Asymptotic Analysis of Stochastic Approximation Algorithms under Violated Kushner-Clark Conditions with Applications

Vladislav Tadić
Department of Electrical and Electronic Engineering, University of Melbourne
Parkville, Victoria 3010, Australia
e-mail: v.tadic@ee.mu.oz.au

Abstract

Motivated by the problem of the asymptotic behavior of temporal-difference learning algorithms with non-linear function approximation, the local almost sure asymptotic properties of stochastic approximation algorithms are analyzed for violated Kushner-Clark conditions. First, the algorithms with additive noise are analyzed for the case where the noise is state-dependent. The obtained results are then applied to the analysis of the algorithms with non-additive noise. Using these general results, the analysis of temporal-difference learning algorithms is carried out for the case of a general non-linear function approximation and under the assumptions allowing the underlying Markov chain to be positive Harris. The general results are also illustrated by an example where the noise is non-additive, correlated and satisfies strong mixing conditions.

1 Introduction

The asymptotic properties of stochastic approximation algorithms (almost sure, mean-square and weak convergence, asymptotic normality and convergence rate) have extensively been considered in a great number of papers (see [2], [11, Part I] and references cited therein). Among them, the almost sure convergence has probably gained the most of attention and research efforts. The most of existing results on the almost sure convergence of stochastic approximation algorithms have been obtained under noise conditions which can be considered as a special case of the Kushner-Clark condition (see [9]). On the other hand, if the Kushner-Clark condition is not satisfied, very few is known about their almost sure asymptotic properties. To the best of the present author's knowledge, these properties have been considered under violated Kushner-Clark conditions only in [4] and [5]. However, the analysis presented in these papers has been carried out under restrictive noise and stability conditions, while the obtained results are not applicable to strongly non-linear stochastic approximation algorithms such as those appearing in the area of reinforcement learning.

Temporal-difference learning algorithms are one of the most important classes of reinforcement learning algorithms and can be considered as a special case of stochastic approximation (see [3] and [13]). Their convergence has been analyzed in a great number of papers (see [3], citesutton and references cited therein). However, the existing results are constrained to the linear function approximation case. On the other hand, as opposed to the linear function approximation case, a suitable Lyapunov function is hard (if possible at all) to be constructed for general temporal-difference learning algorithms with non-linear function approximation. Therefore, the existing results on the almost sure asymptotic properties of stochastic approximation (such as those of [2], [4], [5], [9] and [11, Part I]) cannot be applied to their analysis.

Motivated by the problem of the asymptotic behavior of temporal-difference learning algorithms with non-linear function approximation, the local almost sure asymptotic properties of stochastic approximation algorithms are analyzed for violated Kushner-Clark conditions. First, the algorithms with additive noise are analyzed for the case where the noise is state-dependent (Section 2). The obtained results are then applied to the analysis of the algorithms with non-additive noise (Section 3). Using these general results, the analysis of temporal-difference learning algorithms is carried out for the case of a general non-linear function approximation and under the assumptions allowing the underlying Markov chain to be positive Harris (Section 4). The general results are also illustrated by an example where the noise is non-additive, correlated and satisfies strong mixing conditions (Section 3).

2 Algorithms with Additive Noise

Stochastic approximation algorithms with additive noise analyzed in this section are defined by the fol-

lowing difference equation:

$$\theta_{n+1}^{\varepsilon} = \theta_n^{\varepsilon} + \gamma_{n+1} h(\theta_n^{\varepsilon}) + \gamma_{n+1}\xi_{n+1}^{\varepsilon}, \quad n \geq 0. \qquad (1)$$

$h : R^d \to R^d$ is a Borel-measurable and locally bounded function (i.e., $\sup_{\theta \in Q} \|h(\theta)\| < \infty$ for any compact set $Q \subset R^d$), while $\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals. θ_0^{ε} is an R^d-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, while $\{\xi_n^{\varepsilon}\}_{n\geq 1}$ is an R^d-valued random process defined on the same probability space. $\{\xi_n^{\varepsilon}\}_{n\geq 1}$ is state-dependent, i.e., ξ_{n+1}^{ε} depends on $\theta_0^{\varepsilon}, \ldots, \theta_n^{\varepsilon}$ (e.g., $\xi_{n+1}^{\varepsilon} = \zeta_{n+1}^{\varepsilon}(\theta_0^{\varepsilon}, \ldots, \theta_n^{\varepsilon}, \omega)$), where $\zeta_n^{\varepsilon} : R^{nd} \times \Omega \to R^d$ is $\mathcal{B}^{nd} \times \mathcal{F}$-measurable). $\{\xi_n^{\varepsilon}\}_{n\geq 1}$ is referred to as the additive noise. ε is a non-negative parameter which characterizes the deviation of the asymptotic properties of the additive noise from the Kushner-Clark condition. Throughout this section the following convention is used: ε appears in the superscript of any entity depending on ε.

