Quasi Stochastic Approximation

Darshan Shirodkar and Sean Meyn

***Abstract*— In recent work it was shown that a deterministic analog of stochastic approximation can be formulated to obtain a Q-learning algorithm for approximate optimal control of deterministic and stochastic systems. This paper provides a general foundation for "quasi-stochastic approximation" in which all of the processes under consideration are deterministic, much like quasi-Monte-Carlo for variance reduction in simulation.**

Applications to root finding and to TD-learning are described, and numerical results are presented.

Acknowledgement: Financial support from the AFOSR grant FA9550-09-1-0190 is gratefully acknowledged. The authors thank Profs. Vivek Borkar and Prashant Mehta for many useful discussions.

I. Introduction

The stochastic approximation algorithm is a specially constructed stochastic difference equation with diminishing step sizes. It was introduced in the classic paper of Robbins and Monro [12] as a means to calculate the root of a function, in a setting in which observations of evaluations of the function are corrupted by noise. More specifically, suppose that we wish to solve the equation $\overline{h}(\vartheta) = 0$, with $\vartheta \in \mathbb{R}^d$, and where $\overline{h}\colon \mathbb{R}^d \to \mathbb{R}^d$ is nonlinear. For a given value of ϑ, we can obtain the noise-corrupted value $h(\vartheta, \zeta)$, where ζ is a random variable and

$$\overline{h}(\vartheta) = \mathsf{E}[h(\vartheta, \zeta)]\,.$$

The stochastic approximation algorithm is then given by the recursion,

$$\vartheta_{n+1} = \vartheta_n + a_n h(\vartheta_n, \zeta_n)\,, \qquad n \geq 0, \tag{1}$$

in which $\vartheta_0 \in \mathbb{R}^d$ is given, and ζ_n is identical to ζ in distribution (typically it is assumed to be i.i.d.). Under some general assumptions, the sequence $\boldsymbol{\vartheta}$ converges with probability 1 to a point ϑ^*, where $\overline{h}(\vartheta^*) = 0$. See [2], [7]; an excellent recent reference is [5].

Stochastic approximation has found numerous applications in a wide range of fields. Although its essential purpose in many of these applications is the calculation of roots, the variety stems from the way in which the function $\overline{h}$ is derived and also from the source of randomness. For example, in an application to stochastic control, $\overline{h}(\vartheta)$ is the gradient of the mean square error in a value function approximation, and ζ represents some inherent randomness in the system. This is the underlying idea in Q-learning and TD-learning [3], [5], [9] (see also the example in Section IV-A).

One of the main desirable features of stochastic approximation is that it can be implemented without knowledge of the function h or of the probability distribution of ζ, as long as we have access to the sequence $h(\vartheta_n, \zeta_n)$. This may indeed be the case if the observations $h(\vartheta_n, \zeta_n)$ are derived from some physical process or experiment.

In other applications of this technique, the function h and the distribution of ζ are known, but computation of $\overline{h}$ is infeasible or expensive.[1] That is, we know everything about the system, yet we introduce uncertainty in the algorithm in order to avoid the computational expense of calculating $\overline{h}$. In these cases, stochastic approximation may be regarded as an approach to numerical integration, much like MCMC.

Stochastic approximation can also be regarded as a generalization of the Monte-Carlo algorithm for estimating the mean of a random variable ζ: If $\{\zeta_i\}$ is a stationary sequence, each identical to ζ in distribution, then the Monte-Carlo estimate of $\vartheta^* = \mathsf{E}[\zeta]$ is given by,

$$\vartheta_n = \frac{1}{n}\sum_{i=0}^{n-1} \zeta_i \tag{2}$$

This can be written in the recursive form,

$$\vartheta_{n+1} = \vartheta_n + \frac{1}{n+1}\big(\zeta_n - \vartheta_n\big)$$

which is precisely (1) in the special case $a_n = (n+1)^{-1}$, $h(\vartheta, \zeta) = \zeta - \vartheta$.

In the method of quasi-Monte Carlo, the sequence $\{\zeta_i\}$ is chosen to be *deterministic* in the sample path average (2); in this way the rate at which the estimates $\{\vartheta_n\}$ converge to ϑ^* can be accelerated [1]. In this paper we extend this idea to stochastic approximation. Our goal is to reduce the 'curse of variance' observed in these algorithms, and to reduce computational cost by avoiding the generation of random numbers.

The idea of using deterministic sequences in stochastic approximation has its origins in the works [13] and [14], although in a very specific context. In [13], deterministic sufficient conditions are provided for the perturbation sequences in Random Direction Kiefer Wolfowitz (RDKW) algorithms, while [14] contains deterministic necessary and sufficient conditions for convergence of RDKW and Simultaneous Perturbation Stochastic Approximation (SPSA). To our knowledge, [4] is the first paper claiming an improvement in the rate of convergence due to use of deterministic sequences, again in the context of SPSA. They prove convergence but the rate improvement is demonstrated only through

Authors are with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign (UIUC) shirodk1@illinois.edu; meyn@illinois.edu

[1] In several applications we may not care about the exact distribution of ζ. An example is the approximate dynamic programming problem — see Section IV-A.

numerical experiments. Our work can be considered to be a generalization of these earlier works in a way, though not exactly since our analysis is in continuous time, the justification for which is given next.

Motivation for this work came in part from the Q-learning algorithm for approximate optimal control introduced in [9]. In this prior work the parameter estimate ϑ_t evolves in continuous time. To provide a foundation for [9] and because of its mathematical elegance, in this paper we introduce a differential equation instead of a difference equation in which the 'noise' comes from a deterministic oscillatory signal rather than from a stochastic process.

