THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CARIO, UNIV., VOL.(42) NO.2, 1998

On H –Valued Continuous Robbins - Monro processes

Hoda Mohamed Ali* , *Hala Elnagar
Mathematics Departement , Faculty of science (For Girls)
Al – Azhar University.

Abstract

In this work , we consider a continuous Robbins – Monro process in a Hilbert space , a nonlinear mapping f is treated instead of the usual linearity assumption .
A modified process with varying truncations is analyzed , and the asymptotic properties are investigated and convergence as well as necessary and sufficient conditions are obtained . Yin and Zhu [7] considered the discrete Robbins Monro in a Hilbert space , and we extend their results to the continuous case , by introducting randomly varying truncations and we prove the almost sure convergence of the stochastic approximation algorithms .

1 . Formulation of the problem :

Let x^* be a fixed point of H, but arbitrary. with

$\|x^*\| < M$, for some $M > 0$.

Let M(t) be an increasing positive real function , such that

$M(t) \xrightarrow{t\to\infty} \infty$. Define a sequence of a real - valued random variables by

$$k_1 = 1, \qquad k_{n+1} = k_n + I_{\{t_n < \infty\}}$$

where $I_{\{t_n<\infty\}}$ is the indicator of the set $\{\omega : t_n(\omega) < \infty\}, \omega \in \Omega$, where Ω is the sample space .

$$t_n = \inf\{t : \|x_{k_n}(t)\| > M(k_n)\}$$

Now , we define the continuous version stochastic approximation algorithm with randomly variation truncation as follows

for n = 1 , 2 , ... Let $t_0 = 0$, $x_n(t_{n-1}) = x^*$,

$$\frac{dx_n(t)}{dt} = -a(t) f(x_n(t)) + a(t)\xi(t), \qquad t_{n-1} \le t \le t_n \qquad (1.1)$$

Remark 1 :

It is clear that k_n is monotone increasing random sequence of positive integers , hence either $k_n \to \infty$ or $k_n \to k$, a finite limit .
If finite k exists , then there exists a positive integer N_1 , such that for all

$$t > t_{N_1 - 1} \quad , \left\| x_{N_1}(t) \right\| \le M(k),$$

i.e after finitely many steps x (*t*) will be bounded uniformly for any fixed ω.
Hence our main task to prove the finiteness of the limit k .

2 . *Assumptions :*

A1 : $a(t)$ is a real positive continuous function on $[0, \infty)$,

$$a(t) \to 0 \text{ for } t \to \infty \; ; \quad \int_1^\infty a(t)\,dt = \infty \qquad (2.1)$$

$$\left| \frac{a'(t)}{a(t)} \right| = O(a(t)) \qquad as\ t \longrightarrow \infty \qquad (2.2)$$

Remark 2 :

Let $a(t) = \frac{a}{t}, t \ge 1,$ Then A1 holds .

A2 : $\xi(t)$ is a stochastic process with pathes in H which are integrable on *a* $t-$ finite intervals and

$$\lim_{t \to \infty} a(t) \int_0^t \xi(\tau)\,d\tau = 0, \quad \text{with probability one .}$$

A3 : $f(\,.\,) : H \to H$ is an operator that maps bounded sets into bounded sets and has a zero point $\theta \in H$. f is continuous , and the continuity is uniformly on any bounded subset of H .

A4 : There is a twice continuously , Fre'chet differentiable functional V(Liapunov functional) $V : H \to R$ such that V maps bounded subsets of H into bounded subsets of R with

$$V(\theta) = 0 , \quad \lim_{|x| \to \infty} V(x) = \infty \qquad (2.3)$$

$$V(x) > 0 \quad , V'(x) f(x) > 0 \quad , \forall x \ne \theta$$

where $V'(x)$ and $V''(x)$ denote the first and the second Fre'chet derivative of V (x) , respectively (cf.[6] , [8]).

Remark 3 :

The assumption A3 covers a large class of nonlinear operators . In the following examples , we assume that f has a zero point at θ .

Example 1 :

If f is a linear operator , then A3 is clearly satisfied .

