A Kiefer–Wolfowitz Algorithm with Randomized Differences

H. F. Chen, *Fellow, IEEE*, T. E. Duncan, *Fellow, IEEE,* and B. Pasik-Duncan, *Senior Member, IEEE*

***Abstract*— A Kiefer–Wolfowitz or simultaneous perturbation algorithm that uses either one-sided or two-sided randomized differences and truncations at randomly varying bounds is given in this paper. At each iteration of the algorithm only two observations are required in contrast to 2ℓ observations, where ℓ is the dimension, in the classical algorithm. The algorithm given here is shown to be convergent under only some mild conditions. A rate of convergence and an asymptotic normality of the algorithm are also established.**

***Index Terms*—Kiefer–Wolfowitz algorithm, perturbation algorithm, simultaneous stochastic approximation, stochastic approximation with randomized differences.**

I. Introduction

A Kiefer–Wolfowitz (KW) algorithm [8] is used to find the extrema of an unknown function L: $\mathbb{R}^p \to \mathbb{R}$ which may be observed with some additive noise. If the gradient of L can be observed, then the problem can be solved by a Robbins–Monro (RM) algorithm.

Let x_n be the estimate of the unique extremum of L at the nth iteration. One approach to a KW algorithm is to observe L at the following values:

$$x_n^{i+} = [x_n^1, \cdots, x_n^{i-1}, x_n^i + c_n, x_n^{i+1}, \cdots, x_n^\ell]^T$$
$$x_n^{i-} = [x_n^1, \cdots, x_n^{i-1}, x_n^i - c_n, x_n^{i+1}, \cdots, x_n^\ell]^T$$

for $i = 1, 2, \cdots, \ell$ where $c_n \in \mathbb{R} \setminus \{0\}$.

Consider noisy observations of L so that the processes $(y_{n+1}^{i+}, n \in \mathbb{N}, i \in \{1, \cdots, \ell\})$ and $(y_{n+1}^{i-}, n \in \mathbb{N}, i \in \{1, \cdots, \ell\})$ satisfy

$$y_{n+1}^{i+} = L(x_n^{i+}) + \xi_{n+1}^{i+}$$

and

$$y_{n+1}^{i-} = L(x_n^{i-}) + \xi_{n+1}^{i-}$$

where $(\xi_{n+1}^{i+}, n \in \mathbb{N})$ and $(\xi_{n+1}^{i-}, n \in \mathbb{N})$ are observation noise processes. The ratio

$$\frac{y_{n+1}^{i+} - y_{n+1}^{i-}}{2c_n} \tag{1}$$

can be used as an estimate for the ith component of ∇L. Using these estimates a KW algorithm requires 2ℓ measurements of L. If ℓ is large, for example, the optimization of weights in a neural network, then this KW algorithm can be rather slow.

To reduce the number of measurements that are required, a random difference method can be used. There is a fairly long history of random search or approximation ideas in stochastic approximation (e.g., Fabian [7]). Koronacki [10] introduced a random version of the KW algorithm using a sequence of random unit vectors that are independent and uniformly distributed on the unit sphere or unit cube, and he gave sufficient conditions for the convergence of the algorithm. Koronacki [10], [11] noted that the random direction methods have a better reduction of bias effects caused by the use of finite differences for the derivatives than the nonrandom direction methods and that these methods reduce the required number of observations or measurements. Spall [14] reintroduced a random direction version of the KW algorithm and called it a simultaneous perturbation stochastic approximation (SPSA) algorithm. Using the ordinary differential equation (ODE) method [12] Spall showed the convergence and the asymptotic normality of this modified KW or simultaneous perturbation algorithm though the conditions that he required are restrictive.

Initially Spall's KW or simultaneous perturbation algorithm and the conditions that he uses are described. Let $(\Delta_k^i, i = 1, \cdots, \ell, k = 1, 2, \cdots)$ be a sequence of mutually independent and identically distributed random variables with zero mean. Let Δ_k be given by

$$\Delta_k = [\Delta_k^1, \cdots, \Delta_k^\ell]^T. \tag{2}$$

At each iteration, two measurements are taken:

$$y_{k+1}^+ = L(x_k + c_k\Delta_k) + \xi_{k+1}^+ \tag{3}$$
$$y_{k+1}^- = L(x_k - c_k\Delta_k) + \xi_{k+1}^-. \tag{4}$$

Then the vector symmetric difference

$$\frac{(y_{k+1}^+ - y_{k+1}^-)g_k}{2c_k} \tag{5}$$

where

$$g_k = \left[\frac{1}{\Delta_k^1}, \cdots, \frac{1}{\Delta_k^\ell}\right]^T \tag{6}$$

is used as an estimate for $\nabla L(x_k)$.

The KW or simultaneous perturbation algorithm is formed as follows:

$$x_{k+1} = x_k + a_k \frac{(y_{k+1}^+ - y_{k+1}^-)g_k}{2c_k}. \tag{7}$$

In this form the algorithm seeks the maximum of L. The minimum of L is found by replacing a_k by $-a_k$.

Manuscript received June 2, 1995; revised August 10, 1997 and May 28, 1998. Recommended by Associate Editor, V. Solo. This work was supported in part by the NSF under Grants DMS 9305936 and DMS 9623439 and the National Natural Science Foundation of China.

H. F. Chen is with the Institute of Systems Science, Beijing 100080, China.

T. E. Duncan and B. Pasik-Duncan are with the Department of Mathematics, University of Kansas, Lawrence, KS 66045 USA (e-mail: duncan@math.ukans.edu).

Publisher Item Identifier S 0018-9286(99)02826-3.

For the convergence of the algorithm (5) Spall [14] required the following conditions.

A1) The random variables $(\xi_{k+1}^{+} - \xi_{k+1}^{-},\ k \in \mathbb{N})$ is a martingale difference sequence (mds) with uniformly bounded second moments.

A2) The integrability condition

$$\sup_{k\in\mathbb{N}} E(L^2(x_k + c_k\Delta_k)) < \infty$$

is satisfied.

A3) The sequence $(x_k,\ k \in \mathbb{N})$ is assumed *a priori* to be uniformly bounded, that is,

$$\sup_{k\in\mathbb{N}} \|x_k\| < \eta < \infty \quad \text{a.s.}$$

where $\eta \in \mathbb{R}_+$.

A4) The third derivative of L is bounded.

A5) The point $x^0 \in \mathbb{R}^\ell$ is an asymptotically stable point for the differential equation $dx/dt = -f(x(t))$ where $f = \nabla L$.

A6) The sequence $(x_k,\ k \in \mathbb{N})$ is infinitely often in a compact set that is contained in the domain of attraction of x^0 given in A5).

A7) The sequences $(a_k,\ k \in \mathbb{N})$ and $(c_k,\ k \in \mathbb{N})$ satisfy $a_k > 0$, $c_k > 0$ for all $k \in \mathbb{N}$, $a_k \to 0$ and $c_k \to 0$ as $k \to \infty$, $\sum_{k=1}^{\infty} a_k = \infty$ and $\sum_{k=1}^{\infty} (a_k/c_k)^2 < \infty$.

Furthermore, some conditions are imposed on the sequence $(\Delta_k,\ k \in \mathbb{N})$ in [14], but this sequence can be arbitrarily chosen by the user of the algorithm so these conditions should not be considered as restrictions.

In this paper, both a one-sided randomized difference and a two-sided randomized difference are used, that is,

$$\frac{(y_{k+1}^{+} - y_{k+1}^{0})g_k}{c_k} \tag{8}$$

$$\frac{(y_{k+1}^{+} - y_{k+1}^{-})g_k}{2c_k} \tag{8$'$}$$

and

$$g_k = \left[\frac{1}{\Delta_k^1}, \cdots, \frac{1}{\Delta_k^\ell}\right]^T$$

are used to estimate $f(x_k) = \nabla L(x_k)$ where

$$y_{k+1}^0 = L(x_k) + \xi_{k+1}^0. \tag{9}$$

y_{k+1}^{+}, y_{k+1}^{-}, and g_k are given by (3), (4), and (6), respectively.

The two-sided algorithm is described by (7) and the one-sided algorithm is described by (7) with (8′) replaced by (8). By a modification of the algorithm (7) and the use of a direct approach to verifying convergence, the conditions A2), A3), A5), and A6) are eliminated, A4) and A7) are relaxed, and A1) is weakened to one that provides not only a sufficient but also a necessary condition for convergence. This result is given in Theorem 1. In Theorem 2 the observation noise is modified to one with only bounded second moments that is independent of $(\Delta_k,\ k \in \mathbb{N})$. A convergence rate and an asymptotic normality of the algorithm are given in Theorems 3 and 4, respectively.