The following notation is used throughout the paper. R^+ and R_0^+ are sets of positive and non-negative reals (respectively), while $\|\cdot\|$ denotes the Euclidean vector and Frobenius matrix norm, as well as the total variation of a signed measure. For $\delta \in R^+$, $V(\cdot, \delta)$ and $\bar{V}(\cdot, \delta)$ stand for the open and closed δ-vicinity (respectively) induced by the Euclidean norm, while $d(\cdot, \cdot)$ is the distance induced by the same norm. $\mathcal{Q}^d$ is the family of compact sets from R^d, while $\mathcal{Q}_C^d$ is the family of sets from R^d being both convex and compact. C_L^1 denotes the family of differentiable functions mapping R^d into R and having locally Lipschitz continuous first-order derivatives.

Let $E_* = \{\theta \in R^d : h(\theta) = 0\}$ and $\bar{E}_* = \operatorname{cl} E_*$, while $\mathcal{Q}_*$ is the family of compact subsets of $\bar{E}_*^c$. For $\tau \in R^+$, let $\eta(n, \tau) = \sup\{j \geq n : \sum_{i=n}^{j-1} \gamma_{i+1} \leq \tau\}$, $n \geq 0$. The algorithm (1) is analyzed under the following assumptions:

A1 $\lim_{n\to\infty} \gamma_n = 0$, $\sum_{n=1}^{\infty} \gamma_n = \infty$.

A2 *For all $Q \in \mathcal{Q}^d$, there exists a constant $C_Q \in R^+$ such that the following relation holds for all $\varepsilon \in R_0^+$ and $\tau \in R^+$:*

$$\overline{\lim_{n\to\infty}} \sup_{n \leq j < \eta(n,\tau)} \left\| \sum_{i=n}^{j} \gamma_{i+1} \xi_{i+1}^{\varepsilon} \right\| I_{\{\nu_Q^{\varepsilon} > j\}} \leq C_Q \varepsilon \tau \ w.p.1, \qquad (2)$$

where $\nu_Q^{\varepsilon} = \inf(\{n \geq 0 : \theta_n^{\varepsilon} \notin Q\} \cup \{\infty\})$.

A3 *There exists $v \in C_L^1$ such that:*

(i) $\sup_{\theta \in Q} \dot{v}(\theta) < 0$, $\forall Q \in \mathcal{Q}_*$, where $\dot{v}(\theta) = \nabla^T v(\theta) h(\theta)$, $\theta \in R^d$,

(ii) the Lebesgue measure of $v(\bar{E}_*) \cap v(\bar{E}_*^c)$ is zero.

A1 is a standard assumption for the analysis of stochastic approximation algorithms (see e.g., [2], [11, Part I]).

A2 is a noise condition. It is satisfied if $\xi_n^{\varepsilon} = \tilde{\xi}_n^{\varepsilon} + \hat{\xi}_n^{\varepsilon}$, $n \geq 1$, where $\{\tilde{\xi}_n^{\varepsilon}\}_{n\geq 1}$ and $\{\hat{\xi}_n^{\varepsilon}\}_{n\geq 1}$ are R^d-valued random processes defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and satisfying

$$\lim_{n\to\infty} \sup_{n \leq j < \eta(n,\tau)} \left\| \sum_{i=n}^{j} \gamma_{i+1} \tilde{\xi}_{i+1}^{\varepsilon} \right\| I_{\{\nu_Q^{\varepsilon} > j\}} = 0 \ w.p.1,$$

$$\overline{\lim_{n\to\infty}} \|\hat{\xi}_{n+1}^{\varepsilon}\| I_{\{\nu_Q^{\varepsilon} > n\}} \leq C_Q \varepsilon \ w.p.1,$$

for all $\tau \in R^+$, $Q \in \mathcal{Q}^d$. Moreover, A2 is applicable to the analysis of stochastic approximation algorithms with non-additive noise, which is the subject of the next section. For $\varepsilon = 0$, A2 yields

$$\lim_{n\to\infty} \sup_{n \leq j < \eta(n,\tau)} \left\| \sum_{i=n}^{j} \gamma_{i+1} \xi_{i+1}^{0} \right\| I_{\{\nu_Q^{0} > j\}} = 0 \ w.p.1, \quad (3)$$

