Stability and Instability of Limit Points for Stochastic Approximation Algorithms

Hai-Tao Fang and Han-Fu Chen, *Fellow, IEEE*

***Abstract*—It is shown that the limit points of a stochastic approximation (SA) algorithm compose a connected set. Conditions are given to guarantee the uniqueness of the limit point for a given initial value. Examples are provided wherein $\{x_n\}$ of SA algorithm converges to a limit $\overline{x}$ independent of initial values, but $\overline{x}$ is unstable for the differential equation $\dot{x} = f(x)$ with a nonnegative Lyapunov function. Finally, sufficient conditions are given for stability of $\dot{x} = f(x)$ at $\overline{x}$ if $\{x_n\}$ tends to $\overline{x}$ for any initial values.**

***Index Terms*—Pathwise convergent, stable/instable limit point, stochastic approximation.**

I. Introduction

THE stochastic approximation (SA) algorithm

$$x_{n+1} = x_n + a_n y_{n+1} \tag{1}$$

$$y_{n+1} = f(x_n) + \epsilon_{n+1} \tag{2}$$

proposed by Robbins and Monro [1] is used to search the root set $J \triangleq \{x: f(x) = 0\}$ based on the observations $\{y_n\}$, where $\{\varepsilon_n\}$ is the observation noise. SA methods are widely applied in systems identification [2], adaptive control [3], optimization, neural networks [4], and other fields [5]. In the convergence analysis [6]–[10] of $\{x_n\}$, most published results are concerned with $d(x_n, J) \to 0$, i.e., the distance between $\{x_n\}$ and J tends to zero. Example given in [11] shows that the convergence of $d(x_n, J)$ to zero does not imply the convergence of $\{x_n\}$ itself. In many applications of SA, however, people are not satisfied with $d(x_n, J) \to 0$; they are also interested in the convergence of $\{x_n\}$ itself. For the convergence of $\{x_n\}$, the sufficient conditions are given for the one-dimensional case (i.e., $f(\cdot): \mathbb{R} \to \mathbb{R}$) in [11] and for the multidimensional case in [12].

Under the boundedness assumption on $\{x_n\}$, the asymptotic part of the interpolating function of $\{x_n\}$ with interpolating length $\{a_n\}$ satisfies [2], [6] the following ordinary differential equation:

$$\dot{x} = f(x). \tag{3}$$

It turns out that the stability or instability of the equilibriums of (3) are of crucial importance for the behavior of the SA algorithms (1), (2). The essence of the ordinary differential equation (ODE) method consists in connecting properties of (3) with the convergence analysis of SA algorithms. Under certain conditions on $f(\cdot)$ and $\{\varepsilon_n\}$, the convergence of $\{x_n\}$ and estimation error bounds are established in [13] for the equilibrium being asymptotically stable and exponentially asymptotically stable, respectively. On the other hand, if a equilibrium $\overline{x}$ of (3) is unstable and if $f(\cdot)$ is the gradient of some function $L(\cdot)$ whose extreme is sought for, then $\overline{x}$ may be a saddle point of $L(\cdot)$, which has to be avoided in the optimization problem. Further, if an equilibrium of (3) is unstable, then a finite precision implementation of an SA algorithm might not converge, even though theoretical convergence is guaranteed.

The topics of this paper include general conditions on $f(\cdot)$ and $\{\epsilon_n\}$ to guarantee the convergence of $\{x_n\}$ itself and to establish the relationship between the convergence of $\{x_n\}$ and the stability of an equilibrium of (3).

In order to remove the commonly used conditions, such as the growth rate restriction on $f(\cdot)$ or the boundedness assumption [1], [2], [6], [7] on $\{x_n\}$, the algorithm with randomly varying truncations is defined in Section II. The connectedness of its limit points is also proven there. In Section III, it is shown that, if a point $\overline{x} \in J$ exists that is dominantly stable and if $\overline{x}$ is a limit point of $\{x_n\}$, then x_n must tend to $\overline{x}$. Sections IV and V discuss the converse problem: is the limit of $\{x_n\}$ a stable equilibrium of (3)? At first glance, if $\{x_n\}$ with an arbitrary initial value converges to $\overline{x}$, then (3) must be stable at $\overline{x}$, which is what intuition from ODE tells us. For SA, however, because of the noise, the picture is different from ODE. Examples presented in Section IV show that $\{x_n\}$, with an arbitrary initial value, converges to a limit unstable for (3). In Section V a reasonable condition is proposed to ensure $\overline{x}$ to be a stable equilibrium of (3) if $x_n \to \overline{x}$.

II. The Algorithm and Its Limit Points

To avoid restrictive conditions on $f(\cdot)$ the randomly truncated version of (1) and (2) is considered in [8]–[10] and is described as follows.

Consider $f(\cdot): \mathbb{R}^d \to \mathbb{R}^d$. Let $\{M_n\}$ be a sequence of positive numbers strictly diverging to ∞. Let $x^* \in \mathbb{R}^d$. Consider $\{x_n\}$ generated by the following algorithm truncated at randomly varying bounds:

$$\hat{x}_{n+1} = x_n + a_n y_{n+1} \tag{4}$$

$$x_{n+1} = \hat{x}_{n+1} 1_{\{\|\hat{x}_{n+1}\| \le M_{\sigma_n}\}} + x^* 1_{\{\|\hat{x}_{n+1}\| > M_{\sigma_n}\}} \tag{5}$$

$$\sigma_n = \sum_{i=1}^{n-1} 1_{\{\|\hat{x}_{i+1}\| > M_{\sigma_i}\}}, \qquad \sigma_0 = 0 \tag{6}$$

$$y_{n+1} = f(x_n) + \epsilon_{n+1}. \tag{7}$$

Manuscript received January 7, 1998; revised September 1, 1998 and February 12, 1999. Recommended by Associate Editor, J. C. Spall. This work was supported by the National Key Project of China, the National Natural Science Foundation of China, and the Chinese Academy of Science, under a post-doctoral grant.

The authors are with the Laboratory of Systems and Control, Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080 China (e-mail: hfchen@iss03.iss.ac.cn).

Publisher Item Identifier S 0018-9286(00)02140-1.

The following conditions are needed:

A1) $f(\cdot)$ is measurable and locally bounded;

A2) $a_n > 0$, $a_n \to 0$ and $\sum_{n=1}^{\infty} a_n = \infty$;

A3) a continuously differentiable function (not necessarily being nonnegative) $v(\cdot)$: $\mathbb{R}^d \to \mathbb{R}$ such that

$$\sup_{\Delta \geq d(x,J) \geq \delta} f^\tau(x) v_x(x) < 0 \tag{8}$$

for any $\Delta > \delta > 0$ and $v(J) \triangleq \{v(x): x \in J\}$ is nowhere dense, where $v_x(\cdot)$ denotes the gradient of $v(\cdot)$;

A4) for any convergent subsequence $\{x_{n_k}\}$

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k,t)} a_i \epsilon_{i+1} \right\| = 0, \qquad \forall t \in [0, T],$$

where $m(n, T) = \max\{m: \sum_{i=n}^{m} a_i \leq T\}$.

Remark 1: If $\sum_{i=1}^{\infty} a_i \epsilon_{i+1}$ converges, then A4) is obviously satisfied.

Remark 2: It is worth noting that for the convergence of $\{x_n\}$ a mild constraint exists on x^* stated below in Proposition 1. For example, if $f(x) = \sin x$, then $\cos x$ may serve as $v(x)$ in A3) and any x^* is satisfactory except $x^* = 2k\pi, k = 0, \pm 1, \pm 2, \cdots$. If the noise effect is not strong, then x_n converges to that root of $\sin x$, which is the closest to x^*.

