TECHNICAL
RESEARCH
REPORT

*Institute for
Systems
Research*

On The Convergence and ODE Limit of A Two-Dimensional Stochastic Approximation

by D-J. Ma and A.M. Makowski

*The Institute for Systems
Research is supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University*

TR 93-5

ON THE CONVERGENCE AND ODE LIMIT OF A TWO–DIMENSIONAL STOCHASTIC APPROXIMATION

Dye–Jyun Ma[1] and Armand M. Makowski[2]
DEC and University of Maryland at College Park.

ABSTRACT

We consider a two–dimensional stochastic approximations scheme of the Robbins–Monro type which naturally arises in the study of steering policies for Markov decision processes [6,7]. Making use of a decoupling change of variable, we establish almost sure convergence by ad–hoc arguments that combine standard results on one–dimensional stochastic approximations with a version of the law of large number for martingale differences. Coming full circle, this direct analysis gives clues on how to select the test function which appears in standard convergence results for multi–dimensional schemes. Furthermore, a blind application of the ODE method is not possible here as solutions to the limiting ODE cannot be defined in an elementary way, but the aforementionned change of variasble paves the way for an interpretation of the behavior of solutions to the limiting ODE.

[1] Digital Equipment Corporation, 305 Foster Street, LTN2-2/F12 Littleton, MA 01450. The work of this author was supported partially through NSF Grants ECS-83-51836 and NSFD CDR-85-00108.

[2] Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742. The work of this author was supported partially through NSF Grant NSFD CDR-88-03012 and a grant from AT&T Bell Laboratories.

1. INTRODUCTION

In [8] we introduced a two–dimensional stochastic approximation scheme of the Robbins–Monro type for the purpose of investigating the performance of steering policies in Markov decision processes with a recurrence structure [6,7]. We thought then that the a.s. convergence properties of this scheme were not covered by standard convergence results available in the literature, and we proceeded to analyze it by ad–hoc arguments. In fact, it turns out that conditions due to Blum [1] could have been used to yield the a.s. convergence of the scheme of interest. However, the main difficulty with Blum's approach lies in finding a suitable test function that meets all the requisite conditions, some of them quite stringent. Although we shall eventually exhibit such a test function, it will arise as an easy byproduct of a direct convergence analysis, a situation somewhat similar to that encountered in studying the stability of ODEs by Lyapounov methods. This is perhaps not too surprising in view of the fact that the convergence of stochastic approximations is often investigated by means of an associated ODE, a viewpoint discussed later in this paper.

In this short note, we provide direct convergence arguments which constitute a refinement of those given in [8]. We do so by using an invertible linear change of variables which decouples the original scheme into two one–dimensional stochastic approximations. Their a.s. convergence readily follows from classical results for one–dimensional stochastic approximations [2,10] and from the Stability Theorem for martingale differences [5]. This approach provides a straightforward proof of the a.s. convergence for the original scheme under conditions weaker than those required by [1,10]. It is also of independent interest in that it helps illustrate the interplay between probabilistic and analytic viewpoints in the study of stochastic approximations. First of all, the change of variables provides clues as to how a test function should be selected in order to meet Blum's conditions. Next, the convergence analysis paves the way for a natural interpretation of solutions to the limiting ODE. Indeed, the ODE method cannot be used here (at least in its standard form) [4,9]; this is due to the lack of requisite regularity properties which prevent the limiting ODE to be defined in an elementary way.

The paper is organized as follows: In Section 2, we introduce the two–dimensional stochastic approximation of interest, and give a direct analysis of its convergence properties in Section 3. In Section 4, we discuss the behavior of solutions to the ODE limit, and display the test function that meets Blum's conditions [1].

A word on the notation: The indicator function of any set E is simply denoted by $\mathbf{1}[E]$, and unless otherwise stated, $\lim_n$ is taken with n going to infinity.

