12.8 Convergence for Q-Learning: Discounted Cost

The differential inclusions form of Theorem 3.5 will be applied to prove convergence of the Q-learning algorithm of Section 2.3 for the discounted case $\beta < 1$. The unconstrained algorithm for the decreasing step size case is (2.3.3) or (2.3.4); the reader is referred to that section for the notation. We will work with a practical constrained form and let the step size be ϵ, a constant. The proof for the decreasing step size case is virtually the same. The values of $Q_{n,id}$ will be truncated at $\pm B$ for large B. Thus $H = [-B, B]^r$, where r is the number of possible (state, action) pairs. Define $C = \max\{\bar{c}_{id}\}$. Then the Q-values under any policy are bounded by $C/(1-\beta)$ [This does not imply that the Q-values given by an *unconstrained* stochastic approximation will be no greater than $C/(1-\beta)$.] Let $B > C/(1-\beta)$.

Let ψ_n^ϵ denote the state of the controlled chain at time n. The state will generally depend on the step size ϵ, since it depends on the control policy that, in turn, depends on the current estimate of the optimal Q-values. Recall that the Q-value for the (state, action) pair (i, d) is updated at time $n+1$ (after the next state value is observed) if that pair (i, d) occurs at time n. Thus, we can suppose that the updates are at times $n = 1, 2, \ldots$. Let $\hat{Q}_{n,id}^\epsilon$ denote the Q-values at real time n and let $I_{n,id}^\epsilon$ be the indicator of the event that (i, d) occurs at real time n.

The algorithm can be written as

$$\hat{Q}_{n+1,id}^\epsilon = \Pi_{[-B,B]}\left[\hat{Q}_{n,id}^\epsilon + \epsilon\left(c_{n,id} + \beta \min_{v \in U(\psi_{n+1}^\epsilon)} \hat{Q}_{n,\psi_{n+1}^\epsilon v}^\epsilon - \hat{Q}_{n,id}^\epsilon\right) I_{n,id}^\epsilon\right] \tag{8.1}$$

or, equivalently, as

$$\hat{Q}_{n+1,id}^\epsilon = \Pi_{[-B,B]}\left[\hat{Q}_{n,id}^\epsilon + \epsilon\left(T_{id}(\hat{Q}_n^\epsilon) - \hat{Q}_{n,id}^\epsilon + \delta M_n^\epsilon\right) I_{n,id}^\epsilon\right], \tag{8.2}$$

where the operator T_{id} is defined below (2.3.3) and the noise δM_n^ϵ is defined above (2.3.4).

In the absence of the constraint, the general stability results of Section 7 can be used to prove tightness of $\{\hat{Q}_{n,id}^\epsilon; n, i, d, \epsilon\}$, which can then be used to get the convergence.

Suppose that there are $n_0 < \infty$ and $\delta_0 > 0$ such that for each state pair i, j,

$$\inf P\left\{\psi_{n+k}^\epsilon = j, \text{ some } k \le n_0 \middle| \psi_n^\epsilon = i\right\} \ge \delta_0, \tag{8.3}$$

where the inf is over the time n and all ways in which the experimenter might select the controls. Let $\delta\tau_{n,id}^\epsilon$ denote the time interval between the nth and $(n+1)$st occurrences of the (state, action) pair (i, d), and let $E_{n,id}^{\epsilon,+}$ denote the expectation, conditioned on all the data up to and including the

$(n+1)$st update for the pair (i,d). Define $u^\epsilon_{n,id}$ by

$$E^{\epsilon,+}_{n,id}\delta\tau^\epsilon_{n+1,id} = u^\epsilon_{n+1,id},$$

and suppose that the (possibly random) $\{u^\epsilon_{n,id}; n, \epsilon\}$ are uniformly bounded by a real number $\bar{u}_{id}$ (which must be ≥ 1) and that $\{\delta\tau_{n,id}; n, i, d\}$ is uniformly integrable.

Let $\hat{Q}^\epsilon_{id}(\cdot)$ denote the piecewise constant interpolation of $\{\hat{Q}^\epsilon_{n,id}; n < \infty\}$ in scaled real time, that is, with interpolation intervals of width ϵ. Define $\hat{Q}^\epsilon(\cdot) = \{\hat{Q}^\epsilon_{id}(\cdot);\ i, d\}$. The problem is relatively easy to deal with since the noise terms δM^ϵ_n are martingale differences. Under the given conditions, for any sequence of real numbers T_ϵ, $\{Q^\epsilon_{id}(T_\epsilon + \cdot)\}$ is tight, and (if $T_\epsilon \to \infty$) it will be seen that it converges weakly (as $\epsilon \to 0$) to the process with constant value $\bar{Q}$, the optimal value defined by (2.3.2).

The limit mean ODE will be shown to be

$$\dot{Q}_{id} = D_{id}(t)\,(T_{id}(Q) - Q_{id}) + z_{id}, \quad \text{all } i, d, \tag{8.4}$$

where the $z_{id}(\cdot)$ serve the purpose of keeping the values in the interval $[-B, B]$ and will be shown to be zero. The values of $D_{id}(\cdot)$ lie in the intervals $[1/\bar{u}_{id}, 1]$. We next show that all solutions of (8.4) tend to the unique limit point $\bar{Q}$. Suppose that $Q_{id}(t) = B$ for some t and i, d pair. Note that, by the lower bound on B,

$$\sup_{Q:Q_{id}=B} [T_{id}(Q) - Q_{id}] \leq C + \beta\sum_j p_{ij}(d)B - B = C + (\beta - 1)B < 0.$$

This implies that $\dot{Q}_{id}(t) < 0$ if $Q_{id}(t) = B$. Analogously, $\dot{Q}_{id}(t) > 0$ if $Q_{id}(t) = -B$. Thus, the boundary of the constraint set H is not accessible by a trajectory of (8.4) from any interior point. Now, dropping the z_{id} terms, by the contraction property of $T(\cdot)$, $\bar{Q}$ is the unique limit point of (8.4).

We need only prove that (8.4) is the limit mean ODE by verifying the conditions of the differential inclusions part of Theorem 3.5. The expectation of the observation in (8.2) used at time $(n+1)$, given the past and the fact that the pair (i,d) occurred at time n, is

$$\bar{g}_{id}(\hat{Q}^\epsilon_n) \equiv T_{id}(\hat{Q}^\epsilon_n) - \hat{Q}^\epsilon_{id}.$$

Thus, the $g^\epsilon_{n,id}(\cdot)$ of Theorem 3.5 are all $\bar{g}_{id}(\cdot)$. Thus, (A3.7) holds, with the $\beta^\epsilon_{n,\alpha}$-terms and the delays being zero. [Delays can be allowed, provided that they are uniformly integrable, but in current Q-learning applications, delays are of little significance.]

Condition (A3.9) is obvious (there are no $\xi^\epsilon_{n,\alpha}$). Also the conditions for the differential inclusion result in Theorem 3.5 hold, where $U_{id}(\theta) = U_{id} = [1, \bar{u}_{id}]$. Condition (A3.1′) is obvious. All the other conditions in the theorem are either obvious or not applicable. Thus, Theorem 3.5 holds.

4.1 Martingales, Submartingales, and Inequalities

Let $(\Omega, \mathcal{F}, P)$ denote a probability space, where Ω is the sample space, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a probability measure on $(\Omega, \mathcal{F})$. The symbol ω denotes the canonical point in Ω. All subsequent random variables will be defined on this space. Let $\{M_n\}$ be a sequence of random variables that can be either real or vector-valued and that satisfies $E|M_n| < \infty$ for each n. If

$$E\left[M_{n+1}|M_i, i \le n\right] = M_n \text{ w.p.1 for all } n, \tag{1.1}$$

then $\{M_n\}$ is said to be a *martingale* or a martingale sequence. The difference $\delta M_n = M_{n+1} - M_n$ is called a *martingale difference*. By the definition of a martingale, if $E|M_n|^2 < \infty$ for each n, the martingale differences are uncorrelated in that for $m \neq n$,

$$E\left[M_{n+1} - M_n\right]\left[M_{m+1} - M_m\right]' = 0.$$

To provide a simple example, suppose that a gambler is playing a sequence of card games, and M_n is his "fortune" after the nth game. If $\{M_n\}$ is a martingale process, we say that the game is "fair" in that the expectation of $M_{n+1} - M_n$ conditioned on the past fortunes $\{M_i, i \le n\}$ is zero. In applications, the M_n themselves are often functions of other random variables. For example, we might have $M_{n+1} = f_n(\xi_i, i \le n)$ for some sequence of random variables $\{\xi_n\}$ and measurable functions $\{f_n(\cdot)\}$. In the gambling example, ξ_n might represent the actual sequence of cards within the nth game. Then we could say that the game is fair if $E\left[M_{n+1} - M_n|\xi_i, i \le n, M_0\right] = 0$. This suggests that for use in applications, it is convenient to define a martingale somewhat more generally, as follows. Let $\{\mathcal{F}_n\}$ be a sequence of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, for all n. Suppose that M_n is measurable on $\mathcal{F}_n$ [e.g., in the preceding example, $\mathcal{F}_n$ might be determined by $\{\xi_i, i \le n, M_0\}$]. Write the expectation conditioned on $\mathcal{F}_n$ as $E_{\mathcal{F}_n}$. If

$$E_{\mathcal{F}_n} M_{n+1} = M_n \text{ w.p.1 for all } n, \tag{1.2}$$

then we say that either $\{M_n, \mathcal{F}_n\}$ is a *martingale* or $\{M_n\}$ is an $\mathcal{F}_n$-martingale. If we simply say that $\{M_n\}$ is a martingale and do not specify $\mathcal{F}_n$, then we implicitly suppose (as we can always do) that it is just the σ-algebra generated by $\{M_i, i \le n\}$. Martingales are one of the fundamental types of processes occurring in stochastic analysis, and there is a very large literature dealing with them; see, for example, [21, 133]. Note that if an $\mathcal{F}_n$-martingale is vector-valued, then each of the real-valued components is also an $\mathcal{F}_n$-martingale. Similarly, a finite collection of real-valued $\mathcal{F}_n$-martingales is a vector-valued $\mathcal{F}_n$-martingale.

Let M_n be real-valued, and replace the equality in (1.2) by the inequality $\le$, as in

$$E_{\mathcal{F}_n} M_{n+1} \le M_n \text{ w.p.1 for all } n. \tag{1.3}$$

Then we say either that $\{M_n, \mathcal{F}_n\}$ is a *supermartingale* or that $\{M_n\}$ is an $\mathcal{F}_n$-supermartingale. If the $\mathcal{F}_n$ are understood, then we might just say that $\{M_n\}$ is a supermartingale. If the inequality has the form

$$E_{\mathcal{F}_n} M_{n+1} \geq M_n \text{ w.p.1 for all } n,$$

then the process is called a *submartingale*. In the applications in this book, martingale processes occur when we decompose each member of a sequence of random variables into a part "depending on the past" and an "unpredictable" part. For example, let $\{Y_n\}$ be a sequence of random variables with $E|Y_n| < \infty$ for each n. Write

$$Y_n = (Y_n - E[Y_n|Y_i, i < n]) + E[Y_n|Y_i, i < n].$$

The first "unpredictable" part is a martingale difference, because the process defined by the sum

$$M_n = \sum_{j=0}^{n} (Y_j - E[Y_j|Y_i, i < j])$$

is a martingale. There are many useful inequalities and limit theorems associated with martingale-type processes that facilitate the analysis of their sample paths.

Martingale inequalities and a convergence theorem. Let $\{M_n, \mathcal{F}_n\}$ be a martingale, which is supposed to be real-valued with no loss in generality. The following key inequalities can be found in [21, Chapter 5], [51, Chapter 1] and [133, Chapter IV.5]. Let $q(\cdot)$ be a non-negative nondecreasing convex function. Then for any integers $n < N$ and $\lambda > 0$,

$$P_{\mathcal{F}_n}\left\{\sup_{n\leq m\leq N} |M_m| \geq \lambda\right\} \leq \frac{E_{\mathcal{F}_n} q(M_N)}{q(\lambda)}, \tag{1.4}$$

an inequality that will be useful in providing sharp bounds on the excursions of stochastic approximation processes. Commonly used forms of $q(\cdot)$ are $q(M) = |M|$, $q(M) = |M|^2$, and $q(M) = \exp(\alpha M)$ for positive α. The inequality

$$E_{\mathcal{F}_n}\left[\sup_{n\leq m\leq N} |M_m|^2\right] \leq 4E_{\mathcal{F}_n}|M_N|^2, \tag{1.5}$$

will also be useful.

