SIAM J. CONTROL OPTIM.
Vol. 41, No. 1, pp. 1–22

STOCHASTIC APPROXIMATION FOR NONEXPANSIVE MAPS: APPLICATION TO Q-LEARNING ALGORITHMS*

JINANE ABOUNADI†, DIMITRI P. BERTSEKAS†, AND VIVEK BORKAR‡

Abstract. We discuss synchronous and asynchronous iterations of the form

$$x^{k+1} = x^k + \gamma(k)(h(x^k) + w^k),$$

where h is a suitable map and $\{w^k\}$ is a deterministic or stochastic sequence satisfying suitable conditions. In particular, in the stochastic case, these are stochastic approximation iterations that can be analyzed using the ODE approach based either on Kushner and Clark's lemma for the synchronous case or on Borkar's theorem for the asynchronous case. However, the analysis requires that the iterates $\{x^k\}$ be bounded, a fact which is usually hard to prove. We develop a novel framework for proving boundedness in the deterministic framework, which is also applicable to the stochastic case when the deterministic hypotheses can be verified in the almost sure sense. This is based on scaling ideas and on the properties of Lyapunov functions. We then combine the boundedness property with Borkar's stability analysis of ODEs involving nonexpansive mappings to prove convergence (with probability 1 in the stochastic case). We also apply our convergence analysis to Q-learning algorithms for stochastic shortest path problems and are able to relax some of the assumptions of the currently available results.

Key words. stochastic approximation, Q-learning, neuro-dynamic programming

AMS subject classifications. 62L20

PII. S0363012998346621

1. Introduction.

The motivation for this paper has been the analysis of Q-learning algorithms, which have emerged as a powerful simulation tool for solving dynamic programming problems when a model is not known and/or the problem must be solved on-line as the data become available. Q-learning algorithms were first formulated by Watkins (1989), who gave a partial convergence analysis that was later amplified by Watkins and Dayan (1992). A more comprehensive analysis was given by Tsitsiklis (1994) (also reproduced in Bertsekas and Tsitsiklis (1996)), which made the connection between Q-learning and stochastic approximation. (A related treatment of a class of algorithms that include Q-learning and TD(λ) also appeared around the same time in Jaakola, Jordan, and Singh (1994). It may be recalled here that TD(λ) is a learning scheme for estimating the value function of a policy based on an exponentially weighted average (with weights λ^n for some $\lambda \in (0,1)$) of the so-called n-step truncated returns—see Bertsekas and Tsitsiklis (1996) for a detailed description.) In particular, Q-learning algorithms for discounted cost problems or stochastic shortest path (SSP) problems were viewed as asynchronous stochastic approximation versions of well-known value iteration algorithms in dynamic programming. This connection paved the way for a general analysis based on classic stochastic approximation techniques and dynamic programming-related contraction and monotonicity properties.

*Received by the editors October 29, 1998; accepted for publication August 31, 2001; published electronically March 27, 2002. This research was supported by the NSF under grant 9600494-DMI.
http://www.siam.org/journals/sicon/41-1/34662.html

†Department of Electrical Engineering and Computer Science, M.I.T., Cambridge, MA 02139 (jinane@mit.edu, dimitri@mit.edu).

‡School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India (borkar@titr.res.in). The research of this author was supported in part by the Homi Bhabha Fellowship and by the government of India, Department of Science and Technology grant III 5(12)/96-ET.

A weakness of the methodology developed so far is that it deals in an ad hoc way with the question of boundedness of the Q-learning iterates. In particular, the analysis of Tsitsiklis required a special argument for proving boundedness with probability 1 (w.p.1), and for the case of SSP problems it also required that the cost per stage be nonnegative, unless boundedness is imposed as an assumption (see Bertsekas and Tsitsiklis (1996), Prop. 5.6).

Our purpose in this paper is to provide a new and powerful general framework for establishing boundedness and proving convergence in synchronous and asynchronous stochastic approximation methods involving nonexpansive maps, including as a special case Q-learning algorithms. Our framework relies strongly on nonexpansiveness and combines ideas from several fields, including asynchronous stochastic approximation analysis via the limiting ODE technique and nonlinear analysis of ODEs. Our method for dealing with boundedness bears a similarity to an idea from the paper by Jaakola, Jordan, and Singh (1994), which addressed the convergence of TD(λ) using stochastic approximation methods (see section 2). Also see Csibi (1975) and Gerencser (1992) for work in a similar spirit. As a special case of our analysis, we improve on Tsitsiklis' convergence result by dispensing with the boundedness assumption for the iterates of SSP Q-learning, in the case where the cost per stage may be negative. The methodology developed in this paper also provides an essential foundation for a convergence analysis of Q-learning algorithms for average cost dynamic programming problems given in a companion paper (Abounadi, Bertsekas, and Borkar (2001)).

Our results, in fact, can be cast as a powerful *deterministic* principle, because the conditions on the noise required to ensure its applicability can be cast in simple deterministic terms. These can, in turn, be verified in the almost sure sense for the stochastic approximation algorithms of interest here. The deterministic formulation also requires weaker conditions on the stepsizes. Thus we shall initially state our results in a deterministic framework, enlarging their scope beyond the applications to stochastic approximation.

The general framework that we propose applies to synchronous and asynchronous variants of algorithms of the form

$$x^{k+1} = x^k + \gamma(k)\big(h(x^k) + w^k\big). \tag{1}$$

Here x^k is a sequence in $\Re^n$, w^k is a deterministic noise sequence, h is Lipschitz, $\gamma(k)$ is a positive stepsize sequence, and the aim is to find a solution of the equation $h(x) = 0$. This is the synchronous implementation in which all components are updated together at each time with full information about past iterates. The asynchronous model that we use is based on the formulation of Borkar (1998) and is of the form

$$x_i^{k+1} = x_i^k + \gamma\big(\nu(k,i)\big)\big(h_i(x^k) + w_i^k\big)I(i \in Y^k) \tag{2}$$

for $i = 1, \ldots, n$, where Y^k is the subset of $\{1, 2, \ldots, n\}$ denoting components being updated at time k, $I(\cdot)$ is the indicator function, and $\nu(k,i)$ is the number of times the component x_i of the vector x has been updated by time k.

For the synchronous algorithm (1), a powerful analysis technique is the ODE method introduced by Ljung (1977), formally treated by Kushner and Clark (1978), and Benveniste, Metivier, and Priouret (1990). For the asynchronous algorithm (2), a similar technique has been developed by Borkar (1998). (See also Kushner and Yin (1997) and references therein for related work.) The major idea behind these two techniques is to find a limiting deterministic continuous-time ODE for the stochastic discrete-time processes, using interpolation with the appropriate time scaling. The

main result is that if the ODE has an asymptotically stable equilibrium point, then under appropriate assumptions, which include boundedness of the generated iterates, the discrete-time iteration converges to this point w.p.1. Thus, in ODE techniques, boundedness must be independently verified.

This paper's methodology for dealing with the boundedness issue involves three steps:

1. obtaining a related scaled iteration and establishing its convergence,
2. showing that the sequence $\{x^k\}$ generated by the original iteration is bounded as a consequence of the convergence of the scaled iteration,
3. showing that the boundedness of $\{x^k\}$ implies convergence by invoking a standard ODE limiting argument.

For each of the steps above, we will impose appropriate sufficient conditions on the mapping h, the stepsize, and the noise. A central assumption in our later applications is that the mapping h is of the form $h(x) = T(x) - x$, where the map T is nonexpansive with respect to some norm $\|\cdot\|_p$ with $p \in (1, \infty]$ for the synchronous case, and with respect to the sup-norm $\|\cdot\|_\infty$ for the asynchronous case. To our knowledge, ours is the first general method for dealing with the boundedness issues in the ODE approach where the underlying mapping T is not a contraction. (See, however, the recent work by Borkar and Meyn (2000), which is discussed later in this section.) Note that the class of fixed-point problems which involve nonexpansive mappings arises in a number of different applications (see the book by Bertsekas and Tsitsiklis (1989) and the papers by Tseng, Bertsekas, and Tsitsiklis (1990), Borkar and Soumyanath (1997), and Soumyanath and Borkar (1999)). In particular, it includes value iteration algorithms for various dynamic programming formulations, including Q-learning algorithms.

Step 1 of the scheme described above is carried out by choosing the scaling based on a Lyapunov function of an appropriate ODE. The scaling works like a projection on an appropriate bounded set when the iterates lie outside a certain level set of the Lyapunov function. Note that we do not need to know the Lyapunov function; all we need to know is that such a function exists. For this we will use a general converse Lyapunov theorem that guarantees the existence of a smooth Lyapunov function if the ODE has a globally asymptotically stable equilibrium point (Wilson (1969)). Given this scaling scheme, we will be able to show that the scaled iteration has the same deterministic limiting ODE and hence converges. The argument is similar to the standard limiting ODE argument of Kushner and Clark (1978). We need to consider the Skorohod topology instead of the "uniform convergence on compacts" topology on $C([0, \infty); \Re^n)$. Step 2 involves the idea of comparing the original iteration and its scaled counterpart and showing that the difference between the two is bounded due to the nonexpansiveness of the mapping F. The idea of comparing the two iterations appeared first in Jaakola, Jordan, and Singh (1994) in a more limited setting. Step 3 is an application of standard ODE limiting arguments since boundedness is already established.