II. The Algorithm and its Convergence

Initially the algorithm is precisely described. Let $(\Delta_k^i,\ i = 1, \cdots, \ell, k \in \mathbb{N})$ be a sequence of independent and identically distributed random variables where $|\Delta_k^i| < a$, $|1/\Delta_k^i| < b$, $E(1/\Delta_k^i) = 0$ for all $i \in \{1, \cdots, \ell\}$ and $k \in \mathbb{N}$ and some a, $b \in \mathbb{R}_+$. Furthermore, let Δ_k be independent of $\mathcal{F}_k^\xi = \sigma(\xi_i^+,\ \xi_i^0,\ \xi_i^-,\ i \in \{0, \cdots, k\})$ for $k \in \mathbb{N}$. Define y_{k+1} and ξ_{k+1} by the following equations:

$$y_{k+1} = \frac{(y_{k+1}^{+} - y_{k+1}^{0})g_k}{c_k} \tag{10}$$

$$y_{k+1} = \frac{(y_{k+1}^{+} - y_{k+1}^{-})g_k}{2c_k} \tag{10$'$}$$

$$\xi_{k+1} = \xi_{k+1}^{+} - \xi_{k+1}^{0} \tag{11}$$

$$\xi_{k+1} = \xi_{k+1}^{+} - \xi_{k+1}^{-}. \tag{11$'$}$$

It follows that

$$y_{k+1} = \frac{(L(x_k + c_k\Delta_k) - L(x_k))g_k}{c_k} + \frac{\xi_{k+1}g_k}{c_k} \tag{12}$$

$$y_{k+1} = \frac{(L(x_k + c_k\Delta_k) - L(x_k - c_k\Delta_k))g_k}{2c_k} + \frac{\xi_{k+1}g_k}{2c_k}. \tag{12$'$}$$

Equations (10), (10′), (11), (11′), (12), and (12′) represent some abuse of notation. Equations (10)–(12) are used for the one-sided algorithm and (10′), (11′), and (12′) are used for the two-sided algorithm. In the subsequent description the meaning of y_{k+1} is determined by whether the one-sided or the two-sided algorithm is used. Choose $x^* \in \mathbb{R}^\ell$ and fix it. Define the following KW algorithm with randomly varying truncations and randomized differences:

$$\begin{aligned} x_{k+1} &= (x_k + a_k y_{k+1}) 1_{\{\|x_k + a_k y_{k+1}\| \le M_{\sigma_k}\}} \\ &\quad + x^* 1_{\{\|x_k + a_k y_{k+1}\| > M_{\sigma_k}\}} \end{aligned} \tag{13}$$

$$\sigma_k = \sum_{i=0}^{k-1} 1_{\{\|x_i + a_i y_{i+1}\| > M_{\sigma_i}\}} \tag{14}$$

$$\sigma_0 \equiv 0$$

where $(M_k,\ k \in \mathbb{N})$ is a sequence of strictly positive, strictly increasing real numbers that diverges to $+\infty$. It is clear that σ_k is the number of truncations that have occurred before time k. Clearly the random vector x_k is measurable with respect to $\mathcal{F}_k := \mathcal{F}_k^\xi \vee \mathcal{F}_{k-1}^\Delta$ where $\mathcal{F}_k^\Delta = \sigma(\Delta_i,\ i \in \{0, \cdots, k\})$. Thus the random vector Δ_k is independent of $\sigma(x_i,\ i \le k)$.

The following conditions are imposed on the algorithm.

H1) The function $\nabla L = f$ is locally Lipschitz continuous. There is a unique maximum of L at x^0 that is the only local extremum so that $f(x^0) = 0$ and $f(x) \ne 0$ for $x \ne x^0$. There is a $c_0 \in \mathbb{R}_+$ such that $\|x^*\| < c_0$ and $\sup_{\|x\|=c_0} L(x) < L(x^*)$.

H2) The two sequences of strictly positive real numbers $(a_k,\ k \in \mathbb{N})$ and $(c_k,\ k \in \mathbb{N})$ satisfy $a_k \to 0$ and $c_k \to 0$ as $k \to \infty$, $\sum_{k=1}^{\infty} a_k = \infty$ and there is a $p \in (1, 2]$ such that $\sum_{k=1}^{\infty} a_k^p < \infty$.

Remark: If L is twice continuously differentiable then f is locally Lipschitz continuous. If in H1) x^0 is the unique minimum of L, then in (13) and (14) a_k should be replaced by $-a_k$.

The main result of this section is the following theorem that gives necessary and sufficient conditions for the convergence of the algorithm (13).

Theorem 1: Let H1) and H2) be satisfied and $(x_k,\, k \in \mathbb{N})$ be given by (13) with the conditions for $(\Delta_k^i,\, i = 1, \cdots, \ell;\, k \in \mathbb{N})$ given for the algorithm. The sequence $(x_k,\, k \in \mathbb{N})$ satisfies

$$\lim_{k\to\infty} x_k = x^0 \quad \text{a.s.} \tag{15}$$

where x^0 is given in H1) if and only if for each $k \in \mathbb{N}$ the observation noise ξ_k in (11) or (11′) can be decomposed into the sum of two terms for each $j \in \{1, \cdots, \ell\}$ as

$$\xi_k = e_k^j + \nu_k^j \tag{16}$$

such that

$$\sum_{k=1}^{\infty} \frac{a_k e_{k+1}^j}{c_k \Delta_k^j} < \infty \quad \text{a.s.} \tag{17}$$

and

$$\lim_{k\to\infty} \frac{\nu_{k+1}^j}{c_k \Delta_k^j} = 0 \quad \text{a.s.} \tag{18}$$

for $j = 1, \cdots, \ell$ where Δ_k^j is given in (2).

Before proving the theorem, two lemmas are proved.

To describe easily the replacement of some components of Δ_k by zero, two functions are introduced. Let $\Delta_k(\cdot)$ $\{0, \cdots, \ell\} \to \mathbb{R}^l$ and $\Delta_k^c(\cdot)$ $\{0, \cdots, \ell\} \to \mathbb{R}^\ell$ be given by

$$\Delta_k(s) = [\Delta_k^1, \cdots, \Delta_k^s, 0, \cdots, 0]^T, \qquad \Delta_k(0) = 0 \tag{19}$$

$$\Delta_k^c(s) = [0, \cdots, 0, \Delta_k^{s+1}, \cdots, \Delta_k^\ell]^T, \qquad \Delta_k^c(\ell) = 0. \tag{20}$$

It is clear that

$$\Delta_k = \Delta_k^c(0) = \Delta_k(\ell) \tag{21}$$

and

$$\Delta_k(i-1) + \Delta_k^c(i) = [\Delta_k^1, \cdots, \Delta_k^{i-1}, 0, \Delta_k^{i+1}, \cdots, \Delta_k^\ell]^T. \tag{22}$$

For notational convenience let $\delta_k(i)$ denote a generic $\mathbb{R}^\ell$-valued random vector such that

$$\delta_k(i) = [\underbrace{0, \cdots, 0}_{i-1}, \delta_k^i, \underbrace{0, \cdots, 0}_{\ell-i}]^T \tag{23}$$

where $|\delta_k(i)| \le |c_k a|$, $a \in \mathbb{R}^\ell$ is fixed, and δ_k^i may vary for different applications.

Lemma 1: If H1) is satisfied, then the observation y_{k+1} given by (12) or (12′) can be expressed as

$$y_{k+1} = f(x_k) + \varepsilon_{k+1} \tag{24}$$

where

$$\varepsilon_{k+1} = w_k(1) + w_k(2) + \frac{1}{c_k}\, \xi_{k+1} g_k \tag{25}$$

for y_{k+1} given by (12) and

$$\varepsilon_{k+1} = \frac{1}{2}\left(w_k(1) + w_k(2) + u_k(1) + u_k(2)\, \frac{1}{c_k}\, \xi_{k+1} g_k \right) \tag{26}$$

for y_{k+1} given by (12′) and the ith components of $w_k(1)$ and $u_k(1)$ are, respectively

$$w_{k,i}^i = f_i(x_k + c_k\Delta_k + \delta_k(i)) - f_i(x_k) \tag{27}$$

$$u_{k,i}^i = f_i(x_k + c_k\Delta_k + \delta_i(i)) - f_i(x_k) \tag{28}$$

$$f_i(x_k) = \frac{\partial L(x_k)}{\partial x_k^i}$$

and the ith components of $w_k(2)$ and $u_k(2)$ are $\sum_{j\neq i}^{\ell} h_{kj}(1/\Delta_k^i)$ and $\sum_{j\neq i}^{\ell} g_{kj}(1/\Delta_k^i)$, respectively, and

$$h_{kj} = \begin{cases} f_j(x_k + c_k\Delta_k(j-1) + \delta_k(j))\Delta_k^j \\ \quad j \in \{1, \cdots, i-1\} \\ f_j(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(j)) + \delta_k(j))\Delta_k^j \\ \quad j \in \{i+1, \cdots, \ell\} \end{cases} \tag{29}$$

$$g_{kj} = \begin{cases} f_j(x_k - c_k\Delta_k(j-1) + \delta_k(j))\Delta_k^j \\ \quad j \in \{1, \cdots, i-1\} \\ f_j(x_k - c_k(\Delta_k(i-1) + \Delta_k^c(j)) + \delta_k(j))\Delta_k^j \\ \quad j \in \{i+1, \cdots, \ell\} \end{cases} \tag{30}$$

and h_{kj} and g_{kj} are independent of Δ_k^i.