If $\{M_n, \mathcal{F}_n\}$ is a non-negative supermartingale, then for integers $n < N$

$$P_{\mathcal{F}_n}\left\{\sup_{n\leq m\leq N} M_m \geq \lambda\right\} \leq \frac{M_n}{\lambda}. \tag{1.6}$$

Let y^- denote the absolute value of the negative part of the real number y, defined by $y^- = \max\{0, -y\}$. Let $\{M_n, \mathcal{F}_n\}$ be a real-valued submartingale with $\sup_n E|M_n| < \infty$. Then the martingale convergence theorem [21, Theorem 5.14] states: $\{M_n\}$ *converges with probability one, as* $n \to \infty$. A supermartingale $\{M_n\}$ converges with probability one if $\sup_n E[M_n]^- < \infty$.

Stopping times. Let $\mathcal{F}_n$ be a sequence of nondecreasing σ-algebras. A random variable τ with values in $[0, \infty]$ (the set of extended real numbers) is said to be an $\mathcal{F}_n$-*stopping time* (or simply a stopping time if the σ-algebras are evident) if $\{\tau \le n\} \in \mathcal{F}_n$ for each n. Let $\mathcal{F}_n$ be the σ-algebra determined by a random sequence $\{\xi_i, i \le n\}$. Then, if τ is an $\mathcal{F}_n$-stopping time, whether or not the event $\{\tau \le n\}$ occurred can be "determined" by watching ξ_i up to and including time n. If a stopping time is not defined at some ω, we always set its value equal to infinity at that ω. Let $\{M_n, \mathcal{F}_n\}$ be a martingale (resp., a sub- or supermartingale) and let τ be a bounded (uniformly in ω) $\mathcal{F}_n$-stopping time. Define

$$\tau \wedge n = \min\{\tau, n\}.$$

Then $\{M_{\tau \wedge n}, \mathcal{F}_n\}$ is a martingale (resp., a sub- or supermartingale).

Continuous time martingales. The definitions of martingale and sub- and supermartingale extend to continuous time. Let $M(t)$ be a random process satisfying $E|M(t)| < \infty$ for each $t \ge 0$, and let $\mathcal{F}_t$ be a nondecreasing sequence of σ-algebras such that $M(t)$ is $\mathcal{F}_t$-measurable. If $E_{\mathcal{F}_t}[M(t+s) - M(t)] = 0$ with probability one for each $t \ge 0$ and $s > 0$, then $\{M(t), \mathcal{F}_t\}$ is said to be a martingale. If the sequence of σ-algebras $\mathcal{F}_t$ is understood, then it might be omitted. In this book, essentially all we need to know about continuous parameter martingales are the following facts. Some additional material on the topic of Theorem 1.2 is in Section 7.2.4.

Theorem 1.1. *A continuous time martingale whose paths are locally Lipschitz continuous with probability one on each bounded time interval is a constant with probability one.*

By *locally Lipschitz continuous with probability one*, we mean that for each $T > 0$ there is a random variable $K(T) < \infty$ with probability one such that for $t \le t + s \le T$,

$$|M(t+s) - M(t)| \le K(T)s.$$

The result will be proved (with no loss of generality) for real-valued $M(t)$. First we assume that $E|M(t)|^2 < \infty$ for each t and that the Lipschitz constant is bounded by a constant K, and show that $|M(t)|^2$ is also a martingale. For $\Delta > 0$, we can write

$$\begin{aligned} &E_{\mathcal{F}_t} M^2(t+\Delta) - M^2(t) \\ &\quad = 2E_{\mathcal{F}_t} M(t)\left[M(t+\Delta) - M(t)\right] + E_{\mathcal{F}_t}\left[M(t+\Delta) - M(t)\right]^2. \end{aligned}$$

Then, using the martingale property on the first term on the second line above and the Lipschitz condition on the second term yields that the expression is bounded in absolute value by $K^2\Delta^2$. Now, adding the increments over successive intervals of width Δ and letting $\Delta \to 0$ shows that $E_{\mathcal{F}_t} M^2(t+s) - M^2(t) = 0$ for any $s \geq 0$. Consequently, $|M(T+t) - M(T)|^2, t \geq 0$, is a martingale for each T. Since we have proved that the expectation of this last martingale is a constant and equals zero at $t = 0$, it is identically zero, which shows that $M(t)$ is a constant. If $E|M(t)|^2$ is not finite for all t, then a "stopping time" argument can be used to get the same result; that is, for positive N, work with $M(\tau_N \wedge \cdot)$ where $\tau_N = \min\{t : |M(t)| \geq N\}$ to show that $M(\tau_N \wedge \cdot)$ is a constant for each N. An analogous stopping time argument can be used if the Lipschitz constant is random.

Definition. Let $W(\cdot)$ be an $\mathbb{R}^r$-valued process with continuous paths such that $W(0) = 0$, $EW(t) = 0$, for any set of increasing real numbers $\{t_i\}$ the set $\{W(t_{i+1}) - W(t_i)\}$ is mutually independent and the distribution of $W(t+s) - W(t)$, $s > 0$, does not depend on t. Then $W(\cdot)$ is called a vector-valued *Wiener process* or *Brownian motion*, and there is a matrix Σ, called the covariance, such that $EW(t)W'(t) = \Sigma t$, and the increments are normally distributed [21].

The next theorem gives a convenient criterion for verifying that a process is a Wiener process. The criterion will be discussed further in Chapter 7.

Theorem 1.2. [43, Chapter 5, Theorem 2.12]. *Let $\{M(t), \mathcal{F}_t\}$ be a vector-valued martingale with continuous paths and let there be a matrix Σ such that for each t and $s \geq 0$,*

$$E_{\mathcal{F}_t}\left[M(t+s) - M(t)\right]\left[M(t+s) - M(t)\right]' = \Sigma s \text{ w.p.1.}$$

Then $M(\cdot)$ is a Wiener process with zero mean and covariance parameter Σ.

The Borel–Cantelli Lemma. Let A_n be events (i.e., sets in $\mathcal{F}$) and suppose that

$$\sum_n P\{A_n\} < \infty. \tag{1.7}$$

Then the *Borel-Cantelli Lemma* [21] states that for almost all ω only finitely many of the events A_n will occur.

Inequalities. Let X, Y be real-valued random variables. Then *Chebyshev's inequality* (see [21]) states that for any integer k and $\delta > 0$,

$$P\{|X| \geq \delta\} \leq \frac{E|X|^k}{\delta^k}. \tag{1.8}$$

Hölder's inequality states that for any positive p, q such that $1/p+1/q=1$,

$$E|XY| \leq E^{1/p}|X|^p E^{1/q}|Y|^q. \tag{1.9}$$

The special case $p = q = 2$ is called the *Schwarz inequality.* There is an analogous inequality for sums. Let $1/p + 1/q = 1$ with p, $q > 0$, and let X_n and Y_n be real-valued random variables and let $a_n \geq 0$ with $\sum_n a_n < \infty$. Then

$$E\left|\sum_n a_n X_n Y_n\right| \leq \left(E\sum_n a_n |X_n|^p\right)^{1/p} E\left(\sum_n a_n |Y_n|^q\right)^{1/q}. \tag{1.10}$$

Let $f(\cdot)$ be a convex function and $\mathcal{F}_0$ a σ-algebra, and suppose that $E|X| < \infty$ and $E|f(X)| < \infty$. Then *Jensen's inequality* is

$$\begin{aligned} &Ef(X) \geq f(EX) \quad \text{or with conditioning,} \\ &E_{\mathcal{F}_0} f(X) \geq f(E_{\mathcal{F}_0} X) \text{ w.p.1.} \end{aligned} \tag{1.11}$$

4.2 Ordinary Differential Equations

4.2.1 *Limits of a Sequence of Continuous Functions*

In this section, denote the Euclidean r-space by $I\!R^r$ with x being the canonical point. If $r = 1$, we use $I\!R$ instead of $I\!R^1$. For $b > a$, let $C^r[a,b]$ (resp., $C^r[0,\infty)$, $C^r(-\infty,\infty)$) denote the space of $I\!R^r$-valued continuous functions on the interval $[a,b]$ (resp., on $[0,\infty)$, $(-\infty,\infty)$). The metric used will be the sup norm if the interval is finite and the "local" sup norm if it is infinite: That is, a sequence $\{f_n(\cdot)\}$ in $C^r(-\infty,\infty)$ converges to zero if it converges to zero uniformly on each bounded time interval in the domain of definition.

Definition. Let $\{f_n(\cdot)\}$ denote a set of $I\!R^r$-valued functions on $(-\infty,\infty)$. The set is said to be *equicontinuous* in $C^r(-\infty,\infty)$ if $\{f_n(0)\}$ is bounded and for each T and $\epsilon > 0$, there is a $\delta > 0$ such that for all n

$$\sup_{0 \leq t-s \leq \delta,\ |t| \leq T} |f_n(t) - f_n(s)| \leq \epsilon. \tag{2.1}$$

There is an obviously analogous definition for the other time intervals. Note that (2.1) implies that each $f_n(\cdot)$ is continuous. The *Arzelà-Ascoli* Theorem ([35, p. 266], [155, p. 179]) states the following.

Theorem 2.1. *Let $\{f_n(\cdot)\}$ be a sequence of functions in $C^r(-\infty,\infty)$, and let the sequence be equicontinuous. Then there is a subsequence that converges to some continuous limit, uniformly on each bounded interval.*

There is a simple extension of the concept of equicontinuity to a class of noncontinuous functions that will be useful. Suppose that for each n, $f_n(\cdot)$ is an $\mathbb{R}^r$-valued measurable function on $(-\infty, \infty)$ and $\{f_n(0)\}$ is bounded. Also suppose that for each T and $\epsilon > 0$, there is a $\delta > 0$ such that

$$\limsup_n \sup_{0 \le t-s \le \delta,\ |t| \le T} |f_n(t) - f_n(s)| \le \epsilon. \tag{2.2}$$

Then we say that $\{f_n(\cdot)\}$ is *equicontinuous in the extended sense*. The functions $f_n(\cdot)$ might not be continuous, but we still have the following extension of Theorem 2.1, whose proofs are virtually the same as that of Theorem 2.1.

Theorem 2.2. *Let $\{f_n(\cdot)\}$ be defined on $(-\infty, \infty)$ and be equicontinuous in the extended sense. Then there is a subsequence that converges to some continuous limit, uniformly on each bounded interval.*

Example of equicontinuity in the extended sense. Let $\{f_n(\cdot)\}$ be equicontinuous in the original sense (2.1). Define $\bar{f}_n(\cdot)$ by $\bar{f}_n(t) = f_n(k/n)$ on the interval $[k/n, k/n+1/n)$. Then $\{\bar{f}_n(\cdot)\}$ is not continuous but $\{\bar{f}_n(\cdot)\}$ is equicontinuous in the extended sense.

All convergence theorems use some notion of compactness in one way or another. Equicontinuity is just such a notion, and the Arzelà–Ascoli Theorem played an essential role in obtaining relatively simple proofs of the convergence (with probability one) of stochastic approximation processes in [99]. The same basic idea, which uses a sequence of continuous time interpolations of the stochastic approximation iterates θ_n with interpolation intervals ϵ_n, will play a fundamental role in this book. Define the interpolation $\theta^0(\cdot)$ of the stochastic approximation process θ_n as: $\theta^0(t) = \theta_0$ for $t \le 0$, and for $t \ge 0$,

$$\theta^0(t) = \theta_n \text{ on } [t_n, t_{n+1}), \quad \text{where } t_n = \sum_{i=0}^{n-1} \epsilon_i. \tag{2.3}$$

Define the sequence of *shifted processes* $\theta^n(\cdot) = \theta^0(t_n + \cdot)$. The tail behavior of the sequence $\{\theta_n\}$ is captured by the behavior of $\theta^n(\cdot)$ for large n.

In Chapters 5 and 6 it is shown, under reasonable conditions, that for almost all sample points ω, the set of paths $\{\theta^n(\cdot, \omega)\}$ is equicontinuous in the extended sense. The extended Arzelà–Ascoli Theorem (Theorem 2.2) can then be used to extract convergent subsequences whose limits satisfy the "mean" or "average" ODE. Then the asymptotic properties of the ODE will tell us what we wish to know about the tail behavior of θ_n. This way of getting the ODE was introduced and used heavily in [99]; it is a very useful approach to the analysis of stochastic approximation algorithms. To illustrate the role of equicontinuity in getting useful limit and approximation results, we next apply the Arzelà–Ascoli Theorem to the problem of the existence of a solution to an ODE.