The differential equation that defines the quasi-stochastic approximation (QSA) algorithm has the general form

$$\tfrac{d}{dt}\vartheta(t) = a(t) f(\vartheta(t), \xi(t)) \,. \tag{3}$$

The non-negative function of time $\boldsymbol{a}$ is the 'step size', and $\boldsymbol{\xi}$ is an m-dimensional process that constitutes the 'driving noise'. As in the classical stochastic approximation algorithm, our analysis is based on consideration of the associated ODE,

$$\frac{d}{du}\bar{\theta}(u) = \bar{f}(\bar{\theta}(u)), \tag{4}$$

in which the 'averaged' vector field is given by

$$\bar{f}(\theta) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(\theta, \xi(t))\, dt, \quad \text{for all } \theta \in \mathbb{R}^d. \tag{5}$$

In most applications the process $\boldsymbol{\xi}$ will be constructed so that it is deterministic, but ergodic. Under appropriate conditions on f, this implies that the limit in (5) is defined by an invariant measure π for $\boldsymbol{\xi}$ via

$$\bar{f}(\theta) = \int f(\theta, y)\, \pi(dy) \,. \tag{6}$$

We envision processes $\boldsymbol{\xi}$ obtained as mixtures of sinusoids, or other periodic signals. In this case continuity of f is sufficient to obtain (5) and (6).

The extension of stability and convergence results from the classical stochastic model (1) to the deterministic analog (3) requires some specialized analysis since the standard methods are not directly applicable. In particular, the first step in [5] and most other references is to write (1) in the form,

$$\vartheta_{n+1} = \vartheta_n + a_n\big(\overline{h}(\vartheta_n) + M_n\big) \,,$$

where $\boldsymbol{M}$ is a martingale difference sequence (or a perturbation of such a sequence). This is possible when $\boldsymbol{\zeta}$ is i.i.d., or for certain Markov $\boldsymbol{\zeta}$ [8]. A similar transformation is not possible for any class of deterministic $\boldsymbol{\xi}$.

In addition to convergence, we establish the rate of convergence of $\boldsymbol{\vartheta}$ to its limit ϑ^*. As in [6], we simplify analysis by taking h to be a linear function of its arguments — this can be justified via a Taylor-series approximation once convergence is established. In the classical algorithm, an analog of the Central Limit Theorem holds, so that the rate of convergence is $O(t^{-\frac{1}{2}})$. In this deterministic setting, we find that the convergence is $O(t^{-1})$. Shown in Figure 1 are two

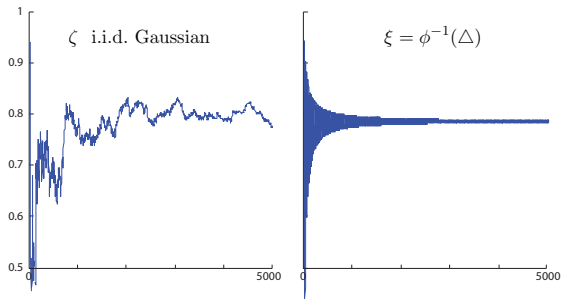

Fig. 1. Finding roots of a function: Comparison of SA and QSA

plots comparing results from stochastic approximation and quasi-stochastic approximation for a particular root-finding application described in Section IV-B. The results from the stochastic recursion on the left converge much more slowly than the output of the QSA algorithm shown on the right.

The remainder of the paper is organized as follows. Stability and convergence of the QSA process is established in Section II under general conditions on the model mirroring those imposed in the usual stochastic approximation algorithm. Section III contains the results on the rate of convergence of QSA, and several applications are presented in Section IV. Conclusions and thoughts on future research are contained in Section V.

II. Convergence of QSA

In the ODE approach for analysis of stochastic approximation algorithms, a continuous time trajectory $\boldsymbol{\theta}$ is obtained from the sequence defined in (1) through linear interpolation: The n^{th} iterate ϑ_n in the recursion corresponds to a point $\theta(u_n)$ in the interpolated trajectory, where

$$u_n = \sum_{i=0}^{n} a_i.$$

The piecewise linear trajectory $\boldsymbol{\theta}$ is then compared with $\bar{\boldsymbol{\theta}}$ that is defined in (4).

A similar construction is used in the analysis here: we substitute in (3) the new time variable u given by

$$u = g(t) := \int_0^t a(r)\, dr, \qquad t \ge 0.$$

The time-scaled process is then defined by $\theta(u) := \vartheta(g^{-1}(u))$. We will shortly impose assumptions on $\boldsymbol{a}$ to ensure g^{-1} is well defined, nonnegative and monotonically increasing, with $g^{-1}(u) \to \infty$ as $u \to \infty$. For example, if $a(r) = (1+r)^{-1}$, then

$$u = \log(1+t) \quad \text{and} \quad \xi(g^{-1}(u))) = \xi(e^u - 1). \tag{7}$$

The chain rule of differentiation gives

$$\frac{d}{du}\vartheta(g^{-1}(u)) = f(\vartheta(g^{-1}(u)), \xi(g^{-1}(u))).$$

That is, the time-scaled process solves the differential equation,

$$\frac{d}{du}\theta(u) = f(\theta(u), \xi(g^{-1}(u))). \tag{8}$$

The two processes $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ differ only in time scale, and hence, proving convergence of one proves that of the other. For the remainder of this section we will deal exclusively with $\boldsymbol{\theta}$ — It is on the 'right' time scale for comparison with $\bar{\boldsymbol{\theta}}$, the solution of (4).

We begin with resolving stability of the algorithm: In the following subsection, we establish boundedness of the trajectory $\boldsymbol{\theta}$ for each initial condition. This is the main requirement in the convergence proof provided in Section II-B. We first state the assumptions under which our analysis is valid:

(A1) The system described by equation (4) has a globally asymptotically stable equilibrium at $\bar{\theta} = \vartheta^*$.