Example 2 :

Let $f: H \to H$ suppose that there exists $c_0 > 0$, $0 < \alpha \le 1$, such that

$$\|f(x) - f(y)\| \le c_0 \|x - y\|^{\alpha}, \qquad \text{for all} \quad x, y \in H$$

i.e. f is a Holder and Lipschitz operator then A3 holds .

Example 3 :

(Uniformly continuous operators).

$f: H \to H$ is uniformly continuous iff for any $\eta > 0$, there exists $\delta(\eta) > 0$, such that $|x - y| < \delta(\eta)$ implies $|f(x) - f(y)| < \eta$ (cf.[6]) .

The uniform continuity of f implies that f maps bounded sets into bounded sets . It follows that if f is uniformly continuous then A3 is satisfied . Many nonlinear operators such as continuous operators with "polynomial growth" and compact operator and integral and differential operators can make A3 holds .

(A4) is a "stability condition" . V can be viewed as Liapunov functional . When we are treating R^r values stochastic approximation problems , this kind of condition is used often (cf. [6]) , the actual form of V need not to be known .

If the operator f is Fre'chet differentiable at θ then a locally quadratic V can be used *i*.e , there is a self – adjoint , postive (linear) operator $Q: H \to H$, such that

$$V(x) = < x - \theta : Q(x - \theta) > + 0(|x - \theta|^2) .$$

(A2) is a "robustness" assumption (robust with respect to the random errors . Conditions such as independence or no correlation need not to be assumed . To see the scope of the assumption A2 (cf. [4],[5]) .

3. Main Results :

In the following , we need to show that the number of truncations is finite and this finiteness is the first step for obtaining the convergence result and we need the following lemma to prove theorem 1 and theorem 2 .

Lemma 3.1 :

Let $\Psi : R^{+} \to H$ be locally Riemann integrable function . Then

$$\lim_{t\to\infty} a(t)\int_0^t \Psi(s)ds = 0, \quad \text{iff}$$

$$\lim_{t\to\infty} \sup_{t\geq\tau} a(t)\int_\tau^t \Psi(s)ds = 0 \quad ,$$

(The proof is direct for R^r and can generalize to H, (cf [4]) .

Theorem 1 :

Let the assumptions A1-A4 hold ; Then

$$k_n \to k \quad \text{as} \quad n \to \infty$$

Proof :

If the truncations are executed infinitely often, *i.e* $k_n \to \infty$. Then there is a positive integer N_1 such that $M(k_{N_1}) > 2M$, then $x(t)$ would across the sphere $\{x;|x| = M\}$ infinitely often . and there would exists δ_1 and δ_2 with $0<\delta_1<\delta_2$, such that $[\delta_1,\delta_2] \subset (V(x^*),d)$, where $d=\inf \{V(x), |x| > M\}$.

Let $D = \{x;\delta_1 \leq V(x) \leq \delta_2\} \cap \{x:|x| \leq M\}$,

which implies that D is a closed set .

Then for every $N \geq N_1$, there exists t_1, t_2 such that $t_{N-1} < t_1 < t_2 < t_N$, be the time of the first entrance from the left and the first exit from the right of the sequence $\{V(x_j)\}$ to the interval $[\delta_1, \delta_2]$. By virtue of (A4), $\{x(t), t_1 \leq t \leq t_2\}$ is bounded .

Now Let $x_{k_n}(t) = x(t)$.

Let

$$T(t_1,\eta) = \max\left\{t; \int_{t_1}^{t} a(s)ds \leq \eta\right\} .$$

Let $T = \min\,[t_2, T(t,\eta)]$

multiply (1.1) by $V'(x(t))$ (inner product) we obtain

$$< V'(x(t)).\dot{x}(t) >= -a(t) < V'(x(t)).f(x(t)) > +a(t) < V'(x(t)).\xi(t) >$$

then

$$\dot{V}(x(t)) = -a(t) < V'(x(t)).f(x(t)) > +a(t)(V'(x(t)).\xi(t)) \tag{3.1}$$

by integration w.r.t. t we obtain

$$V(x(T)) - V(x(t_1)) = -\int_{t_1}^{T} a(u) < V'(x(u)).f(x(u)) > du + G_{t_1}^{T} \tag{3.2}$$