Example: Existence of the solution to an ODE. Given $X(0) \in I\!R$ and a continuous and bounded real-valued function $\bar{g}(\cdot)$ from $I\!R$ to $I\!R$, for $\Delta > 0$ define the sequence $\{X_n^\Delta\}$ by $X_0^\Delta = X(0)$ and

$$X_{n+1}^\Delta = X_n^\Delta + \Delta \bar{g}(X_n^\Delta), \quad n \geq 0.$$

Define the piecewise linear interpolation $X^\Delta(\cdot)$ by

$$X^\Delta(t) = \frac{(t - n\Delta)}{\Delta} X_{n+1}^\Delta + \frac{(n\Delta + \Delta - t)}{\Delta} X_n^\Delta, \quad \text{on } [n\Delta, n\Delta + \Delta). \tag{2.4}$$

Then we can write

$$X^\Delta(t) = X(0) + \int_0^t \bar{g}(X^\Delta(s))ds + \rho^\Delta(t),$$

where the interpolation error $\rho^\Delta(\cdot)$ goes to zero as $\Delta \to 0$. The sequence of functions $\{X^\Delta(\cdot)\}$ is equicontinuous. Hence, by the Arzelà–Ascoli Theorem, there is a convergent subsequence in the sense of uniform convergence on each bounded time interval, and it is easily seen that any limit $X(\cdot)$ must satisfy the ODE $\dot{X} = \bar{g}(X)$, with initial condition $X(0)$.

4.2.2 Stability of Ordinary Differential Equations

Our approach to the study of the asymptotic properties of the stochastic approximation sequence involves, either explicitly or implicitly, the asymptotic properties of an ODE that represents the "mean" dynamics of the algorithm. Thus, we need to say a little about the asymptotic behavior of ODE's. First, we reiterate the intuitive connection in a simple example. Write the stochastic approximation as $\theta_{n+1} = \theta_n + \epsilon_n Y_n$, and suppose that $\sup_n E|Y_n|^2 < \infty$. Suppose also that there is a continuous function $\bar{g}(\cdot)$ such that $\bar{g}(\theta_n) = E[Y_n | Y_i, i < n, \theta_0]$. Then

$$\theta_{n+m+1} - \theta_n = \sum_{i=n}^{m} \epsilon_i \bar{g}(\theta_i) + \sum_{i=n}^{m} \epsilon_i \left[Y_i - \bar{g}(\theta_i)\right].$$

Since the variance of the second term (the "noise" term) is of the order of $\sum_{i=n}^{m} O(\epsilon_i^2)$, we might expect that as time increases, the effects of the noise will go to zero, and the iterate sequence will eventually follow the "mean trajectory" defined by $\bar{\theta}_{n+1} = \bar{\theta}_n + \epsilon_n \bar{g}(\bar{\theta}_n)$. Suppose that this is true. Then, if we start looking at the θ_n at large n when the decreasing ϵ_n are small, the stochastic approximation algorithm behaves similarly to a finite difference equation with small step sizes. This finite difference equation is in turn approximated by the solution to the mean ODE $\dot{\theta} = \bar{g}(\theta)$.

Additionally, when using stability methods for proving the boundedness of the trajectories of the stochastic approximation, the stochastic Liapunov function to be used is close to what is used to prove the stability of the mean ODE. These ideas will be formalized in the following chapters.

Liapunov functions. Let $\bar{g}(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^r$ be a continuous function. The classical method of Liapunov stability [114] is a useful tool for determining the asymptotic properties of the solutions of the ODE $\dot{x} = \bar{g}(x)$.

Suppose that $V(\cdot)$ is a continuously differentiable and real-valued function of x such that $V(0) = 0$, $V(x) > 0$ for $x \neq 0$ and $V(x) \to \infty$ as $|x| \to \infty$. For $\lambda > 0$, define $Q_\lambda = \{x : V(x) \leq \lambda\}$. The time derivative of $V(x(\cdot))$ is given by

$$\dot{V}(x(t)) = V_x'(x(t))\bar{g}(x(t)) \equiv -k(x(t)). \tag{2.5}$$

For a given $\lambda > 0$, let $V(x(0)) \leq \lambda$, and suppose that $k(x) \geq 0$ for $x \in Q_\lambda$. The following conclusions are part of the theory of stability via the Liapunov function method:

- The inequality $\dot{V}(x(t)) = -k(x(t)) \leq 0$ for $x(t)$ in Q_λ implies that $V(x(\cdot))$ is nonincreasing along this trajectory and hence

  $$x(t) \in Q_\lambda \text{ for all } t < \infty.$$

- Similarly, since $V(\cdot)$ is non-negative,

  $$V(x(0)) \geq V(x(0)) - V(x(t)) = -\int_0^t \dot{V}(x(s))ds = \int_0^t k(x(s))ds.$$

  The last equation and the non-negativity of $k(\cdot)$ imply that $0 \leq \int_0^\infty k(x(s))ds < \infty$. Furthermore, since $k(\cdot)$ is continuous and the part of the path of the solution $x(\cdot)$ on $[0,\infty)$ that is in Q_λ is Lipschitz continuous,

  $$x(t) \to \{x \in Q_\lambda : k(x) = 0\}.$$

  If for each $\delta > 0$ there is an $\epsilon > 0$ such that $k(x) \geq \epsilon$ if $|x| \geq \delta$, then the convergence holds even if $k(\cdot)$ is only measurable.

In this book, we will not generally need to know the Liapunov functions themselves, only that they exist and have appropriate properties. For existence theorems under particular definitions of stability, see [85, 186].

Example and extension. Consider the two-dimensional system:

$$\dot{x} = Ax = \begin{bmatrix} x_2 \\ -x_1 - x_2 \end{bmatrix}. \tag{2.6}$$

If $V(x) = |x|^2 = x_1^2 + x_2^2$, then at $x(t) = x$, $\dot{V}(x) = V_x'(x)Ax = -2x_2^2$. By the Liapunov stability theory, we know that $x(t) \to \{x : x_2 = 0\}$. The given ODE is that for an undriven electrical circuit consisting of a resistor, capacitor, and inductor in a loop, where x_1 is the charge on the capacitor and x_2 is the current. The Liapunov function is proportional to the total energy. The Liapunov function argument says that the energy in the system decreases as long as the current through the resistor is nonzero, and that the current goes to zero. It does not say directly that the energy goes to zero. However, we know from the ODE for the circuit that the current cannot be zero unless the charge on the capacitor is zero. It is not hard to show by a direct analysis that $x_2(t) \to 0$ implies that $x_1(t) \to 0$ also. The proof of this follows from the fact that as long as the charge is not zero, the current cannot remain at zero. Thus the system cannot remain forever arbitrarily close to the set where $x_2 = 0$ unless $x_1(t)$ is eventually also arbitrarily close to zero.

The "double" limit problem arises since $k(x) = 0$ does not imply that $x = 0$. The analogous argument for more complicated problems would be harder, and a useful way to avoid it will now be discussed.

Definition. A set $\Lambda \in I\!R^r$ is an *invariant set* for the ODE $\dot{x} = \bar{g}(x)$ if for each $x_0 \in \Lambda$, there is a solution $x(t), -\infty < t < \infty$, that lies entirely in Λ and satisfies $x(0) = x_0$. The *limit set* for a given initial condition $x(0)$ is the set of limit points of the trajectory with initial condition $x(0)$.

If the trajectory is bounded and $\bar{g}(\cdot)$ depends only on x, then the limit set is a compact invariant set [62]. Recall that $x(t)$ does not necessarily converge to a unique point. For example, for the ODE $\ddot{u} + u = 0$ where u is real-valued, the limit set is a circle.

Definition. A point x_0 is said to be *asymptotically stable in the sense of Liapunov* for an ODE if any solution $x(t)$ tends to x_0 as $t \to \infty$, and for each $\delta > 0$ there is an $\epsilon > 0$ such that if $|x(0) - x_0| \le \epsilon$, then $|x(t) - x_0| \le \delta$ for all t. The point x_0 is said to be *locally asymptotically stable in the sense of Liapunov* if the preceding definition holds when the initial condition is in some neighborhood of x_0.

Limit sets and the invariant set theorem. The Liapunov function method for proving stability and exhibiting the limit points works best when $k(x) = 0$ only at isolated points. When this is not the case, the following result, known as *LaSalle's Invariance Theorem* (see [114]), which improves on the assertions available directly from the Liapunov function analysis, often helps to characterize the limit set.

Theorem 2.3. *For given $\lambda > 0$, assume the conditions on the Liapunov function preceding* (2.5) *in the set Q_λ, and let $V(x(0)) \le \lambda$. Then, as*

$t \to \infty$, $x(t)$ converges to the largest invariant set contained in the set $\{x : k(x) = 0, V(x) \le V(x(0))\}$.

Thus, in accordance with LaSalle's Invariance Theorem, we need to look for the largest bounded set B on which $k(x) = 0$ such that for each $x \in B$, there is an entire trajectory on the doubly infinite time interval $(-\infty, \infty)$ that lies all in B and goes through x. To apply this result to the example (2.6), note that there is no bounded trajectory of the ODE that satisfies $x_2(t) = 0$ for all $t \in (-\infty, \infty)$, unless $x(t) \equiv 0$.

Suppose that there is a continuously differentiable real-valued function $f(\cdot)$, bounded below with $f(x) \to \infty$ as $|x| \to \infty$ and such that $\dot{x} = \bar{g}(x) = -f_x(x)$. Then we can say more about the limit sets; namely, that set of stationary points $\{x : \bar{g}(x) = 0\}$ is a collection of disjoint compact sets and the limit trajectory must be contained in one of these sets. For the proof, use $f(\cdot)$ as a Liapunov function so that $k(x) = -|f_x(x)|^2$.

4.3 Projected ODE

In applications to stochastic approximation, it is often the case that the iterates are constrained to lie in some compact set H in the sense that if an iterate ever leaves H, it is immediately returned to the closest (or some other convenient) point in H. A common procedure simply truncates, as noted in the examples in Chapters 1 to 3. In other applications, there are physical constraints that the parameter θ_n must satisfy. Owing to the pervasive practical use of constraints, much of the development in this book will concern algorithms in which the state is constrained to a compact set in some way. The simplest constraint is just a truncation or projection of each component separately. This is condition (A3.1). (A3.2) defines a more general constraint set, and (A3.3) defines a constraint set that is a submanifold of $\mathbb{R}^r$.

(A3.1) H is a hyperrectangle. In other words, there are real numbers $a_i < b_i, i = 1, \ldots, r$, such that $H = \{x : a_i \le x_i \le b_i\}$.

For $x \in H$ satisfying (A3.1), define the set $C(x)$ as follows. For $x \in H^0$, the interior of H, $C(x)$ contains only the zero element; for $x \in \partial H$, the boundary of H, let $C(x)$ be the infinite convex cone generated by the outer normals at x of the faces on which x lies.

Then the *projected* ODE (or ODE whose dynamics are projected onto H) is defined to be

$$\dot{x} = \bar{g}(x) + z, \quad z(t) \in -C(x(t)), \tag{3.1}$$

where $z(\cdot)$ is the *projection* or *constraint term*, the minimum term needed to keep $x(\cdot)$ in H. Let us examine (3.1) more closely. If $x(t)$ is in H^0 on some time interval, then $z(\cdot)$ is zero on that interval. If $x(t)$ is on the interior of a face of H (i.e., $x_i(t)$ equals either a_i or b_i for a unique i) and $\bar{g}(x(t))$ points "out" of H, then $z(\cdot)$ points inward, orthogonal to the face. If $x(t)$ is on an edge or corner of H, with $\bar{g}(x(t))$ pointing "out" of H, then $z(t)$ points inward and takes values in the convex cone generated by the inward normals on the faces impinging on the edge or corner; that is, $z(t)$ takes values in $-C(x(t))$ in all cases. In general, $z(t)$ is the smallest value needed to keep $x(\cdot)$ from leaving H. For example, let $x_i(t) = b_i$, with $\bar{g}_i(x(t)) > 0$. Then, $z_i(t) = -\bar{g}_i(x(t))$.

The function $z(\cdot)$ is not unique in that it is defined only for almost all t. Apart from this it is determined by H and $\bar{g}(\cdot)$. If $\dot{x} = \bar{g}(x)$ has a unique solution for each $x(0)$, then so does (3.1).

Figure 3.1 illustrates the unconstrained and constrained flow lines. In applications, the actual constraint is often flexible, and one tries to use constraints that do not introduce unwanted limit points.

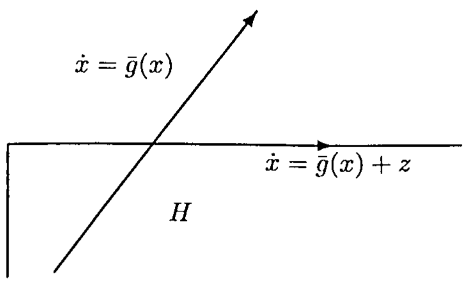

Figure 3.1. Constrained and unconstrained flow lines.