(A2) There exists a continuous function $V : \mathbb{R}^d \to \mathbb{R}_+$ and a constant $c_0 > 0$ such that, for $0 \le T \le 1$, $\|\bar{\theta}(s)\| > c_0$,

$$V(\bar{\theta}(s+T)) - V(\bar{\theta}(s)) \le -T\|\bar{\theta}(s)\|.$$

(A3) There exists a function $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ and a process $\{\xi(t)\}_{t\ge 0}$ that takes values in a compact set $\Omega \subset \mathbb{R}^m$ such that, for some $b_0 < \infty$, and all $\theta \in \mathbb{R}^d$, $T > 0$,

$$\left\| \frac{1}{T}\int_0^T f(\theta, \xi(u))\, du - \bar{f}(\theta) \right\| \le \frac{b_0}{T}(1 + \|\theta\|)$$

(A4) There exists a constant $\ell < \infty$ such that the functions V, $\bar{f}$ and f satisfy the following Lipschitz conditions:

$$\begin{aligned} \|V(\theta') - V(\theta)\| &\le \ell\|\theta' - \theta\|, \\ \|\bar{f}(\theta') - \bar{f}(\theta)\| &\le \ell\|\theta' - \theta\|, \\ \|f(\theta', \xi) - f(\theta, \xi)\| &\le \ell\|\theta' - \theta\|, \qquad \theta',\ \theta \in \mathbb{R}^d. \end{aligned}$$

(A5) The process $\boldsymbol{a}$ is non-negative and monotonically decreasing, and as $t \to \infty$,

$$a(t) \downarrow 0, \qquad \int_0^t a(r)\, dr \to \infty.$$

Assumption (A1) uniquely determines the solution point to which we wish to converge. Assumption (A2) ensures that there is a Lyapunov function V with a strictly negative drift whenever $\bar{\boldsymbol{\theta}}$ escapes a ball of radius c_0. This assumption is required to establish boundedness of the trajectory. Assumptions (A3) and (A4) are technical requirements essential to the proofs; (A3) is only slightly stronger than ergodicity of $\boldsymbol{\xi}$ as given by (5), while (A4) is necessary to control the growth of the respective functions. The process $\boldsymbol{a}$ in (A5) is a continuous time counterpart of the standard step size schedules in stochastic approximation, except that we impose monotonicity in place of square integrability.

A. Stability

Theorem 2.1: The solution to (8) is ultimately bounded: for some $b < \infty$ and for any $\theta(0) = \theta$, there is a $T_\theta < \infty$ such that $\|\theta(t)\| \le b$ for all $t \ge T_\theta$.

We prove Theorem 2.1 by establishing a 'drift condition' similar to (A2) for $\boldsymbol{\theta}$. To do so, we compare the trajectory of $\boldsymbol{\theta}$ with that of $\bar{\boldsymbol{\theta}}$, initialized at some large time s to $\theta(s)$. To this end, let us define $\bar{\theta}^s(t)$, $t \ge s$, to be the unique solution to (4) 'starting' at s:

$$\tfrac{d}{dt}\bar{\theta}^s(t) = \bar{f}(\bar{\theta}^s(t)),\ t \ge s, \qquad \bar{\theta}^s(s) = \theta(s). \tag{9}$$

Lemma 2.4 bounds the difference between $\boldsymbol{\theta}$ and $\bar{\boldsymbol{\theta}}^s$. This bound is then used to establish a drift condition for $\boldsymbol{\theta}$. But first, we assume the required drift condition and prove ultimate boundedness of $\boldsymbol{\theta}$.

Lemma 2.2: The solution to (8) is ultimately bounded if for some $0 < T \le 1$, and $s_0, b < \infty$,

$$V(\theta(s+T)) - V(\theta(s)) \le -T\|\theta(s)\|, \qquad \text{for all } s \ge s_0,\ \|\theta(s)\| > b.$$

Proof: Suppose for some $\theta(0) = \theta$, there is $s \ge s_0$ such that $\|\theta(s)\| > b$. Let $\tau := \min(u \ge 0 : \|\theta(s+u)\| \le b)$; if $\|\theta(s+u)\| > b$ for all $u \ge 0$, set $\tau = \infty$. For $m \in \mathbb{N}$, define $\tau_m = \min\{\tau, m\}$. Then,

$$\begin{aligned} -\tau_m b &\ge -T\int_s^{s+\tau_m} \|\theta(u)\| du \\ &\ge \int_s^{s+\tau_m} (V(\theta(u+T)) - V(\theta(u)))\, ds \\ &= \int_{s+\tau_m}^{s+\tau_m+T} V(\theta(r))\, dr - \int_s^{s+T} V(\theta(r))\, dr \\ &\ge -\int_s^{s+T} V(\theta(r))\, dr. \end{aligned}$$

This establishes a uniform bound on τ_m for all $m \in \mathbb{N}$, thus proving $\tau < \infty$. Also, we have the following bound on the growth of $\boldsymbol{\theta}$ (see Appendix, Section B): For some $b_1 < \infty$,

$$\|\theta(s+u) - \theta(s)\| \le b_1 u\|\theta(s)\| \le b_1\tau\|\theta(s)\| \quad \text{for } u \le \tau.$$

Thus, we have $\|\theta(s)\| \le b(1 + b_1\tau), \quad s \ge s_0$. This proves ultimate boundedness of $\boldsymbol{\theta}$. ∎

Next, we prove a Law of Large Numbers (LLN) for the time scaled process $\{\xi(g^{-1}(u))\}_{u\ge 0}$. Notice the difference from a conventional LLN. Here, the interval of integration is some arbitrary fixed T, and the averaging becomes more accurate as the interval is shifted towards infinity.

Lemma 2.3: For any $s, T > 0$, $\|\theta\| \ge 1$, the function f satisfies the following bound:

$$\left\| \frac{1}{T}\int_s^{s+T} f(\theta, \xi(g^{-1}(u)))\, du \ - \ \bar{f}(\theta) \right\| \le 4b_0\varepsilon(s)\|\theta\|/T, \tag{10}$$

where $\varepsilon(s) \to 0$ as $s \to \infty$.

