BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION
Statistical Research Report Series
No. RR2001/02

Convergence of a Robbins-Monro Algorithm for Recursive Parameter Estimation with Non-Monotone Weights and Multiple Zeros

David F. Findley
Statistical Research Division
Methodology and Standards Directorate
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: June 26, 2001
Report Revised: September 16, 2002

Convergence of a Robbins-Monro Algorithm for Recursive Estimation with Non-Monotone Weights for Functions with a Restricted Domain and Multiple Zeros

David F. Findley
U.S. Census Bureau

September 16, 2002

Abstract

Convergence properties are established for the output of a deterministic Robbins-Monro recursion for functions that can have singularities and multiple zeros. Our analysis is built largely on adaptations of lemmas of Fradkov published in Russian. We present versions of these lemmas in English for the first time. A gap in Fradkov's proof of the final lemma is fixed but only for the scalar case.

1 Introduction

In this note, we present a proposition about the convergence of a sequence θ_t obtained from a deterministic algorithm of Robbins-Monro form,

$$\theta_t = \theta_{t-1} - \delta_t f\left(\theta_{t-1}\right) + \delta_t w_t,\ t \geq 1, \tag{1}$$

where the sequences θ_t and w_t belong to the space R^d of d-dimensional real column vectors, $f\left(\theta\right)$ is an R^d-valued function defined on a subset of R^d containing the θ_t but possibly not on all of R^d, and $\delta_t, t \geq 1$ is a sequence of real numbers satisfying

$$\delta_t \geq 0, \lim_{t\to\infty} \delta_t = 0, \sum_{t=1}^{\infty} \delta_t = \infty. \tag{2}$$

We require the sequence w_t to be bounded,

$$\max_{u \geq 1} |w_u| < \infty, \tag{3}$$

and to satisfy the usual Kushner–Clark condition (10) below. Other hypotheses serve to restrict considerations to the situation in which the set of cluster points of θ_t,

$$K = \left\{ \overline{\theta} : \overline{\theta} = \lim_{t' \to \infty} \theta_{t'} \text{ for some subsequence } \theta_{t'} \text{ of } \theta_t \right\},$$

is contained within a bounded open set Θ on which $f(\theta)$ is continuously differentiable. Then, part (a) of the Proposition explains that the sequence θ_t either converges to some limit $\theta_\infty \in \Theta$,

$$\theta_t \to \theta_\infty, \tag{4}$$

or K is infinite. Part (b) states that if $f(\theta)$ has a "Lyapunov function" $V(\theta)$ with the basic property required in the Proposition (which is always true when $d = 1$), then (4) implies $f(\theta_\infty) = 0$. For the case $d = 1$, part (c) of the Proposition explains that, even when K is infinite,

$$K \subseteq \Theta_0 = \{\theta \in \Theta : f(\theta) = 0\}. \tag{5}$$

The proof of the last result and the foundations of the proofs of parts (a) and (b) have mainly been extracted from the proof of Theorem 3.17 of the monograph by Derevitzkiĭ and Fradkov (1981) (hereafter D&F) concerning the almost sure convergence of a stochastic approximation scheme. In D&F, this theorem, and the sequence of lemmas P.12–P.16 that constitute its proof, are credited to Fradkov. Although formulated differently, Fradkov's Lemma P.12 can be interpreted as reducing the proof of its convergence assertions for the stochastic approximation method considered in his Theorem 3.17 to the proof of the assertion (5) of our Proposition.

Our Proposition avoids the hypotheses of Fradkov's theorem that the weighting sequence δ_t is square summable and monotonically decreasing. D&F explicitly use these hypotheses only in the proof of Lemma P.12, essentially to verify (10) below, and there is only one place in the proof of the subsequent lemmas where we have to provide additional discussion because we do not assume monotonicity. However, there is a gap in the proof of the Fradkov's main supporting result, Lemma P. 16, that cannot be bridged without additional restrictions except in the case of sequences θ_t of dimension $d = 1$. For this reason, the part (c) of the Proposition and our version of this lemma, Lemma 6 below, are restricted to scalar case. The gap is described after Lemma 6's proof. With the exception of Lemmas 2 and 5, the sequence of lemmas and proofs given below follows closely the sequence of lemmas and proofs of D&F. However, our presentation provides greater precision in both statements and proofs of the lemmas and it makes these results available in English.

