'95 대한전기학회 하계학술대회 논문집(1995. 7. 20~22)

DI기법에 의한 스토캐스틱 순환적 알고리즘의 수렴분석

추 연 석
홍익대학교 전자전산공학과

Convergence analysis of stochastic recursive algorithms

Choo, Younseok
Department of Electronic and Computer Engineering, Hongik University

Abstract/ The ordinary differential equation (ODE) method has been widely used for the convergence analysis of stochastic recursive algorithms. The principal objective of this method is to associate to a given algorithm a differential equation with continuous righthand side. Usually some assumptions should be imposed to get such a differential equation. If any of assumptions fails, then the ODE method cannot be used. Recently a new method using differential inclusions (DIs) was introduced in [3], which is useful to deal with those cases. The DI method shares the same idea with the ODE method, but it is different in that a differential inclusion is identified instead of a differential equation with continuous righthand side. In this paper, we briefly review the DI method and then analyze a Robbins and Monro (RM)-type algorithm. Our focus is placed on the projected algorithm.

1 Introduction

Suppose we are given a stochastic recursive algorithm of the form

$$\theta_{n+1} = \theta_n + a_n h(\theta_n, \xi_{n+1}) \tag{1}$$

where $\theta_n \in R^r$, $\xi_n \in R^d$, h is a R^r-valued function and $\{a_n\}$ is a sequence of positive real numbers. The algorithm (1) arises in many applications. For example, (1) serves as an parameter estimator in adaptive control problems.

In this paper we are concerned with the convergence property of (1), i.e., we want to identify the limit points of the sequence $\{\theta_n\}$ generated by (1). The ODE method [4], [5] has been one of the most powerful tools for the convergence analysis of (1). In the ODE method we first associate to (1) a (deterministic) differential equation

$$\dot{\theta}(t) = \overline{h}(\theta(t)) \tag{2}$$

with $\overline{h}$ continuous and then analyze the asymptotic behavior of $\{\theta_n\}$ through the stability analysis of (2). Usually some assumptions should be imposed on h and/or the process $\{\theta_n, \xi_n\}$ to guarantee the continuity of $\overline{h}$, which may restrict the applicability of the ODE method in the sense that it cannot be used if any of assumptions fails.

Recently Choo and Arapostathis [3] introduced a new method using differential inclusions, which is useful to deal with the cases mentioned above. The DI method shares the same idea with the ODE method, but it is different in that we associate to (1) a differential inclusion of the form

$$\dot{\theta}(t) \in H(\theta(t)) \tag{3}$$

where H is a set-valued map. Any function $\theta(\cdot)$ is called a solution to (3) if it is absolutely continuous (therefore its derivative exists a.e.) and satisfies (3) a.e. on a given domain. Of course, the set-valued map H in (3) should be defined in an appropriate way so that (3) can be used for analyzing the asymptotic behavior of $\{\theta_n\}$. Furthermore (3) should be a reasonable generalization of (2).

One natural way is the following [1], [2]: Define the set-valued map H by

$$H(\theta) = \bigcap_{\varepsilon>0} \overline{co}\, \overline{h}(\theta+\varepsilon B) \tag{4}$$

where $\theta+\varepsilon B$ denotes the ε-neighborhood of θ and $\overline{co}A$ is the closed convex hull of A, i.e., the smallest closed convex set containing A. It is easy to see that $\overline{h}(\theta) \in H(\theta)$ for all θ and $H(\theta) = \{\overline{h}(\theta)\}$ whenever $\overline{h}$ is continuous at θ. Also H is upper semicontinuous (u.s.c.), i.e., for each θ and any open set B containing $H(\theta)$, there exists a neighborhood A of θ such that $H(A) \subseteq B$.

In the next section, a RM-type algorithm is analyzed via the DI method for both unprojected (Theorem 1) and projected (Theorem 2) cases. Theorem 1 was presented in [3], but it is repeated in this paper for ease of presentation. Theorem 2 shows that projected algorithms can be handled similarly by defining suitable differential inclusions.

2 A RM-type algorithm

First we consider an unprojected RM-type algorithm given by

$$\theta_{n+1} = \theta_n + a_n h(\theta_n) + v_n \tag{5}$$

where $\theta_n, v_n \in R^r$ and h is a R^r-valued function. The algorithm (5) is a special form of (1) and was analyzed in [4] for the case where h is continuous. Their results were applied to general classes of algorithms by many researchers [6], [7], [8]. We use the same condition on $\{v_n\}$ as in [4] (see [7] also) and identify a differential inclusion limit.