Proof: By comparing (12′) and (24) it follows that the ith component of ϵ_{k+1} is

$$\frac{L(x_k + c_k\Delta_k) - L(x_k - c_k\Delta_k)}{2c_k\Delta_k^i} - f_i(x_k) + \frac{1}{2c_k\Delta_k^i}\, \xi_{k+1}.$$

Therefore, to obtain (26) it suffices to show that

$$\frac{L(x_k + c_k\Delta_k) - L(x_k - c_k\Delta_k)}{2c_k\Delta_k^i} - f_i(x_k) = \frac{1}{2}\left[w_{ki}^i + u_{ki}^i + \sum_{j\neq i}^{l} (w_{ki}^j + u_{ki}^j) \right] \tag{31}$$

where

$$w_{ki}^j = h_{kj}\, \frac{1}{\Delta_k^i} \tag{32}$$

and

$$u_{ki}^j = g_{kj}\, \frac{1}{\Delta_k^i}. \tag{33}$$

It follows that

$$\begin{aligned} &\frac{L(x_k + c_k\Delta_k) - L(x_k - c_k\Delta_k)}{2c_k\Delta_k^i} - f_i(x_k) \\ &\quad = \frac{L(x_k + c_k\Delta_k) - L(x_k)}{2c_k\Delta_k^i} + \frac{L(x_k) - L(x_k - c_k\Delta_k)}{2c_k\Delta_k^i} \\ &\qquad - f_i(x_k) \\ &\quad = \frac{1}{2c_k\Delta_k^i} \Bigg\{ \sum_{j=1}^{i-1} [L(x_k + c_k\Delta_k(j)) \\ &\qquad - L(x_k + c_k\Delta_k(j-1))] + L(x_k + c_k\Delta_k) \\ &\qquad - L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(i))) \\ &\qquad + \sum_{j=i+1}^{l} [L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(j-1))) \\ &\qquad - L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(j)))] \\ &\qquad - L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(j)))] \Bigg\} \end{aligned}$$

$$+\frac{1}{2c_k\Delta_k^i}\Bigg\{\sum_{j=1}^{i-1}[L(x_k-c_k\Delta_k(j-1)) - L(x_k-c_k\Delta_k(j))] + \sum_{j=i+1}^{l}[L(x_k-c_k(\Delta_k(i-1)+\Delta_k^c(j))) - L(x_k-c_k(\Delta_k(i-1)+\Delta_k^c(j-1)))] + L(x_k-c_k(\Delta_k(i-1)+\Delta_k^c(c))) - L(x_k-c_k\Delta_k)\Bigg\} - f_i(x_k). \quad (34)$$

It is clear by the Taylor formula that

$$\frac{1}{2}\Bigg[\frac{L(x_k+c_k\Delta_k)-L(x_k+c_k(\Delta_k(i-1)+\Delta_k^c(i)))}{c_k\Delta_k} + \frac{L(x_k-c_k(\Delta_k(i-1)+\Delta_k^c(i)))-L(x_k-c_k\Delta_k)}{c_k\Delta_k^i} - 2f_i(x_k)\Bigg] = w_{ki}^i + u_{ki}^i \quad (35)$$

and for $j \le i-1$, $L(x_k+c_k\Delta(j)) - L(x_k+c_k\Delta_k(j-1)) = f_j(x_k+c_k\Delta(j-1)+\Delta_k(j))\Delta_k^j = h_{kj}$, which is independent of Δ_k^i.

The other expressions in (29) and (30) are obtained by a similar argument.

For y_{k+1} given by (12), (31) changes to

$$\frac{L(x_k+c_k\Delta_k)-L(x_k)}{c_k\Delta_k^i} - f_i(x_k) = w_{ki}^i + \sum_{j\ne i}^{l} w_{ki}^j \quad (31')$$

and the analysis is the same except that the u_{ki}^j terms are deleted. □

Lemma 2: If H1), H2), (17), and (18) are satisfied, then there is $\Omega' \subset \Omega$ with $P(\Omega')=1$ such that for any fixed $\omega \in \Omega'$ and any convergent subsequence $\{x_{n_k}(\omega),\, k \in \mathbb{N}\}$ there exist positive constants M, T, and k_T such that if $k > k_T$, then

$$\left\|\sum_{i=n_k}^{m+1} a_i y_{i+1}(\omega)\right\| \le M \quad (36)$$

for each $m \in \{n_k, \cdots, m(n_k, T)\}$, where

$$m(n,t) = \max\left\{m\colon \sum_{i=n}^{m} a_i \le t\right\}. \quad (37)$$

Proof: For notational convenience, the evaluation of the random variables at ω is suppressed. For example, $x_{n_k}(\omega)$ is simply denoted as x_{n_k}.

Since h_{kj} and g_{kj} are independent of Δ_k^i and $\sum_{i=1}^{\infty} a_i^p < \infty$ for some $p \in (1, 2]$ it follows by the convergence theorem for mds's [3], [6] that for each $N \in \mathbb{N}$ and $j \ne i$

$$\sum_{k=1}^{\infty} a_k w_{ki}^j 1_{\{|h_{kj}|\le N\}} < \infty, \quad \text{a.s.} \quad (38)$$

$$\sum_{k=1}^{\infty} a_k u_{ki}^j 1_{\{|g_{kj}|\le N\}} < \infty, \quad \text{a.s.} \quad (39)$$

Let $\Omega' \subset \Omega$ be the ω-set such that for each $\omega \in \Omega'$ (17) is finite, (18) is satisfied, and (38) and (39) are finite for all $N \in \mathbb{N}$. It is clear that $P(\Omega')=1$.

If the number of truncations is finite then there is an N such that $\sigma_n = \sigma_N$ for all $n \ge N$ and $\|\sum_{i=n_k}^{m+1} a_i y_{i+1}\| = \|x_{m+2}-x_{n_k}\| \le 2M_{\sigma_N}$ for all $n_k \ge N$. Thus the inequality (36) is satisfied.

Now let $\sigma_n \to \infty$ and assume that (36) is not satisfied. Choose $c > \|\overline{x}\|$ and fix it where $\overline{x}$ is the limit of the subsequence, $(x_{n_k},\, k \in \mathbb{N})$. By the convergence of $(x_{n_k},\, k \in \mathbb{N})$ there is a k_c such that

$$\|x_{n_k}\| < \frac{c+\|\overline{x}\|}{2}$$

for all $k \ge k_c$. Choose a sequence $(T_s,\, s \in \mathbb{N})$ of strictly positive numbers that converges to zero. Since it is assumed that (36) is not satisfied for each T_s, there are a $k_s > s$ and an $m_s \in \{n_{k_s}, \cdots, m(n_{k_s}, T_s)\}$ such that

$$\left\|\sum_{i=n_{k_s}}^{m_s+1} a_i y_{i+1}\right\| > \frac{c-\|\overline{x}\|}{2}.$$

It can be assumed that $k_{s+1} > k_s \ge k_c$ for all $s \in \mathbb{N}$ and let

$$m_s = \inf\left\{m\colon \left\|\sum_{i=n_{k_s}}^{m+1} a_i y_{i+1}\right\| > \frac{c-\|\overline{x}\|}{2}\right\}. \quad (40)$$

Since the sequence $(x_{n_k},\, k \in \mathbb{N})$ is convergent, it follows by the Lipschitz continuity of f that

$$\lim_{k\to\infty} w_{n_k,i}^i = 0$$
$$\lim_{k\to\infty} u_{n_k,i}^i = 0$$

for $i \in \{1, \cdots, l\}$ where $w_{k,i}^i$ and u_{ki}^i satisfy (27) and (28), respectively. It is elementary to verify that

$$\lim_{k\to\infty} a_{n_k} w_{n_k,i}^j = 0$$
$$\lim_{k\to\infty} a_{n_k} u_{n_k i}^j = 0$$

for $i \ne j$. Thus

$$\lim_{s\to\infty} a_{n_{k_s}} y_{n_{k_s}+1} = 0$$

where $m_s \ge n_{k_s}$ and m_s is given by (40). From (40) and the convergence of $(x_{n_k},\, k \in \mathbb{N})$ it follows that

$$\left\|x_{n_{k_s}} + \sum_{i=n_{k_s}}^{m} a_i y_{i+1}\right\| \le c \quad (41)$$

for each $m \in \{n_{k_s}, m_s\}$.

