Studies in Nonlinear Dynamics & Econometrics

Volume 6, *Issue* 1 2002 *Article* 1

Asymptotic Properties of Some Projection−based Robbins−Monro Procedures in a Hilbert Space

Xiaohong Chen* Halbert White†

*Department of Economics, London School of Economics & Political Science
†Department of Economics, University of California, San Diego

Studies in Nonlinear Dynamics & Econometrics is produced by The Berkeley Electronic Press (bepress). http://www.bepress.com/snde

Asymptotic Properties of Some Projection−based Robbins−Monro Procedures in a Hilbert Space

Abstract

Let H be an infinite−dimensional real separable Hilbert space. Given an unknown mapping $M{:}H$ (r)H that can only be observed with noise, we consider two modified Robbins−Monro procedures to estimate the zero point ?o (H of M. These procedures work in appropriate finite dimensional sub−spaces of growing dimension. Almost−sure convergence, functional central limit theorem (hence asymptotic normality), law of iterated logarithm (hence almost−sure loglog rate of convergence), and mean rate of convergence are obtained for Hilbert space−valued mixingale, (−*dependent* error processes.

I. INTRODUCTION

To locate the root θ_o in H (a Hilbert space) of an unknown measurable mapping $M : H \rightarrow H$, one can use the stochastic approximation (SA) method introduced by Robbins and Monro (1951). The Robbins-Monro (RM) procedure recursively approximates θ_o by:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + a_n M_n(Z_n, \hat{\theta}_n) , \quad n = 1, 2, ...,$$

where $\hat{\theta}_1$ is a randomly chosen element in H, a_n is a step size tending to zero, and $M_n(Z_n, \theta)$ is a measurement of $M(\theta)$ at time n , i.e.,

$$M_n(Z_n, \theta) \equiv M(\theta) + U_n(Z_n, \theta) ,$$

where the error term $U_n(Z_n, \theta)$ is inßuenced by random elements Z_n .

Robbins and Monro (1951) treated the case $H = I\!R$ (real line). Since then, various Þnite dimensional SA procedures have been studied intensively. Generally, $I\!R^d$ $(d < \infty)$ –valued SA processes are captured asymptotically by the solutions to deterministic ordinary differential equations (ODE's) in $I\!R^d$: $\dot{\theta}(t) = M(\theta(t))$. Kushner, Ljung, and others have obtained numerous elegant results about almost-sure convergence, asymptotic normality and rate of convergence of such SA procedures under very general but rather abstract conditions. These results have been widely utilized in such different areas as estimation in statistics, adaptive learning in control theory, simulation in computation, and signal processing in engineering.

Recently, these techniques have been applied in various economic theory and econometric contexts. For example, Marcet and Sargent (1989), Woodford (1990), Sargent (1993), Evans and Honkapohja (1995), Crawford (1995), and many others have applied recursive nonlinear least squares (a special case of $I\!R^d$ $(d < \infty)$ –valued RM algorithms) to model boundedly rational economic agents' learning behavior in macroeconomics and game theory. Pakes and McGuire (2001) have applied Q-learning (another special case of $I\!R^d$ $(d < \infty)$ –valued RM algorithms) to solve and estimate complicated empirical IO models. Kuan and White (1993) have established the consistency and asymptotic normality of the $I\!R^d$ –valued recursive m –estimator and then applied these to possibly misspeciÞed nonlinear parametric regression models, including a leading neural network model. Patilea and Renault (2001) have applied the $I\!R^d$ $(d < \infty)$ –valued RM algorithm to perform option pricing with stochastic volatility (latent factor) models.

However, all the preceding applications require correct parametric speciÞcations. Kuan and White (1994) present an example where agents misspecify the form of a heterogeneous variance, and the resulting recursive nonlinear least squares procedure fails to converge to the rational expectations equilibrium. Chen and White (1998b) consider an example where a competitive Þrm misspeciÞes the parametric form of market supply function, and the recursive least squares learning procedure leads to a Þxed point which is not a rational expectations equilibrium market price. Because economic systems are generally too complicated to be plausibly speciÞed correctly as parametric models and because SA is appealing in its simplicity, our goal here is to develop some *nonparametric* SA procedures. By their nature, nonparametric procedures have less scope for misspeciÞcation. For example, Chen and White (1998b), in the example cited above, show that nonparametric learning procedures of the type studied here do in fact converge to the rational expectations equilibrium price.

The results in this paper will be useful for deriving large sample properties for all kinds of nonlinear, nonparametric, or semi-nonparametric recursive moment- or score-type estimators, especially those involving latent state variables such as stochastic volatility models, Garch models, the nonlinear Kalman Þlter, on-line forecasting of density, or on-line regression in a heterogeneous, dependent dynamic environment. Moreover, the results presented in this paper and those in Chen and White (1998b) will allow for nonlinear learning where the agents do not specify the parametric form of a model, but allow for more and more ßexible possibly nonlinear functional forms to be learned as new information arrives as time goes by. Most of the results in this paper are derived under the assumption that the true errors $\{ U_n(Z_n,\theta_o) \}$ are Hilbert-valued mixingale processes or Hilbert-valued near epoch dependent functions of mixing processes (see Chen and White (1996) for deÞnitions). Thus, our results allow for lots of heterogeneity and temporal dependence, which are important features in modeling economic agents' adaptive learning behavior. To avoid further lengthening this paper we do not provide speciÞc examples here, but interested readers can Þnd many examples in Chen and White (1996, 1998a, 1998b), including recursive nonparametric density and regression estimation, nonparametric goodness-of-Þt tests, and nonparametric adaptive learning.

There are already many papers that treat the inÞnite dimensional SA (e.g., Venter (1966), Walk (1977), Berman and Shwartz (1989), Yin and Zhu (1990)). The asymptotic properties are again determined by the associated deterministic ODE. The conditions are similar to those of the Þnite dimensional case. However, most of these results are restricted to the $\theta-$*independent* error case, (i.e., $U_n = F_n(Z_n)$, F_n a Borel-measurable mapping of Z_n , independent of θ) . Also, most of the results assume *a priori* that the elements of the sequence $\{ \hat{\theta}_n \}$ lie in a certain compact subset. An even more serious problem from an applications point of view is that previous results are cast directly in either an inÞnite dimensional Hilbert space or a general Banach space. Since the inÞnite dimensional SA is not computable, it is preferable to develop sieve-like SA procedures for the purposes of estimation and inference. So far, there are three papers in this direction: Goldstein (1988) has proved almost-sure convergence in the norm topology for a modiÞed Kiefer-Wolfowitz (1952) procedure in inÞnite dimensional Hilbert space using a sieve approach; Nixdorf (1984) has shown asymptotic normality for a modiÞed sieve-type RM procedure; and Yin (1992) has proved almost-sure convergence in the weak topology for a sieve-type RM procedure. However, all three papers impose the restrictive $\theta-$*independence* condition on the error terms; and the Þrst two papers require that the error sequence $\{ U_n \}$ is a martingale difference sequence.

This paper combines the direct abstract approach (Venter (1966), Walk (1977), Yin & Zhu (1990), etc.) with the sieve approach (Nixdorf (1984), etc.). In Section II, we present some modiÞed Hilbert-space valued RM procedures, and obtain their almost-sure norm-convergence properties. Our procedures do not require a prior compact subset to which $\{ \hat{\theta}_n \}$ must belong. For $\theta-$*independent* errors, we need not even assume a prior bound on θ_o , exploiting the advances of Yin & Zhu (1990). Under the assumption of a prior bound on θ_o , we relax Yin & Zhu's (1990) conditions to allow a $\theta-$*dependent* error (i.e., $U_n(Z_n, \theta)$, where U_n is a Borel-measurable mapping in both Z_n and θ). Given the existence of a sieve, that is an increasing sequence of Þnite-dimensional subspaces whose union is dense in the estimation space, our Þnite-dimensional estimation procedure delivers a consistent estimator when the errors form a mixingale process, a condition much weaker than that of Goldstein (1988) and Nixdorf (1984). Section III obtains functional central limit theorems (FCLT's) and asymptotic normality for the sieve-based RM procedures when errors are $\theta-$*dependent* , $H-$valued mixingale processes. Our results include Nixdorf's (1984) result for the $\theta-$*independent* , martingale difference error case as a special example, and we need weaker conditions than he does. Section IV gives laws of iterated logarithm (LIL's) (hence almost-sure loglog rate of convergence results), which is a refinement of the asymptotic normality results. The results for the cases of θ*–dependent* , weakly stationary mixingale error processes are new even for the $I\!R^d$ ($d < \infty$)–valued RM procedures. Section V presents mean rates of convergence under conditions similar to those for the almost-sure convergence, which are weaker than the conditions for the FCLT's and LIL's. Like the previous sections, the results are stated for both the direct H-RM and the sieve-based RM procedures when errors are H-valued mixingale processes. The results for the direct H–RM procedures include Yin & Zhu's (1990) as a special case. The results for the sieve-based RM procedure are to our knowledge new. Section VI is a brief summary and indicates some further research directions. A Mathematical Appendix contains the proofs.

II. ALMOST-SURE CONVERGENCE

We need the following definitions throughout the paper: Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. Let $I\!B$ be a generic real separable Banach space with norm $\|\cdot\|$. Let $\mathcal{B}(I\!B)$ be the Borel σ–field generated by the Borel open sets of $I\!B$. We call $W: \Omega \to I\!B$ a $I\!B$–valued random element ($I\!B$*–r.e.*) if W is $\mathcal{F}/\mathcal{B}(I\!B)$–measurable.

A function $X: \Omega \to I\!B$ is *simple* if for an integer m and each $\omega \in \Omega$

$$X(\omega) = \sum_{i=1}^{m} x_i \, 1_{A_i}(\omega), \quad \text{where } x_i \in I\!B,\ A_i \in \mathcal{B},\ \cup_{1 \le i \le m} A_i = \Omega \text{ and } A_i \cap A_j = \emptyset \text{ if } i \ne j\,.$$

The *Bochner* integral of the simple function X is defined as

$$\int_\Omega X(\omega)\, P(d\omega) = \sum_{i=1}^{m} x_i \, P(A_i)\,.$$

A $I\!B$*–r.e.* W is *Bochner integrable* if there exists a sequence of simple functions $\{ X_n : \Omega \to I\!B \}$ such that

$$X_n \to W \quad a.s.\!-\!P \quad \text{and } \int_\Omega \| X_n(\omega) - W(\omega) \| \, P(d\omega) \to 0 \quad \text{as } n \to \infty\,.$$

The *mathematical expectation* of W in the sense of Bochner (or in the strong sense) is defined as the limit of Bochner integrals of simple functions

$$\int_\Omega W(\omega)\, P(d\omega) \equiv \lim_{n \to \infty} \int_\Omega X_n(\omega)\, P(d\omega)\,.$$

It has been shown that a $I\!B$*–r.e.* W has a mathematical expectation in the sense of Bochner if and only if $E[\|W\|] < \infty$.

Let G and H be real separable Hilbert spaces. H is endowed with inner product $(\cdot,\cdot)$, norm $\|x\| = (x, x)^{1/2}$, and identity operator I. Let $\{ M_n : G \times H \to H,\ n = 1,2,\ldots \}$ be a sequence of $\mathcal{B}(G \times H)/\mathcal{B}(H)$–measurable mappings. Let $Z \equiv \{ Z_n : \Omega \to G\ ;\ n = 1,2,\ldots \}$ be a sequence of $\mathcal{F}/\mathcal{B}(G)$–measurable mappings that is generated by nature, and is not Granger-caused by $\{ \hat{\theta}_n \}$ to be defined below.

ASSUMPTION A.1: Let $M: H \to H$ be a Borel measurable mapping such that:
(1) M has a zero point $\theta_o \in H$, i.e., $M(\theta_o) = 0$.
(2) M is uniformly continuous on any norm-bounded subset of H.

Condition A.1(2) implies that M is continuous and maps bounded sets into bounded sets. Yin and Zhu (1990) give examples of M that satisfy Assumption A.1(2). These include continuous linear operators, Hölder and Lipschitz operators, uniformly continuous operators, continuous operators with "polynomial growth", some compact operators, and others.

ASSUMPTION A.2: $\{ a_n ; n = 1,2,... \}$ is a sequence of nonincreasing positive real numbers such that:

$$(1)\ a_n \to 0 \ \ as\ n \to \infty, \ \ \sum_{n=1}^{\infty} a_n = \infty ; \ \text{ and } (2)\ a_n \le a_{n+1}(1 + b\, a_n) \ \text{ for some } 0 < b \le 1 .$$

A Hilbert space-valued RM procedure (RM) is

$$\hat{\theta}_{n+1} = \hat{\theta}_n + a_n M_n(Z_n, \hat{\theta}_n), \quad n = 1,2,...,$$

where $\hat{\theta}_1$ is an arbitrary H–valued random element, denoted $\hat{\theta}_1$ *arb.* H–*r.e.*. Note that for each n, $\hat{\theta}_n$ is an H–*r.e.* by the measurability of M_n.

DeÞne $U_n : G \times H \to H$ as $U_n(\cdot, \theta) \equiv M_n(\cdot, \theta) - M(\theta)$. Then one can rewrite the H–valued RM procedure as :

$$\hat{\theta}_1 \ \ arb.\ H\text{–}r.e., \quad \hat{\theta}_{n+1} = \hat{\theta}_n + a_n [\, M(\hat{\theta}_n) + U_n(Z_n, \hat{\theta}_n)\,], \quad n = 1,2,...,$$

where $U_n(Z_n, \hat{\theta}_n)$ is an H–*r.e.* by the deÞnition.

Let $\{ \mathcal{F}^n \}$ be a family of increasing sub σ–algebras of $\mathcal{F}$ generated as follows:

$$\mathcal{F}^n = \sigma(\emptyset, \Omega) \ \text{ for } n < 0 ; \quad \mathcal{F}^0 = \sigma(\hat{\theta}_1) ; \quad \mathcal{F}^n = \sigma(Z_j, \hat{\theta}_{j+1} ; j \le n) \ \text{ for } n > 0 .$$

Then $\hat{\theta}_n$ is $\mathcal{F}^{n-1}$ – measurable, while Z_n, $U_n(Z_n, \hat{\theta}_n)$ and $\hat{\theta}_{n+1}$ are $\mathcal{F}^n$ – measurable.

There are various methods used to prove almost sure convergence of $\{ \hat{\theta}_n \}$ to θ_o, and there are a variety of corresponding assumptions on the step size $\{ a_n \}$, the error term $\{ U_n(Z_n, \hat{\theta}_n) \}$, the mapping $M(\cdot)$, and the measurement $\{ M_n(Z_n, \hat{\theta}_n) \}$. Nevertheless, all try to establish the following two essential relations:

(i) $\sup_n \| \hat{\theta}_n \| < \infty$ $a.s.\text{–}P$; and (ii) For every $T > 0$,

$$\lim_{n \to \infty} \sup_{n < i \le m(n,T)} \| \Sigma_{n \le j \le i-1}\, a_j\, [M_j(Z_j, \hat{\theta}_j) - M(\hat{\theta}_j)] \| = 0 \qquad a.s.\text{–}P ,$$

where $m(n,T) \equiv \max \{ i : \Sigma_{n \le j \le i-1}\, a_j \le T \}$.

Kushner & Clark's (1978) "weak convergence" methods typically impose weak conditions on M and U_n to obtain (ii). But they often assume that $\{ M_n(Z_n, \theta) \equiv m(Z_n, \theta) \}$, and either directly require (i) or assume that $\{ \hat{\theta}_n \}$ lies in a compact set $a.s.\text{–}P$. Although the compactness assumption is not very restrictive for $I\!R^d$ $(d < \infty)$–valued RM procedures, it is too strong a requirement for $\{ \hat{\theta}_n \}$, generated by an inÞnite-dimensional Hilbert space-valued RM procedure, to belong to a compact set in the norm topology. Berman and Shwartz (1989) have assumed that $\{\hat\theta_n\}$ lies in a convex, compact set under a topology induced by an invariant metric. They get almost-sure convergence in this metric for a Banach space-valued RM algorithm when $\{U_n\}$ are θ*–independent* errors. Although we can follow their method and get a Banach space extension of Kushner and Clark (1978)'s Theorem 2.4.2 for θ*–dependent* $U_n(\cdot,\cdot)$, we do not adopt this approach here, since we are interested in almost-sure convergence in the norm metric.

Metivier & Priouret's (1984) "martingale" methods explicitly assume the existence of a Liapunov functional, and impose stronger conditions on the growth of M and on the error U_n to obtain (i) and (ii). Some typical assumptions are that $M(\theta)$ grows at most linearly in θ, and $\{U_n\}$ is a martingale difference sequence. An important advance is the paper by Yin & Zhu (1990). They establish almost sure convergence for an H–valued RM procedure with θ*–independent* errors, under weak conditions on the growth of $M(\theta)$ and the average of the errors, without an *a priori* assumption on the uniform boundedness of $\{\hat\theta_n\}$. We shall extend Yin & Zhu's results to include finite-dimensional projected RMs and RM procedures with θ*–dependent* errors. We allow weak conditions on $\{U_n\}$ akin to Kushner and Clark's, but we do not assume a prior compact set to which $\{\hat\theta_n\}$ belongs. Also, we allow $M_n(\cdot,\theta)$ to depend on n.

To establish the uniform boundedness of $\{\hat\theta_n\}$ generated by RM procedures, we consider certain truncated RM procedures. Let $\{B_n,\ n=1,2,...\}$ be a sequence of strictly increasing positive real numbers. Define a sequence of positive integer-valued random variables by

$$T(1) = 1, \quad T(n+1) = T(n) + 1\,(J_n^c),$$

where $1\,(A)$ denotes the indicator of the set $A \in \mathcal{F}$, $J_n \equiv \{\|\hat\theta_n + a_n M_n\| \le B_{T(n)}\}$, and J_n^c is the complement of the set J_n. A truncated RM procedure (TRM) is :

$$\hat\theta_1 \ \ arb.\ \ H\text{–}r.e.\ ,$$
$$\hat\theta_{n+1} = [\hat\theta_n + a_n M_n(Z_n,\hat\theta_n)]\,1\,(J_n) + \bar\theta_n\,1\,(J_n^c)\,, \quad n=1,2,...,$$

where $\{\bar\theta_n\}$ is a sequence of arbitrary fixed elements of H such that $\|\bar\theta_n\| < B$ for all n and for some $0 < B < B_1 < \infty$. One example of $\{\bar\theta_n\}$ is as follows: Let $\bar\theta^j$, $j=1,2$ be two arbitrary fixed points of H, with $\bar\theta^1 \ne \bar\theta^2$, and $\|\bar\theta^j\| < B$. Put

$$\bar\theta_n = \bar\theta^1 \ \text{ if } \ T(n) = 2j\,; \quad \bar\theta_n = \bar\theta^2 \ \text{ if } \ T(n) = 2j-1\,; \qquad j=1,2,3,...\,.$$

More generally, we take $\{\bar\theta_n\}$ to be a sequence of arbitrary H–*r.e.* generated by nature, independent of $\hat\theta_1$ and $\{Z_n\}$, such that $\|\bar\theta_n\| < B$ *a.s.–P* for all n. For this we set $\mathcal{F}^n = \sigma(Z_j, \bar\theta_j, \hat\theta_{j+1};\ j \le n)$ for $n > 0$. For example, set $\bar\theta_n \equiv \bar\theta$, for $\bar\theta$ *arb. H–r.e.*, independent of $\hat\theta_1$, with $\|\bar\theta\| < B$ *a.s.–P*.

Depending on whether or not we have prior information on the bounded region to which θ_o belongs, we consider two situations:

(1) If there is no prior information on where θ_o belongs, we adopt a "randomly truncated RM procedure" (RTRM) by choosing $\{B_n\}$ such that $\lim_{n\to\infty} B_n = \infty$. This includes Yin and Zhu's (1990) algorithm as a special case. Yin and Zhu (1990) only consider the case in which $\bar\theta_n = \bar\theta$ for all n. By considering a more general scheme, we lessen the possibility of the algorithm following similar paths to undesirable regions of H.

(2) If there is prior information on where θ_o belongs, e.g., $\|\theta_o\| < B_o < B$, we consider a "bounded truncated RM procedure" (BTRM) by choosing $\{ B_n ; n=1,2,... \}$ such that $\lim_{n\to\infty} B_n = \bar{B} < \infty$ with $\bar{B} > B$. For example, we can choose $B_n = B + 10 \sum_{j=1}^{n} 2^{-j}$ with $\bar{B} = B + 20$.

From now on, we use TRM to denote that conditions or results hold for both RTRM and BTRM.

ASSUMPTION A.3: There is a twice continuously Fréchet differentiable functional $V : H \to I\!R$ such that

$$(1) \quad V(\theta_o) = 0 ; \quad \lim_{\|\theta\|\to\infty} V(\theta) = \infty ; \quad V(\theta) > 0 , \quad (V'(\theta) , M(\theta)) \le 0 \quad \text{for } \theta \ne \theta_o .$$

$$(2) \quad \text{for any } \eta > 0 , \quad \inf [- (V'(\theta) , M(\theta)) : \|\theta - \theta_o\| \ge \eta] > 0 ,$$

where V' denotes the first Fréchet derivative of V.

This assumes the existence of a Liapunov functional V, which implies the asymptotic stability of the solution θ_o for the nonstochastic, Hilbert-space valued ODE $\dot{\theta}(t) = M(\theta(t))$. Note that Assumption A.3 implies that V maps bounded subsets of H into bounded subsets of $I\!R$. When M is Fréchet differentiable at θ_o with first Fréchet derivative A, we can choose V to be a local quadratic form

$$V(\theta) = (\theta - \theta_o , A (\theta - \theta_o)) + o (\|\theta - \theta_o\|^2) .$$

However, as we do not impose differentiability on M, we only assume the existence of V.

LEMMA 2.1: Given TRM with Assumptions A.1, A.2 and A.3(1) holding , if there exists $0 \le \varepsilon < 1$ such that

$$\limsup_{n\to\infty} \| a_n \sum_{j=1}^{n} [M_j(Z_j,\hat{\theta}_j) - M(\hat{\theta}_j)] \| = \varepsilon \qquad a.s.-P ,$$

then there exists a positive integer-valued random variable T such that $P(\sup_n T(n) \le T < \infty) = 1$.

This lemma demonstrates that $\{\hat{\theta}_n\}$ generated by RTRM or BTRM becomes bounded for all large n, i.e., the truncation is only invoked a finite number of times. Therefore, the asymptotic properties of the RTRM and BTRM are the same as that of the original RM with $\sup_n \|\hat{\theta}_n\| < \infty$ $a.s.-P$. This result provides the fundamental property on which all of our subsequent results for the truncated methods rest.

To get an almost sure consistency result, we need stronger assumptions than those in Lemma 2.1.

ASSUMPTION A.4:

$$\limsup_{n\to\infty} \| a_n \sum_{j=1}^{n} U_j(Z_j,\hat{\theta}_j) \| = 0 \qquad a.s.-P .$$

By Kronecker's Lemma and Assumption A.2(1), Assumption A.4 is implied by

$$\limsup_{n\to\infty} \| \sum_{j=1}^{n} a_j U_j(Z_j,\hat{\theta}_j) \| < \infty \qquad a.s.-P \ .$$

The following consistency result is proven by following the proof of Theorem 3.2 of Yin and Zhu (1990).