In what follows, ω denotes a point in the underlying probability space. Then, for a fixed ω and a fixed initial value x_0, the algorithm (4)–(7) defines a trajectory $\{x_n(\omega, x_0)\}$. For simplicity of notations, we always suppress (ω, x_0) and write x_n instead of $x_n(\omega, x_0)$. The noise ε_n may depend not only on ω, but also on the trajectory up to time $n-1$ with x_{n-1} included.

Proposition 1: Assume A1)–A3) hold and that c_0 exists such that $v(x^*) < \inf_{\|x\|=c_0} v(x)$ and $\|x^*\| < c_0$. Let $\{x_n\}$ be defined by (4)–(7) for some initial value x_0. If A4) holds for a given trajectory, then for this trajectory

$$\lim_{n\to+\infty} d(x_n, J) = 0.$$

The proof is given in [8], but for convenience, we attach it in the Appendix. If J is a singleton x^0 and if $f(\cdot)$ is continuous at x^0, then the converse is also true; i.e., "$x_n \to x^0$" implies A4). Under the assumptions of Proposition 1, several conditions are proven to be equivalent to A4) in [14]. We now show that limit points of $\{x_n\}$ compose a connected set under the conditions required in Proposition 1.

Proposition 2: Assume conditions of Proposition 1 are satisfied. Then, for fixed x_0 and ω, a connected subset $J^* \subset \overline{J}$ exists such that

$$d(x_n, J^*) \underset{n\to+\infty}{\longrightarrow} 0$$

where $\overline{J}$ denotes the closure of J and $\{x_n\}$ is generated by (4)–(7).

Proof: Denote by J^* the set of limit points of $\{x_n\}$. Assume the converse: i.e., J^* is disconnected. In other words, closed sets J_1^* and J_2^* exist such that $J^* = J_1^* \cup J_2^*$ and $d(J_1^*, J_2^*) > 0$.

Take $\rho := 3^{-1} d(J_1^*, J_2^*)$. Because $d(x_n, J^*) \underset{n\to+\infty}{\longrightarrow} 0$, an N_0 exists such that

$$x_n \in B(J_1^*, \rho) \cup B(J_2^*, \rho), \qquad \forall n > N_0$$

where $B(A, \rho)$ denotes the ρ-neighborhood of set A. Define

$$n_0 = \inf\{n > N_0, d(x_n, J_1^*) < \rho\}$$
$$m_l = \inf\{n > n_l, d(x_n, J_2^*) < \rho\}$$
$$n_{l+1} = \inf\{n > m_l, d(x_n, J_1^*) < \rho\}.$$

It is clear that $m_l < \infty$, $n_l < \infty$, $\forall l > 0$ and

$$x_{n_l} \in B(J_1^*, \rho), \qquad x_{n_l-1} \notin B(J_1^*, \rho). \tag{9}$$

By Proposition 1, $\{x_n\}$ is bounded, and after a finite time, the algorithm (4)–(7) becomes the one without truncations. Without loss of generality, it may be assumed that $x_{n_l} \underset{l\to+\infty}{\longrightarrow} x_1^*$, $x_{n_l-1} \underset{l\to+\infty}{\longrightarrow} x_2^*$. By (9), it follows that $\|x_1^* - x_2^*\| \geq \rho$.

On the other hand, by A1) and A4)

$$\|x_{n_l} - x_{n_l-1}\| \leq a_{n_l-1} \sup_{n\geq 0} \|f(x_n)\| + \|a_{n_l-1}\epsilon_{n_l}\| \to 0.$$

The obtained contradiction shows that J^* is connected. □

III. Dominant Stability

Under the conditions of Proposition 1, although the distance between $\{x_n\}$ and J tends to zero, $\{x_n\}$ may still not converge, if J is not a singleton. This result is because $\{x_n\}$ may still walk about in J even though the sequence $\{x_n, n \geq m\}$ is contained in J starting from some m, which means that $d(x_n, J) = 0$, $\forall n \geq m$. Let us take the example given in [11], where

$$f(x) = \begin{cases} x, & x \leq 0 \\ 0, & x \in [0, 1], \ J = [0, 1] \\ x - 1, & x \geq 1 \end{cases}$$

$a_0 = 1$, $a_1 = 1, \cdots, a_{2^n} = \cdots = a_{2^{n+1}-1} = \sqrt{1/2^n}$, $\epsilon_1 = 1$, $\epsilon_2 = -1$, $\epsilon_{2^n+1} = \cdots = \epsilon_{2^{n+1}} = (-1)^{n-1}\sqrt{1/2^n}$. Then, $\sum_{i=0}^{n} a_i \epsilon_{i+1}$ changes from one to zero and then from zero to one, and this process repeats forever with decreasing stepsizes. Take $x_0 = 0$. Then, $x_{2^n+m} = \sum_{i=0}^{2^n+m-1} a_i \epsilon_{i+1}$, $m = 0, 1, \cdots$. It is clear that $d(x_n, J) = 0$, and all conditions A1)–A4) are satisfied, but $\{x_n\}$ is dense in [0, 1]. This phenomenon hints that for convergence of $\{x_n\}$ the stability-like condition A3) is not enough; a stronger stability is needed.

Definition 1: A point $\overline{x} \in J$ is called dominantly stable for $f(\cdot)$, if a $\delta_0 > 0$ and a positive measurable function $c(\cdot)$: $(0, +\infty) \to (0, +\infty)$ exist satisfying the following condition that $c(x)$ at the interval $[\delta, \delta_0]$ is bounded from below by a positive constant for any $\delta \in (0, \delta_0)$ such that

$$f^\tau(x)(x - \overline{x}) \leq -c(\|x - \overline{x}\|)\|f(x)\| \tag{10}$$

for all $x \in B(\overline{x}, \delta_0)$.

Remark 3: The dominant stability implies stability. To see this, it suffices to take $v(x) = \|x - \overline{x}\|^2$ as the Lyapunov function. The dominant stability of $\overline{x}$, however, is not necessary for asymptotic stability.

Remark 4: Equation (10) holds for any $x \in J$, whatever $\overline{x}$ is. Therefore, all interior points of J are dominantly stable for $f(\cdot)$. Further, for a boundary point $\overline{x}$ of J to be dominantly stable for $f(\cdot)$, it suffices to verify (10) for $x \in B(\overline{x}, \delta_0) \cap J^c$ with small δ_0, i.e., all x that are close to $\overline{x}$ and outside J.

Theorem 1: Assume A1)–A3) hold. If for a given ω, $\sum_{i=1}^{\infty} a_i \epsilon_{i+1}$ is convergent and a limit point $\overline{x}$ of $\{x_n\}$ generated by (4)–(7) is dominantly stable for $f(\cdot)$, then for this trajectory, $x_n \underset{n\to+\infty}{\to} \overline{x}$.