Proof: Define

$$\hat{f}(\theta, t) := \frac{1}{t}\int_0^t f(\theta, \xi(r))\, dr - \bar{f}(\theta).$$

Note that by assumption, for $\|\theta\| > 1$,

$$\|\hat{f}(\theta, t)\| \le b_0(1 + \|\theta\|)/t \le 2b_0\|\theta\|/t. \tag{11}$$

Consider the following integral which we simplify using integration by parts:

$$\begin{aligned}\int_{t_0}^{t_1} a(t)f(\theta,\xi(t))\,dt &= \Big[a(t)\int_0^t f(\theta,\xi(r))\,dr\Big]_{t_0}^{t_1} \\ &\quad - \int_{t_0}^{t_1} a'(t)\int_0^t f(\theta,\xi(r))\,drdt \\ &= [t_1 a(t_1)\hat{f}(\theta,t_1) - t_0 a(t_0)\hat{f}(\theta,t_0)] \\ &\quad - \int_{t_0}^{t_1} ta'(t)\hat{f}(\theta,t)\,dt \\ &\quad + (g(t_1)-g(t_0))\bar{f}(\theta)\end{aligned}$$

Rearranging and taking norms, we obtain on applying (11) and after some algebra,

$$\begin{aligned}&\Big\|\int_{t_0}^{t_1} a(t)f(\theta,\xi(t))\,dt - (g(t_1)-g(t_0))\bar{f}(\theta)\Big\| \\ &\le 4b_0 a(t_0)\|\theta\|.\end{aligned}$$

We have used the fact that $a(t)$ is decreasing and so $-a'(t) \ge 0$. Letting $t_0 = g^{-1}(s),\ t_1 = g^{-1}(s+T)$ and $t = g^{-1}(u)$ yields

$$\begin{aligned}&\Big\|\frac{1}{T}\int_s^{s+T} f(\theta,\xi(g^{-1}(u)))\,du \ - \ \bar{f}(\theta)\Big\| \\ &\le 4b_0 a(g^{-1}(s))\|\theta\|/T.\end{aligned}$$

Set $\varepsilon(s) := a(g^{-1}(s))$. As $s\to\infty$, $g^{-1}(s)\to\infty$ and hence, $\varepsilon(s)\to 0$. This completes the proof. ■

Lemma 2.4: For some $b<\infty$ and any $0\le T\le 1$,

$$\|\theta(s+T) - \bar{\theta}^s(s+T)\| \le (bT^2 + 4b_0\varepsilon(s))\|\theta(s)\|. \quad (12)$$

Proof: See Apendix. ■

Proof: [Proof of Theorem 2.1] Recall that V is the Lyapunov function assumed in (A2). We now prove a drift condition for $\boldsymbol{\theta}$. For $0\le T\le 1$, $\|\theta(s)\|\ge 1$,

$$\begin{aligned}V(\theta(s+T)) - V(\theta(s)) &= V(\theta(s+T)) - V(\bar{\theta}^s(s+T)) \\ &\quad + V(\bar{\theta}^s(s+T)) - V(\bar{\theta}^s(s)) \\ &\le |V(\theta(s+T)) - V(\bar{\theta}^s(s+T))| \\ &\quad + V(\bar{\theta}^s(s+T)) - V(\bar{\theta}^s(s)) \\ &\le \ell\|\theta(s+T) - \bar{\theta}^s(s+T)\| - T\|\theta(s)\| \\ &\le \ell(bT^2 + 4b_0\varepsilon(s))\|\theta(s)\| - T\|\theta(s)\|,\end{aligned}$$

where the second inequality follows from the Lipschitz assumption on V and the last inequality uses Lemma 2.4. Let us choose T small enough to make $2\ell bT^2 \le T/2$, and then s_0 large enough so that $4b_0\varepsilon(s)\le bT^2$ for all $s\ge s_0$, which leads to

$$V(\theta(s+T)) - V(\theta(s)) \le -\frac{T}{2}\|\theta(s)\|.$$

Lemma 2.2 completes the proof. ■

B. Convergence

We now show that the solution to (8) converges to ϑ^*, the unique asymptotically stable equilibrium point of (4). First, we present a variation of Lemma 2.3. A proof is included in the Appendix.

Lemma 2.5: For any $T>0$,

$$\lim_{s\to\infty}\sup_{t\in[0,T]}\Big\|\int_s^{s+t}\big(f(\theta(u),\xi(g^{-1}(u))) - \bar{f}(\theta(u))\big)\,du\Big\| = 0.$$

■

The next result is very similar to Lemma 1 in Chapter 2 of [5].

Lemma 2.6: For any $T>0$,

$$\lim_{s\to\infty}\sup_{t\in[0,T]}\|\theta(s+t) - \bar{\theta}^s(s+t)\| = 0.$$

Proof: Use Lemma 2.5 and Gronwall inequality. ■

Theorem 2.7: For any initial condition $\theta(0)=\theta$, the solution to (8) converges to ϑ^*.

Proof: By ultimate boundedness of $\theta(u)$, there exists $b<\infty$ such that for $\theta(0)=\theta$, there is a T_θ such that

$$\|\theta(u)\| \le b \ \text{ for all } u\ge T_\theta.$$

Thus, for $s\ge T_\theta$, $\|\bar{\theta}^s(s)\| = \|\theta(s)\|\le b$. By the definition of global asymptotic convergence, for every $\epsilon>0$, there exists a $T^\epsilon>0$, independent of the initial condition $\bar{\theta}^s(s)$, such that

$$\|\bar{\theta}^s(s+u) - \vartheta^*\| < \epsilon \ \text{ for all } u\ge T^\epsilon.$$

From Lemma 2.6, we have

$$\begin{aligned}&\limsup_{s\to\infty}\|\theta(s+T^\epsilon) - \vartheta^*\| \\ &\le \limsup_{s\to\infty}\|\theta(s+T^\epsilon) - \bar{\theta}^s(s+T^\epsilon)\| \\ &\quad + \limsup_{s\to\infty}\|\bar{\theta}^s(s+T^\epsilon) - \vartheta^*\| \\ &\le 0+\epsilon.\end{aligned}$$

Since ϵ is arbitrary, we have the desired limit,

$$\lim_{u\to\infty}\|\theta(u) - \vartheta^*\| = \lim_{s\to\infty}\|\theta(s+T^\epsilon) - \vartheta^*\| = 0.$$

This completes the proof. ■

III. Rates of convergence

The goal in this section is to establish a rate of convergence for the QSA algorithm. In the stochastic approximation algorithm it is known that, under appropriate assumptions, the algorithm satisfies a Central Limit Theorem, generalizing the usual CLT for the Monte-Carlo recursion (2). That is, in the stochastic model (1) we can write,

$$\vartheta_n \approx \vartheta^* + n^{-\frac{1}{2}}W \quad (13)$$

where the approximation is in distribution, and W is a Gaussian random variable with some finite covariance matrix Σ [5]. The proof of this result is based on comparing the stochastic approximation recursion with a linearization about ϑ^*. In this section we take this step for granted, assuming

that the function given in (A3) is in fact linear. A similar simplification is made in the analysis of two-time scale stochastic approximation in [6]. This and further assumptions are collected together here:

(A6) The function f is linear, $f(\theta,\xi) = A\theta + \xi$, and moreover

(i) A is Hurwitz, and each eigenvalue $\lambda(A)$ satisfies $\text{Re}(\lambda) < -1$.