Our weaker assumptions on δ_t and the fact that we allow $f(\theta)$ to have singularities usefully increase the range of applicability of the results. Our Proposition has been applied to the analysis of recursively estimated time series model parameters in the situation of a misspecified moving average model of order one, see Cantor (2001) where almost sure convergence of some well known recursive estimation methods is established showing that an approximating sequence to the sequence of recursive estimates satisfies a stochastic version of (1) with non-monotonic δ_t and w_t converging almost surely to zero. In the incorrect model situation, it is known that under weak assumptions, maximum likelihood and other parameter estimates can converge almost surely to a set of points rather than to a unique limit,(see Section 4.3 of Findley, Pötscher, and Wei (2001) for a survey of the relevant literature) and that this set can be finite see Åström and Söderström (1974), Kabaila (1983) and Tanaka and Huzii (1992), or infinite, see Pötscher (1991). The limit set consists of minimizers of an appropriate function and in Cantor (2001), the functions $f(\theta)$ are the derivatives of the the functions being minimized. It has not been known whether, for a given realization (sample path) of the time series being modeled, the sequence of parameter estimates of an incorrect model converges to a single element of the set or oscillates between distinct elements. When the set Θ_0 of of zeros of $f(\theta)$ is finite, our Proposition shows that each recursive estimate considered by Cantor can converge to different limits for different realizations of the time series but on a given realization it must converge to some zero of $f(\theta)$.

There are stochastic Robbins-Monro algorithms with nonmonotonic but square summable δ_t for which more precise results are available about what values of Θ_0 can be limits when this set is finite with multiple entries, see Nevel'son (1972) and Nevel'son and Has'minskiĭ (1976), but these results assume that $f(\theta)$ is defined and differentiable on all of R^d with a derivative that is bounded (among other restrictions on $f(\theta)$ stronger than those of the Proposition) and therefore exclude the case of interest in Cantor (1991), in which, as in the examples of Åström and Söderström (1974), Kabaila (1983), and Tanaka and Huzii (1992), $f(\theta)$ is a rational function whose denominators have real zeros. Such functions are also excluded by the assumptions of Benaim (1996), a reference with very general results about the limiting behavior of θ_t when δ_t is monotonically decreasing but not necessarily square summable.

2 The Proposition and Lemmas

For a d-dimensional column vector $x = (x_1, \ldots, x_d)^T$, we define $\|x\| = \left(\sum_{i=1}^d x_i^2\right)^{1/2}$. For a $d \times d$ matrix M, we define $\|M\| = \lambda_{\max}^{1/2}\left(M^T M\right)$, where $\lambda_{\max}$ denotes the largest eigenvalue. For any sequence δ_t satisfying (2), we define, for every $\Delta > 0$,

$$t_\Delta = \min\left\{t_0 : \delta_t + \delta_{t+1} \le \Delta \text{ for all } t \ge t_0\right\} \tag{6}$$

and, for every $t \ge t_\Delta$,

$$u_\Delta(t) = \max\left\{u \ge t : \delta_t + \cdots + \delta_u \le \Delta\right\}. \tag{7}$$

Because $\sum_{t=0}^\infty \delta_t = \infty$, we always have $u_\Delta(t) < \infty$. Also, $u_\Delta(t) \ge t+1$, and because $\delta_t \to 0$,

$$\lim_{t\to\infty}\left\{\Delta - \sum_{u=t}^{u_\Delta(t)} \delta_u\right\} = 0. \tag{8}$$

For each $\Delta > 0$ and $t \ge 1$, we define

$$r_\Delta(t) = \begin{cases} 0, & \delta_t + \delta_{t+1} > \Delta \\ \max_{t \le v \le u_\Delta(t)-1} \left\|\sum_{u=t}^{v} \delta_{u+1} w_{u+1}\right\|, & \delta_t + \delta_{t+1} \le \Delta \end{cases}. \tag{9}$$

and require

$$\lim_{t\to\infty} r_\Delta(t) = 0. \tag{10}$$

In the applications in Cantor (1991) that motivated this paper, w_t is transient, meaning $\lim_{t\to\infty} w_t = 0$. In this case, it follows from $r_\Delta(t) \le \Delta \max_{u \ge t} |w_u|$, that (10) holds. A variety of conditions are shown to be equivalent to (10) (without assuming (3)) in Wang, Chong and Kulkarni

(1996): in particular, if (10) holds for one $\Delta > 0$, it holds for all $\Delta > 0$. Theorem 1 of this reference shows that (10) is a necessary and sufficient condition for (4) to hold for all bounded functions on R^d with a single zero, $\Theta_0 = \{\theta_\infty\}$ which are continuous at θ_∞ and have $\|\theta - \theta_\infty\|^{-1} (\theta - \theta_\infty)^T f(\theta)$ bounded above zero in a certain uniform sense for every $\theta \neq \theta_\infty$. Here ^{T} denotes transpose.