2. THE STOCHASTIC APPROXIMATION SCHEME

Consider a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, equipped with a filtration $\{\mathcal{H}_k,\ k = 0, 1, \ldots\}$, which carries an $\mathcal{H}_k$–adapted sequence of $I\!R^2$–valued rvs $\{(\zeta(k), \tau(k)),\ k = 1, 2, \ldots\}$. The two–dimensional stochastic approximation scheme of interest here has output values $\{(Z(k), T(k)),\ k = 1, 2, \ldots\}$ which are recursively generated by

$$\begin{bmatrix} Z(k+1) \\ T(k+1) \end{bmatrix} = \begin{bmatrix} Z(k) \\ T(k) \end{bmatrix} + \frac{1}{k+1} \begin{bmatrix} \zeta(k+1) - Z(k) \\ \tau(k+1) - T(k) \end{bmatrix} \qquad k = 1, 2, \ldots \tag{2.1}$$

with initial condition $(Z(1), T(1)) = (\zeta(1), \tau(1))$. Throughout the discussion we enforce conditions **(H1)–(H3)** which are now presented:

(H1) There exist a scalar V and a pair of probability distributions $\overline{F}$ and $\underline{F}$ on $I\!R^2$ such that

$$\begin{aligned} &\mathbf{P}[\zeta(k+1) \leq z, \tau(k+1) \leq t | \mathcal{H}_k] \\ =&\mathbf{1}[Z(k) \leq VT(k)]\, \overline{F}(z,t) + \mathbf{1}[Z(k) > VT(k)]\, \underline{F}(z,t), \quad (z,t) \in I\!R^2 \end{aligned}$$

for all $k = 1, 2, \ldots$;

(H2) The probability distributions $\overline{F}$ and $\underline{F}$ on $I\!R^2$ are square–integrable.

It will be convenient to denote by F any one of the probability distributions $\overline{F}$ and $\underline{F}$, with a similar convention for quantities derived from these distributions. With this convention in mind, **(H2)** reads

$$\int_{I\!R^2} (z^2 + t^2) dF(z,t) < \infty. \tag{2.2}$$

We also define the first moments $m(\zeta)$ and $m(\tau)$ by

$$m(\zeta) := \int_{I\!R^2} z dF(z,t) \quad \text{and} \quad m(\tau) := \int_{I\!R^2} t dF(z,t). \tag{2.3}$$

(H3) These quantities, which are well defined and finite by virtue of **(H2)**, satisfy the inequalities

$$\underline{m}(\zeta) - V\,\underline{m}(\tau) < 0 < \overline{m}(\zeta) - V\,\overline{m}(\tau). \tag{2.4}$$

Under the assumptions **(H1)–(H3)** the recursive scheme (2.1) is a stochastic approximation of the Robbins–Monro type [4], and we are interested in establishing its a.s. convergence. Standard a.s. convergence results [1,10] for multi–dimensional schemes rely on selecting a test or Lyapounov function which satisfies a set of sometimes stringent conditions. Later on, in Section 4, we shall exhibit such a test function. It is perhaps amusing to note that its discovery passed through a direct proof of the a.s. convergence of the scheme (2.1), the very fact it was supposed to help establish.

3. A CONVERGENCE PROOF

The basis for our arguments is a simple one–to–one change of variables which maps (2.1) into a two–dimensional stochastic approximation scheme which can analyzed by standard techniques: With scalars a and b still to be determined, we define the $I\!R^2$–valued rvs $\{(U(k), V(k)),\ k = 1, 2, \ldots\}$ by the linear transformation

$$\begin{bmatrix} U(k) \\ V(k) \end{bmatrix} = \begin{bmatrix} 1 & -V \\ a & b \end{bmatrix} \begin{bmatrix} Z(k) \\ T(k) \end{bmatrix}. \qquad k = 1, 2, \ldots \quad (3.1)$$

