Notes

1. See Mount and Reiter (1990), Appendix A. For the continuous case, see Definition 3.
2. The space P^2 consists of the collection of lines passing through the origin of Euclidean 3-space R^3. Therefore, P^2 can be considered to be the collection of equivalence classes of the points in E^3-{(0, 0, 0)} under the equivalence relation $(x, y, z) \approx (tx, ty, tz)$ for each $t \neq 0$. The topology used is the quotient topology. If E^3 has coordinates (X, Y, Z), then these coordinates are also called homogeneous coordinates on P^2. The space P^2 is a nonoriented closed C^r-manifold. For a discussion of P^n in general, see Eisenberg (1974).

References

Eisenberg, M. 1974. *Topology*. New York: Holt, Rinehart and Winston, Inc.

Golubitsky, M., and Guillemin, V. 1973. "Stable Mappings and Their Singularities." In *Graduate Texts in Mathematics*, No. 14. New York: Springer-Verlag.

Klein, E., and Thompson, A. C. 1984. *Theory of Correspondences*. Canadian Mathematical Society Series of Monographs and Advanced Texts. New York: John Wiley and Sons.

Michael, E. "Topologies on Spaces of Subsets." *Transactions of the American Mathematical Society* 71, 152–182.

Mount, K. R., and Reiter, S. 1990. "A Model of Computing with Human Agents." Discussion Paper No. 890, The Center for Mathematical Studies in Economics, Northwestern University.

3 CONVERGENCE THEOREMS FOR A CLASS OF RECURSIVE STOCHASTIC ALGORITHMS

Diego Moreno and Mark Walker

Several recent studies of the way individual economic units might learn their parts in an economic or strategic equilibrium have modeled the learning process as a recursive algorithm with stochastic features. The central idea in each of these studies has been to explain or justify the notion of equilibrium by demonstrating that an equilibrium is a stationary point to which the learning process converges. This approach to equilibrium analysis—and in particular the modelling of learning in terms of a recursive stochastic algorithm—seems attractive and powerful, and we expect its use to become more widespread.

Because the central issue in this approach is the convergence of recursive stochastic algorithms, the theory of convergence for such algorithms is clearly an important technical tool. The theoretical framework that these recent studies have used is the one developed by Ljung (1975, 1977). Ljung defines a general class of estimation or forecasting algorithms, and he provides theorems that give conditions under which these algorithms will converge to their stationary points. The algorithms are particularly apt for modelling the learning of equilibrium.

Ljung's contributions, despite their obvious importance, are fraught with a number of difficulties. Our aim here is to clarify and resolve many of these difficulties, and thereby to provide a stronger foundation for the convergence results and to smooth the path for applying the framework and the results. Most of the difficulties in Ljung's analysis arise from his attempt at maximum generality, which has two kinds of consequences. First, the theory is much more difficult to apply than it ought to be; in particular, it is difficult to determine in applications whether Ljung's assumptions are satisfied, and it is easy to apply the theory incorrectly (see, for example, Remark 1 in section 3). And second, it is very difficult to determine whether Ljung's proofs are correct for such convoluted assumptions. (The proofs are even more convoluted than the assumptions. Ljung (1977) acknowledges that some of his assumptions "admittedly look somewhat complex" and that "the many technicalities [in the proofs] tend to obscure the simple idea.")

In order to clarify the convergence theory and to obtain clear, revealing proofs, we will concentrate on a linear form for the algorithms, a somewhat less general form than Ljung's. We will lay out the framework clearly and will use simple, clear assumptions. This enables us to give a straightforward and complete proof, in which it is easy to understand the proof's structure and to determine whether its steps are correct. While all this is done for a less general algorithm than Ljung's, the proof given here provides a clear model for moving to applications that, in one respect or another, fail to fit within this form.

The most serious of the difficulties we will address is that Ljung's proofs are not entirely correct. For the most part, we show how the missteps in his proofs can be avoided. There is one error, however, that does not seem to be correctible in the general case that Ljung treats. We give a correct proof for our linear version of the algorithm, and we indicate how strengthening one of the assumptions will avoid the problem in the general case. It remains an open question whether the theorem is true with Ljung's original weaker assumption (see Remarks 3 and 4 in section 3 for discussion of this issue).

Another of the difficulties with Ljung's proofs concerns his "projection" algorithms, the kind of algorithm that has been used in the economic studies that deal with rational expectations. The proof of convergence that was given for these algorithms is not correct (it appeals incorrectly to lemmas obtained for nonprojection algorithms). We treat projection and nonprojection algorithms together, in a single unified proof, thereby obtaining a clear and correct proof for projection algorithms (see Remark 2 in section 3).

In an attempt to apply projection algorithms to economic situations in which individuals have disparate information and objectives, Marcet and Sargent (1989b) and Moreno and Walker (1991) have defined "decentralized" projection algorithms (Marcet and Sargent do not use this terminology). We indicate in Remark 1 of section 3 why our proof (and Ljung's) does not apply to such algorithms, and we also indicate what would be required in order to devise a correct proof of convergence for such algorithms.

We emphasize that most of the constructions in the proof we provide are taken more or less directly from Ljung (1975, 1977). We show how those constructions can (and cannot) be incorporated into a correct proof of convergence; the clarity of the analysis we provide should make for much greater ease in verifying that the analysis is correct, as well as in applying the algorithms and theorems, and in generalizing them to deal with new situations.

1. The Algorithms and the Theorems

Consider a system, evolving over time, in which estimates or forecasts (of parameters, actions, the state of the system, etc.) must be formed, and in which these estimates determine the current state of the system. We assume that the estimates are formed via a recursive rule, or algorithm; in forming the current estimate, only the preceding estimate and the current state are used. Formally, we have the following difference equation system, which we will call Ljung's Basic Algorithm:

$$q(t) = \mathscr{A}(x(t-1))q(t) + \mathscr{B}(x(t-1))e(t), \quad \text{the state at } t; \tag{1}$$

$$x(t) = x(t-1) + \gamma(t)Q(t, x(t-1), q(t)), \quad \text{the estimate at } t. \tag{2}$$

We assume that $q(t) \in \mathbb{R}^m$, $x(t) \in \mathbb{R}^n$, $e(t)$ is a random variable taking values in $\mathbb{R}^l$, that $\gamma(t)$ is a real number, and that (for each $x \in \mathbb{R}^n$) $\mathscr{A}(x)$ and $\mathscr{B}(x)$ are matrices of appropriate dimension.

It is natural to use this framework to model economic processes that involve interaction among agents who do not know the true parametric structure of the process, but who must nevertheless take actions based upon some estimate of the structure, and in which those actions in turn determine the data which the agents will use in forming their subsequent estimates. Marcet and Sargent (1989a, 1989b) and Woodford (1990) contain applications of the framework to rational expectations (q represents such variables as prices, and x represents individuals' forecasts of those variables); Moreno and Walker (1991) contains an application to Nash equilibrium in a simple duopoly game ($m = n = 2$; q is the list of the firms' output levels; and x is the list of the firms' forecasts of their rivals' output levels). In each of these applications, the question addressed is whether the system will converge to the equilibrium—whether the participants will "learn" to play their parts of an equilibrium. The question whether a system of the form (1) & (2) will converge to a stationary state is therefore of interest to economists.

Ljung's method for studying the convergence properties of recursive stochastic algorithms of the form (1) & (2) is to associate an ordinary differential equation (called the *Associated Differential Equation*, or ADE for short)

$$\frac{dx}{dt} = f(x) \tag{3}$$

with the algorithm; then to show that the algorithm (more specifically, its sequence of estimates $x(t)$) "behaves asymptotically like the differential equation"; and then to study the convergence properties of the differential equation, which is generally far easier than directly studying the convergence properties of the time-varying stochastic difference equation system (1) & (2).

Ljung derives the ADE for a system of the form (1) & (2) as follows. Denote by $\bar{q}(t, x)$ the path of the state $q(t)$ under the restriction that the estimate $x(t)$ is fixed at x instead of updated via (2):

$$\bar{q}(t, x) = \mathscr{A}(x)\bar{q}(t-1, x) + \mathscr{B}(x)e(t), \quad \text{and} \quad \bar{q}(0, x) = 0;$$

and denote by $\bar{Q}(t, x)$ the corresponding path of the value of the updating function Q:

$$\bar{Q}(t, x) = Q(t, x, \bar{q}(t, x)).$$

Clearly, $f(x)$ ought to behave in some respect like $\bar{Q}(t, x)$ if it is going to serve as a surrogate for the algorithm: Each is a description of how the estimate $x(t)$ is updated if its current value is x. $\bar{Q}(t, x)$ tells how the actual $x(t)$ behaves at x, and $f(x)$ tells how the surrogate continuous-time system of the ADE behaves at x. Although the intuition here leaves something to be desired, Ljung nevertheless defines $f(x)$ as

$$f(x) = \lim_{t\to\infty} E\bar{Q}(t, x),$$

and then goes about proving—at great length and difficulty, but for the most part successfully—that indeed the algorithm "behaves asymptotically" like the ADE if the ADE is defined in this way, and that the algorithm's convergence properties do indeed essentially coincide with those of the ADE. (He requires, of course, that the limit that defines $f(x)$ be a well-defined finite vector for each x.) We will assume throughout that the ADE has at least one stationary point x^*, and we denote by $\mathscr{D}(x^*, f)$ the domain of attraction of x^*—i.e., the set of x_0 for which the solution $X(\cdot, t)$ of the ADE satisfies $\lim_{t\to\infty} X(x_0, t) = x^*$.