Proposition *Let $\theta_t, t \geq 0$ be a sequence that satisfies a recursion of the form* (1) *for which* (2) *and (10) hold for each $\Delta > 0$. Suppose there is a bounded open set Θ on which the function $f(\theta)$ in (1) is continuous and which θ_t enters infinitely often, without having a limit point of the boundary $\partial\Theta$ of Θ. Then the sequence θ_t is bounded,* $\sup_{t\geq 0} \|\theta_t\| < \infty$, *and its set K of cluster points is contained in Θ.*

(a) *If $f(\theta)$ is continuously differentiable on this set Θ, then K is either infinite or consists of a single vector, $K = \{\theta_\infty\}$, in which case* $\lim_{t\to\infty} \theta_t = \theta_\infty$ *holds.*

(b) *Suppose in addition that there is a twice continuously differentiable function $V\colon\ \Theta \longmapsto R$ such that, for any $\theta \in \Theta$* for which $f(\theta) \neq 0$, *the derivative $\nabla V = (\partial V/\partial\theta_1, \ldots, \partial V/\partial\theta_d)^T$ has the property*

$$\nabla V(\theta)^T f(\theta) > 0. \tag{11}$$

Then if $\lim_{t\to\infty} \theta_t = \theta_\infty$, *it necessarily follows that* $f(\theta_\infty) = 0$, *i.e.* (5) *holds.*

(c) *If $d = 1$, then any antiderivative $V(\theta)$ of $f(\theta)$ satisfies* (11), *and the inclusion* (5) *holds. Therefore, if Θ_0 is finite, then* $\lim_{t\to\infty} \theta_t$ *exists and is contained in* $\Theta_0 = \{\theta \in \Theta : f(\theta) = 0\}$.

The proof is obtained via a sequence of lemmas and the following observations. First, the assumptions of the Proposition yield that:

(i) There are only finitely many t' such that $\theta_{t'} \in \Theta$ but $\theta_{t'+1} \notin \Theta$. Otherwise, since Θ is bounded, a subsequence $\theta_{t''}$ of the infinite sequence $\theta_{t'}$ would have a limit $\overline{\theta}$ in $\Theta \cup \partial\Theta$. Since $\overline{\theta} \notin \partial\Theta$, by assumption, we would have $\overline{\theta} \in \Theta$. By continuity, $\lim_{t''\to\infty} f(\theta_{t''}) = f\left(\overline{\theta}\right)$, and (1), (2), and (10) would yield $\lim_{t''\to\infty} \{\theta_{t''+1} - \theta_{t''}\} = 0$, and therefore $\lim_{t''\to\infty} \theta_{t''+1} = \overline{\theta}$, and hence that $\theta_{t''+1} \in \Theta$ for t'' sufficiently large (because Θ is open), which contradicts the definition of the $\theta_{t'}$ sequence. Consequently, there is a time t_Θ such that

$$\theta_t \in \Theta,\ t \geq t_\Theta. \tag{12}$$

For the subsequent discussion, we shall always assume that $t \geq t_\Theta$.

(ii) It follows from (12) that the sequence θ_t is bounded. Hence its set of cluster points K is a nonempty compact subset of $\Theta \cup \partial\Theta$, and since $K \cap \partial\Theta$ is empty, $K \subset \Theta$, as asserted.

Consequently, there exists a $\rho > 0$ such that, for each $\overline{\theta} \in K$, the closed ball $B(\overline{\theta}, 2\rho) = \left\{\theta \in \Theta : \left\|\theta - \overline{\theta}\right\| \leq 2\rho\right\}$ satisfies

$$B(\overline{\theta}, 2\rho) \subseteq \Theta. \tag{13}$$

Because K is compact, for any point θ^* on the boundary $\partial\Theta$ of Θ, we have

$$\min_{\overline{\theta} \in K} \min_{\theta \in B(\overline{\theta}, 2\rho)} \|\theta - \theta^*\| > 0.$$

It follows that the closure $\tilde{K} = \tilde{K}(\rho)$ of $\cup_{\overline{\theta} \in K} B(\overline{\theta}, 2\rho)$ is a compact subset of Θ containing K in its interior, $K \subset Int\tilde{K}$. By the continuity of $\nabla f(\theta)$, $L = 1 + \max_{\theta \in \tilde{K}} \|\nabla f(\theta)\|$ is finite. Hence, from the Mean Value Theorem and the convexity of $B(\overline{\theta}, 2\rho)$, we have

$$\|f(\theta) - f(\theta')\| \leq L \|\theta - \theta'\| \tag{14}$$

for every $\theta, \theta' \in B(\overline{\theta}, 2\rho)$ when $\overline{\theta} \in K$.

Lemma 1 *(cf. Lemma P.13 of D&F). For each $\rho > 0$ such that (13) holds for all $\overline{\theta} \in K$, there exists a $\Delta_0 = \Delta_0(\rho)$ such that for every $0 < \Delta \leq \Delta_0$ and any $\overline{\theta} \in K$, if*

$$\left\|\theta_t - \overline{\theta}\right\| \leq \rho, \tag{15}$$

holds for some $t \geq t_\Delta$, then so does

$$\sup_{t \leq u \leq u_\Delta(t)} \left\|\theta_u - \overline{\theta}\right\| \leq 2\rho, \tag{16}$$

for t_Δ and $u_\Delta(t)$ defined in (6) and (7).