These rvs also obey recursions similar to (2.1), namely

$$U(1) = \nu(1), \quad U(k+1) = U(k) + \frac{1}{k+1}[\nu(k+1) - U(k)] \qquad k = 1, 2, \ldots \quad (3.2)$$

$$V(1) = \varepsilon(1), \quad V(k+1) = V(k) + \frac{1}{k+1}[\varepsilon(k+1) - V(k)] \qquad k = 1, 2, \ldots \quad (3.3)$$

where the $I\!R^2$–valued rvs $\{(\nu(k), \varepsilon(k)),\ k = 1, 2, \ldots\}$ are given by

$$\nu(k) := \zeta(k) - V\tau(k) \quad \text{and} \quad \varepsilon(k) := a\zeta(k) + b\tau(k). \qquad k = 1, 2, \ldots \quad (3.4)$$

Let G denote the probability distribution on $I\!R^2$ induced from F by the transformation $(z, t) \to (z - Vt, az + bt)$ appearing in (3.1). We readily check from **(H2)** that

$$\begin{aligned} &\mathbf{P}[\nu(k+1) \leq u, \varepsilon(k+1) \leq v | \mathcal{H}_k] \\ =&\mathbf{1}[U(k) \leq 0]\, \overline{G}(u, v) + \mathbf{1}[U(k) > 0]\, \underline{G}(u, v), \quad (u, v) \in I\!R^2 \end{aligned} \quad (3.5)$$

for all $k = 1, 2, \ldots$, and the combined recursion (3.2)–(3.3) is thus also a stochastic approximation of the Robbins–Monro type (since the rvs $\nu(k)$ and $\varepsilon(k)$ are both $\mathcal{H}_k$–measurable for all $k = 1, 2, \ldots$). We find it convenient to introduce the notation

$$m(\nu) := \int_{I\!R^2} u dG(u, v) = m(\zeta) - Vm(\tau) \quad (3.6)$$

and

$$m(\varepsilon) := \int_{I\!R^2} v dG(u, v) = am(\zeta) + bm(\tau). \quad (3.7)$$

We first show the a.s. convergence of (3.2): Inspection of (3.2) and (3.5) reveals that the recursion (3.2) by itself is a stochastic approximation scheme of the Robbins–Monro type such that

$$\mathbf{E}[\nu(k+1)|\mathcal{H}_k] = \underline{m}(\nu) + \mathbf{1}[U(k) \le 0]\Delta m(\nu) \qquad k = 1, 2, \ldots \tag{3.8}$$

with $\Delta m(\nu) := \overline{m}(\nu) - \underline{m}(\nu)$. Combining (3.8) with (2.4), a condition equivalent to $\underline{m}(\nu) < 0 < \overline{m}(\nu)$, we can now invoke standard convergence results [2, Theorem 1, p. 275] [9] to get the following.

Lemma 3.1. *Under* **(H1)–(H3)**, *the convergence* $\lim_k\ U(k) = 0$ *takes place a.s.*

The a.s. convergence properties of (3.3) are considered next:

Lemma 3.2. *Assume* **(H1)–(H3)**. *For any pair of scalars a and b such that*

$$\underline{m}(\varepsilon) = \overline{m}(\varepsilon), \tag{3.9}$$

the convergence $\lim_k\ V(k) = c$ *takes place a.s. with c denoting the common value in (3.9).*

Proof. First we note that there always exists a pair a and b such that (3.9) holds. Next, for such a pair, we observe from (3.5) and (3.7) that

$$\mathbf{E}[\varepsilon(k+1)|\mathcal{H}_k] = \mathbf{1}[U(k) \le 0]\,\overline{m}(\varepsilon) + \mathbf{1}[U(k) > 0]\,\underline{m}(\varepsilon) = c, \qquad k = 1, 2, \ldots \tag{3.10}$$

and the $\mathcal{H}_k$–adapted rvs $\{\varepsilon(k) - c,\ k = 1, 2, \ldots\}$ thus form a zero–mean $(\mathbf{P}, \mathcal{H}_k)$–martingale difference sequence with bounded second moments; in fact, we have $\sup_k \mathbf{E}[|\varepsilon(k) - c|^2] < \infty$ as a result of **(H1)–(H2)**. Invoking the Stability Theorem for martingale difference [5], we get

$$\lim_k \frac{1}{k}\sum_{i=1}^{k}(\varepsilon(i) - c) = 0 \quad a.s. \tag{3.11}$$

and the conclusion follows from the fact that the output to (3.3) can also be expressed as

$$V(k) = c + \frac{1}{k}\sum_{i=1}^{k}(\varepsilon(i) - c). \qquad k = 1, 2, \ldots \tag{3.12}$$

