Indian J. Pure Appl. Math., **47**(2): 291-299, June 2016

DOI: 10.1007/s13226-016-0188-1

EVENT-DRIVEN STOCHASTIC APPROXIMATION[1]

Vivek S. Borkar, Neeraja Sahasrabudhe and M. Ashok Vardhan[2]

Department of Electrical Engineering, Indian Institute of Technology,
Powai, Mumbai 400 076, *India*
e-mails: {*borkar.vs, neeraja.sigma, ashokevardhan*}*@gmail.com*

(*Received* 8 *June* 2015; *accepted* 26 *September* 2015)

We consider a Robbins-Monro type iteration wherein noisy measurements are event-driven and therefore arrive asynchronously. We propose a modification of step-sizes that ensures desired asymptotic behaviour regardless of this aspect. This generalizes earlier results on asynchronous stochastic approximation wherein the asynchronous behaviour is across different components, but not along the same component of the vector iteration, as is the case considered here.

Key words : Stochastic approximation; asynchronous computation; distributed algorithms; o.d.e. limit; sampling bias

1. Introduction

The classical Robbins-Monro stochastic approximation scheme [11] finds zeros of a nonlinear function $h : \mathcal{R}^d \mapsto \mathcal{R}^d$ given its noisy observations. It is given by the d-dimensional iteration

$$x(n+1) = x(n) + a(n)\left[h(x(n)) + M(n+1)\right], \ n \geq 0, \tag{1}$$

where $\{a(n)\}$ is a positive stepsize sequence satisfying

$$\sum_n a(n) = \infty, \ \sum_n a(n)^2 < \infty. \tag{2}$$

[1]Research of VSB supported in part by a J. C. Bose Fellowship and a grant for 'Approximation of High Dimensional Optimization and Control Problems' from Department of Science and Technology, Government of India.

[2]Ashok Vardhan is now with the Department of Electrical Engineering, University of Illinois at Urbana-Champaign, USA.

While the original analysis of [11] and others used probabilistic tools such as '*almost supermartingales*', an alternative approach due to [7, 10] that has gained currency subsequently treats (1) as a noisy Euler scheme for the ordinary differential equation ('o.d.e.' for short)

$$\dot{x}(t) = h(x(t)),\ t \geq 0. \tag{3}$$

Thus $\{a(n)\}$ are interpreted as slowly decreasing time steps that cover the entire time axis (because of the first condition in (2)), while going to zero at an appropriate rate (because of the second condition in (2)) so as to suppress the errors due to discretization and noise. One can then show that under reasonable technical conditions, the iterates (1) track the asymptotic behaviour of (3) with probability one. See [4] for a detailed treatment of this approach.

In engineering applications, one often has a situation where each component of (1) is computed by a possibly different processor. These processors have their own clocks and communicate with each other with possible delays. In particular, not all components may be updated at each time. Let $Y_n \subset \{1, 2, \cdots, d\}$ denote the set of components updated at time n. For $1 \leq i \leq d, n \geq 0$, define

$$\begin{aligned} \nu(i,n) &= \sum_{m=0}^{n} I\{i \in Y_m\}, \\ \text{for } I\{\cdots\} &= 1 \text{ if '} \cdots \text{' is true,} \\ &= 0 \text{ if not.} \end{aligned} \tag{4}$$

This is the 'local clock' at processor i that counts the number of iterates performed by i till time n. Then i knows $\nu(i,n)$, but need not know the 'global clock' n. In fact the global clock may be a complete artifice, not a constant multiple of a given unit as in a physical clock, as long as causal relationships are respected. Then the ith component iterate may be written as

$$\begin{aligned} x_i(n+1) &= x_i(n) + a(\nu(i,n))I\{i \in Y_n\}\Big[h_i(x_1(n-\tau_{1i}(n)), \cdots, \\ &\quad x_d(n-\tau_{di}(n))) + M_i(n+1)\Big] \end{aligned} \tag{5}$$

Here $\tau_{ij}(n)$ is the random delay with which i's output was received by j at time n. That is, at time n, j has access to $x_i(n-\tau_{ij}(n))$ but not $x_i(m)$ for $m > n - \tau_{ij}(n)$. Additional conditions imposed on $\{a(n)\}$ in [3] were:

- **(A1)** $a(n+1) \leq a(n)$ from some n on.
- **(A2)** There exists $r \in (0,1)$ such that $\sum_n a(n)^{1+q} < \infty\ \forall\ q \geq r$.