Proof. We use induction to establish (16). Let $t \leq v < u_\Delta(t)$ be such that

$$\sup_{t \leq u \leq v} \left\|\theta_u - \overline{\theta}\right\| \leq 2\rho. \tag{17}$$

(This holds for $v = t$ by (15).) From (1),

$$\theta_{v+1} = \theta_t - \sum_{u=t}^{v} \delta_{u+1} f\left(\theta_u\right) + \sum_{u=t}^{v} \delta_{u+1} w_{u+1}$$

$$= \theta_t - \sum_{u=t}^{v} \delta_{u+1} f\left(\overline{\theta}\right) + \sum_{u=t}^{v} \delta_{u+1} \left\{ f\left(\overline{\theta}\right) - f\left(\theta_u\right) \right\} + \sum_{u=t}^{v} \delta_{u+1} w_{u+1}. \qquad (18)$$

It follows from (18) and (14) that

$$\left\|\theta_{v+1} - \overline{\theta}\right\| \leq \left\|\theta_t - \overline{\theta}\right\| + \Delta \left\|f\left(\overline{\theta}\right)\right\| + r_\Delta\left(t\right) + L \sum_{u=t}^{v} \delta_{u+1} \left\|\theta_u - \overline{\theta}\right\|.$$

Therefore, from an induction argument or the Discrete Bellman-Gronwall Lemma (Solo and Kong, 1995 p. 315) and the fact that $e^x \geq 1 + x$ for any $x \geq 0$, we have

$$\begin{aligned} \left\|\theta_{v+1} - \overline{\theta}\right\| &\leq \left\{ \left\|\theta_t - \overline{\theta}\right\| + \Delta \left\|f\left(\overline{\theta}\right)\right\| + r_\Delta\left(t\right) \right\} \prod_{u=t}^{v} \left(1 + L\delta_{u+1}\right) \\ &\leq \left\{ \left\|\theta_t - \overline{\theta}\right\| + \Delta \left\|f\left(\overline{\theta}\right)\right\| + r_\Delta\left(t\right) \right\} e^{L \sum_{u=t}^{v} \delta_{u+1}} \\ &\leq \left\{ \left\|\theta_t - \overline{\theta}\right\| + \Delta \left(\left\|f\left(\overline{\theta}\right)\right\| + \max_{u \geq t} \left|w_u\right| \right) \right\} e^{L\Delta}. \end{aligned} \qquad (19)$$

Define

$$\Delta_0\left(\rho\right) = \min \left\{ \frac{1}{L} \log \frac{4}{3}, \frac{\rho}{2\left(\max_{\overline{\theta} \in K} \left\|f\left(\overline{\theta}\right)\right\| + \max_{u \geq 1} \left|w_u\right| + 1\right)} \right\}. \qquad (20)$$

Then for each $\Delta \leq \Delta_0\left(\rho\right)$ and $t \geq t_\Delta$ for which (15) and (17) hold, it follows from (19) and (15) that $\left\|\theta_{v+1} - \overline{\theta}\right\| \leq 2\rho$. Thus, by induction, (16) holds for all $t \geq t_\Delta, \Delta \leq \Delta_0\left(\rho\right)$ for which (15) holds, as asserted. ■

The next lemma establishes the part (a) of the Proposition.

Lemma 2 *Suppose the set K of cluster points of θ_t is finite. Then θ_t has only one cluster point and hence has a limit. That is, $K = \{\theta_\infty\}$ for some $\theta_\infty \in \Theta$ and $\theta_t \to \theta_\infty$.*

Proof. Suppose to the contrary that K has several elements but is finite. Then there is a $\rho > 0$ satisfying the hypotheses of Lemma 1 such that for each distinct pair $\tilde{\theta}, \bar{\theta} \in K$,

$$\left\|\tilde{\theta} - \bar{\theta}\right\| > 3\rho \tag{21}$$

holds, and there is a $t_\rho \geq 1$, such that for every $t \geq t_\rho$ there exists a $\bar{\theta}(t) \in K$. such that

$$\left\|\theta_t - \bar{\theta}(t)\right\| \leq \rho$$

holds. Define Δ_0 by (20). Because no $\bar{\theta}(t)$ is the limit of θ_t, there is a $t \geq \max\{t_\rho, t_{\Delta_0}\}$ such that $\bar{\theta}(t) \neq \bar{\theta}(t+1)$. Applying Lemma 1, we obtain

$$\begin{aligned}\left\|\bar{\theta}(t+1) - \bar{\theta}(t)\right\| &\leq \left\|\theta_{t+1} - \bar{\theta}(t+1)\right\| + \left\|\theta_{t+1} - \bar{\theta}(t)\right\| \\ &\leq \rho + 2\rho = 3\rho,\end{aligned}$$

contradicting (21). Thus θ_t has a single cluster point and hence a limit. ■

Lemma 3 *(cf. Lemma P.14 of D&F). For $\bar{\theta} \in K$, ρ as in (13) and Δ_0 as in (20), and any $0 < \Delta \leq \Delta_0$ and $t \geq t_\Delta$ such that (15) holds, we have*