where

$$G_{t_1}^{T} =< a(T)u(T); V'(x(T)) > -\int_{t_1}^{T} < u(s), a(s)V'(x(s)) > ds - \int_{t_1}^{T} < u(s), a(s)V''(x(s))\dot{x}(s) > ds \tag{3.3}$$

where $$u(T) = \int_{t_1}^{T} \xi(s)ds \tag{3.4}$$

Let H'_t be the last integral in (3.3), thus by (1.1) we have

$$H'_t = \int_{t}^{t} < u(s), a^2(s)V''(x(s))f(x(s)) > ds - \int_{t}^{t} < u(s), a^2(s)V''(x(s))\xi(s) > ds \tag{3.5}$$

$|a'(s)| = o(a^2(s))$ as $s \to \infty$

using (3.3) to (3.5) we get

$$\|G_{t_1}^{T}\| \leq \| V'(x(T))\| \, \| a(T)u(T)\| +$$

$$+ k_1 \sup_{t \geq t_1} \| a(t)u(t)\| \{ \int_{t_1}^{T} [\|V'(x(s)) + \|V''(x(s))\| f(x(s))\| \, a(s)ds]$$

$$+ \int_{t_1}^{T} \|V''(x(s))\| \, \|\xi(s)\| a(s)ds\}$$

since $\| x \| \leq M$, $V'(x(s)$, $V''(x(s)$, $\int a(s)\xi(s)ds < \infty$,

$$\therefore \left\| G_{t_1}^{T} \right\| \leq k \sup_{t \geq t_1} \| a(t)u(t) \| (k_1\eta + k_2)$$

$$V(x(T)) - V(x(t_1)) \leq -\lambda\eta + \sup_{t \geq t_1} \| a(t)u(t) \| (k_1\eta + k_2)$$

let $\Delta(t_1) = \sup_{t \geq t_1} \| a(t)u(t) \|$; Let T be chosen and hence t_1

such that $\Delta(t_1)(k_1\eta + k_2) \leq \frac{\lambda\eta}{2}$

i.e $T > t_1$ choose t_1 such that

$\Delta(t_1)(k_1\eta + k_2) \leq \frac{\lambda\eta}{2}$ and this is possible and since

$\Delta(t_1) = \sup_{t \geq t_1} \| a(t)u(t) \|$; and this is converges to zero as $t_1 \to \infty$.

[according to lemma (3.1)] and hence $\Delta(t_1)$ can be taken arbitrary small and in this case we take t_1 such that

$$\Delta(t_1) \leq \frac{\lambda\eta}{2(k_1\eta + k_2)}$$

thus

$V(x(T)) - V(x(t_1)) \leq \frac{-\lambda\eta}{2}$, which contradicts that

$\delta_1 \leq V(x(t)) \leq \delta_2$, for $t \in [t_1, t_2]$

Theorem 2 :

Under the assumptions (A1) - (A4)

$x(t)$ converges to θ w.p.1 as $t \to \infty$.

Proof :

It follows from theorem (1) , that there exists $T_2 > T_1$ such that $\| x(t) \| \leq M(k)$ *forall* $t \geq T_2 - 1$

Thus from (3.2)

$$V(x(t)) - V(x(T_2)) = -\int_{T_2}^{t} a(s) < V'(x(s)), f(x(s)) ds + \int_{T_2}^{t} a(s) < V'(x(s)), \xi(s) > ds$$

we prove the theorem by contradiction . Suppose $\lim_{t\to\infty}\|x(t)-\theta\| \neq 0$. i.e there exists a positive number δ such that for every t, there exists t' such that $\|x(t')-\theta\| > \delta$, from the previous assumption $\|x(t)\|$ is finite and bounded after a certain time (finite truncation) .Assume that $\limsup_{t\to\infty} x(t) \neq \liminf_{t\to\infty} x(t) \neq \theta$ and $V(a) \neq V(b) \neq 0$, then there exists $[\alpha_1, \alpha_2] \subset [a, b]$, such that $V(x(t))$ would be across infinitely often, since $\|x(t)\| < M(k)$, and repeat the same arguments in theorem (1) we reach to a contradiction .