Since $\sigma_n \to \infty$ as $n \to \infty$ there is an s_0 such that if $s \ge s_0$ then $M_{\sigma_{n_{k_s}}} > c$. Thus

$$\left\|x_{n_{k_s}} + \sum_{i=n_{k_s}}^{m} a_i y_{i+1}\right\| \le M_{\sigma_{n_{k_s}}} \quad (42)$$

and

$$x_{m+1} = x_m + a_m y_{m+1} \quad (43)$$

for all $m \in \{n_{k_s}, \cdots, m_s\}$. Inequality (41) implies that

$$\|x_m\| \le c \tag{44}$$

and

$$\|f(x_m)\| \le c' \tag{45}$$

for all $m \in \{n_{k_s}, \cdots, m_{s+1}\}$ where $c' \in \mathbb{R}_+$. By the local Lipschitz condition it follows that

$$\lim_{s\to\infty} \max\{|w^j_{mj}| + |u^j_{mj}|: m \in \{n_{k_s}, \cdots, m_s+1\}\} = 0$$

where w^i_{kj} and u^i_{kj} are given by (32) and (33), respectively, and thus

$$\lim_{s\to\infty} \sum_{i=n_{k_s}}^{m_s} a_i w^j_{ij} = 0$$
$$\lim_{s\to\infty} \sum_{i=n_{k_s}}^{m_s} a_i u^j_{ij} = 0$$

and

$$\lim_{s\to\infty} a_{m_s+1} w^j_{m_s+1j} = 0$$
$$\lim_{s\to\infty} a_{m_s+1} u^j_{m_s+1j} = 0.$$

For $N > c'$ it follows from (38), (39), (44), and (45) that

$$\lim_{s\to\infty} \sum_{i=n_{k_s}}^{m_s} a_i w^j_{i\lambda} = 0$$
$$\lim_{s\to\infty} \sum_{i=n_{k_s}}^{m_s} a_i u^j_{i\lambda} = 0$$

and

$$\lim_{s\to\infty} a_{m_s+1} w^j_{m_s+1\lambda} = 0$$
$$\lim_{s\to\infty} a_{m_s+1} u^j_{m_s+1\lambda} = 0$$

for all $\lambda \ne j$ and $j, \lambda \in \{1, \cdots, l\}$. Combining (38), (39), (44), and (45) it follows that

$$\lim_{s\to\infty} \sum_{i=n_{k_s}}^{m_s} a_i \epsilon_{i+1} = 0$$

and

$$\lim_{s\to\infty} a_{m_s+1} \epsilon_{m_s+2} = 0.$$

The following inequalities are elementary:

$$\begin{aligned}\|x_{m_s+1} - x_{n_{k_s}}\| &\le \sum_{i=n_{k_s}}^{m_s} a_i \|f(x_i)\| + \left\| \sum_{i=n_{k_s}}^{m_s} a_i \epsilon_{i+1} \right\| \\ &\le c' T_s + \left\| \sum_{i=n_{k_s}}^{m_s} a_i \epsilon_{i+1} \right\| \end{aligned} \tag{46}$$

and

$$\begin{aligned}&\|x_{m_s+1} - x_{n_{k_s}} + a_{m_s+1} y_{m_s+2}\| \\ &\quad \le \|x_{m_s+1} - x_{n_{k_s}}\| + a_{m_s+1}\|f(x_{m_s}+1)\| \\ &\quad\quad + \|a_{m_s+1}\epsilon_{m_s+2}\|. \end{aligned} \tag{47}$$

The right-hand sides of the inequalities (46) and (47) tend to zero as $s \to \infty$. However by (40) it follows that

$$\|x_{m_s+1} - x_{n_{k_s}} + a_{m_s+1} y_{m_s+2}\| = \left\| \sum_{i=n_{k_s}}^{m_s+1} a_i y_{i+1} \right\| > \frac{c - \|\overline{x}\|}{2} \tag{48}$$

so that (47) and (48) are not compatible and (36) is verified. ☐

Now the proof of Theorem 1 is given.

Proof—Sufficiency: By (36) it follows that for k sufficiently large there is no truncation for all $m \in \{n_k, \cdots, m(n_k, T)\}$ so

$$x_{m+1} = x_m + a_m y_{m+1} \tag{49}$$

and $\|x_m\| \le 2M$, $\|f(x_m)\| \le M'$ where $M' \in \mathbb{R}_+$.

By these last two inequalities, there is an $N > 0$ such that h_{mj} and g_{mj} given by (29) and (30), respectively, satisfy the inequalities

$$|h_{mj}| \le N$$
$$|g_{mj}| \le N$$

for all $m \in \{n_k, \cdots, m(n_k, T)\}$ and $j \in \{1, \cdots, l\}$ for all k sufficiently large. Therefore, for $\lambda \ne j$

$$\begin{aligned}&\lim_{k\to\infty} \left| \sum_{i=n_k}^{m(n_k,t)} a_i w^j_{i\lambda}(\omega) \right| \\ &\quad = \lim_{k\to\infty} \left| \sum_{i=n_k}^{m(n_k,t)} a_i w^j_{i\lambda}(\omega) 1_{\{|h_{ij}| \le N\}} \right| = 0\end{aligned}$$

for $\lambda \in \{1, \cdots, l\}$ by (38) and (39) for $\omega \in \Omega'$ where Ω' is given in Lemma 2. Thus for $\omega \in \Omega'$ and $\lambda \ne j$

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k,t)} a_i w^j_{i\lambda} \right\| = 0 \tag{50}$$

for all $t \in [0, T]$. For $\lambda = j$ it follows that

$$\begin{aligned}&\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left| \sum_{i=n_k}^{m(n_k,t)} a_i w^\lambda_{i\lambda}(\omega) \right| \\ &\quad = \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left| \sum_{i=n_k}^{m(n_k,t)} a_i (f_\lambda(x_i + c_i\Delta_i - \delta_i(\lambda)) - f_\lambda(x_i)) \right| \\ &\quad = \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left| \sum_{i=n_k}^{m(n_k,t)} a_i o(1) \right| = 0 \end{aligned} \tag{51}$$

where $o(1) \to 0$ as $i \to \infty$ because $c_i \to 0$ and $\delta_i(\lambda) \to 0$ as $i \to \infty$ and $\|x_i\| \le 2M$ for all $i \in \{n_k, \cdots, m(n_k, t)\}$. Thus, the noise condition required in Theorem A in the Appendix is satisfied, while for the function v in Theorem A, $-L$ can be selected (cf. Remark 2). Then sufficiency follows from Theorem A.

Necessity: Now it is verified that (15) implies (16)–(18). Assume that (15) is satisfied and let $\Omega \setminus \mathcal{N}$ be the set, where $(x_n, n \in \mathbb{N})$ converges and the series in (38) and (39) are finite for any $N \in \mathbb{N}$, $P(\mathcal{N}) = 0$ and fix $\omega \in \Omega \setminus \mathcal{N}$. Again for notational convenience the evaluations at ω are suppressed. There is a $K \in \mathbb{N}$ such that if $k \geq \sigma_K$ then

$$x_{k+1} = x_k + a_k y_{k+1}.$$

Use the decomposition in [15] so that

$$\epsilon_{k+1} = \frac{x_{k+1} - x_k}{a_k} - f(x_k) = \epsilon_{k+1}^{(1)} + \epsilon_{k+1}^{(2)} \tag{52}$$

where $\epsilon_{k+1}^{(1)} = -f(x_k)$ and $\epsilon_{k+1}^{(2)} = (x_{k+1} - x_k)/a_k$. It is clear that

$$\sum_{k=1}^{\infty} a_k \epsilon_{k+1}^{(2)} < \infty \tag{53}$$

and

$$\lim_{k \to \infty} \epsilon_{k+1}^{(1)} = 0. \tag{54}$$

Since $(x_k, k \in \mathbb{N})$ is convergent and (38) and (39) are satisfied, it follows that $\lim_{k\to\infty} w_{ki}^i = 0$, $\lim_{k\to\infty} u_{ki}^i = 0$ and $\sum_{i=1}^{\infty} a_i w_{i\lambda}^j < \infty$, $\sum_{i=1}^{\infty} a_i w_{i\lambda}^j < \infty$ for all $\lambda \neq j$ and λ, $j \in \{1, \cdots, l\}$. Thus $\lim_{k\to\infty} w_k(1) = 0$, $\lim_{k\to\infty} u_k(1) = 0$ and

$$\sum_{i=1}^{\infty} a_i w_i(2) < \infty \tag{55}$$

$$\sum_{i=1}^{\infty} a_i u_i(2) < \infty \tag{56}$$

where $w_k(1)$, $u_k(1)$, $w_k(2)$, and $u_k(2)$ are defined in Lemma 1. From (26) and (52) it follows that

$$\frac{1}{c_k} \xi_{k+1} g_k = -(2f(x_k) + w_k(1)) + u_k(1)) + 2\epsilon_{k+1}^{(2)} - w_k(2) - u_k(2)). \tag{57}$$

By (35) it follows that

$$2f_i(x_k) + w_{ki}^i + u_{ki}^i = \frac{L(x_k + c_k\Delta_k) - L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(i)))}{2c_k\Delta_k^i} + \frac{L(x_k - c_k(\Delta_k(i-1) + \Delta_k^c(i))) - L(x_k - c_k\Delta_k)}{2c_k\Delta_k^i}.$$

Let ν_{k+1}^i be given by

$$\nu_{k+1}^i = \tfrac{1}{2}[L(x_k + c_k(\Delta_k(i-1) + \Delta_k^c(i))) - L(x_k + c_k\Delta_k) + L(x_k - c_k\Delta_k)) - L(x_k - c_k(\Delta_k(i-1) + \Delta_k^c(i)))]. \tag{58}$$

By (27) and (28) it follows that

$$\frac{\nu_{k+1}^i}{c_k\Delta_k^i} = -f_i(x_k + c_k\Delta_k + \delta_k(i)) - f_i(x_k - c_k\Delta_k + \delta_k(i)) \tag{59}$$

for $i \in \{1, \cdots, l\}$ and

$$\lim_{k\to\infty} f_i(x_k \pm c_k\Delta_k + \delta_k(i)) = f_i(x^0) = 0.$$

Let e_{k+1}^j be given by

$$e_{k+1}^j = \xi_{k+1} - \nu_{k+1}^j.$$

By (57) and (58) it follows that

$$2\epsilon_{k+1}^{(2)} - w_k(2) - u_k(2) = \left[\frac{e_{k+1}^1}{c_k\Delta_k^1}, \cdots, \frac{e_{k+1}^l}{c_k\Delta_k^l}\right]^T.$$

By (53)–(56) it follows that

$$\sum_{k=1}^{\infty} a_k \frac{e_{k+1}^j}{c_k\Delta_k^j} < \infty \quad \text{a.s.} \tag{60}$$

for $j \in \{1, \cdots, l\}$. Equations (58)–(60) give the required decomposition (16)–(18). The equations (16)–(18) can be verified for the one-sided algorithm in a similar fashion. □

Assuming an independence condition the following theorem gives a large family of noise processes that satisfy (16)–(18).