(ii) $\boldsymbol{\xi}$ has zero mean: $\lim_{T\to\infty} \frac{1}{T}\int_0^T \xi(t)\,dt = 0 \in \mathbb{R}^d$.

Assumption (A6) implies $\bar{f}(\theta) = A\theta$, and the unique stable equilibrium point is $\vartheta^* = 0$.

Observe that the assumption (i) of (A6) is slightly stronger than what is imposed in the usual Central Limit Theorem for stochastic approximation, which requires $\text{Re}(\lambda) < -\frac{1}{2}$. We will see that this stronger bound is indeed required to obtain the rate of convergence sought in this section, which is a significant improvement on (13). Under the assumptions of this section we show that for some constant $\bar{\sigma} < \infty$ we have,

$$\limsup_{t\to\infty} t\|\vartheta(t) - \vartheta^*\| \le \bar{\sigma} \tag{14}$$

Theorem 3.1: Suppose that assumptions (A1)-(A6) hold, and that $a(t) = 1/(1+t)$. Then $\boldsymbol{\vartheta}$ converges to 0 at a rate of $1/(1+t)$. That is, (14) holds with $\vartheta^* = 0$.

The proof is given at the end of this section.

Observe that the eigenvalue condition in (A6) can be achieved by scaling: That is, replacing $f(\theta,\xi) = A\theta + \xi$ by $\kappa(A\theta + \xi)$ for a sufficiently large positive constant κ. However, this might result in poor transient behavior. An alternative is to use the two-time-scale approach of Polyak and Juditsky [11], [5]. We propose the following continuous-time counterpart of their algorithm: Fix $\delta \in (0,1)$, and consider the algorithm with $a(t) = 1/(1+t)^\delta$:

$$\frac{d}{dt}\gamma(t) = \frac{1}{(1+t)^\delta} f(\gamma(t),\xi(t)) \tag{15}$$

The output of this algorithm is then averaged as follows:

$$\frac{d}{dt}\vartheta(t) = \frac{1}{1+t}\big(-\vartheta(t) + \gamma(t)\big) \tag{16}$$

Theorem 3.2: Suppose that assumptions (A1)-(A6) hold, so that in particular $f(\gamma(t),\xi(t)) = A\gamma(t) + \xi(t)$ in (15). Then (14) holds with $\vartheta^* = 0$ for the averaged process defined in (16).

The proof of Theorem 3.2 is skipped due to space constraints — It is similar to, but messier than the proof of Theorem 3.1.

Proof of Theorem 3.1: We can use the transformation $u = \log(1+t)$ as before, and equivalently prove

$$\sup_{u\in[0,\infty)} \|e^u\theta(u)\| < \infty. \tag{17}$$

The linear system in the transformed time scale is given by

$$\dot{\theta}(u) = A\theta(u) + \xi(e^u - 1).$$

Let $z(u) := e^u\theta(u)$. Then

$$\dot{z}(u) = (A+I)z(u) + e^u\xi(e^u - 1).$$

For notational convenience, let us define $B := A + I$. Then B is Hurwitz because of the assumption on eigenvalues of A. From linear system theory, we know that

$$z(u) = e^{Bu}\cdot z(0) + \int_0^u e^{B(u-v)}\cdot e^v\xi(e^v - 1)dv.$$

Clearly, the first term decays to zero and hence, is bounded. To show that the second term is also bounded, we use integration by parts and denote $\beta(v) = \int_0^v e^r\xi(e^r - 1)dr$ to obtain

$$\int_0^u e^{B(u-v)}\cdot e^v\xi(e^v - 1)dv = \Big[e^{B(u-v)}\beta(v)\Big]_0^u + \int_0^u e^{B(u-v)}\cdot B\beta(v)dv. \tag{18}$$

Note that by changing the variable of integration to $t = e^r - 1$ in the expression for $\beta(v)$, we obtain

$$\beta(v) = \int_0^{e^v-1} \xi(t)dt,$$

which is bounded as a function of v. Thus, the first term in (18) is bounded. The second term is just the response of the linear system represented by B to the input $B\beta(v)$. Since B is Hurwitz, it represents a BIBO stable system and $B\beta(v)$ is a bounded input. So the second term is also bounded and this completes the proof. ■

IV. Applications

A. Approximating a value function

In this section, we consider an application of the algorithm to approximation of solutions of differential equations by a linear combination of basis functions as in TD-learning [3], [10]. As a specific example, we look at an uncontrolled diffusion process with state dependent cost. Our goal is to approximate an associated discounted-cost value function. We choose basis functions based on heuristic arguments and provide numerical results to demonstrate the accuracy of approximation.

Consider the following scalar diffusion,

$$dX_t = -X_t^3 dt + dW_t,$$

where $\boldsymbol{W}$ is a standard Brownian motion. For a given cost function $c\colon \mathbb{R} \to \mathbb{R}_+$ and discount factor $\gamma > 0$, the value function is given by

$$V(x) = \mathsf{E}\Big[\int_{t=0}^\infty e^{-\gamma t}c(X_t)\,dt \mid X_0 = x\Big].$$

Subject to growth conditions on c, it can be shown that V satisfies the following differential equation:

$$\gamma V(x) = c(x) + \mathcal{D}V(x), \quad x \in \mathbb{R}, \tag{19}$$

where the differential generator is given by $\mathcal{D}V := (-x^3)V' + \frac{1}{2}V''$.

The *discounted-cost dynamic programming equation* (19) could of course be directly solved using numerical integration. To demonstrate the use of our algorithm, we seek to approximate V by a linear combination of basis functions, denoted $\{\phi_1, \phi_2, ..., \phi_d\}$. For $\vartheta \in \mathbb{R}^d$, the approximation to V is given by

$$V_\vartheta(x) = \sum_{i=1}^{d} \vartheta_i \phi_i(x). \tag{20}$$

We want to find ϑ such that $V_\vartheta(x)$ satisfies (19) 'as closely as possible'.