for all $\tau \in R^+$, $Q \in \mathcal{Q}^d$. The above relation is the Kushner-Clark condition slightly modified by the introduction of the stopping time ν_Q^0. The Kushner-Clark condition is the weakest noise condition still allowing the almost sure convergence of stochastic approximation algorithms to be shown. Moreover, under some additional conditions on $h(\cdot)$ (requiring $h(\cdot)$ to be continuous and $d\theta/dt = h(\theta)$ to have a globally asymptotically stable equilibrium), the Kushner-Clark condition is necessary and sufficient for their almost sure convergence (see e.g., [16]). If A1, A3 and (3) hold, then it can be demonstrated that $\lim_{n\to\infty} d(\theta_n^0, E_*) = 0$ w.p.1 on $\{\sup_{0\leq n} \|\theta_n^0\| < \infty\}$ (see Corollary 1). For $\varepsilon > 0$, the Kushner-Clark condition is violated, and consequently, $\lim_{n\to\infty} d(\theta_n^{\varepsilon}, E_*) = 0$ does not necessarily hold w.p.1 on $\{\sup_{0\leq n} \|\theta_n^{\varepsilon}\| < \infty\}$. The parameter ε characterizes the degree of the violation of the Kushner-Clark condition, and the aim of the analysis carried out in this section is determining how ε influences the deviation of the asymptotic behavior of $\{\theta_n^{\varepsilon}\}_{n\geq 0}$ from the convergence towards E_*, i.e., the aim is to determine an almost sure upper bound for $\overline{\lim}_{n\to\infty} d(\theta_n^{\varepsilon}, E_*)$ and to find how this bound depends on ε.

A3 corresponds to the stability properties of $h(\cdot)$, i.e., of $d\theta/dt = h(\theta)$ (provided that $h(\cdot)$ is continuous). The condition (i) of A3 requires $h(\cdot)$ to have a global Lyapunov function. If $h(\cdot)$ is continuous, it reduces to the requirement that $\dot{v}(\theta) < 0$, $\forall \theta \in E_*^c$. In the context of the Lyapunov stability, this requirement represents the weakest condition under which the Lagrange stable solutions of $d\theta/dt = h(\theta)$ tend to E_* (see e.g., [10, Invariance Principle]). The condition (ii) of A3 does not have an interpretation in the context of the Lyapunov stability and is specific for the analysis of stochastic approximation algorithms. It ensures that any closed continuous path starting and ending at the interior of

E^c_* has a subpath also belonging to E^c_* along which $v(\cdot)$ decreases.

In comparison with the assumptions adopted in [4] and [5], A1 – A3 are more general and covers a wider class of stochastic approximation algorithms. The noise conditions of [4] and [5] require $\{\xi^\varepsilon_n\}_{n\geq 1}$ to be exogenous (i.e., not to depend on $\{\theta^\varepsilon_n\}_{n\geq 0}$). The stability conditions of [4] and [5] require $h(\cdot)$ to be continuous and $d\theta/dt = h(\theta)$ to have a globally asymptotically stable equilibrium. Apparently, this is the simplest special case of A3. Moreover, in the context of the Lyapunov stability, A3 is more general than the stability conditions required by the existing results on the convergence of stochastic approximation algorithms — compare A3 with the corresponding assumptions of [2], [9] and [11], as well as with those of [1], [6], [7] and [8]. It is also important to emphasize that A3 covers several classes of strongly non-linear stochastic approximation algorithms to which the existing robustness and convergence results cannot be applied (see Section 4). Furthermore, in the context of the Lyapunov stability, A3 seems to be the weakest condition which still allows obtaining almost sure asymptotic results.

Theorem 1 *Let A1 – A3 hold. Then, for all $Q \in \mathcal{Q}^d$, there exists a non-decreasing function $\phi_Q : R^+_0 \to R^+_0$ such that $\lim_{s\to 0+} \phi_Q(s) = \phi_Q(0) = 0$ and such that $\overline{\lim}_{n\to\infty} d(\theta^\varepsilon_n, E_*) \leq \phi_Q(\varepsilon)$ w.p.1 on $\bigcap_{0\leq n}\{\theta^\varepsilon_n \in Q\}$ for all $\varepsilon \in R^+_0$.*

The proof is given in [14].

Corollary 1 *Let A1 – A3 hold. Then, $\overline{\lim}_{n\to\infty} d(\theta^0_n, E_*) = 0$ w.p.1 on $\{\sup_{0\leq n}\|\theta^0_n\| < \infty\}$.*

3 Algorithms with Non-Additive Noise

Stochastic approximation algorithms with non-additive noise analyzed in this section are defined by the following difference equation:

$$\begin{aligned}\theta^\varepsilon_{n+1} &= \theta^\varepsilon_n + \gamma_{n+1}H_{n+1}(\theta^\varepsilon_n, \tilde{X}^\varepsilon_{n+1}) \\ &\quad +\gamma_{n+1}\rho^\varepsilon_{n+1}(\tilde{\theta}^\varepsilon_n, \tilde{X}^\varepsilon_{n+1}), \quad n \geq 0. \qquad (4)\end{aligned}$$

