsilon(n-1)}.$$

Hence there is no truncation at time n and thus

$$x_n = x_{n-1} + \frac{1}{n-1}(h(x_{n-1}) + \xi_n).$$

Again, using (21) this implies

$$\|x_n - x_{n-1}\| \leq \tfrac{13}{4}\varepsilon^*, \quad x_n \in U, \ x_{n-1} \in U \tag{35}$$

and by (18)

$$v(x_n) < v(x_{n-1}) + \delta_2 - \delta_1 < \delta_2.$$

Now with $m_0 > n_0+2$, we can define

$$n_1 = \max\{i\colon v(x_{i-1}) < \delta_1, n_0+2 \leq i < m_0\},$$

$$n_2 = \min\{i\colon v(x_i) > \delta_2, n_0+2 < i \leq m_0\},$$

and obtain $n_2 > n_1 \geq n_0+2$. Summarizing,

$$\text{(i)} \quad v(x_{n_1-1}) < \delta_1, \quad v(x_{n_2}) > \delta_2, \tag{36}$$

$$\text{(ii)} \quad \delta_1 \leq v(x_i) \leq \delta_2, \quad \forall i\colon \ n_1 \leq i \leq n_2 - 1. \tag{37}$$

Take $n = n_1$ and $T = T^*$ in Lemmas 1 and 2. By (15) we have $T < \Delta = 1/(1+h_1) \leq 1$ and by (5), (37) we know that $\|x_{n_1}\| \leq M$, hence Lemmas 1 and 2 are applicable. By the selection of σ_0 we see that $1/n_1 \leq T^*$, hence the existence of $m(n, T)$ is guaranteed.

By a Taylor expansion we have

$$\begin{aligned} v(x_{m(n_1,T)+1}) - v(x_{n_1}) &= (x_{m(n_1,T)+1} - x_{n_1})^\tau v_x(x_{n_1}) \\ &\quad + \tfrac{1}{2}(x^\tau_{m(n_1,T)+1} - x^\tau_{n_1}) v_{xx}(\eta)(x_{m(n_1,T)+1} - x_{n_1}) \end{aligned} \tag{38}$$

where η is an R^m-vector located on the straight line between x_{n_1} and $x_{m(n_1,T)+1}$.

We aim to show (in (42) below) that $v(x_{m(n_1,T^*)+1}) < \delta_1$, and then demonstrate that this is the final contradiction. To that end, by Lemma 2,

$$\|x_{m(n_1,T)+1} - x_{n_1}\| \leq 3\varepsilon^* + T^*(h_1 + 2\varepsilon^*),$$

$$\|\eta - x_{n_1}\| \leq 3\varepsilon^* + T^*(h_1 + 2\varepsilon^*).$$

Hence $\|\eta\| \leq M + h_1 + 5\varepsilon^* \leq M + h_1 + 5$ so $\eta \in U$, and by (12)

$$\|v_{xx}(\eta)\| \leq r_2. \tag{39}$$

From this and (26), (38), (39) it follows that

$$
\begin{aligned}
&v(x_{m(n_1,T)+1}) - v(x_{n_1}) \\
&\leq \sum_{i=n_1}^{m(n_1,T)} \frac{1}{i}(h^\tau(x_i)+\xi_{i+1}^\tau)v_x(x_{n_1}) + \tfrac{1}{2}r_2[3\varepsilon^* + T(h_1+2\varepsilon^*)]^2 \\
&\leq \sum_{i=n_1}^{m(n_1,T)} \frac{1}{i}h^\tau(x_{n_1})v_x(x_{n_1}) + \sum_{i=n_1}^{m(n_1,T)} \frac{1}{i}(h^\tau(x_i)-h^\tau(x_{n_1}))v_x(x_{n_1}) \\
&\quad + \sum_{i=n_1}^{m(n_1,T)} \frac{1}{i}\xi_{i+1}^\tau v_x(x_{n_1}) + \tfrac{1}{2}r_2[3\varepsilon^* + T(h_1+2\varepsilon^*)]^2
\end{aligned} \tag{40}
$$

and, from (12), (35) with $n = n_1$,

$$
|v(x_{n_1}) - v(x_{n_1-1})| \leq \|v_x(\theta)\| \cdot \|x_{n_1} - x_{n_1-1}\| \leq \tfrac{13}{4} r_1 \varepsilon^*
$$

for some $\theta \in U$.