Proof: For any $\delta \in (0, \delta_0/3)$, define

$$\begin{aligned} n_0 &= \inf\{n > 0,\, x_n \in B(\overline{x}, \delta)\} \\ m_i &= \inf\left\{n > n_i,\, x_n \in \overline{B(\overline{x}, 2\delta)} \setminus B(\overline{x}, \delta)\right\} \\ l_i &= \inf\left\{n > m_i,\, x_n \notin \overline{B(\overline{x}, 2\delta)} \setminus B(\overline{x}, \delta)\right\} \\ n_{i+1} &= \inf\{n \geq l_i,\, x_n \in B(\overline{x}, \delta)\}. \end{aligned}$$

It is clear that n_0 is well defined, because $x_{n_k} \to \overline{x}$ and $x_{n_k} \in B(\overline{x}, \delta)$ for any k greater than some k_0. If for any $\delta > 0$, $m_i = +\infty$ for some i, then $x_n \underset{n\to+\infty}{\to} \overline{x}$ by arbitrariness of δ. Therefore, for proving the theorem, it suffices to show that, for any small $\delta > 0$, an N_0 exists such that "$m_i < +\infty$" implies "$l_i = n_{i+1}$" if $m_i > N_0$.

Assume l_i is large enough so that the truncations no longer exist in (4)–(7). It then follows that

$$\begin{aligned} \|x_{l_i} - \overline{x}\|^2 &= \|x_{l_i} - x_{l_i-1} + x_{l_i-1} - \overline{x}\|^2 \\ &= a_{l_i-1}^2 \|f(x_{l_i-1})\|^2 + 2a_{l_i-1}^2 f^\tau(x_{l_i-1})\epsilon_{l_i} \\ &\quad + 2a_{l_i-1} f^\tau(x_{l_i-1})(x_{l_i-1} - \overline{x}) \\ &\quad + \|x_{l_i-1} - \overline{x} + a_{l_i-1}\epsilon_{l_i}\|^2. \end{aligned} \tag{11}$$

Notice that for any $x \in \overline{B(\overline{x}, 2\delta)} \setminus B(\overline{x}, \delta)$, $c(\|x - \overline{x}\|) \geq c_1 > 0$ and $\|f(x)\|$ is bounded by A1), and hence by (10)

$$\begin{aligned} & a_n^2 \|f(x)\|^2 + 2a_n^2 f^\tau(x)\epsilon_{n+1} + 2a_n f^\tau(x)(x - \overline{x}) \\ &\quad \leq a_n \|f(x)\|(a_n\|f(x)\| + \|2a_n \epsilon_{n+1}\| - 2c(\|x - \overline{x}\|)) \\ &\quad \leq 0 \end{aligned} \tag{12}$$

$\forall\, n > N_1$, for some N_1, because $\sum_{n=1}^{\infty} a_n \varepsilon_{n+1}$ is convergent and $a_n \underset{n\to+\infty}{\to} 0$. Combining (11) and (12) leads to

$$\|x_{l_i} - \overline{x}\|^2 \leq \|x_{l_i-1} - \overline{x} + a_{l_i-1}\epsilon_{l_i}\|^2$$

for $m_i > N_1$.

Further

$$\begin{aligned} & \|x_{l_i-1} - \overline{x} + a_{l_i-1}\epsilon_{l_i}^2\| \\ &\quad = \|x_{l_i-1} - x_{l_i-2} + x_{l_i-2} + \overline{x} + a_{l_i-1}\epsilon_{l_i}\|^2 \\ &\quad = \|a_{l_i-2} f(x_{l_i-2}) + x_{l_i-2} - \overline{x} + a_{l_i-2}\epsilon_{l_i-1} + a_{l_i-1}\epsilon_{l_i}\|^2. \end{aligned}$$

The similar treatment yields

$$\begin{aligned} & \|x_{l_i-1} - \overline{x} + a_{l_i-1}\epsilon_{l_i}\|^2 \\ &\quad \leq \|x_{l_i-2} - \overline{x} + a_{l_i-2}\epsilon_{l_i-1} + a_{l_i-1}\epsilon_{l_i}\|^2. \end{aligned}$$

Inductively, it then follows that

$$\begin{aligned} \|x_{l_i} - \overline{x}\|^2 &\leq \left\| x_{m_i} - \overline{x} + \sum_{k=m_i}^{l_i-1} a_k \epsilon_{k+1} \right\|^2 \\ &\leq 3 \left\| \sum_{k=m_i}^{l_i-1} a_k \epsilon_{k+1} \right\|^2 + \tfrac{3}{2}\|x_{m_i} - \overline{x}\|^2 \\ &\leq 3 \left\| \sum_{k=m_i}^{l_i-1} a_k \epsilon_{k+1} \right\|^2 + 15 a_{m_i-1}^2 (\|f(x_{m_i-1})\|^2 \\ &\quad + \|\epsilon_{m_i}\|^2) + \tfrac{15}{8}\delta^2 \end{aligned}$$

where the elementary inequality

$$\|x + y\|^2 \leq (1 + a)\|x\|^2 + \left(1 + \frac{1}{a}\right)\|y\|^2, \qquad \forall\, a > 0$$

is used. Because $\|f(x_{m_i-1})\|$ is bounded, $\|\sum_{k=m_i}^{l_i-1} a_k \epsilon_{k+1}\| \to 0$ and $\|a_{m_i-1}\epsilon_{m_i}\| \underset{i\to+\infty}{\to} 0$, an N_2 exists such that

$$\|x_{l_i} - \overline{x}\|^2 \leq \tfrac{31}{16}\delta^2, \qquad \forall\, m_i > N_2.$$

This process means that $l_i = n_{i+1}$. □

Remark 5: The convergence of x_n is also considered in [15], but under different conditions. We now make a comparison between conditions used here and in [15]. First, [15, Theorem 2] requires a condition called A-stability, which implies the boundedness of $\{x_n\}$, while here (Theorem 1) does not make any *a priori* assumption on $\{x_n\}$. Second, concerning the noise $\{\epsilon_i\}$ [15] requires that it can be decomposed into a sum $\epsilon_n = e_n + r_n$ such that $\sum_{i=1}^{\infty} a_i e_{i+1} < \infty$ and $r_n \underset{n\to\infty}{\to} 0$. Obviously, this decomposition implies A4). Furthermore, [15] requires an additional condition (see [15, (10)] for the case $\delta_p = 0$)

$$\sum_p \left[a_p \left| \sum_{i=p}^{\infty} a_i e_{i+1} \right| + a_p^2 + a_p |r_{p+1}| \right] < \infty$$

which, however, is rather difficult to check. Third, both [15] and this paper require some attractiveness of J: in [15] Condition B) is used, and here the dominant stability is applied.

Example 1: Consider (4)–(7) with

$$f(x) = -\nabla L(x)$$

where

$$L(x) = \begin{cases} (\|x\|^2 - 1)^2, & \text{if } \|x\| > 1 \\ 0, & \text{otherwise} \end{cases}$$

i.e., (4)–(7) is used for seeking the extrema of $L(x)$ based on the noisy observations

$$y_{n+1} = -\nabla L(x_n) + \varepsilon_{n+1}.$$

Then, all points of $J = \{x \in \mathbb{R}^d,\, \|x\| \leq 1\}$ are dominantly stable for $f(\cdot)$.

As mentioned before, it suffices to show that all $\overline{x}$ with $\|\overline{x}\| = 1$ are dominantly stable for $f(\cdot)$. For any x with $\|x\| > 1$, it is seen that

$$\begin{aligned} f^\tau(x)(x-\overline{x}) &= -2(\|x\|^2-1)x^\tau(x-\overline{x}) \\ &= -\|f(x)\|\frac{x^\tau(x-\overline{x})}{\|x\|} \\ &= -\|f(x)\|\,\|x-\overline{x}\|\cos\angle(x,\,x-\overline{x}) \end{aligned}$$

where $\angle(x,\,x-\overline{x})$ denotes the angle between the vectors x and $x-\overline{x}$. It is clear that

$$\inf_{\|x-\overline{x}\|=\delta,\|x\|>1}\cos\angle(x,x-\overline{x})>0$$

for all small enough $\delta > 0$. This process verifies that all points in J are dominantly stable for $f(\cdot)$.