$H_n : R^d \times R^{nd'} \to R^d$ and $\rho^\varepsilon_n : R^{nd} \times R^{nd'} \to R^d$ are Borel-measurable functions. θ^ε_0 and $\{\gamma_n\}_{n\geq 1}$ have the same meaning as in Section 2. $\{X^\varepsilon_n\}_{n\geq 0}$ is an $R^{d'}$-valued random process defined on $(\Omega, \mathcal{F}, \mathcal{P})$, while $\tilde{\theta}^\varepsilon_n = (\theta^\varepsilon_0, \ldots, \theta^\varepsilon_n)$ and $\tilde{X}^\varepsilon_n = (X^\varepsilon_0, \ldots, X^\varepsilon_n)$, $n \geq 0$. $\{X^\varepsilon_n\}_{n\geq 0}$ is referred to as the non-additive noise. Similarly as in Section 2, ε is a non-negative parameter which characterizes the deviation of the asymptotic properties of the non-additive noise from those ensuring almost sure convergence. Throughout this section, the following convention is used: ε appears in the superscript of any entity depending on ε.

The algorithm (4) is analyzed under the assumptions A1, A3 (given in Section 2) and B1 – B3 (given below).

B1 *$h : R^d \to R^d$ is a locally Lipschitz continuous function. For all $Q \in \mathcal{Q}^d$, there exists a constant $C'_Q \in R^+$ such that the following relation holds for all $\varepsilon \in R^+_0$, $\tau \in R^+$, $\theta \in Q$:*

$$\begin{aligned}&\overline{\lim_{n\to\infty}} \sup_{n\leq j<\eta(n,\tau)} \left\| \sum_{i=n}^{j} \gamma_{i+1}(H_{i+1}(\theta, \tilde{X}^\varepsilon_{i+1}) - h(\theta)) \right\| \\ &\leq C'_Q \varepsilon\tau \ w.p.1.\end{aligned}$$

B2 *For all $Q \in \mathcal{Q}^d$, there exist Borel-measurable functions $\psi^Q_n : R^{nd'} \to R^+_0$, $n \geq 1$, such that*

$$\|H_n(\theta, x)\| \leq \psi^Q_n(x); \quad \forall\theta \in Q, \ \forall x \in R^{nd'}, \ n \geq 1,$$

$$\begin{aligned}&\|H_n(\theta, x) - H_n(\theta', x)\| \leq \psi^Q_n(x)\|\theta - \theta'\|; \\ &\forall\theta, \theta' \in Q, \ \forall x \in R^{nd'}, \ n \geq 1,\end{aligned}$$

$$\lim_{t\to 0+} \overline{\lim_{n\to\infty}} \sum_{i=n}^{\eta(n,t)-1} \gamma_{i+1}\psi^Q_{i+1}(\tilde{X}^\varepsilon_{i+1}) = 0 \ w.p.1.$$

B3 *For all $Q \in \mathcal{Q}^d$, there exists a constant $C''_Q \in R^+$ such that the following relation holds for all $\varepsilon \in R^+_0$, $\tau \in R^+$:*

$$\begin{aligned}&\overline{\lim_{n\to\infty}} \sum_{i=n}^{\eta(n,\tau)-1} \gamma_{i+1}\|\rho^\varepsilon_{i+1}(\tilde{\theta}^\varepsilon_i, \tilde{X}^\varepsilon_{i+1})\| I_{\{\nu^\varepsilon_Q > i\}} \\ &\leq C''_Q \varepsilon\tau \ w.p.1,\end{aligned}$$

where $\nu^\varepsilon_Q = \inf(\{n \geq 0 : \theta^\varepsilon_n \notin Q\} \cup \{\infty\})$.

For $\varepsilon = 0$, B1 and B3 yield

$$\begin{aligned}&\lim_{n\to\infty} \sup_{n\leq j<\eta(n,\tau)} \left\| \sum_{i=n}^{j} \gamma_{i+1}(H_{i+1}(\theta, \tilde{X}^0_{i+1}) - h(\theta)) \right\| \\ &= 0 \ w.p.1; \quad \forall\tau \in R^+, \ \forall\theta \in R^d, \qquad (5)\end{aligned}$$

$$\begin{aligned}&\lim_{n\to\infty} \sum_{i=n}^{\eta(n,\tau)-1} \gamma_{i+1}\|\rho^0_{i+1}(\tilde{\theta}^0_i, \tilde{X}^0_{i+1})\| I_{\{\nu^0_Q > i\}} \\ &= 0 \ w.p.1, \quad \forall\tau \in R^+. \qquad (6)\end{aligned}$$

Assumptions similar to B2, (5) and (6) have been used in [9] and [12]. However, only the convergence has been considered therein. Moreover, the stability conditions adopted in [9] and [12] are much more restrictive than

A3 — they require $d\theta/dt = h(\theta)$ to have an asymptotically stable equilibrium. If A3, B2, (5) and (6) hold, then it can be demonstrated that $\lim_{n\to\infty} d(\theta_n^0, E_*) = 0$ w.p.1 on $\{\sup_{0\le n} \|\theta_n^0\| < \infty\}$ (see Corollary 2). If $\varepsilon > 0$, (5) and (6) are not satisfied any more, and $\lim_{n\to\infty} d(\theta_n^0, E_*) = 0$ does not hold necessarily w.p.1 on $\{\sup_{0\le n} \|\theta_n^0\| < \infty\}$. The parameter ε characterizes the degree of the violation of (5) and (6), and the aim of the analysis carried out in this section is determining how ε influences the deviation of the asymptotic behavior of $\{\theta_n^\varepsilon\}_{n\ge 0}$ from the convergence towards E_*.