$$\theta_{u_\Delta(t)} = \theta_t - f\left(\bar{\theta}\right)\Delta + q_1(t, \Delta) + q_2(t, \Delta), \tag{22}$$

where $q_1(t, \Delta)$ has the property that there exist constants C_1, C_2 such that

$$\left\|q_1(t, \Delta)\right\| \leq C_1\Delta\left\|\theta_t - \bar{\theta}\right\| + C_2\Delta^2, \tag{23}$$

and $q_2(t, \Delta)$ satisfies

$$\lim_{t\to\infty} q_2(t, \Delta) = 0. \tag{24}$$

Proof. From (1), we obtain

$$\begin{aligned}\theta_{u_\Delta(t)} &= \theta_t - f\left(\bar{\theta}\right)\Delta + \left(\Delta - \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1}\right) f\left(\bar{\theta}\right) \\ &\quad + \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1}\left\{f\left(\bar{\theta}\right) - f\left(\theta_u\right)\right\} + \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} w_{u+1}.\end{aligned}$$

Set

$$q_1(t,\Delta) = \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} \left\{ f\left(\overline{\theta}\right) - f\left(\theta_u\right) \right\},$$

and

$$q_2(t,\Delta) = \left(\Delta - \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} \right) f\left(\overline{\theta}\right) + \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} w_{u+1}. \tag{25}$$

From (14) we have

$$\left\| q_1(t,\Delta) \right\| \le L \sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} \left\| \theta_u - \overline{\theta} \right\|, \tag{26}$$

and from (19) and (20),

$$\sum_{u=t}^{u_\Delta(t)-1} \delta_{u+1} \left\| \theta_u - \overline{\theta} \right\| = \delta_{t+1} \left\| \theta_t - \overline{\theta} \right\| + \sum_{u=t+1}^{u_\Delta(t)-1} \delta_{u+1} \left\| \theta_u - \overline{\theta} \right\|$$

$$\begin{aligned} \le \; & \delta_{t+1} \left\| \theta_t - \overline{\theta} \right\| + e^{L\Delta} \sum_{u=t+1}^{u_\Delta(t)-1} \delta_{u+1} \left\| \theta_t - \overline{\theta} \right\| \\ & + \Delta e^{L\Delta} \left(\sum_{u=t+1}^{u_\Delta(t)-1} \delta_{u+1} \right) \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \ge t} \left| w_{u+1} \right| \right) \end{aligned}$$

$$\le \frac{4}{3} \Delta \left\| \theta_t - \overline{\theta} \right\| + \frac{4}{3} \Delta^2 \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \ge 1} \left| w_u \right| \right).$$

This yields (23) with

$$C_1 = \frac{4}{3} L, \quad C_2 = \frac{4}{3} L \left(\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| + \max_{u \ge 1} \left| w_u \right| \right). \tag{27}$$

The assertion (24) concerning $q_2(t,\Delta)$ follows from (8), (3), $\delta_t \to 0$, $\max_{\overline{\theta} \in K} \left\| f\left(\overline{\theta}\right) \right\| < \infty$, and (10). ■

The final three lemmas use the properties of $V(\theta)$.

Lemma 4 *(cf. Lemma P.15 of D&F). Suppose $\overline{\theta} \in K$ is such that $f\left(\overline{\theta}\right) \neq 0$. Then for each subsequence $\theta_{t'}$ converging to $\overline{\theta}$, there exist $\Delta_0 > 0$ and $\eta > 0$ with the following property: for each $0 < \Delta \leq \Delta_0$ there is a $t'(\Delta)$ such that for all $t' \geq t'(\Delta)$, the inequality*

$$V\left(\theta_{u_\Delta(t')}\right) < V\left(\theta_{t'}\right) - \eta\Delta \tag{28}$$

holds.

Proof. Let $\rho > 0$ be such that (13) holds. For each $0 < \varepsilon < \rho$, define $\Delta_0(\varepsilon)$ by (20) with ρ is replaced by ε. For $\Delta \leq \Delta_0(\varepsilon)$ consider a $t' \geq t_\Delta$ such that $\left\|\theta_{t'} - \overline{\theta}\right\| \leq \varepsilon$ holds, and therefore $\left\|\theta_{u_\Delta(t')} - \overline{\theta}\right\| \leq 2\varepsilon$ by Lemma 1. To simplify notation, set $\theta' = \theta_{t'}$ and $\theta'' = \theta_{u_\Delta(t')}$. By taking Taylor expansions of V and ∇V, we obtain

$$V(\theta'') - V(\theta') = \nabla V(\zeta)^T (\theta'' - \theta')$$

$$\begin{aligned} &= \nabla V\left(\overline{\theta}\right)^T (\theta'' - \theta') + \left[\nabla V(\zeta) - \nabla V\left(\overline{\theta}\right)\right]^T (\theta'' - \theta') \\ &= \nabla V\left(\overline{\theta}\right)^T (\theta'' - \theta') + \left(\zeta - \overline{\theta}\right)^T \nabla^2 V(\zeta')(\theta'' - \theta') \end{aligned} \tag{29}$$