To quantify this approximation we use the *Bellman error* [3]. For $x \in \mathbb{R}$, $\vartheta \in \mathbb{R}^d$ this is given by,

$$\mathcal{L}(x, \vartheta) := \gamma V_\vartheta(x) - c(x) + \mathcal{D}V_\vartheta(x). \tag{21}$$

On denoting $c_\vartheta(x) = \mathcal{L}(x, \vartheta) + c(x)$, $x \in \mathbb{R}$, we see that V_ϑ solves the dynamic programming equation,

$$\gamma V_\vartheta(x) = c_\vartheta(x) + \mathcal{D}V_\vartheta(x), \quad x \in \mathbb{R}.$$

Subject to general conditions, this implies that V_ϑ is the value function associated with c_ϑ. Hence we obtain a good approximation if $c \approx c_\vartheta$, which means that the Bellman error is small.

As a scalar mismatch criterion we consider the mean-square Bellman error: We assume that a probability measure μ on $\mathbb{R}$ is given, and denote

$$\mathcal{E}_{\text{Bell}}(\vartheta) := \int \big(\mathcal{L}(x, \vartheta)\big)^2 \mu(dx) \tag{22}$$

The choice of μ will affect the approximation: If we obtain a small mean-squared error, then we can expect the approximation $V_\vartheta \approx V$ to be reasonably good over the support of μ. In the experiments described below we took μ to be a uniform distribution on a finite interval.

Next, we provide a rationale for our choice of basis functions. We now fix the cost function to be $c(x) = x^2$. In this case the cost function as well as the system are symmetric about the origin. So the basis functions should be even functions of x. For large x, the solution to (19) can be approximated by $\log(|x|)$. To make the function well defined at the origin, we can use $\frac{1}{2}\log(1 + x^2)$. For x near the origin, the term $(-x^3)V'(x)$ is neglible. With this term neglected, the solution is of the form $c_1 x^2 + c_2$. To make this part of the solution bounded for large x, we propose $x^2/(1 + x^4)$ and 1 as basis functions. In summary, we have

$$\phi_1(x) = \log(1+x^2),\ \phi_2(x) = x^2/(1+x^4),\ \text{ and } \ \phi_3(x) \equiv 1.$$

In particular, if $\vartheta = (\frac{1}{2}, 0, 0)^\intercal$, then $V_\vartheta(x) = \frac{1}{2}\log(1 + x^2)$, and in this case simple calculations show that $|\mathcal{L}(x, \vartheta)|$ is bounded by $\frac{1}{2}\gamma \log(1 + x^2)$ plus a constant.

To apply our algorithm, we denote

$$f(\vartheta, x) := \nabla \mathcal{E}_{\text{Bell}}(\vartheta, x), \ \ \bar{f}(\vartheta) := \nabla_\vartheta \bar{B}_e(\vartheta) = \int_{\mathbb{R}} f(\vartheta, x)\, \mu(dx)$$

Our goal is to find $\vartheta^* \in \mathbb{R}^3$ that solves $\bar{f}(\vartheta^*) = 0$. We need a process $\{\xi(t)\}_{t \ge 0}$ that satisfies

$$\bar{f}(\vartheta) = \lim_{T \to \infty} \frac{1}{T} \int_{t=0}^{T} f(\vartheta, \xi(t))\, dt, \quad \vartheta \in \mathbb{R}^3.$$

In fact, we need something stronger, since we require the assumptions of (A3). Since μ is uniform on some interval $[-I, I]$, we chose $\boldsymbol{\xi}$ to be a triangular wave with amplitude I. Let $\boldsymbol{\Delta}$ denote the non-negative triangular wave process with period τ:

$$\Delta(t) = \begin{cases} 2t/\tau, & t \le \tau/2 \mod \tau, \\ 2 - 2t/\tau, & t > \tau/2 \mod \tau. \end{cases} \tag{23}$$

The process $\boldsymbol{\xi}$ was taken to be the scaled process,

$$\xi(t) = 2I(\Delta(t) - \tfrac{1}{2}) \quad t \ge 0,$$

We chose the period $\tau = 40$, a discount factor $\gamma = 0.5$, and stepsize $a(t) = 1/(1 + t)$.

Due to space constraints, we did not display the plot showing convergence of trajectories of ϑ_1, ϑ_2 and ϑ_3. The parameters converge to the solution $(\vartheta_1^*, \vartheta_2^*, \vartheta_3^*) = (0.49, -0.02, 0.81)$. The error $\mathcal{E}_{\text{Bell}}(\vartheta^*)$ was numerically computed to be 0.008, which suggests an extremely close approximation. The approximation at large x taken to be $\frac{1}{2}\log(1+x^2)$ is corroborated by the coefficient of $\vartheta_1^* = 0.49$ for the $\phi_1(x) = \log(1 + x^2)$ basis term.

Shown on the left in Figure 2 is the Bellman error obtained using ϑ^*. For comparison we include results from two other experiments in which only the distribution μ was varied through choice of I: In the central figure we took $I = 5$, and in the final figure $I = 1$, giving μ uniform on, respectively, $[-5, 5]$ and $[-1, 1]$. The plots show that the approximation is good on the support of μ, and may be very poor outside of this support. Moreover, with a smaller support we obtain a tighter approximation on the support.

B. Finding roots of a function

As mentioned in the Introduction, the original motivation for stochastic approximation was finding the roots of a function based on its noisy measurements. In particular, we can solve $\bar{f}(x) = 0$ using only noise corrupted observations $f(x, \zeta)$ where

$$\bar{f}(x) = \mathsf{E}[f(x, \zeta)], \quad \text{for all } x \in \mathbb{R}^d.$$

Let G be the distribution function of the random variable ζ. One of the advantages of stochastic approximation is that we can apply it without the knowledge of G. In many applications however, the complete statistics are known, but the the main hindrance to knowing $\bar{f}$ is computational: Stochastic approximation is used to approximate the expectation, by generating a sequence of independent random variables $\{\zeta_n\}$ according to G. An alternative is to make use of the deterministic algorithm given by (3) for a suitable process $\xi(t)$.