THEOREM 2.2: If TRM and Assumptions A.1 - A.4 hold, then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

Note that our error terms $\{ U_j \}$ can depend on $\{ \hat{\theta}_j \}$, while the error terms in Yin and Zhu (1990)'s paper are independent of $\{ \hat{\theta}_j \}$. In their case, Yin and Zhu (1990) have shown that the error condition A.4 is a sort of necessary and sufficient condition for the almost sure convergence result to occur, given the other assumptions . This is true here as well, a fortiori, but we will not give a formal statement of this fact. Yin and Zhu show that martingale difference, moving average, and stationary ϕ –mixing error sequences $\{ U_n \}$ satisfy assumption A.4. Here we provide a weaker set of sufficient conditions for A.4 based on Chen and White (1996)'s results on H-valued, L_p-mixingale processes. Let $\|\cdot\|_p$ denote the L_p –norm for an $H-r.e.$ X, $\|X\|_p \equiv [E\|X\|^p]^{1/p}$, $1 \le p < \infty$. Let $\{ W_n ; -\infty < n < \infty \}$ be a sequence of $H-r.e.\acute{}s$ with finite L_p –norms, $1 \le p < \infty$. Let $\{ \mathcal{A}^n \}$ be a filtration of $\mathcal{F}$. Then $\{ W_n , \mathcal{A}^n \}$ is an H –valued L_p *–mixingale* sequence if there exist sequences of finite nonnegative constants $\{ c_n ; n \ge 1 \}$ and $\{ \psi_m ; m \ge 0 \}$ with $\psi_m \to 0$ as $m \to \infty$ such that the following two inequalities hold for all $n \ge 1$, $m \ge 0$:

$$\| E (W_n \mid \mathcal{A}^{n-m}) \|_p \le \psi_m c_n ;$$

$$\| W_n - E (W_n \mid \mathcal{A}^{n+m}) \|_p \le \psi_{m+1} c_n .$$

If, in addition, W_n is $\mathcal{A}^n$ –measurable, then $\{ W_n , \mathcal{A}^n \}$ is an *adapted* H –valued L_p –mixingale sequence.

An H –valued L_p –mixingale with $1 \le p < \infty$ has zero mean. We can choose $\{ \psi_m ; m \ge 0 \}$ to be non-increasing in m when $\{ W_n , \mathcal{A}^n \}$ is an adapted L_p – mixingale ($p \ge 1$). We say that ψ_m is of size $-a$ if $\sum_{m=0}^{\infty} [\psi_m]^\delta < \infty$ or $\psi_m = o (m^{-1/\delta})$ for some $a < (1/\delta)$ or $\psi_m = O (m^\lambda)$ for some $\lambda < -a$.

θ*–independent* error case : Suppose for each $\theta \in H$, $U_n(Z_n,\theta) \equiv F_n(Z_n)$, where $F_n : G \to H$ is a Borel measurable mapping, $n = 1,2,\ldots$.

ASSUMPTION A.5: $\{ F_n(Z_n) , \mathcal{F}^n \}$ is an H –valued L_p –mixingale, $1 < p < \infty$, with parameters $\{ \psi_m \}$ and $\{ c_n \}$ satisfying either

$$(i) \ \sum_{i=1}^{\infty} (c_i\, a_i)^2 < \infty \ \text{ and } \ \sum_{m=1}^{\infty} (\psi_m)^2 < \infty \ \text{ if } p \ge 2; \quad \text{or}$$

$$(ii) \ \sum_{i=1}^{\infty} (c_i a_i)^p < \infty \text{ and } \sum_{m=1}^{\infty} \psi_m < \infty \ \text{ if } 1 < p < 2 \ .$$

The following corollary is a consequence of Theorem 2.2 and Chen and White's (1996) Corollaries 3.8 and 3.9.

COROLLARY 2.3: Let $\{ \hat{\theta}_n \}$ be given by RTRM . Suppose Assumptions A.1-A.3 and A.5 hold for θ –*independent* errors $\{ U_n \}$. Then $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.-P$.

Gyorfi and Masry (1990) establish consistency and convergence rates for a class of recursive kernel estimators for $Z_n \in I\!R^d$ $(d < \infty)$. A consequence of Corollary 2.3 is an extension of their consistency result to $Z_n \in G$, where G is not necessarily a finite dimensional space and $\{ F_n(Z_n) \}$ is not necessarily an L_2 –mixingale.

ASSUMPTION A.6: $\{ F_n(Z_n) , \mathcal{F}^n \}$ is an adapted H –valued L_p –mixingale , $p \geq 1$ with $\sup_n E[\| F_n(Z_n) \|^r] < \infty$ for some $r \geq p$, $r > 1$ and $\psi_m = O((\log m)^{-2-\beta})$ for some small $\beta > 0$.

The following corollary is a simple consequence of Theorem 2.2 and Chen and White's (1996) Theorem 3.10.

COROLLARY 2.4: Let $\{ \hat{\theta}_n \}$ be given by RTRM . Suppose Assumptions A.1-A.3 and A.6 hold for θ –*independent* errors $\{ U_n \}$. If $a_n = O(n^{-1} \log n)$, then $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.-P$.

θ –*dependent* error case : The error terms U_n are influenced by both Z_n and $\hat{\theta}_n$, where U_n is $\mathcal{B}(G \times H) / \mathcal{B}(H)$ –measurable.

First we adapt the "martingale" approach to get almost-sure convergence results for the RM procedure by imposing stronger versions of Assumptions A.1 - A.3. Some common conditions are : (a) $M(\theta)$ and $M_n(Z_n, \theta)$ ($a.s.-P$) grow at most linearly in θ ; (b) the Liapunov functional is taken to be $V(\theta) \equiv \| \theta - \theta_o \|^2$; (c) some control is imposed on the noise via $\{ a_n \}$; or (d) some restrictions are imposed on the inner products of both $(V'(\cdot) , M(\cdot))$ and $(V'(\cdot) , U_n(Z_n, \cdot))$.

If we do not impose linear growth restrictions on M_n and M , we need to assume that $\| \theta_o \| < B_o < B$ and use the BTRM. The key point here is that truncations are only invoked finitely many times, given proper control of the growth of the average of the errors (via Lemma 2.1). The following three corollaries yield almost-sure convergence results of this sort by providing sufficient conditions for Assumption A.4.

COROLLARY 2.5: Let $\{ \hat{\theta}_n \}$ be given by BTRM. Suppose Assumptions A.1 - A.3 and the following conditions hold :

(1) There exist sequences of mappings $\{ \overline{M}_n : H \to H \}$ and $\{ s_n : H \to (0,\infty) \}$ such that, for each n , $\overline{M}_n$ is $\mathcal{B}(H) / \mathcal{B}(H)$ –measurable, s_n is $\mathcal{B}(H) / \mathcal{B}(I\!R)$ –measurable, and for any ξ that is $\mathcal{F}^n$ –measurable,

$$E[M_n(Z_n, \xi) \mid \mathcal{F}^n] = \overline{M}_n(\xi) \quad a.s.-P \quad ; \qquad E[\| M_n(Z_n, \xi) \|^2 \mid \mathcal{F}^n] = s_n(\xi) \quad a.s.-P \ .$$

For any $0 < K < \infty$ and for all n , define

$$b_{K,n} \equiv \sup_{\| \theta \| \leq K} \| \overline{M}_n(\theta) - M(\theta) \| \ , \qquad c_{K,n} \equiv \sup_{\| \theta \| \leq K} [s_n(\theta)] \ .$$

(2) $\lim_{n\to\infty} b_{K,n} = 0$; $\sum_{j=1}^{\infty} (a_j\, b_{K,j}) < \infty$.

(3) $c_{K,n} < \infty$; $\sum_{j=1}^{\infty} (a_j^2\, c_{K,j}) < \infty$.

Then $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.-P$.

Remark: Here we can identify

$$U_n(Z_n,\hat{\theta}_n) \equiv [\, \overline{M}_n(\hat{\theta}_n) - M(\hat{\theta}_n)\,] + [\, M_n(Z_n,\hat{\theta}_n) - \overline{M}_n(\hat{\theta}_n)\,] \;.$$

Condition 2.5(1) assumes that both $M_n(Z_n,\hat{\theta}_n)$ and $\| M_n(Z_n,\hat{\theta}_n) \|^2$ have a kind of Markovian structure, which is a typical assumption in the stochastic approximation literature. This corollary includes the classical Robbins-Monro algorithm as a special case. Metivier and Priouret (1984) present an $I\!R^d$ $(d<\infty)$ –valued classical RM algorithm of this type without truncation. They choose a quadratic Liapunov functional $V(\theta) = \| \theta - \theta_o \|^2$. Under an additional condition on the growth of s_n , namely $s_n(\theta) \le c_n\, (1 + \| \theta - \theta_o \|^2)$, they get almost sure convergence. Our corollary relaxes these conditions and extends their result to H , although we need truncation but they do not.

The following is a less abstract sufficient condition for Assumption A.4:

ASSUMPTION A.7:

(1) (a) For any $\theta \in H$ and all n , $E\,[\, U_n(\cdot, \theta)\,] = 0$; and (b) For any $K > 0$,

$$\lim_n \sup_{\|\theta\| \le K} \| a_n \sum_{j=1}^{n} U_j(Z_j,\theta) \| = 0 \;\; a.s.-P \;.$$

(2) There exists a sequence of $\mathcal{B}(G) / \mathcal{B}(I\!R)$ –measurable functions $\{ h_n : G \to [0,\infty) \}$ and a constant $h < \infty$ such that:

(a) $\sup_n E\,[h_n(Z_n)] \le h$;

$$\lim_n a_n \sum_{j=1}^{n} (\, h_j(Z_j) - E\,[\, h_j(Z_j)\,]\,) = 0 \quad a.s.-P \quad ;$$

(b) for all $z \in G$, $\theta, \theta' \in H$,

$$\| U_n(\, z , \theta\,) - U_n(\, z , \theta'\,) \| \le h_n(z) \| \theta - \theta' \| .$$

(3) For any $K > 0$, there exists a sequence of $\mathcal{B}(G) / \mathcal{B}(I\!R)$ –measurable functions $\{ g_{K,n} : G \to [0,\infty) \}$ and a constant $g_K = O\,(\,K\,)$ such that:

(a) $\sup_n E\,[\, g_{K,n}(Z_n)\,] \le g_K$; and

$$\lim_n a_n \sum_{j=1}^{n} (g_{K,j}(Z_j) - E\,[\, g_{K,j}(Z_j)\,]\,) = 0 \quad a.s.-P \quad ;$$

(b) $\sup_{\|\theta\| \le K} \| U_n(Z_n,\theta) \| \le g_{K,n}(Z_n)$ $a.s.-P$.

(4) $\sup_{\|\theta\| \le K} \| M(\theta) \| = O(K)$.

This assumption is satisfied by many dependent Hilbert space-valued random sequences.

COROLLARY 2.6: Let $\{ \hat{\theta}_n \}$ be given by the BTRM with $B_{n+1} \ge B_n (1 + c\, a_n)$ for some positive number c . Suppose Assumptions A.1 - A.3 and A.7 hold when $\{ U_n \}$ are θ *–dependent* errors. Then $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.{-}P$.

This corollary includes linear algorithms as special cases. It also includes an $I\!R^d$ $(d < \infty)$ -valued RM algorithm driven by an ergodic process as in Metivier and Priouret (1984).

COROLLARY 2.7: Let $\{ \hat{\theta}_n \}$ be given by the BTRM with $B_{n+1} \ge B_n (1 + c\, a_n)$ for some positive number c . Suppose Assumptions A.1 - A.3, A.7(2)(b), A.7(3)(b), and A.7(4) hold when $\{ U_n \}$ are θ *–dependent* errors. Suppose further the following conditions hold:

(1) For each $\theta \in H$, $\{ U_n(Z_n,\theta) , \mathcal{F}^n \}$ is an adapted H –valued L_2 –mixingale sequence with $\{ \psi_m \}$ of size $-1/2$ and $\sum_{n=1}^{\infty} (a_n c_n)^2 < \infty$.

(2) $\sup_n \| h_n(Z_n) \|_2 < \infty$ and $\{ h_n(Z_n) - E[h_n(Z_n)] , \mathcal{F}^n \}$ is an $I\!R$ –valued L_2 –mixingale with $\{ \psi_m \}$ and $\{ c_n \}$ as in (1) .

(3) $\sup_n \| g_{K,n}(Z_n) \|_2 < \infty$ and $\{ g_{K,n}(Z_n) - E[g_{K,n}(Z_n)] , \mathcal{F}^n \}$ is an $I\!R$ –valued L_2 –mixingale with $\{ \psi_m \}$ and $\{ c_n \}$ as in (1) .

Then $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.{-}P$.

Note that the conditions in this corollary are similar to Kuan & White's (1994) sufficient conditions for almost sure convergence of an $I\!R^d$ $(d < \infty)$ –valued RM with θ *–dependent* errors.

For estimation purposes, we consider finite dimensional approximation modifications of the above RM and TRM procedures. The separability of H implies the existence of a complete orthonormal basis $\{ e_j , j = 1,2,\ldots \}$. Let $P_{k(n)} : H \to H_{k(n)}$ be the orthogonal projection of H onto $H_{k(n)}$, where $H_{k(n)}$ is the closure of linear space spanned by $(e_1 , \ldots , e_{k(n)})$, and $\{ k(n), n = 1,2,\ldots \}$ is an integer-valued sequence such that:

$$1 \le k(n) \le k(n+1) \le k(n)+1 \; ; \quad \text{and} \qquad \lim_{n \to \infty} k(n) = \infty .$$

From this it follows that $\dim(H_{k(n)}) = k(n) \le n$.

An RM procedure with orthonormal projection (RMP) can be defined as

$$\hat{\theta}_1 \;\; arb. \;\; H_{k(1)} - r.e. \; ,$$

$$\hat{\theta}_{n+1} = [\hat{\theta}_n + a_n P_{k(n+1)} M_n(Z_n, \hat{\theta}_n)] \; .$$

A truncated RM with projection (TRMP) is

$$\hat{\theta}_1 \;\; arb. \;\; H_{k(1)} - r.e. \; ,$$

$$\hat{\theta}_{n+1} = [\hat{\theta}_n + a_n P_{k(n+1)} M_n(Z_n, \hat{\theta}_n)] \, 1(J_n) + \theta^*_{n+1} \, 1(J_n^c) \; ,$$

where $J_n \equiv \{ \|\hat{\theta}_n + a_n P_{k(n+1)} M_n \| \le B_{T(n)} \}$; $J_n^c \equiv \{ \|\hat{\theta}_n + a_n P_{k(n+1)} M_n \| > B_{T(n)} \}$. Here $\{ \theta^*_{n+1} \}$ is either a sequence of arbitrary fixed (nonrandom) elements in $H_{k(n+1)}$, with $\|\theta^*_{n+1}\| < B$ for all n; or a sequence of arbitrary $H_{k(n+1)}$ –r.e.´s which are generated by nature and independent of $\hat{\theta}_1$ and $\{ Z_n \}$, with $\|\theta^*_{n+1}\| < B$ $a.s.-P$ for all n.

Again, depending on whether or not there is prior information on where θ_o belongs, we can specify TRMP to be either a "randomly truncated RM with projection" (RTRMP) or a "bounded truncated RM with projection" (BTRMP).

The following Assumptions A.3P-A.7P play the same roles for RMP and TRMP as do Assumptions A.3-A.7 for RM and TRM :

ASSUMPTION A.3P: There is a twice continuously Fréchet differentiable functional $V : H \to I\!R$ such that

(1) $V(\theta_o) = 0$; $\lim_{\|\theta\|\to\infty} V(\theta) = \infty$; and $V(\theta) > 0$, $(V'(\theta), P_{k(n)} M(\theta)) \le 0$, for all $\theta \in H_{k(n)}$, $\theta \ne \theta_o$;

(2) There exists a finite integer N_o such that for all $n \ge N_o$, for any $\eta > 0$,

$$\inf [-(V'(\theta), P_{k(n)} M(\theta)) : \|\theta - \theta_o\| \ge \eta ,\ \theta \in H_{k(n)}] > 0 .$$

ASSUMPTION A.4P:

$$\limsup_{n\to\infty} \| a_n \sum_{j=1}^{n} P_{k(j+1)} U_j(Z_j, \hat{\theta}_j) \| = 0 \quad a.s.-P.$$

ASSUMPTION A.5P: $\{ P_{k(n+1)} F_n(Z_n) , \mathcal{F}^n \}$ is an adapted H –valued L_p –mixingale with $1 < p < \infty$, and with parameters $\{ \psi_m \}$ and $\{ c_n \}$ as in A.5.

ASSUMPTION A.7P: Assumption A.7 holds with U_n replaced by $P_{k(n+1)} U_n$.

It is easy to prove that A.5 implies A.5P and A.7 implies A.7P .

We can now state some consistency results for TRMP.

THEOREM 2.8 (Corresponding to Theorem 2.2) : Given Assumptions A.1, A.2, A.3P, A.4P, let $\{ \hat{\theta}_n \}$ be given by the TRMP. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

COROLLARY 2.9 (Corresponding to Corollary 2.3) : Let $\{ \hat{\theta}_n \}$ be given by RTRMP when $\{ U_n \}$ are θ *–independent* errors. Suppose Assumptions A.1, A.2, A.3P, and A.5P hold. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

COROLLARY 2.10 (Corresponding to Corollary 2.6) : Let $\{ \hat{\theta}_n \}$ be given by the BTRMP with $B_{n+1} \ge B_n (1 + c\, a_n)$ for some positive number c . Suppose A.1, A.2, A.3P and A.7P hold when $\{ U_n \}$ are θ *–dependent* errors. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

Assumption A.3P plays an important role for the exact analog of Corollaries 2.8, 2.9 and 2.10 to Theorem 2.2, Corollaries 2.3 and 2.6, respectively. Since A.3P(2) might be unduly restrictive, now we consider a weaker assumption:

ASSUMPTION A.8: There is a twice continuously Fréchet differentiable functional $V : H \to I\!R$, and there is a sequence of elements $\{ \theta^o_n \in H_{k(n)} \}$ such that :

(1) There exists a finite integer N_o such that for all $n \ge N_o$, for any $\eta > 0$,

$$\inf [-(V'(\theta) , P_{k(n)}M(\theta)) : \|\theta - \theta^o{}_n\| \geq \eta \ , \ \theta \in H_{k(n)}] > 0 .$$

(2) $\sum_{n=1}^{\infty} a_n \|\theta^o{}_{n+1} - \theta_o\| < \infty$.

THEOREM 2.11 (Corresponding to Theorem 2.2) : Given Assumptions A.1, A.2, A.3P(1), A.4P and A.8, let $\{ \hat{\theta}_n \}$ be given by TRMP. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

COROLLARY 2.12 (Corresponding to Corollary 2.9) : Let $\{ \hat{\theta}_n \}$ be given by RTRMP when $\{ U_n \}$ are θ *–independent* errors. Suppose Assumptions A.1, A.2, A.3P(1), A.5P and A.8 hold. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

COROLLARY 2.13 (Corresponding to Corollary 2.10) : Let $\{ \hat{\theta}_n \}$ be given by the BTRMP with $B_{n+1} \geq B_n (1 + c\, a_n)$ for some positive number c . Suppose A.1, A.2, A.3P(1), A.7P and A.8 hold when $\{ U_n \}$ are θ *–dependent* errors. Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

The comparison of Assumptions A.3P and A.8 remains valid for the "martingale" approach. Here we present two propositions as an illustration. Again, the linear growth restriction on $P_{k(n+1)}M_n$ and the quadratic form of V guarantee the $a.s.-P$ uniformly boundedness of $\{ \hat{\theta}_n \}$ generated by RMP; hence, no truncation is needed.

PROPOSITION 2.14: Let $\{ \hat{\theta}_n \}$ be given by RMP satisfying $E[\|\hat{\theta}_1 - \theta_o\|^2] < \infty$ when $\{ U_n \}$ are θ *–dependent* errors. Suppose A.1(1), A.2(1) and the following two conditions hold:

(1) For any $\eta > 0$,

$$\liminf_n (\inf [-(\theta - \theta_o , P_{k(n+1)}M_n(Z_n,\theta)) : \|\theta - \theta_o\| \geq \eta \ , \ \theta \in H_{k(n+1)}]) > 0 \quad a.s.-P .$$

(2) There exist sequences of $\mathcal{B}(G)$ –measurable functions $\{ h_n : G \to [0,\infty) \}$ and $\{ g_n : G \to [0,\infty) \}$ such that,

(a) for each $z \in G$, $\theta \in H_{k(n+1)}$, $\|P_{k(n+1)}M_n(z,\theta)\|^2 \leq h_n(z) + g_n(z)\|\theta - \theta_o\|^2$

(b) $\sum_{n=1}^{\infty} a_n^2 E[h_n(Z_n) \mid \mathcal{F}^{n-1}] < \infty$ $a.s.-P$; and $\sum_{n=1}^{\infty} a_n^2 E[g_n(Z_n) \mid \mathcal{F}^{n-1}] < \infty$ $a.s.-P$.

Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

PROPOSITION 2.15: Let $\{ \hat{\theta}_n \}$ be given by RMP satisfying $E[\|\hat{\theta}_1 - P_{k(1)}\theta_o\|^2] < \infty$ when $\{ U_n \}$ are θ *–dependent* errors. Suppose A.1(1), A.2(1) and the following three conditions hold:

(1) For any $\eta > 0$,

$$\liminf_n (\inf [-(\theta - P_{k(n+1)}\theta_o , M_n(Z_n,\theta)) : \|\theta - P_{k(n+1)}\theta_o\| \geq \eta \ , \ \theta \in H_{k(n+1)}]) > 0 \quad a.s.-P .$$

(2) There exist sequences of $\mathcal{B}(G)$ –measurable functions $\{ h_n : G \to [0,\infty) \}$ and $\{ g_n : G \to [0,\infty) \}$ such that,

(a) for each $z \in G$, $\theta \in H_{k(n+1)}$, $\|P_{k(n+1)}M_n(z,\theta)\|^2 \leq h_n(z) + g_n(z)\|\theta - P_{k(n+1)}\theta_o\|^2$.

$(b)\quad \sum_{n=1}^{\infty} a_n^2\, E[\, h_n(Z_n) \mid \mathcal{F}^{n-1}\,] < \infty \quad a.s.-P$; and $\sum_{n=1}^{\infty} a_n^2\, E[\, g_n(Z_n) \mid \mathcal{F}^{n-1}\,] < \infty \quad a.s.-P$.

(3) $\sum_{n=1}^{\infty} \| P_{k(n)}\theta_o - P_{k(n+1)}\theta_o \| < \infty$.

Then $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

Remarks: (i) In these two propositions, $V(\theta) \equiv \|\theta - \theta_o\|^2/2$, and $\theta^o{}_n = P_{k(n)}\theta_o$ in Proposition 2.15. (ii) Since for any $\theta \in H_{k(n+1)}$,

$$\|\theta - P_{k(n+1)}\theta_o\|^2 + \|(I - P_{k(n+1)})\theta_o\|^2 \equiv \|\theta - \theta_o\|^2 \qquad \text{and}$$

$$(\theta - P_{k(n+1)}\theta_o\, ,\, M_n(\cdot\, ,\theta)) = (\theta - \theta_o\, ,\, P_{k(n+1)}M_n(\cdot\, ,\theta)) \ ,$$

condition 2.14(1) implies 2.15(1). Since $\lim_{n\to\infty} k(n) = \infty$, $\lim_{n\to\infty} \|\theta_o - P_{k(n+1)}\theta_o\|^2 = 0$, 2.14(2) and 2.15(2) are equivalent.

III. ASYMPTOTIC NORMALITY

This section presents asymptotic normality results for the RMP and TRMP algorithms. Nixdorf (1984) has obtained an asymptotic normality result for an RMP algorithm when $\{ U_n \}$ is a sequence of H–valued martingale difference, θ–*independent* errors. Here we follow Nixdorf's approach, but relax his noise conditions to allow $\{ U_n \}$ to be a sequence of H–valued mixingale θ–*dependent* errors. Our improvements are applications of Walk's (1987) results and Chen and White's (1998a) new central limit theorems for near epoch dependent functions of H–valued mixing processes.

In this section, we always assume that A.1, $a_n = 1/n$, A.3P and A.4P hold. Hence, $a.s.-P$, $\{ \hat{\theta}_n \}$ converges to θ_o in norm, and is uniformly bounded for all $n \geq n_o$. For simplicity and without loss of generality, we set $n_o = 1$ from now on. Also, we only consider the θ–*dependent* error case since it includes θ–*independent* errors as a special case. Thus, we study limiting distribution properties of the following algorithm :

$$\hat{\theta}_{n+1} = \hat{\theta}_n + n^{-1} P_{k(n+1)} M(\hat{\theta}_n) + n^{-1} P_{k(n+1)} U_n(Z_n, \hat{\theta}_n) \ .$$

We begin with the following smoothness condition.