■

We are now ready to state and give a proof of the main convergence result of this note. To simplify the notation, we write

$$\Delta m(\zeta) := \overline{m}(\zeta) - \underline{m}(\zeta) \quad \text{and} \quad \Delta m(\tau) := \overline{m}(\tau) - \underline{m}(\tau). \tag{3.13}$$

Theorem 3.3. *Under* **(H1)–(H3)**, *we have the a.s. convergence results*

$$\lim_k Z(k) = \underline{m}(\zeta) + r^* \Delta m(\zeta) \tag{3.14}$$

and

$$\lim_k T(k) = \underline{m}(\tau) + r^* \Delta m(\tau) \tag{3.15}$$

with

$$r^* = -\frac{\underline{m}(\nu)}{\Delta m(\nu)}. \tag{3.16}$$

Proof. By Lemmas 3.1 and 3.2, we already know that the rvs $\{(U(k), V(k)),\ k = 1, 2, \ldots\}$ converge a.s. to the vector $(0, c)$. The a.s. convergence of the sequence $\{(Z(k), T(k)),\ k = 1, 2, \ldots\}$ would then follow if we could select the scalars a and b so that the linear transformation $(z, t) \to (z - Vt, az + bt)$ is invertible *and* the moment condition (3.9) holds. The first requirement is equivalent to the determinant condition $b + aV \neq 0$, while the second reduces to

$$a\Delta m(\zeta) = -b\Delta m(\tau). \tag{3.17}$$

If $\Delta m(\tau) \neq 0$, we choose $a = 1$ and $b = -\frac{\Delta m(\zeta)}{\Delta m(\tau)}$, with the determinant condition satisfied by virtue of (2.4). If $\Delta m(\tau) = 0$, then $\Delta m(\zeta) \neq 0$ owing to (2.4), whence $a = 0$ by (3.17) and the determinant condition reduces to $b \neq 0$; the choice $a = 0$ and $b = 1$ does the job! The limiting values in (3.14)–(3.16) are now readily obtained by solving (in (z, t)) the system of linear equations $z - Vt = 0$ and $az + bt = c$. ∎

Combining Lemma 3.1 and the convergence results (3.14)–(3.16), we get the equality

$$\underline{m}(\zeta) + r^* \Delta m(\zeta) = V\left(\underline{m}(\tau) + r^* \Delta m(\tau)\right) \tag{3.18}$$

which was found useful in [8]. Moreover, inspection of the proof of Theorem 3.3 reveals that this convergence result actually holds under slightly weaker conditions than the square–integrability condition **(H2)** used here. In fact, the entire analysis can still be carried out under the conditions

$$\int_{\mathbb{R}^2} (z - Vt)^2 dF(z, t) < \infty \quad \text{and} \quad \int_{\mathbb{R}^2} (az + bt)^{1+p} dF(z, t) < \infty \tag{3.19}$$

for some $p > 0$. The first condition is sufficient for obtaining Lemma 3.1 while the Stability Theorem for martingale difference in the form (3.11) will hold under the second part of (3.19) [3].