It is instructive to compare this approach with that of Borkar and Meyn (2000). While both are motivated by the same class of algorithms, viz., Q-learning, they exploit different features of the latter. While our approach is solely based on the nonexpansivity of an associated map, Borkar and Meyn use a scaling limit of this map, in the spirit of fluid models in queueing theory. To underscore the difference, note that the stochastic gradient scheme can be viewed as a fixed-point seeking iteration of an L_2-nonexpansive map when the associated Hessian is uniformly bounded—see section III.B of Soumyanath and Borkar (1999). Thus it comes under the purview of

the present scheme, but not under that of Borkar and Meyn (2000) in the absence of any specification of how the gradient in question behaves near infinity. On the other hand, the requirement that a convenient scaling limit hold in their sense can be met without the map being nonexpansive: the former concerns only the behavior near infinity, but the latter is a global requirement. Thus the approach of Borkar and Meyn and that of the present paper are quite distinct, and given the paucity of general purpose criteria for the stability of stochastic recursions of this type, both are of interest, despite the fact that currently they are aimed at broadly the same class of problems. More generally, our scheme will work (under mild technical assumptions) for the recursions wherein the distance between iterates for two instantiations of the algorithm with the same random inputs, but with two different initial conditions, remains bounded by a function of the initial conditions.

Finally, we note that for recursive algorithms the idea of using projection as a way of forcing boundedness is not new. The difference in our approach is that the use of scaling is only a method of proof, and the objective is to establish the boundedness of the original iteration without altering the iterates by forcing them to be bounded.

2. Boundedness lemmas. The results in this paper will be divided into two parts: the boundedness lemmas and the convergence analysis of appropriately scaled synchronous and asynchronous iterations. The boundedness lemmas are given in the present section, and rely on the nonexpansiveness property of the concerned map with respect to some norm $\|\cdot\|_p$, $p \in (1, \infty]$, for the synchronous case, and the sup-norm for the asynchronous case. The convergence of the scaled iteration is analyzed in the next section.

For a set $\mathcal{A}$ of $\Re^n$, we denote by $\partial\mathcal{A}$ and $\bar{\mathcal{A}}$ the boundary and closure of $\mathcal{A}$, respectively (i.e., $\bar{\mathcal{A}} = \mathcal{A} \cup \partial\mathcal{A}$). We introduce via scaling a map that "projects" any point onto a bounded and open set $\mathcal{B}$ that contains the origin. This is done each time the point leaves a given set $\mathcal{C}$ that contains $\mathcal{B}$. The map is defined as follows.

DEFINITION 2.1. *Let $\mathcal{B}$ be an open and bounded subset of $\Re^n$ containing the origin, and let $\mathcal{C}$ be a subset of $\Re^n$ that contains $\mathcal{B}$. We define the mapping $\Pi_{\mathcal{B},\mathcal{C}} : \Re^n \mapsto \bar{\mathcal{B}}$ by*

$$\Pi_{\mathcal{B},\mathcal{C}}(x) = \gamma_{\mathcal{B},\mathcal{C}}(x) \cdot x,$$

where $\gamma_{\mathcal{B},\mathcal{C}} : \mathbb{R}^n \to (0, 1]$ is given by

$$\gamma_{\mathcal{B},\mathcal{C}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{C}, \\ \max\{\beta > 0 : \beta x \in \bar{\mathcal{B}}\} & \text{if } x \notin \mathcal{C}. \end{cases}$$

Since $\bar{\mathcal{B}}$ is compact, it can be seen that $\Pi_{\mathcal{B},\mathcal{C}}$ is well defined as a real-valued function. If $\mathcal{B}$ is an open ball with respect to the Euclidean norm centered at the origin, the map $\Pi_{\mathcal{B},\mathcal{C}}$ is like a projection on $\mathcal{B}$, but the decision to project depends on whether the point is outside the larger set $\mathcal{C}$.

Our first result is inspired by a lemma of Jaakola, Jordan, and Singh (1994), which guarantees convergence of an iteration as long as a scaled version converges. Their lemma uses a strong homogeneity assumption, which is unnecessary for our purposes.

LEMMA 2.1. *Let $\mathcal{B}$ be an open and bounded subset of $\Re^n$ containing the origin, and let $\mathcal{C}$ be a subset of $\Re^n$ that contains $\mathcal{B}$. Consider the algorithm*

$$x^{k+1} = G^k(x^k, \xi^k), \tag{3}$$

where we assume the following:

1. *$\{\xi^k\}$ is a sequence in a measurable space $(\Omega, \mathcal{F})$.*
2. *G^k is nonexpansive in x with respect to some norm $\|\cdot\|$ for every $\xi \in \Omega$:*

$$\|G^k(x,\xi) - G^k(y,\xi)\| \le \|x - y\| \qquad \forall\ x, y, \xi.$$

3. *The sequence $\{\tilde{x}^k\}$ generated by the scaled iteration*

$$\tilde{x}^{k+1} = G^k\big(\Pi_{\mathcal{B},\mathcal{C}}(\tilde{x}^k), \xi^k\big), \qquad \tilde{x}^0 = x^0,$$

converges to some vector $x^* \in \mathcal{B}$.

Then $\{x^k\}$ is bounded.

Proof. Since $\mathcal{B}$ is open, there exists a large enough $\overline{k}$ such that $\tilde{x}^k \in \mathcal{B}$ for $k \ge \overline{k}$. In other words, there exists a large enough $\overline{k}$ such that

$$\gamma_{\mathcal{B},\mathcal{C}}(\tilde{x}^k) = 1 \qquad \forall\ k \ge \overline{k}, \tag{4}$$

and hence

$$\tilde{x}^{k+1} = G^k(\tilde{x}^k, \xi^k) \qquad \forall\ k \ge \overline{k}. \tag{5}$$

Therefore, for $k \ge \overline{k}$,

$$\|x^{k+1} - \tilde{x}^{k+1}\| = \|G^k(x^k, \xi^k) - G^k(\tilde{x}^k, \xi^k)\| \le \|x^k - \tilde{x}^k\| \le \cdots \le \|x^{\overline{k}} - \tilde{x}^{\overline{k}}\|.$$

Since $\{\tilde{x}^k\}$ is bounded, it follows that $\{x^k\}$ is bounded. □

3. Analysis of the scaled iteration. Our objective is to apply Lemma 2.1 to the synchronous and asynchronous algorithms given by (1) and (2). To this end, we will first establish the convergence of scaled versions of iterations (1) and (2) by using ODE-type arguments and conclude boundedness of the unscaled versions. However, the scaling (i.e., the sets $\mathcal{B}$ and $\mathcal{C}$ in Lemma 2.1) must be chosen so that we can find a limiting ODE that is easily analyzed. In particular, if the scaling is not done appropriately, the scaled iteration might not converge. The iterates could, for example, keep hitting the boundary of $\mathcal{B}$ infinitely often and thus never converge, or the scaling could generate additional fixed points at the boundary that the iterates might converge to.

Given an ODE $\dot{x} = h(x)$ in $\Re^n$ with a global asymptotically stable equilibrium point x^*, a smooth Lyapunov function $V : \Re^n \mapsto \Re$ is a continuously differentiable function satisfying $V(x^*) = 0$, $V(x) > 0$ for all $x \ne x^*$, and such that the inner product of its gradient $\nabla V(x)$ and $h(x)$ is negative for all $x \ne x^*$. A necessary and sufficient condition for x^* to be a global asymptotically stable equilibrium point is the existence of a corresponding Lyapunov function (see Yoshizawa (1966)). Using some smoothing techniques, Wilson showed that the Lyapunov function can be taken to be smooth (in fact, infinitely differentiable; see Theorem 3.2 in Wilson (1969)). The following lemma will be useful to us.

LEMMA 3.1. *Let $\dot{x} = h(x)$ be an ODE with a global asymptotically stable equilibrium point x^*. Let V be a smooth Lyapunov function for the ODE. For any $R > 0$, there is a $C > 0$ such that the closed ball $\bar{B}(x^*, R)$ of radius R centered at x^* is in the interior of the level set $L = \{x \in \Re^n : V(x) \le C\}$.*

Proof. Consider the closure $\bar{B}(x^*, R)$ of $B(x^*, R)$. Since V is continuous and $\bar{B}(x^*, R)$ is compact, the maximum of V over $\bar{B}(x^*, R)$ is attained. Let $\overline{C} = \max_{x \in \bar{B}(x^*, R)} V(x)$. Any level set of the form $L = \{x \in \Re^n : V(x) \le C\}$, where $C > \overline{C}$, contains $\bar{B}(x^*, R)$ in its interior. □

3.1. Analysis of the scaled iteration-synchronous case. The scaled version of the synchronous algorithm of (1) is given by

$$\begin{aligned} \tilde{x}^{k+1} &= x^k + \gamma(k)\big(h(x^k) + w^k\big), \\ x^{k+1} &= \Pi_{\mathcal{B},\mathcal{C}}(\tilde{x}^{k+1}). \end{aligned} \tag{6}$$

We first show, under appropriate conditions, that this iteration converges w.p.1 to the unique equilibrium point of an appropriate ODE. The scaled iteration (6) can be written as

$$x^{k+1} = x^k + \gamma(k)\big(h(x^k) + w^k\big) + g^k, \tag{7}$$

where

$$g^k = \Pi_{\mathcal{B},\mathcal{C}}\big[x^k + \gamma(k)\big(h(x^k) + w^k\big)\big] - \big[x^k + \gamma(k)\big(h(x^k) + w^k\big)\big]. \tag{8}$$

We formally state our assumptions, which for completeness include some of our earlier assertions on the existence of a global asymptotically stable equilibrium x^*, the choice of the sets $\mathcal{B}$ and $\mathcal{C}$, etc., as follows.