Ljung lays out several sets of assumptions on the algorithm and on the stochastic process $\{e(t)\}$, but in his effort to achieve maximum generality, he makes the assumptions extremely complicated and unintuitive. He then proves several theorems describing the algorithm's convergence properties under these assumptions. We will concentrate upon his two main theorems (Theorems 1 and 4 in Ljung, 1977), which we will refer to as Ljung's First Theorem and Ljung's Second Theorem.

Ljung's First Theorem: Under Ljung's assumptions, if a sequence $\{x(t)\}$ which is generated by the Basic Algorithm has a bounded subsequence that lies within $\mathscr{D}(x^*, f)$, then $\{x(t)\}$ converges almost surely to x^*.

The requirement that the sequence $\{x(t)\}$ have a bounded subsequence is not entirely satisfactory. If we convert it from a condition on individual sequences to a condition on the algorithm, we have: "If the algorithm almost surely generates sequences that have bounded subsequences, then the algorithm converges almost surely to x^*." This condition is generally very difficult to verify; in most applications, it may be no easier to verify it than to establish convergence by direct analysis of the difference equation system.

In order to overcome the difficulties with the boundedness condition, Ljung introduces a second algorithm, called a "projection algorithm," which bounds the path of the first algorithm by replacing "outlier" estimates with substitute estimates that lie in a bounded set. He proves that this projection algorithm behaves asymptotically like the original algorithm (i.e., the replacement, or "projection," is invoked only finitely many times), and, therefore, he has the following result (we will properly define the projection algorithm shortly):

Ljung's Second Theorem: Under Ljung's assumptions, if the Projection Algorithm satisfies Condition L (see below), then the algorithm converges almost surely to x^*.

Our own analysis will be carried out entirely for a linear form of the Basic Algorithm and the Projection Algorithm; for the remainder of this section we will also restrict our attention to the deterministic (nonstochastic) case. The Basic Algorithm will henceforth be described by the following difference equation system:

$$q(t) = Bx(t-1) \tag{4}$$

$$x(t) = x(t-1) + \frac{1}{t}[Gq(t) + Hx(t-1)],$$

where $q(t) \in \mathbb{R}^m$, $x(t) \in \mathbb{R}^n$ and B, G, and H are $m \times n$, $n \times m$, and $n \times n$ matrices. This algorithm has Ljung's form, as follows:

$$\mathscr{A}(x) = 0, \quad \mathscr{B}(x) = Bx, \quad \gamma(t) = \frac{1}{t}$$

$$Q(t, x, q) = Gq(t) + Hx(t-1), \quad e(t) \equiv 1.$$

Therefore, we have $f(x) = (GB + H)x$. We write A for the $n \times n$ matrix $GB + H$, and we have $f(x) = Ax$. Note that if A is nonsingular, then $x^* = 0$ is the unique stationary point of both the algorithm and its ADE; and that x^* is globally stable under f (i.e., $\mathscr{D}(x^*, f) = \mathbb{R}^n$) if A is a stable matrix (i.e., if all its eigenvalues are negative), and that otherwise $\mathscr{D}(x^*, f)$ is a proper subspace of $\mathbb{R}^n$, and x^* is an unstable stationary point of the ADE. (We have $x^* = 0$ as the stationary point because of the system's homogeneity; this is equivalent to the system being nonhomogeneous and to x and q being the *deviations* of the estimate and state from their equilibrium values.)

For the Basic Algorithm (4) & (5), Ljung's Boundedness Condition and his First Theorem can be given the following simple forms:

Condition B: An algorithm satisfies Condition B if, for any $x(0) \in \mathbb{R}^n$, the sequence $\{x(t)\}$ generated by the algorithm has a bounded subsequence.

First Convergence Theorem (Deterministic Version): If the Basic Algorithm (4) & (5) satisfies Conditions B, and if x^* is an asymptotically stable stationary point of the algorithm's ADE (i.e., if $\mathscr{D}(x^*, f)$ is an open set), then $\{x(t)\}$ converges to x^* for every $x(0) \in \mathbb{R}^n$.

We have already suggested that Condition B is generally difficult to verify. Ljung therefore introduces a *projection operator* which replaces "outlier" estimates with well-defined substitutes, thereby forcing the algorithm's estimates to always lie in a bounded set. Formally, the projection operator is defined as follows, where D is an arbitrary *bounded* subset of $\mathbb{R}^n$:

Let D be a subset of $\mathbb{R}^n$ with a nonempty interior D°; (5′a)

Let C be a nonempty closed subset of $D^\circ \cap \mathscr{D}(x^*, f)$; (5′b)

For any $x(0) \in \mathbb{R}^n$, $q(0) \in \mathbb{R}^m$, and $p(0) \in \mathbb{R}^n$, define $\{\tilde{x}(t)\}$, $\{p(t)\}$, and $\{x(t)\}$ as follows:

$$\tilde{x}(t) = x(t-1) + \frac{1}{t}[Gq(t) + Hx(t-1)], \tag{5′c}$$

$\{p(t)\}$ is an arbitrary sequence in C, and (5′d)

$$x(t) = \begin{cases} \tilde{x}(t) & \text{if } \tilde{x}(t) \in D \quad (5'e) \\ p(t) & \text{if } \tilde{x}(t) \notin D \quad (5'f) \end{cases}$$

The projection operator (5′) replaces (5) in the Basic Algorithm (4) & (5), giving us the Projection Algorithm (4) & (5′). We say that "the projection operator is invoked at t" in a sequence $\{x(t)\}$ if $\tilde{x}(t) \notin D$, or equivalently, if $x(t) \neq \tilde{x}(t)$.

The Projection Algorithm certainly satisfies Condition B; indeed, the entire sequence $\{x(t)\}$ will always be bounded. But so long as the projection operator is being invoked, the Projection Algorithm will not behave like its ADE (which, in D°, is the same as the ADE for the Basic Algorithm): It will be repeatedly "jumping." Thus, the First Convergence Theorem will not apply to the Projection Algorithm. It can be shown, however, that if the Projection Algorithm satisfies the following condition, then its projection operator will be invoked only finitely many times (for any $x(0) \in R^n$), so that eventually the Projection Algorithm will behave like the Basic Algorithm after all, and it will therefore converge to x^*.

Condition L: The algorithm (4) & (5′) is said to satisfy Condition L if there is a twice continuously differentiable function U: $\mathbb{R}^n \to \mathbb{R}_+$ for which

$\forall x \in D^\circ \setminus C$: $U'(x)f(x) < 0$ and

$\exists c_1, c_2 \in \mathbb{R}$ such that $0 \leqslant c_1 < c_2$ and $\forall x \in C$: $U(x) \leqslant c_1$ and

$\forall x \notin D$: $U(x) \geqslant c_2$,

where $U'(x)$ denotes the derivative (i.e., the gradient) of U at x.

Condition L requires the existence of a function U that behaves in the set $D^\circ \setminus C$ like a Lyapunov function for the ADE. It implies, in particular, that trajectories of the ADE that begin in C never leave the set D. The effect of Condition L is as follows: Because the "step size" $\tilde{x}(t) - x(t-1)$ in the Projection Algorithm goes to zero as t grows large, the step size eventually becomes so small that if $x(t) \in C$, as it is when the projection operator has been invoked, then $\tilde{x}(t+1)$ must still lie in D°; and because the algorithm behaves (in D°) asymptotically like the ADE, Condition L ensures that if the projection operator is invoked when t is large enough, then $\{x(t)\}$ will be trapped in D°, and the projection operator will therefore never again be invoked. If the set C is chosen sufficiently close to D, then Condition L will be satisfied if the trajectories of the ADE point inward on the boundary of D, which is often easier to verify than Condition B.

Second Convergence Theorem (Deterministic Version): If the Projection Algorithm ((4) & (5′) with D bounded) satisfies Condition L, and if x^* is an asymptotically stable stationary point of the algorithm's ADE, then $\{x(t)\}$ converges to x^* for every $x(0) \in \mathbb{R}^n$.