4. THE ODE LIMIT AND BLUM'S CONDITIONS

We complete the discussion of the convergence of the scheme (2.1) by considering its ODE limit [4,9]: We see that, were it to exist, the limiting ODE should be of the form

$$\begin{aligned}\dot{Z}(t) &= -Z(t) + \mathbf{1}[Z(t) \leq VT(t)]\,\overline{m}(\zeta) + \mathbf{1}[Z(t) > VT(t)]\,\underline{m}(\zeta)\\ \dot{T}(t) &= -T(t) + \mathbf{1}[Z(t) \leq VT(t)]\,\overline{m}(\tau) + \mathbf{1}[Z(t) > VT(t)]\,\underline{m}(\tau)\end{aligned} \qquad t \geq 0 \qquad (4.1)$$

with given initial condition $(Z(0), T(0))$. The right handside of (4.1) is not Lipschitz continuous on *all* of $I\!R^2$ since discontinuous on the one–dimensional manifold $M = \{(z,t) \in I\!R^2 : z = Vt\}$, and therefore a global solution to (4.1) cannot be guaranteed by elementary results on the existence and uniqueness of solutions to ODEs. However, here we expect the following scenario to unfold: Any solution to (4.1) with initial condition not in M will eventually hit M in finite time, and thereafter the trajectory will chatter along M while drifting alongside M towards an equilibrium point on it.

Some care clearly needs to be exercised in formally defining a solution to (4.1), with the net result that it does not seem possible to use the ODE method in its standard form [4,9]. We now show that the transformation (3.1), with (3.9), does provide a natural way for indirectly apprehending the definition of a solution to (4.1), and its asymptotic behavior. Indeed, the ODE limit for the scheme (3.2)–(3.3) should have the form

$$\dot{U}(t) = -U(t) + \mathbf{1}[U(t) \leq 0]\,\overline{m}(\nu) + \mathbf{1}[U(t) > 0]\,\underline{m}(\nu) \qquad (4.2a)$$

$$\dot{V}(t) = -V(t) + c \qquad (4.2b)$$

with given initial condition $(U(0), V(0))$. Note that (4.2) could also be interpreted as the ODE formally obtained from (4.1) by the change of variables $(u, v) = (z - Vt, az + bt)$ appearing in the proof of Theorem 3.3. The structure of solutions to (4.2) is now much more transparent because this linear transformation yields a *decoupled* system composed of two independent one–dimensional ODEs; this decoupling was already apparent in the probabilistic analysis of Section 3. For every initial condition $(U(0), V(0))$ with $U(0) \neq 0$, we can readily construct a solution for (4.2): First, the solution to (4.2b) is simply $V(t) = c + (V(0) - c)e^{-t}$, $t \geq 0$; it is defined for all times with $\lim_t V(t) = c$. On the other hand, care is still required when defining a solution trajectory to (4.2a) since the right handside of this ODE exhibits a discontinuity at the origin. This single point of discontinuity should however be contrasted against the one–dimensional discontinuity manifold M associated with (4.1), so that we should now expect a simpler description of the trajectory's behavior.

We see, at least initially, that

$$U(t) = \begin{cases} \overline{m}(\nu) + (U(0) - \overline{m}(\nu))e^{-t} & \text{if } U(0) < 0 \\ \\ \underline{m}(\nu) + (U(0) - \underline{m}(\nu))e^{-t} & \text{if } U(0) > 0 \end{cases} \tag{4.3}$$

until hitting the origin, an event which occurs in finite time. Thereafter, the solution will dance around in a vanishingly small neighborhood of the origin for the rest of times. This description leads to $(0, c)$ as the asymptotically stable point of the ODE (4.2), in agreement of course with the earlier probabilistic analysis.

The behavior near the origin of solution trajectories to (4.2a) is perhaps best understood through the corresponding *local time* process $\{R(k),\ k = 1, 2, \ldots\}$ associated with the recursion (3.2), where we define

$$R(k) := \frac{1}{k} \sum_{i=1}^{k} \mathbf{1}[U(i) \leq 0]. \qquad k = 1, 2, \ldots \tag{4.4}$$

In [8] we combined Lemma 3.1 with the Stability Theorem for martingale differences in order to get the following result on the behavior of (3.2) near the origin.