ASSUMPTION 3.1. *The stepsizes $\gamma(k)$ satisfy*

$$0 < \gamma(k) \to 0, \qquad \sum_{k=0}^{\infty} \gamma(k) = \infty.$$

ASSUMPTION 3.2.

1. *There exists D such that $\|w^k\| \leq D$ for all k.*
2. $\lim_{k\to\infty} \sum_{m=k}^{m_T(k)} \gamma(m) w^m = 0$ *for all T, where*

$$m_T(k) = \min\left\{ m \geq k : \sum_{l=k}^{m} \gamma(l) \geq T \right\}.$$

3. *h is Lipschitz continuous; i.e., for some $L > 0$,*

$$\|h(x) - h(y)\| \leq L\|x - y\|.$$

4. *The ODE $\dot{x} = h(x)$ has a globally asymptotically stable equilibrium point x^*.*

Remark 3.1. Note that the boundedness condition in Assumption 3.2 is for the rescaled iterations, *not* for the original iterations. For the applications we have in mind, $||w^k||$ will be bounded by an affine function of $||x^k||$ and therefore will be bounded whenever the latter is. But the latter is bounded, by construction, for the rescaled iterations, and thus Assumption 3.2.1 is satisfied. It is being neither assumed nor implied a priori that the noise sequence $\{w^k\}$ in the original iterations is bounded; this will, in fact, be a consequence of our stability result. More generally, it will suffice to have $||w^k||$ bounded by a continuous function of x^k.

Remark 3.2. This remark concerns Assumption 3.2. The important thing to note here is that we are imposing this assumption on the *projected* algorithm, for which the boundedness of iterates is true by construction, not for the original scheme, whose stability we intend to prove.

In our analysis, we will use Lemma 2.1 with $\mathcal{B} = B(0, R)$ and $\mathcal{C} = \{x \in \Re^n : V(x) < C\}$, where $R > \|x^*\|$, V is a smooth Lyapunov function for the ODE $\dot{x} = h(x)$, and the constant C is large enough so that $\mathcal{C}$ contains $B(0, \bar{R})$ for some $\bar{R} > R$. Note

that the vector field of the ODE is transversal to the level sets of V, implying that if $x \in \partial\mathcal{C}$, then $(x + \Delta h(x)) \in \mathcal{C}$ for small enough $\Delta > 0$. This motivates the choice of the scaling sets $\mathcal{B}$ and $\mathcal{C}$ above. Intuitively, if the stepsize is small enough, we can think of the algorithm as starting at the boundary of $\mathcal{B}$ and moving around initially in $\mathcal{C}$. As it approaches the boundary of $\mathcal{C}$, it gets pushed back to the interior of $\mathcal{C}$, thanks to the fact that the vector field of the ODE on the boundary points inward and in spite of the noise term.

In order to proceed with our convergence analysis, we need to define piecewise linear or piecewise constant interpolated processes based on the iterates $\{x^k\}$. Let

$$t_k = \sum_{m=0}^{k-1} \gamma(m), \qquad k \geq 1,$$

with $t_0 = 0$. Let

$$\tilde{x}^{k+1} = x^k + \gamma(k)\big(h(x^k) + w^k\big), \qquad k \geq 0,$$

$$X_l(t) = \begin{cases} x^k & \text{for } t = t_k, \\ \big(1 - \frac{t-t_k}{\gamma(k)}\big)x^k + \frac{t-t_k}{\gamma(k)}\tilde{x}^{k+1} & \text{for } t \in [t_k, t_{k+1}), \end{cases}$$

$$X_c(t) = x^k, \qquad k \geq 0, \qquad \text{for } t \in [t_k, t_{k+1}),$$

$$G_c(t) = \sum_{m=0}^{k-1} g^m \qquad \text{for } t \in [t_k, t_{k+1}),$$

$$W_l(t) = \begin{cases} \sum_{m=0}^{k-1} \gamma(m) w^m & \text{for } t = t_k, \\ \big(1 - \frac{t-t_k}{\gamma(k)}\big)W_l(t_k) + \frac{t-t_k}{\gamma(k)} W_l(t_{k+1}) & \text{for } t \in [t_k, t_{k+1}). \end{cases}$$

Thus $X_l(\cdot)$ is right-continuous with left limits (r.c.l.l., for short); that is, $X_l(t^+) = \lim_{\delta \downarrow 0} X_l(t + \delta)$ and $X_l(t^-) = \lim_{\delta \downarrow 0} X_l(t - \delta)$ are well defined, with $X_l(t) = X_l(t^+)$. In fact, $X_l(\cdot)$ is piecewise linear and continuous everywhere, except at times t_k for which $g^k \neq 0$, where it has a jump discontinuity. Define the left-shifted versions of these processes as follows, for $t \geq 0$:

$$X_l^k(t) = X_l(t + t_k),$$

$$W_l^k(t) = W_l(t + t_k) - W_l(t_k),$$

$$X_c^k(t) = X_c(t + t_k),$$

$$G_c^k(t) = G_c(t + t_k) - G_c(t_k).$$

Then it is easy to see that for $t \geq -t_k$

$$\begin{aligned} X_l^k(t) &= X_l^k(0) + \int_0^t h\big(X_c^k(\tau)\big)d\tau + W_l^k(t) + G_c^k(t) \\ &= X_l^k(0) + \int_0^t h\big(X_l^k(\tau)\big)d\tau + W_l^k(t) + G_c^k(t) + e^k(t), \end{aligned}$$

where

$$e^k(t) = \int_0^t h\big(X_c^k(\tau)\big)d\tau - \int_0^t h\big(X_l^k(\tau)\big)d\tau.$$

By Assumption 3.2, $\{W_l^k(\cdot)\}$ converges to zero uniformly on finite intervals as $k \to \infty$. We show next that $\{e^k(\cdot)\}$ and $\{G_c^k(\cdot)\}$ behave analogously.

LEMMA 3.2. *For any $T > 0$, $\sup_{t\in[0,T]} \|e^k(t)\| \to 0$ as $k \to \infty$.*

Proof. By Assumptions 3.2,

$$\|e^k(t)\| \le \int_0^t \big\|h\big(X_c^k(\tau)\big) - h\big(X_l^k(\tau)\big)\big\|d\tau \le L\int_0^t \big\|X_c^k(\tau) - X_l^k(\tau)\big\|d\tau.$$

Letting

$$m_T(k) = \min\left\{m \ge k \ : \ \sum_{l=k}^{m}\gamma(m) \ge T\right\},$$

we have

$$\begin{aligned}
\sup_{t\in[0,T]} \|e^k(t)\| &\le L\int_0^T \big\|X_c^k(\tau) - X_l^k(\tau)\big\|d\tau \\
&\le \sum_{m=k}^{m_T(k)} \gamma(m)L \sup_{\tau\in[t_m,t_{m+1})} \big\|X_c^k(\tau) - X_l^k(\tau)\big\| \\
&\le \sum_{m=k}^{m_T(k)} \gamma(m)L(t_{m+1} - t_m)\big\|h(x^m) + w^m\big\| \\
&\le \sum_{m=k}^{m_T(k)} \gamma^2(m)LD',
\end{aligned}$$

where

$$D' = D + \sup_{x\in C}\|h(x)\|,$$

and the second inequality is a consequence of the definitions of $X_l^k(\cdot)$ and $X_c^k(\cdot)$. By Assumption 3.1, we have $\sum_{m=k}^{m_T(k)} \gamma^2(m) \to 0$ as $k \to \infty$, implying the result. □

To analyze the r.c.l.l. processes $X_l^k(\cdot)$, $G_c^k(\cdot)$, we recall from Billingsley (1968) the space $D([0,T];\Re^n)$ of r.c.l.l. functions from $[0,T]$ to $\Re^n$ (where $T > 0$), equipped with the Skorohod topology. This topology is defined so that $f^k(\cdot) \to f(\cdot)$ in $D([0,T];\Re^n)$ if and only if there exist continuous, nondecreasing, onto functions $\lambda^k : [0,T] \to [0,T]$ such that $f^k(\lambda^k(t)) \to f(t)$ and $\lambda^k(t) \to t$, uniformly on $[0,T]$. We denote by $D([0,\infty);\Re^n)$ the space of r.c.l.l. functions from $[0,\infty)$ to $\Re^n$, defined such that $f^k(\cdot) \to f(\cdot)$ in $D([0,\infty);\Re^n)$ if and only if their respective restrictions to $[0,T]$ converge in $D([0,T];\Re^n)$ for every $T > 0$. Both $D([0,T];\Re^n)$ and $D([0,\infty);\Re^n)$ are separable and metrizable with a complete metric.

We recall from Billingsley (1968, p. 118) the following characterization of relative compactness in $D([0,T];\Re^n)$: a set $A \subset D([0,T];\Re^n)$ is relatively compact if and only if

$$\sup_{x(\cdot)\in A}\sup_{t\in[0,T]}\|x(t)\| < \infty \tag{9}$$

and

$$\lim_{\delta\to 0}\sup_{x(\cdot)\in A}\sup_{t_1\le t\le t_2,\, t_2-t_1\le\delta}\min\bigl\{\|x(t)-x(t_1)\|,\,\|x(t_2)-x(t)\|\bigr\}=0, \tag{10a}$$

$$\lim_{\delta\to 0}\sup_{x(\cdot)\in A}\sup_{t_1,t_2\in[0,\delta)}\|x(t_2)-x(t_1)\|=0, \tag{10b}$$

$$\lim_{\delta\to 0}\sup_{x(\cdot)\in A}\sup_{t_1,t_2\in[T-\delta,T)}\|x(t_2)-x(t_1)\|=0. \tag{10c}$$

This generalizes the well-known Arzelà–Ascoli theorem for $C([0,T];\Re^n)$, the space of continuous functions from $[0,T]$ to $\Re^n$ with the sup-norm.