Suppose we had assumed that $D = \mathbb{R}^n$ in (5′), instead of that D is bounded. Then the projection operator would never be invoked, the set C and the sequences $\{p(t)\}$ would be irrelevant, and we would always have $x(t) = \tilde{x}(t)$. In other words, if $D = \mathbb{R}^n$, then (5′) is the same as (5). This observation leads us to adopt a unified treatment of the two algorithms: The Basic Algorithm is defined by (4) & (5′) with $D = \mathbb{R}^n$, and the Projection Algorithm is defined by (4) & (5′) with D bounded. This yields the following convergence theorem, which contains the First and Second Theorems as special cases:

Convergence Theorem (Deterministic Version): If the algorithm (4) & (5′) satisfies either

(a) $D = \mathbb{R}^n$ (the Basic Algorithm) and Condition B, or
(b) D is bounded (the Projection Algorithm) and Condition L,

and if x^* is an asymptotically stable stationary point of the algorithm's ADE, then $\{x(t)\}$ converges to x^* for every $x(0) \in \mathbb{R}^n$.

Section 3 contains a proof of the Convergence Theorem.

2. The Stochastic Convergence Theorem

Although it is not very clear in Ljung's papers, his approach to the convergence theory for recursive stochastic algorithms is essentially deterministic, with the proof for the stochastic case built upon the deterministic proof. We make this relation between the deterministic and the stochastic cases more explicit, and this enables us to provide, in section 4, a brief but complete proof for the stochastic case.

Assume that $(\Omega, \mathscr{E}, P)$ is a probability space, in which Ω consists of all the sequences $\{\omega(t)\}_1^\infty$ of points in $\mathbb{R}^m$; $\mathscr{E}$ is the set of Borel-measurable subsets ("events") of Ω; and P is a probability measure defined on $\mathscr{E}$. We redefine the algorithm to include, at each time t, the random term $\omega(t)$, as follows:

$$q(t) = Bx(t-1) + \omega(t). \tag{4′}$$

It is not necessary that the random vector have the same dimension as q. We could instead have $q(t) = Bx(t-1) + Su(t)$, where $u(t) \in \mathbb{R}^\ell$, S is $m \times \ell$, and $\omega(t) = Su(t)$. We take the sequences ω to be the elementary events of the probability model (the elements of Ω), but we also treat each $\omega(t)$ as a random variable that takes its values in $\mathbb{R}^n$). (In other words, we write $\omega(t)$ both for the tth random variable and for the random variable's value at a realization $\omega \in \Omega$.) We denote the rth moment of $\omega(t)$ by $\mu_r(t)$, when it exists.

We make the following assumptions on the random process ω:

(S1) The random variables $\omega(t)$ are independent.
(S2) The first four moments of the random variables $\omega(t)$ are bounded:

$$\exists \bar{\mu}_r \in \mathbb{R}: \forall t \in \mathbb{N}: \mu_r(t) \leq \bar{\mu}_r, \quad \text{for } r = 1, 2, 3, 4.$$

(S3) Each random variable has mean zero: $\forall t \in \mathbb{N}: \mu_1(t) = 0$.

Assumption S3 is made for convenience only: It allows us to continue working with the same ADE, but it is not essential. Assumption S2 cannot be relaxed to require only that, say, the first two or three moments of the process be bounded. On the other hand, it is possible to generalize the form of the algorithm a bit if *higher*-order moments of the stochastic process are assumed to be bounded (see Remark 6 in section 4). It may be possible to relax somewhat the independence assumption S1 (see Remark 5 in section 4).

A proof of the following theorem is given in section 4:

Convergence Theorem (Stochastic Version): Assume that the algorithm (4′) & (5′) satisfies either

(a) $D = \mathbb{R}^n$ (the Basic Algorithm) and Condition B almost surely, or
(b) D is bounded (the Projection Algorithm) and Condition L.

If x^* is an asymptotically stable stationary point of the algorithm's ADE, and if the stochastic process ω satisfies S1, S2, and S3, then $\{x(t)\}$ converges almost surely to x^* for every $x(0) \in \mathbb{R}^n$.

3. Proof of the Deterministic Convergence Theorem

We emphasize again that most of the constructions in the proof we are about to provide are adapted from Ljung (1975, 1977). In section 4, we will show that the deterministic proof given in this section is essentially the proof for the theorem's stochastic version as well. In particular, the lemmas that appear in the current section as steps in the deterministic proof will appear again in the stochastic proof, where their statements will be exactly the same, except for the addition of the phrase "almost surely" in the right places. Therefore, in order to simplify things when we come to the stochastic proof, and in order to emphasize the essentially deterministic nature of the proof even for the stochastic case, we include in the statements of this section's lemmas the necessary additional phrases "almost surely," but we place them in brackets: [a.s.].

We denote the norm of a point $x \in \mathbb{R}^n$ by $|x|$, and we denote the norm of a matrix A by $\|A\|$, defined in the usual way: $\|A\| = \max_{|z|=1}|Az|$. We will use the notation $B(x, \rho)$ for the open ball about x of radius ρ, i.e., $B(x, \rho) = \{z \in \mathbb{R}^n |\ |z - x| < \rho\}$. We assume throughout that x^* is an asymptotically stable stationary point of the algorithm's ADE, i.e., that $\mathscr{D}(x^*, f)$ is an open set. (For our linear ADE, this is equivalent to global stability, i.e., to $\mathscr{D}(x^*, f) = \mathbb{R}^n$.) We should perhaps note that while Ljung may not seem to assume that x^* is asymptotically stable, he requires (equivalently) that $\mathscr{D}(x^*, f)$ be open when he invokes the "converse" stability theorems to obtain a Lyapunov function.

The key idea in Ljung's approach is that the sequence $\{x(t)\}$ "behaves asymptotically like" the ADE $\dot{x} = f(x)$. More precisely, for any point $\bar{x} \in D^\circ$ and any sufficiently small positive number τ, if $x(t)$ is near $\bar{x}$ and if t is sufficiently large, then

1. a certain number of subsequent terms, say $x(t + 1), \ldots, x(m)$, where m depends upon τ, will also be near $\bar{x}$; and moreover
2. the cumulative movement of the sequence between term t and term m will be approximately $\tau f(\bar{x})$.

Following Ljung, we formally define the number m that appears in (1) and (2) by a function m: $\mathbb{N} \times \mathbb{R}_{++} \to \mathbb{N}$ which, for each $\tau > 0$, satisfies

$$\lim_{t\to\infty} \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} = \tau.$$

Because Ljung does not address the question whether such a function $m(t, \tau)$ exists, we provide an explicit example: Given $(t; \tau)$, if $t + 1 < \frac{1}{\tau}$, then take $m(t, \tau) = t + 1$; and when $t + 1 \geqslant \frac{1}{\tau}$, take $m(t, \tau)$ to be the first index such that

$$\sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} \leqslant \tau \quad \text{and} \quad \sum_{s=t+1}^{m(t,\tau)+1} \frac{1}{s} > \tau.$$

The function is well-defined, because $\sum_{s=t+1}^{\infty} \frac{1}{s} = \infty$, and it has the desired asymptotic property because, when $t + 1 \geqslant \frac{1}{\tau}$, we have

$$0 \leqslant \tau - \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} < \frac{1}{m(t, \tau) + 1}.$$

Note that this particular function m has the additional property, which we will use, that the sum is never larger that τ (when t is large enough).

We can now state the properties (1) and (2) precisely; they will be used throughout the proof. (Henceforth, the symbol τ will denote only strictly positive real numbers.)

Property 1: Let $\bar{x} \in \mathbb{R}^n$ and let $\rho > 0$. A sequence $\{x(t)\}$ is said to have Property 1 for $\bar{x}$ and ρ if $\exists \tau_1 > 0$: $\forall \tau < \tau_1$: $\exists T$: $\forall t > T$: if $x(t) \in B(\bar{x}, \rho)$ then $t \leqslant s \leqslant m(t, \tau) \Rightarrow x(s) \in B(\bar{x}, 2\rho)$.

Property 2: Let $\bar{x} \in \mathbb{R}^n$ and let $\rho > 0$. A sequence $\{x(t)\}$ is said to have Property 2 for $\bar{x}$ and ρ if $\exists \tau_2 > 0$: $\forall \tau < \tau_2$: $\exists T > 0$: $\forall t > T$: if $x(t) \in B(\bar{x}, \rho)$ then $|x(m(t, \tau)) - x(t) - \tau f(\bar{x})| < \|A\| \rho\tau$.

We will develop the proof of the theorem in a series of four lemmas. Lemmas 1 and 2 establish conditions under which the sequence $\{x(t)\}$ will have Properties 1 and 2. Lemma 3 then uses Properties 1 and 2 to establish that at any accumulation point of $\{x(t)\}$ other than the stationary point x^*, any function that is "Lyapunov-like" with respect to the ADE at x^* must satisfy an important auxiliary inequality. Lemma 3 is used to prove Lemma 4, which establishes that the projection operator is never invoked infinitely often (so that the Projection Algorithm eventually behaves like the Basic Algorithm), and then Lemma 3 is used again to complete the proof by establishing that every subsequence generated by the Basic Algorithm must converge to x^*.