A1) and A3) are automatically satisfied in this example, because $L(x)$ may serve as $v(x)$ in A3).

Therefore, if A2) holds and $\sum_{i=1}^\infty a_i\varepsilon_{i+1}$ is convergent for some trajectory, then for this trajectory $\{x_n\}$ given by (4)–(7) converges to a point $\overline{x}$ belonging to J.

Remark 6: If the results given in [12] are applied to Example 1, then for convergence of $\{x_n\}$ one needs to impose rather restrictive conditions on $\{\varepsilon_n\}$ and to have $a_n = c/n + \sigma(1/n)$.

IV. Instability of the Limit

In this section, three examples of $f(\cdot)$ and $\{\varepsilon_n\}$ are given. For each of them, the corresponding stochastic approximation algorithm converges to a limit $\overline{x}$ independent of initial value, but $\overline{x}$ is not stable for (3). In the first example, the stability-like condition A3) is not satisfied, and for the remaining examples, A3) is satisfied with $v(\cdot)$ even nonnegative.

Example 2: Let $x \triangleq (x_1,\,x_2,\,x_3)^\tau$. Here, we use x_1, x_2, and x_3 to denote the components of x just for simplicity with the hope that they will not be confused with the estimates $\{x_n\}$. Let $f(\cdot)$: $\mathbb{R}^3 \to \mathbb{R}^3$ be defined as follows:

$$f(x) = f_1(x)+f_2(x)+f_3(x)+f_4(x)$$

where

$$\begin{aligned} f_1(x) &= \varphi(1-\|x\|^2)x \\ f_2(x) &= \psi(\|x\|^2)x_3^2(x_3,\,0,\,-x_1)^\tau \\ f_3(x) &= \psi(\|x\|^2)x_2(x_2,\,-x_1,\,0)^\tau \\ f_4(x) &= \psi(\|x\|^2)x_2^3\begin{pmatrix} -x_1x_2\left(x_3^2-x_3\right) \\ x_3\left(x_1^2x_3+x_2^2+x_3^3\right) \\ -x_2\left(x_1^2+x_2^2+x_3^3\right)\end{pmatrix} \end{aligned}$$

where $\psi(t)$ is a $C^2([0,\,+\infty),[0,\,+\infty))$ decreasing function satisfying

$$\psi(t) = \begin{cases} 1, & t<2 \\ 0, & t>3 \end{cases}$$

and $\varphi(t)$ is a $C^2(\mathbb{R},\,\mathbb{R})$ increasing function satisfying

$$\varphi(t) = \begin{cases} -2, & t<-2 \\ t, & t>-1. \end{cases}$$

It is easy to check that $f_i(x)$, $i = 1, 2, 3, 4$ are mutually orthogonal, $J = \{(0,\,0,\,0),\,(\pm1,\,0,\,0)\}$ and

$$\begin{aligned} \left.\frac{\partial f_1(x)}{\partial x}\right|_{x=(\pm1,0,0)} &= \begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \left.\frac{\partial f_1(x)}{\partial x}\right|_{x=(0,0,0)} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ \left.\frac{\partial f_3(x)}{\partial x}\right|_{x=(\pm1,0,0)} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mp1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \left.\frac{\partial f_2(x)}{\partial x}\right|_{x=(\pm1,0,0),(0,0,0)} &= \left.\frac{\partial f_3(x)}{\partial x}\right|_{x=(0,0,0)} \\ &= \left.\frac{\partial f_4(x)}{\partial x}\right|_{x=(\pm1,0,0),(0,0,0)} = 0. \end{aligned}$$

Therefore, all three points in J are unstable for (3).

Consider the SA algorithm

$$\begin{cases} y_{n+1} = f(x_n)+\varepsilon_{n+1} \\ x_{n+1} = x_n + a_n y_{n+1} \end{cases} \tag{13}$$

with $a_n = \alpha/n$, $\alpha < 1/4$. Let ε_n be i.i.d. and uniformly distributed in $[-1, 1]$. By [12] $x_n \to \{(\pm1,\,0,\,0),\,(0,\,0,\,0)\}$, and the results in [16] show that

$$P\{x_n \to \{(-1,\,0,\,0),\,(0,\,0,\,0)\}\} = 0.$$

Hence, for any $x_0 \in \mathbb{R}^3$, $x_n \to (1,\,0,\,0)$ a.e. In other words, the limit of $\{x_n\}$ is unstable.

We might say that this phenomenon might be caused by the nonexistence of a stable attractor for (3). The following two examples show that even though a stable attractor for (3) exists, $\{x_n\}$ may still converge to a unstable $\overline{x}$.

Example 3: Let $f(x) = \sin x$, $a_n = 1/(n+1)$, $n \ge 0$, $\varepsilon_1 = -(x+\sin x)$, $\varepsilon_i = 0$, $\forall\, i > 1$. For any $x_0 \in \mathbb{R}$, take $M_0 = \|x_0\| + 1$. It is clear that A1)–A4) hold with $v(x) = \cos x$, and $\{k\pi/2,\, k = \pm1, \pm2, \cdots\}$ are stable attractors of (3). However, $x_n \underset{n\to\infty}{\to} 0$ for any $x_0 \in \mathbb{R}$ and zero is not stable for $f(\cdot)$.

In this example, ε_1 strongly depends on x and thus is rather special. In the following example, $\{\varepsilon_n\}$ is general, a stable attractor for (3) exists, and A1)–A4) hold with $v(\cdot)$ even nonnegative.

Example 4: Let

$$f(x) = \begin{cases} -x, & x \le 0 \\ x, & 0 < x \le \frac{1}{2} \\ -x+1, & x > \frac{1}{2}. \end{cases}$$

It is straightforward to check that

$$v(x) = \begin{cases} x^2+\frac{1}{2}, & x \le 0 \\ -x^2+\frac{1}{2}, & 0 < x \le \frac{1}{2} \\ (x-1)^2, & x > \frac{1}{2} \end{cases}$$

satisfies A3). Take $a_n = 1/(n+1)$, $n \ge 0$, $\varepsilon_1 = -1$, $\varepsilon_n = \rho_n < 0$, $\forall\, n \ge 1$, where $\{\rho_n\}$ is a sequence of mutually independent random variables such that $\sum_{n=0}^\infty a_n\rho_{n+1} < +\infty$ a.s.

Then, $J = \{0, 1\}$ with 1 being a stable attractor for (3) and all A1)–A4) are satisfied. Take $M_0 = \|x_0\| + 1$. Then, by Proposition 1 and by the fact that $\varepsilon_n < 0$, it follows that $x_n \underset{n\to\infty}{\longrightarrow} 0$ a.s. Zero, however, is unstable for $f(\cdot)$.

V. Stability of the Limit

In Examples 3 and 4, A1)–A4) are satisfied, and $\{x_n\}$ converges to a limit, which is independent of initial value and unstable. This strange phenomenon happens because $F_n(x_n) \triangleq x_n + a_n f(x_n) + a_n \varepsilon_{n+1}$ as a function of x_n is singular for some n, $n = 0, 1, \cdots$ in the sense that it restricts the algorithm in a certain set in $\mathbb{R}^d$. Therefore, in order for the limit of $\{x_n\}$ to be stable, certain regularity conditions on $F_n(x_n)$ and some restrictions on noises are unavoidable.