with $\zeta \in [\theta', \theta'']$ and $\zeta' \in \left[\zeta, \overline{\theta}\right]$. ($[\theta', \theta''] = \{\alpha\theta' + (1-\alpha)\theta'' : 0 \leq \alpha \leq 1\}$, etc.) Since $B\left(\overline{\theta}, 2\varepsilon\right)$ is convex, $\zeta, \zeta' \in B\left(\overline{\theta}, 2\varepsilon\right)$. From Lemma 3,

$$\theta'' - \theta' = -f\left(\overline{\theta}\right)\Delta + q_1(t', \Delta) + q_2(t', \Delta), \tag{30}$$

where

$$\|q_1(t, \Delta)\| \leq C_1\Delta\varepsilon + C_2\Delta^2$$

with C_1, C_2 given by (27), and where $\lim_{t'\to\infty} q_2(t', \Delta) = 0$. Since $\overline{\theta} \notin \Theta_0$, it follows from (11) that $\nabla V\left(\overline{\theta}\right)^T f\left(\overline{\theta}\right) = \eta_1 > 0$. Let η satisfy $0 < \eta < \eta_1$ and set $\tilde{\eta} = \eta_1 - \eta$. Substituting (30) into (29), we obtain

$$\begin{aligned} V(\theta'') - V(\theta') &= -\eta_1\Delta \\ &- \left(\zeta - \overline{\theta}\right)^T \nabla^2 V(\zeta') f\left(\overline{\theta}\right)\Delta \\ &+ \left[\nabla V\left(\overline{\theta}\right)^T + \left(\zeta - \overline{\theta}\right)^T \nabla^2 V(\zeta')\right] q_1(t', \Delta) \\ &+ \left[\nabla V\left(\overline{\theta}\right)^T + \left(\zeta - \overline{\theta}\right)^T \nabla^2 V(\zeta')\right] q_2(t', \Delta). \end{aligned} \tag{31}$$

Set

$$L_1 = \max_{\theta \in \tilde{K}} \|\nabla V(\theta)\|, \quad L_2 = \max_{\theta \in \tilde{K}} \left\|\nabla^2 V(\theta)\right\| \tag{32}$$

and $C_3 = L_1 + 2\rho L_2$. Now choose ε small enough that

$$2\varepsilon L_2 \max_{\theta \in \tilde{K}} |f(\theta)| < \frac{\tilde{\eta}}{3}, \;\; C_1 C_3 \varepsilon < \frac{\tilde{\eta}}{6}$$

and also so that

$$C_2 C_3 \Delta_0(\varepsilon) < \frac{\tilde{\eta}}{6}.$$

Set $\Delta_0 = \Delta_0(\varepsilon)$. Then, for any $0 < \Delta \leq \Delta_0$, if we choose $t'(\Delta) \geq t_\Delta$ so that $t \geq t'(\Delta)$ implies

$$\left\|\theta_{t'} - \overline{\theta}\right\| \leq \varepsilon$$

and

$$C_3 q_2(t, \Delta) < \frac{\tilde{\eta}}{3}\Delta,$$

it follows from (31) that

$$V(\theta'') < V(\theta') - \eta_1 \Delta + \tilde{\eta}\Delta = V(\theta') - \eta\Delta,$$

holds when $t' \geq t'(\Delta)$, as asserted. ■

Now we obtain (b) of the Proposition.

Lemma 5 *Under the hypotheses of Lemma 4, if $\theta_t \to \theta_\infty \in \Theta$, then $f(\theta_\infty) = 0$, i.e. $\theta_\infty \in \Theta_0$.*

Proof. Suppose to the contrary that $\theta_\infty \notin \Theta_0$. Then by Lemma 4, for all sufficiently large t, $V\left(\theta_{u_\Delta(t)}\right) < V(\theta_t) - \eta\Delta$ holds for some $\eta > 0, \Delta > 0$. Since $u_\Delta(t) \geq t+1 \to \infty$ as $t \to \infty$, it follows from $\theta_t \to \theta_\infty$ that $\theta_{u_\Delta(t)} \to \theta_\infty$, and therefore from the continuity of V, that $V(\theta_\infty) \leq V(\theta_\infty) - \eta\Delta$, which is impossible. Hence, $\theta_\infty \in \Theta_0$. ■

The final lemma yields (c) of the Proposition. It is the first to require $d = 1$. Under this condition, for every $\overline{\theta} \in \Theta$ for which $f\left(\overline{\theta}\right) \neq 0$, it follows from (11), the Mean Value Theorem, and the continuity of $\nabla V(\theta)$ that there exist $m > 0, \rho > 0$ such that

$$\left|V(\theta) - V\left(\overline{\theta}\right)\right| \geq m\left\|\theta - \overline{\theta}\right\| \tag{33}$$

holds for all $\theta \in B\left(\overline{\theta}, 2\rho\right)$.