Theorem 2: Let H1) and H2) be satisfied. If $\sum_{k=1}^{\infty}(a_k^2/c_k^2) < \infty$ and the observation noise $(\xi_k, k \in \mathbb{N})$ has the property that $\xi(k+1)$ is independent of $(\Delta_j, j = 1, 2, \cdots, k)$ for each $k \in \mathbb{N}$ and satisfies one of the following two conditions:

1) $\sup_k |\xi_k| \leq \xi$ a.s. where ξ is a random variable;
2) $\sup_k E\xi_k^2 < \infty$;

then

$$\lim_{k\to\infty} x_k = x^0 \quad \text{a.s.} \tag{61}$$

where x_k is given by (13).

Proof: Initially it should be noted that ξ_k may depend arbitrarily on $(\xi_j, j \neq k)$ and may not be zero mean, e.g., a sequence of bounded deterministic observation errors satisfies the conditions.

It is only necessary to verify (16)–(18). Let 1) be satisfied, that is, there is a random variable ξ such that $\sup_{k\in\mathbb{N}} |\xi_k| \leq \xi < \infty$. For $j \in \{1, \cdots, \ell\}$ and $k \in \mathbb{N}$ let $\mathcal{F}_k^j$ be given by

$$\mathcal{F}_k^j = \sigma(\Delta_i^j, \ i \in \{0, \cdots, k\}, \xi_p, p \in \{0, \cdots, k+2\}).$$

By the definition of Δ_k, it is independent of ξ_{k+1} and $\mathcal{F}_{k-1}^j$ so that

$$E\left[\frac{\xi_{k+1}}{\Delta_k^j}\middle|\mathcal{F}_{k-1}^j\right] = \xi_{k+1} E\left[\frac{1}{\Delta_k^j}\middle|\mathcal{F}_{k-1}^j\right] = 0 \quad \text{a.s.}$$

and

$$\sum_{k=1}^{\infty} E\left[\frac{a_k^2}{c_k^2}\frac{\xi_{k+1}^2}{(\Delta_k^j)^2}\middle|\mathcal{F}_{k-1}^j\right] \leq \sum_{k=1}^{\infty}\frac{a_k^2}{c_k^2} b^2\xi^2 < \infty \quad \text{a.s.}$$

By the convergence theorem for mds's [3], [6], it follows that

$$\sum_{k=1}^{\infty}\frac{a_k}{c_k}\frac{\xi_{k+1}}{\Delta_k^j} < \infty \quad \text{a.s.}$$

for all $j \in \{1, \cdots, \ell\}$. Thus in (16) it can be assumed that $\nu_k^j \equiv 0$ and $e_k^j = \xi_k$ for $j = 1, \cdots, \ell$.

Now assume that $\sup_{k\in\mathbb{N}} E\xi_k^2 < \infty$ is satisfied.

By the independence assumption it follows that for $j < k$, Δ_k is independent of $(\Delta_k, \xi_{k+1}, \xi_{j+1})$ so that

$$E\left[\frac{\xi_{j+1}\xi_{k+1}}{\Delta_j^i \Delta_k^i}\right] = E\left[\frac{\xi_{j+1}\xi_{k+1}}{\Delta_k^i}\right] E\left[\frac{1}{\Delta_j^i}\right] = 0$$

for all $i \in \{1, \cdots, \ell\}$ so that

$$E\left[\left(\sum_{k=1}^{\infty} \frac{a_k}{c_k} \frac{\xi_{k+1}}{\Delta_k^j}\right)^2\right] \le b^2 \sup_{k\in s} E\xi_{k+1}^2 \sum_{k=1}^{\infty} \frac{a_k^2}{c_k^2} < \infty$$

where b is an upper bound of $|1/\Delta_k^i|$. It follows directly that

$$\sum_{k=1}^{\infty} \frac{a_k}{c_k} \frac{\xi_{k+1}}{\Delta_k^j} < \infty \quad \text{a.s.}$$

for all $j = 1, \cdots, \ell$. To verify (17) and (18) it suffices to choose $\nu_k^j \equiv 0$ and $e_k^j = \xi_k$. □

Remark 2: Using Theorem A in the Appendix, Theorems 1 and 2 can be extended to the case where $f(x) = 0$ for all $x \in J$ and J is not a singleton. In this case H1) is replaced by C2 2) and 3) in Theorem A where $v = -L$ and f is locally Lipschitz continuous.

III. Rate of Convergence and Asymptotic Normality

In the following theorem a rate of convergence of the algorithm (13) is given.

Theorem 3: Assume the hypotheses of Theorem 2 and that

$$\lim_{n\to\infty} (a_{n+1}^{-1} - a_n^{-1}) = \alpha \ge 0 \tag{62}$$

$$c_k = o(a_k^\delta)$$

$$\sum_{j=1}^{\infty} a_j^{2(1-\delta)}/c_j^2 < \infty \tag{63}$$

for some $\delta \in (0, 1)$ and

$$f(x) = F(x - x^0) + \gamma(x) \tag{64}$$

where $F \in L(\mathbb{R}^\ell, \mathbb{R}^\ell)$, $\gamma(x) = o(\|x - x^0\|)$ and $F + \alpha\gamma I$ is stable. Then $(x_n, n \in \mathbb{N})$ given by (13) satisfies

$$\|x_n - x^0\| = o(a_n^\delta) \quad \text{a.s.} \tag{65}$$

for δ given in (63).

Proof: By Theorem B in the Appendix it suffices to show that $(\varepsilon_n, n \in \mathbb{N})$ given by (25) or (26) can be represented as $\varepsilon_n = e_n + \nu_n$ where $\nu_n = o(a_n^\delta)$ as $n \to \infty$ and $\sum_{k=1}^{\infty} a_k^{1-\delta} e_{k+1} < \infty$ a.s. It suffices to verify this result for (25), because the proof for (26) is similar. It follows by Theorem 2 that $\lim_{n\to\infty} x_n = x^0$ a.s. By the local Lipschitz condition on f it follows from (31) that

$$\|w_k(1)\| = O(c_k) = o(a_k^\delta) \tag{66}$$

as $k \to \infty$. Since $c_k = o(a_k^\delta)$ it follows that

$$a_k^{1-\delta} = o\left(\frac{a_k}{c_k}\right)$$

and

$$\sum_{k=1}^{\infty} a_k^{2(1-\delta)} < \infty.$$

It follows that $|h_{kj}(\omega)|$ in (38) is uniformly bounded for $j \in \{1, \cdots, \ell\}$ and $k \in \mathbb{N}$ for each ω where $(x_n, n \in \mathbb{N})$ converges to x^0. By the convergence theorem for mds's it follows that

$$\sum_{j=1}^{\infty} a_j^{1-\delta} w_j(2) < \infty \quad \text{a.s.}$$

As in the proof of Theorem 2 it follows from (63) that

$$\sum_{j=1}^{\infty} a_j^{1-\delta} \frac{\xi_{j+1}}{c_j} g_j < \infty \quad \text{a.s.}$$

Using (66) and the last two inequalities verifies (65). □

Remark 3: If $a_n = 1/n$ and $c_n = 1/n^v$ for some $v \in (0, 1/2)$ and all $n \in \mathbb{N}$, then the conditions on $(a_n, n \in \mathbb{N})$ and $(c_n, n \in \mathbb{N})$ in Theorem 3 are satisfied.

The following result is an asymptotic normality property of $(x_n, n \in \mathbb{N})$ given by the algorithm (13).