To illustrate the application of this approach, we consider a model with ζ a scalar valued random variable. Observe that

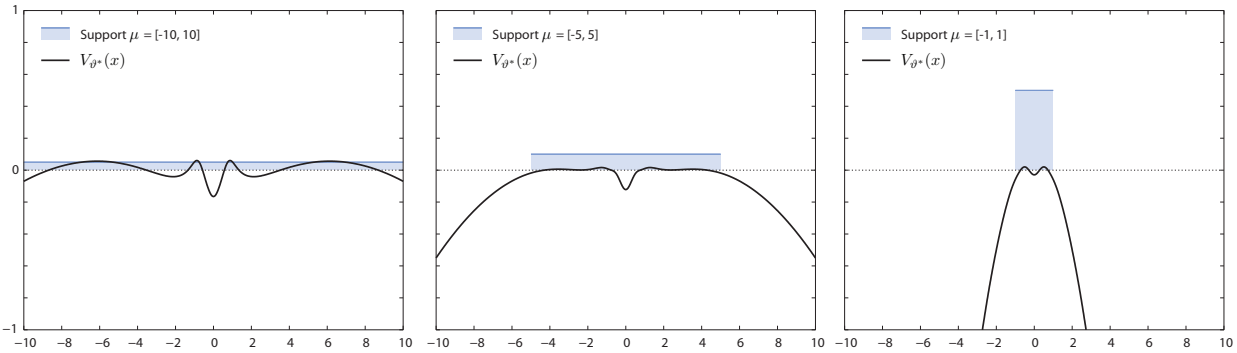

Fig. 2. Learning where you look: Bellman error is small on the support of μ

the process $\boldsymbol{\Delta}$ defined in (23) takes on values uniformly over $[0, 1]$, for any value of τ. We can then use $\xi(t) = G^{-1}(\Delta(t))$ in (3) to obtain a process with the desired distribution: $\boldsymbol{\xi}$ has distribution G in the sense that for any continuous function $h\colon \mathbb{R} \to \mathbb{R}$,

$$\mathsf{E}[h(\zeta)] = \lim_{T\to\infty} \frac{1}{T} \int_{t=0}^{T} h(\xi(t))\, dt.$$

However, for the purpose of comparison with stochastic approximation, we used the following discrete time version of (3):

$$\vartheta_{n+1} = \vartheta_n + a_n f(\vartheta_n, \xi_n). \tag{24}$$

The specific function used for the numerical experiment was $f(\vartheta, \zeta) = -\tan(\vartheta) + \zeta$, where ζ was taken to be a normal random variable with mean $\mu = 1$ and standard deviation $\sigma = 3$. Thus, $\bar{f}(\vartheta) = \mathsf{E}[-\tan(\vartheta) + \zeta] = -\tan(\vartheta) + 1$, which gives the equilibrium point $\vartheta^* = \pi/4$. The step size schedule used in both (1) and (24) was $a_n = 1/(n+1)$. The samples ξ_n were generated using $\xi_n = \phi^{-1}(\Delta_n)$, where ϕ is the Gaussian distribution function and $\{\Delta_n\}$ is a sampled triangular wave with a period of 40 samples. Of course Δ_n must be truncated so that it is never exactly 0 or 1, which would result in $\phi^{-1}(\Delta_n)$ becoming $-\infty$ or $+\infty$ respectively.

Figure 1 shows a sample trajectory each of SA and QSA. QSA clearly converges faster and more smoothly than SA. Even after $10,000$ samples, SA hasn't exactly converged to the required equilibrium point.

We also experimented with the period of $\boldsymbol{\Delta}$ and it was found that for longer periods, the trajectory was more oscillatory and convergence was slower. This is to be expected since averaging for QSA takes place over a full period; so when we are in the first half of the triangular wave $\boldsymbol{\Delta}$, we have introduced a bias in one direction, which gets nullified during the second half. So it would seem that the shorter the period, the better. But we also need enough resolution of samples over a period. For example, in the extreme case of a period of 2 samples, the samples Δ_n would be all zeros. An intermediate value needs to be found, mainly through experimentation, which may depend on the specific application.

V. Conclusions

This paper was initially motivated by our recent paper [9] concerning Q-learning for nonlinear, deterministic systems. We wrote that a "proof of convergence will be straightforward following standard arguments from stochastic approximation theory". With significantly more work than anticipated this conjectured turned out to be correct.

Our main current interest is the development of algorithms for approximate dynamic programming based on these techniques. In particular,

(i) Techniques for basis selection based on approximate models.

(ii) The special case of large interconnected models in which mean-field theory suggests simplified basis functions [9].

(iii) Extensions to partially observed models, distributed models, and dynamic games.

Appendix

Here we collect together the more technical arguments.

A. Bounds using Gronwall's inequality

The well known Gronwall inequality can be found in any of the stochastic approximation texts referenced, such as [5]. The following result is an immediate consequence.

Lemma 1.1: For $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ satisfying (4) and (8) respectively, there exists $b < \infty$ such that for any $s, u \geq 0$, and $\|\bar{\theta}(s)\|, \|\theta(s)\| \geq 1$, we have

$$\|\bar{\theta}(s+u) - \bar{\theta}(s)\| \leq bue^{\ell u}\|\bar{\theta}(s)\|, \tag{25}$$

$$\|\theta(s+u) - \theta(s)\| \leq bue^{\ell u}\|\theta(s)\|, \tag{26}$$

where ℓ is the Lipschitz constant introduced in (A4).
Proof: Write expressions for the left hand sides, use Lipschitz properties of f and $\bar{f}$ and apply Gronwall inequality. ■

B. Proof of Lemma 2.4

Using Lemma 1.1, we have for some $b < \infty$ and any $0 \leq T \leq 1$,

$$\begin{aligned}\|\bar{\theta}(s+T) - \bar{\theta}(s)\| &\leq bT\|\theta(s)\|e^{\ell T} \\ &\leq b_1 T\|\theta(s)\|, \quad b_1 := be^{\ell}.\end{aligned}$$