In what follows, we specify the noise $\varepsilon_{n+1}(\cdot, \cdot)$ in observation

$$y_{n+1} = f(x_n) + \varepsilon_{n+1}(\omega, x_n)$$

as a Borel function defined on the product space $\Omega \times \mathbb{R}^d$, where Ω denotes the basic probability space and set

$$F_n(x) = x + a_n f(x) + a_n \epsilon_{n+1}(\omega, x). \tag{14}$$

Let us introduce the following conditions.

A0) For a given ω, $F_n(x)$ is a surjection for any $n = 0, 1, \cdots$.

A4′) For any ω and $n, \varepsilon_n(\omega, x) \in C(\mathbb{R}^d, \mathbb{R}^d)$ and for any ω, $x \in \mathbb{R}^d$ and $\delta > 0$,

$$\lim_{T\to 0} \lim_{n\to\infty} \frac{1}{T} \sup_{\substack{t\in[0,T] \\ x_n \in B(x,\delta)}} \left\| \sum_{i=n}^{m(n,t)} a_i \varepsilon_{i+1}(\omega, x_i) \right\| = 0 \tag{15}$$

where $B(x, \delta)$ denotes the ball with radius δ centered at x.

Remark 7: A4′) is equivalent to the following condition: For any ω and any compact set $K \subset \mathbb{R}^d$

$$\lim_{T\to 0} \lim_{n\to\infty} \frac{1}{T} \sup_{\substack{t\in[0,T] \\ x_n \in K}} \left\| \sum_{i=n}^{m(n,t)} a_i \varepsilon_{i+1}(\omega, x_i) \right\| = 0.$$

Remark 8: If ε_i does not depend on x_{i-1}, then (15) is equivalent to condition A4). Condition A4) is applied to the case in which the initial value of $\{x_n\}$ is arbitrary but fixed. In this case a convergent subsequence $\{x_{n_k}\}$ is automatically located in a compact set. Theorem 2, however, will consider the case in which the initial value arbitrarily varies, and hence x_n for any fixed n may be any point in $\mathbb{R}^d$. If x_n in A4′) were not restricted to a compact set [i.e., with "$x_n \in B(x, \delta)$" removed in (15)], then the resulting condition would be too strong. Therefore, to put "$x_n \in B(x, \delta)$" in (15) is to make the condition reasonable.

Remark 9: If $F_n(\cdot)$ is continuous and if $x^\tau F(x)/\|x\| \to \infty$ as $\|x\| \to \infty$, then $F_n(\cdot)$ is a surjection by [17, Theorem 3.3]. Using this property, we can show that $F_n(\cdot)$ is a surjection for a large class of $f(\cdot)$. To see this, let ϵ_n be free of x. In the case in which the growth rate of $\|f(x)\|$ is not faster than linearly as $\|x\| \to \infty$, then $\{a_n\}$ can be selected such that A2) holds and $x^\tau F_n(x)/\|x\| \underset{\|x\|\to\infty}{\longrightarrow} \infty$ for all $n \geq 0$. Hence, A0) holds. In the case in which the growth rate of $\|f(x)\|$ is faster than linearly as $\|x\| \to \infty$ and $x^\tau f(x) \geq 0\ \forall x$: $\|x\| \geq K$ for some $K > 0$, then $x^\tau F_n(x)/\|x\| \underset{\|x\|\to\infty}{\longrightarrow} \infty$ for all $n \geq 0$ and A0) is satisfied.

Theorem 2: Assume A1)–A4′) hold and that c_0 exists such that $v(x^*) < \inf_{\|x\|=c_0} v(x)$ and $\|x^*\| < c_0$ and for a given ω A0) holds. If $\{x_n\}$ defined by (4)–(6) and (14) with any initial value x_0 converges to a limit $\overline{x}$ independent of x_0, then $\overline{x}$ belongs to the unique stable set of (3).

Proof: Because $v(x^*) < \inf_{\|x\|=c_0} v(x)$ and $\|x^*\| < c_0$, by the continuity of $v(x)$, $\hat{x}$ exists with $\|\hat{x}\| < c_0$ such that $v(\hat{x}) = \min_{\|x\|\leq c_0} v(x)$. Thus, $v_x(\hat{x}) = 0$. By A3) and the continuity of $f(x)$, $f^\tau(x) v_x(x) < 0$ for any x for which $f(x) \neq 0$. Thus, we must have $f(\hat{x}) = 0$, i.e. $\hat{x} \in J$. Denote by J_0 the connected component of J that contains $\hat{x}$. We know that $v(J_0)$ is a connected set and $v(J)$ is nowhere dense. Then $v(J_0)$ is a constant. By the Lyapunov Theorem, J_0 is a stable set; i.e., any x belongs to J_0 is stable.

Let A_h be the connected component of $\{x \in \mathbb{R}^d, v(x) < v(\hat{x}) + h, h > 0\}$ such that A_h contains J_0. By the continuity of $v(x)$ for an arbitrary small $\delta > 0$, $h_2 > h_1 > 0$ exist such that $h_2 < \delta, A_{h_2} \subset B(J_0, \delta)$ and the distance between the interval $[v(\hat{x}) + h_1, v(\hat{x}) + h_2]$ and the set $v(J)$ is positive; i.e., $d([v(\hat{x}) + h_1, v(\hat{x}) + h_2], v(J)) > 0$. For the given ω, we show that an n_0 exists such that $x_n(\omega, x_0) \in A_{h_2}$ for any $n > n_0$ whenever $x_{n_0}(\omega, x_0) \in A_{h_1}$ for some $x_0 \in \mathbb{R}^d$. Here, the trajectories $\{x_n(\omega, x_0)\}$ are considered for the same fixed ω, but with different initial values. We first note that with ω fixed, for any n_0 by A0) x_0 exists such that $x_{n_0}(\omega, x_0) \in A_{h_1}$. By arbitrary smallness of $\delta > 0$, from this it follows that $d(x_n, J_0) \underset{n\to\infty}{\longrightarrow} 0$. Therefore, $\overline{x} \in J_0$, which means $\overline{x}$ is stable. If another stable set J_0' existed such that $J_0' \cap J_0 = \emptyset$, then by the same argument $\overline{x}$ would belong to J_0'. The contradiction shows the uniqueness of the stable set.

We now prove the existence of n_0 such that $x_n(\omega, x_0) \in A_{h_2}$ for any $n > n_0$ if $x_{n_0}(\omega, x_0) \in A_{h_1}$. We first show that, for any $\omega, x \in \mathbb{R}^d$ and $\hat{\delta} > 0$, $c_1 > 0$, $T_1 > 0$, and n_T exist such that, for any $n > n_T$ if $x_n(x_0) \in B(x, \hat{\delta})$, then

$$\|x_m(x_0) - x_n(x_0)\| \leq c_1 T$$
$$\forall T \in (0, T_1],\ \forall m: n \leq m \leq m(n, T) \tag{16}$$

where, and hereafter, $x_n(\omega, x_0)$ is written as $x_n(x_0)$ for simplicity.