LEMMA 3.3. *The sequences* $\{X_l^k(\cdot)\}$ *and* $\{G_c^k(\cdot)\}$ *are relatively compact in* $D([0,\infty);\Re^n)$.

Proof. It suffices to check the relative compactness of their restrictions to $[0,T]$ in $D([0,T];\Re^n)$ for arbitrary $T>0$. Let us fix $T>0$. Since $\{x^k\}$ and $\{g^k\}$ are bounded, so are the sequences $\{X_l^k(\cdot)\}$ and $\{G_c^k(\cdot)\}$. Thus (9) above holds. It is easy to see that (10a)–(10c) will follow if any two discontinuity points of $x(\cdot)\in A$ are separated by at least some $\Delta>0$. For the processes under consideration, discontinuities occur at some of the t_k's. Let there be a discontinuity at t_k for some k. Then $g^{k-1}\neq 0$ and $x^k\in\partial B$. Let

$$d=\min_{x\in\partial B,\, y\in\partial C}\|x-y\|>0,$$

and define

$$m(k)=\max\left\{j\ :\ \sum_{i=0}^{j}\gamma(k+i)\le\frac{d}{D'}\right\},$$

where D' is as before. We claim that $x^{k+1},x^{k+2},\ldots,x^{k+m(k)}$ are in the interior of C. To see this, notice that if

$$\gamma(k)<\frac{d}{D'},$$

then

$$\|\tilde{x}^{k+1}-x^k\|<d,$$

implying that $\tilde{x}^{k+1}$ is in the interior of C and thus $x^{k+1}=\tilde{x}^{k+1}$. Therefore, $g^k=0$, implying no discontinuity at t_{k+1}. Similarly, if

$$\sum_{i=0}^{j-1}\gamma(k+i)<\frac{d}{D'},$$

then x^{k+i} is in the interior of C for $i=1,\ldots,j$. This implies the claim that there are no discontinuities in the interval $[t_k,t_k+d/D')$. Let $\Delta=d/2D'$. □

Let $K=\{k\ :\ g^k=0\}$. Let $\{X_l^k(\cdot)\}$ and $\{G_c^k(\cdot)\}$ converge in $D([0,T];\Re^n)$ to some $X(\cdot)$ and $G(\cdot)$, respectively, along a subsequence of K. (From the above proof, it is easy to see that K will be infinite: once k is large enough so that $\gamma(k)<\frac{d}{D'}$, each k with $g^k\neq 0$ will lead to $g^{k+1}=0$.) Then the limits must satisfy

$$X(t)=X(0)+\int_0^t h\bigl(X(\tau)\bigr)d\tau+G(t).$$

Furthermore, from the nature of our notion of convergence in $D([0,T];\Re^n)$, it is clear that $G(\cdot)$ is piecewise constant r.c.l.l. with $G(0)=0$ and that any two discontinuities of $G(\cdot)$ (hence of $X(\cdot)$) are separated by at least Δ on the time axis. Recall that for $x(\cdot)\in D([0,T];\Re^n)$, $x(t^+)=\lim_{t<s\to t} x(s)$ and $x(t^-)=\lim_{t>s\to t} x(s)$.

LEMMA 3.4. *We have $G(\cdot)\equiv 0$, implying $\dot{X}(t)=h(X(t))$.*

Proof. Let

$$\tau=\inf\{t>0\ :\ X(t^+)\neq X(t^-)\}.$$

By the right continuity at 0 and the fact that any two discontinuity points are separated by at least $\Delta>0$, it follows that $\tau>0$. Let $\|X(\tau^+)-X(\tau^-)\|=\delta>0$. Then, by our notion of convergence, we can find $\tau_k<\tau_k'$, $k\geq 0$, such that $\tau_k'-\tau_k\to 0$ and

$$\left\|X_l^{n(k)}(\tau_k')-X(\tau^+)\right\|\to 0, \tag{11}$$

$$\left\|X_l^{n(k)}(\tau_k)-X(\tau^-)\right\|\to 0. \tag{12}$$

Recall that $\|h(\cdot)\|$ is bounded on C and that $e^k(\cdot)$ and $W_l^k(\cdot)$ converge to 0 uniformly on compact sets. Also, any two discontinuities of $X_l^n(\cdot)$ must be at least Δ apart. Thus, for sufficiently large k, there must exist a $\hat{\tau}_k\in[\tau_k,\tau_k']$ such that

$$\left\|X_l^{n(k)}(\hat{\tau}_k)-X_l^{n(k)}(\hat{\tau}_k^-)\right\|\geq\frac{\delta}{2}.$$

But then $X_l^{n(k)}(\hat{\tau}_k^+)\in\partial B$, and $X_l^{n(k)}(\hat{\tau}_k^-)$ is not in the interior of C. Once again, using (11) and (12) and the fact that two discontinuities of $X_l^n(\cdot)$ must be at least Δ apart, we conclude that $X(\tau^+)\in\partial B$ and $X(\tau^-)\in\partial C$. But then $X(\cdot)$ satisfies $\dot{X}(t)=h(X(t))$ on $[0,\tau)$ (since $G(\cdot)\equiv 0$ on $[0,\tau)$), and therefore an interior trajectory of this ODE in C hits ∂C, a contradiction of our choice of C. (Since C is a level set of the Lyapunov function $V(\cdot)$, $h(\cdot)$ is transversal to ∂C everywhere and is directed towards the interior.) This contradiction proves that $G(\cdot)\equiv 0$. □

The preceding lemma allows us to prove the following proposition, the proof of which proceeds along standard lines; see, e.g., Kushner and Clark (1978), Benveniste, Metivier, and Priouret (1990).

PROPOSITION 3.1. *Let Assumptions* 3.1 *and* 3.2 *hold. The scaled synchronous algorithm* (6) *converges to x^*.*

3.2. Analysis of the scaled iteration-asynchronous case. The scaled version of the asynchronous algorithm of (2) is given by

$$\begin{aligned}\tilde{x}_i^{k+1}&=x_i^k+\gamma\big(\nu(k,i)\big)\left(h_i(x^k)+w_i^k\right)I(i\in Y^k),\\ x^{k+1}&=\Pi_{\mathcal{B},\mathcal{C}}(\tilde{x}^{k+1}).\end{aligned} \tag{13}$$

We confine ourselves to nonexpansive mappings with respect to the sup-norm. We also impose a further assumption on the stepsize. In particular, we will use the following assumptions in place of Assumption 3.1. We use $[a]$ to denote the integer part of a real number a.

ASSUMPTION 3.3. *The stepsizes $\gamma(k)$ are eventually nonincreasing and satisfy*

$$0<\gamma(k)\to 0,\qquad \sum_{k=0}^{\infty}\gamma(k)=\infty.$$

In addition, for all $\beta \in (0,1)$,

$$\sup_k \frac{\gamma([k\beta])}{\gamma(k)} < \infty$$

and

$$\lim_{k\to\infty} \frac{\sum_{m=0}^{[k\overline{\beta}]} \gamma(m)}{\sum_{m=0}^{k} \gamma(m)} = 1, \qquad \textit{uniformly in } \overline{\beta} \in [\beta, 1].$$

ASSUMPTION 3.4. *There exists a* $\Gamma > 0$ *such that for all* i

$$\liminf_{k\to\infty} \frac{1}{k+1} \nu(k,i) \geq \Gamma.$$

Furthermore, for all $T > 0$, *the limit*

$$\lim_{n\to\infty} \frac{\sum_{k=\nu(n,i)}^{\nu(m_T(n),i)} \gamma(k)}{\sum_{k=\nu(n,j)}^{\nu(m_T(n),j)} \gamma(k)}$$

exists for all i, j.

Theorem 3.2 of Borkar (1998) implies that the above limit will in fact be 1, a fact we use later. In addition, we change Assumption 3.2 to the following.

ASSUMPTION 3.2′. *For* $T, m_T(k)$ *as before,*

$$\lim_{k\to\infty} \sum_{m=k}^{m_T(k)} \gamma(m) w^{l(m)} = 0,$$

where $\{l(m)\}$ *is any increasing sequence of nonnegative integers satisfying* $l(m) \geq m$ *for all* m.

Examples of stepsizes that satisfy Assumption 3.3 include $\gamma(k) = 1/k$, $\gamma(k) = 1/(k \log k)$, etc., for $k \geq 2$, with suitable modifications for $k = 0, 1$. The essential meaning of Assumption 3.4 is that all components are updated comparably often.

Under Assumptions 3.2′, 3.3, and 3.4, the analysis closely mimics that of the synchronous case, except that the ODE-based convergence analysis of Kushner and Clark (1978) and Benveniste, Metivier, and Priouret (1990) is replaced by the corresponding analysis of Borkar (1998). In order to avoid undue repetition, we shall provide only a brief sketch. The key result of Borkar (1998) that is used here is briefly described in the appendix.