Thus by (36)

$$
v(x_{n_1}) \leq v(x_{n_1-1}) + \tfrac{13}{4} r_1 \varepsilon^* < \delta_1 + \tfrac{13}{4} r_1 \varepsilon^*. \tag{41}
$$

From this and (40) by (13), (21) and the fact that $\sum_{i=n_1}^{m(n_1,T)+1} 1/i > T$ we find

$$
\begin{aligned}
v(x_{m(n_1,T)+1}) &\leq \delta_1 + \tfrac{13}{4} r_1 \varepsilon^* - aT - \frac{h^\tau(x_{n_1})v_x(x_{n_1})}{m(n_1,T)+1} + \mathrm{T} r_1 \max_{n_1 \leq i \leq m(n_1,T)} \|h(x_{n_1}) - h(x_i)\| \\
&\quad + r_1\left[\left\|\frac{1}{m(n_1,T)} S_{m(n_1,T)}\right\| + \left\|\frac{1}{n_1} S_{n_1-1}\right\| + \left\|\sum_{i=n_1}^{m(n_1,T)-1} \frac{S_i}{i}\frac{1}{i+1}\right\|\right] \\
&\quad + \tfrac{1}{2}r_2[3\varepsilon^* + T(h_1+2\varepsilon^*)]^2 \\
&\leq \delta_1 + \tfrac{13}{4} r_1 \varepsilon^* - aT + \frac{r_1 h_1}{n_1} + \mathrm{T} r_1 \max_{n_1 \leq i \leq m(n_1,T)} \|h(x_i) - h(x_{n_1})\| \\
&\quad + r_1[\tfrac{11}{4}\varepsilon^* + \tfrac{11}{8}\varepsilon^* T] + \tfrac{1}{2}r_2[3\varepsilon^* + T(h_1+2\varepsilon^*)]^2 \\
&\leq \delta_1 - aT + 8r_1\varepsilon^* + Tr_1 \max_{n_1 \leq i \leq m(n_1,T)} \|h(x_i) - h(x_{n_1})\| \\
&\quad + \tfrac{1}{2}r_2[3\varepsilon^* + T(h_1+2\varepsilon^*)]^2.
\end{aligned}
$$

Paying attention to (15)-(17) and setting $T = T^*$ we finally have

$$
\begin{aligned}
v(x_{m(n_1,T^*)+1}) &\leq \delta_1 - aT^* + 8r_1\varepsilon^* + T^*(\alpha_1 - \beta) + \tfrac{1}{2}r_2(T^*)^2(h_1+2)^2 \\
&\quad + \tfrac{9}{2}r_2(\varepsilon^*)^2 + 3r_2\varepsilon^* T^*(h_1+2) \\
&\leq \delta_1 - T^*\beta + [8r_1 + r_2(3h_1+11)]\varepsilon^* < \delta_1
\end{aligned} \tag{42}
$$

On the other hand, by Lemma 2 and (12), (19) and (41) it follows that

$$
\begin{aligned}
v(x_m) &= v(x_{n_1}) + v_x^\tau(\phi)(x_m - x_{n_1}), \quad \phi \in U, \\
&\leq \delta_1 + 4r_1\varepsilon^* + r_1[3\varepsilon^* + T(h_1+2\varepsilon^*)] < \delta_2
\end{aligned}
$$

for all $m: n_1 \leq m \leq m(n_1, T)+1$. From this and (36), (37) we conclude that $m(n_1, T^*)+1 < n_2$, so

$$v(x_{m(n_1,T^*)+1}) \geq \delta_1.$$

But this contradicts (42), proving the theorem.

4. Proofs of Theorems 1 and 2

Proof of Theorem 1. By Theorem 3 we know that there is an N_1 independent of ε and $N_1 > N$ such that

$$\|x_n\| \leq M_\sigma,$$

and

$$x_{n+1} = x_n + \frac{1}{n}(h(x_n) + \xi_{n+1}) \tag{43}$$

for any $n \geq N_1$ and any $\varepsilon \in [0, \varepsilon^*)$.