ASSUMPTION B.1: $M(\cdot)$ is continuously Fréchet differentiable at θ_o with ﬁrst derivative A .

Note that whereas M could be any uniformly continuous (possibly nonlinear) operator on any bounded set in the previous section, the present assumption requires M to be a locally linear operator around the true root θ_o . Under this local smoothness assumption, RMP or TRMP can be translated into a recursive procedure of Fabian's (1968):

$$\hat{\theta}_{n+1} - \theta_o = (\, I + n^{-1} A_n\,)\,(\hat{\theta}_n - \theta_o) + n^{-(1+\beta)/2}\, \nu_n + n^{-1-(\beta/2)}\, T_{1n} + n^{-1-(\beta/2)}\, T_{2n} \ ,$$

where $\beta > 0$,

$$A_n \equiv A + P_{k(n+1)} \left(F[\hat{\theta}_n] - F[\theta_o] \right) ,$$

with $F : H \to L(H,H)$, such that for any $y \in H$,

$$F[\theta_o](y) = A\,y \qquad \text{for } \theta = \theta_o \quad ,$$

$$F[\theta](y) = A\,y + [\, M(\theta) - A\,(\theta-\theta_o)\,]\,(\theta-\theta_o\,,y\,)\,/\,(\theta-\theta_o\,,\theta-\theta_o\,) \qquad \text{for } \theta \neq \theta_o\,;$$

$$\nu_n \equiv n^{(\beta-1)/2}\, P_{k(n+1)}\, [\, U_n(Z_n,\theta_o) - EU_n(Z_n,\theta_o)\,]\,;$$

$$T_{1n} \equiv n^{\beta/2}\, P_{k(n+1)}\, [\, U_n(Z_n,\hat{\theta}_n) - U_n(Z_n,\theta_o) + EU_n(Z_n,\theta_o)\,]\ ;$$

$$T_{2n} \equiv n^{\beta/2}\, (\, P_{k(n+1)}\, A - A\,)\,(\,\hat{\theta}_n - \theta_o\,)$$

$$\equiv n^{\beta/2}\,(\, P_{k(n+1)} - I\,)\, A\, P_{k(n+1)}(\,\hat{\theta}_n - \theta_o\,) - n^{\beta/2}\,(\, P_{k(n+1)} - I\,)\, A\,(\,\theta_o - P_{k(n+1)}\theta_o\,)\,.$$

Under Assumption B.1, F is continuous at θ_o . If in addition, we have $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$, then $\|A_n - A\| \to 0$ $a.s.-P$.

ASSUMPTION B.2: $\sigma^* \equiv \inf\,[\,\mathrm{Re}\,\gamma : \gamma \in \operatorname{spec}(-A)\,] > \beta/2$, where $\operatorname{spec}(-A)$ is the spectrum of $-A$. Equivalently, for any $u \in (0,\infty)$, $\|\exp(\,u\,A\,)\| < \exp(\,-u\,\sigma^*\,)$ with $\sigma^* > \beta/2$.

This is a stability assumption. Under this assumption, the solution to the differential equation should be insensitive to "small" perturbations of A_n , T_{1n} , T_{2n} and the partial sum $\sum_{j=1}^{n} \nu_j$. Given $T_{1n} \to 0$, $T_{2n} \to 0$ in a suitable sense, results like the central limit theorem (CLT), functional central limit theorem (FCLT), and law of iterated logarithm (LIL) for the sequence $\{\,\hat{\theta}_n\,\}$ are consequences of the corresponding results for the partial sum $\sum_{j=1}^{n} \nu_j$. Berger’s (1986) and Walk’s (1987) results for general Fabian-type recursive schemes in a Banach space are applicable.

Assumption B.2 requires that the spectrum of A is contained in $\{\gamma :\ \mathrm{Re}\,\gamma < -\sigma^*\}$, which is in turn contained in $\{\gamma :\ \mathrm{Re}\,\gamma < -\beta/2\}$. Hence A cannot be a compact operator when H is an inÞnite-dimensional Hilbert space or any general inÞnite-dimensional Banach space. This means that A cannot be the operator norm-limit of any sequence of Þnite rank operators . In particular, this assumption rules out that $\|P_{k(n+1)}\, A - A\| \to 0$ as $n \to \infty$. Nevertheless, we can always rewrite A in the form $a\,I + b\,C$, where $a \neq 0$, a , b are constants, and C is a compact operator.

Before proceeding to asymptotic normality results obtained using Walk (1977,1987) and Berger (1986), we need assumptions ensuring that " $T_{1n} \to 0$, $T_{2n} \to 0$ in a suitable sense," and that " the partial sum $\sum_{j=1}^{n} \nu_j$ follows a FCLT."

ASSUMPTION B.3:

(1) $\| (I - P_{k(n)})\, A\, P_{k(n)} \| \to 0$ as $n \to \infty$.

(2) $n^{\beta} \| \theta_o - P_{k(n)}\theta_o \|^2 \to 0$ as $n \to \infty$ for some $\beta \in (0, 1]$.

Note that B.3(2) relaxes Nixdorf's (1984) requirement that $\beta = 1$. Assumption B.3 and $\| \hat{\theta}_n - \theta_o \| \to 0$ *a.s.–P* imply that $T_{2n} \to 0$ as $n \to \infty$ *almost in first mean,* i.e., for any $\varepsilon > 0$, there exists $\Omega' \in \mathcal{F}$ with $P(\Omega') \geq 1 - \varepsilon$ such that $\int_{\Omega'} \| T_{2n} \| \, dP \to 0$ as $n \to \infty$. This notion of convergence is weaker than convergence *a.s.* and convergence in first mean, but stronger than convergence in probability. (See Nixdorf (1984), Berger (1986), Walk (1987).)

ASSUMPTION B.4: Let $\delta_o \equiv \max(0, (1+\beta)/2 - \sigma^*)$.

(1) There exists $\delta \in (\delta_o, 1/2)$ such that

$$n^{-3/2} \sum_{l=1}^{n} (n/l)^{\delta+1} \| \sum_{j=1}^{l} \nu_j \| = O_P(1) .$$

(2) There exists $\delta^* \in (\delta_o, 1/2)$ such that

$$n^{-1/2} \max_{1 \leq l \leq n} (n/l)^{\delta^*} \| \sum_{j=1}^{l} j^{-1/2} T_{1j} \| \to 0 \quad \textit{in Prob.} \quad \text{as } n \to \infty .$$

A sequence of $\mathbb{B}$*–r.e.'s* $\{ W_n \}$ *converges in distribution* to a $\mathbb{B}$*–r.e.* W ($W_n \Rightarrow W$ in $\mathbb{B}$) if the sequence of distributions of $\{ W_n \}$ *converges weakly* to the distribution of W on $\mathbb{B}$, i.e., if for all bounded and (norm -) continuous functionals $F : \mathbb{B} \to \mathbb{R}$,

$$\lim_{n \to \infty} \int_B F(x)\, \eta_n(dx) = \int_B F(x)\, \eta(dx) ,$$

where the distributions $\{ \eta_n \}$ and η are given by, for any $\Delta \in \mathcal{B}(\mathbb{B})$,

$$\eta_n(\Delta) \equiv P[\omega \in \Omega : W_n(\omega) \in \Delta] \quad \text{and} \quad \eta(\Delta) \equiv P[\omega \in \Omega : W(\omega) \in \Delta] .$$

Let $C_H[0,1]$ denote the space of continuous mappings from $[0,1]$ to H with sup-norm $\| X \|_\infty \equiv \sup\{ \| X(t) \| : t \in [0,1] \}$. This is a real separable Banach space. Let $X = \{ X(t) ; t \in [0,1] \}$ be a $C_H[0,1]$ – valued random element. Define the $C_H[0,1]$ – valued random elements $X_n = \{ X_n(t) ; t \in [0,1] \}$, $Y_n = \{ Y_n(t) ; t \in [0,1] \}$, and $Y = \{ Y(t) ; t \in [0,1] \}$ as follows:

$$X_n(t) \equiv n^{-1/2} \sum_{j=1}^{[nt]} \nu_j + n^{-1/2} (nt - [nt])\, \nu_{[nt]+1} , \quad t \in [0,1] ;$$

$$Y_n(t) \equiv n^{-1/2} R_{[nt]} + n^{-1/2} (nt - [nt])\, (R_{[nt]+1} - R_{[nt]}) , \quad t \in [0,1] \qquad \text{with} \quad R_n \equiv n^{(1+\beta)/2} (\hat{\theta}_{n+1} - \theta_o) ;$$

$$Y(t) \equiv X(t) + [A + I(1+\beta)/2] \int_{(0,1]} s^{-[A+I(3+\beta)/2]} X(st)\, ds \quad , \quad t \in [0,1] \ .$$

In the deÞnition of Y and elsewhere, for $t > 0$ and $\Gamma, \Lambda \in L(H,H)$, we deÞne $t^{\Gamma} \equiv \exp((\log t)\,\Gamma)$ and $\exp(\Lambda) \equiv \sum_{j=0}^{\infty} \Lambda^j / j!$. The integrals are Bochner integrals in $L(H,H)$. Assumption B.2 ensures that the integral in the deÞnition of Y exists $a.s.-P$.

ASSUMPTION B.5: For some $X \in C_H[0,1]$, $X_n => X$ (as $n \to \infty$) in $C_H[0,1]$.

THEOREM 3.1: Suppose $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$ and Assumptions B.1-B.5 hold. Then $Y_n => Y$ (as $n \to \infty$) in $C_H[0,1]$.

This theorem is proved by mimicing the proofs of theorem 1 in Walk (1987), corollary 2.12 in Berger (1986) and theorem 1 in Nixdorf (1984).

Before we state our next result, we need some deÞnitions and notations. (Details can be found in Chen and White (1998a)). An $H-r.e.$ W (or the distribution μ of an $H-r.e.$ W) is *weakly second order* if $E[(W, h)^2] \equiv \int_H (w, h)^2 \mu(dw) < \infty$ for all $h \in H$. For a weakly second order $H-r.e.$ W with expectation $E[W]$, we deÞne the *covariance* of W, Cov W: $H \times H \to I\!R$, as

$$\text{Cov } W(x,y) \equiv E[(W - E[W], x)(W - E[W], y)] , \quad \text{for any } x, y \in H .$$

This is a symmetric, positive, and continuous bilinear form. Alternatively, we can deÞne a *covariance operator* $S_W : H \to H$ as $S_W x \equiv E((W - E[W], x)[W - E[W]])$ for any $x \in H$. By the deÞnitions we have

$$(S_W x)(y) \equiv (S_W x, y) = \text{Cov } W(x,y) \ \text{ for any } x, y \in H .$$

Let $L(H,H)$ denote the space of bounded linear operators with the operator norm $\|\cdot\|$ deÞned as $\|S\| = \sup[\|S x\| : \|x\| \le 1 ,\ x \in H]$. For a compact operator S in $L(H,H)$, if $\sum_{j=1}^{\infty} |(S e_j, e_j)| < \infty$ and $\sum_{j=1}^{\infty} (S e_j, e_j)$ is independent of the choice of complete orthonormal system (cons) $\{e_j\}$, we call S a *nuclear* operator and $\text{tr}(S) \equiv \sum_{j=1}^{\infty} (S e_j, e_j)$ the (matrix) *trace* of S. When a nuclear operator S is self-adjoint and positive, we have $\text{tr}(S) = \sum_{j=1}^{\infty} \lambda_j(S)$, where $\{\lambda_j(S), j \ge 1\}$ are the eigenvalues of S. Let $\mathcal{S}(H)$ be the set of all self-adjoint positive nuclear operators. This is a Polish space under the metric $d(S,J) \equiv \|S - J\|_{tr} \equiv \text{tr}([(S-J)^*(S-J)]^{1/2})$, where $(S-J)^*$ is the adjoint of $(S-J)$, for any $S, J \in \mathcal{S}(H)$. In probability theory, $\mathcal{S}(H)$ consists of all covariance operators of Gaussian measures on H.

An $H-r.e.$ $\mathcal{N}$ has a *Gaussian distribution on* H if for all $h \in H$, the real-valued random variable $(h, \mathcal{N})$ has a Gaussian distribution on $I\!R$, or equivalently, an arbitrary Þnite set of coordinates of $\mathcal{N}$ (in an arbitrary cons) has a Þnite-dimensional Gaussian distribution. We call $\mathcal{N}$ an H*-valued Gaussian*, and $\mathcal{N}(0, S)$ an H-valued Gaussian with zero mean and covariance $S \in \mathcal{S}(H)$.

A *Brownian motion* (BM) in H is a $C_H[0,1]-r.e.$ X satisfying the following conditions: (a) $X(0) = 0$; (b) the increments on disjoint time intervals are independent; (c) for all

$0 \le t < t+s \le 1$, the increment $X(t+s) - X(t)$ has a Gaussian distribution on H with mean zero and covariance operator $s\,S$, where $S \in \mathcal{S}(H)$, does not depend on t, s.

COROLLARY 3.2: In Theorem 3.1, if B.5 holds for X a Brownian motion in H with $X(0) = 0$, $E\,X(1) = 0$ and Cov $X(1) = S$, then:
(i) $Y_n => Y$ where Y is a Brownian motion in H with $Y(0) = 0$, $E\,Y(1) = 0$ and Cov $Y(1) = K$, where K is the unique solution of the operator equation $\bar{A}\,K + K\,\bar{A}^* = -S$ in $L\,(H\,,H)$, with $\bar{A} \equiv (\beta/2)\,I + A$,

$$K \equiv \int_{(0,1]} s^{-(A+I)}\, S\, s^{-(A^*+I)}\, ds \equiv \int_{(0,\infty)} \exp(\bar{A}u)\, S \exp(\bar{A}^* u)\, du \quad ; \text{ and}$$

(ii) $n^{\beta/2}\,(\hat{\theta}_{n+1} - \theta_o) => \mathcal{N}(0\,,K)$ (as $n \to \infty$) in H.

Assumptions B.4 and B.5 are abstract conditions. In the rest of this section, we provide sufficient conditions in terms of various dependent, possibly heterogeneous H–valued random sequences. In particular, we consider a general class of H–valued mixingales. Before we provide sufficient conditions for Assumption B.4, we state a lemma that contains some general criteria for B.4(1) and B.4(2). These results can be found in Berger (1986) and Walk (1987).

LEMMA 3.3: Let $\{\nu_n\}$ and $\{T_n\}$ be any sequences of H–$r.e.\acute{}s$.

(i) If there exists $\delta' \in (\delta_o\,, 1/2)$ such that

$$\text{B.4(1}')\quad n^{-1/2} \max_{1 \le l \le n} (n/l)^{\delta'} \left\| \sum_{j=1}^{l} \nu_j \right\| = O_P(1)\,,$$

then $\{\nu_n\}$ satisfies B.4(1).

(ii) If B.4(1''): $E \left\| \sum_{j=1}^{n} \nu_j \right\| = O(n^{1/2})$, then B.4(1') holds.

(iii) If B.4(2'): $n^{-1} \sum_{j=1}^{n} \| T_j \| \to 0$ almost in first mean, then $\{T_n\}$ satisfies B.4(2).

(iv) If $T_n \to 0$ almost in first mean, then B.4(2') is satisfied.

(v) If B.4(2''): $n^{-1} \left\| \sum_{j=1}^{n} T_j \right\| \to 0$ $a.s.-P$, then $\{T_n\}$ satisfies B.4(2).

To verify the conditions in the above lemma for an H–valued mixingale process, we can apply Chen and White's (1996) maximal inequalities and law of large numbers for $L_p(H)$–mixingales.

ASSUMPTION B.6:

(1) For any $\theta \in H$ and all n, $E\,[\,P_{k(n+1)}\,U_n(\cdot\,,\theta)\,] = 0$.

(2) For every bounded set $\Theta \subset H$, there is a sequence of nonnegative square integrable functions $\{\phi_{\Theta,n}\}$ such that $\sup_{l \in \mathbb{N}} E[\phi^2_{\Theta,l}] < \infty$ and for all $\theta\,,\theta' \in \Theta$, $n \in \mathbb{N}$,

$$\| P_{k(n+1)} U_n(\cdot\,,\theta) - P_{k(n+1)} U_n(\cdot\,,\theta') \| \le \phi_{\Theta,n} \|\theta - \theta'\|.$$

(3) For some $r \in (2,\infty)$, for any bounded subset $\Theta \subset H$,

$$\sup_{n \in \mathbb{N}} E\,[\,\sup_{\theta \in \Theta} \| P_{k(n+1)} U_n(\cdot\,,\theta)\|^r\,] < \infty\,.$$

(4) $\{ P_{k(n+1)} U_n(Z_n, \theta_o) , \mathcal{F}^n \}$ is an adapted L_2–mixingale sequence with $\{ \psi_m \}$ of size $-1/2$, and $\sum_{j=1}^{n} j^{\beta-1} c_j^{\,2} = O(n)$.

The following is a set of sufficient conditions for B.6. In particular, for j=1,2,3,4, B.7(j) implies B.6(j) respectively.

ASSUMPTION B.7:

(1) For any $\theta \in H$ and all n , $E[U_n(\cdot , \theta)] = 0$.

(2) For every bounded set $\Theta \subset H$, there is a sequence of nonnegative square integrable functions $\{ \phi_{\Theta,n} \}$ such that $\sup_{l \in \mathbb{N}} E[\phi^2_{\Theta,l}] < \infty$ and for all $\theta , \theta' \in \Theta$, $n \in \mathbb{N}$,

$$\| U_n(\cdot , \theta) - U_n(\cdot , \theta') \| \le \phi_{\Theta,n} \| \theta - \theta' \| .$$

(3) For some $r \in (2,\infty)$, for any bounded subset $\Theta \subset H$,

$$\sup_{n \in \mathbb{N}} E[\sup_{\theta \in \Theta} \| U_n(\cdot , \theta) \|^r] < \infty .$$

(4) $\{ U_n(Z_n, \theta_o) , \mathcal{F}^n \}$ is an adapted L_2–mixingale sequence with $\{ \psi_m \}$ of size $-1/2$, and $\sum_{j=1}^{n} j^{\beta-1} c_j^{\,2} = O(n)$.

PROPOSITION 3.4: (i) Conditions B.6(1) and B.6(4) imply B.4(1’’);
(ii) Conditions B.6(1) - B.6(3) and $\| \hat{\theta}_n - \theta_o \| \to 0 \quad a.s. - P$ imply B.4(2).

Before we present sufficient conditions for B.5, we need the definition of $L_p(H)$–near epoch dependent processes and related results; see Chen and White (1998a) for details.

DEFINITION 3.5:

(1) Let $\mathcal{A}$, $\mathcal{G}$ be two σ–subfields on the probability space $(\Omega , \mathcal{F} , P)$. Define two measures of dependence as :

$$\alpha (\mathcal{A} , \mathcal{G}) \equiv \sup [\, | P(A C) - P(A) P(C) | : A \in \mathcal{A} , C \in \mathcal{G}] ;$$

$$\phi (\mathcal{A} , \mathcal{G}) \equiv \sup [\, | P(C \mid A) - P(C) | : A \in \mathcal{A} , P(A) > 0 , \; C \in \mathcal{G}] .$$

(2) Let $\{ D_n \}$ be a sequence of $\mathbb{B}$ –r.e.´s defined on the probability space $(\Omega , \mathcal{F} , P)$, and denote $\mathcal{A}_a^{\,b} \equiv \sigma (D_j ; \; a \le j \le b)$. Define

$$\alpha(m) \equiv \sup_n [\alpha (\mathcal{A}_{-\infty}^{\;n} , \mathcal{A}_{n+m}^{\;\infty})] ; \quad \phi(m) \equiv \sup_n [\phi (\mathcal{A}_{-\infty}^{\;n} , \mathcal{A}_{n+m}^{\;\infty})] .$$

If $\lim_{m \to \infty} \alpha(m) = 0$, then $\{ D_n \}$ is called an α–*mixing* sequence. If $\lim_{m \to \infty} \phi(m) = 0$, then $\{ D_n \}$ is called a ϕ–*mixing* sequence.

(3) Let $\{ D_n ; \; -\infty < n < \infty \}$ be a $\mathbb{B}$–r.e. sequence and $\{ W_n ; \; -\infty < n < \infty \}$ be an H –r.e. sequence. Then $\{ W_n \}$ is called $L_p(H)$ –*near epoch dependent* (NED) on $\{ D_n \}$ if $\| W_n \|_p < \infty$, $1 \le p < \infty$, and there exist constants $\{ \mu_m \ge 0 ; \; m \ge 0 \}$ with μ_m decreasing to zero as $m \to \infty$ and $\{ d_n \ge 0 ; \; n \ge 1 \}$ with $\sup_n d_n < \infty$ such that

$$\| W_n - E [W_n \mid \mathcal{A}_{n-m}{}^{n+m}] \|_p \le \mu_m \, d_n \,, \quad \text{where } \mathcal{A}_a{}^b \text{ is as in (2)} .$$

Let $\{ e_j \,;\, j \ge 1 \}$ be an arbitrary cons of H, and let H_k be the closed linear span of $[e_j \,;\, 1 \le j \le k]$. Let $P_k : H \to H_k$ be the orthonormal projection operator. Let $\{ W_n \}$ be an $H-r.e.$ sequence. Let S_n be the covariance operator of $n^{-1/2} \sum_{j=1}^{n} W_j$. Define $S_n{}^k \equiv P_k \, S_n \, P_k$ and

$$X_n(t) \equiv n^{-1/2} \sum_{i=1}^{[nt]} W_i + n^{-1/2} \, (nt - [nt]) \, W_{[nt]+1} \;, \quad t \in [0,1] .$$

Let $S_{[nt]}{}^k \equiv P_k \, S_{[nt]} \, P_k$ be the covariance matrix of $n^{-1/2} \sum_{j=1}^{[nt]} P_k W_j$.

Now we are ready to provide sufficient conditions for Assumption B.5, using the following lemma, which is Theorem 4.14 in Chen and White (1992).

LEMMA 3.6: Suppose that $\{ W_n = n^{(\beta-1)/2} \, P_{k(n+1)} U_n(Z_n, \theta_o) \,,\, \mathcal{F}^n \}$ has zero means and uniformly bounded $r-th$ moments $(r > 2)$ and satisfies

(1) (a) $\{ W_n \}$ is $L_2(H)$ –NED on $\{ D_n \}$ with μ_m of size $-1/2$ and $d_n \equiv 1$;

(b) $\{ D_n \}$ is a mixing $\mathbb{B}-r.e.$ sequence with either $\alpha(m)$ of size $-r/(r-2)$ or $\phi(m)$ of size $-r/2(r-1)$.

(2) Suppose there exists $S \in \mathcal{S}(H)$, $S \ne 0$ such that:

(a) For each $k \ge 1$, let $\{ \lambda_l (S_n{}^k) \,;\, 1 \le l \le k \}$ be the eigenvalues of $S_n{}^k$ in nonincreasing order. Then

$$\text{diag} \, [\lambda_1{}^{-1} (S_n{}^k) \,, \; \cdots \;, \lambda_k{}^{-1} (S_n{}^k) \,] = O\,(1) \;;$$

(b) For each $k \ge 1$, $S_{[nt]}{}^k \to t \, P_k S P_k$ as $n \to \infty$ for any $t \in [0,1]$; and

(c) $\limsup_n \mid \text{tr}(S_{[nt]}) - \text{tr}(P_k S_{[nt]} P_k) \mid \to 0$ uniformly in $t \in [0,1]$ *as* $k \to \infty$.

Then Assumption B.5 is satisfied when X is a Brownian motion in H with $X(0) = 0$, $E\,X(1) = 0$ and $\text{Cov}\,X(1) = S$.

The following result summarizes Corollary 3.2, Proposition 3.4 and Lemma 3.6. It gives a CLT and an FCLT for the RMP and TRMP algorithms when errors are $\theta-$*dependent* and $L_2(H)$–NED on some mixing processes.

COROLLARY 3.7: Suppose $\| \hat{\theta}_n - \theta_o \| \to 0 \;\, a.s. - P$ and Assumptions B.1-B.3, B.6(1) - B.6(3) hold. If $\{ W_n = \nu_n = n^{(\beta-1)/2} \, P_{k(n+1)} U_n(Z_n, \theta_o) \,,\, \mathcal{F}^n \}$ satisfies conditions 3.6(1) and 3.6(2), then all the conclusions in Corollary 3.2 hold.