Let $t_n = \sum_{i=0}^{n-1} a_i$ and let $\theta^0(\cdot) := l.i.\{(\theta_n, t_n), t\}$, i.e., a piecewise linear interpolation function with interpolation steps $\{a_n\}$ such that $\theta^0(t_n) = \theta_n$ and $\theta^0(t) = \theta_0$ for $t < 0$. Define a sequence $\{\theta^n(\cdot)\}$ of left shifts of $\theta^0(\cdot)$ by $\theta^n(t) = \theta^0(t + t_n)$. $\overline{\theta}^0(\cdot)$ is a piecewise constant interpolation function such that $\overline{\theta}^0(t) = \theta_n$ for $t \in [\,t_n, t_{n+1})$, and $\overline{\theta}^0(t) = 0$ for $t < 0$. Define $\{\overline{\theta}^n(\cdot)\}$ in a similar manner.

Given a set-valued map F from an interval I to the subsets of R^r, define [1]

$$\int_I F(t)dt := \{\int_I f(t)dt : f \text{ is an integrable selection from } F\}.$$

The following assumptions will be used for the algorithm (5).

(A1) $a_n > 0$, $a_n \to 0$ and $\sum_{n=0}^{\infty} a_n = \infty$.

(A2) h is Borel-measurable and bounded on bounded sets.

(A3) $\{\theta_n\}$ is bounded a.s.

(A4) For each $T > 0$

$$\sup\{\,|\sum_{i=n}^{k} v_i| : n < k \le m(n, T)\} \to 0 \text{ a.s}$$

where $m(n, T) = \max\{k\ k > n, \sum_{i=n}^{k} a_i \le T\}$.

Define a set-valued map H by

$$H(\theta) = \bigcap_{\varepsilon > 0} co\, h(\theta + \varepsilon B). \tag{6}$$

Then we have

Theorem 1: For the algorithm defined by (5), assume (A1)-(A4) hold. Then each sequence in $\{\theta^n(\cdot)\}$ has a convergent subsequence and the limit of any convergent subsequence is a solution to the differential inclusion

$$\dot{\theta}(t) \in H(\theta(t)) \tag{7}$$

on $(-\infty, \infty)$ a.s.

Proof: See [3].

Next we consider a projected RM-type algorithm. Let G be a constraint set satisfying KC below.

(KC) $G = \{\theta\ q_i(\theta) \le 0,\ 1 \le i \le s\}$. G is bounded, convex and $q_i(\cdot)$ is continuously differentiable for each $1 \le i \le s$. At each $\theta \in \partial G$, the gradients of the active constraints are linearly independent.

The the projected version of (5) is

$$\theta_{n+1} = \pi_G(\theta_n + a_n h(\theta_n) + v_n) \tag{8}$$

where $\pi_G(\cdot)$ is the projection operator onto G. To deal with (8), we need an operator describing the motion of the limit process on the boundary. Therefore define [4]

$$\widehat{\pi}(h(\theta)) = \lim_{\delta \downarrow 0} \frac{\pi_G(\theta + \delta h(\theta)) - \theta}{\delta}.$$

Then $\widehat{\pi}(h(\theta))$ is a projection of the vector field h onto G at θ and defines a tangent vector at $\theta \in \partial G$ when $h(\theta)$ points outside of G.

Define a set-valued map $H_{\bar{\pi}}$ by

$$H_{\bar{\pi}}(\theta) = \bigcap_{\varepsilon > 0} co\, \widehat{\pi}(h(\theta + \varepsilon B)). \tag{9}$$

Then $H_{\bar{\pi}}$ is u.s.c. in θ with convex values. Using the idea in [4], we can prove the following theorem. Since G is compact, (A3) is not needed.

Theorem 2: Consider the projected algorithm (8) subject to (A1), (A2), (A4) and (KC). Then the conclusions of Theorem 1 continue to hold with the differential inclusion (7) replaced by

$$\dot{\theta}(t) \in H_{\bar{\pi}}(\theta(t)) \tag{10}$$

where $H_{\bar{\pi}}$ is as defined in (9).