Theorem 4: Assume that the conditions H1) and H2) are satisfied and that:

1) $\lim_{n\to\infty}(a_{n+1}^{-1} - a_n^{-1}) = \alpha > 0$ and $c_n = a_n^\gamma$ for some $\gamma \in (1/4, 1/2)$;
2) $\|f(x) - F(x - x^0)\| \le b\|x - x^0\|^{1+\beta}$ for some $\beta > 0$ and $b > 0$;
3) $F + \alpha\gamma I$ is stable and $\sum_{j=1}^{\infty} a_j^{2(1-\gamma-\delta)} < \infty$ for some $\delta \in (\gamma/(1+\beta), \gamma)$;
4) $\xi_n = \sum_{i=0}^{r} b_i w_{n-i}$ if $\{y_n\}$ is given by (12), and $\xi_n = 2\sum_{i=0}^{r} b_i w_{n-i}$ if $\{y_n\}$ is given by (12′) where $w_i = 0$ for $i < 0$, $(b_i, i \in \{0, 1, \cdots, r\})$ is a sequence of real numbers, $r \in \mathbb{N}$ is fixed and $(w_i, \mathcal{F}_i^w, i \in \mathbb{N})$ is an mds that is independent of $(\Delta_i, i \in \mathbb{N})$ and satisfies
$$E[w_i^2|\mathcal{F}_{i-1}^w] \le \sigma_0$$
for all $i \in \mathbb{N}$ where $\sigma_0 \in \mathbb{R}_+$
$$\lim_{i\to\infty} E[w_i^2|\mathcal{F}_{i-1}^w] = \sigma^2$$
where $\sigma^2 \in \mathbb{R}_+$ and
$$\lim_{N\to\infty} \sup_{i\in\mathbb{N}} E[w_i^2 1_{\{|w_i|>N\}}] = 0.$$

Then for both the one-sided and the two-sided algorithms

$$a_n^{-\mu}(x_n - x_0) \xrightarrow{d} Z \tag{67}$$

where $\mu = (1/2) - \gamma$ and Z is an $N(0, S)$ random variable

$$S = \sigma^2\sigma_\Delta^2 \left(\sum_{i=0}^{r} b_i\right)^2 \int_0^\infty e^{t(F+\alpha\mu I)} e^{t(F+\alpha\mu I)^T}\, dt$$

$$\sigma_\Delta^2 = E\left[\frac{1}{(\Delta_i^1)^2}\right] \tag{68}$$

and r is given in 4).

Proof: Since $\gamma > \delta$, it follows that $c_k = a_k^\gamma = o(a_k^\delta)$, $\sum_{j=1}^\infty a_j^{2(1-\delta)}/c_j^2 = \sum_{j=1}^\infty a_j^{2(1-\gamma-\delta)} < \infty$ and by Theorem 3

$$\|x_n - x^0\| = o(a_n^\mu) \tag{69}$$

and after a finite number of iterations of (13) there are no subsequent truncations.

Since $\mu = (1/2) - \gamma$, $\gamma \in (1/4, 1/2)$, $0 < \mu < 1/4$, $\gamma > \mu$, and $F + \alpha\mu I$ is stable, it follows directly that

$$\begin{aligned}\frac{a_{n+1}^{-\mu}}{a_n^{-\mu}} &= \left(1 + \frac{a_{n+1}^{-1} - a_n^{-1}}{a_n^{-1}}\right)^\mu\\ &= (1 + \alpha a_n + a_n(a_{n+1}^{-1} - a_n^{-1} - \alpha))^\mu\\ &= 1 + \alpha\mu a_n + o(a_n).\end{aligned}$$

Let u_n be given by

$$u_n = a_n^{-\mu}(x_n - x^0)$$

so that

$$\begin{aligned}u_{n+1} &= (1 + \alpha\mu a_n + o(a_n)) \times \Bigg(u_n + a_n a_n^{-\mu}\\ &\quad \times \bigg(F(x_n - x^0) + f(x_n) - F(x - x^0)\\ &\quad + w_n(1) + w_n(2) + \frac{1}{c_n}\xi_{n+1} g_n\bigg)\Bigg)\\ &= [I + a_n(F + \alpha\mu I + o(1))]u_n\\ &\quad + (1 + O(a_n))a_n^{(1/2)+\gamma} b\|x_n - x^0\|^{1+\beta}\\ &\quad + (1 + O(a_n))a_n^{(1/2)+\gamma}(w_n(1) + w_n(2))\\ &\quad + (1 + O(a_n))a_n^{(1/2)}\xi_{n+1} g_n\end{aligned} \tag{70}$$

where subsequently the verification is made only for (12) because for (12′) it is only necessary to replace ξ_{n+1} by $\xi_{n+1}/2$, and $w_n(1)$, $w_n(2)$, and g_n are given in Lemma 1.

Let Φ_{nk} be given by

$$\begin{aligned}\Phi_{nk} &= \prod_{i=k}^n (I + a_i(F + \alpha\mu I) + o(1))\\ \Phi_{j,\,j+1} &= I\end{aligned}$$

where $k < n$. Since $F + \alpha\mu I$ is stable it follows that for n large

$$\|\Phi_{nk}\| \le \lambda_1 \exp\left(-\lambda \sum_{i=k}^n a_i\right) \tag{71}$$

where λ_1 and λ are strictly positive constants.

Consider the following algorithm:

$$\begin{aligned}\hat\eta_{n+1} &= (I + a_n(F + \alpha\mu I + o(1)))\eta_n + a_n^{(1/2)+\gamma} w_n(2)\\ \eta_{n+1} &= \hat\eta_{n+1} 1_{\{\|\hat\eta_{n+1}\| \le M_{\sigma_n}\}} + \eta^* 1_{\{\|\hat\eta_{n+1}\| > M_{\sigma_n}\}}\\ \sigma_n &= \sum_{i=0}^{n-1} 1_{\{\|\hat\eta_i\| > M_{\sigma_n}\}}\end{aligned}$$

where $M_k > 0$ for all $k \in \mathbb{N}$ and $M_k \uparrow \infty$ as $k \to \infty$. Since $\sum_{k=1}^\infty a_k^{2-2\gamma} < \infty$ and $\gamma > 1/4$, $\sum_{k=1}^\infty a_k^{1+2\gamma} < \infty$. By the martingale convergence theorem $\sum_{k=1}^\infty a_k^{(1/2)+\gamma} w_k(2) < \infty$ a.s.

Similar to the proof of Lemma 2, it can be verified that for any convergent subsequence $\{\eta_{n_k}\}$ there are positive constants M, T, and k_T such that if $k > k_T$, then

$$\left\|\sum_{i=n_k}^{m+1} a_i[(F + \alpha\mu I + o(1))\eta_i + a_i^{\gamma-(1/2)} w_i(2)]\right\| \le M.$$

Therefore, for any fixed ω there is $c(\omega)$ such that $\|\eta_i(\omega)\| \le c(\omega)$, $\forall i$: $n_k \le i \le m(n_k, T)$, and

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T}\left\|\sum_{i=n_k}^{m(n_k,\,t)} (a_i o(1)\eta_i + a_i^{(1/2)+\gamma} w_i(2))\right\| = 0.$$

Since $F + \alpha\mu I$ is stable, by Theorem A, $\eta_n \to 0$ a.s. as $n \to \infty$. Therefore, the algorithm for η_n becomes the one without truncations starting from some n_0, i.e.,

$$\eta_{n+1} = (I + a_n(F + \alpha\mu I + o(1))\eta_n + a_n^{(1/2)+\gamma} w_n(2)$$

for $n \ge n_0$ and hence

$$\eta_{n+1} = \Phi_{n,\,n_0}\eta_{n_0} + \sum_{i=n_0}^n \Phi_{n,\,i+1} a_i^{(1/2)+\gamma} w_i(2).$$

By (71) and the convergence of $(\eta_n,\ n \in \mathbb{N})$ to zero it follows that

$$\sum_{i=n_0}^n \Phi_{n,\,i+1} a_i^{(1/2)+\gamma}(2) \to 0, \quad \text{a.s. as } n \to \infty$$

and hence $\sum_{i=1}^n \Phi_{n,\,i+1} a_i^{(1/2)+\gamma} w_i(2) \to 0$ a.s. as $n \to \infty$, which implies

$$\lim_{n\to\infty} \sum_{i=1}^n \Phi_{n,\,i+1}(1 + O(a_i))a_i^{(1/2)+\gamma} w_i(2) = 0 \quad \text{a.s.}$$

By (69) it follows that

$$\begin{aligned}&(1 + O(a_n))a_n^{(1/2)+\gamma} b\|x_n - x^0\|^{1+\beta}\\ &\quad \le b(1 + O(a_n))a_n^{(1/2)+\gamma} a_n^{(1+\beta)\delta} = O(a_n^{1+\epsilon})\end{aligned}$$

where $\epsilon \triangleq (1+\beta)\delta + (1/2) + \gamma > (1/2) + 2\gamma \triangleq \epsilon' > 0$, and from (27), it follows that

$$(1 + O(a_n))a_n^{(1/2)+\gamma} w_n(1) = O(a_n^{(1/2)+\gamma} O(c_n)) = O(a_n^{1+\epsilon'}).$$

Solving the recursion (70) it follows that

$$\begin{aligned}u_{n+1} &= \Phi_{n1} u_1 + \sum_{i=1}^n \Phi_{ni+1} o(a_i)\\ &\quad + \sum_{i=1}^n \Phi_{n,\,i+1}(1 + O(a_i))a_i^{(1/2)+\gamma} w_i(2)\\ &\quad + \sum_{i=1}^n \Phi_{ni+1} a_i^{1/2} \xi_{i+1} g_i.\end{aligned}$$