More general constraint sets. In the preceding discussion, we let the constraint set H be a hyperrectangle in order to simplify the discussion that introduced the projected ODE. In applications, one might have additional hard constraints on the state. The following two cases will be of particular importance, the first being an extension of (A3.1).

(A3.2) Let $q_i(\cdot), i = 1, \cdots, p$, be continuously differentiable real-valued functions on $\mathbb{R}^r$, with gradients $q_{i,x}(\cdot)$. Without loss of generality, let $q_{i,x}(x) \neq 0$ if $q_i(x) = 0$.

Define $H = \{x : q_i(x) \leq 0, i = 1, \cdots, p\}$ and assume that it is connected, compact and nonempty. A constraint $q_i(\cdot)$ is said to be *active* at x if $q_i(x) = 0$. Define $A(x)$, the set of (indices of the) *active constraints* at x, by $A(x) = \{i : q_i(x) = 0\}$. Define $C(x)$ to be the convex cone generated by the set of outward normals $\{y : y = q_{i,x}(x), i \in A(x)\}$. Suppose that for each x with nonempty $A(x)$, the set $\{q_{i,x}(x), i \in A(x)\}$ is linearly independent. If there are no active constraints at x, then $C(x)$ contains only the zero element.

(A3.3) H is an $\mathbb{R}^{r-1}$ dimensional connected compact surface with a continuously differentiable outer normal. In this case, define $C(x), x \in H$, to be the linear span of the outer normal at x.

If one cares to define the associated stochastic approximation algorithm, any of the usual types of constraints used in the theory of nonlinear programming can be handled, including (A3.2) applied to the manifold in (A3.3).

Note. If under (A3.2), there is only one active constraint (say, that indexed by i) at t, and $\bar{g}(x(t))$ points out of H, then the right-hand side of (3.1) is just the projection of $\bar{g}(x(t))$ onto the tangent plane to the surface at $x(t)$; that is, onto the plane which is orthogonal to the gradient $q_{i,x}(x(t))$ at $x(t)$. Let $\bar{g}(x) = -f_x(x)$, for a continuously differentiable real-valued $f(\cdot)$. Then the constraints in (A3.1) or (A3.2) can give rise to spurious stationary points on the boundary, but this is the only type of singular point which can be introduced by the constraint. The function $f(\cdot)$ can still be used as a Liapunov function, since the derivative along the trajectory at a point x is $f_x'(x)[-f_x(x) + z] \leq 0$.

Note on the invariant set theorem for the projected ODE. The result on invariant set Theorem 2.3 holds for the constrained process (3.1), but is less useful because the entire set H might be an invariant set if H is bounded. For this reason, when working with the constrained algorithm, we simply use the limit points of (3.1) rather than the largest bounded invariant set. Let L_H denote the set of limit points of (3.1), over all initial conditions in H. Let S_H denote the set of points in H where $\bar{g}(x) + z = 0$. These are called the *stationary points*. Interior to H, the stationarity condition is just $\bar{g}(x) = 0$. On the boundary, the condition is $\bar{g}(x) \in C(x)$.

Upper semicontinuity of $C(x)$. An "infinitesimal" change in x cannot increase the number of active constraints. Thus, loosely speaking, this property can be stated as $\lim_{y \to x} C(y) \subset C(x)$. More precisely, let $N_\delta(x)$ be a δ-neighborhood of x. Then

$$\bigcap_{\delta>0} \bigcup_{y \in N_\delta(x)} C(y) = C(x). \tag{3.2}$$

A set-valued function $C(\cdot)$ satisfying (3.2) is said to be *upper semi-continuous*.

In the analysis of the stochastic approximation algorithms, (3.1) appears in the integral form:

$$\begin{aligned} x(t+\tau) &= x(t) + \int_t^{t+\tau} \bar{g}(x(s))ds + Z(t+\tau) - Z(t), \\ x(t) &\in H, \quad \text{for all } t, \ Z(0) = 0, \end{aligned} \tag{3.3}$$

where $x(\cdot)$ is Lipschitz continuous and the following conditions hold for $\tau > 0$:

$$Z(t+\tau) - Z(t) = 0 \text{ if } x(s) \in H^0 \text{ for almost all } s \in [t, t+\tau], \tag{3.4}$$

$$Z(t+\tau) - Z(t) \in -\text{co}\left[\bigcup_{t \le s \le t+\tau} C(x(s))\right], \tag{3.5}$$

where $\text{co}(A)$ denotes the closed convex hull of the set A. The following theorem relating the forms (3.3) and (3.1) will be needed.

Theorem 3.1. *Assume one of the constraint set conditions* (A3.1), (A3.2), *or* (A3.3). *Let $\bar{g}(\cdot)$ be bounded on H and let* (3.3)–(3.5) *hold, where $x(\cdot)$ is Lipschitz continuous. Then $Z(\cdot)$ is absolutely continuous. There is a measurable function $z(\cdot)$ such that $z(t) \in -C(x(t))$ for almost all t and*

$$Z(t+\tau) - Z(t) = \int_t^{t+\tau} z(s)ds. \tag{3.6}$$

4.4 Stochastic Stability and Perturbed Stochastic Liapunov Functions

The following theorem is a direct application of the martingale probability inequalities and the martingale convergence theorem of Section 1, and it is a natural analog of the Liapunov function theorem for ODEs. It is one of the original theorems in stochastic stability theory, first proved in [25, 78, 87, 88]. The basic idea will be adapted in various ways to prove convergence or to get bounds on the trajectories of the stochastic approximation sequences.

Theorem 4.1. *Let $\{X_n\}$ be a Markov chain on $\mathbb{R}^r$. Let $V(\cdot)$ be a real-valued and non-negative function on $\mathbb{R}^r$ and for given $\lambda > 0$ define the set $Q_\lambda = \{x : V(x) \le \lambda\}$. Suppose that for all $x \in Q_\lambda$*

$$E\left[V(X_{n+1}) | X_n = x\right] - V(x) \le -k(x) \tag{4.1}$$

for all n, where $k(x) \ge 0$ and is continuous on Q_λ. Then, with probability one, for each integer ν

$$P\left\{\sup_{\nu \le m < \infty} V(X_m) \ge \lambda \big| X_\nu\right\} I_{\{X_\nu \in Q_\lambda\}} \le \frac{V(X_\nu)}{\lambda}. \tag{4.2}$$

Equivalently, if $X_\nu \in Q_\lambda$ then the path on $[\nu, \infty)$ stays in Q_λ with a probability at least $1 - V(X_\nu)/\lambda$. Let Ω_λ denote the set of paths that stay in Q_λ from some time ν on. Then for almost all $\omega \in \Omega_\lambda$, $V(X_m)$ converges and

$X_m \to \{x : k(x) = 0\}$; that is, for each $\omega \in \Omega_\lambda - N$ where N is a null set, the path converges to a subset of $\{x : k(x) = 0\}$, which is consistent with $V(x)$ =constant.

Outline of proof. (See [1, 78, 88] for more detail.) The argument is a stochastic analog of what was done in the deterministic case. For notational simplicity, we suppose that $\nu = 0$ and $X_0 = x \in Q_\lambda$. Define $\tau_{Q_\lambda} = \min\{n : X_n \notin Q_\lambda\}$ (or infinity, if $X_n \in Q_\lambda$ for all n). Since we have not assumed that $k(x)$ is non-negative for $x \notin Q_\lambda$, it is convenient to work with the stopped process $\tilde{X}_n$ defined by $\tilde{X}_n = X_n$ for $n < \tau_{Q_\lambda}$, and $\tilde{X}_n = X_{\tau_{Q_\lambda}}$ for $n \geq \tau_{Q_\lambda}$. Thus, from time τ_{Q_λ} on, the value is fixed at $X_{\tau_{Q_\lambda}}$, which is the value attained on first leaving Q_λ. Define $\tilde{k}(x)$ to equal $k(x)$ in Q_λ and to equal zero for $x \notin Q_\lambda$. Now,

$$E\left[V(\tilde{X}_{n+1})|\tilde{X}_n\right] - V(\tilde{X}_n) \leq -\tilde{k}(\tilde{X}_n) \tag{4.3}$$

for all n. Thus, $\{V(\tilde{X}_n)\}$ is a non-negative supermartingale and (4.2) is implied by (1.6). By the supermartingale convergence theorem, $\{V(\tilde{X}_n)\}$ converges to some random variable $\tilde{V} \geq 0$. Iterating (4.3) and using the non-negativity of $V(\cdot)$, for all x we have

$$V(x) \geq V(x) - E_x V(\tilde{X}_n) \geq E_x \sum_{m=0}^{n-1} \tilde{k}(\tilde{X}_m), \tag{4.4}$$

where E_x denotes the expectation given that $X_0 = x$. In view of (4.3), $E_x \sum_{m=0}^{\infty} \tilde{k}(\tilde{X}_m) < \infty$. Thus, by the Borel–Cantelli Lemma, with probability one for any $\epsilon > 0$, $\tilde{X}_n$ can spend only a finite amount of time more than a distance of ϵ from $\{x : \tilde{k}(x) = 0\}$. □

Sometimes the right side of (4.1) is replaced by $-k(x) +$ "small term," and the "small term" can be used to guide the construction of a "perturbed" Liapunov function for which the conditional difference is nonpositive. The proof of Theorem 4.1 yields the following result.

Theorem 4.2. *Let $\{X_n\}$ be a Markov chain and $V(\cdot)$ a real-valued and non-negative function on $\mathbb{R}^r$. Let $\{\mathcal{F}_n\}$ be a sequence of σ-algebras, which is nondecreasing and where $\mathcal{F}_n$ measures at least $\{X_i, i \leq n\}$. Let δV_n be $\mathcal{F}_n$-measurable random variables such that $\delta V_n \to 0$ with probability one and $E|\delta V_n| < \infty$ for each n. Define $V_n(x) = V(x) + \delta V_n$. Suppose that*

$$E\left[V_{n+1}(X_{n+1}) - V_n(X_n)|\mathcal{F}_n\right] \leq -k(X_n) \leq 0,$$

where $k(\cdot)$ is continuous and positive for $x \neq 0$. Then $X_n \to 0$ with probability one.

The following extension of Theorem 4.2 will be useful in the stochastic approximation problem. In the theorem, $\{X_n\}$ is an $\mathbb{R}^r$-valued stochastic process, not necessarily a Markov process. Let $\{\mathcal{F}_n\}$ be a sequence of non-decreasing σ-algebras, with $\mathcal{F}_n$ measuring at least $\{X_i, i \le n\}$, and let E_n denote the expectation conditioned on $\mathcal{F}_n$. If $X_n = \theta_n$, then the form (4.5) arises from the truncated Taylor expansion $E[V(\theta_{n+1})|\theta_0, Y_i, i < n] - V(\theta_n)$ in stochastic approximation problems.

Theorem 4.3. *Let $V(\cdot)$ be a non-negative real-valued continuous function on $I\!R^r$ that is positive for $x \ge 0$, $V(0) = 0$, and with the property that for each $\epsilon > 0$ there is a $\delta > 0$ such that $V(x) \ge \delta$ for $|x| \ge \epsilon$. Suppose that $EV(X_0) < \infty$ and δ does not decrease as ϵ increases. Let there be random variables Y_n and a non-negative function $k(\cdot)$ such that for each $\epsilon > 0$ there is a $\delta > 0$ satisfying $k(x) \ge \delta$ for $|x| \ge \epsilon$. Let there be a $K_1 < \infty$ such that*

$$E_n V(X_{n+1}) - V(X_n) \le -\epsilon_n k(X_n) + K_1 \epsilon_n^2 E_n |Y_n|^2, \tag{4.5}$$

where ϵ_n are positive $\mathcal{F}_n$-measurable random variables which tend to zero with probability one and $\sum_n \epsilon_n = \infty$ with probability one. Suppose that $Ek(X_n) < \infty$ if $EV(X_n) < \infty$ and that there are $K_2 < \infty$ and $K < \infty$ such that

$$E_n |Y_n|^2 \le K_2 k(X_n), \quad \text{when } |X_n| \ge K. \tag{4.6}$$

Let

$$E \sum_{i=0}^{\infty} \epsilon_i^2 |Y_i|^2 I_{\{|X_i| \le K\}} < \infty. \tag{4.7}$$

Then $X_n \to 0$ with probability one.