Now, assume that $\lim_{t\to\infty} x(t)$ exists $= \delta \neq \theta$

Then for $t > T_3 - 1$, $\|x(t) - \theta\| > \dfrac{\delta}{2}$

Now

$$V(x(t)) - V(x(T_3)) = - \int_{T_3}^{t} a(s) < V'(x(s)), f(x(s)) > ds$$

$$+ \int_{T_3}^{t} a(s) < V'(x(s)), \xi(s) > ds$$

since $< V'(x(s)), f(x(s)) > \; \geq 0$

then $\inf_{T_3 \leq s \leq t} < V'(x(s)), f(x(s)) >$

$\dfrac{\delta}{2} \leq \|x(s) - \theta\| \leq M(k)$

exists and is biger than zero.

Let $\lambda = \inf_{T_3 \leq s \leq t} < V'(x(s)), f(x(s)) >$

$\dfrac{\delta}{2} \leq \|x(s) - \theta\| \leq M(k)$

since $< V'(x(s)), f(x(s)) > \; \geq \lambda$

then

$- < V'(x(s)), f(x(s)) > \; \leq -\lambda$

then

$$V(x(t)) - V(x(T_3)) \le -\lambda \int_{T_3}^{t} a(s)ds + \int_{T_3}^{t} a(s) V'(x(s)), \xi(s) > ds \quad (3.6)$$

and since the last term integral in equation (3.6) is convergent from theorem (1) and $\int_{T_3}^{t} a(s)ds$ is divergent .

Then $\sup_{t \to \infty} V(x(t)) \le -\infty$,

which contradicts $V(x(t)) > 0$,

$\|x(t)\|$ is bounded , which completes the proof of the theorem .

4. Asymptotic property :

For the usual stochastic approximation algorithms , the normalized error has normal limiting distributions. For the varying truncation algorithms, the asymptotic normality still holds.

Let us examine $(x(t) - \theta) / \sqrt{a(t)}$ as $t \to \infty$ there exists N_1 , such that for all $t > t_{N_1 - 1}$

$$p\left\{ \sup_{t \ge N_{1-1}} \|x_{N_1}(t)\| \le M(k) \right\} > 1 - \varepsilon \text{ for every } \varepsilon > 0$$

Let $y(t)$ denotes the solution obtained from the usual stochastic approximation algorithm without truncation and with approximation condition leads to asymptotic normality . Thus for any bounded subset B of H , we have

$$\limsup_{t} p\left\{ \frac{x(t) - \theta}{\sqrt{a(t)}} \in B \right\} \le \limsup_{t} p\left\{ \frac{x(t) - \theta}{\sqrt{a(t)}} \in B,\ \sup_{t \ge N_1 - 1} \|x(t) - y(t)\| = 0 \right\}$$

$$+ \limsup_{t} p\left\{ \frac{x(t) - \theta}{\sqrt{a(t)}} \in B,\ \sup_{t \ge N_1 - 1} \|x(t) - y(t)\| > 0 \right\}$$

$$\le \limsup_{t} \left\{ \frac{y(t) - \theta}{\sqrt{a(t)}} \in B \right\} + \varepsilon$$

Similarly

$$\liminf_{t \to \infty} p\left\{ \frac{y(t) - \theta}{\sqrt{a(t)}} \in B \right\} \le \liminf_{t \to \infty} p\left\{ \frac{x(t) - \theta}{\sqrt{a(t)}} \in B \right\} + \varepsilon$$

since ε is arbitrary

$$\lim_{t \to \infty} p\left\{ \frac{y(t) - \theta}{\sqrt{a(t)}} \in B \right\} = \lim_{t \to \infty} p\left\{ \frac{x(t) - \theta}{\sqrt{a(t)}} \in B \right\}$$

which proves the asymptotic normality of the truncation process .

REFERENCES

1- Clark, S.D (1981). *"A stochastic difference Equation"*. Journal of differenetial equations. 40, 71-93 .

2- Doob, J.L. (1953). *"Stochastic processes"* John Wiley and Son, New York.