Similarly, $\|\theta(s+T)-\theta(s)\| \le b_1 T\|\theta(s)\|$. On combining the above bounds with Lemma 2.3, we obtain

$$\begin{aligned}
&\|\theta(s+T)-\bar\theta^s(s+T)\| \\
&= \|\int_s^{s+T} \left(f(\theta(t),\xi(g^{-1}(t))) - \bar f(\bar\theta^s(t))\right)\, dt\| \\
&\le \|\int_s^{s+T} \left(f(\theta(t),\xi(g^{-1}(t))) - f(\theta(s),\xi(g^{-1}(t)))\right)\, dt\| \\
&\quad + \|\int_s^{s+T} \left(\bar f(\bar\theta^s(t)) - \bar f(\bar\theta^s(s))\right)\, dt\| \\
&\quad + \left\|\int_s^{s+T} \left(f(\theta(s),\xi(g^{-1}(t))) - \bar f(\theta(s))\right)\, dt\right\| \\
&\le 2\ell\left(\int_s^{s+T} b_1(t-s)\, dt\right)\|\theta(s)\| + 2b_0\varepsilon(s)\|\theta(s)\| \\
&\le (b_2T^2 + 2b_0\varepsilon(s))\|\theta(s)\|,
\end{aligned}$$

where $\varepsilon(s)$ is defined in Lemma 2.3, and $b_2 := \ell b_1$. ∎

C. Proof of Lemma 2.5

First let us assume $t \in [\eta, T]$ for some $\eta > 0$. Let the interval $[0,t]$ be divided into n subintervals of length t/n each. Let us introduce the following notation:

$$\Delta_{[s,s+r]} := \left\|\int_s^{s+r} \left(f(\theta(u),\xi(g^{-1}(u))) - \bar f(\theta(u)\right)\, du\right\|.$$

Then, we have

$$\Delta_{[s,s+t]} \le \Delta_{[s,s+t/n]} + ... + \Delta_{[s+(n-1)t/n,s+t]}. \qquad (27)$$

Assume s is large enough so that $\|\theta(u)\| \le b$ for $u \ge s$, as proved in the Section II-A. Also assume n is large enough so that $T/n \le 1$.

Consider the first term in the sum (27):

$$\begin{aligned}
\Delta_{[s,s+t/n]} &\le \left\|\int_s^{s+t/n} \left(f(\theta(s),\xi(g^{-1}(u))) - \bar f(\theta(s))\right)\, du\right\| \\
&+ \left\|\int_s^{s+t/n} \left(f(\theta(u),\xi(g^{-1}(u))) - f(\theta(s),\xi(g^{-1}(u)))\right) du\right\| \\
&+ \left\|\int_s^{s+t/n} \left(\bar f(\theta(u) - \bar f(\theta(s))\right)\, du\right\| \\
&\le \frac{T}{n}\left\|\frac{n}{t}\int_s^{s+t/n} \left(f(\theta(s),\xi(g^{-1}(u))) - \bar f(\theta(s))\right)\, du\right\| \\
&\quad + 2\ell\int_s^{s+t/n} \|\theta(u)-\theta(s)\| du \\
&\le \frac{T}{n}\cdot\frac{2bb_0\varepsilon(s)n}{t} + 2\ell\int_s^{s+T/n} bb_2(u-s)\, dt \\
&\le \frac{T}{n}\cdot\frac{2bb_0\varepsilon(s)n}{\eta} + 2\ell bb_2\frac{T^2}{n^2},
\end{aligned}$$

where the second to last inequality follows from Lemma 2.3, Lemma 2.4, and the ultimate boundedness of $\boldsymbol{\theta}$. The same derivation applies for each subinterval of length t/n. Adding up, we obtain

$$\Delta_{[s,s+t]} \le 2bb_0T\varepsilon(s)n/\eta + 2\ell bb_2T^2/n, \text{ for all } t \in [\eta, T].$$

Given $\delta > 0$, let us choose n such that $2\ell bb_2T^2/n < \delta/2$, and then s_0 such that for all $s \ge s_0$, $2bb_0T\varepsilon(s)n/\eta < \delta/2$, which yields

$$\sup_{t\in[\eta,T]} \Delta_{[s,s+t]} \le \delta, \text{ for all } s \ge s_0. \qquad (28)$$

Also, $\sup_{t\in[0,\eta]} \Delta_{[s,s+t]}$ is bounded above by,

$$\int_s^{s+\eta} \left(\|f(\theta(u),\xi(g^{-1}(u)))\| + \|\bar f(\theta(u))\|\right)\, du \le bb_1\eta.$$

Letting $\eta < \delta/(bb_1)$ and using (28), we obtain Lemma 2.5. ∎

REFERENCES

[1] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, New York, 2007.

[2] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

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

[4] Shalabh Bhatnagar, Michael C. Fu, Steven I. Marcus, and I-Jeng Wang. Two-timescale simultaneous perturbation stochastic approximation using deterministic perturbation sequences. *ACM Trans. Model. Comput. Simul.*, 13(2):180–209, 2003.

[5] V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Hindustan Book Agency and Cambridge University Press (jointly), Delhi, India and Cambridge, UK, 2008.

[6] V. R. Konda and J. N. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *Ann. Appl. Probab.*, 14(2):796–819, 2004.

[7] H. J. Kushner and G. G. Yin. *Stochastic approximation algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1997.

[8] D.-J. Ma, A. M. Makowski, and A. Shwartz. Stochastic approximations for finite-state Markov chains. *Stochastic Process. Appl.*, 35(1):27–45, 1990.

[9] P. G. Mehta and S. P. Meyn. Q-learning and Pontryagin's minimum principle. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 3598–3605, Dec. 2009.

[10] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, Cambridge, 2007.

[11] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992.

[12] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[13] Sandilya S. and Kulkarni S. R. Deterministic sufficient conditions for convergence of simultaneous perturbation stochastic approximation algorithms. In *Proceedings of the 9th INFORMS Applied Probability Conference*, 1997.

[14] I.-J. Wang and E.K.P. Chong. A deterministic analysis of stochastic approximation with randomized directions. *IEEE Trans. Automat. Control*, 43(12):1745 –1749, dec. 1998.