The first simplifying assumption that we make is that Y^k is a singleton for all k; i.e., only one component is updated at a time. This is justified as in Borkar (1998), the idea being that one unfolds a single iteration that updates d components, $d \geq 2$, into d iterations, in which each iteration updates a single component. There is, however, a complication in that this artificially introduces bounded delays; that is, the update of the ith component at time $k+1$ may use the value of the jth component updated not at time k, but at time $k-m$ for some $m \leq n$. These delays can be handled as in Borkar (1998). For simplicity of exposition, we ignore the delays here.

Thus we have $Y^k = \{\phi^k\}$, where ϕ^k is the index of the component updated at time k, and the iteration (13) is written as

$$x^{k+1} = x^k + D^k(h(x^k) + w^k) + g^k,$$

where

$$D^k = \mathrm{diag}\big[\gamma(\nu(k,1))I(\phi^k = 1), \ldots, \gamma(\nu(k,n))I(\phi^k = n)\big]$$

and

$$g^k = \Pi_{\mathcal{B},\mathcal{C}}\big[x^k + D^k(h(x^k) + w^k)\big] - \big[x^k + D^k(h(x^k) + w^k)\big].$$

Let us denote

$$\bar{\mu}^k = \big[I(\phi^k = 1), \ldots, I(\phi^k = n)\big]$$

and set $\bar{\gamma}(m,j) = \gamma(\nu(m,j))$, $\hat{\gamma}(m) = \bar{\gamma}(m, \phi^m)$, $t_0 = 0$, and $t_k = \sum_{m=0}^{k-1} \hat{\gamma}(m)$, $k \geq 1$. Let us define piecewise linear and piecewise constant processes as follows:

$$\mu(t) = \bar{\mu}^k \qquad \text{for } t \in [t_k, t_{k+1}),$$

$$X_c(t) = x^k \qquad \text{for } t \in [t_k, t_{k+1}),$$

$$G_c(t) = \sum_{m=0}^{k-1} g^m \qquad \text{for } t \in [t_k, t_{k+1}),$$

$$X_l(t) = \begin{cases} x^k & \text{for } t = t_k, \\ \big(1 - \frac{t-t_k}{\gamma(k)}\big)x^k + \frac{t-t_k}{\gamma(k)}\tilde{x}^{k+1} & \text{for } t \in [t_k, t_{k+1}), \end{cases}$$

where

$$\tilde{x}^{k+1} = x^k + D^k\big(h(x^k) + w^k\big),$$

$$W_l(t) = \begin{cases} \sum_{m=0}^{k-1} D^m w^m & \text{for } t = t_k, \\ \big(1 - \frac{t-t_k}{\gamma(\nu(k,\phi^k))}\big)W_l(t_k) + \frac{t-t_k}{\gamma(\nu(k,\phi^k))}W_l(t_{k+1}) & \text{for } t \in [t_k, t_{k+1}). \end{cases}$$

Define the corresponding left-shifted processes as follows, for $t \geq 0$:

$$X_l^k(t) = X_l(t + t_k),$$

$$X_c^k(t) = X_c^k(t + t_k),$$

$$W_l^k(t) = W_l(t + t_k) - W_l(t_k),$$

$$G_c^k(t) = G_c(t + t_k) - G_c(t_k),$$

$$\mu^k(t) = \mu(t + t_k).$$

For an n-dimensional probability vector $p = [p_1, \ldots, p_n]$, let $\mathrm{diag}(p)$ denote the diagonal matrix whose ith diagonal entry is p_i. Then, letting μ^* denote the uniform probability vector $[1/n, \ldots, 1/n]$, we have, for $t \geq 0$,

$$X_l^k(t) = X_l^k(0) + \int_0^t \mathrm{diag}(\mu^*)h\big(X_l^k(\tau)\big)d\tau + W_l^k(t) + G_c^k(t) + e^k(t) + \eta^k(t),$$

$$\eta^k(t) = \int_0^t \big(\mathrm{diag}(\mu^k(\tau)) - \mathrm{diag}(\mu^*)\big) h\big(X_l^k(\tau)\big) d\tau,$$

$$e^k(t) = \int_0^t \mathrm{diag}\big(\mu^k(\tau)\big)\big(h\big(X_c^k(\tau)\big) - h\big(X_l^k(\tau)\big)\big) d\tau.$$

The convergence of $\{W_l^k(\cdot)\}$ to 0 follows from Assumption 3.2′. Convergence of $\{e^k(\cdot)\}$ to 0 follows along the lines of the preceding subsection. The proof of Lemma 3.3 now goes through as before, with $D' = \sup_{z\in\mathcal{C}} \max_i |h_i(z)| + D$. We also have the following.

LEMMA 3.5. *For each $T > 0$,*

$$\lim_{k\to\infty} \sup_{t\in[0,T]} ||\eta^k(t)|| = 0.$$

Proof. As before, one verifies that the set $\{\{X_l^k(t), t \in [0,T]\}, k \geq 1\}$ is relatively compact in $D([0,T];\Re^n)$. Thus one may drop to a subsequence of $\{k\}$, denoted by $\{k\}$ again by abuse of notation, such that $X_l^k(\cdot) \to Z(\cdot)$ for some $Z(\cdot) \in D([0,T];\Re^n)$. Since the map $x(\cdot) \in D([0,T];\Re^n) \to x(t) \in \Re^n$ for any $t \in [0,T]$ is continuous at $z(\cdot)$ if $z(\cdot)$ is continuous at t (see Billingsley (1968, p. 121)), and also any $x(\cdot) \in D([0,T];\Re^n)$ has at most countably many points of discontinuity (see Borkar (1998, p. 119)), it follows that $X_l^k(t) \to Z(t)$ for almost every $t \in [0,T]$. By the dominated convergence theorem, one then has

$$\lim_{k\to\infty} \int_0^t \big(\mathrm{diag}(\mu^k(\tau)) - \mathrm{diag}(\mu^*)\big)\big(h(X_l^k(\tau)) - h(Z(\tau))\big) d\tau = 0.$$

Since the left-hand side (L.H.S.) has a bounded derivative in t, it is equicontinuous. It is clearly bounded for each fixed t. Thus a straightforward application of the Arzelà–Ascoli theorem shows that the above convergence is uniform in $t \in [0,T]$. Therefore the claim would follow if we show that

$$\lim_{k\to\infty} \int_0^t \big(\mathrm{diag}(\mu^k(\tau)) - \mathrm{diag}(\mu^*)\big) h(Z(\tau)) d\tau = 0,$$

uniformly in $[0,T]$. The uniformity of convergence over $[0,T]$ will follow as before from the Arzelà–Ascoli theorem if we prove pointwise convergence on $[0,T]$. In turn, the latter follows if we show that for each t

$$\lim_{k\to\infty} \int_0^t \big(\mathrm{diag}(\mu^k(\tau)) - \mathrm{diag}(\mu^*)\big) f(\tau) d\tau = 0$$

for any $f \in L_2([0,T];\Re^n)$. Consider $\mu^k(\cdot)$, $k \geq 1$, as elements of the space $\mathcal{U}$ of measurable maps from $[0,\infty)$ to the space of probability vectors in $\Re^n$, with the coarsest topology that renders continuous the maps $\mu(\cdot) \in \mathcal{U} \to \int_0^t \langle \mu(s), f(s)\rangle ds$ for all $t > 0$ and f as above. It is easy to deduce from the Banach–Alaoglu theorem that $\mathcal{U}$ is compact metrizable. Let $\bar{\mu}(\cdot)$ be any limit point of $\{\mu^k(\cdot)\}$ in $\mathcal{U}$ as $k \to \infty$. It follows from Theorem 3.2 of Borkar (1998) that $\bar{\mu} = \mu^*$. The claim follows. □

The proof of Lemma 3.4 now goes through as before. Thus the asynchronous iterates, suitably interpolated, track the ODE $\dot{x}(t) = (1/n)h(x(t))$, which has the same qualitative behavior as $\dot{x}(t) = h(x(t))$—the difference is a mere time scaling. As in Borkar (1998), we then obtain the following proposition.

PROPOSITION 3.2. *Let Assumptions* 3.2′*,* 3.3*, and* 3.4 *hold. The scaled asynchronous algorithm* (6) *converges to* x^*.

The only difference with Borkar (1998) will be that we are dealing with the projected algorithm here; therefore we have to allow for discontinuous trajectories. But this can be dealt with exactly as in the synchronous case.

4. Convergence theorems for stochastic approximation. Now we consider the situation in which $\{w^k\}$ is a random noise sequence. Specifically, we assume that it is adapted to a family of increasing σ-fields $\{\mathcal{F}^{k+1}\}$ to which $\{x^{k+1}\}$ is also adapted and satisfies

$$E[w^k/\mathcal{F}^k] = 0$$

for $k \geq 1$. We strengthen Assumption 3.1 to include

$$\sum_0^\infty \gamma(k)^2 < \infty,$$

which is a standard assumption in stochastic approximation theory. We further assume, in place of Assumptions 3.1 and 3.2′, that

$$E[||w^k||^2/\mathcal{F}^k] \leq H(x^k)$$

for some continuous $H(\cdot)$. Assumptions 3.3, 3.4 remain as before. We shall refer to the modified Assumptions 3.1, 3.2 as Assumptions 3.1(m), 3.2(m), respectively.

Note that the only use of Assumption 3.2 has been to ensure that there exists a $\Delta > 0$ such that consecutive jump times of $X_l(\cdot)$ are at least Δ apart. However, this Δ can depend on sample path in the present case without affecting the proof in any way. Since we are seeking almost sure convergence, it suffices to show the following.