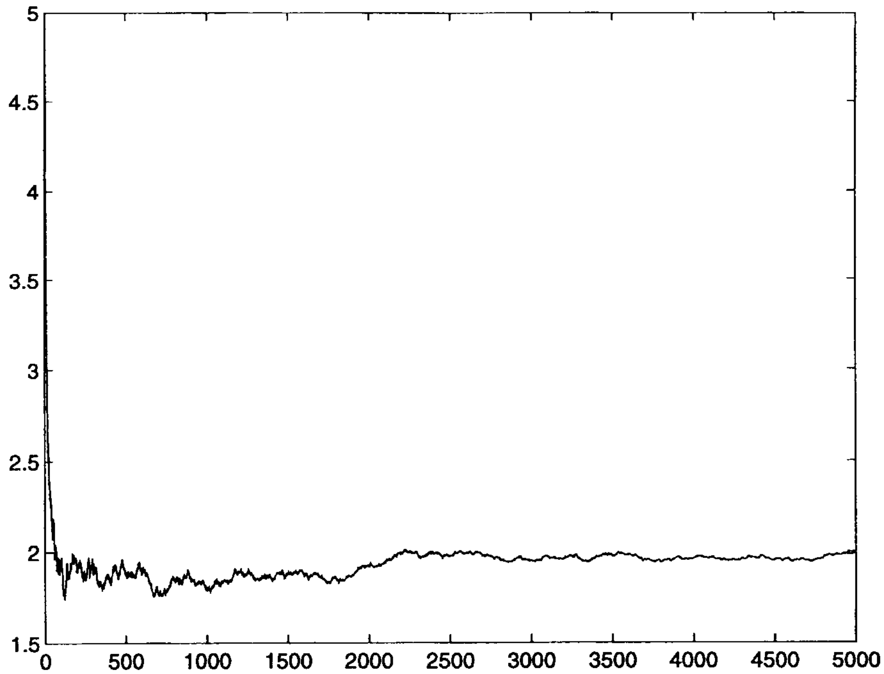

Fig. 1. The path of x_k with $x^* = 5$ and initial value $x_0 = 3$.

Since

$$\sum_{i=1}^{n} a_i \exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right)$$

$$\leq \frac{1}{\lambda} \sum_{i=1}^{n} \left(1 - e^{-\lambda a_i} + \frac{(\lambda a_i)^2}{2}\right) \exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right)$$

$$= \frac{1}{\lambda} \sum_{i=1}^{n} \left(\exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right) - \exp\left(-\lambda \sum_{j=i}^{n} a_j\right)\right)$$

$$+ \frac{\lambda}{2} \sum_{i=1}^{n} a_i^2 \exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right)$$

it follows that

$$\sum_{i=1}^{n} \left(1 - \frac{\lambda a_i}{2}\right) a_i \exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right) \leq \frac{1}{\lambda}.$$

Since $a_n \to 0$ as $n \to \infty$ it follows that

$$\sum_{i=1}^{n} a_i \exp\left(-\lambda \sum_{j=i+1}^{n} a_j\right)$$

is uniformly bounded for $n \in \mathbb{N}$. This boundedness and (71) imply that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \Phi_{n,\, i+1} \cdot (a_i) = 0.$$

Thus

$$\lim_{n\to\infty} \left(u_{n+1} - \sum_{i=1}^{n} \Phi_{n,\, i+1} a_i^{1/2} \xi_{i+1} g_i\right) = 0 \quad \text{a.s.}$$

Since $(\Delta_n,\ n \in \mathbb{N})$ is independent of $(\xi_n,\ n \in \mathbb{N})$ it follows by a standard method [5], [8], which is sketched in the Appendix, that

$$\sum_{i=1}^{n} \Phi_{n,\, i+1} a_i^{1/2} \xi_{i+1} g_i \xrightarrow{d} Z$$

where Z is $N(0, S)$. □

IV. A Numerical Example

An elementary numerical example is given of the one-sided algorithm (10), (13). Let $L(x) = -x^2 + 4x + 2$, which has a unique maximum at $x = 2$. The random variable Δ_k^i is uniformly distributed on $[-1, -0.5] \cup [0.5, 1]$ for $i \in \{1, \cdots, \ell\}$. The noise processes $(\xi_k^+,\ k \in \mathbb{N})$ and $(\xi_k^0,\ k \in \mathbb{N})$ are independent white Gaussian processes and the constants a_k, c_k, and M_k are $a_k = 1/(k+20)$, $c_k = 1/k^{1/5}$, and $M_k = 2k$ for $k \in \mathbb{N}$. In Fig. 1 is a graph of $\{x_k,\ 0 \leq k \leq 5000\}$ with $x_0 = 3$ and $x^* = 5$. In Figs. 2 and 3 are graphs of $(y_k^+,\ 0 \leq k \leq 5000)$ and $(y_k^0,\ 0 \leq k \leq 5000)$, respectively.

V. Conclusion

The classical Kiefer–Wolfowitz algorithm has been modified in two ways: 1) the one-sided and the two-sided randomized differences are used instead of the two-sided deterministic differences and 2) the estimates are truncated at randomly varying bounds. For the convergence analysis, a direct method is used rather than the classical probabilistic method or the ODE method. By the algorithm modifications 1) and 2) and a different approach to algorithm analysis, the following algorithm improvements have been made: 1) some restrictive

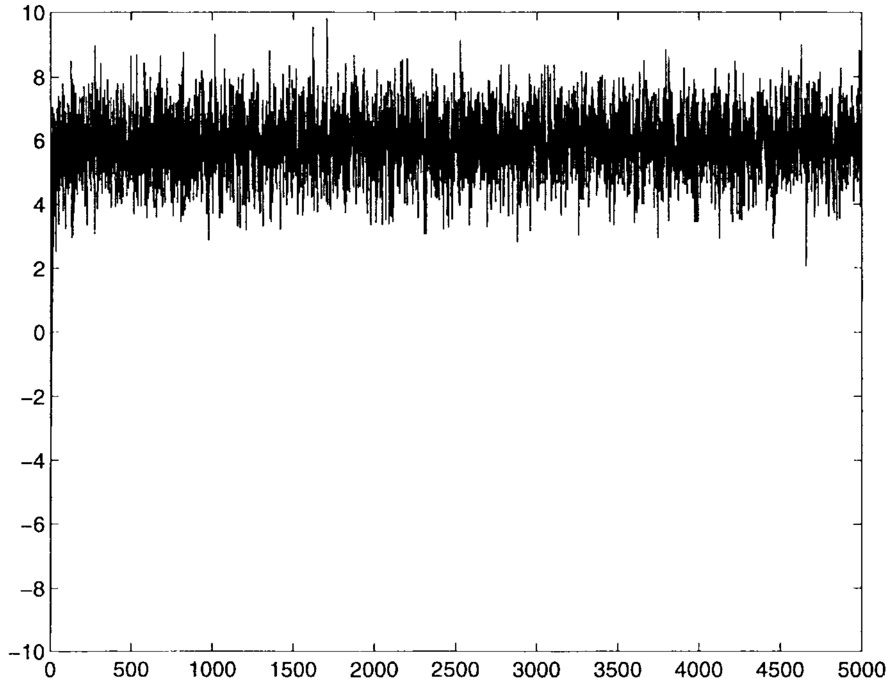

Fig. 2. The path of the observation y_{k+1}^{+}.

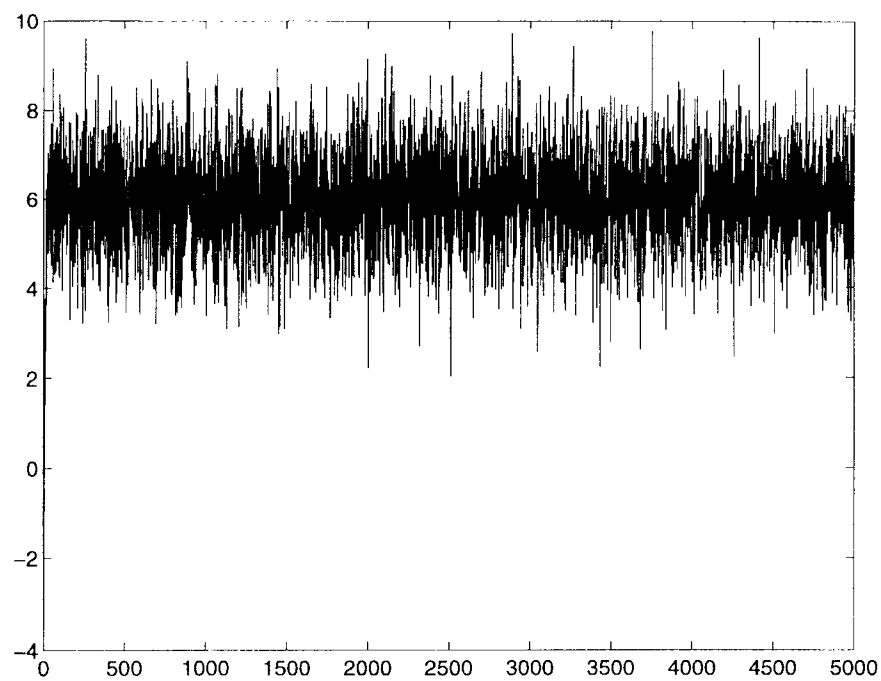

Fig. 3. The path of the observation y_{k+1}^{0}.

conditions on the function L or some boundedness assumptions on the estimates have been removed; 2) some restrictive conditions on the noise process have been removed; and 3) the number of required observations at each iteration have been reduced. If the function L has many extrema then the algorithm may become stuck at a local extremum. To obtain the almost sure convergence to global extrema, some methods that combine search and stochastic approximation are needed, but it seems that there is a lack of a complete theory for this approach.

APPENDIX

Theorem A: Assume that the following conditions are satisfied.