Note that

$$x_m(x_0) - x_n(x_0) = \sum_{i=n}^{m-1} a_i f(x_i(x_0)) + \sum_{i=n}^{m-1} a_i \varepsilon_{i+1}(\omega, x_i(x_0)). \tag{17}$$

For any $M > 0$, let $c = \max_{y\in \overline{B(x, M+\hat{\delta})}} |f(y)|$. By A4′), sufficiently small $T_1 < M/3c$ and n_T exist such that for any $n > n_T$

$$\left\| \sum_{i=n}^{m-1} a_i \varepsilon_{i+1}(\omega, x_i(x_0)) \right\| < cT,$$
$$\forall T \in [0, T_1],\ \forall m: n < m \leq m(n, T).$$

If $x_i \in B(x, M + \hat{\delta})$ for $n \leq i \leq m - 1, \forall n$, then (16) immediately follows by setting $c_1 = 2c$. Assume $x_i \notin B(x, M + \hat{\delta})$ for some i: $n < i \leq m$. Let $i_0 > n$ be the first one. Then

$$\|x_{i_0}(x_0) - x_n(x_0)\| > M. \tag{18}$$

By (17), however

$$\begin{aligned}\|x_{i_0}(x_0) - x_n(x_0)\| &\le \sum_{i=n}^{i_0-1} a_i\|f(x_i(x_0))\| + \left\|\sum_{i=n}^{i_0-1} a_i\varepsilon_{i+1}(\omega, x_i(x_0))\right\| \\ &\le cT + cT \le M\end{aligned}$$

which is contradictory with (18). Thus, $x_i \in B(x, \hat{\delta} + M)$, $\forall\, i \le m$, and (16) is verified.

Assume the assertion above is not true; i.e., for any N, x_0^N exists such that $x_N(x_0^N) \in A_{h_1}$ and $x_{N'}(x_0^N) \notin A_{h_2}$ for some $N' > N$. Suppose $N \le k_N < k'_N \le N'$ and $x_{k_N}(x_0^N) \in \overline{A}_{h_1}$, $x_{k'_N}(x_0^N) \in \overline{A}_{h_2}^c$, $x_i(x_0^N) \in A_{h_2}\backslash\overline{A}_{h_1}$, i.e., $v(\hat{x}) + h_1 \le v(x_i(x_0^N)) \le v(\hat{x}) + h_2\ \forall\, i, k_N < i < k'_N$. Because

$$x_{k_N}\left(x_0^N\right) \in B(J_0, \delta) \subset \{x \in \mathbb{R}^d, \|x\| \le c_0 + \delta\}$$

a subsequence of $\{x_{k_N}(x_0^N)\}$ exists, also denoted by $\{x_{k_N}(x_0^N)\}$ for notational simplicity, such that $x_{k_N}(x_0^N) \underset{N\to+\infty}{\longrightarrow} \tilde{x}$. By the continuity of $v(x)$, $v(\tilde{x}) = v(\hat{x}) + h_1$. Hence, $d(\tilde{x}, J) := \tilde{\delta} > 0$. By (16) and the fact $x_{k_N}(x_0^N) \to \tilde{x}$, we can choose sufficiently small T and large enough N such that

$$d\left(x_m\left(x_0^N\right), J\right) \ge \frac{\tilde{\delta}}{2} \tag{19}$$

and $m(k_N, T) < k'_N$, i.e.,

$$v(\hat{x}) + h_1 \le v\left(x_m\left(x_0^N\right)\right) \le v(\hat{x}) + h_2 \tag{20}$$

for any m: $k_N \le m \le m(k_N, T)$. By (16), ξ exists with the property $\|\xi - x_{k_N}(x_0^N)\| \le c_1T$ such that

$$\begin{aligned}&v\left(x_{m(k_N,T)+1}\left(x_0^N\right)\right) - v\left(x_{k_N}\left(x_0^N\right)\right)\\ &\quad= \left(x_{m(k_N,T)+1}\left(x_0^N\right) - x_{k_N}\left(x_0^N\right)\right)^\tau v_x(\tilde{x})\\ &\qquad+ \left(x_{m(k_N,T)+1}\left(x_0^N\right) - x_{k_N}\left(x_0^N\right)\right)^\tau (v_x(\xi) - v_x(\tilde{x})).\end{aligned} \tag{21}$$

Because $x_{k_N}(x_0^N) \underset{N\to\infty}{\longrightarrow} \tilde{x}$, for sufficiently large N, $\|\xi - \tilde{x}\| \le 2c_1T$, by (16) the last term of (21) is $o(T)$. Then

$$\begin{aligned}&v\left(x_{m(k_N,T)+1}\left(x_0^N\right)\right) - v\left(x_{k_N}\left(x_0^N\right)\right)\\ &\quad= \sum_{i=k_N}^{m(k_N,T)} a_iy_{i+1}^\tau v_x(\tilde{x}) + (T)\\ &\quad= \sum_{i=k_N}^{m(k_N,T)} a_if^\tau\left(x_i\left(x_0^N\right)\right)v_x\left(x_i\left(x_0^N\right)\right)\\ &\qquad+ \sum_{i=k_N}^{m(k_N,T)} a_iv_x^\tau(\tilde{x})\varepsilon_{i+1}(\omega, x_i)\\ &\qquad+ \sum_{i=k_N}^{m(k_N,T)} a_if^\tau\left(x_i\left(x_0^N\right)\right)\\ &\qquad\times\left(v_x(\tilde{x}) - v_x\left(x_i\left(x_0^N\right)\right)\right) + o(T).\end{aligned} \tag{22}$$

By (16) and the continuity of $v_x(x)$, the third term on the right-hand side of (22) is $o(T)$, and by A4′), the norm of the second term on the right-hand side of (22) is also $o(T)$ as $N \to \infty$. Hence, by A3) and (19), some $\alpha > 0$ exists such that the right-hand side of (22) is less than $-\alpha T$ for all sufficient large N if T is small enough, and its left-hand side is nonnegative by (20) and the selection of k_N. The obtained contradiction shows that, for any x_0, an N exists such that $x_n(x_0) \in A_{h_2}$ for any $n \ge N$ if $x_N(x_0) \in A_{h_1}$. □

VI. Concluding Remarks

In this paper, we have given conditions to guarantee the uniqueness of the limit point of the SA algorithm. The limit, however, may be unstable, as shown by examples. Sufficient conditions are also given for stability of the limit. All of these conditions are reasonable, but may not be the weakest ones. It is of interest to consider the possibility of weakening conditions in Theorem 2 if we extend the stability notion of a point so that it has attraction domain with positive Lebesgue measure.

Appendix

A. Proof of Proposition 1

In what follows, the analysis is deterministic.

1) Let $\{x_{n_k}\}$ be a convergent subsequence of $\{x_n\}$, $x_{n_k} \to \overline{x}$. It is shown that constants $M > 0$, $\Delta > 0$ and $k_T > 0$ exist such that, for any $k > k_T$ and $T \in [0, \Delta)$

$$\left\|\sum_{i=n_k}^{m+1} a_iy_{i+1}\right\| \le M, \qquad \forall\, m: n_k \le m \le m(n_k, T). \tag{23}$$

We now prove this. If $\sigma_n = \sigma_N, \forall\, n \ge N$ for some N, then the algorithm is bounded and

$$\left\|\sum_{i=n_k}^{m+1} a_iy_{i+1}\right\| = \|x_{m+2} - x_{n_k}\| \le 2M_{\sigma_N} \quad \text{whenever } n_k \ge N.$$

Consider the case in which $\sigma_n \underset{n\to\infty}{\longrightarrow} \infty$. Assume A1) is not true. Take $c > \|\overline{x}\|$. k_c exists such that

$$\|x_{n_k}\| \le (c + \|\overline{x}\|)/2, \qquad \forall\, k \ge k_c. \tag{24}$$