LEMMA 4.1. *There exists w.p.*1 *a (possibly sample path dependent)* Δ *with the above property.*

Proof. Suppose that the claim is not true for some sample path. Let $\{t^{m(k)}\}$ denote the successive jump times, with $+\infty$ being a possible value for these. (In particular, $t^{m(k)} = \infty$ for $k > k_0$ if there are only k_0 jumps.) Then for the sample path under consideration, these are all finite, and moreover, there exist consecutive jump times $t^{m(k(l)+1)} > t^{m(k(l))}$ such that $t^{m(k(l)+1)} - t^{m(k(l))} \to 0$ as $l \to \infty$. Let $K = \sup_{x\in\mathcal{C}}||h(x)||$. Since the iterates move from $\partial\mathcal{B}$ to $\partial\mathcal{C}$ between $(t^{m(k(l))})^+$ and $(t^{m(k(l)+1)})^-$, we must have

$$\left\| \sum_{i=m(k(l))}^{m(k(l)+1)-1} \gamma(i)w^i \right\| \geq d - \left(t^{m(k(l)+1)} - t^{m(k(l))}\right) K \geq \frac{d}{2}$$

for l sufficiently large. Letting Ψ^l denote the L.H.S. above, it then follows that $\Psi^l \geq \frac{d}{2}$ infinitely often (i.o.). We shall prove that

$$P\left(\Psi^l \geq \frac{d}{2}, \text{ i.o.}\right) = 0,$$

which will imply the desired claim. By the Chebyshev inequality, we have

$$\sum_k P\left(\psi^k \geq \frac{d}{2}\right) \leq \sum_k \frac{4E[||\sum_{i=m(k)}^{m(k+1)-1} \gamma(i)w^i||^2 I(t^{m(k)} < \infty)]}{d^2}.$$

Summing over k, the R.H.S. sums to a quantity bounded by

$$\frac{4\sum_i \gamma(i)^2 E[||w^i||^2]}{d^2} \leq \frac{4(\sum_i \gamma(i)^2)\sup_{x\in\mathcal{C}}|H(x)|}{d^2} < \infty,$$

in view of our hypotheses on $\{w^k\}$. The claim follows from the Borel–Cantelli lemma. □

Our hypotheses also ensure that Assumption 3.2 holds a.s. To see this, let $M^k = \sum_{i=0}^k \gamma(i)w^i$ for $i \geq 0$. Then $(M^k, \mathcal{F}^{k+1})$ is a square-integrable martingale. Its quadratic variation process is

$$\sum_{i=0}^k \gamma(i)^2 \left(E\left[\frac{||w^i||^2}{\mathcal{F}^{i-1}}\right] - \left\| E\left[\frac{w^i}{\mathcal{F}^{i-1}}\right]\right\|^2 \right),$$

which is bounded by the finite quantity $2\sup_{x\in\mathcal{C}}|H(x)|\sum_i \gamma(i)^2$. By Theorem 3.3.4, p. 53, of Borkar (1995), $\{M^k\}$ converges a.s. It then follows that Assumption 3.2 holds a.s. Hence we have the following counterpart of Proposition 3.1.

LEMMA 4.2. *Under the above hypotheses, the scaled synchronous algorithm* (5) *converges to x^* a.s.*

For the asynchronous case, note that $(\sum_{m=0}^k \gamma(\nu(m,i))w_i^m, \mathcal{F}^k)$ is a (square-integrable) martingale for each i. Considerations similar to those above then lead to the following stochastic counterpart of Proposition 3.2.

LEMMA 4.3. *Under the above hypotheses, the scaled asynchronous algorithm converges to x^* a.s.*

We now specialize to algorithms of the form

$$x^{k+1} = x^k + \gamma(k)(F(x^k, \xi^k) - x^k)$$

in synchronous form and

$$x_i^{k+1} = x_i^k + \gamma(\nu(k,i))(F_i(x^k, \xi^k) - x_i^k)I(i \in Y^k)$$

in asynchronous form, where $\{\xi^k\}$ is an independently and identically distributed (i.i.d.) stochastic noise sequence taking values in some measurable space, and the function $F(\cdot,\cdot)$ is assumed to satisfy the nonexpansivity property:

$$||F(x,u) - F(y,u)||_p \leq ||x-y||_p$$

for some $p \in (0,\infty]$ and all x, y, u. Let $T(x) = E[F(x,\xi^k)]$. Then

$$||T(x) - T(y)||_p \leq ||x-y||_p.$$

The aim is to find a fixed point x^* of $T(\cdot)$, i.e., a point x^* satisfying $x^* = T(x^*)$, which we assume to exist uniquely. Define $h(x) = T(x) - x$ and $w^k = F(x^k,\xi^k) - T(x^k)$, which casts this algorithm into the form analyzed above. Note, in particular, that in view of our hypotheses on F, $E[||w^k||^2/\mathcal{F}^k] \leq c(||x^k||^2 + 1)$ for some $c > 0$. The foregoing then leads to the following.

PROPOSITION 4.1. *Let $\{x^k\}$ be generated by the synchronous stochastic approximation algorithm* (1). *Let Assumptions* 3.1(m) *and* 3.2(m) *hold. Then the sequence $\{x^k\}$ converges to x^* w.p.*1.

Proof. The theorem is an application of Lemmas 2.1 and 4.2, the global asymptotic stability of the equilibrium x^* for the ODE $\dot{x}(t) = T(x(t)) - x(t)$ being proved in Borkar and Soumyanath (1997). □

PROPOSITION 4.2. *Let $\{x^k\}$ be generated by the asynchronous version of the above algorithm. Let Assumptions* 3.1(m)*,* 3.2(m)*,* 3.3*, and* 3.4 *hold with the modifications stated above. Then the sequence $\{x^k\}$ converges to x^* w.p.*1.

Proof. The theorem is an application of Lemmas 2.1 and 4.3, the global asymptotic stability of the ODE $\dot{x}(t) = (1/n)(T(x(t)) - x(t))$ being ensured as before by observing that the scalar $1/n$ on its R.H.S. represents a mere time scaling. □

5. Analysis of *Q*-learning algorithms. The convergence theorems above are directly applicable to the analysis of Q-learning algorithms for discounted and SSP dynamic programming problems. As discussed in Bertsekas (2001, Vol. 1), discounted cost problems can be formulated as SSP problems. We will therefore restrict ourselves to SSP problems. Here we have a controlled discrete-time dynamic system where at state i the use of a control u specifies the transition probability $p_{ij}(u)$ to the next state j. There are a finite number of states. At state i, the control u is constrained to take values from a given finite control set $U(i)$. The cost of using u at state i and moving to state j is denoted by $g(i, u, j)$. We assume that there is a special cost-free termination state 0. Once the system reaches that state, it remains there at no further cost; that is, $p_{00}(u) = 1$ for all u. We denote by $1, \ldots, n$ the states other than the termination state 0.

The total expected cost associated with an initial state i and a policy $\pi = \{\mu_0, \mu_1, \ldots\}$, where each μ_k maps states i into controls $\mu_k(i) \in U(i)$, is

$$J_\pi(i) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N} g\big(x_k, \mu_k(x_k), x_{k+1}\big) \;\Big|\; x_0 = i\right\}.$$

Note that the discounted cost problem with discount factor $\alpha \in (0, 1)$ and states $i = 1, \ldots, n$ is obtained as the special case of an SSP problem, where $p_{i0}(u) = 1 - \alpha$ and $g(i, u, 0) = 0$ for all $i = 1, \ldots, n$ and $u \in U(i)$.

A stationary policy is a policy of the form $\pi = \{\mu, \mu, \ldots\}$, and its corresponding cost function is denoted by $J_\mu(i)$. We call a stationary policy π *proper* if there exists an integer m such that

$$\max_{i=1,\ldots,n} P\{x_m \neq 0 \mid x_0 = i, \pi\} < 1,$$

and call π *improper* otherwise. We assume the following.

ASSUMPTION 5.1. *There exists at least one proper policy.*

ASSUMPTION 5.2. *Every improper policy results in infinite expected cost from at least one initial state.*

These assumptions, introduced by Bertsekas and Tsitsiklis (1991), have become standard in the analysis of SSP problems and are sufficient to show the validity of the major types of dynamic programming results. For example, the value iteration method converges to the optimal cost function J^*, which is the unique solution of Bellman's equation

$$J^*(i) = \min_{u\in U(i)} \sum_{j=0}^{n} p_{ij}(u)\big(g(i, u, j) + J^*(j)\big), \qquad i = 1, \ldots, n,$$

$$J^*(0) = 0.$$

Q-learning algorithms update estimates of the Q-factors, defined for all pairs (i,u) by

$$Q^*(i,u) = \sum_{j=0}^{n} p_{ij}(u)\big(g(i,u,j) + J^*(j)\big).$$

From this definition and Bellman's equation, we see that the Q-factors are the unique solution of the following system of equations:

$$Q(i,u) = \sum_{j=0}^{n} p_{ij}(u)\left(g(i,u,j) + \min_{v\in U(j)} Q(j,v)\right), \qquad i = 1,\dots,n,\ u \in U(i),$$

$$Q(0,u) = 0,$$

which may be viewed as Bellman's equation for Q-factors.