Lemmas 1, 2, and 3 provide two alternative conditions under which their conclusions hold. This may seem odd, especially since the Basic Algorithm clearly satisfies both conditions. The conditions are quite different, however, for the Projection Algorithm. One of the conditions (essentially, that we are not examining boundary points of D) is appropriate for applying the lemmas when it is not known (as it will not be until Lemma 4) whether $\tilde{x}(t)$ must eventually lie in D. The other condition (that $\tilde{x}(t)$ is indeed eventually in D) allows us, once we have obtained Lemma 4, to use Lemmas 1, 2, and 3 to examine boundary points of D that may occur as accumulation points of $\{x(t)\}$.

Lemma 1: Let $\bar{x} \in \mathbb{R}^n$ and $\rho > 0$. If $B(\bar{x}, 2\rho) \subseteq D$ (and a fortiori $\bar{x} \in D^\circ$), or if $\tilde{x}(t)$ is [a.s.] eventually in D, then [a.s.] $\{x(t)\}$ has Property 1 for $\bar{x}$ and ρ.

Proof. If τ is small enough and t large enough, then for each k that satisfies $t + 1 \leqslant k \leqslant m(t, \tau)$ we will show that if $x(s) \in B(\bar{x}, 2\rho)$ for $s = t, t + 1, \ldots, k - 1$, then also $x(k) \in B(\bar{x}, 2\rho)$. Property 1 assumes that we already have $x(t) \in B(\bar{x}, \rho) \subseteq B(\bar{x}, 2\rho)$; therefore this recursive argument will establish that $x(t + 1), x(t + 2), \ldots, x(m) \in B(\bar{x}, 2\rho)$, as required. Thus, assume that

$$t \leqslant s \leqslant k - 1 \Rightarrow x(s) \in B(\bar{x}, 2\rho); \tag{6}$$

we will show that $\tilde{x}(k) \in B(\bar{x}, 2\rho)$, from which it follows that $x(k) = \tilde{x}(k) \in B(\bar{x}, 2\rho)$.

If $\bar{x} \in D$ and $B(\bar{x}, 2\rho) \subseteq D$, then (6) implies that the projection operator is never invoked during the periods t to $k - 1$. The lemma's alternative assumption is that $\tilde{x}(t)$ is eventually in D, in which case, we let t be large enough that $\tilde{x}(t') \in D$ for $t' \geqslant t$, and again the projection operator is never invoked during the periods t to $k - 1$. In either case, then, we have $x(s) = \tilde{x}(s) \in D$ for $t \leqslant s \leqslant k - 1$. Therefore,

$$\begin{aligned} \tilde{x}(k) &= x(k-1) + \frac{1}{k}[Gq(k) + Hx(k-1)] \\ &= x(t) + \sum_{s=t+1}^{k} \frac{1}{s}[Gq(s) + Hx(s-1)] \\ &= x(t) + \sum_{s=t+1}^{k} \frac{1}{s}[A\, x(s-1)], \end{aligned} \tag{7}$$

and thus

$$\begin{aligned} |\tilde{x}(k) - x(t)| &\leqslant \sum_{s=t+1}^{k} \frac{1}{s}\|A\|\, |x(s-1)| \\ &\leqslant \|A\| \max_{t<s\leqslant k} |x(s-1)| \sum_{s=t+1}^{k} \frac{1}{s} < \|A\|\,(2\rho + |\bar{x}|) \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s}, \end{aligned}$$

where the last inequality follows from the assumption that $x(s) \in B(\bar{x}, 2\rho)$ whenever $t \leqslant s < k$. Define τ_1 as follows:

$$\tau_1 = \frac{\rho}{2\|A\|\,(2\rho + |\bar{x}|)}, \tag{8}$$

and given $\tau < \tau_1$, let T be such that $\forall t > T$

$$\left| \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} - \tau \right| < \tau. \tag{9}$$

Then we have

$$|\tilde{x}(k) - x(t)| < \|A\|\,(2\rho + |\bar{x}|)2\tau < \rho,$$

and therefore

$$|\tilde{x}(k) - \bar{x}| = |\tilde{x}(k) - x(t) + x(t) - \bar{x}| \leqslant |\tilde{x}(k) - x(t)| + |x(t) - x| < 2\rho,$$

completing the proof of Lemma 1. ■

Lemma 2: Let $\bar{x} \in \mathbb{R}^n$ and $\rho > 0$. If $B(\bar{x}, 2\rho) \subseteq D$ (and a fortiori, $\bar{x} \in D^\circ$), or if $\tilde{x}(t)$ is [a.s.] eventually in D, then [a.s.] $\{x(t)\}$ has Property 2 for $\bar{x}$ and ρ.

Proof. Lemma 1 ensures that for any sufficiently small τ we can choose t large enough that we will have

$$x(s) = x(s-1) + \frac{1}{s}[Gq(s) + Hx(s-1)]$$

for all s satisfying $t \leq s \leq m(t, \tau)$. We therefore have

$$\begin{aligned} x(m(t,\tau)) &= x(t) + \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s}[Gq(s) + Hx(s-1)] \\ &= x(t) + \tau f(\bar{x}) + f(\bar{x})\left[\sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} - \tau\right] \\ &\quad + \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s}[Gq(s) + Hx(s-1) - f(\bar{x})] \\ &= x(t) + \tau f(\bar{x}) + f(\bar{x})\left[\sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} - \tau\right] \\ &\quad + A\sum_{s=t+1}^{m(t,\tau)} \frac{1}{s}(x(s-1) - \bar{x}). \end{aligned} \tag{10}$$

Denote the third and fourth terms of (10) by z_1 and z_2. According to the definition of the function $m(\cdot, \cdot)$, for any ρ and τ we can choose t large enough that $|z_1| < \frac{1}{2}\|A\|\rho\tau$. We also have

$$|z_2| \leq \|A\| \max_{t \leq s < m(t,\tau)} |x(s) - \bar{x}| \sum_{s=t+1}^{m(t,\tau)} \frac{1}{s} \leq \tau\|A\| \max_{t \leq s < m(t,\tau)} |x(s) - \bar{x}|.$$

Lemma 1 ensures that for each ρ there is a τ_1 such that

$$\forall \tau < \tau_1 \colon \exists T \colon t > T \Rightarrow |x(s) - \bar{x}| < \frac{\rho}{2},$$

and therefore, for such τ and t, we have

$$|z_1| + |z_2| < \frac{1}{2}\|A\|\rho\tau + \tau\|A\|\frac{1}{2}\rho = \|A\|\rho\tau. \quad \blacksquare \tag{11}$$

Lemma 3: Let $V\colon \mathbb{R}^n \to \mathbb{R}_+$ be a twice continuously differentiable function and let $\bar{x} \in \mathbb{R}^n$ be a point at which $V'(\bar{x})f(\bar{x}) < 0$. If $\bar{x} \in D^\circ$, or if $\bar{x}(t)$ is [a.s.] eventually in D, then [a.s.] if $\{x(t_k)\}$ is a subsequence of $\{x(t)\}$ that converges to $\bar{x}$, then $\exists \delta > 0$, $\tau_0 > 0$: $\forall \tau < \tau_0$: $\exists K$: $\forall k > K$: $V(x(m(t_k, \tau))) < V(\bar{x}) - \delta\tau$.

Proof. Let the function V, the point $\bar{x}$, and the subsequence $\{x(t_k)\}$ be as described in the lemma, and let $\delta = -\frac{1}{2}V'(\bar{x})f(\bar{x})$. For any k and any positive number τ, two succesive applications of the Mean Value Theorem

yield the following, where ζ is a convex combination of $x(m(t_k, \tau))$ and $x(t_k)$, and ξ is a convex combination of ζ and $\bar{x}$:

$$\begin{aligned} V(x(m(t_k,\tau))) - V(x(t_k)) &= V'(\zeta)[x(m(t_k,\tau)) - x(t_k)] \\ &= V'(\bar{x})[x(m(t_k,\tau)) - x(t_k)] \\ &\quad + (\zeta - \bar{x})^T V''(\xi)[x(m(t_k,\tau) - x(t_k)]. \\ &= V'(\bar{x})\tau f(\bar{x}) + V'(\bar{x})[x(m(t_k,\tau)) - x(t_k) - \tau f(x)] \\ &\quad + (\zeta - \bar{x})^T V''(\xi)[x(m(t_k,\tau) - x(t_k)]. \\ &= -2\delta\tau + R(t_k, \tau, \bar{x}), \end{aligned} \tag{12}$$

where $R(t_k, \tau, \bar{x})$ is given by

$$\begin{aligned} R(t_k,\tau,\bar{x}) &= V'(\bar{x})[x(m(t_k,\tau)) - x(t_k) - \tau f(x)] \\ &\quad + (\zeta - \bar{x})^T V''(\xi)[x(m(t_k,\tau)) - x(t_k)]. \end{aligned} \tag{13}$$