Using the same approach, we can also obtain a CLT and an FCLT for the RM and TRM algorithms when errors are $\theta-$*dependent* and $L_2(H)$ –NED on some mixing processes. In particular, we consider the case where errors are weakly stationary mixingale processes and $\{ \hat{\theta}_n \}$ is given by the RM algorithm:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + n^{-1} \, M(\hat{\theta}_n) + n^{-1} \, U_n(Z_n \,, \hat{\theta}_n) \;.$$

We obtain the following result using Theorem 3.9 in Chen and White (1998a).

COROLLARY 3.8: Given the RM as above, suppose $\|\hat{\theta}_n - \theta_o\| \to 0$ *a.s.* $-P$ and Assumptions B.1, B.2, B.7(1) - B.7(3) hold. Suppose $\{ W_n = \nu_n = n^{(\beta-1)/2} U_n(Z_n, \theta_o) , \mathcal{F}^n \}$ is weakly stationary and satisfies the following conditions:
(1) (a) $\{ W_n \}$ is $L_2(H)$ $-$NED on $\{ D_n \}$ with μ_m of size -1 and $d_n \equiv 1$;
(b) $\{ D_n \}$ is a mixing $\mathbb{B}-r.e.$ sequence with either $\alpha(m)$ of size $-2r/(r-2)$ or $\phi(m)$ of size $-r/(r-1)$;
(2) For each $k \geq 1$, $\min [(P_k S P_k x , x) : \|x\| = 1 , x \in H] > 0$,

Then all the conclusions in Corollary 3.2 hold.

In practice, A is unknown and S is unobservable; hence, the corresponding K is unknown. The following result is useful for statistical inference.

LEMMA 3.9: Let $\{ \tilde{\Gamma}_n \}$ be a random sequence in $L(H,H)$ and let $\{ \tilde{S}_n \}$ be a random sequence in $\mathcal{S}(H)$. Suppose for each n , spec $(\tilde{\Gamma}_n) \subseteq [\gamma \in \mathbb{C} : \mathrm{Re}\gamma > -1/2]$ *a.s.* $-P$, and let

$$\tilde{K}_n \equiv \int_{(0,1]} s^{\tilde{\Gamma}_n} \tilde{S}_n s^{\tilde{\Gamma}_n^*} ds .$$

If there exist a nonrandom Γ in $L(H, H)$ with spec $(\Gamma) \subseteq [\gamma \in \mathbb{C} : \mathrm{Re}\gamma > -1/2]$ and a nonrandom S in $\mathcal{S}(H)$ such that $\lim_n \|\tilde{\Gamma}_n - \Gamma\| = 0$ *in Prob.* (resp. *a.s.-P*), and $\lim_n \|\tilde{S}_n - S\|_{tr} = 0$ *in Prob.* (resp. *a.s.-P*), then $\lim_n \|\tilde{K}_n - K\|_{tr} = 0$ *in Prob.* (resp. *a.s.* $-P$), where $K \equiv \int_{(0,1]} s^{\Gamma} S s^{\Gamma^*} ds$.

Given Corollary 3.7 (or Corollary 3.8) and Lemma 3.9, we can then apply Dippon's (1991) Theorem 1 to construct asymptotic confidence regions for $\|\hat{\theta}_n - \theta_o\|$.

IV. LAWS OF ITERATED LOGARITHM

This section derives a law of iterated logarithm (LIL) for the RMP and TRMP algorithms when $\{ U_n \}$ is a sequence of $H$$-$valued NED, $\theta-$*dependent* errors. Our approach is akin to that for the asymptotic normality results; thus the basic setup of section III remains valid.

Just as a CLT for the partial sum $\{ \sum_{j=1}^{n} \nu_j \}$ implies a CLT for $\{ \hat{\theta}_n \}$, an LIL for $\{ \sum_{j=1}^{n} \nu_j \}$ implies an LIL for $\{ \hat{\theta}_n \}$. The following result is a simple corollary of Walk's (1987) Theorem 2.

COROLLARY 4.1: Suppose $\|\hat{\theta}_n - \theta_o\| \to 0$ *a.s.* $-P$, and Assumptions B.1, B.2, and the following conditions hold :

$$(1) \quad (n \log \log n)^{-1/2} \left\| \sum_{j=1}^{n} \nu_j - BM(n) \right\| \to 0 \quad \text{as } n \to \infty \quad a.s. -P ,$$

where $BM = \{ BM(t) ; t \in [0, \infty) \}$ is a Brownian motion in H with covariance operator S ;

$$(2)\ n^{-1} (\log\log n)^{-1/2} \| \sum_{j=1}^{n} T_{1j} \| \to 0 \quad \text{as } n \to \infty \quad a.s.-P ;$$

$$(3)\ n^{-1} (\log\log n)^{-1/2} \| \sum_{j=1}^{n} T_{2j} \| \to 0 \quad \text{as } n \to \infty \quad a.s.-P .$$

Then (i) $(t \log\log t)^{-1/2} \| [t]^{(1+\beta)/2} (\hat{\theta}_{[t]+1} - \theta_o) - GM(t) \| \to 0$ as $t \to \infty$ $a.s.-P$,

and (ii) $\limsup_{t\to\infty} (2\lambda\ t \log\log t)^{-1/2} \| [t]^{(1+\beta)/2} (\hat{\theta}_{[t]+1} - \theta_o) \| = 1$ $a.s.-P$,

where $\lambda \equiv \sup [(K h, h) : \|h\| \le 1, h \in H]$ is the largest eigenvalue of K, as deÞned in Corollary 3.2, and $GM = \{ GM(t) ; t \in [0, \infty) \}$ is a Gaussian Markov process in H with $GM(t) \equiv BM(t) + (A + I(1+\beta)/2) \int_{(0,\infty)} s^{-(A+I(3+\beta)/2)} BM(st)\, ds$.

Now we give sufÞcient conditions for 4.1(1), 4.1(2) and 4.1(3). Condition 4.1(1) assumes an LIL for the partial sum $\{ \sum_{j=1}^{n} v_j \}$. The following LIL for a general adapted $L_2(H)$–mixingale of size -1 is a consequence of Philipp's (1986) theorem.

LEMMA 4.2: Let $d \equiv \dim(H) \le \infty$, and let $\{ W_n ; n \ge 1 \}$ be an H–valued sequence with $\sup_n E[\| W_n \|^{2+\delta}] < \infty$ for some $\delta > 0$. Suppose the following conditions hold:
(1) $\{ W_n , \mathcal{F}^n ; n \ge 1 \}$ is an adapted $L_p(H)$–mixingale of size -1 for some $p \ge 1$.
(2) There exist $\delta > 0$ and an $S \in \mathcal{S}(H)$, $S \ne 0$, such that $E[\| C_{n,m} - n S \|_{tr}] = O(n^{1-\delta})$ uniformly in $m \ge 0$, where $C_{n,m} : H \to H$ is deÞned as

$$C_{n,m}\, h \equiv E[(h, \Sigma_{m+1 \le j \le m+n} W_j) \Sigma_{m+1 \le j \le m+n} W_j \mid \mathcal{F}^m], \quad h \in H .$$

Then without changing its distribution we can redeÞne the sequence $\{ W_n \}$ on a richer probability space on which there exists an H–valued *i.i.d.* Gaussian sequence $\{ \mathcal{N}_j(0, S) \}$ such that

$$(i)\ \| \sum_{j=1}^{n} W_j - \sum_{j=1}^{n} \mathcal{N}_j(0, S) \| = O(n^{(1/2)-\eta}) \quad a.s.-P \quad \text{if } d \equiv \dim(H) < \infty ,$$

where $\eta > 0$ is a constant depending only on r and d ; and

$$(ii)\ \| \sum_{j=1}^{n} W_j - \sum_{j=1}^{n} \mathcal{N}_j(0, S) \| = o([n \log\log n]^{1/2}) \quad a.s.-P \quad \text{if } d \equiv \dim(H) = \infty .$$

The following LIL for a weakly stationary, adapted $L_2(H)$–NED sequence is an application of Lemma 4.2, Chen and White's (1996) Theorem 3.7 and Lemma 4.2, and Chen and White's (1998a) Lemma 3.8.

LEMMA 4.3: Let $d \equiv \dim(H) \le \infty$ and let $\{ W_n \}$ be an adapted weakly stationary H–*r.e.*

sequence

with zero means and uniformly bounded L_r –norm ($r > 2$) satisfying condition 3.8(1). Let S_n be the covariance operator of $n^{-1/2} \sum_{j=1}^{n} W_j$ with $\| n\, S_n \|_{tr} \to \infty$ as $n \to \infty$. Then there exists $S \in \mathcal{S}(H)$, $S \neq 0$, such that $\| S_n - S \|_{tr} \to 0$ as $n \to \infty$. Moreover, without changing its distribution we can redefine the sequence $\{ W_n \}$ on a richer probability space on which there exists an H –valued *i.i.d.* Gaussian sequence $\{ \mathcal{N}_j (0 , S) \}$ such that the conclusions of Lemma 4.2 and the following hold:

$$\lim sup_{n \to \infty} [\, 2 \tau_n\, n \log \log n \,]^{-1/2} \| \sum_{j=1}^{n} W_j \| = 1 \quad a.s.\!-\!P \;,$$

where $\tau_n \equiv \sup [\, (S_n\, h\, , h) : \| h \| \leq 1\, ,\; h \in H \,]$ is the largest eigenvalue of S_n .

REMARK 4.4: If $\{ W_n = \nu_n = n^{(\beta-1)/2} P_{k(n+1)} U_n(Z_n,\theta_o) \}$ satisfies all the conditions of Lemma 4.2, then without changing its distribution we can redefine the sequence $\{ \nu_n \}$ on a richer probability space on which 4.1(1) holds.

We can apply Chen and White (1996) to provide sufficient conditions for 4.1(2); one such example is as follows.

ASSUMPTION C.1:
(1) For each $\theta \in H$, $\{ P_{k(n+1)} U_n(Z_n,\theta) , \mathcal{F}^n \}$ is an adapted H –valued L_2 –mixingale sequence with $\{ \psi_m \}$ of size $-1/2$ and $\sum_{n=1}^{\infty} n^{-(2-\beta)} (\log \log n)^{-1} (c_n)^2 < \infty$.
(2) There exists a sequence of $\mathcal{B}(G) / \mathcal{B}(I\!R)$ –measurable functions $\{ h_n : G \to [0,\infty) \}$ such that for all $z \in G$, $\theta, \theta' \in H$,

$$\| P_{k(n+1)} U_n(z , \theta) - P_{k(n+1)} U_n(z , \theta') \| \leq h_n(z) \| \theta - \theta' \| .$$

(3) $\sup_n \| h_n(Z_n) \|_2 < \infty$ and $\{ h_n(Z_n) - E[\, h_n(Z_n) \,] , \mathcal{F}^n \}$ is an $I\!R$ –valued L_2 –mixingale with $\{ \psi_m \}$ of size $-1/2$.

Assumption C.1 is implied by the following stronger assumption.

ASSUMPTION C.2:
(1) For each $\theta \in H$, $\{ U_n(Z_n,\theta) , \mathcal{F}^n \}$ is an adapted H –valued L_2 –mixingale sequence with $\{ \psi_m \}$ of size $-1/2$ and $\sum_{n=1}^{\infty} n^{-(2-\beta)} (\log \log n)^{-1} (c_n)^2 < \infty$.
(2) There exists a sequence of $\mathcal{B}(G) / \mathcal{B}(I\!R)$ –measurable functions $\{ h_n : G \to [0,\infty) \}$ such that for all $z \in G$, $\theta, \theta' \in H$,

$$\| U_n(z , \theta) - U_n(z , \theta') \| \leq h_n(z) \| \theta - \theta' \| .$$

(3) $\sup_n \| h_n(Z_n) \|_2 < \infty$ and $\{ h_n(Z_n) - E[\, h_n(Z_n) \,] , \mathcal{F}^n \}$ is an $I\!R$ –valued L_2 –mixingale with $\{ \psi_m \}$ of size $-1/2$.

The proof of the following result is similar to those of Corollaries 2.6 and 2.7.

PROPOSITION 4.5: Given the RMP or TRMP as in Section III, A.3(2), C.1 and $\| \hat{\theta}_n - \theta_o \| \rightarrow 0 \quad a.s.-P$, then 4.1(2) is satisfied.

The following assumption together with $\{ \| \hat{\theta}_n \| \}$ uniformly bounded $a.s.-P$ implies condition 4.1(3).

ASSUMPTION C.3:

$$(1) \quad n^{-1} (\log \log n)^{-1/2} \| \sum_{j=1}^{n} j^{\beta/2} (I - P_{k(j+1)}) A P_{k(j+1)} \| \rightarrow 0 \quad \text{as } n \rightarrow \infty ;$$

$$(2) \quad n^{-1} (\log \log n)^{-1/2} \| \sum_{j=1}^{n} j^{\beta/2} (\theta_o - P_{k(j+1)} \theta_o) \| \rightarrow 0 \quad \text{as } n \rightarrow \infty .$$

This assumption only imposes very mild restrictions on the choice of the projection subspaces. By Kronecker's Lemma, Assumption C.3 is implied by

ASSUMPTION C.4:

$$n^{\beta/2} \| (I - P_{k(n+1)}) A P_{k(n+1)} \| = O(1) ; \qquad n^{\beta/2} \| \theta_o - P_{k(n+1)} \theta_o \| = O(1) .$$

The following result summarizes Corollary 4.1, Remark 4.4 and Proposition 4.5. It gives an LIL for the RMP and TRMP algorithms when errors are θ*–dependent* , $L_p(H)$–mixingales ($p \geq 1$).

COROLLARY 4.6: Given the RMP or TRMP as in Section III, B.1, B.2, C.1, C.3, and $\| \hat{\theta}_n - \theta_o \| \rightarrow 0 \quad a.s.-P$, if $\{ W_n = \nu_n = n^{(\beta-1)/2} P_{k(n+1)} U_n(Z_n, \theta_o) \}$ satisfies conditions 4.2(1) - 4.2(2), then the conclusions of Corollary 4.1 hold.

Finally, we present an LIL for the RM and TRM algorithms when $\{ U_n \}$ is a weakly stationary θ*–dependent* process, $L_2(H)$–NED on some mixing processes.

COROLLARY 4.7: Given the RM or TRM as in Section III, B.1, B.2, C.2, and $\| \hat{\theta}_n - \theta_o \| \rightarrow 0 \quad a.s.-P$, if $\{ W_n = \nu_n = n^{(\beta-1)/2} U_n(Z_n, \theta_o) \}$ satisfies condition 3.8(1), then the conclusions of Corollary 4.1 hold.

REMARK 4.8: When $H = \mathbb{R}^d$, $d < \infty$, the approximation error (result 4.2(i)) is small enough to directly deliver a CLT and an FCLT in addition to the almost-sure loglog rate result 4.3(i). Heunis (1992) has followed this approach to get an almost-sure loglog rate of convergence and an FCLT for a special RM algorithm in $\mathbb{R}^d$ ($d < \infty$) (i.e., the Wiener-Hopf problem) when the error sequence $\{ U_n \}$ is a linearly θ*–dependent* , strictly stationary mixing process. Our Corollaries 3.8 and 4.7 and Lemma 4.3 extend most of his results to various RM algorithms in $\mathbb{R}^d$ ($d < \infty$) when errors $\{ U_n \}$ are θ*–dependent* , weakly stationary NED functions of some mixing processes, and nonlinear in θ .

REMARK 4.9: When $\dim (H) = \infty$, the approximation error in result 4.2(ii) is not small enough to directly imply a CLT and an FCLT. But it is easy to verify that all conditions of Chen and White's (1992) Theorem 4.10 are satisfied given the assumptions of Lemma 4.2 above. Thus an FCLT and a CLT still hold, although the conditions for the LIL are in general stronger than those needed for an FCLT when $\dim(H) = \infty$ (Max Stinchcombe has constructed a clever example which satisfies an FCLT trivially yet fails an LIL). We note also that the mixingale rate condition in Lemma 4.2 is stronger than that in Lemma 3.6; therefore the results of Section III are not redundant.

V. MEAN RATE OF CONVERGENCE

Section IV provides an almost-sure loglog rate of convergence for $\{\hat{\theta}_n\}$ when $M(\cdot)$ is locally linearizable about θ_o. This section investigates another type of convergence rate property without the smoothness assumption B.1. In particular, we obtain some results on the order of magnitude of $E[V(\hat{\theta}_n)]$, where V is a Liapunov functional. The method of proof combines Venter's (1966) lemma and Kushner's (1984) perturbed Liapunov functional technique. The conditions are slightly stronger than those specified for almost-sure convergence, but are weaker than those for asymptotic normality and law of iterated logarithm. They cover many different types of noise assumptions, including adapted H–valued mixingale sequences (e.g., near epoch dependent (NED) functions of α – or ϕ – mixing sequences) as special cases. Since our approach is valid for direct H–valued RM and "truncated" RM (TRM) procedures, as well as sieve-based RM projected (RMP) and TRM projected (TRMP) procedures, we only provide the conditions, results and proofs for the sieve-based RMP (TRMP) in detail and state those for the direct H–valued RM (TRM) in brief. We specialize our results to get rates of convergence for $E[\|\hat{\theta}_n - \theta_o\|^2]$ under two kinds of Liapunov functional assumptions.

Assumptions A.1, A.3P and A.4P are always in force in this section. We also set $a_n = 1/n$ from now on. Hence we always have $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$. By an argument similar to Yin and Zhu's (1990) Theorem 3.3, we can rewrite Assumption A.4P as

$$\lim_{n\to\infty} \| n^{-1} \sum_{j=1}^{n} P_{k(j+1)}\, U_j(Z_j, \hat{\theta}_j) \| = 0 \quad a.s.-P \ .$$

We use the following Assumption AP to simplify notation.

ASSUMPTION AP: Assumptions A.1 and A.3P(1) hold, $a_n = 1/n$, and $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

Although we can work directly with $E[V(\hat{\theta}_n)]$, we consider instead the behavior of the expectation of a perturbed Liapunov functional to allow for highly dependent error processes. In particular, we consider the perturbed Liapunov functional $S(n) \equiv V(\hat{\theta}_n) + \bar{V}(\hat{\theta}_n, n)$, where

$$\bar{V}(\theta, n) \equiv \sum_{j=n}^{\infty} j^{-1} E[\, (V'(\theta), P_{k(j+1)}\, U_j(Z_j, \theta)) \mid \mathcal{F}^{n-1}] \ ,$$

where as before $\mathcal{F}^{n-1}$ is the σ–algebra generated by $\{ Z_j, \hat{\theta}_{j+1} ; j \le n-1 \}$. Denote

$$\overline{DV}_n \equiv E[\, \bar{V}(\hat{\theta}_{n+1}, n+1) - \bar{V}(\hat{\theta}_n, n+1) \mid \mathcal{F}^{n-1}] \ .$$

By definition, $\overline{V}(\hat{\theta}_n, n)$, $\overline{DV}_n$, and $S(n)$ are measurable$-\mathcal{F}^{n-1}$.

ASSUMPTION D.1P:

(1) There exists a sequence $\{ c_n > 0 \}$ with $\lim_{n\to\infty} c_n = c$ for some $0 < c < \infty$, and a finite integer N_o, such that for all $n \geq N_o$,

$$(V'(\theta), P_{k(n)} M(\theta)) \leq - c_n V(\theta), \quad \text{for any } \theta \in H_{k(n)}, \ \theta \neq \theta_o .$$

(2) $\| V'(\theta) \|^2 \leq K (1 + V(\theta))$ for all $\theta \in H$, and some $0 < K < \infty$.

Condition D.1P(1) is stronger than A.3P(2), and is crucial for rate of mean convergence results. Condition D.1P(2) is very mild and is merely a convenient way to deliver a nonnegative perturbed Liapunov functional.

LEMMA 5.1: Given TRMP, suppose Assumptions AP and D.1P(1) hold. Suppose further the following conditions hold for some $a, b \in (0, 1]$ $\beta \in [0, 1)$ and all sufficiently large n:

$$(1)\ E[\overline{V}(\hat{\theta}_n, n)] \leq O(n^{-a}); \quad (2)\ E[\overline{DV}_n] \leq O(n^{-(1+b)});$$

$$(3)\ E[S(n)] \geq 0; \quad \text{and} \quad (4)\ E[\| P_{k(n+1)} U_n(Z_n, \hat{\theta}_n) \|^2 \mid \mathcal{F}^{n-1}] = O(n^{\beta}).$$

Denote $q \equiv \min(a, b, 1-\beta)$. Then for n sufficiently large,

$$E[V(\hat{\theta}_n)] = O(n^{-q}) \quad \text{for } c > q; \quad E[V(\hat{\theta}_n)] = O(n^{-q} \log n) \quad \text{for } c = q;$$

$$\text{and} \quad E[V(\hat{\theta}_n)] = O(n^{-c}) \quad \text{for } 0 < c < q.$$

The key point is that $E[S(n)]$ and $E[V(\hat{\theta}_n)]$ have the same convergence rate, while $E[\overline{V}(\hat{\theta}_n, n)]$ will go to zero at the same or faster rate. Conditions 5.1(1) - 5.1(4) are satisfied by many kinds of dependent random sequences. We now give some sufficient conditions.

If $\{ U_n \equiv F_n(Z_n) \}$ is $\theta-$*independent*, we have the following simple sufficient conditions :

ASSUMPTION D.2P:

$$(1) \quad \| E[\sum_{j=n}^{\infty} j^{-1} P_{k(j+1)} F_j(Z_j) \mid \mathcal{F}^{n-1}] \|_2 \leq O(n^{-b}) \text{ for some } b > 0 \text{ and all sufficiently large } n.$$

$$(2) \quad \sup_n E[\| P_{k(n+1)} F_n(Z_n) \|^2] < \infty .$$

By Minkowski's inequality, D.2P(1) is implied by

$$(3) \quad \sum_{j=n}^{\infty} j^{-1} \| E[P_{k(j+1)} F_j(Z_j) \mid \mathcal{F}^{n-1}] \|_2 \leq O(n^{-b}) \text{ for some } b > 0 \text{ and all sufficiently large } n.$$

COROLLARY 5.2: Given Assumptions AP, D.1P, D.2P and TRMP with $\theta-$*independent* errors $\{ U_n \}$, then all the conclusions of Lemma 5.1 hold.

The following assumption is a stronger version of A.5P:

ASSUMPTION A.5P': $\{ P_{k(n+1)}F_n(Z_n) , \mathcal{F}^n \}$ is an adapted L_p–mixingale sequence of H–r.e.´s with uniformly bounded second moments, and $\{ \psi_m \}$ satisfies either

(1) $\{ \psi_m \}$ is of size -1 for $1 < p \leq 2$; or

(2) $\{ \psi_m \}$ is of size $-a$, $1/2 \leq a < 1$ for $p = 2$.

THEOREM 5.3: Given AP, D.1P, and TRMP with θ–*independent* errors $\{ U_n \}$,

(i) if A.5P'(1) is satisfied, then all the conclusions of Lemma 5.1 hold with $q = 1$.

(ii) if A.5P'(2) is satisfied, then all the conclusions of Lemma 5.1 hold with $q = a \in [\, 1/2 , 1 \,)$.

If $\{ U_n \equiv U_n(Z_n,\theta) \}$ is θ–*dependent* , we have the following sufficient conditions :

ASSUMPTION D.3P:

(1) For any $\theta \in H$, some $b > 0$ and all sufficiently large n,

$$\| E [\sum_{j=n}^{\infty} j^{-1} P_{k(j+1)}U_j(Z_j,\theta) \mid \mathcal{F}^{n-1}] \|_2 \leq O\,(n^{-b}) .$$

(2) For any $K > 0$,

$$\sup_n E\,[\sup_{\|\theta\| \leq K} \| P_{k(n+1)}U_n(Z_n,\theta) \|^2] < \infty \quad .$$

(3) For every $K > 0$, there is a sequence of nonnegative $\mathcal{B}(G)$–measurable functions $\{ h_{K,n} : G \to [\, 0 , \infty) \}$ such that

$$\Sigma_{j \geq n+1}\, j^{-1} E[\, h_{K,j}(Z_j) \mid \mathcal{F}^n] = O\,(\, n^{-b} \,)$$

for some $b > 0$ and all sufficiently large n $a.s.$–P, and

$$\| P_{k(n+1)}U_n(z,\theta) - P_{k(n+1)}U_n(z,\theta') \| \leq h_{K,n}(z) \| \theta - \theta' \| ,$$

for all $z \in G$, $\|\theta\| \leq K, \|\theta'\| \leq K$, $n \in \mathbb{N}$.