- **(A3)** For $\alpha \in (0,1)$, $\sup_n \frac{a(\lceil \alpha n \rceil)}{a(n)} < \infty$.
- **(A4)** For $x \in (0,1)$ and $A(n) := \sum_{m=0}^{n} a(m)$, $\frac{A(\lceil yn \rceil)}{A(n)} \to 1$ uniformly in $y \in [x,1]$.

Step-sizes that satisfy this are $a(n) = \frac{1}{n}, \frac{1}{n \log n}, \frac{\log n}{n}$, etc. with suitable modification for $n = 0, 1$ as needed. The delays $\tau_{ij}(n) \in \{0, 1, \cdots, n\}$ were assumed to satisfy

$$E\left[\tau_{ij}(n)^b | X(m), Y(m), M(m), \tau_{ij}(m), i, j \in \{1, \cdots, d\}, m \leq n\right] < \infty \tag{6}$$

for some $b > \frac{r}{1-r}$ with r as in **(A1)**, and all $n \geq 0$. Suppose

$$\sup_n \|x(n)\| < \infty \text{ a.s.} \tag{7}$$

(i.e., the iterates remain bounded with probability one) and (3) has a globally asymptotically stable equilibrium x^*. Furthermore, assume that

$$\liminf_{n\to\infty} \frac{\nu(i,n)}{n} > 0, \ a.s. \tag{8}$$

That is, all components are updated 'comparably often'. Then we have:

***Theorem* 1.1** — $x(n) \to x^*$ *a.s.*

PROOF : This is a special case of Theorem 3.2 of [3]. As in *ibid.*, by unrolling the nth iterate in $|Y_n|$ separate iterates in each of which only one component is updated, we may suppose without loss of generality that Y_n is a singleton $\forall\, n$. To prove the claim, first note that global asymptotic stability of x^* implies by the converse Lyapunov theorem [9] the existence of a continuously differentiable Lyapunov function $V : \mathcal{R}^d \mapsto \mathcal{R}^+$ such that $\lim_{\|x\|\uparrow\infty} V(x) = \infty$ and $\langle \nabla V(x), h(x) \rangle < 0$ for $x \neq x^*$. Furthermore, in the notation of [3], $a(n,j) = a(n)$ for all $1 \leq j \leq d$ and thus $\bar{\beta}(i)$ (in the notation of [3]) is $\frac{1}{d}$. Therefore by Theorem 3.2 of [3], $x(n) \to x^*$ a.s. □

As already observed, this in fact is a very special case of Theorem 3.2 of [3], stated here to motivate the subsequent development. Our main point of departure is to assume that $h(\cdot) = [h_1(\cdot), \cdots, h_d(\cdot)]^T$ is of the form $h_i = \sum_{j\in\mathcal{N}(i)} h_{ij}$ where $\mathcal{N}(i)$ are prescribed nonempty subsets of a prescribed finite set $\mathcal{S}$ with $|\mathcal{S}| = r$ and the h_{ij}'s are Lipschitz. The intuition behind this is: the set $\mathcal{S}$ corresponds to 'sources' or 'measurement devices' that supply processor i with the relevant data in an episodic manner, with possible communication delays. The algorithm then is:

$$\begin{aligned} x_i(n+1) &= x_i(n) + \sum_{j\in\mathcal{N}(i)} a(\nu(i,j,n))\xi_{ij}(n+1)\Big[h_{ij}(x_1(n-\tau_{1i}(n)), \cdots, \\ &\qquad x_d(n-\tau_{di}(n))) + M_i(n+1)\Big], \end{aligned} \tag{9}$$

where $\{\xi_{ij}(n)\}$ are independent $\{0,1\}$-valued random variables that are also identically distributed for each choice of i, j, with $P(\xi_{ij}(n) = 1) = p_{ij} > 0$. The idea is that the ith component receives 'inputs' from agents $j \in \mathcal{N}(i)$. The latter may not always be active. The random variable $\xi_{ij}(n)$ is 1 if i did receive an input from j at time n, 0 if not. This justifies the terminology 'event-driven': the measurements are episodic. The 'clocks' $\{\nu(i,j,n)\}$ are now defined as $\nu(i,j,n) := \sum_{m=0}^{n} \xi_{ij}(m)$ and are seen to satisfy

$$\lim_{n\uparrow\infty} \frac{\nu(i,j,n)}{n} = \lim_{n\uparrow\infty} \frac{\sum_{m=0}^{n} \xi_{ij}(n)}{n} = p_{ij} > 0 \; a.s., \tag{10}$$

which replaces (8). The conditions on the delays $\{\tau_{ij}(n)\}$ will be specified later.