Proof. The hypotheses imply that $EV(X_n) < \infty$ for all n; we leave the proof of this fact to the reader. Since $\epsilon_n \to 0$ with probability one, $\epsilon_n^2 \ll \epsilon_n$ for large n. Now, since $EV(X_n) < \infty$ for all n, by shifting the time origin we can suppose without loss of generality that $\epsilon_n^2 \ll \epsilon_n$ for all n. In particular, we suppose that $K_1 K_2 \epsilon_n^2 < \epsilon_n/2$. Define

$$\delta V_n = K_1 E_n \sum_{i=n}^{\infty} \epsilon_i^2 |Y_i|^2 I_{\{|X_i| \le K\}},$$

and the perturbed Liapunov function $V_n(X_n) = V(X_n) + \delta V_n$. Note that $V_n(X_n) \ge 0$ and

$$E_n \delta V_{n+1} - \delta V_n = -K_1 \epsilon_n^2 E_n |Y_n|^2 I_{\{|X_n| \le K\}}.$$

This, together with (4.5) and (4.6), yields

$$E_n V_{n+1}(X_{n+1}) - V_n(X_n) \le -\epsilon_n k(X_n)/2, \tag{4.8}$$

which implies that $\{V_n(X_n)\}$ is an $\mathcal{F}_n$-supermartingale sequence. By the supermartingale convergence theorem, there is a $\tilde{V} \ge 0$ such that $V_n(X_n) \to$

$\tilde{V}$ with probability one. Since (4.7) implies that $\delta V_n \to 0$ with probability one, $V(X_n) \to \tilde{V}$ with probability one.

For integers N and m, (4.8) yields

$$E_N V_{N+m}(X_{N+m}) - V_N(X_N) \leq - \sum_{i=N}^{N+m-1} E_N \epsilon_i k(X_i)/2. \tag{4.9}$$

As $m \to \infty$, the left side of (4.9) is bounded below by $-V_N(X_N)$. Suppose that $\tilde{V} > 0$ with positive probability. Then, by the properties of $V(\cdot)$, X_N is asymptotically outside some small neighborhood of the origin, with a positive probability. This and the fact that $\sum \epsilon_i = \infty$ with probability one and the properties of $k(\cdot)$ imply that the sum on the right side of (4.9) goes to infinity with a positive probability, leading to a contradiction. Thus $\tilde{V} = 0$ with probability one. □

These theorems are concerned with convergence with probability one. The Liapunov function method is used simply to prove recurrence. With the recurrence given, other methods might be used to prove convergence. Then the following special case of Theorem 4.3 is useful.

Theorem 4.4. *Let X_n, $\mathcal{F}_n$, and ϵ_n be as in Theorem 4.3. Let $V(x) \geq 0$ and suppose that there are $\delta > 0$ and compact $A \subset \mathbb{R}^r$ such that for all n*

$$E_n V(X_{n+1}) - V(X_n) \leq -\delta < 0, \quad \text{for } x \notin A.$$

Then the set A is recurrent for $\{X_n\}$ in that $X_n \in A$ for infinitely many n with probability one.

5

Convergence with Probability One: Martingale Difference Noise

5.0 Outline of Chapter

Much of the classical work in stochastic approximation dealt with the situation where the "noise" in each observation Y_n is a martingale difference, that is, where there is a function $g_n(\cdot)$ of θ such that $E[Y_n|Y_i, i < n, \theta_0] = g_n(\theta_n)$ [17, 40, 45, 47, 56, 79, 86, 132, 154, 159, 169, 181]. Then we can write $Y_n = g_n(\theta_n) + \delta M_n$, where δM_n is a martingale difference. This "martingale difference noise" model is still of considerable importance. It arises, for example, where Y_n has the form $Y_n = F_n(\theta_n, \psi_n)$ where ψ_n are mutually independent. The convergence theory is relatively easy in this case, because the noise terms can be dealt with by well-known and relatively simple probability inequalities for martingale sequences. This chapter is devoted to this martingale difference noise case. Nevertheless, the ODE, compactness, and stability techniques to be introduced are of basic importance for stochastic approximation, and will be used in subsequent chapters.

A number of definitions that will be used throughout the book are introduced in Section 1. In particular, the general "ODE" techniques used in the rest of the book are based on the analysis of continuous time interpolations of the stochastic approximation sequence. These interpolations are defined in Section 1. The general development in the book follows intuitively reasonable paths but cannot be readily understood unless the definitions of the interpolated processes are understood.

Section 2 gives a fundamental convergence theorem and shows how the stochastic approximation sequence is related to a "mean limit" ODE that characterizes the asymptotic behavior. The Arzelà–Ascoli theorem is crucial to getting the ODE since it guarantees that there will always be convergent subsequences of the set of interpolated processes. The limits of any of these subsequences will satisfy the "mean limit" ODE. The first theorem (Theorem 2.1) uses a simple constraint set to get a relatively simple proof and allows us to concentrate on the essential structure of the "ODE method"-type proofs. This constraint set is generalized in Theorem 2.3, where a method for characterizing the reflection terms is developed, which will be used throughout the book. All the results carry over to the case where the constraint set is a smooth manifold of any dimension.

The conditions used for the theorems in Section 2 are more or less classical. For example, square summability of the step sizes ϵ_n is assumed. The square summability, together with the martingale noise property and a stability argument, can be used to get a simpler proof if the algorithm is unconstrained. However, the type of proof given readily generalizes to one under much weaker conditions. The set to which the iterates converge is a limit or invariant set for the mean limit ODE. These limit or invariant sets might be too large in that the convergence can only be to a subset. Theorem 2.5 shows that the only points in the limit or invariant set that we need to consider are the "chain recurrent" points, an idea due to Benaim [6].

The conditions are weakened in Subsection 3.1, which presents the "final form" of the martingale difference noise case in terms of conditions that require the "asymptotic rates of change" of certain random sequences to be zero with probability one. These conditions are satisfied by the classical case of Section 2. They are phrased somewhat abstractly but are shown to hold under rather weak and easily verifiable conditions in Subsection 3.2. Indeed, these "growth rate" conditions seem to be nearly minimal for convergence, and they hold even for "very slowly" decreasing step sizes. The conditions have been proved to be necessary in certain cases. The essential techniques of this chapter originated in [99].

A stability method for getting convergence, when there are no *a priori* bounds on the iterates, is in Section 4. A stochastic Liapunov function method is used to prove recurrence of the iterates, and then the ODE method takes over in the final stage of the proof. This gives a more general result than one might obtain with a stability method alone and is more easily generalizable. Section 5 concerns "soft" constraints, where bounds on functionals of the iterate are introduced into the algorithm via a penalty function. The results in Section 6 on the random directions Kiefer–Wolfowitz method and on the minimization of convex functions are suggestive of additional applications. Section 7 gives the proof of convergence for the "lizard learning" problem of Section 2.1 and the pattern classification problem of Section 1.1. When using stochastic approximation for function minimization, where the function has more than one local minimum, one would like to assure at least that convergence to other types of stationary points (such as local maxima or saddles) is impossible. One expects that the noise in the algorithm will destabilize the algorithm around these "undesirable" points. This is shown to be the case for a slightly perturbed algorithm in Section 8.

5.1 Truncated Algorithms: Introduction

To develop the basic concepts behind the convergence theory in a reasonably intuitive way, we will first work with a relatively simple form and then systematically generalize it.

An important issue in applications of stochastic approximation concerns the procedure to follow if the iterates become too large. Practical algorithms tend to deal with this problem via appropriate adjustments to the basic algorithm, but these are often ignored in the mathematical developments, which tend to allow unbounded iterate sequences and put various "stability" conditions on the problem. However, even if these stability conditions do hold in practice, samples of the iterate sequence might get large enough to cause concern. The appropriate procedures to follow when the parameter value becomes large is, of course, dependent on the particular problem and the form of the algorithm that has been chosen, and it is unfortunate that there are no perfectly general rules to which one can appeal. Nevertheless, the useful parameter values in properly parameterized practical problems are usually confined by constraints of physics or economics to some compact set. This might be given by hard physical constraint that requires that, say, a dosage be less than a certain number of milligrams or a temperature set point in a computer simulation of a chemical process be less than 200°C. There are also implicit bounds in most problems. If θ_n is the set point temperature in a chemical processor and it reaches the temperature at the interior of the sun, or if the cost of setting the parameter at θ_n reaches the U.S. gross national product, then something is very likely wrong with the model or with the algorithm or with both. The models used in simulations are often inaccurate representations of physical reality at excessive values of the parameter (or of the noise), and so a mathematical development that does not carefully account for the changes in the model as the parameter (and the noise) values go to infinity might well be assuming much more than is justified. The possibility of excessive values of θ_n is a problem unique to computer simulations, because any algorithm that is used on a physical process would be carefully controlled.

Excessively large values of θ_n might simply be a consequence of poor choices for the algorithm structure. For example, instability can be caused by values of ϵ_n that are too large or values of finite difference intervals that are too small. The path must be checked for undesirable behavior, whether or not there are hard constraints. If the algorithm appears to be unstable, then one could reduce the step size and restart at an appropriate point or even reduce the size of the constraint set. The path behavior might suggest a better algorithm or a better way of estimating derivatives. Conversely, if the path moves too slowly, we might wish to increase the step sizes. If the problem is based on a simulation, one might need to use a cruder model, with perhaps fewer parameters and a more restricted constraint set, to get a rough estimate of the location of the important values of the parameters. Even hard constraints are often somewhat "flexible," in that they might be intended as rough guides of the bounds, so that if the iterate sequence "hugs" a bounding surface, one might try to slowly increase the bounds, or perhaps to test the behavior via another simulation. In practice, there is generally an upper bound, beyond which the user will not allow the iterate sequence to go. At this point, either the iterate will be truncated in some way by the rules of the algorithm or there will be external intervention.

Much of the book is concerned with projected or truncated algorithms, where the iterate θ_n is confined to some bounded set, because this is a common practice in applications. Allowing unboundedness can lead to needless mathematical complications because some sort of stability must be shown or otherwise assumed, with perhaps artificial assumptions introduced on the behavior at large parameter values, and it generally adds little to the understanding of practical algorithms.

Many practical variations of the constraints can be used if the user believes they will speed convergence. For example, if the iterate leaves the constraint set, then the projection need not be done immediately. One can wait several iterates. Also, larger step sizes can be used near the boundary, if desired.

Throughout the book, the step size sequence will satisfy the fundamental condition

$$\sum_{n=0}^{\infty} \epsilon_n = \infty, \ \epsilon_n \geq 0, \ \epsilon_n \to 0, \quad \text{for } n \geq 0; \quad \epsilon_n = 0, \ \text{for } n < 0. \tag{1.1}$$

When *random* ϵ_n are used, it will always be supposed that (1.1) holds *with probability one*. Let $Y_n = (Y_{n,1}, \ldots, Y_{n,r})$ denote the $I\!R^r$-valued "observation" at time n, with the real-valued components $Y_{n,i}$.

Many of the proofs are based on the ideas in [99]. To facilitate understanding of these ideas, in Section 2 we start with conditions that are stronger than needed, and weaken them subsequently. The basic interpolations and time scalings will also be used in the subsequent chapters. In Theorem 2.1, we let the ith component of the state θ_n be confined to the interval $[a_i, b_i]$, where $-\infty < a_i < b_i < \infty$. Then the algorithm is

$$\theta_{n+1,i} = \Pi_{[a_i,b_i]} \left[\theta_{n,i} + \epsilon_n Y_{n,i}\right], \quad i = 1, \ldots, r. \tag{1.2}$$

We will write this in vector notation as

$$\theta_{n+1} = \Pi_H \left[\theta_n + \epsilon_n Y_n\right], \tag{1.3}$$

where Π_H is the projection onto the constraint set $H = \{\theta : a_i \le \theta^i \le b_i\}$. Define the *projection* or "correction" term Z_n by writing (1.3) as

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n + \epsilon_n Z_n. \tag{1.4}$$

Thus $\epsilon_n Z_n = \theta_{n+1} - \theta_n - \epsilon_n Y_n$; it is the vector of shortest Euclidean length needed to take $\theta_n + \epsilon_n Y_n$ back to the constraint set H if it is not in H.

To get a geometric feeling for the Z_n terms, refer to Figures 1.1 and 1.2. In situations such as Figure 1.1, where only one component is being truncated, Z_n points inward and is orthogonal to the boundary at θ_{n+1}. If more than one component needs to be truncated, as in Figure 1.2, Z_n again points inward but toward the corner, and it is proportional to a convex combination of the inward normals at the faces that border on that corner. In both cases, $Z_n \in -C(\theta_{n+1})$, where the cone $C(\theta)$ determined by the outer normals to the active constraint at θ was defined in Section 4.3.

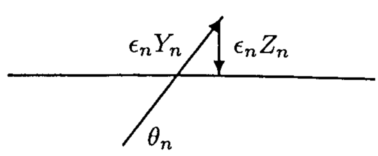

Figure 1.1. A projection with one violated constraint.