Take $T_s \to 0$, $T_s > 0$. By the converse assumption for any s, k_s and $m_s : n_{k_s} \le m_s \le m(n_{k_s}, T_s)$ exist such that

$$\left\|\sum_{i=n_{k_s}}^{m_s+1} a_iy_{i+1}\right\| > (c - \|\overline{x}\|)/2. \tag{25}$$

Without loss of generality, we may assume $k_s > k_{s-1} > k_c$, $\forall\, s \ge 1$ and

$$m_s = \inf\left\{m: \left\|\sum_{i=n_{k_s}}^{m+1} a_iy_{i+1}\right\| > (c - \|\overline{x}\|)/2\right\}. \tag{26}$$

Then, for any m: $n_{k_s} \le m \le m_s$, from (24) and (26), it follows that

$$\left\|x_{n_{k_s}} + \sum_{i=n_{k_s}}^{m} a_iy_{i+1}\right\| \le c. \tag{27}$$

Because $\sigma_n \to \infty$, s_0 exists such that $M_{\sigma_{n_{k_s}}} > c\ \forall\, s \ge s_0$. Then, from (27), it follows that

$$x_{m+1} = x_m + a_my_{m+1}, \qquad \forall\, m: n_{k_s} \le m \le m_s \tag{28}$$

and by (24), (27), and (28) $\|x_m\| \le c$; hence

$$\|f(x_m)\| \le c', \qquad \forall\, m: n_{k_s} \le m \le m_{s+1}. \tag{29}$$

For any $T > 0$, we have $T_s < T$ and $m_s + 1 < m(n_{k_s}, T)$ if s is large enough. From condition A4), it follows that

$$\lim_{s\to\infty} a_{m_s+1}\epsilon_{m_s+2} = 0.$$

By (29), we then have

$$\left\|x_{m_s+1} - x_{n_{k_s}}\right\| \le \sum_{i=n_{k_s}}^{m_s} a_i\|f(x_i)\| + \left\|\sum_{i=n_{k_s}}^{m_s} a_i\epsilon_{i+1}\right\| \underset{s\to\infty}{\longrightarrow} 0$$

and

$$\begin{aligned}&\left\|x_{m_s+1} - x_{n_{k_s}} + a_{m_s+1}y_{m_s+2}\right\| \\ &\quad\le \left\|x_{m_s+1} - x_{n_{k_s}}\right\| + \|a_{m_s+1}y_{m_s+2}\| \underset{s\to\infty}{\longrightarrow} 0.\end{aligned}$$

On the other hand, by (26)

$$\begin{aligned}&\left\|x_{m_s+1} - x_{n_{k_s}} + a_{m_s+1}y_{m_s+2}\right\| \\ &\quad= \left\|\sum_{i=n_{k_s}}^{m_s+1} a_i y_{i+1}\right\| > (c - \|\overline{x}\|)/2.\end{aligned}$$

The obtained contradiction proves (23).

2) $c_1 > 0$ and k_T exist such that

$$\begin{aligned}&\|x_{m+1} - x_{n_k}\| \le c_1 T, \\ &\qquad \forall m\colon n_k \le m \le m(n_k, T),\ \forall k \ge k_T. \end{aligned} \tag{30}$$

We now prove (30).

In the case in which $\lim_{n\to\infty}\sigma_n = \infty$, k_0 and $c_2 > 0$ exist such that $M_{\sigma_{n_k}} > M + 1 + \|\overline{x}\|$, $\|x_{n_k}\| \le 1 + \|\overline{x}\|$, $\forall k \ge k_0$. By (23), we have

$$\begin{aligned}&\left\|x_{n_k} + \sum_{i=n_k}^{m} a_i y_{i+1}\right\| \le M + 1 + \|\overline{x}\| \le M_{\sigma_{n_k}}, \\ &\qquad\qquad \forall T \in (0, \Delta),\ \forall m\colon n_k \le m \le m(n_k, T),\end{aligned}$$

if $k \ge k_T$. Consequently

$$\begin{aligned}&x_{m+1} = x_m + a_m y_{m+1} \\ &\|x_{m+1}\| \le M + 1 + \|\overline{x}\| \\ &\|f(x_m)\| \le c_3, \qquad \forall m\colon n_k \le m \le m(n_k, T)\end{aligned} \tag{31}$$

and, hence

$$\left\|\sum_{i=n_k}^{m} a_i f(x_i)\right\| \le c_3 \sum_{i=n_k}^{m} a_i \le c_3 T. \tag{32}$$

This process with Condition A4) leads to (30).

If $\{\sigma_n\}$ is bounded, then $\{x_n\}$ is bounded and (32) is still valid. Hence, we also have (30).

3) i) For any interval $[\delta_1, \delta_2]$ with $\delta_1 < \delta_2$ and $d([\delta_1, \delta_2], v(J)) > 0$, the sequence $\{v(x_n)\}$ cannot cross $[\delta_1, \delta_2]$ infinitely many times with $\|x_{n_k}\|$ bounded, where by "crossing $[\delta_1, \delta_2]$ by $v(x_{n_k}), \cdots, v(x_{m_k})$" we mean that $v(x_{n_k}) \le \delta_1, v(x_{m_k}) \ge \delta_2$, and $\delta_1 < v(x_i) < \delta_2, \forall i$: $n_k < i < m_k$.

ii) If $v(x_n) \underset{n\to\infty}{\longrightarrow} \delta_1$ with $d(\delta_1, v(J)) > 0$, then no convergent subsequence of $\{x_n\}$ occurs.

We first prove i). Assume the converse: infinitely many crossings $v(x_{n_k}), \cdots, v(x_{m_k})$, $k = 1, 2, \cdots$, occur and $\{\|x_{n_k}\|\}$ is bounded. Without loss of generality, we may assume $x_{n_k} \underset{k\to\infty}{\longrightarrow} \overline{x}$. Because $x_{n_k+1} - x_{n_k} = a_{n_k}(f(x_{n_k}) + \epsilon_{n_k+1}) \to 0$ and $v(x_{n_k+1}) > \delta_1 \ge v(x_{n_k})$, we have $v(x_{n_k}) \to \delta_1 = v(\overline{x})$ and $d(\overline{x}, J) \triangleq \delta > 0$. By (30), we may take T small enough such that

$$d(x_m, J) \ge \frac{\delta}{2}, \qquad \forall m\colon n_k \le m \le m(n_k, T) \tag{33}$$

for all sufficiently large k. Then, a ξ exists with $\|\xi - x_{n_k}\| \le c_1 T$ such that

$$\begin{aligned}&v\left(x_{m(n_k,T)+1}\right) - v\left(x_{n_k}\right) \\ &\quad= \left(x_{m(n_k,T)+1} - x_{n_k}\right)^{\tau} v_x(\overline{x}) \\ &\qquad+ \left(x_{m(n_k,T)+1} - x_{n_k}\right)^{\tau} (v_x(\xi) - v_x(\overline{x})).\end{aligned} \tag{34}$$

By (30), (31), and the continuity of $v_x(\cdot)$, the last term of (34) is of $o(T)$ and

$$\begin{aligned}&v\left(x_{m(n_k,T)+1}\right) - v\left(x_{n_k}\right) \\ &\quad= \sum_{i=n_k}^{m(n_k,T)} a_i y_{i+1}^{\tau} v_x(\overline{x}) + o(T) \\ &\quad= \sum_{i=n_k}^{m(n_k,T)} a_i f^{\tau}(x_i) v_x(x_i) \\ &\qquad+ \sum_{i=n_k}^{m(n_k,T)} a_i f^{\tau}(x_i)(v_x(\overline{x}) - v_x(x_i)) \\ &\qquad+ \sum_{i=n_k}^{m(n_k,T)} a_i v_x^{\tau}(\overline{x})\epsilon_{i+1} + o(T).\end{aligned} \tag{35}$$