An example of such a situation is the scheme discussed in [8] for rating 'experts' whose opinions are sought on certain outcomes of interest. The algorithm is a stochastic approximation scheme which incrementally updates their ratings based on observed performance. As noted in [8], if one were to use a common step-size, the scheme will put more weight on and therefore favor the experts who opine more frequently. With this in mind, [8] uses the step-size schedule analyzed here. More examples from crowdsourcing and network economics applications are conceivable.

We analyze this scheme next.

2. Convergence Analysis

For simplicity, we first consider the case $\tau_{ij}(n) \equiv 0$, i.e., there are no communication delays. Assume that (7) holds and define

$$\bar{a}(n) := \max_{i,j} \Big(a(\nu(i,j,n))\xi_{ij}(n+1) \Big), \; n \geq 0,$$

which is ≥ 0 and is > 0 if and only if at least one $\xi_{ij}(n+1)$ equals 1. Then

$$\sum_{n} \bar{a}(n) \geq \sum_{n} a(\nu(i,j,n))\xi_{ij}(n+1) \; \forall \, i, j.$$

By Theorem 5.28, p. 96, [6], $\sum_n a(\nu(i,j,n))\xi_{ij}(n+1)$ and $\sum_n a(\nu(i,j,n))p_{ij}$ converge or diverge together a.s. But by (10),

$$\sum_{n} a(\nu(i,j,n))p_{ij} = p_{ij} \sum_{n} a(n) = \infty,$$

implying that $\sum_n \bar{a}(n) = \infty$ a.s. Now,

$$\begin{aligned}\sum_n \bar{a}(n)^2 &\leq \sum_n (\sum_{i,j} a(\nu(i,j,n))\xi_{ij}(n+1))^2 \\ &\leq K \sum_n a(n)^2 < \infty\end{aligned}$$

for a suitable $K \in (0, \infty)$. Rewrite (9) as

$$x_i(n+1) = x_i(n) + \bar{a}(n) \left(\sum_{j \in \mathcal{N}(i)} q(i,j,n) \left[h_{ij}(x(n)) + M_i(n+1) \right] \right),$$

where

$$q(i,j,n) := \frac{a(\nu(i,j,n))}{\bar{a}(n)} \xi_{ij}(n+1) \ \in \ [0,1] \ \forall \, n.$$

Next define:

1. $t(0) = 0, \ t(n) = \sum_{m=0}^{n} \bar{a}(m), \ n \geq 1$ (the algorithm's time scale),
2. $\bar{x}(\cdot) : [0, \infty) \mapsto \mathcal{R}^d$ by $\bar{x}(t(n)) = x(n)$ with linear interpolation on $[t(n), t(n+1)] \ \forall n$ (which makes it continuous and piecewise linear),
3. $\lambda_{ij}(\cdot) : [0, \infty) \mapsto [0, 1]$ by $\lambda_{ij}(t) = q(i,j,n) \ \forall \ i, j, n, \ t \in [t(n), t(n+1))$,
4. $\Lambda(\cdot) := [[\lambda_{ij}(\cdot)]] : [0, \infty) \mapsto \mathcal{R}^{d \times r}$, $\Lambda_i(\cdot) :=$ the ith row of $\Lambda(\cdot)$,
5. $H_i(\cdot) := [h_{i1}(\cdot), \cdots, h_{ir}(\cdot)]^T : \mathcal{R}^d \mapsto \mathcal{R}^r$ (column vector) where $h_{ij}(\cdot) \equiv 0$ for $j \notin \mathcal{N}(i)$,
6. $J_i = [z_{i1}, \cdots, z_{ir}] \in \mathcal{R}^r$ (row vector) where $z_{ij} = 1$ if $j \in \mathcal{N}(i)$, 0 otherwise,
7. $x^s(t), t \geq s \geq 0$, the solution to the o.d.e.

$$\dot{x}_i^s(t) = \Lambda_i(t) H_i(x(t)), \ t \geq s, \ x^s(s) = \bar{x}(s).$$

Then by standard arguments based on the Gronwall inequality as in, e.g., Chapter 7 of [4], we have:

Lemma 2.1 — For any $T > 0$, $\lim_{s \uparrow \infty} \sup_{t \in [s, s+T]} \|\bar{x}(t) - x^s(t)\| = 0$ a.s.