Let us generically denote by Q the vector of Q-factors. The synchronous version of Q-learning is given by

$$Q^{k+1} = Q^k + \gamma(k)\big(F(Q^k,\xi^k) - Q^k\big), \tag{14}$$

where $\{\xi^k\}$ is a sequence of independent vector-valued random variables taking the values $0,1,\dots,n$, with probabilities Prob $(\xi^k_{iu} = j) = p_{ij}(u)$ for all k,

$$F(Q,\xi)(i,u) = g(i,u,\xi_{iu}) + \min_{v\in U(\xi_{iu})} Q(\xi_{iu},v).$$

The initial condition is assumed to satisfy $Q^0(0,u) = 0$, which ensures that $Q^k(0,u) = 0$ for all k. Also, for $i = 0$, $\xi_{iu} = 0$ w.p.1. Thus $g(i,u,\xi_{iu}) = g(0,u,0) = 0$ (because 0 is a cost-free state) and $Q(\xi_{iu},u) = Q(0,u) = 0$ for all u. Thus $F(Q,\xi)(0,u) = 0$ for all u. In fact, this permits us to consider the iteration of $Q^k(i,u)$ for $1 \le i \le n$ alone, which we denote again by Q^k by abuse of notation. Define

$$T(Q)(i,u) = \sum_{j=1}^{n} p_{ij}(u)F(Q,j)$$

and

$$w^k = F(Q^k,\xi^k) - T(Q^k).$$

Assumption 3.2 applies to the stepsize $\gamma(k)$ and the noise w^k for the *rescaled* iterates.

The following two properties of the mapping T are significant for our purposes:

1. T is nonexpansive with respect to the sup-norm.
2. The unique fixed point Q^* of the mapping T is a global asymptotically stable equilibrium of the ODE $\dot{Q} = T(Q) - Q$.

Property 1 follows from the nonexpansiveness of F, which can be verified by noting that for all $Q_1, Q_2 \in \mathbb{R}^{n+m}$ we have

$$\begin{aligned} F(Q_1,\xi)(i,u) - F(Q_2,\xi)(i,u) &= \min_{u'} Q_1(\xi_{iu},u') - \min_{u'} Q_2(\xi_{iu},u') \\ &\le Q_1(\xi_{iu},u_2) - Q_2(\xi_{iu},u_2) \\ &\le \max_{(i,u)} |Q_1(i,u) - Q_2(i,u)| \\ &\le \|Q_1 - Q_2\|_\infty, \end{aligned}$$

where u_2 achieves the minimum in $\min_{u'} Q_2(\xi_{iu}, u')$. A symmetric argument shows that

$$F(Q_2, \xi)(i,u) - F(Q_1, \xi)(i,u) \le \|Q_1 - Q_2\|_\infty.$$

Property 2 follows from the analysis of Bellman's equation for SSP problems (see e.g., Bertsekas (2001, Vol. 2)), and from the analysis of ODE maps involving nonexpansive mappings in Borkar and Soumyanath (1997). Using the facts that Q^* is the unique fixed point of T and that T is nonexpansive, it follows that any solution trajectory $Q(t)$ converges to Q^*. Moreover, the analysis in Borkar and Soumyanath (1997) implies that $\|Q(t) - Q^*\|_\infty$ is nonincreasing, establishing that Q^* is a global asymptotically stable equilibrium point for the ODE.

The mapping F, in addition to being nonexpansive, satisfies

$$E\big[F(Q^k, \xi^k) \mid \mathcal{F}^k\big] = T(Q^k),$$

where

$$\mathcal{F}^k = \sigma(x^k, \ldots, x^0, \xi^{k-1}, \ldots, \xi^0).$$

The properties above are sufficient to show that all of the assumptions of Proposition 4.1 are satisfied, thus implying the following convergence result.

PROPOSITION 5.1. *The sequence* $\{Q^k\}$ *generated by the synchronous Q-learning iteration* (14) *converges to* Q^* *w.p.*1.

5.1. Analysis of the SSP asynchronous *Q*-learning. The asynchronous version of (14) is what is usually referred to as the Q-learning algorithm. It is written as

$$Q^{k+1}(i,u) = Q^k(i,u) + \gamma\big(\nu(k,\phi^k)\big)\big(F(Q^k,\xi^k)(i,u) - Q^k(i,u)\big) I\big((i,u) = \phi^k\big), \tag{15}$$

where $\{\xi^k\}$ is as defined above and $\{\phi^k\}$ is a random process. Again we impose Assumption 3.2′ on the stepsize, and we assume in addition that

1.
$$\liminf_{k\to\infty} \frac{1}{k+1}\nu(k,i,a) \ge \Delta \qquad \text{for some } \Delta > 0.$$

Furthermore, for all $T > 0$, the limit

$$\lim_{n\to\infty} \frac{\sum_{k=\nu(n,i,a)}^{\nu(m_T(n),i,a)} \gamma(k)}{\sum_{k=\nu(n,j,b)}^{\nu(m_T(n),j,b)} \gamma(k)}$$

exists w.p.1 for all i, j, a, b.

2. $\{\gamma(k)\}$ is as in Assumption 3.3.

Again the mapping F satisfies

$$E\big[F(Q^k, \xi^k) \mid \mathcal{F}^k\big] = T(Q^k),$$

with

$$\mathcal{F}^k = \sigma(x^k, \ldots, x^0, \xi^{k-1}, \ldots, \xi^0, \phi^k, \ldots, \phi^0).$$

Similarly, the assumptions of Proposition 4.2 are satisfied, and we have the following.

PROPOSITION 5.2. *The sequence* $\{Q^k\}$ *generated by the asynchronous* Q*-learning iteration* (15) *converges to* Q^* *w.p.*1.

As already mentioned, the case in which more than one component is updated at a time can be reduced to the one above, modulo bounded delays, which can be separately taken care of as in Borkar (1998).

Remark 5.1. The usual formalism for Q-learning algorithms (see, e.g., Bertsekas and Tsitsiklis (1996)) presupposes the availability of a simulation device that generates independent random variables $\{\xi^k\}$ as above. Alternatively, one may consider it as an on-line scheme where the samples are generated by a single simulation or actual run $\{X^k\}$ of the controlled Markov chain with the control process $\{Z^k\}$. Then it is asynchronous, with $\phi^k = (X^k, Z^k)$. The above framework still applies if we use the representation $X^{k+1} = f(X^k, Z^k, \xi^k)$, where $\{\xi^k\}$ are i.i.d. and f is a suitable map. Such a representation is always possible (albeit on a possibly augmented probability space) by the stochastic realization theoretic results of Borkar (1993). See Kifer (1986) for the uncontrolled case.

6. Some extensions. This section points out some important extensions of the preceding analysis. The first is an extension of Lemma 2.1. It is possible to replace the assumption of nonexpansivity with respect to a norm there by nonexpansivity with respect to the span seminorm $||\cdot||_s$, defined by

$$||x||_s = \max_{i=1,\dots,n} x_i - \min_{i=1,\dots,n} x_i,$$

where $x_1, \dots, x_n$ are the components of x. In this case, however, a weaker boundedness result is obtained, which is the subject of the following lemma. This lemma is used crucially in our companion paper on Q-learning in average cost control (Abounadi, Bertsekas, and Borkar (2001)).

LEMMA 6.1. *Let* $\mathcal{B}$ *be an open and bounded subset of* $\Re^n$ *containing the origin, and let* $\mathcal{C}$ *be a subset of* $\Re^n$ *that contains* $\mathcal{B}$*. Consider the algorithm*

$$x^{k+1} = G^k(x^k, \xi^k),$$

where we assume the following:

1. $\{\xi^k\}$ *is a sequence in a measurable space* $(\Omega, \mathcal{F})$.
2. G^k *is nonexpansive in* x *with respect to the span seminorm; i.e., for every* $\xi \in \Omega$,

$$||G^k(x,\xi) - G^k(y,\xi)||_s \le ||x - y||_s \qquad \forall\ x, y, \xi.$$

3. *The sequence* $\{\tilde{x}^k\}$ *generated by the scaled iteration*

$$\tilde{x}^{k+1} = G^k(\Pi_{\mathcal{B},\mathcal{C}}(\tilde{x}^k), \xi^k), \qquad \tilde{x}^0 = x^0,$$

converges to some vector $x^* \in \mathcal{B}$.

Then $\{||x^k||_s\}$ *remain bounded.*

Proof. The proof is identical to that of Lemma 2.1. □

The second extension relates to the Q-learning schemes described above. One can also allow for random costs under mild technical conditions. Thus, let a real or simulated transition from i to j under control u at time k lead to a random cost ζ_{iuj}^{k+1}. We suppose that $E[\zeta_{iuj}^{k+1} \mid \mathcal{F}^{k+1}] = g(i,u,j)$ and $E[(\zeta_{iuj}^{k+1})^2 \mid \mathcal{F}^{k+1}] \le M$ w.p.1 for some constant $M < \infty$. (Compare with Remark 3.2.) Then the foregoing analysis

goes through exactly as before with one modification: the "martingale difference" sequence w^k gets replaced by $\hat{w}^k$, defined as follows: its (iu)th component is $\hat{w}^k_{iu} = w^k_{iu} + \zeta^{k+1}_{iu\xi^k_{iu}} - g(i,u,\xi^k_{iu})$, where w^k_{iu} is the (iu)th component of w^k. Note that $\{\hat{w}^k\}$ is also a martingale difference sequence. An example is the case in which $\zeta^{k+1}_{iuj} = g(i,u,j)+\psi^{k+1}_{iuj}$, where $\{\psi^n_{iuj}\}$ are i.i.d. zero mean, bounded variance random variables representing additive noise.