Theorem 2 *Let A1, A3 and B1 – B3 hold. Then, for all $Q \in \mathcal{Q}^d$, there exists a non-decreasing function $\phi_Q : R_0^+ \to R_0^+$ such that $\lim_{s\to 0+} \phi_Q(s) = \phi_Q(0) = 0$ and such that $\overline{\lim}_{n\to\infty} d(\theta_n^\varepsilon, E_*) \le \phi_Q(\varepsilon)$ w.p.1 on $\bigcap_{0\le n}\{\theta_n^\varepsilon \in Q\}$ for all $\varepsilon \in R_0^+$.*

The proof is given in [14].

Corollary 2 *Let A1, A3 and B1 – B3 hold. Then, $\overline{\lim}_{n\to\infty} d(\theta_n^0, E_*) = 0$ w.p.1 on $\{\sup_{0\le n} \|\theta_n^0\| < \infty\}$.*

Now, the results of Theorem 2 are illustrated by an example where the non-additive noise is correlated and satisfies strong mixing conditions. The following subclass of the algorithms (4) is analyzed in the rest of the section:

$$\theta_{n+1}^\varepsilon = \theta_n^\varepsilon + \gamma_{n+1} H(\theta_n^\varepsilon, X_{n+1}^\varepsilon), \quad n \ge 0. \tag{7}$$

$H : R^d \times R^{d'} \to R^d$ is a Borel-measurable function, while θ_0^ε, $\{\gamma_n\}_{n\ge 1}$ and $\{X_n^\varepsilon\}_{n\ge 0}$ have the same meaning as in the case of the algorithm (4).

The algorithm (7) is analyzed for the case where $\{X_n^\varepsilon\}_{n\ge 0}$ satisfies strong mixing conditions. Let $\mu_n^\varepsilon(\cdot)$ be the probability measure of X_n^ε, $n \ge 0$, while $\mathcal{F}_0^\varepsilon = \sigma\{\theta_0^\varepsilon, X_0^\varepsilon, \ldots, X_n^\varepsilon,\}$, $n \ge 0$. The analysis is carried out under assumption A1 (given in Section 2) and C1 – C4 (given below).

C1 $0 < \lim_{n\to\infty} n\gamma_n < \infty$.

C2 *For all $Q \in \mathcal{Q}^d$, there exists a Borel-measurable function $\psi_Q : R^{d'} \to R_0^+$ satisfying*

$$\|H(\theta, x)\| \le \psi_Q(x); \quad \forall \theta \in Q,\ \forall x \in R^{d'},$$

$$\|H(\theta, x) - H(\theta', x)\| \le \psi_Q(x)\|\theta - \theta'\|; \quad \forall \theta, \theta' \in Q,\ \forall x \in R^{d'}.$$

C3 *There exist constants $p,\ r \in (1,\infty)$, $q \in (p,\infty)$ and a sequence $\{\alpha_n\}_{n\ge 1}$ of positive reals satisfying $pq^{-1} + r^{-1} = 1$, $\sum_{n=1}^\infty \alpha_n^p n^{-1} \log^{4q} n < \infty$, $\sup_{0\le n} \int \psi_Q^{2r}(x)\mu_n^\varepsilon(dx) < \infty$, $\forall Q \in \mathcal{Q}^d$, and*

$$E|\mathcal{P}(X_j^\varepsilon \in B|\mathcal{F}_n^\varepsilon) - \mu_j^\varepsilon(B)| \le \alpha_{j-n}; \quad \forall \varepsilon \in R_0^+,\ \forall B \in \mathcal{B}^{d'},\ 0 \le n < j$$

($\|\mu_n^\varepsilon - \mu\|$ denotes the total variation of the signed measure $\mu_n^\varepsilon - \mu$).