By Conditions A3) and A4), from this it follows that $\alpha > 0$ and $T > 0$ exist such that

$$v\left(x_{m(n_k,T)+1}\right) - v(x_{n_k}) \le -\alpha T \tag{36}$$

for sufficiently large k. This process implies that

$$\limsup_{k\to\infty} v\left(x_{m(n_k,T)+1}\right) \le \delta_1 - \alpha T.$$

By (30), however, $\max_{n_k\le m\le m(n_k,T)} |v(x_m) - v(x_{n_k})| \underset{T\to 0}{\longrightarrow} 0$, which implies $m(n_k, T) + 1 < m_k$ for small enough T. This result means that $v(x_{m(n_k,T)+1}) \in [\delta_1, \delta_2)$, which contradicts (36).

For ii), let us assume the converse: a convergent subsequence $x_{n_k} \underset{k\to\infty}{\longrightarrow} \overline{x}$ occurs. By the same argument, we arrive at (36). By assumption, however, the left-hand-side of (36) tends to zero, which leads to a contradiction.

4) We now show $d(x_k, J) \underset{k\to\infty}{\longrightarrow} 0$.

By assumptions of Proposition 1, a nonempty interval $[\delta_1, \delta_2]$ exists such that $[\delta_1, \delta_2] \subset (v(x^*), \inf_{\|x\|=c_0} v(x))$ and $d([\delta_1, \delta_2], v(J)) > 0$. If $\sigma_n \to \infty$, then x_n, starting from x^*, will cross the sphere $\{x : \|x\| = c_0\}$ infinitely many times and, hence, $v(x_n)$ will cross $[\delta_1, \delta_2]$ infinitely often with $\{x_{n_k}\}$ bounded. We have shown this process is impossible. Therefore, starting from some n_0, the algorithm (4)–(7) will have no truncations and $\{x_n\}$ is bounded.

Let $v_1 \triangleq \liminf_{n\to\infty} v(x_n) \le \limsup_{n\to\infty} v(x_n) \triangleq v_2$. If $v_1 = v_2$, then $v_1 \in v(J)$ because otherwise a contradiction with 3) ii) would occur. If $v_1 < v_2$ and one of v_1 and v_2 does not belong to $v(J)$, then $[\delta_1, \delta_2] \subset (v_1, v_2)$ exists such that $d([\delta_1, \delta_2], v(J)) > 0$ and $\delta_2 > \delta_1$. By 3) i) this process is impossible. So, both v_1 and v_2 belong to $v(J)$

and $\lim_{n\to\infty} d(v(x_n), v(J)) = 0$. Assume $v_1 \neq v_2$, because $x_{k+1} - x_k = a_k y_{k+1} \underset{k\to\infty}{\longrightarrow} 0, \{v(x_k)\}$ is dense in $[v_1, v_2]$. Hence, $v(J)$ is dense in $[v_1, v_2]$, which contradicts to the assumption. Hence, $v(x_n)$ converges.

For proving $d(x_n, J) \underset{n\to\infty}{\longrightarrow} 0$, it suffices to show that any limit points of $\{x_n\}$ belong to J. Assume the converse: $x_{n_k} \to x^0 \notin J$, $d(x^0, J) \triangleq \delta > 0$. By (30), we have

$$d(x_m, J) \geq \frac{\delta}{2}, \qquad \forall\, m\colon n_k \leq m \leq m(n_k, T)$$

for all large k if T is small enough. By (8), we then have

$$v_x^\tau(x_m) f(x_m) < -b < 0, \qquad \forall\, m\colon n_k \leq m \leq m(n_k, T)$$

and from (35) $v(x_{m(n_k,T)+1}) - v(x_{n_k}) \leq -bT/2$ for small enough T. This process leads to a contradiction because $v(x_n)$ converges. The proof of Proposition 1 is completed.

References

[1] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[2] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551–575, 1977.

[3] H. F. Chen and X. R. Cao, "Controllability is not necessary for adaptive pole placement control," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1222–1297, 1997.

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

[5] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*. Berlin: Springer-Verlag, 1997.

[6] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Berlin: Springer-Verlag, 1978.

[7] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer-Verlag, 1990.

[8] H. F. Chen, "Stochastic approximation and its new applications," in *Proc. 1994 Hong Kong Int. Workshop New Directions Contr. Manufact.*, 1994, pp. 2–12.

[9] ——, "Recent developments in stochastic approximation," in *Preprints 13th IFAC World Congr.*, vol. D, San Francisco, CA, 1996, pp. 375–380.

[10] H. F. Chen and Y. M. Zhu, "Stochastic approximation procedures with randomly varying truncations," *Scientia Sinica*, ser. A, vol. 29, pp. 914–926, 1986.

[11] G. J. Wang and H. F. Chen, "Behavior of the stochastic approximation algorithm in the root set of regression function," *Proc. Chinese Contr. Conf.*, pp. 1018–1021, 1996.

[12] M. Benaim, "A dynamical system approach to stochastic approximations," *SIAM J. Contr. Optim.*, vol. 34, pp. 437–472, 1996.

[13] V. S. Borkar and S. P. Meyn, "Stability and convergence of stochastic approximation using ODE method,", Tech. Rep..

[14] I. J. Wang, E. K. P. Chong, and S. R. Kulkarni, "Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms," *Adv. Appl. Probab.*, vol. 23, pp. 784–801, 1996.

[15] B. Delyon, "General results on the convergence of stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1245–1255, 1996.

[16] R. Pemantle, "Non convergence to unstable points in urn models and stochastic approximations," *Ann. Probab.*, vol. 18, pp. 689–712, 1990.

[17] K. Deimling, *Nonlinear Functional Analysis*. Berlin: Springer-Verlag, 1985.

Hai-Tao Fang received the B.S. degree in probability and statistics in 1990, M.S. degree in applied mathematics in 1993 and the Ph.D. degree in 1996, respectively, from the Peking University, Tsinghua University and Peking University. He now is with the Laboratory of Systems and Control, Institute of Systems Science, the Chinese Academy of Sciences as an Associate Professor.

From 1996–1998, he was a Postdoc at the Institute of Systems Science and joined the Institute as an Assistant Professor in 1998. During 1998–1999, he was with the Hong Kong University of Science and Technology as a Research Associate. His current research interests include stochastic optimization and systems control.

Han-Fu Chen (SM'94–F'97) received the Diploma from the Department of Mathematics and Mechanics, Leningrad (St. Petersburg) University, Russia.

He joined the Institute of Mathematics, Chinese Academy of Sciences, in 1961. Since 1979, he has been with the Institute of Systems Science, Chinese Academy of Sciences, where he is a Professor now. His research interests include stochastic systems, system indentification, adaptive control, recursive estimation and stochastic approximation and its applications. He has authored and co-authored more than 130 journal papers.

He is the Editor of both "Control Theory an Applications" and "Systems Science and mathematical Sciences" and is involved in the editorial boards of several international and domestic journals. He was elected to an Academician of the Chinese Academy of Sciences in 1993. He now serves as the President of the Chinese Association of Automation and serves also on the Technical Board of IFAC.