Set

$$u_{n+1} = \frac{1}{n}\sum_{i=1}^{n} \xi_{i+1}.$$

Then

$$u_{n+1} = u_n - \frac{1}{n}u_n + \frac{1}{n}\xi_{n+1} \tag{44}$$

and

$$\limsup_{n\to\infty} \|u_n\| = \varepsilon$$

by Condition A. Hence

$$\|u_n\| \leq 2\varepsilon \quad \text{for any } n \geq N_2 \tag{45}$$

where N_2 is assumed to be greater than N_1 and sufficiently large.

From (43), (44) we have

$$x_{n+1} - u_{n+1} = x_n - u_n + \frac{1}{n}h(x_n) + \frac{1}{n}u_n, \quad \forall n > N_2$$

and

$$\begin{aligned} v(x_{n+1} - u_{n+1}) &= v(x_n - u_n) + \frac{1}{n}(h(x_n) + u_n)^\tau v_x(x_n - u_n) \\ &\quad + O\left(\frac{\|h(x_n)\|^2}{n^2} + \frac{\|u_n\|^2}{n^2}\right) \\ &\leq v(x_n - u_n) + \frac{1}{n}h^\tau(x_n)(v_x(x_n) - v_{xx}(\xi)u_n) + O\left(\frac{\varepsilon}{n}\right) + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Thus

$$v(x_{n+1}-u_{n+1}) \leqslant v(x_n-u_n)+\frac{1}{n}h^\tau(x_n)v_x(x_n)+\mathrm{O}\left(\frac{\varepsilon}{n}\right)+\mathrm{O}\left(\frac{1}{n^2}\right). \tag{46}$$

Now, define for any $\delta \geqslant 0$

$$\alpha(\delta)=\min_{\substack{\|x-x^0\|\geqslant\delta \\ \|x\|\leqslant M_\sigma}}\left(-h^\tau(x)v_x(x)/\max_{\|x\|\leqslant M_\sigma+2\varepsilon^*}v(x)\right). \tag{47}$$

Then we have

$$\alpha(\delta)v(x_n-u_n)\leqslant -h^\tau(x_n)v_x(x_n), \quad \text{for } \|x_n-x^0\|\geqslant\delta,$$

Since $\|x_n-u_n\|\leqslant M_\sigma+2\varepsilon^*$.

Noticing $h^\tau(x_n)v_x(x_n)\leqslant 0$, from (46) we know that

$$\begin{aligned} v(x_{n+1}-u_{n+1}) &\leqslant v(x_n-u_n)\left(1-\frac{\alpha(\delta)}{n}I_{[\|x_n-x^0\|\geqslant\delta]}\right)+\mathrm{O}\left(\frac{\varepsilon}{n}\right)+\mathrm{O}\left(\frac{1}{n^2}\right) \\ &=\left(1-\frac{\alpha(\delta)}{n}\right)v(x_n-u_n)+\frac{\alpha(\delta)}{n}v(x_n-u_n)I_{[\|x_n-x^0\|<\delta]} \\ &\quad +\mathrm{O}\left(\frac{\varepsilon}{n}\right)+\mathrm{O}\left(\frac{1}{n^2}\right). \end{aligned}$$

So

$$\begin{aligned} v(x_{n+1}-u_{n+1}) &\leqslant \prod_{i=N_2}^{n}\left(1-\frac{\alpha(\delta)}{i}\right)v(x_{N_2}-u_{N_2})+\sum_{i=N_2}^{n}\prod_{j=i+1}^{n}\left(1-\frac{\alpha(\delta)}{j}\right) \\ &\quad \times\left[\frac{\alpha(\delta)}{i}L\delta+\mathrm{O}\left(\frac{\varepsilon}{i}\right)+\mathrm{O}\left(\frac{1}{i^2}\right)\right] \end{aligned} \tag{48}$$

where

$$L=\max_{\|x\|\leqslant M_\sigma+\|x^0\|+2}\|v_x(x)\|,$$

and the last inequality follows since

$$\begin{aligned} v(x_n-u_n) &\leqslant v(x^0)+\|v_x(\xi_n)\|\cdot\|x_n-x^0-u_n\| \\ &\leqslant L(\|x_n-x^0\|+2\varepsilon). \end{aligned}$$

Here the η_n are vectors with components located in between corresponding components of x_n-u_n and x^0.