Lemma 6 *(cf. Lemma P.16 of D&F). Under the assumptions of the Proposition, no point $\overline{\theta} \in \Theta$ for which $f\left(\overline{\theta}\right) \neq 0$ can be a limit point of θ_t. Therefore $K \subseteq \Theta_0$.*

Proof. Suppose, to the contrary, that there is a $\overline{\theta} \in \Theta$ with $f\left(\overline{\theta}\right) \neq 0$ (and therefore with $\eta_1 = \nabla V\left(\overline{\theta}\right)^T f\left(\overline{\theta}\right) > 0$) that is a limit point of θ_t. Choose ρ so that (13) is satisfied, and also so that

$$f(\theta) \neq 0 \tag{34}$$

and (33) hold for all $\theta \in B\left(\overline{\theta}, 2\rho\right)$. For Δ_0 as in the proof of Lemma 4, choose $\Delta, \eta_0 > 0$ so that

$$\Delta \leq \min\{\Delta_0, \rho m\}, \quad \eta_0 < \min\{\eta_1, 1\}. \tag{35}$$

Then Lemma 4's proof shows that, for any subsequence $\theta_{t'}$ that converges to $\overline{\theta}$, the inequality

$$V\left(\theta_{u_\Delta(t')}\right) < V(\theta_{t'}) - \eta_0 \Delta \tag{36}$$

holds for all t' large enough. The sequence of values $\theta_{u_\Delta(t')}$ appearing on the l.h.s. of (36) does not necessarily change with t'. Since δ_t need not be monotonically decreasing, all that can be asserted is that for $t'' > t' \geq t_\Delta$, one has $u_\Delta(t'') \geq u_\Delta(t')$, with $u_\Delta(t'') > u_\Delta(t')$ holding for $t'' \geq u_\Delta(t')$. The latter inequality guarantees that $\theta_{u_\Delta(t')}$ takes on infinitely many values of θ_t. From this fact and (36), and from $V(\theta_{t'}) \to V\left(\overline{\theta}\right)$, we can conclude that, for a given $0 < \eta < \eta_0$, the sequence θ_t enters each of the disjoint sets

$$R_{\eta\Delta} = \left\{\theta \in B\left(\overline{\theta}, 2\rho\right) : V(\theta) \leq V\left(\overline{\theta}\right) - \eta\Delta\right\}$$

and

$$S_{\frac{1}{2}\eta\Delta} = \left\{\theta \in B\left(\overline{\theta}, 2\rho\right) : V(\theta) > V\left(\overline{\theta}\right) - \frac{\eta}{2}\Delta\right\}$$

infinitely often. Let $\theta_{\tau'}$ denote the subsequence of last values of θ_t in $R_{\eta\Delta}$ before a next entry $\theta_{\tau'+n'}$ in $S_{\frac{1}{2}\eta\Delta}$. Some subsequence $\theta_{\tau''}$ of $\theta_{\tau'}$ must have a limit $\tilde{\theta}$. Since $V(\theta_{\tau''+1}) > V\left(\overline{\theta}\right) - \eta\Delta$ and $V(\theta_{\tau''+1}) - V(\theta_{\tau''}) \to 0$, we must have $V\left(\tilde{\theta}\right) = V\left(\overline{\theta}\right) - \eta\Delta$. Therefore, from (33) and (35),

$$\left\|\tilde{\theta} - \overline{\theta}\right\| < \rho. \tag{37}$$

Thus, $f\left(\tilde{\theta}\right) \neq 0$ by (34). With $\Delta_0(\rho)$ as in (20) and L_1 as in (32), we can conclude from Lemma 4 that there exist $\widetilde{\Delta} > 0$ satisfying

$$\widetilde{\Delta} < \min\left\{\frac{\rho}{2}, \frac{\eta}{6L_1}\Delta, \Delta_0(\rho)\right\}, \tag{38}$$

and $0 < \widetilde{\eta} < 1$ and $\tau''\left(\widetilde{\Delta}\right) \geq t_{\widetilde{\Delta}}$ such that $V\left(\theta_{u_{\widetilde{\Delta}}(\tau'')}\right) < V(\theta_{\tau''}) - \widetilde{\eta}\widetilde{\Delta}$ holds for all $\tau'' \geq \tau''\left(\widetilde{\Delta}\right)$. Hence

$$V\left(\theta_{u_{\widetilde{\Delta}}(\tau'')}\right) < V\left(\overline{\theta}\right) - \eta\Delta - \widetilde{\eta}\widetilde{\Delta}. \tag{39}$$

Because $\theta_{\tau''} \to \tilde{\theta}$, we can, by taking a larger $\tau''\left(\widetilde{\Delta}\right)$ if necessary, further obtain

$$\left\|\theta_{\tau''} - \tilde{\theta}\right\| \leq \widetilde{\Delta}, \tag{40}$$

for all $\tau'' \geq \tau''\left(\widetilde{\Delta}\right)$, and therefore, from Lemma 1, also

$$\max_{\tau'' \leq u \leq u_{\widetilde{\Delta}}(\tau'')} \left\|\theta_u - \tilde{\theta}\right\| \leq 2\widetilde{\Delta} < \rho. \tag{41}$$