This leads to the key result:

Lemma 2.2 — Almost surely, any limit point of $\bar{x}(s + \cdot)$ in $C([0, \infty); \mathcal{R}^d)$ as $s \uparrow \infty$ is a solution of the o.d.e.

$$\dot{x}_i(t) = \alpha(t) J_i H_i(x(t)) = \alpha(t) h_i(x(t)), \ 1 \leq i \leq d, \tag{11}$$

where $\alpha(\cdot) : \mathcal{R}^+ \mapsto \mathcal{R}^+$ satisfies $\Delta \geq \alpha(t) \geq \delta \; \forall \, t \geq 0$ for suitable $\infty > \Delta > \delta > 0$.

PROOF : View $\Lambda(\cdot)$ as an element of the space $\mathcal{U}$ of measurable maps $u(\cdot) = [[u_{ij}(\cdot)]] : [0, \infty) \mapsto [0,1]^{d \times r}$ with the coarsest topology that renders continuous the maps $U(\cdot) \mapsto \int_0^T g(t) u_{ij}(t) dt$ $\forall \, T > 0, 1 \leq i \leq d, \; 1 \leq j \leq s, \; g(\cdot) \in L_2[0,T]$. Using Banach-Alaoglu theorem, one can prove that $\mathcal{U}$ is compact and metrizable, therefore Polish. From (7) and the Lipschitz condition on $h(\cdot)$, it follows that $\bar{x}(t + \cdot), t \geq 0$, are pointwise bounded and equicontinuous, hence relatively compact in $C([0,\infty); \mathcal{R}^d)$. Let $(\Lambda^*(\cdot) = [[\lambda^*_{ij}(\cdot)]], x^*(\cdot))$ denote a limit point of $(\Lambda(t + \cdot), \bar{x}(t + \cdot))$ in $\mathcal{U} \times C([0,\infty); \mathcal{R}^d)$ as $t \uparrow \infty$, along a subsequence, say, $t_n \uparrow \infty$. By Lemma 2.1, for $s > s' > 0$,

$$\bar{x}_i(t_n + s) - \bar{x}_i(t_n + s') = \int_{s'}^{s} \sum_{j \in \mathcal{N}(i)} \lambda_{ij}(t_n + y) h_{ij}(\bar{x}(t_n + y)) dy + o(1).$$

Letting $n \uparrow \infty$, we have

$$x^*(s) - x^*(s') = \int_{s'}^{s} \sum_{j \in \mathcal{N}(i)} \lambda^*_{ij}(y) h_{ij}(x^*(y)) dy.$$

Now for $j \notin \mathcal{N}(i)$, $\lambda_{ij}(\cdot) \equiv 0 \Longrightarrow \lambda^*_{ij}(\cdot) \equiv 0$. Define

$$N(n,s) := \min\{m > n : \sum_{k=n+1}^{m} \bar{a}(k) > s\}.$$

Let $[t]$ denote the unique integer such that $[t] \leq t < [t] + 1$. Then for $j \in \mathcal{N}(i), \ell \in \mathcal{N}(k)$ and $t, s > 0$,

$$\begin{aligned}
\frac{\int_t^{t+s} \lambda^*_{ij}(y) dy}{\int_t^{t+s} \lambda^*_{k\ell}(y) dy} &= \lim_{n \uparrow \infty} \frac{\int_t^{t+s} \lambda_{ij}(t_n + y) dy}{\int_t^{t+s} \lambda_{k\ell}(t_n + y) dy} \\
&= \lim_{n \uparrow \infty} \frac{\sum_{m=[t_n+t]}^{N([t_n+t],s)} \frac{a(\nu(i,j,m))\xi_{ij}(m)}{\bar{a}(m)} \bar{a}(m)}{\sum_{m=[t_n+t]}^{N([t_n+t],s)} \frac{a(\nu(k,\ell,m))\xi_{k\ell}(m)}{\bar{a}(m)} \bar{a}(m)} \\
&= \lim_{n \uparrow \infty} \frac{\sum_{m=\nu(i,j,[t_n+t])}^{\nu(i,j,N([t_n+t],s))} a(m)}{\sum_{m=\nu(k,\ell,[t_n+t])}^{\nu(k,\ell,N([t_n+t],s))} a(m)} \\
&= 1
\end{aligned} \tag{12}$$