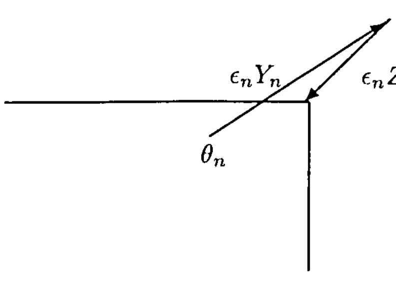

Figure 1.2. A projection with two violated constraints.

Martingale difference noise. In this chapter we will suppose that there are measurable functions $g_n(\cdot)$ of θ and random variables β_n such that Y_n can be decomposed as

$$Y_n = g_n(\theta_n) + \delta M_n + \beta_n, \quad \delta M_n = Y_n - E\left[Y_n | \theta_0, Y_i, i < n\right]. \tag{1.5}$$

The sequence $\{\beta_n\}$ will be "asymptotically negligible" in a sense to be defined. The sequence $\{\delta M_n\}$ is a martingale difference (with respect to the sequence of σ-algebras $\mathcal{F}_n$ generated by $\{\theta_0, Y_i, i < n\}$). The martingale difference assumption was used in the earliest work in stochastic approximation [17, 36, 40, 42, 45, 79, 132, 154]. Our proofs exploit the powerful ideas of the ODE methods stemming from the work of Ljung [119, 120] and Kushner [93, 99, 102]. In many of the applications of the Robbins–Monro or Kiefer–Wolfowitz algorithms, Y_n has the form $Y_n = F_n(\theta_n, \psi_n) + \beta_n$ where $\{\psi_n\}$ is a sequence of mutually independent random variables, $\{F_n(\cdot)\}$ is a sequence of measurable functions, $\beta_n \to 0$ and $E[F_n(\theta_n, \psi_n)|\theta_n = \theta] = g_n(\theta)$. For the Kiefer–Wolfowitz algorithm (see (1.3.1)–(1.3.4)), β_n represents the finite difference bias. The function $g_n(\cdot)$ might or might not depend on n. In the classical works on stochastic approximation, there was no n-dependence. The n-dependence occurs when the successive iterations are on different components of θ, the experimental procedure varies with n, or variance reduction methods are used, and so on. In the introductory result (Theorem 2.1), it will be supposed that $g_n(\cdot)$ is independent of n to simplify the development.

Definitions: Interpolated time scale and processes. The definitions and interpolations introduced in this section will be used heavily throughout the book. They are basic to the ODE method, and facilitate the effective exploitation of the time scale differences between the iterate process and the driving noise process. The ODE method uses a continuous time interpolation of the $\{\theta_n\}$ sequence. A natural time scale for the interpolation is defined in terms of the step size sequence. Define $t_0 = 0$ and $t_n = \sum_{i=0}^{n-1} \epsilon_i$. For $t \geq 0$, let $m(t)$ denote the unique value of n such that $t_n \leq t < t_{n+1}$. For $t < 0$, set $m(t) = 0$. Define the *continuous time interpolation* $\theta^0(\cdot)$ on $(-\infty, \infty)$ by $\theta^0(t) = \theta_0$ for $t \leq 0$, and for $t \geq 0$,

$$\theta^0(t) = \theta_n, \qquad \text{for } t_n \leq t < t_{n+1}. \tag{1.6}$$

For later use, define the sequence of *shifted* processes $\theta^n(\cdot)$ by

$$\theta^n(t) = \theta^0(t_n + t), \quad t \in (-\infty, \infty). \tag{1.7}$$

Figures 1.3 and 1.4 illustrate the functions $m(\cdot)$, $m(t_n + \cdot)$, and interpolations $\theta^0(\cdot)$, and $\theta^n(\cdot)$.

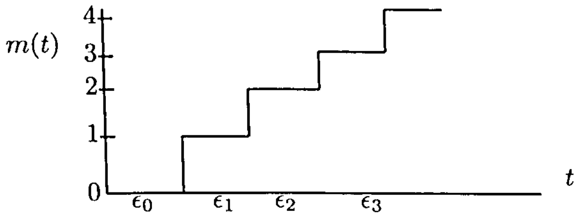

Figure 1.3a. The function $m(\cdot)$.

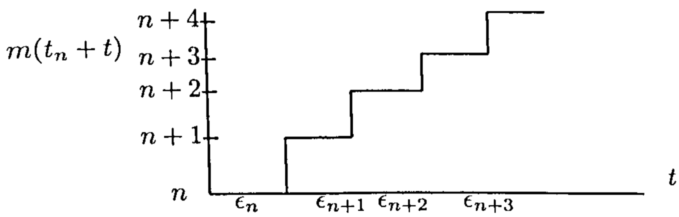

Figure 1.3b. The function $m(t_n + t)$.

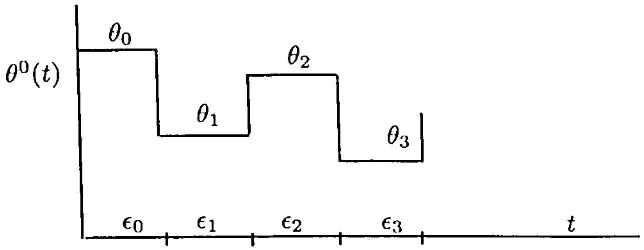

Figure 1.4a. The function $\theta^0(\cdot)$.

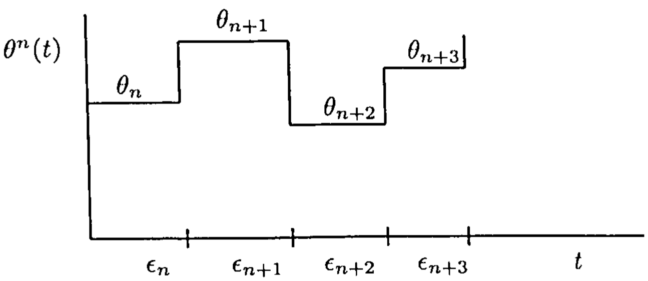

Figure 1.4b. The function $\theta^n(\cdot)$.

Let $Z_i = 0$ and $Y_i = 0$ for $i < 0$. Define $Z^0(t) = 0$ for $t \leq 0$ and

$$Z^0(t) = \sum_{i=0}^{m(t)-1} \epsilon_i Z_i, \quad t \geq 0.$$

Define $Z^n(\cdot)$ by

$$\begin{aligned} Z^n(t) &= Z^0(t_n + t) - Z^0(t_n) = \sum_{i=n}^{m(t_n+t)-1} \epsilon_i Z_i, \quad t \geq 0, \\ Z^n(t) &= -\sum_{i=m(t_n+t)}^{n-1} \epsilon_i Z_i, \quad t < 0. \end{aligned} \tag{1.8}$$

Define $Y^n(\cdot), M^n(\cdot)$, and $B^n(\cdot)$ analogously to $Z^n(\cdot)$ but using Y_i, δM_i, and β_i, resp., in lieu of Z_i. By the definitions (recall that $m(t_n) = n$)

$$\theta^n(t) = \theta_n + \sum_{i=n}^{m(t_n+t)-1} \epsilon_i \left[Y_i + Z_i\right] = \theta_n + Y^n(t) + Z^n(t), \quad t \geq 0, \tag{1.9a}$$

$$\theta^n(t) = \theta_n - \sum_{i=m(t_n+t)}^{n-1} \epsilon_i \left[Y_i + Z_i\right] = \theta_n + Y^n(t) + Z^n(t), \quad t < 0. \tag{1.9b}$$

For simplicity, we always write the algorithm as (1.9a), *whether t is positive or negative, with the understanding that it is to be interpreted as* (1.9b) *if* $t < 0$. All the above interpolation formulas will be used heavily in the sequel.

Note that the time origin of the "shifted" processes $\theta^n(\cdot)$ and $Z^n(\cdot)$ is time t_n for the original processes, the interpolated time at the nth iteration. The step sizes ϵ_n used in the interpolation are natural intervals for the continuous time interpolation. Their use allows us to exploit the time scale differences between the mean terms and the noise terms under quite general conditions. We are concerned with the behavior of the tail of the sequence $\{\theta_n\}$. Since this is equivalent to the behavior of $\theta^n(\cdot)$ over any finite interval for large n, a very effective method (introduced in [99]) of dealing with the tails works with these shifted processes $\theta^n(\cdot)$.

Note on piecewise linear vs. piecewise constant interpolations. The basic stochastic approximation (1.3) is defined as a discrete time process. We have defined the continuous time interpolations $\theta^n(\cdot), Z^n(\cdot)$ to be piecewise constant with interpolation intervals ϵ_n. We could have defined the interpolations to be piecewise linear in the obvious way by simply interpolating linearly between the "break" or "jump" points $\{t_n\}$. Nevertheless, there are some notational advantages to the piecewise constant interpolation. In the proofs in this chapter and Chapter 6, it is shown that for almost all sample paths the set $\{\theta^n(\omega, \cdot)\}$ is equi- (actually Lipschitz) continuous in the extended sense (see Theorem 4.2.2). Thus, the set of piecewise linear interpolations is equicontinuous.

5.2 The ODE Method: A Basic Convergence Theorem

5.2.1 *Assumptions and the Main Convergence Theorem*

One way or another, all methods of analysis need to show that the "tail" effects of the noise vanish. This "tail" behavior is essentially due to the martingale difference property and the fact that the step sizes ϵ_n decrease to zero as $n \to \infty$.

Definition. Recall that L_H denotes the set of limit points of the mean limit ODE (2.1) in H, over all initial conditions where z is the minimum force needed to keep the solution in H. By invariant set, we always mean a two-sided invariant set; that is, if $x \in I$, an invariant set in H, then there is a path of the ODE in H on the time interval $(-\infty, \infty)$ that goes through x at time 0. If there is a constraint set, then the set of limit points might be smaller than the largest two-sided invariant set.

Definitions. Let E_n denote the expectation conditioned on the σ-algebra $\mathcal{F}_n$, generated by $\{\theta_0, Y_i, i < n\}$. When it is needed, the definition will be changed to a larger σ-algebra.

A set $A \subset H$ is said to be *locally asymptotically stable in the sense of Liapunov* for the ODE

$$\dot{\theta} = \bar{g}(\theta) + z, \quad z \in -C(\theta), \tag{2.1}$$

if for each $\delta > 0$ there is $\delta_1 > 0$ such that all trajectories starting in $N_{\delta_1}(A)$ never leave $N_\delta(A)$ and ultimately stay in $N_{\delta_1}(A)$.

Assumptions. The assumptions listed here, which will be used in Theorem 2.1, are more or less classical except for the generality of the possible limit set for the mean limit ODE and the use of a constraint set. All the conditions will be weakened in subsequent theorems. The proof is more complicated than the "minimal" convergence proof, since the algorithm is not necessarily of the gradient descent form and we do not insist that there be a unique limit point, but allow the algorithm to have a possibly complicated asymptotic behavior. Also, the proof introduces decompositions and interpolations that will be used in the sequel, as well as the basic idea of the ODE method for the probability one convergence. Condition (A2.2) simply sets up the notation, where β_n satisfies (A2.5). Condition (A2.6) (and (A2.6′) as well) is intended to describe the limits of θ_n in terms of the limit points of the ODE. The motivating example is where the set L_H^1 in (A2.6) is either empty or a set of unstable or marginally stable limit points, and the set of remaining limit points, A_H, is asymptotically stable in the sense of Liapunov. The condition is not needed in the "gradient descent" case, and stronger results are obtained without it in Subsection 2.2, where it is shown that under essentially the other conditions of this subsection, the process converges to the subset of the limit points consisting of "chain recurrent" points, a natural set.

(A2.1) $\sup_n E|Y_n|^2 < \infty$.

(A2.2) There is a measurable function $\bar{g}(\cdot)$ of θ and random variables β_n such that

$$E_n Y_n = E\left[Y_n|\theta_0, Y_i, i < n\right] = \bar{g}(\theta_n) + \beta_n.$$

(A2.3) $\bar{g}(\cdot)$ is continuous.

(A2.4) $\sum_i \epsilon_i^2 < \infty$.

(A2.5) $\sum_i \epsilon_i|\beta_i| < \infty$ w.p.1.

The following condition will sometimes be used.

(A2.6) Let L_H^1 be a subset of L_H and A_H a set that is locally asymptotically stable in the sense of Liapunov. Suppose that for any initial condition not in L_H^1 the trajectory of (2.1) goes to A_H.

For later use in this chapter, we will restate (A2.6) when a differential inclusion replaces the ODE.

(A2.6′) Let L_H now denote the (compact) set of limit points of the differential inclusion

$$\dot{\theta} \in G(\theta) + z, \quad z(t) \in -C(\theta(t)),$$

over all initial conditions in H, where z is the minimum force needed to keep the solution in H. Let L_H^1 be a subset of L_H and A_H a set that is locally asymptotically stable in the sense of Liapunov. Suppose that for any initial condition not in L_H^1 the trajectory goes to A_H.