Due to (37), the last inequality shows that

$$\theta_u \in B\left(\overline{\theta}, 2\rho\right), \ \tau'' \leq u \leq u_{\widetilde{\Delta}}(\tau''). \tag{42}$$

With (39), this yields the key result: $\theta_{u_{\widetilde{\Delta}}(\tau'')} \in R_{\eta\Delta}$. Since $\theta_{\tau''}$ is a last value in $R_{\eta\Delta}$ before an entry in $S_{\frac{1}{2}\eta\Delta,}$ at time $\tau''+n''$, we must have $\tau''+n'' < u_{\widetilde{\Delta}}(\tau'')$ whenever $\tau'' \geq \tau''\left(\widetilde{\Delta}\right)$. For these τ'', it follows from (40), and (41) that $\|\theta_{\tau''} - \theta_{\tau''+n''}\| \leq 3\tilde{\Delta}$. Therefore, from (38), we have

$$\begin{aligned} V(\theta_{\tau''+n''}) &\leq V(\theta_{\tau''}) + |V(\theta_{\tau''}) - V(\theta_{\tau''+n''})| \\ &\leq V\left(\overline{\theta}\right) - \eta\Delta + L_1\|\theta_{\tau''} - \theta_{\tau''+n''}\| \\ &\leq V\left(\overline{\theta}\right) - \eta\Delta + 3\tilde{\Delta}L_1 \\ &< V\left(\overline{\theta}\right) - \frac{\eta}{2}\Delta, \end{aligned}$$

by virtue of (38). But this contradicts $\theta_{\tau''+n''} \in S_{\frac{1}{2}\eta\Delta}$. Thus the proofs of Lemma 6 and the Proposition are complete. ■

Remark. The gap in the proof given in D&F is the lack of verification of (42), which seems to require a condition that forces $\left\|\theta - \overline{\theta}\right\|$ to be small when $\left|V(\theta) - V\left(\overline{\theta}\right)\right|$ is, as (33) does. No such condition is imposed in D&F. When $d > 1$ and $\overline{\theta} \in \Theta$ is such that $\nabla V\left(\overline{\theta}\right) \neq 0$, the level sets $\left\{\theta \in \Theta : V(\theta) = V\left(\overline{\theta}\right)\right\}$ will be nonempty, so (33) will fail for every $m > 0$.

Acknowledgement. The author is indebted to James Cantor and John Aston for their comments on earlier drafts of this paper, to Ching-Zong Wei for calling attention to Benaim (1996), and to Victor Solo for suggesting examination of the work of Kulkarni and coauthors.

References

[1] Åström, K. J. and T. Söderström (1974). Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model, *IEEE Trans. Aut. Contr.* 19, 769-773.

[2] Benaim, M. (1996). A dynamical system approach to stochastic approximations. *SIAM J. Control and Optimization* 34 ,437-472.

[3] Cantor, J. L. (2001). *Recursive and Batch Estimation for Misspecified ARMA Models.* Ph.D. Dissertation, Columbian School of Arts and Science, George Washington University.

[4] Derevitzkiĭ, D. P. and A. L. Fradkov (1981). *Applied Theory of Discrete Adaptive Control Systems.* (In Russian) Nauka: Moscow

[5] Findley, D. F., B. M. Pötscher and C.-Z. Wei (2002). Modeling of time series arrays by multistep prediction or likelihood methods. *Journal of Econometrics* (forthcoming).

[6] Kabaila, P. V. (1983). Parameter value of ARMA models minimizing the one-step-ahead prediction-error when the true system is not in the model set. *J. Appl. Prob.* 20, 405-408.

[7] Nevel'son, M. B. (1972). The convergence of continuous and discrete Robbins-Monro procedures in the case of several roots of the regression equation. *Problems of Information Transmission* 7, 215-223.

[8] Nevel'son, M. B and R. Z. Has'minskiĭ (1976). *Stochastic Approximation and Recursive Estimation.* Translations of Mathematical Monographs, v. 47, American Mathematical Society.

[9] Pötscher, B. M. (1991). Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models. *Econometric Theory* 7, 435-449. Corrections: *Econometric Theory* 10, 811.

[10] Solo, V. and X. Kong (1995). *Adaptive Signal Processing Algorithms: Stability and Performance.* Prentice Hall: Englewood Cliffs.

[11] Tanaka, M. and M. Huzii (1992). Some properties of moving average model fittings. *J. Japan Statist. Soc. 22, 33-44.*

[12] Wang, I.-J., E. K. P. Chong, and S. R. Kulkarni (1996). Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms. *Adv. Appl. Prob.* 28, 784-801.