By Minkowski's inequality, a sufficient condition for D.3P(1) is

(4) For any $\theta \in H$, some $b > 0$ and all sufficiently large n,

$$\sum_{j=n}^{\infty} j^{-1} \| E[\, P_{k(j+1)}U_j(Z_j,\theta) \mid \mathcal{F}^{n-1}] \|_2 \leq O\,(n^{-b}) .$$

Comparing the sufficient conditions for the θ–*dependent* error case with those for the θ–*independent* error case, it is obvious that we need some local Lipschitz condition (say D.3P(3)) on the θ–*dependent* errors $\{ U_n(Z_n,\theta) \}$. Notice that D.3P(1) is an analog of D.2P(1), whereas D.3P(2) implies 5.1(4) . Since θ–*independent* errors $\{ U_n \equiv F_n(Z_n) \}$ satisfy D.3P(3) automatically, we can regard θ–*independent* errors as a special case of θ–*dependent* errors.

COROLLARY 5.4: Given TRMP with θ–*dependent* errors $\{ U_n \}$, suppose Assumptions AP, D.1P and D.3P hold. Then all the conclusions of Lemma 5.1 hold.

In the same fashion, we can obtain rates of mean convergence for direct H–valued RM or RTRM procedures:

ASSUMPTION A: Assumptions A.1 and A.3(1) hold, $a_n = 1/n$, and $\| \hat{\theta}_n - \theta_o \| \to 0$ as $n \to \infty$ $a.s.-P$.

ASSUMPTION D.1:

(1) $(V'(\theta) , M(\theta)) \leq - c\ V(\theta)$ for some $0 < c < \infty$ and all $\theta \in H$, $\theta \neq \theta_o$.

(2) $\| V'(\theta) \|^2 \leq K\ (1 + V(\theta))$ for all $\theta \in H$, and some $0 < K < \infty$.

ASSUMPTION A.5’: $\{ U_n \equiv F_n(Z_n) , \mathcal{F}^n \}$ is an adapted L_p–mixingale sequence ($1 < p \leq 2$) of $H-r.e.'s$ with uniformly bounded second moments, and $\{ \psi_m \}$ satisfies either A.5P’(1) or A.5P’(2).

It is again easy to prove that A.5’ implies A.5P’.

ASSUMPTION D.3:

(1) For any $\theta \in H$, some $b > 0$ and all sufficiently large n,

$$\| E [\sum_{j=n}^{\infty} j^{-1}\ U_j(Z_j,\theta) \mid \mathcal{F}^{n-1}] \|_2 \leq O\ (n^{-b}) .$$

(2) For any $K > 0$,

$$\sup_n E\ [\sup_{\|\theta\| \leq K} \| U_n(Z_n,\theta) \|^2] < \infty \quad .$$

(3) For every $K > 0$, there is a sequence of nonnegative $\mathcal{B}(G)$–measurable functions $\{ h_{K,n}:G \to [0 ,\infty)\}$ such that $\Sigma_{j \geq n+1}\ j^{-1}\ E[h_{K,j}(Z_j) \mid \mathcal{F}^n] = O\ (\ n^{-b}\ )$ for some $b > 0$ and all sufficiently large n $a.s.-P$, and

$$\| U_n(z,\theta) - U_n(z,\theta') \| \leq h_{K,n}(z) \| \theta - \theta' \| \text{ for all } z \in G\ ,\ \|\theta\| \leq K\ , \|\theta'\| \leq K\ ,\ n \in I\!N \quad .$$

A sufficient condition for D.3(1) is

(4) For any $\theta \in H$, some $b > 0$ and all sufficiently large n,

$$\sum_{j=n}^{\infty} j^{-1} \| E[\ U_j(Z_j,\theta) \mid \mathcal{F}^{n-1}] \|_2 \leq O\ (n^{-b}) .$$

COROLLARY 5.5: Given Assumptions A, D.1 and D.3 with the TRM, let $q \equiv \min (1 , b)$. Then for all n sufficiently large,

$$E[\ V(\hat{\theta}_n)] = O(\ n^{-q}\ ) \quad \text{for } c > q\ ; \quad E[\ V(\hat{\theta}_n)]\ = O(\ n^{-q} \log n\ ) \quad \text{for } c = q\ ;$$

$$\text{and} \quad E[\ V(\hat{\theta}_n)] = O(\ n^{-c}\ ) \quad \text{for } 0 < c < q\ .$$

EXAMPLE 5.6: Given Assumptions A, D.1 and A.5’ with the RTRM when $\{ U_n \}$ are θ–*independent* errors, then all the conclusions of Corollary 5.5 hold.

Note that this example includes Proposition 5.1 of Yin and Zhu (1990) as a special case. There, Yin and Zhu assume θ–*independent* errors $\{ U_n \}$ to be a stationary ϕ–mixing sequence with $\sum_{m=1}^{\infty} [\phi(m)]^{(r-2)/r} < \infty$, and $\sup_n E[\| U_n \|^r] < \infty$, where $r > 2$. Our Example 5.6 relaxes their conditions in three ways : (i) we do not impose stationarity; (ii) we allow $r \geq 2$; (iii) we permit mixingales of size -1. Yin and Zhu (1990) require $\phi(m)$ of size $-a\,r/(r-1)$ where $a \equiv (r-1)/(r-2) > 1$, corresponding to an L_2–mixingale of size $-a$, $a > 1$.

Our rate of mean convergence results depend heavily on the properties of the error terms $\{ U_n \}$, and the choice of a Liapunov functional, which is in general related to the choice of the projections $\{ P_n \}$ and the properties of M. We have shown that our error conditions, such as D.3P, are reasonably mild. In the rest of this section, we illustrate effects of the choice of Liapunov functional. In particular, we consider Assumption D.1P when $V(\theta)$ can be some local quadratic form, and obtain convergence rates for $E[\|\hat{\theta}_n - \theta_o\|^2]$ corresponding to two versions of assumption D.1P. As long as M is relatively smooth at θ_o, we can choose a quadratic form as a Liapunov functional. The following corollaries are thus applicable in many situations.

COROLLARY 5.7: Given TRMP, suppose Assumptions AP, D.3P with $b = 1$ hold and that

(1) there exists a sequence $\{ c_n > 0 \}$ with $\lim_{n\to\infty} c_n = c$ for some $0 < c < \infty$ such that

$$2\,(\theta - \theta_o\,,\, P_{k(n)}M(\theta)\,) \leq -\,c_n\,\|\theta - \theta_o\|^2\,,\ \text{ for all } \theta \in H_{k(n)}\,.$$

Then for all n sufficiently large,

$$E[\|\hat{\theta}_n - \theta_o\|^2] = O(1/n) \quad \text{for } c > 1\,; \quad E[\|\hat{\theta}_n - \theta_o\|^2] = O(n^{-1}\log n) \quad \text{for } c = 1\,;$$

$$\text{and} \qquad E[\|\hat{\theta}_n - \theta_o\|^2] = O(n^{-c}) \quad \text{for } 0 < c < 1\,.$$

Note that 5.7(1) is just Assumption D.1P with $V(\theta) = \|\theta - \theta_o\|^2$.

COROLLARY 5.8: Given TRMP, suppose Assumptions AP, D.3P with $b = 1$ hold and that

(1) there exists a sequence $\{ c_n > 0 \}$ with $\lim_{n\to\infty} c_n = c$ for some $0 < c < \infty$ such that

$$2\,(\theta - P_{k(n)}\theta_o\,,\, M(\theta)\,) \leq -\,c_n\,\|\theta - P_{k(n)}\theta_o\|^2\,,\quad \text{for all } \theta \in H_{k(n)}\,;\ \text{and}$$

(2) $\sum_{n=1}^{\infty} n^{-1}\,\|\theta_o - P_{k(n+1)}\theta_o\| < \infty$.

(i) Then for all n sufficiently large,

$$E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(1/n) \quad \text{for } c > 1\,; \quad E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(n^{-1}\log n) \quad \text{for } c = 1\,;$$

$$\text{and} \qquad E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(n^{-c}) \quad \text{for } 0 < c < 1\,.$$

If we slightly strengthen (2) to

(3) $\|\theta_o - P_{k(n)}\theta_o\| = O(n^{-\alpha})$ for some $0 < \alpha \leq 1/2$,

(ii) then for all n sufficiently large,

$$E[\|\hat{\theta}_n - \theta_o\|^2] = O(n^{-2\alpha}) \quad \text{for } c > 1\,;$$

$$E[\|\hat{\theta}_n - \theta_o\|^2] = O(\max[\,n^{-1}\log n\,,\, n^{-2\alpha}\,]) \quad \text{for } c = 1\,;$$

and $E[\|\hat{\theta}_n - \theta_o\|^2] = O(n^{-\delta})$ for $0 < c < 1$ and $\delta = \min(c, 2\alpha)$.

Note that 5.7(1) implies 5.8(1), and that 5.8(1) and 5.8(2) together imply Assumption A.8.

Comparing Corollary 5.7 and Corollary 5.8, we can see that the rate of convergence heavily depends on assumption D.1P. If one only imposes the weaker version of Assumption D.1P, (e.g., 5.8(1)), one needs additional assumptions on the rate of increase of $k(n)$, (e.g., 5.8(3)), in order to get a comparable rate of convergence for the RMP estimators.

VI. SUMMARY

This paper has improved the current asymptotic theory on Hilbert-space valued Robbins-Monro procedures in the following ways:

(1) $M : H \to H$ is allowed to be an operator uniformly continuous on any norm-bounded set. This includes all bounded linear operators, all uniformly continuous nonlinear operators, some compact nonlinear operators, and continuous nonlinear operators with "polynomial growth" (i.e., $\|M(\theta)\| \le K[1 + \|\theta\|^p]$ for $\theta \in H$, $K > 0$, $p \in \mathbb{N}$).

(2) $U_n \equiv M_n(Z_n,\theta) - M(\theta)$ is allowed to be a Borel mapping from $G \times H$ to H , i.e., error terms $U_n(Z_n,\hat{\theta}_n)$ can be inßuenced by both the random process Z_n generated by nature, and the process θ_n generated by our various RM procedures.

(3) A large class of Hilbert space-valued dependent random processes is permitted by our assumptions on { U_n }.

(4) Our modiÞed RM procedures work without an *a priori* bound on { $\hat{\theta}_n$ } .

(5) Our modiÞed RM procedures work with Þnite-dimensional approximations.

Our almost-sure norm-convergence and mean rate of convergence results are inspired by those of Yin and Zhu (1990). We generalize their results to allow θ *–dependent* errors, to permit Hilbert space-valued mixingale error processes, and to Þnite-dimensional approximations. Our functional central limit theorems and asymptotic normality results are inspired by Nixdorf’s (1984) . We generalize his result to cover θ *–dependent* errors by following Berger’s (1986) approach. We relax his error assumptions to allow Hilbert space-valued NED functions of mixing processes by applying Walk’s (1987) and Chen and White’s (1998a) functional central limit theorems for a general class of Hilbert space-valued random processes. Our law of iterated logarithm for the θ *–dependent* mixingale error processes is new, to the best of our knowledge.

Our modiÞed RM procedures can be applied to nonparametric recursive m–estimation. We will give examples in another paper. In concrete situations, involving for example the Wiener-Hopf equation, we might have more information about the functional form of $M_n(Z_n, \theta)$, hence the error $U_n(Z_n, \theta)$. We then can make some reasonable assumptions for just the data process { Z_n } and derive the dependence structure for U_n as a consequence.

This paper ignores several important issues, such as covariance estimation, optimal stopping rules, and the possibility of multiple roots. We will address some of these problems in our future research.

All the results in this paper are proved under the assumption that $\{\hat{\theta}_n : n \ge 1\}$ does not affect $\{Z_n : n \ge 1\}$. Since it is very natural in economic time series analysis and learning models for economic agents that $\{\hat{\theta}_n\}$ affects $\{Z_n\}$ (feedback), we need to extend our current results to cover the case of $\{Z_n\}$ being Granger-caused by $\{\hat{\theta}_n\}$. Kushner's and others' methods for studying the $I\!R^d$, $(d < \infty)$ –valued stochastic approximation algorithms with feedback may well be helpful in studying convergence of inÞnite-dimensional RM algorithms with feedback. Chen and White (1998b) established almost sure convergence for H valued RM algorithms with feedbacks. It will be important to establish convergence rates and limiting distribution for the feedback case as well. The results in this paper, and those in Chen and White (1998b), allow for the study of representative agents' nonparametric adaptive learning behavior in the sense that the representative agents do not need to specify a Þxed parametric model while they are learning as new information arrives. Thus our work extends the parametric recursive least squares learning models studied by Marcet and Sargent and others. Moreover, our inÞnite-dimensional RM procedure also allows for heterogeneous agents' learning -- each can in principle use a different learning algorithm (countably many). See Chen and White (1998b) for an example. In this sense our procedure should thus have some connection with the algorithms studied by Spear (1989), and we plan to look into this in our future research.

VII. MATHEMATICAL APPENDIX

PROOF OF LEMMA 2.1: By the deÞnition of $\bar{\theta}_n$, there exists $\Omega_o \in \mathcal{F}$ with $P(\Omega_o) = 1$ such that for all $\omega \in \Omega_o$, $\|\bar{\theta}_n(\omega)\| < B$. Fix an $\omega \in \Omega_o$, and suppose that $T = \infty$; then $\hat{\theta}_n$ would cross the sphere $\{\theta : \|\theta\| = B\}$ inÞnitely often. Let $d_1 \equiv \sup\{V(\theta) : \|\theta\| < B\}$ and $d_2 \equiv \inf\{V(\theta) : \|\theta\| > B_1\}$. Then there exist $0 < \delta_1 < \delta_2$ such that $[\delta_1, \delta_2] \subset (d_1, d_2)$. Let $D \equiv \{\theta : \delta_1 \le V(\theta) \le \delta_2\} \bigcap \{\theta : \|\theta\| \le B_1\}$. Then D is a closed set. Now we can follow the proof of Proposition 4.1 in Yin and Zhu (1990) for both the RTRM and the BTRM, except that we only need Assumption A.3(1) .

PROOF OF COROLLARY 2.3: It sufÞces to show that Assumption A.5 implies A.4 when $\{U_n\}$ are θ*–independent* errors. By Assumption A.5 and the deÞnition of mixingales, we know that $\{a_n F_n(Z_n), \mathcal{F}^n\}$ is an adapted L_2– mixingale with parameters $\{\psi_m\}$ and $\{a_n c_n\}$. Let $\{b_n^{-1} = a_n\}$ in Corollary 3.8 or Corollary 3.9 in Chen and White (1996). Then by Assumption A.5, we get A.4: $\limsup_n a_n \|\sum_{j=1}^{n} F_j(Z_j)\| = 0$ $a.s.-P$.

PROOF OF COROLLARY 2.4: It sufÞces to show that Assumption A.6 and $a_n = O(n^{-1}\log n)$ imply A.4 when $\{U_n\}$ are θ*–independent* errors. By Assumption A.6, we can set $c_n = \|F_n(Z_n)\|_p$ and get $\sup_n c_n < \infty$. Let

$$b_n = a_n^{-1} \quad , \quad m(n) = O(n^{\alpha}) \text{ for some } 0 < \alpha < 1/2 \quad , \quad B_n^{1-r} = O((\log n)^{-2}) \ .$$

Then all three conditions of Theorem 3.10 in Chen and White (1996) are satisÞed, which gives us A.4 with $a_n = O(n^{-1}\log n)$.

PROOF OF COROLLARY 2.5: It sufÞces to show that conditions 2.5(1) - 2.5(3) imply A.4. By the triangle inequality we have

$$\|\sum_{j=1}^{n} a_j U_j(Z_j, \hat{\theta}_j)\| \le A_1 + A_2 , \quad \text{where}$$

$$A_1 \equiv \|\sum_{j=1}^{n} a_j [\bar{M}_j(\hat{\theta}_j) - M(\hat{\theta}_j)]\|, \qquad A_2 \equiv \|\sum_{j=1}^{n} a_j [M_j(Z_j, \hat{\theta}_j) - \bar{M}_j(\hat{\theta}_j)]\|.$$

Since $\{\hat{\theta}_n\}$ is generated by the BTRM, $\|\hat{\theta}_n\| \le \bar{B}$ $a.s.-P$ for all n . Now let $K = \bar{B}$ in Assumptions 2.5(2) and 2.5(3). By 2.5(2), $\lim_{n\to\infty} A_1 \le \sum_{j=1}^{\infty} a_j b_{\bar{B},j} < \infty$ $a.s.-P$.

By the deÞnition of $\bar{M}_n(\hat{\theta}_n)$, we have that $\{M_n(Z_n, \hat{\theta}_n) - \bar{M}_n , \mathcal{F}^n\}$ is a H–valued martingale . It is a L_2–martingale by assumption 2.5(3). Now Doob's inequality , the conditional Jensen's inequality, and 2.5(3) imply $\limsup_{n\to\infty} A_2 < \infty$ $a.s.-P$ Hence A.4 is satisÞed.

The next lemma is used in the proof of Corollary 2.6.

LEMMA A.1: Let $\{a_n\}$ satisfy Assumption A.2. For a given index set L and every $\rho \in L$, let $\{ y_{\rho,n} \}$ be a sequence of H –r.e.´s such that

$$\lim_n \sup_{\rho\in L} \| a_n \sum_{j=1}^{n} y_{\rho,n} \| = 0 \quad a.s.-P \ .$$

Then for all $\alpha > 0$ there exists an integer-valued random variable N such that for all $m > n > N$,

$$\sup_{\rho\in L} \| \Sigma_{n \le j \le m-1} a_j y_{\rho,j} \| \le \alpha \ (1 + \Sigma_{n \le j \le m-1} a_j \ ) \quad a.s.-P \ .$$

Remark: This lemma extends lemma D of Metivier and Priouret (1984, page 147) for a Þnite-dimensional space to an inÞnite-dimensional Hilbert space. The proof is similar.

PROOF OF COROLLARY 2.6: First we prove that there exists an integer-valued random variable N_o such that, for all $n \ge N_o$, there is no truncation $a.s.-P$, i.e.,

$$\text{for all } n \ge N_o \ , \ \hat\theta_{n+1} = \hat\theta_n + a_n \ [\ M(\hat\theta_n) + U_n(Z_n,\hat\theta_n) \ ] \quad a.s.-P \ .$$

Fix an $\omega \in \Omega$. We have that for all n, $B_{T(n)} \le B_n$, and by A.7(3) and A.7(4),

$$\|\hat\theta_n(\omega) + a_n \ [\ M(\hat\theta_n(\omega)) + U_n(Z_n(\omega),\hat\theta_n(\omega)) \ ] \| \le B_n + a_n \ O(B_n) \ .$$

Because $\{ \hat\theta_n \}$ is generated by the BTRM and

$$\lim_{n\to\infty} B_n = \bar{B} \ , \quad \lim_{n\to\infty} a_n = 0 \ , \quad B_{n+1} \ge B_n \ (1 + c \ a_n \ ) \ ,$$

there exists an integer $N_o(\omega)$ such that for all $n \ge N_o(\omega)$, $B_n + a_n \ O(B_n) \le B_{n+1}$. Hence there will be no truncation invoked from $N_o(\omega)$ on.

We must now show that A.4 is satisÞed, as this ensures the almost sure convergence by Theorem 2.2. To verify A.4, it sufÞces to show the following:

$$\text{for each } T > 0 \ , \quad \lim_{n\to\infty} \sup_{n < i \le m(n,T)} \| \Sigma_{n \le j \le i-1} a_j U_j(Z_j,\hat\theta_j) \| = 0 \ .$$

To prove this, we apply Lemma A.1 and follow an argument similar to that in Metivier and Priouret (1984) for the $I\!R^d$ $(d < \infty)$-valued RM. We Þx $\alpha \in (0 , 1)$, and set $\tau \equiv \alpha^{1/2} \ T$. DeÞne by recurrence $i_0 = n,\ldots, \ i_r = m \ (i_{r-1} \ , \tau \ ),\ldots,$

$$S(p,q) \equiv \Sigma_{p \le j \le q-1} \ a_j \ U_j(Z_j,\hat\theta_j) = S_1(p,q) + S_2(p,q) \ ;$$

$$S_1(p,q) \equiv \Sigma_{p \le j \le q-1} \ a_j \ U_j(Z_j,\hat\theta_p) \ ; \quad S_2(p,q) \equiv \Sigma_{p \le j \le q-1} \ a_j \ [\ U_j(Z_j,\hat\theta_j) - U_j(Z_j,\hat\theta_p) \ ] \ .$$

By the BTRM, $\|\hat\theta_{i_r}\| \le \bar{B}$ for every r. By Lemma A.1 and A.7(1), there exists $N_1(\alpha,\omega) \ge N_o(\omega)$ such that $n > N_1(\alpha,\omega)$ implies that for every i_r, and any $i \in (\ i_r \ , \ i_{r+1} \ ]$, $\| S_1(i_r,i) \| < \alpha \ (\ 1 + \tau \ )$; while A.7(2)(b) gives

$$\| S_2(i_r,i) \| \le \Sigma_{i_r \le j \le i-1} \ a_j \ \| \hat\theta_j - \hat\theta_{i_r} \| \ h_j(Z_j) \ .$$

But if $i_r < j \le i-1 < i_{r+1}$, there is no truncation, and we have

$$\|\hat{\theta}_j - \hat{\theta}_{i_r}\| \le \Sigma_{i_r \le l \le j-1}\, a_l \,\| M(\hat{\theta}_l) + U_l(Z_l,\hat{\theta}_l)\|$$

$$\le \Sigma_{i_r \le l \le j-1}\, a_l \,\| M(\hat{\theta}_l)\| + \Sigma_{i_r \le l \le j-1}\, a_l \,\| U_l(Z_l,\hat{\theta}_l)\| \equiv A_1 + A_2 \,.$$

As $\|\hat{\theta}_l\| \le \bar{B}$, by A.1(2), $\| M(\hat{\theta}_l)\| \le C_{\bar{B}}$, a constant. Since $\tau \ge \Sigma_{i_r \le l \le i_{r+1}-1}\, a_l$ by deÞnition, we have $A_1 \le \tau\, C_{\bar{B}}$. By A.7(3),

$$A_2 \le \Sigma_{i_r \le l \le j-1}\, a_l \,(\, g_{\bar{B},l}(Z_l) - E[g_{\bar{B},l}(Z_l)]\,) + g_{\bar{B}}\, \Sigma_{i_r \le l \le j-1}\, a_l \,.$$

By Lemma A.1 and A.7(3), there exists $N_2(\alpha,\omega) \ge N_o(\omega)$ such that $n > N_2(\alpha,\omega)$ implies that for every i_r, and any $j \in (\, i_r, i-1\,] \subset (\, i_r, i_{r+1}\,)$,

$$\|\hat{\theta}_j - \hat{\theta}_{i_r}\| < \alpha\,(\,1+\tau\,) + \tau\,(\,C_{\bar{B}} + g_{\bar{B}}\,) \le \alpha + \tau\,(\,1 + C_{\bar{B}} + g_{\bar{B}}\,).$$

Hence by A.7(2),

$$\| S_2(i_r,i)\| \le [\,\alpha + \tau(1+C_{\bar{B}}+g_{\bar{B}})\,]\, \Sigma_{i_r \le j \le i-1}\, a_j h_j(Z_j)$$

$$\le [\,\alpha + \tau(1+C_{\bar{B}}+g_{\bar{B}})\,]\, \Sigma_{i_r \le j \le i-1}\, a_j \,(\, h_j(Z_j) - E[h_j(Z_j)]\,) + [\,\alpha + \tau(1+C_{\bar{B}}+g_{\bar{B}})\,]\, h\,\tau \,.$$

By Lemma A.1 and A.7(2), there exists $N_3(\alpha,\omega) \ge N_o(\omega)$ such that $n > N_3(\alpha,\omega)$ implies that for every i_r, and any $i \in (\, i_r, i_{r+1}\,]$,

$$\| S_2(i_r,i)\| \le [\,\alpha + \tau(1+C_{\bar{B}}+g_{\bar{B}})\,]\,[\,\alpha(1+\tau) + h\,\tau\,] \,.$$

Hence there exists a constant $C(\bar{B},g,h)$ such that

$$\text{for every } i_r,\ i \in (\, i_r, i_{r+1}\,],\quad \| S_2(i_r,i)\| \le [\,\alpha + \tau\, C(\bar{B},g,h)\,]^2 \,.$$

Hence

$$\sup_{i_r < i \le i_{r+1}} \| S(i_r,i)\| \le \alpha\,(1+\tau) + (\alpha + c\tau)^2 = \alpha\,(1+\alpha^{1/2}T) + (\alpha + c\,\alpha^{1/2}T)^2 \le C\,\alpha \,,$$

where c, C are constants depending only on $\bar{B}$, T.
By the deÞnitions of t_n, $m(n,T)$, i_r and τ, there exists $N(\alpha,\omega)$ such that for all $n > N(\alpha,\omega)$,

$$\sup_{n < i \le m(n,T)} \| \Sigma_{n \le j \le i-1}\, a_j U_j(Z_j,\hat{\theta}_j)\| \le (\,1 + 2\,\alpha^{-1/2}\,)\, C\,\alpha \,.$$

Letting $\alpha \to 0$, we get

$$\text{for each } T > 0,\quad \lim_{n\to\infty} \sup_{n < i \le m(n,T)} \| \Sigma_{n \le j \le i-1}\, a_j U_j(Z_j,\hat{\theta}_j)\| = 0 \,.$$

This completes the proof.