3- Robbins, H, and Monro (1951). *"Stochastic approximation method"* Annal Math. Statistic . vol 22 . PP. 400-407 .

4- Sorour, E. S. (1985). *" A continuous stochastic approximation* procedure ". Proc. Conf. Oper. Res. Of Math. Methods . Bull Fac. Alex. Volume 26 A(1),155 -165 .

5- Chen. H. F. And Zhu. Y. M. (1986) . *" Stochastic with rondomly* varying truncations" . Sci. Scinica Ser. A 29, 914-926 .

6- Berger M. (1977). *"Nonlinearity and functional Analysis"* . Academi Press, New York .

7- Yin, G and Zhu Y. M. *"On H-valued Robbins - Monro Processes"* (1990). Multivariate Analysis Journal, Vol 34, No 1, 116-140 .

8- Harry . Hcchstadt . (1975) . *"Differential Equations ; A modern approach"*. Dover publications , Inc . New York.

9- G .Yin .(1991). "O*n extensions of Polyak's averaging approach to stochastic approximation."Stochastic and Stochastic Reports*, 36 , pp.245- 264.

10- G.Yin. (1992). *"Stochastic Approximation Via Averaging Polyak 's Approach Revisited"*, Berlin: Springer-Verlag.vol.374 pp.119-134.

11- B.Delyon, A.B.Juditsky,(1992). *"Stochastic approximation with optimization of the trajectories." Stochastics and Stochastic Reports*, vol.39,pp.107-118 .

12- B.T.Polyak and A.B.Juditsky.(1992). *"Acceleration of stochastic approximation by averaging"*.SIAM J.Contr.optim.vol.30. pp.838-855.

13- H.J .Kushner,J.Yang.(1993). *"Stochastic approximation with averaging and feedback"*. Division of Applied Mathematics.Brown University,Providence,R.I.

14-B.T.Polyak,(1993).“*Accelerated stochastic approximation*”,S.I.A.M. Journal Optim., vol.3,no.4, pp.868-881.

15- A.Le Breton , (1994). “ *About gaussian schemes in stochastic approximation , stochastic process*” , App. , 50 , PP. 101 – 115 .

16- A.Le Breton and A.A. Novikov,(1994). “*Averaging for estimating covariances in stochastic approximation*” , Mathematical methods of statistics ,3,no.3,pp.244-266.

17- Harold J. Kushner,(1995). “*Stochastic Approximation with Averaging and Feedback*“ :Rapidly convergent “On-line” Algorithms”. IEEE, Transactions on Automatic Control-vol.40 No.1.PP 24-34.

18- A.V. Melnikov and E.Valkeila,(1996). “ *A new look on the method of stochastic approximation*”, preprint no.107, Univeirsity of Helsinki.

19- J.Poix,(1996). “*Stochastic approximation for semimartingale error*”, Mathematical Methods of Statistics,5,no.2. pp.227-236.

20- J.Poix,(1996). “*Stochastic approximation for local martingale error and random gain process*”.

21- Gang George Yin, KewenYin,(1996). ”*Passive stochastic approximation with constant step size and window width*” . IEEE Transaction on Automatic Control. vol.41. no.1.pp.90-106

22- H.J.kushner and G.Yin, (1997). “*Stochastic Approximation Algorithm and Application*” ,Springer-Verlag, New York.

23- P.L’Ecuyer and G.Yin,(1998). “*Budget-dependent rate of stochastic approximation*” . SIAM.J. Optim. 8,217-247.

24- G.Yin and Q.Zhang,(1998). “*Continuous-Time Markov chains and Applications*” , A Singular Perturbation Approach, Springer-Verlag,New York.

25- G.Yin. (1998). “Rates of convergence for a class of global stochastic optimization algorithms”. Department of Math. Wayne State Univ. pp.1-25.July 6.

26- K.J.Astrom G.C. Goodwin P.R Kumar. (1998). “*Adaptive Control , Filtering , and Signal Processing*” , Springer – Verlag . New York – Berlin – Heidelberg – London – Paris – Toky. PP. 374 – 396 .