Suppose that there is a continuously differentiable real-valued function $f(\cdot)$ such that $\bar{g}(\cdot) = -f_\theta(\cdot)$. Then the points in L_H are the stationary points, called S_H; they satisfy the stationarity condition

$$\bar{g}(\theta) + z = 0 \quad \text{for almost all } t, \quad z \in -C(\theta). \tag{2.2}$$

The set of stationary points can be divided into disjoint compact and connected subsets $S_i, i = 0, \ldots$ The following unrestrictive condition is needed.

(A2.7) Let $\bar{g}(\cdot) = -f_\theta(\cdot)$ for continuously differentiable real-valued $f(\cdot)$. Then $f(\cdot)$ is constant on each S_i.

If $f(\cdot)$ and the $q_i(\cdot)$ in (A4.3.2) (which define the constraint set) are twice continuously differentiable, then (A2.7) holds.

Comment on equality constraints and smooth manifolds. The equality constrained problem and the case where the constraint set H is a smooth manifold in $\mathbb{R}^{r-1}$ are covered by the results of the book. A convenient alternative approach that works directly on the manifold and effectively avoids the reflection terms can be seen from the following comments. The reader can fill in the explicit conditions that are needed. Suppose that the constraint set H is a smooth manifold. The algorithm $\theta_{n+1} = \Pi_H(\theta_n + \epsilon_n Y_n)$ can be written as

$$\theta_{n+1} = \theta_n + \epsilon_n \gamma(\theta_n) Y_n + \epsilon_n \beta_n,$$

where $\gamma(\cdot)$ is a smooth function and $\epsilon_n \gamma(\theta_n) Y_n$ is the projection of $\epsilon_n Y_n$ onto the orthogonal complement of the normal hyperplane (or line, depending on the case) to H at the point θ_n, and $\epsilon_n \beta_n$ represents the "error." Under reasonable conditions on the smoothness and on the sequence $\{Y_n\}$, the sequences $\{\gamma(\theta_n)Y_n, \beta_n\}$ will satisfy the conditions required by the $\{Y_n, \beta_n\}$ in the theorems. The mean limit ODE will be $\dot{\theta} = \gamma(\theta)\bar{g}(\theta)$. Similar comments hold when the ODE is replaced by a differential inclusion, for the correlated noise case of Chapter 6 and the various weak convergence cases of Chapters 7 and 8. The results can be extended to the case where H is the intersection of the $\mathbb{R}^{r-1}$-dimensional manifold defined by (A4.3.3) and a set satisfying (A4.3.2) or (A4.3.1).

Theorem 2.1. *Let* (1.1) *and* (A2.1)–(A2.5) *hold for algorithm* (1.3). *Then there is a set N of probability zero such that for $\omega \notin N$, the set of functions $\{\theta^n(\omega,\cdot), Z^n(\omega,\cdot), n < \infty\}$ is equicontinuous. Let $(\theta(\omega,\cdot), Z(\omega,\cdot))$ denote the limit of some convergent subsequence. Then this pair satisfies the projected ODE* (2.1), *and $\{\theta_n(\omega)\}$ converges to some invariant set of the ODE in H. If the constraint set is dropped, but $\{\theta_n\}$ is bounded with probability one, then for almost all ω, the limits $\theta(\omega,\cdot)$ of convergent subsequences of $\{\theta^n(\omega,\cdot)\}$ are trajectories of*

$$\dot{\theta} = \bar{g}(\theta) \tag{2.3}$$

in some bounded invariant set and $\{\theta_n(\omega)\}$ converges to this invariant set. Let p_n be integer-valued functions of ω, not necessarily being stopping times or even measurable, but that go to infinity with probability one. Then the conclusions concerning the limits of $\{\theta^n(\cdot)\}$ hold with p_n replacing n. If $\bar{\theta}$ is an asymptotically stable point of (2.1) *and θ_n is in some compact set in the domain of attraction of $\bar{\theta}$ infinitely often with probability $\geq \rho$, then $\theta_n \to \bar{\theta}$ with at least probability ρ.*

Assume (A2.6). *Then the limit points are in $L_H^1 \cup A_H$ with probability one.*

Suppose that (A2.7) *holds. Then, for almost all* ω, $\{\theta_n(\omega)\}$ *converges to a unique* S_i.

Remark. In many applications where $-\bar{g}(\cdot)$ is a gradient and the truncation bounds are large enough, there is only one stationary point of (2.1), and that is globally asymptotically stable. Then $\{\theta_n\}$ converges w.p.1 to that point. For simplicity, we use *equicontinuity* to mean "equicontinuity in the extended sense," as defined in the definition preceding Theorem 4.2.2.

Proof: Part 1. Convergence of the martingale and equicontinuity. Recall that $\delta M_n = Y_n - \bar{g}(\theta_n) - \beta_n$, and decompose the algorithm (1.3) as

$$\theta_{n+1} = \theta_n + \epsilon_n \bar{g}(\theta_n) + \epsilon_n Z_n + \epsilon_n \delta M_n + \epsilon_n \beta_n. \tag{2.4}$$

Then we can write

$$\begin{aligned} \theta^n(t) = \theta_n &+ \sum_{i=n}^{m(t+t_n)-1} \epsilon_i \bar{g}(\theta_i) + \sum_{i=n}^{m(t+t_n)-1} \epsilon_i Z_i \\ &+ \sum_{i=n}^{m(t+t_n)-1} \epsilon_i \delta M_i + \sum_{i=n}^{m(t+t_n)-1} \epsilon_i \beta_i. \end{aligned} \tag{2.5}$$

Define $M_n = \sum_{i=0}^{n-1} \epsilon_i \delta M_i$. This is a martingale sequence (with associated σ-algebras $\mathcal{F}_n$), since we have centered the summands about their conditional expectations, given the "past." By (4.1.4), for each $\mu > 0$,

$$P\left\{ \sup_{n \geq m} |M_n| \geq \mu \right\} \leq \frac{E\left|\sum_{i=m}^{n} \epsilon_i \delta M_i\right|^2}{\mu^2}.$$

By (A2.1), (A2.4), and the fact that $E\delta M_i \delta M_j' = 0$ for $i \neq j$, the right side is bounded above by $K \sum_{i=m}^{\infty} \epsilon_i^2$, for some constant K. Thus, for each $\mu > 0$,

$$\lim_m P\left\{ \sup_{n \geq m} |M_n| \geq \mu \right\} = 0. \tag{2.6}$$

Since $\theta^n(\cdot)$ is piecewise constant, we can rewrite (2.5) as

$$\theta^n(t) = \theta_n + \int_0^t \bar{g}(\theta^n(s))ds + Z^n(t) + M^n(t) + B^n(t) + \rho^n(t), \tag{2.7}$$

where $\rho^n(t)$ is due to the replacement of the first sum in (2.5) by an integral. $\rho^n(t) = 0$ at the times $t = t_k - t_n$ at which the interpolated processes have jumps, and it goes to zero uniformly in t as $n \to \infty$. By (2.6) and (A2.5), there is a null set N such that for $\omega \notin N$, $M^n(\omega, \cdot)$ and $B^n(\omega, \cdot)$ go to zero uniformly on each bounded interval in $(-\infty, \infty)$ as $n \to \infty$.

Let $\omega \notin N$. Then, $M^n(\omega, \cdot)$ and $B^n(\omega, \cdot)$ go to zero on any finite interval as $n \to \infty$. By the definition of N, for $\omega \notin N$ the functions of t on the right side of (2.7) (except for $Z^n(\cdot)$) are equicontinuous in n and the limits of $M^n(\cdot)$, $B^n(\cdot)$, and $\rho^n(\cdot)$ are zero. It will next be shown that the equicontinuity of $\{Z^n(\omega,\cdot), n < \infty\}$ is a consequence of the fact that

$$Z_n(\omega) \in -C(\theta_{n+1}(\omega)). \tag{2.8}$$

For $\omega \notin N$, $\theta_{n+1}(\omega) - \theta_n(\omega) \to 0$. If $Z^n(\omega,\cdot)$ is not equicontinuous, then there is a subsequence that has a jump asymptotically; that is, there are integers $\mu_k \to \infty$, uniformly bounded times s_k, $0 < \delta_k \to 0$ and $\rho > 0$ (all depending on ω) such that

$$|Z^{\mu_k}(\omega, s_k + \delta_k) - Z^{\mu_k}(\omega, s_k)| \geq \rho.$$

The changes of the terms other than $Z^n(\omega,t)$ on the right side of (2.7) go to zero on the intervals $[s_k, s_k + \delta_k]$. Furthermore $\epsilon_n Y_n(\omega) = \epsilon_n \bar{g}(\theta_n(\omega)) + \epsilon_n \delta M_n(\omega) + \epsilon_n \beta_n \to 0$ and $Z_n(\omega) = 0$ if $\theta_{n+1}(\omega) \in H^0$, the interior of H. Thus, this jump cannot force the iterate to the interior of the hyperrectangle H, and it cannot force a jump of the $\theta^n(\omega,\cdot)$ along the boundary either. Consequently, $\{Z^n(\omega,\cdot)\}$ is equicontinuous.

Part 2. Characterizing the limit of a convergent subsequence: Applying the Arzelà–Ascoli Theorem. Let $\omega \notin N$, and let n_k denote a subsequence such that

$$\{\theta^{n_k}(\omega,\cdot), Z^{n_k}(\omega,\cdot)\}$$

converges, and denote the limit by $(\theta(\omega,\cdot), Z(\omega,\cdot))$. Then

$$\theta(\omega,t) = \theta(\omega,0) + \int_0^t \bar{g}(\theta(\omega,s))ds + Z(\omega,t). \tag{2.9}$$

Note that $Z(\omega,0) = 0$ and $\theta(\omega,t) \in H$ for all t. To characterize $Z(\omega,t)$, use (2.8) and the fact that $\theta_{n+1}(\omega) - \theta_n(\omega) \to 0$. These facts, together with the upper semicontinuity property (4.3.2) and the continuity of $\theta(\omega,\cdot)$, imply that (4.3.4) and (4.3.5) hold. In fact, it follows from the method of construction of $Z(\omega,\cdot)$ that the function simply serves to keep the dynamics $\bar{g}(\cdot)$ from forcing $\theta(\omega,\cdot)$ out of H. Thus, for $s > 0$, $|Z(\omega,t+s) - Z(\omega,t)| \leq \int_t^{t+s} |\bar{g}(\theta(\omega,u)|du$. Hence $Z(\omega,\cdot)$ is Lipschitz continuous, and $Z(\omega,t) = \int_0^t z(\omega,s)ds$, where $z(\omega,t) \in -C(\theta(\omega,t))$ for almost all t.

Recall the definition of the set A_H in (A2.6). Suppose that $\{\theta_n(\omega)\}$ has a limit point $x_0 \notin L_H^1 \cup A_H$. Then there is a subsequence m_k such that $\theta^{m_k}(\omega,\cdot)$ converges to a solution of (2.1) with initial condition x_0. Let $\delta > \delta_1 > 0$ be arbitrarily small. Since the trajectory of (2.1) starting at x_0 ultimately enters $N_{\delta_1}(A_H)$ by (A2.6), $\theta_n(\omega)$ must be in $N_{\delta_1}(A_H)$ infinitely often. It will be seen that escape from $N_\delta(A_H)$ infinitely often is impossible. Suppose that escape from $N_{\delta_1}(A_H)$ occurs infinitely often. Then, since

$\theta_{n+1}(\omega) - \theta_n(\omega) \to 0$, there are integers n_k such that $\theta_{n_k}(\omega)$ converges to some point x_1 on $\partial N_{\delta_1}(A_H)$, and the path $\theta^{n_k}(\omega, \cdot)$ converges to a solution of (2.1) starting at x_1. The path of (2.1) starting at x_1 (whatever the chosen convergent subsequence) never leaves $N_\delta(A_H)$ and ultimately stays in $N_{\delta_1}(A_H)$. This implies that $\theta_n(\omega)$ cannot exit $N_\delta(A_H)$ infinitely often.

Whether or not there is a constraint set H, if boundedness with probability one of the sequence $\{\theta_n\}$ is assumed, then the preceding arguments show that (with probability one) the limits of $\{\theta^n(\omega, \cdot)\}$ are bounded solutions to (2.1) (which is (2.3) if there is no constraint) on the doubly infinite time interval $(-\infty, \infty)$. Thus the entire trajectory of a limit $\theta(\omega, \cdot)$ must lie in a bounded invariant set of (2.1) by the definition of an invariant set. The fact that $\{\theta_n(\omega)\}$ converges to some invariant set of (2.1) then follows; otherwise there would be a limit of a convergent subsequence satisfying (2.1) but not lying entirely in an invariant set.