7. Conclusions. In this paper we have studied the convergence of synchronous and asynchronous algorithms involving nonexpansive maps and additive deterministic or stochastic noise. We have used the ODE approach, but we have dispensed with the restrictive boundedness assumption on the generated iterates that this approach requires. The nonexpansiveness property ensures that the distance between the iterates of two instantiations of the algorithm, driven by the same noise sequence and differing only in the initial conditions, remains bounded. In fact, our arguments will work for any algorithm for which this is true, and the associated ODE has a globally asymptotically stable equilibrium, under mild technical conditions on noise as above. As a special case of our analysis, we have discussed Q-learning algorithms for SSP problems, and we have refined the assumptions under which convergence can be proved. Our results used Lemma 2.1 for the boundedness argument. We can likewise use Lemma 6.1 to prove boundedness for certain Q-learning algorithms for the average cost dynamic programming problem. The analysis of these algorithms requires considerable additional machinery and is given separately in a companion paper (Abounadi, Bertsekas, and Borkar (2001)).

Appendix. Here we briefly recall the main results of Borkar (1998) that are used in the paper. Let $F(\cdot,\cdot) = [F_1(\cdot,\cdot),\ldots,F_d(\cdot,\cdot)]^T : \mathcal{R}^d \times \mathcal{R}^m \to \mathcal{R}^d$ be Lipschitz in its first argument uniformly w.r.t. the second. Consider the stochastic approximation algorithm of the form

$$x^{k+1} = x^k + \gamma(k)F(x^k,\xi^k), \quad k \geq 0,$$

for $x^k = [x^k_1,\ldots,x^k_d]$. Let $h(x) = E[F(x,\xi^1)]$. We assume that the ODE $\dot{x}(t) = h(x(t))$ has a globally asymptotically stable equilibrium x^*. The asynchronous version of this algorithm is given by

$$x^{k+1}_i = x^k_i + \gamma\big(\nu(k,i)\big)I(i \in Y^k)F_i(x_1^{k-\tau_{1i}(k)},\ldots,x_d^{k-\tau_{di}(k)},\xi^k), \quad 1 \leq i \leq d,$$

for $k \geq 0$, where

(1) $\{Y^k\}$ is a set-valued random process taking values in the subsets of the set $\{1,\ldots,d\}$, representing the components that do get updated at time k.
(2) $\{\tau_{ij}(k), 1 \leq i,j \leq d, k \geq 0\}$ are bounded random delays. One usually takes $\tau_{ii}(k) = 0$ for all i, though this is not necessary. (Borkar (1998) also relaxes the boundedness condition on delays to a conditional moment bound.)
(3) $\nu(k,i) = \sum_{m=0}^{k} I(i \in Y^m)$ denotes the number of times component i gets updated until time k.

Let Assumptions 3.1–3.4 hold. The main result of Borkar (1998) is the following.

THEOREM A.1. *If $\{x^k\}$ remain w.p.1 bounded, they converge to x^* w.p.1.*

We shall briefly describe what the proof entails, using the notation of section 3.2 above. The intuition behind why the bounded delays don't affect the asymptotics is simple. Recall that the passage from the discrete iteration to an interpolated "approximation to ODE" involves the time scaling $k \to t^k$. This scaling shrinks the

time axis more and more as k increases, because $t^{k+1} - t^k \to 0$. If K denotes a bound on the delays, the intervals $[k, k+1, \ldots, k+K]$ map to $[t^k, t^k + \sum_{m=k}^{k+K-1} \gamma(m)]$, which become smaller and smaller as k increases, because of which the delays as seen by the ODE approximation on the rescaled time become smaller and smaller, becoming asymptotically negligible. This intuition can be made precise quite easily. In fact, it simply contributes one additional asymptotically negligible error term to the usual ODE analysis of stochastic approximations. See Lemma 3.3 of Borkar (1998) for details.

The harder problem is to deal with the Y^k's, i.e., with the fact that not all components are getting updated at each step. As in section 3.2 above, one has $X_l^k(t) = X_l^k(0) + \int_0^t \mathrm{diag}(\mu^k(\tau))h(X_l^k(\tau))d\tau$ + error terms, the latter going to zero w.p.1 as $k \to \infty$. View $\mu^k(\cdot)$ as elements of the space of measurable maps $[0, \infty) \to \{d$-dimensional probability vectors$\}$, with the coarsest topology that renders continuous the maps $\mu(\cdot) \to \int_0^T \langle \mu(t), g(t) \rangle dt$ for any $T > 0$ and any $g : [0, T] \to \mathcal{R}^d$ that satisfies $\int_0^T ||g(t)||^2 dt < \infty$. (Recall Lemma 3.5 above.) This is a compact metrizable topology. Let $\mu^k(\cdot)$ converge along a subsequence to some $\hat{\mu}(\cdot)$ in this topology. Then the limiting trajectory of $X_l^k(\cdot)$ along this subsequence will satisfy the nonautonomous ODE

$$\dot{x}(t) = \mathrm{diag}(\hat{\mu}(t))h(x(t)).$$

The additional conditions on $\gamma(k)$ stipulated in Assumptions 3.3 and 3.4 are required to further ensure that $\hat{\mu}(t)$ in fact equals μ^* for almost every t. See Borkar (1998) for details.

One can, in fact, work with the nonautonomous ODE itself to draw the same conclusions by using Lemma 2.4 of Borkar (1998), the only requirement being that the components of $\hat{\mu}(t)$ remain uniformly bounded away from zero from below for almost every t. This is a weaker version of the statement "all components get updated comparably often." Unfortunately, no simple transparent sufficient condition to ensure this (short of Assumptions 3.3, 3.4) seems available.

Acknowledgment. Thanks are due to John Tsitsiklis, whose suggestions resulted in important simplifications of the lemmas in section 2.

REFERENCES

J. Abounadi, D. P. Bertsekas, and V. S. Borkar (2001), *Learning algorithms for Markov decision processes with average cost*, SIAM J. Control Optim., 40, pp. 681–698.

D. P. Bertsekas and J. N. Tsitsiklis (1989), *Parallel and Distributed Computation: Numerical Methods*, Prentice–Hall, Englewood Cliffs, NJ.

D. P. Bertsekas and J. N. Tsitsiklis (1991), *An analysis of stochastic shortest path problems*, Math. Oper. Res., 16, pp. 580–595.

D. P. Bertsekas and J. N. Tsitsiklis (1996), *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.

D. P. Bertsekas (2001), *Dynamic Programming and Optimal Control*, 2nd ed., Athena Scientific, Belmont, MA.

A. Benveniste, M. Metivier, and P. Priouret (1990), *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York.

P. Billingsley (1968), *Convergence of Probability Measures*, John Wiley, New York.

V. S. Borkar (1993), *White noise representations in stochastic realization theory*, SIAM J. Control Optim., 31, pp. 1093–1102.

V. S. Borkar (1995), *Probability Theory: An Advanced Course*, Springer-Verlag, New York.

V. S. Borkar (1998), *Asynchronous stochastic approximations*, SIAM J. Control Optim., 36, pp. 840–851. Correction note in ibid, 38 (2000), pp. 662–663.

V. S. Borkar and S. P. Meyn (2000), *The O.D.E. method for convergence of stochastic approximation and reinforcement learning*, SIAM J. Control Optim., 38, pp. 447–469.

V. S. Borkar and K. Soumyanath (1997), *An analog parallel scheme for fixed point computation—Part* I*: Theory*, IEEE Trans. Circuits Systems I Fund. Theory Appl., 44, pp. 351–355.

S. Csibi (1975), *Learning under computational constraints from weakly dependent samples*, Prob. Control Inform. Theory, 4, pp. 3–21.

L. Gerencsér (1992), *Rate of convergence of recursive estimators*, SIAM J. Control Optim., 30, pp. 1200–1227.

T. Jaakola, M. I. Jordan, and S. P. Singh (1994), *On the convergence of stochastic iterative dynamic programming algorithms*, Neural Computation, 6, pp. 1185–1201.

Y. Kifer (1986), *Ergodic Theory of Random Transformations*, Birkhäuser Boston, Cambridge, MA.

H. J. Kushner and D. S. Clark (1978), *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, New York.

H. J. Kushner and G. Yin (1997), *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York.

L. Ljung (1977), *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Control, 22, pp. 551–575.

K. Soumyanath and V. S. Borkar (1999), *An analog scheme for fixed point computation—Part* II*: Applications*, IEEE Trans. Circuits Systems I Fund. Theory Appl., 46, pp. 442–451.

P. Tseng, D. P. Bertsekas, and J. N. Tsitsiklis (1990), *Partially asynchronous parallel algorithms for network flow and other problems*, SIAM J. Control Optim., 28, pp. 678–710.

J. N. Tsitsiklis (1994), *Asynchronous stochastic approximation and Q-learning*, Machine Learning, 16, pp. 185–202.

C. J. C. H. Watkins (1989), *Learning from delayed rewards*, Ph.D. thesis, Cambridge University, Cambridge, England.

C. J. C. H. Watkins and P. Dayan (1992), *Q-learning*, Machine Learning, 8, pp. 279–292.

F. W. Wilson (1969), *Smoothing derivatives of functions and applications*, Trans. Amer. Math. Soc., 139, pp. 413–428.

T. Yoshizawa (1966). *Stability Theory by Lyapunov's Second Method*, Mathematical Society of Japan, Tokyo.