a.s. by **(A4)** and (8). Thus for $x > 0$,

$$\begin{aligned}
\lim_{t \uparrow \infty} \frac{\int_0^x \int_0^t \lambda^*_{ij}(s+y) ds dy}{\int_0^x \int_0^t \lambda^*_{kl}(s+y) ds dy} &= \lim_{t \uparrow \infty} \frac{\int_0^t \int_0^x \lambda^*_{ij}(s+y) dy ds}{\int_0^t \int_0^x \lambda^*_{kl}(s+y) dy ds} \\
&= \lim_{t \uparrow \infty} \frac{\int_0^t \left[\left(\frac{\int_0^x \lambda^*_{ij}(s+y) dy}{\int_0^x \lambda^*_{kl}(s+y) dy} \right) \int_0^x \lambda^*_{kl}(s+y) dy \right] ds}{\int_0^t \int_0^x \lambda^*_{kl}(s+y) dy ds} \\
&= 1 \qquad \text{a.s.}
\end{aligned}$$

By l'Hôspital's rule,

$$\lim_{t\uparrow\infty}\frac{\int_0^x \lambda^*_{ij}(t+y)dy}{\int_0^x \lambda^*_{kl}(t+y)dy} = 1 \quad \text{a.s.}$$

Every limit point satisfies the above equation for $t \to \infty$. Since $x > 0$ was arbitrary, we conclude that for $t \geq 0$,

$$\int_t^{t+s} \lambda^*_{ij}(y)dy = \int_t^{t+s} \lambda^*_{k\ell}(y)dy \;\forall\; i, j, k, \ell,$$

for all $t, s > 0$, implying by Lebesgue's theorem that $\lambda^*_{ij}(\cdot) = \alpha(\cdot)$ (say) a.e. for some $\alpha(\cdot) \geq 0$ independent of i, j. We drop the qualification 'a.e.' by choosing a suitable version. It is easily verified from the definition of $\Lambda(\cdot)$ that

$$\Delta := d\sum_i |\mathcal{N}(i)| \geq \sum_{i,j\in\mathcal{N}(i)} \lambda_{ij}(t) \geq 1,$$

leading to

$$\Delta \geq \sum_{i,j\in\mathcal{N}(i)} \lambda^*_{ij}(t) = (d\sum_i |\mathcal{N}(i)|)\alpha(t) \geq 1.$$

Hence $\infty > \Delta \geq \alpha(t) \geq \delta$ for a suitable $\delta > 0$. This completes the proof. □

This brings us to our main result. Say that a set $\mathcal{A}$ is an internally chain transitive invariant set for (3) if for every $x \in \mathcal{A}$, the trajectory $x(t)$ of (3) with $x(0) = x$ remains in $\mathcal{A}$ for all $t \in \mathcal{R}$ and for any $x, y \in \mathcal{A}$ and $T, \epsilon > 0$, there exist $n \geq 1, x_0 = x, x_1, \cdots, x_{n-1}, x_n = y$ such that the trajectory of (3) initiated at $x_i, 0 \leq i < n$, intersects with the open ϵ-ball centered at x_{i+1} at some $t \geq T$. We then have the following extension of the celebrated result of Benaim for the classical Robbins-Monro scheme.

***Theorem* 2.1** — *Almost surely, $x(n) \to$ a nonempty compact connected internally chain transitive invariant set of (3). In particular, if (3) has a unique globally asymptotically stable attractor C, then C is the only such set and $x(n) \to C$ a.s.*

PROOF : Let $\gamma(t) := \int_0^t \alpha(s)ds$ and $\tilde{x}(t) := x(\gamma(t))$ where $x(\cdot)$ satisfies (3). Then

$$\dot{\tilde{x}}(t) = \alpha(t)h(\tilde{x}(t)).$$

Thus $x(\cdot)$ and $\tilde{x}(\cdot)$ have identical trajectories which differ only in time scaling. Furthermore, the time scaling function $\gamma(\cdot)$ along with its inverse has a bounded slope. Thus the two o.d.e.s have the same asymptotic behavior. Also, any limit point as $t \uparrow \infty$ of $\bar{x}(\gamma^{-1}(t + \cdot))$ will be a solution of (3). Now mimic the proof of Theorem 2, pp. 15-16, [4] with minor modifications to conclude. □