Because $\{x(t_k)\} \to \bar{x}$, there is a number k such that

$$k > K \Rightarrow V(x(t_k)) < V(\bar{x}) + \frac{\delta\tau}{2};$$

combining this with (12) yields $V(x(m(t_k, \tau)) < V(\bar{x}) + \frac{\delta\tau}{2} - 2\delta\tau + R(t_k, \tau, \bar{x})$ for all $k > K$. In order to complete the proof, we therefore need only show that if τ is chosen small enough, then k can be chosen sufficiently large that

$$k > K \Rightarrow |R(t_k, \tau, \bar{x})| < \frac{\delta\tau}{2}. \tag{14}$$

In order to establish (14), we will need to apply Lemmas 1 and 2 to $\bar{x}$ and a carefully selected $\rho > 0$. Assume that $\rho > 0$, and assume either that $\bar{x}(t)$ is eventually in D, or else that $\bar{x} \in D^\circ$ and that ρ is no greater than half the distance from $\bar{x}$ to the boundary of D—i.e., that $B(\bar{x}, 2\rho) \subseteq D$. Then Lemmas 1 and 2 guarantee that there are a $\tau_0(\rho) = \min\{\tau_1(\rho), \tau_2(\rho)\}$ and, for each $\tau < \tau_0(\rho)$, a $K(\tau, \rho) = \max\{K_1(\tau, \rho), K_2(\tau, \rho)\}$ such that $\forall k > K(\tau, \rho)$:

$$\text{If} \quad x(t_k) \in B(\bar{x}, \rho) \quad \text{and} \quad t_k \leq s \leq m(t_k, \tau),$$

$$\text{then} \quad |x(s) - \bar{x}| < 2\rho \quad \text{and} \quad |x(m(t_k,\tau)) - x(t_k) - \tau f(\bar{x})| < \|A\|\rho\tau.$$

And because $\{x(t_k)\}$ converges to $\bar{x}$, $K(\tau, \rho)$ can be chosen large enough that indeed $x(t_k) \in \mathrm{B}(\bar{x}, \rho)$ for every $k > K(\tau, \rho)$; therefore $\forall k > K(\tau, \rho)$:

$$t_k \leq s \leq m(t_k, \tau) \Rightarrow [|x(s) - \bar{x}| < 2\rho$$

$$\text{and} \quad |x(m(t_k,\tau)) - x(t_k) - \tau f(\bar{x})| < \|A\|\rho\tau].$$

Let $c(\rho)$ denote the number $\sup\{\|V''(x)\| \mid x \in B(\bar{x}, 2\rho)\}$. For $\tau < \tau_0(\rho)$ and $k > K(\tau, \rho)$, the ζ and ξ in (12) and (13) both lie in $B(\bar{x}, 2\rho)$, and therefore $|(\zeta - \bar{x})^T V''(\xi)| < 2c(\rho)\rho$. Substituting this inequality into (13), we obtain for every $\rho > 0$ the following: $\exists\tau_0(\rho)$: $\forall\tau < \tau_0(\rho)$: $\exists K(\tau, \rho)$: $\forall k > K(\tau, \rho)$:

$$\begin{aligned}|R(t_k, \tau, \bar{x})| &< |V'(\bar{x})|\,|x(m(t_k, \tau) - x(t_k) - \tau f(\bar{x})| + 2c(\rho)\rho|\,x(m(t_k, \tau)) - x(t_k)| \\ &\leqslant (|V'(\bar{x})| + 2c(\rho)\rho)|x(m(t_k, \tau)) - x(t_k) - \tau f(\bar{x})| + 2c(\rho)\rho\tau|f(\bar{x})| \\ &< (|V'(\bar{x})| + 2c(\rho)\rho)\,\|A\|\,\rho\tau + 2c(\rho)|f(\bar{x})|\rho\tau \\ &= [|V'(\bar{x})|\ \|A\| + 2c(\rho)(\rho)\ (\rho\|A\| + |f(\bar{x})|)]\rho\tau.\end{aligned}$$

Now if we choose ρ small enough that $[|V'(\bar{x})|\ \|A\| + 2c(\rho)(\rho\|A\| + |f(\bar{x})|)]\rho < \frac{\delta}{2}$, and we choose τ smaller than $\tau_0(\rho)$, and we choose K at least as large as $K(\tau, \rho)$, then (14) will be satisfied and the proof completed. ■

Lemma 4: If the algorithm satisfies Condition L, then the projection operator is [a.s.] invoked only finitely many times; i.e., $\tilde{x}(t)$ is [a.s.] eventually in D.

Proof. The lemma is, of course, true by definition for the Basic Algorithm (with or without Condition L), so we will assume throughout this proof that we are working with the Projection Algorithm, i.e., that D is bounded. Let e_1 and e_2 be real numbers that satisfy $c_1 < e_1 < e_2 < c_2$, where c_1 and c_2 are the numbers specified in Condition L; let I denote the closed interval $[e_1, e_2]$; and suppose that the projection operator is invoked infinitely many times. We will show that then there are infinitely many terms of $\{U(x(t))\}$ on each side of I, i.e., that

$$U(x(t)) < e_1 \quad \text{infinitely often, and} \tag{15}$$

$$U(x(t)) > e_2 \quad \text{infinitely often,} \tag{16}$$

and that this leads to a contradiction. To establish (15) and (16), note that every time the projection operator is invoked we have a t for which $\tilde{x}(t) \notin D$ and $x(t) \in C$—and therefore, according to Condition L, we also have $U(\tilde{x}(t)) \geqslant c_2$ and $U(x(t)) \leqslant c_1$. Thus, if the projection operator is invoked infinitely often, (15) is obvious; and (16) follows from the continuity of U, together with the fact that the "step size" between the forecast and the subsequent unprojected forecast goes to zero:

$$|\tilde{x}(t) - x(t-1)| = \frac{1}{t}|q(t) - x(t-1)| \tag{17}$$

$$= \frac{1}{t}|A(x(t-1)| \leqslant \frac{1}{t}\,\|A\| \sup_{x \in D}|x|.$$

Now form a subsequence of $\{x(t)\}$, denoted $\{x(t_k)\}$, by taking *every* term of $\{x(t)\}$ that satisfies *both* $U(x(t)) \geqslant e_1$ and $U(x(t-1)) < e_1$. (The existence of such a subsequence is guaranteed by (15) and (16).) We have

$$\lim_{k \to \infty} u(x(t_k)) = e_1, \tag{18}$$

for if any of the terms were to satisfy $\tilde{x}(t_k) \notin D$, so that $x(t_k) \in C$, then we would have $U(x(t_k)) \leqslant c_1 < e_1$, which violates the definition of the subsequence. Therefore, every term satisfies $x(t_k) = \tilde{x}(t_k) \in D$, and (18) is then a direct consequence of (17) and the definition of the subsequence. The important consequences of (18) are that (a) any accumulation point of $\{x(t_k)\}$ must be in D°, and (b) the sequence $\{U(x(t_k))\}$ eventually remains inside the interval I, and therefore the sequence $\{x(t_k)\}$ eventually remains inside the set $D \setminus C$.

For each index t, let t^* denote the first integer $s \geqslant t$ for which $U(x(s)) \notin I$, i.e., for which $U(x(s)) < e_1$ or $U(x(s)) > e_2$. Note that t^* is well-defined, according to either (15) or (16). And note that because $\{U(x(t_k))\}$ eventually remains inside the interval I, none of the terms t_k^* coincides with any of the terms t_k. Now form a subsequence of $\{x(t_k)\}$, denoted $\{x(t_k')\}$, by choosing just those terms t_k for which $U(x(t_k^*)) > e_2$, omitting the terms for which $U(x(t_k^*)) < e_1$. Because each term $x(t_k)$, and *a fortiori* each term $x(t_k')$, is in $D \setminus C$, and because D is bounded, $\{x(t_k')\}$ has a subsequence, say $\{x(t_k'')\}$, that converges; denote its limit by $\bar{x}$, and note that (18) yields $U(\bar{x}) = e_1$. Thus, we *also* have $\bar{x} \in D^\circ$. Because $c_1 < U(\bar{x}) < c_2$, Condition L implies that $\bar{x} \in D \setminus C$. Let ρ be small enough that

$$x \in B(\bar{x}, 2\rho) \Rightarrow U(x) < e_2, \tag{19}$$

which also ensures that $B(\bar{x}, 2\rho) \subset D$. Lemma 1 then guarantees that $\{x(t)\}$ has Property 1 for $\bar{x}$ and ρ; therefore, let τ be small enough and K large enough that if $k \geqslant K$ and $x(t_k) \in B(\bar{x}, \rho)$, then each term $x(t_k)$, $x(t_k + 1), \ldots, x(m(t_k, \tau))$ lies in $B(\bar{x}, 2\rho)$; and since $\{x(t_k'')\}$ converges to $\bar{x}$, choose K large enough that in fact $x(t_k'') \in B(\bar{x}, \rho)$ for each $k \geqslant K$. Consequently, for each $k \geqslant K$ and for each s between t_k'' and $m(t_k'', \tau)$, we have $x(s) \in B(\bar{x}, 2\rho)$, and therefore, according to (19),

$$\text{If} \quad k \geqslant K \quad \text{and} \quad t_k'' \leqslant s \leqslant m(t_k'', \tau), \quad \text{then} \quad U(x(s)) < e_2. \tag{20}$$