Theorem 4.1. *Under assumptions* **(H1)–(H3)**, *we have* $\lim_k R(k) = r^*$ *a.s., with* r^* *given by (3.16).*

In a sense this result provides an indication of the long–term frequency of finding the solution trajectory in either half–lines $(-\infty, 0)$ or $(0, \infty)$.

Finally, coming full circle, it is worth pointing out that the underlying linear transformation $(z, t) \to (z - Vt, az + bt)$ also provides the clue for choosing the Lyapounov function behind Blum's conditions [1]. in doing so, there is no loss of generality in assuming $c = 0$; this is already apparent in the proof of Lemma 3.2. Here, in the notation of [1], we have

$$M(z,t) := -\begin{bmatrix} z \\ t \end{bmatrix} + \mathbf{1}[z \leq Vt] \begin{bmatrix} \overline{m}(\zeta) \\ \overline{m}(\tau) \end{bmatrix} + \mathbf{1}[z > Vt] \begin{bmatrix} \underline{m}(\zeta) \\ \underline{m}(\tau) \end{bmatrix}, \quad (z,t) \in I\!R^2. \tag{4.5}$$

Motivated by the discussion given above, we choose the test function $f : I\!R^2 \to I\!R$ to be

$$f(z,t) = \frac{1}{2}(z - Vt)^2 + \frac{1}{2}(az + bt)^2, \quad (z,t) \in I\!R^2. \tag{4.6}$$

It is easy to see with this choice that all of Blum's conditions, except for his condition **(A3)**, are trivially satisfied. To show that condition **(A3)** of Blum also holds, we note that the gradient function D of f is given by

$$D(z,t) = (z - Vt + a(az + bt), -V(z - Vt) + b(az + bt)), \quad (z,t) \in I\!R^2 \tag{4.7}$$

so that after some uninteresting calculations, we get

$$
\begin{aligned}
U(z,t) &= < D(z,t), M(z,t) > \\
&= -2f(z,t) + \left(\mathbf{1}[z \leq Vt]\,\overline{m}(\nu) + \mathbf{1}[z > Vt]\,\underline{m}(\nu)\right)(z - Vt)
\end{aligned}
\tag{4.8}
$$

for all (z,t) in $I\!R^2$. In view of (2.4) we can conclude that

$$
U(z,t) \leq -2(z - Vt)^2 - 2(az + bt)^2, \quad (z,t) \in I\!R^2 \tag{4.9}
$$

whence for every $\varepsilon > 0$,

$$
\sup\{U(z,t) : \; z^2 + t^2 \geq \varepsilon^2\} < 0 \tag{4.10}
$$

with the stricy inequality following from the invertibility of the linear transformation $(z,t) \to (z - Vt, az + bt)$, and condition **(A3)** is satisfied.

5. ACKNOWLEDGEMENTS

The second author thanks Alan Weiss for some insightful remarks on the behavior of the ODE limit (4.1).

6. REFERENCES

[1] J.R. Blum, "Multidimensional stochastic approximation methods," *Annals of Mathematical Statistics* **25** (1954), pp. 737–744.

[2] E. G. Gladyshev, "On stochastic approximation," *Theo. Prob. Appl.* **10** (1965), pp. 275–278.

[3] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Applications*, J. Wiley & Sons, New York (NY) (1980).

[4] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Applied Mathematical Sciences **26**, Springer–Verlag, Berlin (1978).

[5] M. Loéve, *Probability Theory I*, Fourth Edition, Springer–Verlag, Berlin (1977).

[6] D.-J. Ma, *A Simple Problem of Flow Control: Optimality and Adaptive Implementations*, Ph.D. Thesis, Electrical Engineering Department, University of Maryland, College Park (MD) (1988).

[7] D.-J. Ma and A. M. Makowski, "A class of steering policies under a recurrence condition," in Proceedings of the 27th IEEE Conference on Decision and Control, Austin (TX), December 1988, pp. 1192–1197.