C4 *There exists a probability measure $\mu(\cdot)$ on $(R^{d'}, \mathcal{B}^{d'})$ such that $\overline{\lim}_{n\to\infty} \|\mu_n^\varepsilon - \mu\| \le \varepsilon$, $\forall \varepsilon \in R_0^+$.*

Theorem 3 *Let C1 – C4 hold and let A3 be satisfied with $h(\theta) = \int H(\theta, x)\mu(dx)$, $\theta \in R^d$. Then, for all $Q \in \mathcal{Q}^d$, there exists a non-decreasing function $\phi_Q : R_0^+ \to R_0^+$ such that $\lim_{s\to 0+} \phi_Q(s) = \phi_Q(0) = 0$ and such that $\overline{\lim}_{n\to\infty} d(\theta_n^\varepsilon, E_*) \le \phi_Q(\varepsilon)$ w.p.1 on $\bigcap_{0\le n}\{\theta_n^\varepsilon \in Q\}$ for all $\varepsilon \in R_0^+$.*

The proof is given in [14].

Corollary 3 *Let A3 and C1 – C4 hold. Then, $\overline{\lim}_{n\to\infty} d(\theta_n^0, E_*) = 0$ w.p.1 on $\{\sup_{0\le n} \|\theta_n^0\| < \infty\}$.*

4 Temporal-Difference Learning

Using the results of Theorem 2, temporal-difference learning algorithms with non-linear function approximation are analyzed in this section. These algorithms are defined by the following difference equations:

$$\theta_{n+1}^\lambda = \theta_n^\lambda + \gamma_{n+1}\delta_{n+1}^\lambda \varepsilon_{n+1}^\lambda, \quad n \ge 0, \tag{8}$$

$$\delta_{n+1}^\lambda = g(X_n, X_{n+1}) + \alpha f(\theta_n^\lambda, X_{n+1}) - f(\theta_n^\lambda, X_n), \quad n \ge 0, \tag{9}$$

$$\varepsilon_{n+1}^\lambda = \sum_{i=0}^n (\alpha\lambda)^{n-i} \nabla_\theta f(\theta_i^\lambda, X_i), \quad n \ge 0. \tag{10}$$

$f : R^d \times R^{d'} \to R$ and $g : R^{d'} \times R^{d'} \to R$ are Borel-measurable functions, while $f(\cdot, x)$, $x \in R^{d'}$, is differentiable. $\alpha \in (0,1)$ is a constant, while $\lambda \in [0,1]$ is the algorithm parameter. $\{\gamma_n\}_{n\ge 0}$ has the same meaning as in Section 3. θ_0^λ is an R^d-valued random variable defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, while $\{X_n\}_{n\ge 0}$ is an $R^{d'}$-valued random process defined on the same probability space. Throughout this section the following convention is used: λ appears in the superscript of any entity depending on λ.

The algorithm (8) – (10) is analyzed for the case where $\{X_n\}_{n\ge 0}$ is a homogeneous Markov chain having a unique invariant probability measure. Let $P(x, \cdot)$,

$x \in R^{d'}$, and $\mu(\cdot)$ be the transition and invariant probability measure of $\{X_n\}_{n\geq 0}$ (respectively), while $f_*(x) = E\,(\sum_{n=0}^{\infty} \alpha^n g(X_n, X_{n+1})|X_0 = x)$, $x \in R^{d'}$, is a discounted cost-to-go function associated to $\{X_n\}_{n\geq 0}$. For $\theta \in R^d$ and $x \in R^{d'}$, let $J(\theta, x) = 2^{-1}(f_*(x) - f(\theta, x))^2$ and $J_*(\theta) = \int J(\theta, x)\mu(dx)$, while $E_* = \{\theta \in R^d : \nabla J_*(\theta) = 0\}$ (provided that $f_*(\cdot)$ and $\nabla J_*(\cdot)$ are well-defined and finite). The algorithm (8) – (10) aims at determining $\theta \in R^d$ such that $f(\theta, \cdot)$ fits $f_*(\cdot)$. If $\lambda = 1$, then the algorithm (8) – (10) determines $\theta \in R^d$ such that $f(\theta, \cdot)$ approximates $f_*(\cdot)$ optimally in the $L^2(\mu)$-sense, i.e., it minimizes $J_*(\cdot)$ (see Corollary 3).

Let $\varepsilon = 1 - \lambda$. For $\theta \in R^d$, $\vartheta_i \in R^d$, $x_i \in R^{d'}$, $0 \leq i \leq n+1$, $n \geq 0$, and $\vartheta = (\vartheta_0, \ldots, \vartheta_n)$, $x = (x_0, \ldots, x_{n+1})$, let

$$H_{n+1}^{\varepsilon}(\theta, x) = (g(x_n, x_{n+1}) + \alpha f(\theta, x_{n+1}) - f(\theta, x_n)) \cdot \sum_{i=0}^{n} \alpha^{n-i} \nabla_\theta f(\theta, x_i), \tag{11}$$

$$\rho_{n+1}^{\varepsilon}(\vartheta, x) = (g(x_n, x_{n+1}) + \alpha f(\vartheta_n, x_{n+1}) - f(\vartheta_n, x_n)) \cdot \sum_{i=0}^{n} \alpha^{n-i} (\lambda^{n-i} \nabla_\theta f(\vartheta_i, x_i) - \nabla_\theta f(\vartheta_n, x_i)). \tag{12}$$

Then, it can easily be verified that the algorithm (8) – (10) is of the same form as the algorithm (4).