Applying the elementary inequality

$$\prod_{j=i+1}^{n}\left(1-\frac{\alpha}{j}\right)\leqslant\frac{(i+1)^\alpha}{(n+1)^\alpha} \quad (0\leqslant\alpha\leqslant i+1)$$

to (48) with $N_2 > \alpha(\delta)$ we obtain

$$v(x_{n+1}-u_{n+1}) \leq \left(\frac{N_2}{n+1}\right)^{\alpha(\delta)} v(x_{N_2}-u_{N_2}) + \frac{1}{(n+1)^{\alpha(\delta)}} \sum_{i=N_2}^{n} \frac{(i+1)^{\alpha(\delta)}}{i}$$
$$\times \left[\alpha(\delta)L\delta + \mathrm{O}(\varepsilon) + \mathrm{O}\left(\frac{1}{i}\right)\right]$$
$$= \left(\frac{N_2}{n+1}\right)^{\alpha(\delta)} v(x_{N_2}-u_{N_2}) + \frac{\mathrm{O}(1)}{(n+1)^{\alpha(\delta)}}$$
$$\times \left[(n^{\alpha(\delta)} - N_2^{\alpha(\delta)})L\delta + \frac{\mathrm{O}(\varepsilon)}{\alpha(\delta)}(n^{\alpha(\delta)} - N_2^{\alpha(\delta)})\right] + \mathrm{O}\left(\frac{1}{n(1-\alpha(\delta))}\right)$$
$$\leq \frac{\delta_n}{1-\alpha(\delta)} + c_1\delta + c_2\frac{\varepsilon}{\alpha(\delta)} \tag{49}$$

for all sufficiently small $\delta > 0$, where $c_1 > 0$, $c_2 > 0$ and $\delta_n \to_{n\to\infty} 0$ do not depend on ε.

By the selection of ε^* (20), for any $t \in (0, \varepsilon^*]$ we can define the function

$$\gamma(t) = \min\{\delta: \alpha(\delta) = t\},$$

since $\alpha(\delta)$ is continuous and nondecreasing. Clearly, $\gamma(t) \to 0$ as $t \to 0$.

Set $\delta = \gamma(\sqrt{\varepsilon})$ for any $\varepsilon \in (0, \varepsilon^*)$. Then $\alpha(\delta) = \sqrt{\varepsilon}$ and

$$v(x_{n+1}-u_{n+1}) \leq \frac{\delta_n}{1-\sqrt{\varepsilon}} + c_1\gamma(\sqrt{\varepsilon}) + c_2\sqrt{\varepsilon}. \tag{50}$$

Further, define

$$m(r) = \min_{\|x-x^0\| \geq r} v(x), \quad r \geq 0. \tag{51}$$

Clearly, $m(r)$ is a nondecreasing function of r tending to zero as $r \to 0$ and $m(0) = 0$.

Take r_n' such that

$$m(r_n') = 2\left(\frac{\delta_n}{1-\sqrt{\varepsilon}} + c_1\gamma(\sqrt{\varepsilon}) + c_2\sqrt{\varepsilon}\right).$$

By (50), (51) we see that

$$v(x_{n+1}-u_{n+1}) \leq \frac{m(r_n')}{2} < m(r_n')$$

and by (51)

$$\|x_{n+1}-u_{n+1}-x^0\| < r_n' = f\left(2\left(\frac{\delta_n}{(1-\sqrt{\varepsilon})} + c_1\gamma(\sqrt{\varepsilon}) + c_2\sqrt{\varepsilon}\right)\right) \tag{52}$$

where $f(t) = \min\{r: m(r) = t\}$ which, clearly, is left-continuous and increasing.