PROOF OF COROLLARY 2.7: It suffices to show that all conditions in Corollary 2.6 are satisfied.

Because an adapted L_p–mixingale ($1 \le p < \infty$) has zero mean, 2.7(1) implies A.7(1)(a). By Corollary 3.8 in Chen and White (1996) with $\{ b_n^{-1} = a_n \}$, 2.7(1) implies A.7(1)(b), 2.7(2) implies A.7(2)(a), and 2.7(3) implies A.7(3)(a). Hence all conditions in Corollary 2.6 are satisfied.

We use the following lemma to prove Theorem 2.8.

LEMMA A.2: Suppose Assumptions A.1, A.2 and A.3P(1) hold for TRMP. If there exists $0 \le \varepsilon < \infty$ such that

$$\limsup_{n\to\infty} \| a_n \sum_{j=1}^{n} P_{k(j+1)}[M_j(Z_j, \hat{\theta}_j) - M(\hat{\theta}_j)] \| = \varepsilon \qquad a.s.-P \ ,$$

then there exists a positive integer-valued random variable T such that: $P\,(\sup_n T(n) \le T < \infty\,) = 1$.

PROOF OF LEMMA A.2: In the proof of Lemma 2.1, $P_{k(j+1)}M(\hat{\theta}_j)$ replaces $M(\hat{\theta}_j)$; and $P_{k(j+1)}U_j$ replaces U_j . We also use the relation

$$\| P_{k(j+1)}M(\hat{\theta}_j) - P_{k(u+1)}M(\hat{\theta}_u) \| \le \| P_{k(j+1)}M(\hat{\theta}_j) - P_{k(j+1)}M(\hat{\theta}_u) \|$$
$$+ \| P_{k(j+1)}M(\hat{\theta}_u) - P_{k(u+1)}M(\hat{\theta}_u) \| \le \| M(\hat{\theta}_j) - M(\hat{\theta}_u) \| + 2 \| M(\hat{\theta}_u) \| .$$

PROOF OF THEOREM 2.8: First, by Lemma A.2 , there is an $n_o \in I\!N$ such that for all $n \ge n_o$, the truncations are terminated. The RTRMP procedure becomes

$$(a.1) \quad \text{for } all\ n \ge n_o\ , \quad \hat{\theta}_{n+1} = \hat{\theta}_n + a_n\, P_{k(n+1)}\, M_n(Z_n\, , \hat{\theta}_n) = \hat{\theta}_n + a_n\, P_{k(n+1)}\, [M(\hat{\theta}_n) + U_n]\ ,$$

and the sequence $\{ \hat{\theta}_n\ ;\ n \ge n_o \}$ is bounded $a.s.-P$ (i.e., $\|\hat{\theta}_n\| \le B_T$, $a.s.$, for all $n \ge n_o$).
Secondly, define

$$(a.2) \quad \mu_{n+1} \equiv \mu_n - a_n \mu_n + a_n P_{k(n+1)} U_n(Z_n, \hat{\theta}_n)\ .$$

Then Assumptions A.2 and A.4P imply

$$(a.3) \quad \limsup_{n\to\infty} \|\mu_n\| = 0 \qquad a.s.-P\ .$$

$(a.1)$ and $(a.2)$ imply that for all $n \ge n_o$,

$$(a.4) \quad \hat{\theta}_{n+1} - \mu_{n+1} = \hat{\theta}_n - \mu_n + a_n P_{k(n+1)} M(\hat{\theta}_n) + a_n \mu_n\ .$$

By Taylor expansion in a Hilbert space, $(a.3)$, and the boundedness of $M, V^{\prime}$, and $V^{\prime\prime}$, we get

$$(a.5) \quad V(\hat{\theta}_{n+1} - \mu_{n+1}) \le V(\hat{\theta}_n - \mu_n) + a_n\ (\ V^{\prime}(\hat{\theta}_n)\ ,\ P_{k(n+1)} M(\hat{\theta}_n)\ ) + O(\ a_n^2\ ) \qquad a.s.-P\ ,$$

which corresponds to Yin and Zhu's (1990) inequality (4.15).
For $\theta \in H_{k(n+1)}$ and any $\eta > 0$, define

$$\lambda_{n+1}(\eta) \equiv \min [-(V'(\theta) , P_{k(n+1)}M(\theta)) : \|\theta - \theta_o\| \geq \eta , \|\theta\| \leq B_T] \; / \; \max [V(\theta) : \|\theta\| \leq B_T + 1] ,$$

and $\lambda(\eta) \equiv \inf [\lambda_{n+1}(\eta) : n \geq N_1]$, where N_1 is an integer-valued random variable. Then $\lambda(\eta) > 0$ by Assumption A.5P.2. We can now follow Yin and Zhu's (1990) proof of their theorem 3.1 (from page 127(4.16) on), and obtain the $a.s. - P$ convergence result.

PROOF OF COROLLARY 2.9: In the proof of Corollary 2.3, $P_{k(j+1)}F_j(Z_j)$ replaces $F_j(Z_j)$, and Assumption A.4P and A.5P replace A.4 and A.5 respectively.

PROOF OF COROLLARY 2.10: In the proof of Corollary 2.6, $P_{k(j+1)}U_j(Z_j,\hat{\theta}_j)$ replaces $U_j(Z_j,\hat{\theta}_j)$ and Assumption A.7P replaces A.7.

PROOF OF THEOREM 2.11: By Lemma A.2 , the truncation is invoked only Þnitely many times. The proof is similar to that for Theorem 2.8 except with the following changes:
For $\theta \in H_{k(n+1)}$ and any $\eta > 0$, deÞne

$$\lambda_{n+1}(\eta) \equiv \min [-(V'(\theta) , P_{k(n+1)}M(\theta)) : \|\theta - \theta^o{}_{n+1}\| \geq \eta , \|\theta\| \leq B_T] / \max [V(\theta) : \|\theta\| \leq B_T + 1]$$

and $\lambda(\eta) \equiv \infty [\lambda_{n+1}(\eta) : n \geq N_1]$, where N_1 is an integer-valued random variable. Then, $\lambda(\eta) > 0$ by Assumption A.8(1).

With $\|\hat{\theta}_n - \theta^o{}_{n+1}\| \leq \eta$,

$$V(\hat{\theta}_n - \mu_n) = V(\theta^o{}_{n+1}) + R_1(\theta^o{}_{n+1} , \hat{\theta}_n - \mu_n - \theta^o{}_{n+1}) \leq V(\theta^o{}_{n+1}) + K_3\,\eta + O(\varepsilon) .$$

Since A.8(2) and A.2(1) imply that $\theta^o{}_n \to \theta_o$ as $n \to \infty$, given A.3(1), we have for n large enough, say $n \geq N_2 \geq N_1$,

$$V(\theta^o{}_{n+1}) \leq K_4 \|\theta^o{}_{n+1} - \theta_o\| .$$

Hence for all $n \geq N_2$ and $\|\hat{\theta}_n - \theta^o{}_{n+1}\| \leq \eta$, $V(\hat{\theta}_n - \mu_n) \leq K_5\,\eta + O(\varepsilon)$ for some constant $K_5 > 0$. Instead of Yin and Zhu's (1990) inequality (4.16), we have

$$(a.6) \quad V(\hat{\theta}_n - \mu_n) \leq \Pi_{N_2 \leq j \leq n} (1 - a_j\,\lambda(\eta)) \, V(\hat{\theta}_{N_2} - \mu_{N_2})$$
$$+ \Sigma_{N_2 \leq i \leq n} \Pi_{i+1 \leq j \leq n} (1 - a_j\,\lambda(\eta)) [K_5\, a_i\, \eta\, \lambda(\eta) + O(a_i\, \varepsilon) + O(a_i^2)]$$
$$+ \Sigma_{N_2 \leq i \leq n} \Pi_{i+1 \leq j \leq n} (1 - a_j\,\lambda(\eta)) K_6\, a_i \|\theta^o{}_{i+1} - \theta_o\| .$$

Since

$$\Pi_{i+1 \leq j \leq n} (1 - a_j\,\lambda(\eta)) \leq \exp(-\lambda(\eta)\, \Sigma_{i \leq j \leq n}\, a_j) \to 0 \quad \text{as } n \to \infty ,$$

we have by A.8(2),

$$\lim_n \Sigma_{N_2 \leq i \leq n} \Pi_{i+1 \leq j \leq n} (1 - a_j\,\lambda(\eta)) \, a_i \|\theta^o{}_{i+1} - \theta_o\| = 0 .$$

We can now follow Yin and Zhu (1990)'s proof of their Theorem 3.1 (from page 127(4.17) on), and obtain the $a.s.-P$ convergence result.

PROOF OF COROLLARY 2.12: When $\{U_n\}$ is a sequence of θ–independent errors, A.5P implies A.4P. Hence the result follows from Theorem 2.11.

PROOF OF COROLLARY 2.13: When $\{U_n\}$ is a sequence of θ–dependent errors, A.7P implies A.4P just as A.7 implies A.4 in the proof of Corollary 2.6. Hence the result follows from Theorem 2.11.

We use the next lemma to prove Proposition 2.14.

LEMMA A.3: Let x_n, u_n, v_n, and w_n be nonnegative $\mathcal{A}^n$–measurable real valued random variables, $n = 1, ...,$ deÞned on a complete probability space $(\Omega, \mathcal{F}, P)$, where $\{\mathcal{A}^n\}$ is an increasing family of subσ-algebras of $\mathcal{F}$. Suppose

$$\sum_{n=1}^{\infty} u_n < \infty \quad a.s.-P \quad ; \qquad \sum_{n=1}^{\infty} v_n < \infty \quad a.s.-P \quad ;$$

$$E[\,x_1^{\,2}\,] < \infty \quad \text{and} \quad E[\,x_{n+1}^{\,2} \mid \mathcal{A}^n\,] \le (1+u_n)\,x_n^{\,2} + v_n - b_n w_n \quad a.s.-P\ ,$$

where $\{b_n\,,\ n=1,2,...\}$ is non-random real sequence such that $b_n > 0\,,\ \sum_{n=1}^{\infty} b_n = \infty\,.$

Then there exists a random variable x on $(\Omega, \mathcal{F}, P)$, and a subsequence $\{n(j),\ j=1,2,...\}$ such that:

$$x_n \to x \quad as\ n \to \infty \quad a.s.-P \qquad \text{and} \qquad w_{n(j)} \to 0 \quad as\ j \to \infty \quad a.s.-P\ .$$

PROOF: This result is a slight modiÞcation of Goldstein's (1988) Lemma 4.1. It can be proved by mimicing the proof of Goldstein's (1988) Theorem 4.1.

PROOF OF PROPOSITION 2.14: By deÞnition,

$$\hat\theta_{n+1} - \theta_o = [(\hat\theta_n - \theta_o)\,] + [a_n\,P_{k(n+1)}\,M_n(Z_n\,,\hat\theta_n)]\,.$$

Taking the inner product on both sides gives

$$\|\hat\theta_{n+1} - \theta_o\|^2 = \|\hat\theta_n - \theta_o\|^2 + \|a_n\,P_{k(n+1)}M_n(Z_n\,,\hat\theta_n)\|^2$$
$$+\,2\,a_n\,(\hat\theta_n - \theta_o\,,\,P_{k(n+1)}M_n(Z_n,\hat\theta_n)\,)\quad .$$

We shall take the conditional expectation $E[\,\cdot \mid \mathcal{F}^{n-1}\,]$ on both sides, and make use of the following facts:

(1) Since $\hat\theta_n$ is $\mathcal{F}^{n-1}$ – measurable;

$$E[\,\|\hat\theta_n - \theta_o\|^2 \mid \mathcal{F}^{n-1}] = \|\hat\theta_n - \theta_o\|^2 \equiv x_n^{\,2}\ ;$$

(2) by condition 2.14(2),

$$E[\,\|a_n P_{k(n+1)} M_n(Z_n,\hat\theta_n)\|^2 \mid \mathcal{F}^{n-1}] \le v_n + u_n\,x_n^{\,2}\ ,$$

where $v_n \equiv a_n^2\, E[\, h_n(Z_n) \mid \mathcal{F}^{n-1}\,] \geq 0 \;\; a.s.-P$; and $u_n \equiv a_n^2\, E[\, g_n(Z_n) \mid \mathcal{F}^{n-1}\,] \geq 0 \;\; a.s.-P$;

(3) By condition 2.14(1),

$$w_n \equiv -E[\,(\hat{\theta}_n - \theta_o\,,\, P_{k(n+1)} M_n(Z_n, \hat{\theta}_n)\,) \mid \mathcal{F}^{n-1}] \geq 0 \;\; a.s.-P \,.$$

Letting $b_n \equiv 2\, a_n$ and $x_n \equiv \| \hat{\theta}_n - \theta_o \| \geq 0 \;\; a.s.-P$, we get

$$E\,[x_{n+1}{}^2 \mid \mathcal{F}^{n-1}\,] \leq (\,1 + u_n\,)\, x_n{}^2 + v_n - b_n\, w_n \quad a.s.-P \,.$$

By condition 2.14(2)(b) and $E[\,x_1{}^2\,] < \infty$, all hypotheses of Lemma A.3 are satisÞed. Hence there exists a random variable x on $(\,\Omega, \mathcal{F}, P\,)$, and a subsequence $\{n(j),\, j=1,2,...\}$ such that:

$$x_n \to x \quad as\; n \to \infty \quad a.s.-P \qquad \text{and} \qquad w_{n(j)} \to 0 \quad as\; j \to \infty \quad a.s.-P \,.$$

Further, we can show that $x = 0 \;\; a.s.-P$ (see Goldstein, 1988), i.e., $\| \hat{\theta}_n - \theta_o \| \to 0 \;\; as\;\; n \to \infty \;\; a.s.-P$.

PROOF OF PROPOSITION 2.15: By deÞnition,

$$\hat{\theta}_{n+1} - P_{k(n+1)}\theta_o = \hat{\theta}_n - P_{k(n+1)}\theta_o + a_n P_{k(n+1)} M_n(Z_n, \hat{\theta}_n) \,.$$

Let $x_{n+1} \equiv \| \hat{\theta}_{n+1} - P_{k(n+1)}\theta_o \|$. Then

$$x_{n+1}^2 = \| \hat{\theta}_n - P_{k(n+1)}\theta_o \|^2 + a_n{}^2 \| P_{k(n+1)} M_n(Z_n, \hat{\theta}_n) \|^2 + 2\, a_n\, (\,\hat{\theta}_n - P_{k(n+1)}\theta_o\,,\, P_{k(n+1)} M_n(Z_n, \hat{\theta}_n)\,) \,.$$

We shall take the conditional expectation $E[\,\cdot \mid \mathcal{F}^{n-1}\,]$ on both sides, and make use of the following facts:

(1) since $\hat{\theta}_n$ is $\mathcal{F}^{n-1}-$ measurable, so is $\hat{\theta}_n - P_{k(n+1)}\theta_o$, therefore,

$$E[\, \| \hat{\theta}_n - P_{k(n+1)}\theta_o \|^2 \mid \mathcal{F}^{n-1}] = \| \hat{\theta}_n - P_{k(n+1)}\theta_o \|^2 \,;$$

(2) by condition 2.15(2),

$$E[\, \| a_n P_{k(n+1)} M_n(Z_n, \hat{\theta}_n) \|^2 \mid \mathcal{F}^{n-1}] \leq v_{1n} + u_{1n} \| \hat{\theta}_n - P_{k(n+1)}\theta_o \|^2 \,,$$

where $v_{1n} \equiv a_n^2\, E[\, h_n(Z_n) \mid \mathcal{F}^{n-1}\,] \geq 0 \;\; a.s.-P$; and $u_{1n} \equiv a_n^2\, E[\, g_n(Z_n) \mid \mathcal{F}^{n-1}\,] \geq 0 \;\; a.s.-P$;

(3) by condition 2.15(1),

$$w_n \equiv -E[\,(\hat{\theta}_n - P_{k(n+1)}\theta_o\,,\, M_n(Z_n, \hat{\theta}_n)\,) \mid \mathcal{F}^{n-1}] \geq 0 \;\; a.s.-P \,.$$

Letting $b_n \equiv 2\, a_n$ and $x_n \equiv \| \hat{\theta}_n - P_{k(n)}\theta_o \| \geq 0 \;\; a.s.-P$, we get

$$E\,[x_{n+1}{}^2 \mid \mathcal{F}^{n-1}\,] \leq (\,1 + u_{1n}\,) \| \hat{\theta}_n - P_{k(n+1)}\theta_o \|^2 + v_{1n} - b_n\, w_n \quad a.s.-P \,.$$

By the triangle inequality,

$$\|\hat{\theta}_n - P_{k(n+1)}\theta_o\| \leq \|\hat{\theta}_n - P_{k(n)}\theta_o\| + \|P_{k(n)}\theta_o - P_{k(n+1)}\theta_o\| \equiv x_n + r_n .$$

Hence,

$$\|\hat{\theta}_n - P_{k(n+1)}\theta_o\|^2 \leq x_n^2 + r_n^2 + 2\, x_n\, r_n \leq (1+r_n)\, x_n^2 + (r_n + r_n^2) .$$

(The second inequality is due to $r_n \geq 0$ and $0 \leq x_n \leq 2\, x_n \leq x_n^2 + 1$). Hence,

$$E[\, x_{n+1}^2 \mid \mathcal{F}^{n-1}\,] \leq (1+u_{1n})\,(1+r_n)\, x_n^2 + (1+u_{1n})\,(r_n + r_n^2) + v_{1n} - b_n w_n .$$

Let

$$u_n \equiv u_{1n} + r_n + u_{1n}\, r_n \;;\; v_n \equiv (1+u_{1n})\,(r_n + r_n^2) + v_{1n} .$$

By condition 2.15(3), $\sum_{n=1}^{\infty} r_n < \infty$; since $r_n \geq 0$ and condition 2.15(2)(b) holds, we have $\sum_{n=1}^{\infty} u_n < \infty$ $a.s.-P$ and $\sum_{n=1}^{\infty} v_n < \infty$ $a.s.-P$. Since $E[\, x_1^2\,] < \infty$, all hypotheses of Lemma A.3 are satisfied. Hence there exists a random variable x on $(\Omega, \mathcal{F}, P)$, and a subsequence $\{n(j),\ j=1,2,...\}$ such that:

$$x_n \to x \quad as\ n \to \infty \quad a.s.-P \quad \text{and} \quad w_{n(j)} \to 0 \quad as\ j \to \infty \quad a.s.-P .$$

Further, we can show that $x = 0$ $a.s.-P$ (see Goldstein, 1988), i.e., $\|\hat{\theta}_n - P_{k(n)}\theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$. By the triangle inequality:

$$\|\hat{\theta}_n - \theta_o\| \leq \|\hat{\theta}_n - P_{k(n)}\theta_o\| + \|P_{k(n)}\theta_o - \theta_o\| .$$

We conclude $\|\hat{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ $a.s.-P$.

PROOF OF THEOREM 3.1: We verify that all conditions of Walk's (1987) Theorem 1 are satisfied as follows: First, Walk's (1987) condition (1) is satisfied with $T_n \equiv T_{1n} + T_{2n}$. Second, our B.2 satisfies his condition (2). Third, our B.1 and $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$ imply that $\|A_n - A\| \to 0$ $a.s.-P$, hence his condition (3) is satisfied; this together with B.4(1) imply his condition (4). Fourth, our B.5 is his condition (5). Finally, our B.3 and $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$ imply that $T_{2n} \to 0$ almost in first mean; this together with B.4(2) imply Walk's (1987) condition (6) with $T = 0$ via Lemma 3.3. Thus, the result follows from Walk's (1987) Theorem 1.

We prove that $T_{2n} \to 0$ almost in first mean as follows: for any fixed $0 < \varepsilon < 1$, $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$ and Egorov's theorem imply that there exists $\Omega' \in \mathcal{F}$ with $P(\Omega') \geq 1-\varepsilon$ such that $\|\hat{\theta}_n - \theta_o\| \to 0$ uniformly in Ω' , i.e., for any $\eta > 0$, there exists $m \in I\!N$ such that for any $n \geq m$ and $\omega \in \Omega'$, $\|\hat{\theta}_n(\omega) - \theta_o\| \leq \eta$. We now define a new sequence $\{\ddot{\theta}_n\}$ by

$$\ddot{\theta}_1 \equiv \hat{\theta}_{m+1}\, 1\,(\|\hat{\theta}_{m+1} - \theta_o\| \leq \eta\,) ,$$

$$\ddot{\theta}_{n+1} \equiv \ddot{\theta}_n - (n+m)^{-1}\, P_{k(n+m+1)}\, [\, \tilde{M}(\ddot{\theta}_n) + U_{n+m}(Z_n, \ddot{\theta}_n)\,] ,$$

where $\tilde{M}(\theta) \equiv M(\theta)$ if $\|\theta - \theta_o\| \leq \eta$, and $\tilde{M}(\theta) \equiv A\,\theta$ otherwise. By the deÞnition, $\ddot{\theta}_n$ and $\hat{\theta}_{m+n}$ coincide on Ω', hence $\|\ddot{\theta}_n - \theta_o\| \to 0$ as $n \to \infty$ uniformly in Ω'. For η small enough, we follow Walk's (1977) proof of his equation 44 (p.149) and get $n^{\beta} E[\|\ddot{\theta}_n - \theta_o\|^2] = O(1)$. Given the deÞnition of T_{2n} and assumption B.3, we get

$$[\int_{\Omega'} \|T_{2,\,n+m}\| \, d\,P\,]^2 \leq o(1)\,(n+m)^{\beta} E[\|\ddot{\theta}_n - \theta_o\|^2] + o(1) \to 0 \quad \text{as } n \to \infty .$$

Hence $T_{2n} \to 0$ almost in Þrst mean.

PROOF OF COROLLARY 3.2: (i) is obviously true given Theorem 3.1. By the continuous mapping theorem we get result (ii).