But $\bar{x} \in D^\circ \setminus C$, and Condition L therefore implies that $U'(\bar{x})f(\bar{x}) < 0$, and Lemma 3 therefore yields $U(x(m(t_k'', \tau))) < e_1$, which implies that

$U(x(m(t''_k, \tau))) \notin I$. Clearly, then, t''^*_k is between t''_k and $m(t''_k, \tau)$, and (20) therefore yields $U(x(t''^*_k)) < e_2$, which is inconsistent with the definition of the subsequence $\{x(t'_k)\}$ and its subsequence $\{x(t''_k)\}$, thereby completing the lemma's proof. ∎

Remark 1: It is of fundamental importance, as the proof of Lemma 4 makes clear, that the target set C in the Projection Algorithm be a closed subset of *the interior* of D. More specifically, it is critical (for the *known* proof that the Projection Algorithm converge, at any rate) that the projection sequence $\{p(t)\}$ be bounded away from the boundary of D. If this is not the case, then (15) cannot be guaranteed; therefore, the limit points, such as $\bar{x}$, which result from invoking $p(t)$ infinitely often might *all* lie in the boundary of D; and, therefore, Lemmas 1 and 3 cannot be applied to them. For example, Marcet and Sargent (1989b) and Moreno and Walker define "decentralized" projection algorithms, in which each participant i projects $\tilde{x}_i(t)$ to $p_i(t) \in C_i$ independently (i.e., only when his *own* estimate $x_i(t)$ lies outside his own set D_i), but there is no guarantee that in these algorithms the *realized* projections, say $p^*(t)$, will indeed be bounded away from the exterior of D, and therefore we seem to have no assurance that such algorithms will converge to a stationary point (for more on this, see Moreno and Walker, 1991). A generalization of Ljung's Second Theorem to include such "decentralized" projection algorithms could be achieved if one could devise a proof of Lemma 4 that does not require the existence of this limit point $\bar{x}$ in the interior of D.

Remark 2: The remainder of the proof of the Convergence Theorem will rely on the application of Lemma 3 to a Lyapunov function at subsequential limit points (accumulation points) of the sequence $\{x(t)\}$. When D is bounded (i.e., when we are analyzing the Projection Algorithm) one or more of these limit points might lie in the boundary of D, and we would not be able to use Lemma 3 to analyze these boundary limit points if it were not for Lemma 4. Ljung's proof does not take account of this fact: He establishes analogues of Lemmas 1, 2, and 3 only for the case in which $\tilde{x}(t)$ is always (or eventually) in D, but not for the case (required in order to obtain Lemma 4) in which one does not yet know whether $\{\tilde{x}(t)\}$ eventually remains in D, but one does know that $x \in D$.

Because x^* is an asymptotically stable stationary point of the ADE, there is a twice continuously differentiable Lyapunov function W: $\mathbb{R}^n \to \mathbb{R}_+$ that satisfies $W(x) > 0$ and $W'(x)f(x) < 0$ for all x in $\mathbb{R}^n$ except x^*, and $W(x^*) = W'(x^*)f(x^*) = 0$. Because f is linear, W can be taken to be quadratic (see, e.g., Hahn, 1967, p. 117). Let X denote the set of subsequential limits of $\{x(t)\}$, and let $W(X)$ denote the image of X under the function W. The set X is clearly closed. It is also nonempty: For the Projection Algorithm, this follows from the boundedness of D; for the Basic Algorithm, it is an immediate consequence of Condition B (this is the only use of Condition B in the proof). Let $\underline{w}$ and $\bar{w}$ denote $\inf W(X)$ and $\sup W(X)$. We will prove that $\underline{w} = 0$ and then that $\bar{w} = \underline{w}$. Thus, $\bar{w} = 0$, that is $\{x(t)\}$ converges to x^*. The step $\underline{w} = 0$ is the difficult one; $\bar{w} = \underline{w}$ will require only a repetition of the proof of Lemma 4.

Remark 3: For the Projection Algorithm, in which D is bounded, it is clear that X is nonempty and bounded, and therefore that X is a nonempty compact set. Hence $W(X)$ is compact as well, and in particular, $\underline{w} \in W(X)$. Now if one assumes (in order to obtain a contradiction) that $\underline{w} > 0$, the fact that $\underline{w} \in W(X)$ makes it easy to apply Lemma 3 to obtain the contradiction. Ljung makes it appear equally easy to show that $\underline{w} = 0$ for the Basic Algorithm: he states (1977, p. 568) that X $\cap$ $\mathscr{D}(x^*, f)$ is compact. However, there seems to be no justification for this statement when D is unbounded, as in the Basic Algorithm. For more on this problem, see Remark 4, which follows the completion of the theorem's proof.

Completion of the Theorem's Proof

(*a*) *Proof that* $\underline{w} = 0$:

For each $w \in \mathbb{R}_+$, let $L(w)$ denote the W-lower-contour set of w, i.e., $L(w) = \{x \in \mathbb{R}^n \mid W(x) \leqslant w\}$. For every $w \geqslant 0$, $L(w)$ is nonempty (because $W(x^*) = 0$), closed (because W is continuous), and bounded (because W is quadratic). In other words, each set $L(w)$ is a nonempty compact set.

According to the definition of $\underline{w}$, for every positive integer n there is a point $\bar{x}(n)$ which satisfies $W(\bar{x}(n)) \leqslant \underline{w} + 1/n$ and which is also a limit point of $\{x(t)\}$—say, $\bar{x}(n) = \lim_{j\to\infty} x(t(j, n))$ for a subsequence $\{x(t(\cdot, n))\}$ of $\{x(t)\}$. The sequence $\{\bar{x}(n)\}$ of limit points is therefore bounded, because each of its terms lies in the lower-contour set $L(W(\bar{x}(1)))$. Therefore, the sequence $\{\bar{x}(n)\}$ has a convergent subsequence—say $\lim_{k\to\infty}\bar{x}(n_k)$ converges to $\bar{x}$. And because X is closed, $\bar{x} \in X$; that is there is a subsequence $\{x(t_k)\}$ of $\{x(t)\}$ that converges to $\bar{x}$, and $W(\bar{x}) = \underline{w}$.

Because $\{x(t_k)\}$ converges to $\bar{x}$, we can apply Lemma 3 to the function W, the subsequence $\{x(t_k)\}$, and its limit $\bar{x}$: For some positive

δ and all sufficiently small values of τ, each term of the sequence $\{x(m(t_k, \tau))\}$ satisfies

$$W(x(m(t_k, \tau))) < W(\bar{x}) - \delta\tau < \underline{w}. \tag{21}$$

(Note that, for the Projection Algorithm, we must appeal to Lemma 4 in order to apply Lemma 3.) The sequence $\{x(t_k)\}$ is bounded, and therefore (according to Lemma 1) the sequence $\{x(m(t_k, \tau))\}$ is bounded as well and has a convergent subsequence, which is also a subsequence of $\{x(t)\}$. Write it as $\{x(t'_k)\}$, with, say, $\lim_{k\to\infty} x(t'_k) = x'$. Clearly, (21) implies that $W(x') < W(\bar{x}) = \underline{w}$; but this is inconsistent with the definition of $\underline{w}$ as $\inf W(X)$, and this contradiction establishes that $\underline{w} = 0$.

(*b*) *Proof that* $\bar{w} = \underline{w}$:

Suppose, by way of contradiction, that $\underline{w} < \bar{w}$. Let e_1 and e_2 be real numbers that satisfy $\underline{w} < e_1 < e_2 < \bar{w}$, and let $I = [e_1, e_2]$. Clearly, the sequence $\{W(x(t))\}$ is infinitely often on each side of I, just as in the proof of Lemma 4. The argument used there applies here as well, with the exception of one detail that does not carry over immediately: In the proof of Lemma 4, the conclusion (14) depends upon the *supremum* in (17) being taken over a bounded set (the proof of Lemma 4 concerned only the Projection Algorithm, so we could use the set D there). Here we must establish that for the Basic Algorithm the *supremum* in (17) can similarly be taken over a bounded set. When, as immediately following (17), we form the subsequence $\{x(t_k)\}$, in which, for each k, $W(x(t_k - 1)) < e_1$ and $W(x(t_k)) > e_1$ (the existence of such a subsequence is immediate here, because (15) and (16) are immediate), we only require that (17) hold for the terms of the subsequence, and therefore we now take the *supremum* in (17) over the W-lower-contour set $L(e_1)$, which is bounded because W is quadratic. The remainder of the proof of Lemma 4 applies here without change, yielding a contradiction, and thereby completing the proof. ■

Remark 4: We noted in the preceding remark that Ljung's proof for the Basic Algorithm is incomplete, because he assumes without justification that $X \cap \mathscr{D}(x^*, f)$ is compact. We have established that $\underline{w} \in W(X)$, and also that $\bar{w} = \underline{w}$, by appealing instead to the boundedness of the lower-contour sets of the quadratic Lyapunov function W, and we know that W can be taken to be quadratic because the ADE for our algorithm is linear. It is not clear to us how one can construct a proof without a Lyapunov function in which the lower-contour sets are bounded.