For $x \in R^{d'}$ and $B \in \mathcal{B}^{d'}$, let $\tilde{g}(x) = \int g(x, x')P(x, dx')$, while $P_0(x, B) = I_B(x)$ and $P_{n+1}(x, B) = \int P(x', B)P_n(x, dx')$, $n \geq 0$. Let $I_n = \{1, \ldots, d\}^n$, $n \geq 1$. For $\theta \in R^d$ and $a = (a_1, \ldots, a_n) \in I_n$, $n \geq 1$, let D_θ^a denote $\partial^n / \partial t_{a_1} \cdots \partial t_{a_n}$, where t_i is the i-th component of θ. The algorithm (8) – (10) is analyzed under the assumptions C1 (given in Section 3) and D1 – D3 (given below).

D1 *There exist a Borel-measurable function* $\varphi : R^{d'} \to R_0^+$ *and a constant* $A \in [1, \infty)$ *such that* $\int \varphi^2(x)\mu(dx) < \infty$ *and*

$$\int g^2(x, x')P(x, dx') \leq \varphi^2(x), \quad \forall x \in R^{d'},$$

$$\sum_{n=0}^{\infty} \alpha^n (P_n \varphi^2)(x) < \infty, \quad \forall y \in R^{d'}.$$

D2 $f(\cdot, x)$, $x \in R^{d'}$, *is* $(d+1)$*-times differentiable. For all* $Q \in \mathcal{Q}^d$, *there exist a Borel-measurable function* $\varphi_Q : R^{d'} \to R_0^+$ *and a constant* $A_Q \in [1, \infty)$ *such that* $\int \varphi_Q^2(x)\mu(dx) < \infty$ *and*

$$|f(\theta, x)| \leq \varphi_Q(x); \quad \forall \theta \in Q,\ \forall x \in R^{d'},$$

$$\sup_{a \in I_n} |D_\theta^a f(\theta, x)| \leq \varphi_Q(x); \quad \forall \theta \in Q,\ \forall x \in R^{d'},\ 1 \leq n \leq d+1,$$

$$\sum_{n=0}^{\infty} \alpha^n (P_n \varphi_Q^2)(x) < \infty, \quad \forall x \in R^{d'}.$$

D3 *For all* $\theta \in R^d$,

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} g(X_{i+j}, X_{i+j+1}) \nabla_\theta f(\theta, X_i) = \int (P_j \tilde{g})(x) \nabla_\theta f(\theta, x)\mu(dx)\ w.p.1, \quad j \geq 0,$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f(\theta, X_{i+j}) \nabla_\theta f(\theta, X_i) = \int (P_j f)(\theta, x) \nabla_\theta f(\theta, x)\mu(dx)\ w.p.1, \quad j \geq 0.$$

Assumptions D1 and D2 hold if $f(\cdot, \cdot)$, $\nabla_\theta f(\cdot, \cdot)$ and $D_\theta^a f(\cdot, \cdot)$, $a \in I_n$, $1 \leq n \leq d+1$, are locally bounded and if there exists a constant $K \in R^+$ such that $\|X_n\| \leq K$ w.p.1, $n \geq 0$. From the application point of view, this is probably the most important case. On the other hand, it can be shown that D3 is satisfied if D1 and D2 hold and if $\{X_n\}_{n\geq 0}$ is positive Harris (see [14]).

The convergence of temporal-difference learning algorithms has been analyzed in a great number of papers (see [3], [13] and references cited therein). Among the existing results, those of [15] are probably the strongest. It can easily be shown that the assumptions adopted in [15] are just a special case of D1 – D3. Moreover, the assumptions of [15] practically cover only the case where $\{X_n\}_{n\geq 0}$ is geometrically ergodic and are not satisfied if $\{X_n\}_{n\geq 0}$ is positive Harris. Furthermore, the analysis presented in [15] is constrained to the linear function approximation case, i.e., to the case where $f(\theta, x) = \theta^T \phi(x)$, $\forall \theta \in R^d$, $\forall x \in R^{d'}$, and where $\phi : R^{d'} \to R^d$ is a Borel-measurable function. On the other hand, the analysis given in this section is carried out for the general case $\lambda \in (0, 1]$ and under the assumptions requiring the approximator $f(\cdot, \cdot)$ only to be sufficiently smooth with respect to the first argument. It is particularly important to emphasize that as opposed to the linear function approximation case, a suitable Lyapunov function is hard (if possible at all) to be constructed for the case of $\lambda \in (0, 1)$ and general non-linear approximator. Therefore, the existing results on the asymptotic properties of stochastic approximation cannot be applied to this case. However, using Theorem 2, the following results on the almost sure asymptotic properties of the algorithm (8) – (10) are obtained:

Theorem 4 *Let C1 and D1 – D3 hold. Then, for all* $Q \in \mathcal{Q}^d$*, there exists a non-decreasing function* $\phi_Q : [0,1] \to R_0^+$ *such that* $\lim_{s\to 0+} \phi_Q(s) = \phi_Q(0) = 0$ *and such that* $\overline{\lim}_{n\to\infty} d(\theta_n^\lambda, E_*) \leq \phi_Q(1-\lambda)$ *w.p.1 on* $\bigcap_{0\leq n}\{\theta_n^\lambda \in Q\}$ *for all* $\lambda \in [0,1]$.

The proof is given in [14].

Corollary 4 *Let C1 and D1 – D3 hold. Then,* $\overline{\lim}_{n\to\infty} d(\theta_n^1, E_*) = 0$ *w.p.1 on* $\{\sup_{0\leq n} \|\theta_n^1\| < \infty\}$.

5 Conclusion

Motivated by the problem of the asymptotic behavior of temporal-difference learning algorithms with non-linear function approximation, the local almost sure asymptotic properties of stochastic approximation algorithms have been analyzed for violated Kushner-Clark conditions in this paper. First, the algorithms with additive noise have been analyzed for the case where the noise is state-dependent (Section 2). The obtained results have then been applied to the analysis of the algorithms with non-additive noise (Section 3). On the basis of these general results, the temporal-difference learning algorithms have been analyzed for the case of a general non-linear function approximation (Section 4). The general results have also been illustrated by an example where the noise is non-additive, correlated and satisfies strong mixing conditions (Section 3).

The results of this paper can be considered as an extension and generalization of the results of [1], [4] – [8] and [15]. The results of Sections 2 and 3 have been obtained under stability conditions which are more general than those of [1] and [4] – [8], while the noise condition adopted in these sections are an extension of the noise conditions of [4] and [5]. Moreover, the results of Sections 2 and 3 have successfully been applied to the analysis of temporal-difference learning algorithms with non-linear function approximation. On the other hand, due to the fact that there does not exist a suitable Lyapunov function for these learning algorithms in a general case, the previously available results on the asymptotic properties of stochastic approximation are not applicable to their analysis. Furthermore, the previously available convergence results on temporal-difference learning are constrained to the case where the approximator is linear and where the underlying Markov chain is geometrically ergodic, while the assumptions of Section 4 require the approximator only to be sufficiently smooth and allows the underlying chain to be positive Harris.

References

[1] M. Benaim, "A dynamical system approach to stochastic approximation," *SIAM Journal on Control and Optimization*, vol. 34, pp. 437–472, 1996.

[2] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*, Springer Verlag, 1990.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

[4] H.-F. Chen, L. Guo, A.-J. Gao, "Convergence and robustness of the Robins-Monro algorithm truncated at randomly varying bounds," *Stochastic Processes and their Applications*, 1988, vol. 27, pp. 217–231.

[5] H.-F. Chen, A.-J. Gao, "Robustness analysis for stochastic approximation algorithms" *Stochastics and Stochastics Reports*, 1989, vol. 26, pp. 3–20.

[6] H.-F. Chen, "Recent developments in stochastic approximation," in *Preprints of the 13th IFAC World Congress*, 1996, vol. C, pp. 375–380.

[7] B. Delyon, "General results on the convergence of stochastic approximation," *IEEE Transactions on Automatic Control*, vol. 41, pp. 1245–1255, 1996.

[8] J.-C. Fort and G. Pages, "Convergence of stochastic approximation algorithms: from the Kushner-Clark theorem to the Lyapunov functional method," *Advances of Applied Probability*, vol. 28, pp. 1072–1094, 1996.

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

[10] J. P. LaSalle, *The Stability of Dynamical Systems*, Society for Industrial and Applied Mathematics, 1976.

[11] L. Ljung, G. Pflug, and H. Walk, *Stochastic Approximation and Optimization of Random Systems*, Birkhäuser Verlag, 1992.

[12] M. Metivier and P. Priouret, "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms," *IEEE Transactions on Information Theory*, vol. 30, pp. 140–151, 1984.

[13] R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[14] V. Tadić, "Asymptotic analysis of stochastic iterative algorithms under violated Kushner-Clark conditions with applications to machine learning," unpublished manuscript.

[15] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Transactions on Automatic Control*, vol. 42, pp. 674–690, 1997.

[16] I.-J. Wang, E. K. P. Chong, and S. R. Kulkarni, "Equivalent and sufficient conditions on noise sequences for stochastic approximation algorithms," *Advances in Applied Probability*, vol. 28, pp. 784–801, 1996.