PROOF OF PROPOSITION 3.4: (i) Under B.6(1), $\nu_j = j^{(\beta-1)/2} P_{k(j+1)} U_j(Z_j,\theta_o)$. B.6(4) implies that $\{\nu_n, \mathcal{F}^n\}$ is an adapted $L_2(H)$ – mixingale sequence with parameters $\{\psi_m\}$ of size $-1/2$, and $\{n^{(\beta-1)/2} c_n\}$. By Chen and White (1996, Theorem 3.7) and assumption B.6(4),

$$E\,[\|\sum_{j=1}^{n} \nu_j\|^2] = O\,(\sum_{j=1}^{n} j^{\beta-1} c_j^{\,2}) = O\,(n) .$$

Hence B.4(1'') is satisÞed, and so is B.4(1) via Lemma 3.3(ii).

(ii) Under B.6(1),

$$T_{1n} = n^{\beta/2} P_{k(n+1)} [\, U_n(Z_n,\hat{\theta}_n) - U_n(Z_n,\theta_o)\,] .$$

Since $\{\psi_m\}$ is decreasing and is of size $-\beta$, there exists $\mu : I\!N \to I\!N$ such that

$$1 < \mu(j) < j \quad \text{for } j > 2 , \quad \mu(j) = o\,(j^{1/2}) , \quad \psi_{\mu(j)} = o\,(j^{-\beta/2}) .$$

Take $\varepsilon \in (0, 1/2)$. As $\|\hat{\theta}_n - \theta_o\| \to 0$ $a.s.-P$, there exists $K(\varepsilon) > 0$ such that

$$P\,[\,\omega : \sup_n \|\hat{\theta}_n - \theta_o\| \geq K(\varepsilon)\,] \leq \varepsilon .$$

DeÞne

$$\hat{T}_{1n} \equiv n^{\beta/2} P_{k(n+1)} [\, U_n(Z_n , \hat{\theta}_n 1[\|\hat{\theta}_n - \theta_o\| \leq K(\varepsilon)]\,) - U_n(Z_n , \theta_o)\,] .$$

Then $P\,(\bigcup_{n\geq 1} [\,\hat{T}_{1n} \neq T_{1n}\,]) \leq \varepsilon$. Hence it sufÞces to show that there exists $\delta^* \in (\delta_o , 1/2)$ such that

$$n^{-1/2} \max_{3\leq j\leq n} (n/j)^{\delta^*} \|\sum_{i=3}^{j} i^{-(1/2)} \hat{T}_{1i}\| \to 0 \quad \text{in Prob.} \quad \text{as } n \to \infty .$$

Write $\hat{T}_{1n} \equiv n^{1/2} \sum_{j=1}^{3} D_{jn}$, where

$$D_{1n} \equiv \bar{T}_n - E(\,\bar{T}_n \,|\, \mathcal{F}^{n-\mu(n)}) \ ; \quad D_{2n} \equiv E(\,\bar{T}_n \,|\, \mathcal{F}^{n-\mu(n)}) \ ;$$

$$\overline{T}_n \equiv n^{(\beta-1)/2}\,[\,P_{k(n+1)}\,U_n(\cdot\,,\hat{\theta}_{n-\mu(n)}1[\,\|\hat{\theta}_{n-\mu(n)}-\theta_o\,\|\le K(\varepsilon)]\,)-P_{k(n+1)}U_n(\cdot\,,\theta_o)\,]\;;$$

$$D_{3n} \equiv n^{(\beta-1)/2}\,P_{k(n+1)}U_n(\cdot\,,\hat{\theta}_n 1[\,\|\hat{\theta}_n-\theta_o\,\|\le K(\varepsilon)]\,)$$
$$-\,n^{(\beta-1)/2}\,P_{k(n+1)}\,U_n(\cdot\,,\hat{\theta}_{n-\mu(n)}1[\,\|\hat{\theta}_{n-\mu(n)}-\theta_o\,\|\le K(\varepsilon)]\,)\,.$$

By Lemma 3.3(iv), it suffices to show that: (a) $D_{1n}\to 0$ almost in first mean; (b) $D_{2n}\to 0$ almost in first mean; and (c) $D_{3n}\to 0$ almost in first mean.

Since B.6(2) implies that $P_{n+1}U_n(\omega\,,\cdot)$ is continuous for each $\omega\in\Omega$, and since $\|\hat{\theta}_n-\theta_o\|\to 0$ $a.s.-P$ and $\beta\in(0,1]$, we get $\|\overline{T}_n\|\to 0$ $a.s.-P$. This together with B.6(3) implies that $E\,[\,\|\overline{T}_n\|^r\,]\to 0$ (see Serfling, 1980, p.11, theorem 1.3.7). Hence $E[\,\|\overline{T}_n\|\,]\to 0$. By the conditional Jensen's inequality,

$$E[\,\|D_{2n}\|\,]=E[\,\|E(\overline{T}_n\,|\,\mathcal{F}^{n-\mu(n)})\|\,]\le E[\,\|\overline{T}_n\|\,]\to 0 \quad \text{as } n\to\infty\,,$$

this implies (b). Further

$$E[\,\|D_{1n}\|\,]\le E[\,\|\overline{T}_n\|\,]+E[\,\|D_{2_n}\|\,]\to 0 \quad \text{as } n\to\infty\,,$$

which gives us (a).

Let $\Omega_\varepsilon\equiv\Omega\setminus[\,\omega:\sup_n\|\hat{\theta}_n-\theta_o\|\ge K(\varepsilon)\,]$. Then $P[\,\omega\in\Omega_\varepsilon\,]\ge 1-\varepsilon$.

Define $m(j)=\max[\,i\in\mathbb{N}: j+i-\mu(j+i)<j\,]$. By assumption B.6(2),

$$E[\,\|D_{3n}\|\,1(\Omega_\varepsilon)]$$
$$\le n^{(\beta-1)/2}\sup_n[\,(\,E\phi^2_{B(\varepsilon),n}\,)^{1/2}\,(\,E[\,\|\tilde{\theta}_n-\tilde{\theta}_{n-\mu(n)}\|^2]\,1(\Omega_\varepsilon)\,)^{1/2}\,]\,,$$

where

$$\tilde{\theta}_n=\hat{\theta}_n\,1(\,\|\hat{\theta}_n-\theta_o\|\le K(\varepsilon)\,)\;;\;\tilde{\theta}_{n-\mu(n)}=\hat{\theta}_{n-\mu(n)}\,1(\,\|\hat{\theta}_{n-\mu(n)}-\theta_o\|\le K(\varepsilon)\,)\,,$$

and $B(\varepsilon)\equiv[\,\theta\in H:\|\theta-\theta_o\|\le K(\varepsilon)\,]$. Hence we get (c) provided

$$(d)\quad \sup_{j\le m(n)}[\,(\,E\,\|\hat{\theta}_n-\hat{\theta}_{n-j}\|^2\,1(\Omega_\varepsilon)\,)^{1/2}\,]=o\,(\,n^{-\beta/2}\,)\,.$$

Because

$$\hat{\theta}_{n+1}-\hat{\theta}_n=n^{-1}A_n(\hat{\theta}_n-\theta_o)+n^{-1}P_{k(n+1)}U_n(\cdot\,,\hat{\theta}_n)+n^{-1-(\beta/2)}T_{2n}\,,$$

and $T_{2n}\to 0$ *in first mean* implies $T_{2n}\to 0$ *in Prob*, we get

$$(e)\quad (\,E\,\|\hat{\theta}_{n+1}-\hat{\theta}_n\|^2\,1(\Omega_\varepsilon)\,)^{1/2}\le n^{-1}\,(\,E\,\|A_n(\tilde{\theta}_n-\theta_o)\|^2\,1(\Omega_\varepsilon)\,)^{1/2}$$

$$+\,n^{-1}\,(\,E\,\|P_{k(n+1)}U_n(\cdot\,,\tilde{\theta}_n)\|^2\,1(\Omega_\varepsilon)\,)^{1/2}+o\,(\,n^{-1}\,)=O\,(\,n^{-1}\,)\,,$$

where the last relation is implied by B.6(3) and the definition of Ω_ε. Since

$$m(n) = \max [\, j \in I\!N : n + j - \mu(n+j) < n \,] = \max [\, j \in I\!N : j < \mu(n+j) \,] ,$$

and $\mu(i) = o\,(i^{1/2})$, $1 < \mu(i) < i$ for all $i > 2$, we have $m(n) = o\,(n^{1/2})$. This, (e), and $\beta \in (0,1]$ give us (d), hence (c). This completes the proof.

PROOF OF LEMMA 3.6: Immediate, as it follows from Chen and White's (1992), Theorem 4.14.

PROOF OF COROLLARY 3.7: By Lemma 4.2 in Chen and White (1996), B.6(1), B.6(3) and 3.6(1) imply that B.6(4) is satisÞed. This together with B.6(2) and $\|\hat{\theta}_n - \theta_o\| \to 0 \;\; a.s.-P$ imply that B.4 is satisÞed via Proposition 3.4. B.6(1) and B.6(3) imply that $\{ W_n = v_n = n^{(\beta-1)/2} P_{k(n+1)} U_n(Z_n, \theta_o) \}$ has zero means and uniformly bounded $r-th$ moments. This together with 3.6(1) and 3.6(2) imply by Lemma 3.6 that $\{ v_n = n^{(\beta-1)/2} P_{k(n+1)} U_n(Z_n, \theta_o) \}$ satisÞes B.5 when X is a Brownian motion in H with $X(0) = 0$, $EX(1) = 0$ and Cov $X(1) = S$. Now all the conditions of Corollary 3.2 are satisÞed and the results follows.

PROOF OF COROLLARY 3.8: The proof is similar to that for Corollary 3.7 except with $P_{k(n+1)}$ omitted, and using Chen and White's (1998a) Theorem 3.9 or Chen and White's (1992) Corollary 4.16 instead of Chen and White's (1992) Theorem 4.14.

PROOF OF LEMMA 3.9: We write $\tilde{K}_n - K \equiv a_{1n} + a_{2n} + a_{3n}$, where

$$a_{1n} \equiv \int_{(0,1]} s^{\tilde{\Gamma}_n}\, \tilde{S}_n\, [\, s^{\tilde{\Gamma}_n^*} - s^{\Gamma^*} \,]\, ds ;$$

$$a_{2n} \equiv \int_{(0,1]} [\, s^{\tilde{\Gamma}_n} - s^{\Gamma} \,]\, \tilde{S}_n\, s^{\Gamma^*}\, ds ;$$

$$a_{3n} \equiv \int_{(0,1]} s^{\Gamma}\, [\, \tilde{S}_n - S \,]\, s^{\Gamma^*}\, ds .$$

Since

$$\| a_{1n} \|_{tr} \le \| \tilde{S}_n \|_{tr} \int_{(0,1]} \| s^{\tilde{\Gamma}_n} \| \| s^{\tilde{\Gamma}_n^*} - s^{\Gamma^*} \| ds ,$$

$$\| a_{2n} \|_{tr} \le \| \tilde{S}_n \|_{tr} \int_{(0,1]} \| s^{\tilde{\Gamma}_n} - s^{\Gamma} \| \| s^{\Gamma^*} \| ds ,$$

$$\| a_{3n} \|_{tr} \le \| \tilde{S}_n - S \|_{tr} \int_{(0,1]} \| s^{\Gamma} \|^2 ds ,$$

under the assumptions, we have

$$\lim_{n\to\infty} \| a_{jn} \|_{tr} = 0 \quad \textit{in Prob.} \quad \text{or} \quad a.s.-P \quad \text{for } j = 1, 2, 3 .$$

This gives us $\lim_n \| \tilde{K}_n - K \|_{tr} = 0$ *in Prob.* or $a.s.-P$.

PROOF OF COROLLARY 4.1: It is easy to verify that all conditions of Walk's (1987) Theorem 2 are satisÞed. Hence the result 4.1(i) follows. Since

$$a_1(t) - a_2(t) \le [\, 2\lambda\, t \log\log t \,]^{-1/2} \| [t]^{(1+\beta)/2} (\hat{\theta}_{[t]+1} - \theta_o) \| \le a_1(t) + a_2(t) ,$$

where $a_1(t) \equiv [\, 2\lambda\, t \log\log t \,]^{-1/2} \| GM(t) \|$, and

$$a_2(t) \equiv [\, 2\lambda\, t \log\log t \,]^{-1/2} \| [t]^{(1+\beta)/2} (\hat{\theta}_{[t]+1} - \theta_o) - GM(t) \| ,$$

we have

$$\limsup_{t\to\infty} a_1(t) - \liminf_{t\to\infty} a_2(t)$$
$$\leq \limsup_{t\to\infty} [\, 2\lambda\ t \log\log t\,]^{-1/2} \|\, [t]^{(1+\beta)/2} \,(\hat{\theta}_{[t]+1} - \theta_o)\, \|$$
$$\leq \limsup_{t\to\infty} a_1(t) + \limsup_{t\to\infty} a_2(t)\,.$$

Einmahl's (1991) theorem (p. 1228) gives that

$$\limsup_{t\to\infty} [\, 2\lambda\ t \log\log t\,]^{-1/2} \|\, GM(t)\, \| = 1 \quad a.s.\!-\!P\ ,$$

where $\lambda \equiv \sup [\, (\, K h\, ,\, h\,) : \|\, h\, \| \leq 1\ ,\ h \in H\,]$ is the largest eigenvalues of K. Also, result 4.1(i) implies

$$\liminf_{t\to\infty} a_2(t) = \limsup_{t\to\infty} a_2(t) = \lim_{t\to\infty} a_2(t)$$

$$= [\, 2\lambda\,]^{-1/2} \lim_{t\to\infty} [\, t \log\log t\,]^{-1/2} \|\, [t]^{(1+\beta)/2} \,(\hat{\theta}_{[t]+1} - \theta_o) - GM(t)\, \| = 0 \quad a.s.\!-\!P\ .$$

This gives us result 4.1(ii).

PROOF OF LEMMA 4.2: We simply verify that all conditions of Philipp's (1986) Theorem 1 are satisÞed by our random sequence $\{\, W_n\ ;\ n \geq 1\, \}$. The uniformly bounded L_r –norm ($r > 2$) of $\{\, W_n\, \}$ is Philipp's condition (2.1). Since $\{\, W_n\, \}$ is an L_p–mixingale of size –1 $(1 \leq p < \infty)$, Philipp's condition (2.2) is satisÞed. Condition 4.2(2) is Philipp's condition (2.3). Now Philipp's (1986) theorem 1 gives results (i) and (ii).

PROOF OF LEMMA 4.3: Condition 3.8(1) and Lemma 4.2 in Chen and White (1996) imply 4.2(1). Also the assumptions and Lemma 3.8 in Chen and White (1992) imply 4.2(2). Thus results 4.2(i) and 4.2(ii) hold. The proof of result 4.3(i) is akin to that for result 4.1(ii). Because

$$b_1(n) - b_2(n) \leq [\, 2\tau_n\ n \log\log n\,]^{-1/2} \|\, \sum_{j=1}^{n} W_j\, \| \leq b_1(n) + b_2(n)\ ,$$

where

$$b_1(n) \equiv [\, 2\tau_n\ n \log\log n\,]^{-1/2} \|\, \sum_{j=1}^{n} \mathcal{N}_j\ (0,S)\, \|\ ,$$
$$b_2(n) \equiv [\, 2\tau_n\ n \log\log n\,]^{-1/2} \|\, \sum_{j=1}^{n} W_j - \sum_{j=1}^{n} \mathcal{N}_j\ (0,S)\, \|\ ,$$

we have

$$\limsup_{n\to\infty} b_1(n) - \liminf_{n\to\infty} b_2(n) \leq \limsup_{n\to\infty} [\, 2\tau_n\ n \log\log n\,]^{-1/2} \|\, \sum_{j=1}^{n} W_j\, \|$$
$$\leq \limsup_{n\to\infty} b_1(n) + \limsup_{n\to\infty} b_2(n)\,.$$

Einmahl's (1991) theorem (p. 1228) gives

$$\limsup_{n\to\infty} [\, 2\tau\, n \log\log n \,]^{-1/2} \| \sum_{j=1}^{n} \mathcal{N}_j(\,0\,,\,S\,) \| = 1 \quad a.s.-P \;\; ,$$

where $\tau \equiv \sup [\,(\,S\,h\,,\,h\,) : \|h\| \le 1\,,\; h \in H\,]$ is the largest eigenvalues of S. Because $\lim_{n\to\infty} \tau_n = \tau$ by Lemma 2.1(ii) in Chen and White (1998a), and $0 < \tau < \infty$, we have $\limsup_{n\to\infty} b_1(n) = 1 \;\; a.s.-P$. Also Lemma 4.2(i) and 4.2(ii) imply

$$\liminf_{n\to\infty} b_2(n) = \limsup_{n\to\infty} b_2(n) = \lim_{n\to\infty} b_2(n)$$

$$= [\, 2\tau \,]^{-1/2} \lim_{n\to\infty} [\, n \log\log n \,]^{-1/2} \| \sum_{j=1}^{n} W_j - \sum_{j=1}^{n} \mathcal{N}_j\,(0,S) \| = 0 \quad a.s.-P \;.$$

This gives us result 4.3(i).

PROOF OF PROPOSITION 4.5: The proof is similar to those of Corollaries 2.6 and 2.7, except with the following changes: Let $a_n = n^{-1}\,(\,\log\log n\,)^{-1/2}\, n^{\beta/2}$ and

$$S(p,q) \equiv \Sigma_{p \le j \le q-1}\; a_j\; P_{k(j+1)}\; [\, U_j(Z_j\,,\hat{\theta}_j) - U_j(Z_j\,,\theta_o)\,] \equiv S_1(p,q) + S_2(p,q)\;,$$

where,

$$S_1(p,q) \equiv \Sigma_{p \le j \le q-1}\; a_j\; P_{k(j+1)}\; [\, U_j(Z_j\,,\hat{\theta}_p) - U_j(Z_j\,,\theta_o)\,] \;\; ,$$

and

$$S_2(p,q) \equiv \Sigma_{p \le j \le q-1}\; a_j\; P_{k(j+1)}\; [\, U_j(Z_j\,,\hat{\theta}_j) - U_j(Z_j\,,\hat{\theta}_p)\,] \;.$$

Now Assumption C.1 and $\|\hat{\theta}_n - \theta_o\| \to 0 \;\; a.s.-P$ and the algorithm specified at the beginning of Section III allow us to follow the proof of Corollaries 2.6 and 2.7 to yield:

$$\limsup_{n\to\infty} n^{-1}\,(\,\log\log n\,)^{-1/2} \| \sum_{j=1}^{n} T_{1j} \| = 0 \quad a.s.-P \;.$$

By a similar proof to that of Yin & Zhu's (1990) Theorem 3.3, since $\|\hat{\theta}_n - \theta_o\| \to 0 \;\; a.s.-P$, we get

$$\lim_{n\to\infty} n^{-1}\,(\,\log\log n\,)^{-1/2} \| \sum_{j=1}^{n} T_{1j} \| = 0 \quad a.s.-P \;,$$

This completes the proof.

PROOF OF COROLLARY 4.7: Condition 4.1(1) is satisfied due to Lemma 4.3. Condition 4.1(2) is satisfied as the similar proof of Proposition 4.5. Condition 4.1(3) is trivially satisfied since $T_{2n} \equiv 0$ in the RM and TRM algorithms. The result follows.

The next two lemmas are used in the proof of Lemma 5.1.

LEMMA A.4: Given TRMP, Assumptions AP and D.1P(1), if for all n,

$$E[\, \| P_{k(n+1)} U_n(Z_n,\hat{\theta}_n) \|^2 \,] < \infty\;, \quad E[\; |\,\overline{V}(\hat{\theta}_n,n)\,|\;] < \infty\;, \;\text{ and }\; E[\; |\,\overline{DV_n}\,|\;] < \infty \;\;,$$

then there exists $n_o \in I\!N$, such that for all $n \geq n_o$,

$$E[\,S(n+1)\,] \leq (\,1 - n^{-1} c_n)\, E[\,S(n)\,] + n^{-1}\, c_n\, E[\,\overline{V}(\hat{\theta}_n\,,\,n)\,] + O\,(n^{-2}) + O(n^{-2})\, E[\,\| P_{k(n+1)} U_n(Z_n, \hat{\theta}_n) \|^2\,] + E[\,\overline{DV_n}\,]\ .$$

PROOF OF LEMMA A.4: First, substitute $a_n = n^{-1}$ into $(a.1)$ to give,

$$(a.7) \quad \text{for all } n \geq n_o\,, \quad \hat{\theta}_{n+1} = \hat{\theta}_n + n^{-1} P_{k(n+1)}\, [M(\hat{\theta}_n) + U_n(Z_n, \hat{\theta}_n)]\ .$$

Given Assumptions AP and D.1P(1), the sequence $\{\hat{\theta}_n\,;\ n \geq n_o\}$ is bounded $a.s.-P$ and converges to θ_o in norm $a.s.-P$. Hence there exists a positive non-random real number $\bar{B}$ such that $\|\hat{\theta}_n\| \leq \bar{B}$ $a.s.-P$. By Taylor expansion in Hilbert space,

$$V(\hat{\theta}_{n+1}) = V(\hat{\theta}_n) + n^{-1}\,(\,V'(\hat{\theta}_n)\,,\, P_{k(n+1)} M(\hat{\theta}_n)\,)$$

$$+\, n^{-1}\,(\,V'(\hat{\theta}_n)\,,\, P_{k(n+1)}\, U_n(Z_n, \hat{\theta}_n)\,) + R_2(\,\hat{\theta}_n\,,\, \hat{\theta}_{n+1} - \hat{\theta}_n\,)\ ,$$

where $R_2(\,x\,,\,h\,) = \int_{[0,1]}\,(\,1-s\,)\, V''(\,x + sh\,)\, h^2\, d\,s$, and

$$\| R_2(\,\hat{\theta}_n\,,\, \hat{\theta}_{n+1} - \hat{\theta}_n\,) \| \leq \sup_{\|\theta\| \leq \bar{B}} \| V''(\theta) \| \| \hat{\theta}_{n+1} - \hat{\theta}_n \|^2$$

$$\leq K\, n^{-2} \sup_{\|\theta\| \leq \bar{B}} \| V''(\theta) \| \,[\, \| M(\hat{\theta}_n) \|^2 + \| U_n(Z_n, \hat{\theta}_n) \|^2\,] \quad a.s.-P \quad \text{for some } K > 0\ .$$

The uniform boundedness of V', V'', M imply that there exist finite positive non-random real numbers K_1, K_2, K_3 such that for any $\|\theta\| \leq \bar{B}$,

$$(a.8.i) \quad \| V'(\theta) \| \leq K_1\ ,\quad \| V''(\theta) \| \leq K_2\ ,\quad \| M(\theta) \| \leq K_3\ .$$

Thus $V(\hat{\theta}_n)$, $V'(\hat{\theta}_n)$, and $M(\hat{\theta}_n)$ have finite first and second moments. Furthermore,

$$(a.8.ii) \quad E[\, \| V'(\hat{\theta}_n) \| \,] \leq K_1\ ,\quad E[\, \| M(\theta) \| \,] \leq K_3\ ,$$

$$\text{and} \quad E[\, \| R_2(\,\hat{\theta}_n\,,\, \hat{\theta}_{n+1} - \hat{\theta}_n\,) \| \,] \leq O\,(\,n^{-2}\,) + O\,(\,n^{-2}\,)\, E[\, \| U_n(Z_n, \hat{\theta}_n) \|^2\,]\ .$$

By Assumption D.1P(1), for all $n \geq n_o$,

$$(a.9) \quad E\,[\,V(\hat{\theta}_{n+1})\,] \leq (1 - n^{-1} c_n)\, E\,[\,V(\hat{\theta}_n)\,] + n^{-1}\, E\,[(\,V'(\hat{\theta}_n)\,,\, P_{k(n+1)}\, U_n(Z_n, \hat{\theta}_n)\,)\,] + O(n^{-2}) + O(n^{-2})\, E[\, \| P_{k(n+1)} U_n(Z_n, \hat{\theta}_n) \|^2\,]\,.$$

By the definition of $\overline{V}$,

$$(a.10) \quad E[\,\overline{V}(\hat{\theta}_n\,,\, n+1)\ |\ \mathcal{F}^{n-1}\,] - \overline{V}(\hat{\theta}_n\,,\, n) = -\,n^{-1}\,(\,V'(\hat{\theta}_n)\,,\, P_{k(n+1)}\, E[\,U_n(Z_n, \hat{\theta}_n)\ |\ \mathcal{F}^{n-1}\,]\,)\ .$$

By the definitions of $S(n)$ and $\overline{DV}_n$,

$$(a.11)\ \ E[\,S(n+1)\mid \mathcal{F}^{n-1}\,] - S(n) \equiv E\,[\,V(\hat{\theta}_{n+1}\mid \mathcal{F}^{n-1}\,] - V(\hat{\theta}_n) + E[\,\bar{V}(\hat{\theta}_n\,,n+1)\mid \mathcal{F}^{n-1}\,] - \bar{V}(\hat{\theta}_n\,,n) + \overline{DV}_n \quad .$$

Substituting (a.9) and the expectation of (a.10) into the expectation of (a.11), we get that for all $n \geq n_o$,

$$E[\,S(n+1)\,] \leq (1-n^{-1}c_n)\,E[\,S(n)\,] + n^{-1}\,c_n\,E[\,\bar{V}(\hat{\theta}_n\,,n)\,] + O\,(n^{-2}) + O(n^{-2})\,E[\,\|P_{k(n+1)}U_n(Z_n,\hat{\theta}_n)\|^2\,] + E\,[\,\overline{DV}_n\,]\ .$$

LEMMA A.5 (Venter, 1966):
Let $\{ s_n \}$ be a sequence of non-negative numbers such that for all n large enough

$$s_{n+1} \leq (1 - n^{-1}\,c_n)\,s_n + d\,n^{-(1+q)}\ ,$$

where $d > 0$ and $c_n \rightarrow c$ as $n \rightarrow \infty$. Then for n sufficiently large,

$s_n = O(n^{-q})$ if $c > q > 0$; $s_n = O(n^{-c}\log n)$ if $c = q > 0$; and $s_n = O(n^{-c})$ if $q > c > 0$.