Proof: For simplicity, assume $s = 1$. For the case $s > 1$, see [4]. Let Ω_o be the null set where (A4) is not satisfied and fix $\omega \notin \Omega_o$. Write (8) as

$$\theta_{n+1} = \theta_n + a_n h(\theta_n) + v_n + e_n$$

where e_n is the projection error given by

$$e_n = \pi_G(\theta_n + a_n h(\theta_n) + v_n) - (\theta_n + a_n h(\theta_n) + v_n).$$

Let $E^0(\cdot) := l.i.\{(\sum_{i=0}^{n-1} e_i, t_n), t\}$ and define $\{E^n(\cdot)\}$ to be a sequence of left shifts of $E^0(\cdot)$ such that $E^n(t) = E^0(t + t_n) - E^0(t_n)$, for $t \ge -t_n$, and $E^n(t) = -E^{0}(t_{n})$ for $t < -t_n$. Define $V^0(\cdot)$ and $\{V^n(\cdot)\}$ exactly in the same way. Then for $t \ge -t_n$

$$\theta^n(t) = \theta^n(0) + \int_0^t h(\overline{\theta}^n(s))ds + V^n(t) + E^n(t). \tag{11}$$

It is not difficult to see that $\{\theta^n(\cdot), V^n(\cdot), E^n(\cdot)\}$ is equicontinuous and pointwise bounded on $(-\infty, \infty)$. Let $\{\theta(\cdot), V(\cdot), E(\cdot)\}$ be the limit of a convergent subsequence (again indexed by n). We consider two cases. First, assume $\theta(t) \in G^o$ on $[\,t_1, t_2]$. Then, by the uniform convergence of $\theta^n(\cdot)$ to $\theta(\cdot)$ on finite intervals, there exists N such that $\theta^n(t) \in G^o$ for all

$t \in [t_1, t_2]$ and $n \geq N$, which in turn implies $\theta(\cdot)$ is a solution to the differential inclusion $\dot{\theta}(t) \in H(\theta(t))$ on $[t_1, t_2]$ by Theorem 1. Second, assume $\theta(t) \in \partial G$ on $[t_3, t_4]$. Let

$$\Delta_n(t) = h(\overline{\theta}^n(t)) - \widehat{\pi}(h(\overline{\theta}^n(t)).$$

Then

$$\theta^n(t_4) - \theta^n(t_3) = \int_{t_3}^{t_4} \widehat{\pi}(h(\overline{\theta}^n(s)))ds + \int_{t_3}^{t_4} \Delta_n(s)ds$$

$$+ V^n(t_4) - V^n(t_3) + E^n(t_4) - E^n(t_3). \quad (12)$$

$V^n(\cdot)$ converges to the zero process uniformly on finite intervals by (A4). Taking limits on both sides of (12) and using the fact that $\Delta_n(t)$ is perpendicular to ∂G for all t, we have

$$\theta(t_4) - \theta(t_3) = \lim_{n \to \infty} \int_{t_3}^{t_4} \widehat{\pi}(h(\overline{\theta}^n(s)))ds.$$

Clearly $\theta(\cdot)$ is Lipschitz continuous. Let $\varepsilon > 0$ and let $\dot{\theta}(t)$ exist at $t \in (t_1, t_2)$. Then there exists $\delta > 0$ such that for all $t' \in (t, t+\delta)$

$$\theta(t') - \theta(t) \in \int_t^{t'} (H_{\pi}(\theta(t)) + \varepsilon B)ds$$

$$= (t' - t)(H_{\pi}(\theta(t)) + \varepsilon B).$$

Therefore $\dot{\theta}(t) \in H_{\pi}(\theta(t))$ and this completes the proof. **Q.E.D.**

3 Conclusions

In this paper we briefly reviewed the DI method, which is quite useful to deal with stochastic recursive algorithms to which the ODE method cannot be applied. The main result of this paper is the analysis of the projected RM-type algorithm given in (8). Theorem 2 reveals that projected algorithms can be handled similarly by defining suitable differential inclusions.

References

[1] J.P. Aubin and A. Cellina, Differential inclusions, Springer-Verlag, 1984.

[2] A.F. Filippov, Differential equations with discontinuous righthand sides, Kluwer Academic Publishers, Netherlands, 1988.

[3] Y. Choo and A. Arapostathis, Some results on convergence of stochastic approximations by the differential inclusion methods, Proc. 33rd IEEE CDC, Lake Buena Vista, FL, Dec. 14-16, 1994.

[4] H.J. Kushner and D.S. Clark, Stochastic approximation method for constrained and un constrained systems, Applied Math. Sci., vol. 26, Springer-Verlag, N.Y., 1978.

[5] L. Ljung, Analysis of recursive stochastic algorithms, IEEE Trans. Automat. Contr., AC-22, pp. 551-575, 1977.

[6] D.-J. Ma, A.M. Makowski and A. Shwartz, Stochastic approximations for finite-state Markov chains, Stoch. Proc. Appl., vol. 35, pp. 27-45, 1990.

[7] M. Metivier and P. Priouret, Applications of a Kushner and Clark lemma to general classes of stochastic algorithms, IEEE Trans. Inform. Th., IT-30, pp. 140-151, 1984.

[8] J.C. Spall, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, IEEE Trans. Automat. Contr., AC-37, pp. 332-341, 1992.