Ljung, too, relies on the boundedness of the Lyapunov function's lower-contour sets. (Our proof that $\bar{w} = \underline{w}$ is essentially the same as Ljung's proof.) But this assumptions on the Lyapunov function is not justified when f is not linear: Barbashin and Krasovskii (see Hahn, 1967, p. 109) provide an example of an asymptotically stable differential equation $\dot{x} = f(x)$ for which no Lyapunov function with bounded lower-contour sets exists. Of course, this problem can be completely circumvented by strengthening Condition B so as to specify that $\{x(t)\}$ is itself a bounded sequence. And that may be a reasonable approach: In applications, it may be no more difficult to establish that each $\{x(t)\}$ is bounded than to establish that each one has a bounded subsequence (see, for example, Moreno and Walker, 1991).

4. Proof of the Stochastic Convergence Theorem

The key result in moving from the deterministic theorem to the stochastic theorem is the following lemma, which establishes that, with probability one, the stochastic terms that appear in the lemmas' proofs will all vanish asymptotically. With this lemma in hand, the stochastic proof is virtually identical to the deterministic proof.

Lemma 5: If Assumptions S1, S2, and S3 are satisfied, then $\forall\tau, \varepsilon > 0$: for almost every $\omega \in \Omega$: $\exists T(\tau, \varepsilon, \omega)$: $\forall t > T$: if $t \leqslant k \leqslant m(t_k, \tau)$ then

$$\left|\sum_{s=t+1}^{k} \frac{\omega(s)}{s}\right| < \varepsilon.$$

Proof. We must show that for all τ and ε,

$$P\left\{\text{For infinitely many } t\text{: } \exists k\text{: } t \leqslant k \leqslant m(t_k, \tau) \text{ and } \left|\sum_{s=t+1}^{k} \frac{\omega(s)}{s}\right| > \varepsilon\right\} = 0.$$

Thus, according to the Borel–Cantelli Lemma, it will be sufficient to show that

$$\sum_{t=1}^{\infty} P\left\{\omega \,\middle|\, \max_{t+1\leqslant k\leqslant m} \left|\sum_{s=t+1}^{k} \frac{\omega(s)}{s}\right| > \varepsilon\right\} < \infty. \tag{22}$$

For each t, let $K(t, \tau)$ denote the index k (between $t + 1$ and $m(t, \tau)$) for which

$$\left|\sum_{s=t+1}^{k} \frac{\omega(s)}{s}\right|$$

is maximized. Chebyshev's inequality yields

$$P\left\{\omega \left| \left| \sum_{s=t+1}^{K} \frac{\omega(s)}{s} \right| > \varepsilon \right.\right\} \leqslant \frac{1}{\varepsilon^4} E \left| \sum_{s=t+1}^{K} \frac{\omega(s)}{s} \right|^4, \tag{23}$$

and straightforward calculation (see Remark 5, immediately following) yields

$$\begin{aligned} E\left[\sum_{s=t+1}^{K} \frac{\omega(s)}{s}\right]^4 &= \sum_{t+1}^{K} E\left(\frac{\omega(s)}{s}\right)^4 + 6 \sum_{\substack{t+1 \\ s<s'}}^{K} E\left(\frac{\omega(s)}{s}\right)^2 \left(\frac{\omega(s')}{s'}\right)^2 \\ &\leqslant (\bar{\mu}_4)\left(\sum_{t+1}^{K} \frac{1}{s^4} + 6 \sum_{\substack{t+1 \\ s<s'}}^{K} \frac{1}{s^4}\right) \\ &\leqslant \bar{\mu}_4\left(\frac{m-t}{t^4} + 6\left(\frac{1}{2}\right)\frac{1}{t^4}(m-t)(m-t-1)\right) \\ &< \bar{\mu}_4\left(\frac{1}{t^3} + \frac{3}{t^2}\right) \end{aligned} \tag{24}$$

where the last inequality follows from the fact that $m(t, t) - \tau < t$, which can be easily shown as follows: $\sum_{t+1}^{m} \frac{1}{s} \leqslant \tau$, by definition of $m(t, \tau)$; and $\sum_{t+1}^{m} \frac{1}{s} \geqslant \frac{m-t}{m}$; combining these two inequalities yields $m - t \leqslant m\tau$ or equivalently $m \leqslant \frac{t}{1-\tau}$, or $m - t \leqslant \frac{\tau}{1-\tau} t < t$ when $\tau < \frac{1}{2}$. Combining (23 and (24), it is clear that (22) is satisfied. ■

Remark 5: Writing ξ_s for $\omega(s)/s$, and making use of the independence assumption S1, the first step in (24) is as follows:

$$\begin{aligned} E\left(\sum_{s=t+1}^{K} \xi_s\right)^4 &= \sum \frac{4!}{4!} E(\xi_s)^4 + \sum_{r \neq s} \frac{4!}{3!1!} E(\xi_r)^3 E(\xi_s) + \sum_{r<s} \frac{4!}{2!2!} E(\xi_r)^2 E(\xi_s)^2 \\ &+ \sum_{r<s<u} \frac{4!}{2!1!1!} E(\xi_r)^2 E(\xi_s) E(\xi_u) \\ &+ \sum_{r<s<u<v} \frac{4!}{1!1!1!1!} E(\xi_r) E(\xi_s) E(\xi_u) E(\xi_v). \end{aligned}$$

It follows from the zero-mean assumption S3 that the second, fourth, and fifth terms on the right-hand side are all zero. (If S3 were not assumed, the ADE would be altered, and instead of $\omega(t)$, we would always be working with $\omega(t) - \mu_1(t)$, that is, with the difference between $\omega(t)$ and its mean, which behaves as ω does in the zero-mean case.) This remark is the only use of the independence assumption S1 in the proof of the Stochastic Convergence Theorem. Thus, it may be possible to relax S1 in a way that does not allow the relevant sums to grow at a rate faster than t^2.

Remark 6: The proof of Lemma 5 that we have just given is essentially (for the case $\gamma(t) = 1/t$) the proof in Ljung (1975), except that (a) because of his weaker assumptions on $\gamma(t)$, Ljung's proof is much longer and more complex, and (b) Ljung's proof is not completely correct. It is also worth noting, as Ljung points out (1975, pp. 11 and 55), that there is a trade-off between the restrictiveness of the assumptions on the asymptotic behavior of $\gamma(t)$ on the one hand, and, on the other hand, the number of moments of the random variables $\omega(t)$ that are assumed to be bounded in t.

Remark 7: For a more compact notation, let $S(t, k, \omega)$ denote the sum $\sum_{t+1}^{K} \frac{\omega(s)}{s}$, and let $E^*(\tau, \varepsilon)$ denote the set $\{\omega \in \Omega | \exists T: [t \geqslant T \;\&\; t \leqslant k \leqslant m(t, \tau)] \Rightarrow |S(t, k, \omega)| < \varepsilon\}$. Then Lemma 5 states that for every positive τ and ε, $P(E^*(\tau, \varepsilon)) = 1$. The set-valued function $E^*(\tau, \varepsilon)$ is decreasing in τ and increasing in ε: $(\tau < \tau' \;\&\; \varepsilon < \varepsilon') \Rightarrow E^*(\tau', \varepsilon) \subseteq E^*(\tau, \varepsilon) \subseteq E^*(\tau, \varepsilon')$.

Let $\tilde{E}$ denote the set of ω for which the sequence $\{\tilde{x}(t, \omega)\}$ is eventually in D. For each $x \in \mathbb{R}^n$ and each $\rho > 0$, let $E_1(x, \rho)$ and $E_2(x, \rho)$ be the sets of realizations ω for which $\{x(t, \omega)\}$ has Property 1 and Property 2, respectively. For any $\bar{x}$ and V as in Lemma 3, let $E_3(\bar{x}, V)$ be the set of ω for which the conclusion of Lemma 3 is valid. The stochastic versions of Lemmas 1, 2, and 3 state that for any $\bar{x}$, ρ, and V, the events E_1, E_2, and E_3 all have probability one (under the conditions assumed in the lemmas); Lemma 4 establishes that $\tilde{E}$ has probability one.