It is easy to see that if the right hand side of (12) were something other than 1, say $\kappa_{ijk\ell} > 0$, then we would have to replace (3) in the above statement by

$$\dot{x}_i(t) = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}(t) h_{ij}(x(t)),$$

where the $\alpha_{ij}(\cdot)$'s reflect the sampling bias. This in general would have a different asymptotic behavior than (3).

Recall also that we have ignored delays. We shall replace the condition (6) of [3] by the following conditions from [4], Chapter 7. These are more intuitive and allow for a much easier analysis:

- **(A5)** $\frac{n-\tau_{ij}(n)}{n} \to 0$ a.s.
- **(A6)** $\{\tau_{ij}(n)\}$ are stochastically dominated by a random variable τ satisfying $E\left[\tau^{\frac{1}{\eta}}\right] < \infty$ for some $\eta > 0$ such that $a(n) = o(n^{-\eta})$.

Both are very reasonable assumptions. The first does not allow for arbitrarily large delays whereas the second ensures that the delay distributions have uniformly well-behaved tails in a certain sense. In fact, if $\{\tau_{ij}(n),\ n \geq 0\}$ are identically distributed for some i, j, and $\eta = 1$, then $E[\tau_{ij}(k)] = \sum_n P(\tau_{ij}(k) \geq n) = \sum_n P(\tau_{ij}(n) \geq n) < \infty$, and a simple application of the Borel-Cantelli lemma shows that **(A6)** $\Longrightarrow$ **(A5)**.

Theorem* 2.1** — *Theorem 2.1 continues to hold for random delays satisfying* ***(A5)-(A6).

PROOF : This follows by the arguments of pp. 82-84, [4]. □

Finally, we shall state without proof a sufficient condition for (7) along the lines of [5], [2] (see also Theorem 7, pp. 26-27, [4]). The proof goes along the same lines as Theorem 7, pp. 26-27, [4].

***Theorem* 2.2** — *Suppose* $h_\infty(x) := \lim_{c \uparrow \infty} \frac{h(cx)}{c}$ *is well defined and the o.d.e.*

$$\dot{x}(t) = h_\infty(x(t))$$

has the origin as its unique globally asymptotically stable equilibrium. Then (7) holds.

REFERENCES

1. M. Benaim, A dynamical system approach to stochastic approximation, *SIAM Journal of Control and Optimization*, **34** (1996), 437-472.

2. S. Bhatnagar, The Borkar-Meyn theorem for asynchronous stochastic approximation, *Systems & Control Letters*, **60** (2011), 472-478.
3. V. S. Borkar, Asynchronous stochastic approximation, *SIAM Journal of Control and Optimization*, **36** (1998), 840-851 (Correction note in: 'Erratum: Asynchronous Stochastic Approximation', *SIAM Journal of Control and Optimization*, **38** (2000), 662-663).
4. V. S. Borkar, *Stochastic approximation: A dynamical systems viewpoint*, Hindustan Publising Agency, New Delhi, and Cambridge University Press, Cambridge, UK, 2008.
5. V. S. Borkar and S. P. Meyn, The O.D.E. method for convergence of stochastic approximation and reinforcement learning, *SIAM Journal of Control and Optimization*, **38** (2000), 447-469.
6. L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.
7. D. P. Derevitskii and A. I. Fradkov, Two models for analysing the dynamics of adaptation algorithms, *Automation and Remote Control*, **35** (1974), 59-67.
8. R. Dwivedi and V. S. Borkar, Removing sampling bias in networked stochastic approximation, *Proceedings of International Conference on Signal Processing and Communications (SPCOM)*, July 22-24, 2014, Bangalore.
9. N. N. Krasovskii, *Stability of motion*, Stanford Uni. Press, Stanford, CA, 1963.
10. L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Transactions on Automatic Control*, **22** (1977), 551-575.
11. H. Robbins and J. Monro, A stochastic approximation method, *Annals of Mathematical Statistics*, **22** (1951), 400-407.