From (45) and (52) it follows that

$$\|x_{n+1}-x^0\| \leqslant f\left(2\left(\frac{\delta_n}{(1-\sqrt{\varepsilon})}+c_1\gamma(\sqrt{\varepsilon})+c_2\sqrt{\varepsilon}\right)\right)+2\varepsilon$$

and

$$\limsup_{n\to\infty}\|x_n-x^0\| \leqslant g(\varepsilon), \quad \varepsilon>0, \tag{53}$$

where

$$g(\varepsilon)=f(2c_1\gamma(\sqrt{\varepsilon})+2c_2\sqrt{\varepsilon})+2\varepsilon.$$

If $\varepsilon=0$, then by the arbitrariness of $\delta>0$ from (49) we have $x_n \to_{n\to\infty} x^0$. Thus the theorem has been proved for both the $\varepsilon>0$ and $\varepsilon=0$ case.

Proof of Theorem 2. *Sufficiency.* Sufficiency is implied in Theorem 1 by taking $\varepsilon=0$.

Necessity. Assume $x_n \to x^0$. The x_n cannot take on both x_1^* and x_2^* infinitely often since $x_1^* \neq x_2^*$. Therefore the number of truncations is finite, and there is an n_0 such that

$$\left\|x_n+\frac{1}{n}y_{n+1}\right\| \leqslant M_{\delta(n)} \quad \text{for } n \geqslant n_0.$$

By (7) we have

$$x_{n+1}=x_n+\frac{1}{n}h(x_n)+\frac{1}{n}\xi_{n+1}, \quad n \geqslant n_0$$

or

$$ix_{i+1}=ix_i+h(x_i)+\xi_{i+1}, \quad i \geqslant n_0. \tag{54}$$

Summing up both sides of (54) we obtain

$$nx_{n+1}=\sum_{i=n_0}^{n}x_i+\sum_{i=n_0}^{n}h(x_i)+\sum_{i=n_0}^{n}\xi_{i+1}+(n_0-1)x_{n_0}$$

and

$$x_{n+1}=\frac{1}{n}\sum_{i=n_0}^{n}x_i+\frac{1}{n}\sum_{i=n_0}^{n}h(x_i)+\frac{1}{n}\sum_{i=n_0}^{n}\xi_{i+g}1+\mathrm{o}(1). \tag{55}$$

Since $h(\cdot)$ is continuous and $x_n \to_{n\to\infty} x^0$ we see that $h(x_n) \to_{n\to\infty} 0$. Then from (55) it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\xi_{i+1} \xrightarrow[n\to\infty]{} 0$$

and Condition A holds with $\varepsilon=0$.

Acknowledgement

Particular thanks are due to the referee for a detailed review of the manuscript and for his useful suggestions and comments which have helped improve the quality of the paper.

References

[1] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statist.* 22(1) (1951) 400–407.
[2] M.B. Nevel'son and R.Z. Has'minskii, Stochastic Approximation and Recursive Estimation (American Mathematical Society, 1973).
[3] L. Ljung, Analysis of recursive stochastic algorithms, IEEE Trans. AC-22(4) (1977) 551–575.
[4] L. Ljung, Strong convergence of a stochastic approximation algorithm, Ann. Statist. 6(3) (1978) 680–696.
[5] H.J. Kushner and D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems (Springer, New York, 1978).
[6] V. Solo, Stochastic approximation with dependent noise, Stochastic Process. Appl. 13 (1982) 157–170.
[7] H.F. Chen, Stochastic approximation with ARMA measurement errors, J. of Systems Science and Mathematical Sciences 2(3) (1982) 227–239.
[8] H.F. Chen and Y.M. Zhu, Stochastic approximation procedure with randomly varying truncations, Scientia Sinica (Series A), 29(9), 914–926.
[9] H.F. Chen, Stochastic approximation with randomly varying truncations for the optimization problem, Acta Mathematicae Applicatae Sinica, 10(1) (1987) 58–67.
[10] S.H. Deng, Constrained stochastic approximation algorithm with randomly varying truncations, Acta Mathematicae Applicatae Sinica, 1986.
[11] H.F. Chen and L. Guo, Simultaneous estimation of both the zero of regression function and parameters in noise, J. of Systems Science and Mathematical Sciences, 7(2) (1987) 119–128.
[12] S.H. Deng, A combined algorithm for identification and approximation, Acta Mathematicae Applicatae Sinica (1987 to appear).
[13] D.S. Clark, Necessary and sufficient conditions for the Robbins-Monro method, Stochastic Process. Appl. 17 (1984) 359–367.