Proof of Lemma 1 (Stochastic version)

Let $\omega \in \Omega$, and assume either that $x \in D$ and $B(x, 2\rho) \subseteq D$, or else that $\omega \in \tilde{E}$ and $P(\tilde{E}) = 1$. The only changes in the deterministic proof are as follows: The right-hand side (RHS) of (7) has the additional term $S(t, k, \omega)$, and each of the expressions following (7) has the additional term $|S(t, k, \omega)|$. Define $\tau_1(\rho)$ to be, say, half the value that (8) assigns to it in the deterministic proof; let $\omega \in E^*\left(\tau_1(\rho), \frac{\rho}{2}\right)$; and (invoking Remark 7) for each $\tau < \tau_1(\rho)$, let $T(\tau, \rho)$ be such that

$$[t > T(\tau, \rho) \ \&\ t \leqslant k \leqslant m(t, \tau)] \Rightarrow \left[(9) \text{ and } |S(t, k, \omega)| < \frac{\rho}{2}\right].$$

Then we have $|\tilde{x}(k, \omega) - x(t, \omega)| < \frac{\rho}{2} + \frac{\rho}{2} = \rho$ and $|\tilde{x}(k, \omega) - x| < 2\rho$— either for all ω in $E^*\left(\tau_1(\rho), \frac{\rho}{2}\right)$ or for all ω in $\tilde{E} \cap E^*\left(\tau_1\left(\rho, \frac{\rho}{2}\right)\right)$, i.e., with probability one. ■

Proof of Lemma 2 (Stochastic version)

It is shown that for every x and ρ: (a) $E_1(x, \rho) \cap \tilde{E} \subseteq E_2(x, \rho) \cap \tilde{E}$; and (b) if $B(x, 2\rho) \subseteq$ D, then $E_1(x, \rho) \subseteq E_2(x, \rho)$. It follows, then, from Lemma 1 that $P(E_2(x, \rho)) = 1$ if either $B(x, 2\rho) \subseteq$ D or $P(\tilde{E}) = 1$. The only changes in the deterministic proof are as follows: The expression in (10) has a fifth term, $S(t, m(t, \tau), \omega)$, the absolute value of which can, for any $\omega\varepsilon E_1$, be made smaller than $\rho\tau/2$ by choosing t large enough; $|z_1|$ can be made smaller than $\rho\tau/2$; and the inequality in (11) holds for $\tau < \tau_1(\rho)$. ■

Proof of Lemma 3 (Stochastic version)

It is shown that for any $\bar{x}$ and V as in the lemma, if ρ is small enough, then (a) $E_2(\bar{x}, \rho) \cap \tilde{E} \subseteq E_3(\bar{x}, V) \cap \tilde{E}$; and (b) If $\bar{x} \in D°$, then $E_2(\bar{x}, \rho) \subseteq E_3(\bar{x}, V)$. It follows, then, from Lemma 2 that $P(E_3(\bar{x}, V)) = 1$ if either $\bar{x} \in D°$ or $P(\tilde{E}) = 1$. The only change in the deterministic proof is that now we must show that, *with probability one*, τ can be chosen small enough and K large enough that (14) holds. Consequently, after taking an arbitrary positive ρ, we assume that $\omega \in E_2(\bar{x}, \rho)$ and either that $B(\bar{x}, 2\rho) \subseteq$ D, or else that $\omega \in \tilde{E}$; the proof is otherwise unchanged. ■

Proof of Lemma 4 (Stochastic version)

The only change in the deterministic proof is that (17) becomes

$$\frac{1}{t}|Ax(t-1) + \omega(t)| \leqslant \frac{1}{t}\|A\| \sup_{x \in D}|x| + \frac{1}{t}|\omega(t)|. \qquad (17')$$

Exactly as in the proof of Lemma 5, we appeal to Chebyshev's Inequality and the Borel–Cantelli Lemma to verify that, for any given ε, there is a set of probability one on which t can be chosen large enough to make the second term on the RHS of (17′) smaller than ε. By a result similar to the Borel–Cantelli Lemma (Shiryayev, 1984, p. 253), this is sufficient to guarantee that the second term converges almost surely to zero; i.e., there is a set $\tilde{E}$ for which $P(\tilde{E}) = 1$ and on which $|\omega(t)|/t$ converges to zero. The deterministic proof establishes that if $\omega \notin \tilde{E}$, then there is a ρ sufficiently small that $\omega \notin \tilde{E} \cap E_1(\bar{x}, \rho) \cap E_3(\bar{x}, U)$, where $\bar{x}$ is the limit point defined just before (19). Lemmas 1 and 3 then yield $P(\tilde{E}) = 1$. ■

For the remainder of the proof, we choose an arbitrary $\omega \in \Omega$. This gives us a sequence $\{x(t, \omega)\}$, for which we define $X(\omega)$, $\underline{w}(\omega)$, and $\bar{w}(\omega)$ just as in the deterministic proof. With no other changes, the deterministic proof verifies that, unless ω lies outside one or more of the several probability-one sets we have defined, we must have $\bar{w}(\omega) = 0$—i.e., $\{x(t, \omega)\}$ converges to x^*, and the proof of the stochastic version of the theorem is complete. ■

5. Concluding Remarks

The convergence proof we have given for linear recursive stochastic algorithms serves several functions. It seems to be the only complete, correct proof of convergence for algorithms of this form. Moreover, it shows how a number of flaws in Ljung's proofs for more general algorithms can be corrected, and, for one seemingly uncorrectible error in Ljung's analysis, we indicate a strengthening of assumptions that will circumvent the problem. Perhaps the greatest value of the proof is its potential as a model for proving similar results for algorithms that do not have the exact form of those with which we deal. The proof makes the role of each of the assumptions clear and provides insight into how the various assumptions can and cannot be altered.

In this connection, we have shown, in particular, that if the idea of a projection operator is to be extended to "decentralized" projection, as suggested in Marcet and Sargent (1989a) and Moreno and Walker (1991), then the existing convergence proofs (including the proof given here) do not guarantee convergence. The interplay between Condition L and the definition of the projection operator is quite delicate, and it seems likely that some change in Condition L would be required in order to yield Lemma 4 (that the projection operator is invoked only finitely many times). The failure of the convergence results to cover decentralized estimation when it involves projection operators is an important limitation which, if no way can be found to circumvent the obstacle we have described, narrows considerably the scope for applications of the projection idea to economic problems.

The proof we have given emphasizes the essentially deterministic character of the ordinary differential equation method pioneered by Ljung. If the deterministic results are developed with the stochastic case in mind, as we have done here, then the stochastic extension is as straightforward as Ljung suggests. Developing the proof in such a way that the stochastic extension can be carried out explicitly raises the possibility that the convergence results might hold under much weaker assumptions on the stochastic process than the assumptions which we (or Ljung) have used. In particular, Remark 5 suggests that the independence assumption S1 is stronger than necessary: It ought to be possible to allow for some kinds of interperiod dependence in the stochastic process. Extensions of the convergence theorems in this direction might give them much wider applicability in economic models.

References

Barbashin, E., and Krasovskii, N. 1952. "On the Stability of Motion in the Large." *Doklady Akad. Nauk. SSSR* 86, 453–456.

Hahn, W. 1967. *Stability of Motion*. Heidelberg: Springer-Verlag.

Ljung, L. 1975. "Theorems for the Asymptotic Analysis of Recursive Stochastic Algorithms." Report No. 7522, Department of Automatic Control, Lund Institute of Technology.

Ljung L. 1977. "Analysis of Recursive Stochastic Algorithms." *IEEE Transactions in Automatic Control*, AC-22, 551–575.

Marcet, A., and Sargent, T. 1989a. "Convergence of Least-squares Learning Mechanims in Self-referential Linear Stochastic Models." *Journal of Economic Theory* 48, 337–368.

Marcet, A., and Sargent, T. 1989b. "Convergence of Least-squares Learning in Environments with Hidden State Variables and Private Information." *Journal of Political Economy* 97, 1306–1322.

Moreno, D., and Walker, M. 1991. "On Recursive Learning in Noncooperative Games." Working paper, Department of Economics, University of Arizona.

Moreno, D., and Walker, M. 1994. "Two Problems in Applying Ljung's 'Projection Algorithms' to the Analysis of Decentralized Learning." *Journal of Economic Theory* 62, 420–427.

Shiryayev, A. N. 1984. *Probability*. New York: Springer-Verlag.

Woodford, M. 1990. "Learning to Believe in Sunspots." *Econometrica* 58, 277–307.

4 EFFICIENCY IN PARTNERSHIP WHEN THE JOINT OUTPUT IS UNCERTAIN

Steven R. Williams and Roy Radner

This paper concerns a model of partnership in which each partner privately chooses his input into a joint production process. The partners' inputs determine a probability distribution over a set of alternative output levels. Earlier work suggests that because of moral hazard, there cannot exist a rule for fully sharing the joint output that sustains the efficient inputs as a Nash equilibrium. We show that the existence of such a rule depends critically upon attitudes towards risk: the first-order conditions for existence are solvable generically if the partners are risk neutral, but are unsolvable generically if the partners are risk averse. Robust examples in the case of risk neutrality are then constructed in which such rules exist.

1. Introduction

A *partnership* is a group of agents who jointly produce some observable output. A production plan is *efficient* for the partnership if it is Pareto optimal given the disutility to each partner of his input into the collective effort. Moral hazard may exist when each partner's input is not fully observable, for a partner may have an incentive to contribute less of his