These arguments do not depend on how the "sections" of $\theta^0(\omega, \cdot)$ are chosen. Any set of "sections" other than $\theta^n(\omega, \cdot)$ could have been used, as long as the initial times went to infinity. The statement of the theorem concerning $\{p_n\}$ then follows from what has been done.

Part 3. The case when $-\bar{g}(\cdot)$ is a gradient. Now assume (A2.7) and suppose that $\bar{g}(\cdot) = -f_\theta(\theta)$ for some continuously differentiable function $f(\cdot)$. As will be shown, the conclusion concerning the limits actually follows from what has been done.

We continue to work with $\omega \notin N$. Suppose for simplicity that there are only a finite number of S_i, namely, $S_0, \ldots, S_M$. In (2.1), $|z(t)| \le |\bar{g}(\theta(t))|$. Thus, if $\bar{g}(\cdot) = -f_\theta(\cdot)$, the derivative along the solution of (2.1) at $\theta \in H$ is $f_\theta'(\theta)\left[-f_\theta(\theta) + z\right] \le 0$, and we see that all solutions of (2.1) tend to the set of stationary points defined by (2.2). For each c, the set $\{\theta : f(\theta) \le c\}$ is locally asymptotically stable in the sense of Liapunov, assuming that it is not empty. Then the previous part of the proof implies that $f(\theta_n(\omega))$ converges to some constant (perhaps depending on ω), and $\theta_n(\omega)$ converges to the set of stationary points.

It remains to be shown that $\{\theta_n(\omega)\}$ converges to a unique S_i. If the claimed convergence does not occur, the path will eventually oscillate back and forth between arbitrarily small neighborhoods of distinct S_i. This implies that there is a limit point outside the set of stationary points. □

An elaboration of the proof for the gradient descent case. For future use, and as an additional illustration of the ODE method, we will elaborate the proof for the case where $\bar{g}(\cdot)$ is a gradient. The ideas are just those used in the previous proof. The details to be given are of more general applicability and will be used in Theorems 4.2 and 4.3 in combination with a Liapunov function technique.

We start by supposing that the path $\theta^0(\omega, t)$ oscillates back and forth between arbitrarily small neighborhoods of distinct S_i. This will be seen to contradict the "gradient descent" property of the ODE (2.1). The proof simply sets up the notation required to formalize this idea.

Since $\{\theta_n(\omega)\}$ converges to $S_H = \cup_i S_i$, there is a subsequence m_k such that $\theta_{m_k}(\omega)$ tends to some point $x_0 \in S_H$. Suppose that $x_0 \in S_0$. We will show that $\theta_n(\omega) \to S_0$. Suppose that this last convergence hypothesis is false. Then there is an x_1 in some $S_i, i \neq 0$ (call it S_1 for specificity), and a subsequence $\{q_k\}$ such that $\theta_{q_k}(\omega) \to x_1$.

Continuing this process, let $S_0, \ldots, S_R, R > 0$, be all the sets that contain limit points of the sequence $\{\theta_n(\omega)\}$. Order the sets such that $f(S_R) = \liminf_n f(\theta_n(\omega))$, and suppose that S_R is the (assumed for simplicity) unique set on which the liminf is attained. The general (nonunique) case requires only a slight modification. Let $\delta > 0$ be such that $f(S_R) < f(S_i) - 2\delta, i \neq R$. For $\rho > 0$, define the ρ-neighborhood $N_\rho^f(S_R)$ of S_R by $N_\rho^f(S_R) = \{x : f(x) - f(S_R) < \rho\}$. By the definition of δ, $N_{2\delta}^f(S_R)$ contains no point in any $S_i, i \neq R$. By the hypothesis that more than one S_i contains limit points of $\{\theta_n(\omega)\}$, the neighborhood $N_\delta^f(S_R)$ of S_R is visited infinitely often and $N_{2\delta}^f(S_R)$ is exited infinitely often by $\{\theta_n(\omega)\}$. Thus there are $\nu_k \to \infty$ (depending on ω) such that $\theta_{\nu_k - 1}(\omega) \in N_\delta^f(S_R)$, $\theta_{\nu_k}(\omega) \notin N_\delta^f(S_R)$, and after time ν_k the path does not return to $N_\delta^f(S_R)$ until after it leaves $N_{2\delta}^f(S_R)$.

By the equicontinuity of $\{\theta^n(\omega, \cdot)\}$, there is a $T > 0$ such that $\theta_{\nu_k}(\omega) \to \partial N_\delta^f(S_R)$, the boundary of $N_\delta^f(S_R)$, and for large k, $\theta^{\nu_k}(\omega, t) \notin N_\delta^f(S_R)$ for $t \in [0, T]$.

There is a subsequence $\{\mu_m\}$ of $\{\nu_k\}$ such that $(\theta^{\mu_m}(\omega, \cdot), Z^{\mu_m}(\omega, \cdot))$ converges to some limit $(\bar\theta(\omega, \cdot), \bar Z(\omega, \cdot))$ that satisfies (2.1) with $\bar\theta(\omega, 0) \in \partial N_\delta^f(S_R)$, the boundary of $N_\delta^f(S_R)$, and $\bar\theta(\omega, t) \notin N_\delta^f(S_R)$ for $t \leq T$. This is a contradiction because, by the gradient descent property of (2.1) (with $\bar g(\cdot) = -f_\theta(\cdot)$) and the definitions of δ and $N_\delta^f(S_R)$, any solution to (2.1) starting on $\partial N_\delta^f(S_R)$ must stay in $N_\delta^f(S_R)$ for all $t > 0$. □

The argument in Part 3 of the preceding proof is of more general use and leads to the following result.

Theorem 2.2. *Let $\{\theta^n(\omega, \cdot), Z^n(\omega, \cdot)\}$ be equicontinuous, with all limits satisfying* (2.1). *Suppose that $\{\theta_n(\omega)\}$ visits a set A_0 infinitely often, where A_0 is locally asymptotically stable in the sense of Liapunov. Then $\theta_n(\omega) \to \bar A_0$, the closure of A_0.*

Remark on the proof of Theorem 2.1. Let us review the structure of the proof. First, the increment was partitioned to get the convenient representation (2.7), with which we could work on one part at a time. Then it was shown that with probability one the martingale term M_n converged, and this implied that $M^n(\cdot)$ converged with probability one to the "zero" process. Then the probability one convergence to zero of the bias $\{B^n(\cdot)\}$

was shown. The asymptotic continuity of $Z^n(\cdot)$ was obtained by a direct use of the properties of the Z_n as reflection terms. Then, by fixing ω not in some "bad" null set, and taking convergent subsequences of $\{\theta^n(\omega,\cdot), Z^n(\omega,\cdot)\}$, we were able to characterize the limit as a solution to the mean limit ODE. It then followed that the sequence $\{\theta_n(\omega)\}$ converged to some invariant set of the ODE.

A more general constraint set. Using the same basic structure of the proof, Theorem 2.1 can be readily generalized in several useful directions with little extra work. (A2.1) and (A2.4) will be weakened in the next section. Appropriate dependence on n of $\bar{g}(\cdot)$ can be allowed, and the hyperrectangle H can be replaced by a more general constraint set. The techniques involved in the required minor alterations in the proofs will be important in the analysis of the "dependent" noise case in Chapter 6.

In Theorem 2.3, the constraint form (A4.3.2) or (A4.3.3) will be used, where (A4.3.2) includes (A4.3.1). For $\theta \in I\!R^r$, let $\Pi_H(\theta)$ denote the closest point in H to θ. If the closest point is not unique, select a closest point such that the function $\Pi_H(\cdot)$ is measurable. We will work with the algorithm

$$\theta_{n+1} = \Pi_H\left[\theta_n + \epsilon_n Y_n\right], \tag{2.10a}$$

that will be written as

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n + \epsilon_n Z_n, \tag{2.10b}$$

where Z_n is the projection term. Recall the definition of $C(\theta)$ from Section 4.3. It follows from the calculus that $Z_n \in -C(\theta_{n+1})$, under (A4.3.3). Under (A4.3.2), this is proved by applying the Kuhn–Tucker Theorem of nonlinear programming to the problem of minimizing $|x - (\theta_n + \epsilon_n Y_n)|^2$ subject to the constraints $q_i(x) \leq 0, i \leq p$, where $x = \theta_{n+1}$. That theorem says that there are $\lambda_i \geq 0$ with $\lambda_i = 0$ if $q_i(x) < 0$ such that

$$(x - (\theta_n + \epsilon_n Y_n)) + \sum_i \lambda_i q_{i,x}(x) = 0,$$

which implies that $Z_n \in -C(\theta_{n+1})$.

The following assumption generalizes (A2.2) and will in turn be relaxed in the next section.

(A2.8) There are functions $g_n(\cdot)$ of θ, which are continuous uniformly in n, a continuous function $\bar{g}(\cdot)$ and random variables β_n such that

$$E_n Y_n = g_n(\theta_n) + \beta_n, \tag{2.11}$$

and for each $\theta \in H$

$$\lim_n \left| \sum_{i=n}^{m(t_n+t)} \epsilon_i \left[g_i(\theta) - \bar{g}(\theta)\right] \right| \to 0 \tag{2.12}$$

for each $t > 0$. (In other words, $\bar{g}(\cdot)$ is a "local average" of the functions $g_n(\cdot)$.)

Dependence of $\bar{g}(\cdot)$ on the past. Note that in all the algorithmic forms, $g_n(\theta_n)$ can be replaced by dependence on the past of the form $g_n(\theta_n, \ldots, \theta_{n-K})$ provided that the continuity of $g_n(\cdot)$ is replaced by the continuity in x of $g_n(x_0, \ldots, x_K)$ on the "diagonal" set $x = x_0 = \cdots = x_K$, uniformly in n.

Theorem 2.3. *Assume the conditions of Theorem* 2.1 *but use the algorithm* $\theta_{n+1} = \Pi_H[\theta_n + \epsilon_n Y_n]$ *with any of the constraint set conditions* (A4.3.1), (A4.3.2), *or* (A4.3.3) *holding, and* (A2.8) *with* $\beta_n \to 0$ *with probability one replacing* (A2.2) *and* (A2.5). *Then the conclusions of Theorem* 2.1 *continue to hold.*

Remark on the proof. The proof is essentially the same as that of Theorem 2.1, and we concentrate on the use of (A2.8) and the equicontinuity of $\{Z^n(\cdot)\}$ with probability one. The equicontinuity proof exploits the basic character of Z_n as *projection terms* to get the desired result, and the proof can readily be used for the general cases of Section 4 and Chapter 6.

Proof. Define

$$\bar{G}^n(t) = \sum_{i=n}^{m(t_n+t)-1} \epsilon_i \bar{g}(\theta_i), \quad \tilde{G}^n(t) = \sum_{i=n}^{m(t_n+t)-1} \epsilon_i \left[g_i(\theta_i) - \bar{g}(\theta_i)\right].$$

For simplicity, we only work with $t \geq 0$. With these definitions,

$$\theta^n(t) = \theta_n + \bar{G}^n(t) + \tilde{G}^n(t) + B^n(t) + M^n(t) + Z^n(t). \tag{2.13}$$

As in Theorem 2.1, (A2.1) and (A2.4) imply that $M^n(\cdot)$ converges to the "zero" process with probability one as $n \to \infty$. Since $\beta_n \to 0$ with probability one, the process $B^n(\cdot)$ also converges to zero with probability one. Since $g_n(\cdot)$ and $\bar{g}(\cdot)$ are uniformly bounded on H, the set $\{\bar{G}^n(\omega, \cdot), \tilde{G}^n(\omega, \cdot)\}$ is equicontinuous for each ω. These bounds and convergences imply that the jumps in $\bar{G}^n(\cdot) + \tilde{G}^n(\cdot) + M^n(\cdot) + B^n(\cdot)$ on any finite interval go to zero with probability one as $n \to \infty$. Consequently, with probability one the distance of $\theta_n + \epsilon_n Y_n$ to H goes to zero as $n \to \infty$.

Now fix attention on the case where H satisfies (A4.3.2). [The details under (A4.3.3) are left to the reader.] Then, if we were to ignore the effects of the terms $Z^n(\cdot)$, $\{\theta^n(\omega, \cdot)\}$ would be equicontinuous on $(-\infty, \infty)$ for ω not in a null set N, the set of nonconvergence to zero of $B^n(\omega, \cdot)$ or of $M^n(\omega, \cdot)$. Thus the only possible problem with equicontinuity would originate from Z_n. That this is not possible follows from the following argument.