PROOF OF LEMMA 5.1: Lemma A.4, conditions 5.1(1), 5.1(2) and 5.1(4) imply

$$(a.12)\ \ E[\,S(n+1)\,] \leq (1-n^{-1}c_n)\,E[\,S(n)\,] + O\,(\,\max\,[\,n^{-(1+b)}\,,n^{-2}\,]\,)\ .$$

Now 5.1(3), Lemma A.5 and $(a.12)$ imply

$E[\,S(n)\,] = O(\,n^{-q}\,)$ for $c > q$;

$E[\,S(n)\,] = O(\,n^{-q}\log n\,)$ for $c = q$;

$E[\,S(n)\,] = O(\,n^{-c}\,)$ for $q > c > 0$.

Given the definition of $S(n)$ and condition 5.1(1), we get

$E[\,V(\hat{\theta}_n)\,] = O(\,n^{-q}\,)$ for $c > q$;

$E[\,V(\hat{\theta}_n)\,] = O(\,n^{-q}\log n\,)$ for $c = q$;

$E[\,V(\hat{\theta}_n)\,] = O(\,n^{-c}\,)$ for $q > c > 0$.

PROOF OF COROLLARY 5.2: It suffices to verify that all conditions of Lemma 5.1 are satisfied with $U_n \equiv F_n(Z_n)$. It is obvious that D.2P(2) implies 5.1(4). Since $\hat{\theta}_n$ is $\mathcal{F}^{n-1}$ –measurable, by the definition of $\bar{V}$, we have

$$\bar{V}(\hat{\theta}_n\,,n) = \sum_{j=n}^{\infty} j^{-1}\,(\,V'(\hat{\theta}_n)\,,E[\,P_{k(j+1)}U_j\mid \mathcal{F}^{n-1}\,]\,) = (\,V'(\hat{\theta}_n)\,,E[\,\sum_{j=n}^{\infty} j^{-1}\,P_{k(j+1)}U_j\mid \mathcal{F}^{n-1}\,]\,)\ .$$

By the Cauchy-Schwartz inequality, D.1P(2), D.2P(1) and $(a.8)$, we get

$$E[\,|\,\bar{V}(\hat{\theta}_n, n)\,|\,] \le \| V'(\hat{\theta}_n) \|_2 \, \| E[\sum_{j=n}^{\infty} j^{-1} P_{k(j+1)} U_j \mid \mathcal{F}^{n-1}] \|_2$$
$$\le K_4 \, (1 + E[V(\hat{\theta}_n)])^{1/2} \, O(n^{-b}) \le O(n^{-b}) \; .$$

Hence $E[\bar{V}(\hat{\theta}_n, n)]$ is deÞned and satisÞes 5.1(1). Also, by the deÞnition of $S(n)$, we have that $E[S(n)]$ is deÞned and satisÞes 5.1(3).

Since $\hat{\theta}_{n+1}$ and $\hat{\theta}_n$ are $\mathcal{F}^n$ –measurable, by the deÞnition of $\overline{DV}_n$ we have

$$\overline{DV}_n \equiv E[\, \Sigma_{j \ge n+1} \, j^{-1} \, E[\, (V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n), P_{k(j+1)} U_j) \mid \mathcal{F}^n \,] \mid \mathcal{F}^{n-1} \,]$$
$$= E[\, (V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n), E[\, \Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} U_j \mid \mathcal{F}^n \,]) \mid \mathcal{F}^{n-1} \,] \; .$$

By Taylor expansion,

$$V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n) = \int_{[0,1]} V''(\hat{\theta}_n + s(\hat{\theta}_{n+1} - \hat{\theta}_n))(\hat{\theta}_{n+1} - \hat{\theta}_n)\, ds \; .$$

By $(a.7)$,

$$\overline{DV}_n = n^{-1} E[\, ((\int_{[0,1]} V''(\hat{\theta}_n + s(\hat{\theta}_{n+1} - \hat{\theta}_n))\, ds) \, P_{k(n+1)} (M(\hat{\theta}_n) + U_n),$$
$$\Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} E[U_j \mid \mathcal{F}^n] \,) \mid \mathcal{F}^{n-1} \,] \; .$$

Denote $\overline{DV}_{1n}$ and $\overline{DV}_{2n}$ as

$$\overline{DV}_{1n} = E[\, ((\int_{[0,1]} V''(\hat{\theta}_n + s(\hat{\theta}_{n+1} - \hat{\theta}_n))\, ds) \, (P_{k(n+1)} M(\hat{\theta}_n)),$$
$$\Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} E[U_j \mid \mathcal{F}^n] \,) \mid \mathcal{F}^{n-1} \,] \; .$$

$$\overline{DV}_{2n} = E[\, ((\int_{[0,1]} V''(\hat{\theta}_n + s(\hat{\theta}_{n+1} - \hat{\theta}_n))\, ds) \, (P_{k(n+1)} U_n),$$
$$\Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} E[U_j \mid \mathcal{F}^n] \,) \mid \mathcal{F}^{n-1} \,] \; .$$

D.2P(1) and $(a.8)$ imply that for a Þnite constant K_5

$$E[\,|\, \overline{DV}_{1n} \,|\,] \le K_5 \, \| P_{k(n+1)} M(\hat{\theta}_n) \|_2 \, \| E[\, \Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} U_j \mid \mathcal{F}^n] \|_2 = O(n^{-b}) \; .$$

Further D.2P(1), D.2P(2) and $(a.8)$ imply that for a Þnite constant K_6

$$E[\,|\, \overline{DV}_{2n} \,|\,] \le K_6 \, \| P_{k(n+1)} U_n \|_2 \, \| E[\, \Sigma_{j \ge n+1} \, j^{-1} P_{k(j+1)} U_j \mid \mathcal{F}^n] \|_2 = O(n^{-b}) \, .$$

Because $\overline{DV}_n = n^{-1} [\, \overline{DV}_{1n} + \overline{DV}_{2n} \,]$,

$$E[\, \overline{DV}_n \,] \le E[\,|\, \overline{DV}_n \,|\,] \le n^{-1} (\, E[\,|\, \overline{DV}_{1n} \,|\,] + E[\,|\, \overline{DV}_{2n} \,|\,] \,) \le O(n^{-(1+b)}) \; .$$

This gives 5.1(2). Now the conclusion follows from Lemma 5.1.

PROOF OF THEOREM 5.3: By the conditional Jensen's inequality, with $1 < p \le 2$

$$\| E[P_{k(j+1)}U_j \mid \mathcal{F}^{n-1}] \|^p \le E[\| P_{k(j+1)}U_j \|^p \mid \mathcal{F}^{n-1}] \quad \text{for any } j \ge n \text{ , hence}$$

$$\| E[P_{k(j+1)}U_j \mid \mathcal{F}^{n-1}] \|_p \le \| P_{k(j+1)}U_j \|_p \quad \text{for any } j \ge n \ .$$

Since $\sup_n \| P_{k(n+1)}U_n \|_2 \le \Delta < \infty$, we can pick mixingale parameters $\{ c_n \}$ such that $c_n \le \Delta$. Hence, $\sum_{i=1}^{\infty} (c_i/i)^2 < \infty$. That A.5P' implies A.4P is a simple consequence of Corollaries 3.8 and 3.9 in Chen and White (1996).

By the deÞnition of L_p –mixingales with ψ_m of size -1, we have

$$\sum_{j=n}^{\infty} j^{-1} \| E[P_{k(j+1)}U_j \mid \mathcal{F}^{n-1}] \|_p \le \sum_{j=n}^{\infty} j^{-1}\psi_{j+1-n}c_j$$

$$\le n^{-1}\Delta \sum_{j=n}^{\infty} \psi_{j+1-n} = n^{-1}\Delta \sum_{m=1}^{\infty} \psi_m = O(n^{-1}) \ .$$

Hence A.5P'(1) implies D.2P(3) with $b = 1$, and we get result (i) .

By the deÞnition of L_2 – mixingale with ψ_m of size $-a$, $(1/2 \le a < 1)$, we have from the Hölder inequality that

$$\sum_{j=n}^{\infty} j^{-1} \| E[P_{k(j+1)}U_j \mid \mathcal{F}^{n-1}] \|_2 \le \sum_{j=n}^{\infty} j^{-1}\psi_{j+1-n}c_j$$

$$\le \Delta \Big(\sum_{j=n}^{\infty} \psi^{1/a}{}_{j+1-n} \Big)^a \Big(\sum_{j=n}^{\infty} j^{-1/(1-a)} \Big)^{1-a}$$

$$\le \Delta \Big(\sum_{m=1}^{\infty} \psi^{1/a}{}_m \Big)^a \big(O(n^{[-1/(1-a)]+1}) \big)^{1-a} = O(n^{-a}) \ .$$

Hence A.5P'(2) implies D.2P(3) with $b = a$, and we get result (ii) .

PROOF OF COROLLARY 5.4: It sufÞces to verify that all conditions of Lemma 5.1 are satisÞed .

First we show that D.3P(2) implies 5.1(4). Since $\hat{\theta}_n$ is $\mathcal{F}^{n-1}$ – measurable, and $\sup_n \| \hat{\theta}_n \| \le \bar{B}$ $a.s.-P$,

$$\sup_n E[\| P_{k(n+1)}U_n(Z_n,\hat{\theta}_n) \|^2] = \sup_n E[E[\| P_{k(n+1)}U_n(Z_n,\hat{\theta}_n) \|^2 \mid \mathcal{F}^{n-1}]]$$

$$\le \sup_n E[E[\sup_{\|\theta\| \le \bar{B}} \| P_{k(n+1)}U_n(Z_n,\theta) \|^2 \mid \mathcal{F}^{n-1}]]$$

$$= \sup_n E[\sup_{\|\theta\| \le \bar{B}} \| P_{k(n+1)}U_n(Z_n,\theta) \|^2] < \infty \ .$$

Hence D.3P(2) implies 5.1(4).

The veriÞcation for conditions 5.1(1) and 5.1(3) is similar to the corresponding proof for Corollary 5.2. By D.3P(1), D.1P(2) and $(a.8)$, since $\hat{\theta}_n$ is $\mathcal{F}^{n-1}$ –measurable, we get

$$E[\, | \bar{V}(\hat{\theta}_n, n) | \,] \le \| V'(\hat{\theta}_n) \|_2 \, \| E[\sum_{j=n}^{\infty} j^{-1} P_{k(j+1)}U_j(Z_j,\hat{\theta}_n) \mid \mathcal{F}^{n-1}] \|_2$$

$$\leq [\, 1 + E\, V(\hat{\theta}_n)\,]^{1/2}\, O(n^{-b}) \leq O(n^{-b}) \quad .$$

Hence $E[\,\overline{V}(\hat{\theta}_n\,,n\,)\,]$, $E[\,\underline{S(n)}\,]$, and $E[\,\overline{\underline{DV_n}}\,]$ are defined, and 5.1(1) and 5.1(3) are also satisfied. By the definition of DV_n, we have $DV_n \equiv E\,[\,D_{1n} + D_{2n} \mid \mathcal{F}^{n-1}\,]$, where

$$D_{1n} \equiv \Sigma_{j \geq n+1}\, j^{-1}\, E[\,(\,V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n)\,,\, P_{k(j+1)} U_j(Z_j, \hat{\theta}_{n+1})\,) \mid \mathcal{F}^n\,]\,,$$

$$D_{2n} \equiv \Sigma_{j \geq n+1}\, j^{-1}\, E[\,(\,V'(\hat{\theta}_n)\,,\, P_{k(j+1)}[\,U_j(Z_j, \hat{\theta}_{n+1}) - U_j(Z_j, \hat{\theta}_n)\,]\,) \mid \mathcal{F}^n\,]\ .$$

(Comparing this with the proof of Corollary 5.2, here $E[\,D_{1n} \mid \mathcal{F}^{n-1}\,]$ corresponds to the entire expression for DV_n there, while there is no D_{2n} term in the θ *–independent* error case.)
First we bound D_{2n} as

$$|D_{2n}| \leq \| V'(\hat{\theta}_n) \| \times \Sigma_{j \geq n+1}\, j^{-1} \| E[\, P_{k(j+1)} U_j(Z_j, \hat{\theta}_{n+1}) - P_{k(j+1)} U_j(Z_j, \hat{\theta}_n) \mid \mathcal{F}^n\,] \| \quad a.s.\!-\!P\ .$$

Because $\hat{\theta}_{n+1}$ and $\hat{\theta}_n$ are $\mathcal{F}^n$ –measurable, $(a.8)$ and condition D.3P(3) imply

$$|D_{2n}| \leq O(1) \| \hat{\theta}_{n+1} - \hat{\theta}_n \| \Sigma_{j \geq n+1}\, j^{-1}\, E[\, h_{\overline{B},j}(Z_j) \mid \mathcal{F}^n\,] \leq O\,(\,n^{-b}\,) \| \hat{\theta}_{n+1} - \hat{\theta}_n \| \quad a.s.\!-\!P\ .$$

Hence, $E[\,|D_{2n}|\,] \leq O(n^{-b})\, E[\,\| \hat{\theta}_{n+1} - \hat{\theta}_n \|\,]$. Now $(a.7)$ and the triangle inequality imply

$$\| \hat{\theta}_{n+1} - \hat{\theta}_n \|_2 \leq n^{-1} \| P_{k(n+1)} M(\hat{\theta}_n) \|_2 + n^{-1} \| P_{k(n+1)} U_n(Z_n, \hat{\theta}_n) \|_2\ .$$

Next, $(a.8)$ implies $\| M(\hat{\theta}_n) \|_2 \leq O(1)$, and we have shown that D.3P(4) and $(a.8)$ imply 5.1(4). Hence

$$(a.13) \quad E[\,\| \hat{\theta}_{n+1} - \hat{\theta}_n \|\,] \leq \| \hat{\theta}_{n+1} - \hat{\theta}_n \|_2 \leq O(n^{-1})\,,$$

so that $E[\,|D_{2n}|\,] \leq O(\,n^{-(1+b)}\,)$ for $n \geq n_o$.

Now we bound D_{1n}. (The proof is akin to that for Corollary 5.2; hence we just record the main steps).
Since $\hat{\theta}_{n+1}$ and $\hat{\theta}_n$ are $\mathcal{F}^n$ – measurable,

$$E[\,|D_{1n}|\,] \leq \| V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n) \|_2 \| E[\,\Sigma_{j \geq n+1}\, j^{-1}\, P_{k(j+1)} U_j(Z_j, \hat{\theta}_{n+1}) \mid \mathcal{F}^n\,] \|_2\ .$$

By D.3P(1),

$$(a.14) \quad E[\,|D_{1n}|\,] \leq \| V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n) \|_2\, O(\,n^{-b}\,)\,.$$

By Taylor expansion,

$$V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n) = \int_{[0,1]} V''(\hat{\theta}_n + s\,(\hat{\theta}_{n+1} - \hat{\theta}_n))\,(\hat{\theta}_{n+1} - \hat{\theta}_n)\, ds\ .$$

Then $(a.8)$ and $(a.13)$ imply

$$\| V'(\hat{\theta}_{n+1}) - V'(\hat{\theta}_n) \|_2 = O\,(\,n^{-1}\,) \qquad \text{for } n \geq n_o\ .$$

Substituting this into ($a.$ 14), we get $E[\,|D_{1n}|\,] \leq O(n^{-(1+b)})$ for $n \geq n_o$ Hence,

$$|E[\overline{DV}_n]| \leq |E[D_{1n}]| + |E[D_{2n}]| \leq E[\,|D_{1n}|\,] + E[\,|D_{2n}|\,] = O(n^{-(1+b)}) .$$

This gives 5.1(2). The conclusion now follows from Lemma 5.1.

PROOF OF COROLLARY 5.7: Notice that 5.7(1) implies D.1P with $V(\theta) = \|\theta - \theta_o\|^2$. Then all conditions of Corollary 5.4 are satisfied with $b = 1$, and the result follows.

PROOF OF COROLLARY 5.8: Notice that 5.8(1) and 5.8(2) imply A.8, so we get $\|\hat{\theta}_n - \theta_o\| \to 0 \;\; a.s.-P$. Following the proof of Lemma 5.1 and Corollary 5.4 with $V_n(\theta) = \|\theta - P_{k(n)}\theta_o\|^2$, we get for n sufficiently large,

$$E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(1/n) \quad \text{for } c > 1 ; \quad E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(n^{-1}\log n) \quad \text{for } c = 1 ;$$

$$\text{and} \qquad E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] = O(n^{-c}) \quad \text{for } 0 < c < 1 .$$

Thus result (i) holds. Now result (ii) follows due to 5.8(3) and the following relation :

$$E[\|\hat{\theta}_n - \theta_o\|^2] = E[\|\hat{\theta}_n - P_{k(n)}\theta_o\|^2] + \|\theta_o - P_{k(n)}\theta_o\|^2 .$$

ACKNOWLEDGEMENTS

We thank the editor, an associate editor, two referees, Max Stinchcombe, Elias Masry, and Kurt Hornik for helpful comments. Errors are the authors' responsibility. This work was supported by NSF Grants SES-8921382, SES-9209023, and SES-0111238.

REFERENCES

1. Berger, E. (1986): "Asymptotic Behaviour of a Class of Stochastic Approximation Procedures," *Probability Theory and Related Fields*, 71, 517-552.

2. Berman, N. and A. Shwartz (1989): "Abstract Stochastic Approximations and Applications," *Stochastic Processes and their Applications*, 31, 133-149.

3. Chen, X. and H. White (1992): "Central Limit and Functional Central Limit Theorems for Hilbert Space-Valued Dependent Processes." UCSD Department of Economics Discussion Paper 92-35.

4. Chen, X. and H. White (1996): "Laws of Large Numbers for Hilbert Space - Valued Mixingales with Applications," *Econometric Theory*, 12, 284-304.

5. Chen, X. and H. White (1998a): "Central Limit and Functional Central Limit Theorems for Hilbert Space - Valued Dependent Heterogeneous Arrays with Applications," *Econometric Theory*, 14, 260-284.

6. Chen, X. and H. White (1998b): "Nonparametric Adaptive Learning With Feedback," *Journal of Economic Theory*, 82, 190-222.

7. Crawford, V. (1995): "Adaptive Dynamics in Coordination Games," *Econometrica*, 63, 103-144.

8. Dippon, J. (1991): "Asymptotic ConÞdence Regions of Stochastic Approximation Procedures in Hilbert Spaces," *Journal of Theoretical Probability*, 4, 37-52.

9. Einmahl, U. (1991): "On the Almost Sure Behavior of Sums of i.i.d. Random Variables in Hilbert Space," *Annals of Probability*, 19, 1227-1263.

10. Evans, G. and S. Honkapohja (1995): "Local Convergence of Recursive Learning to Steady States and Cycles in Stochastic Nonlinear Models," *Econometrica*, 63, 195-206.

11. Fabian, V. (1968): "On Asymptotic Normality in Stochastic Approximation," *Annals of Mathematical Statistics*, 39, 1327-1332.

12. Goldstein, L. (1988): "Minimizing Noisy Functionals in Hilbert Space: An Extension of the Kiefer-Wolfowitz Procedure," *Journal of Theoretical Probability*, 1, 189-204.

13. GyorÞ, L. and E. Masry (1990): "The L_1 and L_2 Strong Consistency of Recursive Kernel Density Estimation from Dependent Samples," *IEEE Transactions on Information Theory*, 36, 531-539.

14. Heunis, A. (1992): "Rate of Convergence for an Adaptive Filtering Algorithm Driven by Stationary Dependent Data," University of Waterloo Department of Electrical Engineering Discussion Paper.

15 Kiefer, J. and J. Wolfowitz (1952): "Stochastic Estimation of the Maximum of a Regression Function," *Annals of Mathematical Statistics*, 23, 462-466.

16. Kuan, C.M. and H. White (1993): "ArtiÞcial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1-91.

17. Kuan, C.M. and H. White (1994): "Adaptive Learning With Nonlinear Dynamics Driven by Dependent Processes," *Econometrica*, 62, 1087-1114.

18. Kushner, H.J. (1984): *Approximation and Weak Convergence Methods for Random Processes*. Cambridge: MIT Press.

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

20. Marcet, A. and T.J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48, 337-368.

21. Metivier, M. and P. Priouret (1984): "Applications of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms," *IEEE Trans. Information Theory,* 30, 140-151.

22. Nixdorf, R. (1984): "An Invariance Principle for a Finite Dimensional Stochastic Approximation Method in a Hilbert Space," *Journal of Multivariate Analysis*, 15, 252-260.

23. Pakes, A. and P. McGuire (2001): "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality," *Econometrica*, 69, 1261-1282.

24. Patilea, V. and E. Renault (2001): "Iterative and Recursive Estimation of Structural Non-Adaptive Models," University of Montreal Working Paper.

25. Philipp, W. (1986): "A Note on the Almost Sure Approximation of Weakly Dependent Random Variables," *Mh. Math. 102, 227-236.*

26. Robbins, H. and S. Monro (1951): "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, 22, 400-407.

27. Sargent, T. (1993): *Bounded Rationality in Macroeconomics*, Clarendon Press, Oxford.

28. Serßing, R. (1980): *Approximation Theorems of Mathematical Statistics*, New York: Springer-Verlag.

29. Spear, S.E. (1989): "Learning Rational Expectations Under Computability Constraints," *Econometrica*, 57, 889-910.

30. Venter, J.H. (1966): "On Dvoretzky Stochastic Approximation Theorems," *Annals of Mathematical Statistics*, 37, 1534-1544.

31. Walk, H. (1977): "An Invariance Principle for the Robbins-Monro Process in a Hilbert Space," *Zeitschrift für Warscheinlichkeitstheorie und Verwandte Gebiete*, 39, 135-150.

32. Walk, H. (1987): "Limit Behaviour of Stochastic Approximation Processes," *Statistics and Decisions*, 6, 109-128.

33. Woodford, M. (1990): "Learning to Believe in Sunspots," *Econometrica*, 58, 277-307.

34. Yin, G. and Y.M. Zhu (1990): "On H-Valued Robbins-Monro Processes," *Journal of Multivariate Analysis*, 34, 116-140.

35. Yin, G. (1992): "On H-Valued Stochastic Approximation: Finite Dimensional Projections," *Stochastic Analysis and Applications*, 10, 363-377.