- C0) The function f: $\mathbb{R}^\ell \to \mathbb{R}$ is Borel measurable and locally bounded and $J = \{x: f(x) = 0\}$.
- C1) The sequence of strictly positive numbers $(a_n, n \in \mathbb{N})$ converges to zero and $\sum_{n=1}^{\infty} a_n = \infty$.
- C2) There is a continuously differentiable function v: $\mathbb{R}^\ell \to \mathbb{R}$ such that

1) $\sup_{\delta \le d(x, J) \le \Delta} f^T(x) v_x(x) < 0$ for all $\Delta > \delta > 0$ where v_x is the gradient of v and $d(x, S) = \inf\{\|x - y\|: y \in S\}$.
2) $v(J)$ is nowhere dense.
3) There is a $c_0 > 0$ such that $\|x^*\| < c_0$

$$v(x^*) < \inf_{\|x\|=c_0} v(x).$$

Then $(x_n, n \in \mathbb{N})$ defined by (13) with

$$y_{n+1} = f(x_n) + \varepsilon_{n+1} \tag{72}$$

converges to the set J on A where

$$A = \left\{ \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k, t)} a_i \varepsilon_{i+1} \right\| = 0 \right.$$
$$\text{for all } t \in [0, T] \text{ whenever } (n_k, k \in \mathbb{N}) \text{ is such}$$
$$\left. \text{that } (x_{n_k}, k \in \mathbb{N}) \text{ is convergent} \right\}$$

and

$$m(k, T) = \max\left\{ m: \sum_{i=k}^{m} a_i \le T \right\}.$$

The proof of this result is given in [1] or [2].

Remark A1: Given the conditions of Theorem A, $(x_n, n \in \mathbb{N})$ converges so that

$$A = A'$$
$$=: \left\{ \lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=k}^{m(k, T)} a_i \varepsilon_{i+1} \right\| = 0 \right.$$
$$\left. \text{for all } t \in [0, T] \right\}.$$

However, it should be noted that the equality $A = A'$ is satisfied only with the conditions C0)–C2). Without these conditions there is only the inclusion $A' \subset A$. If ε_{i+1} depends on x_i then since the behavior of $(x_n, n \in \mathbb{N})$ is not known it is difficult to verify if a point is in A' while to verify a point is in A is easier because only convergent subsequences need to be considered.

Remark A2: If $J = \{x^0\}$ then C2-2) is immediately satisfied and C2-3) becomes $\|x^*\| < c_0$ and $v(x^*) < \inf_{\|x\|=c_0} v(x)$ for some c_0.

Theorem B: If the conditions C0)–C2) of Theorem A are satisfied where $J = \{x^0\}$ and

$$\lim_{n\to\infty} (a_{n+1}^{-1} - a_n^{-1}) = \alpha \ge 0$$

$\varepsilon_n = e_n + \nu_n$, $\nu_n = o(a_n^\gamma)$, $\sum_{i=1}^{\infty} a_i^{1-\gamma} e_{i+1} < \infty$, $H + \alpha\gamma I$ is stable for some $\delta \in (0, 1]$ and as $x \to x^0$, $f(x) = H(x - x^0) + \delta(x)$, $\delta(x^0) = 0$, and $\delta(x) = o(\|x - x^0\|)$ then

$$\|x_n - x^0\| = o(a_n^\delta)$$

as $n \to \infty$ where $(x_n, n \in \mathbb{N})$ is given by (13) and y_{n+1} is given by (71).

The proof of this theorem is given in [4].

A sketch of the proof that $\sum_{i=1}^{n} \Phi_{n, i+1} a_i^{1/2} \xi_{i+1} g_i \xrightarrow{d} Z$ is given. Let $\Psi_{n, k} = \prod_{i=k}^{n} (I + a_i(F + \alpha\mu I))$, $\Psi_{j, j+1} = I$.

Then

$$\lim_{n\to\infty} \sum_{i=1}^{n} \Phi_{n, i+1} a_i^{1/2} \xi_{i+1} g_i - \sum_{i=1}^{n} \Psi_{n, i+1} a_i^{1/2} \xi_{i+1} g_i = 0 \quad \text{a.s.}$$

Therefore, it suffices to show that

$$\sum_{i=1}^{n} \Psi_{n, i+1} a_i^{1/2} \xi_{i+1} g_i \xrightarrow{d} N(0, S).$$

By using the Central Limit theorem for random vectors with two indices in [13], it is straightforward to verify that all of the conditions required for $\xi_{n, i+1} \triangleq \Psi_{n, i+1} a_i^{1/2} \cdot (\sum_{j=0}^{r} b_j g_{i+j}) w_{i+1}$ are satisfied.

Hence

$$\sum_{i=1}^{n} \Psi_{n, i+1} a_i^{1/2} \left(\sum_{j=0}^{r} b_j g_{i+j} \right) w_{i+1} \xrightarrow{d} N(0, S).$$

It now remains to show that

$$\lim_{n\to\infty} \eta_n \triangleq \sum_{i=1}^{n} \Psi_{n, i+1} a_i^{1/2} \left(\sum_{j=0}^{r} b_j w_{i+1-j} g_i \right) - \sum_{i=1}^{n} \Psi_{n, i+1} a_i^{1/2} \left(\sum_{j=0}^{r} b_j g_{i+j} \right) w_{i+1} = 0$$

in probability.

Write η_n as

$$\eta_n = \sum_{j=0}^{r} \sum_{i=1}^{j} \Psi_{n, i+1} a_i^{1/2} b_j w_{i+1-j} g_i - \sum_{j=0}^{r} \sum_{i=n-j+1}^{n} \Psi_{n, i+1} a_i^{1/2} b_j w_{i+1} g_{i+j} - \sum_{j=0}^{r} \sum_{i=1}^{n-j} (a_i^{1/2} \Psi_{n, i+1} - a_{i+j}^{1/2} \Psi_{n, i+j+1}) b_j w_{i+1} g_{i+j}.$$

By (71) the first term on the right-hand side of η_n tends to zero a.s., while the second term tends to zero in probability since the expectation of its norm converges to zero as $n \to \infty$.

The last term in the expression of η_n equals

$$\sum_{j=0}^{r} \sum_{i=1}^{n-j} a_i^{1/2} \sum_{k=1}^{j} \Psi_{n, i+k+1} a_{i+k} (F + \alpha\mu I) b_j w_{i+1} g_{i+j} + \sum_{j=0}^{r} \sum_{i=1}^{n-j} \left(\left(\frac{a_i}{a_{i+j}} \right)^{1/2} - 1 \right) a_{i+j}^{1/2} \Psi_{n, i+j+1} b_j w_{i+1} g_{i+j}.$$

Taking expectation of the norm of each term and noting that $\sup_n \sum_{k=1}^{n} a_k \|\Psi_{n, k+1}\|^r < \infty$, $\forall r > 0$, and $(a_n/a_{n+1})^{1/2} = 1 + (1/2)\alpha a_n + o(a_n)$, it is concluded that both terms converge to zero in probability. Therefore, $\eta_n \to 0$ in probability as $n \to \infty$. $\square$

ACKNOWLEDGMENT

The authors would like to thank J. Spall and the referees for their useful comments.

REFERENCES

[1] H. F. Chen, "Stochastic approximation and its new applications," in *Proc. Hong Kong Int. Workshop on New Directions of Control and Manufacturing*, 1994, pp. 2–12.

[2] H. F. Chen, T. E. Duncan, and B. Pasik-Duncan, "On the ODE approach to parameter identification and its application to adaptive control," to be published.

[3] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control.* Boston, MA: Birkhauser, 1991.

[4] H. F. Chen and Y. M. Zhu, "Asymptotical properties of randomly truncated stochastic approximation procedures," *Acta Math. Scientia,* vol. 7, pp. 431–441, 1986.

[5] ——, *Stochastic Approximation.* Shanghai Sci., 1996.

[6] Y. S. Chow, "Local convergence of martingales and the law of large numbers," *Ann. Math. Stat.*, vol. 36, pp. 552–558, 1965.

[7] V. Fabian, "Stochastic approximation," in *Optimizing Methods in Statistics,* J. S. Rustagi, Ed. New York: Academic, 1971, pp. 439–470.

[8] P. Hall and C. C. Heude, *Martingale Limit Theory and Its Applications.* New York: Academic, 1980.

[9] J. Kiefer and J. Wolfowitz, "Stochastic approximation of a regression function," *Ann. Math. Stat.*, vol. 23, pp. 462–466, 1952.

[10] J. Koronacki, "Random-seeking methods for the stochastic unconstrained optimization," *Int. J. Contr.*, vol. 21, pp. 517–527, 1975.

[11] ——, "Some remarks on stochastic approximation methods," in *Numerical Techniques for Stochastic Systems,* F. Archetti and M. Cugiani, Eds. Amsterdam, The Netherlands: North-Holland, 1980, pp. 395–406.

[12] H. J. Kushner and D. A. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* New York: Springer, 1978.

[13] M. B. Nevelson and R. Z. Hasminski, *Stochastic Approximation and Recursive Estimation*, vol. 47, Translations of Math. Monograph. AMS, 1996.

[14] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Contr.,* vol. 37, pp. 332–341, 1992.

[15] I. J. Wang, E. K. P. Chong, and S. R. Kulkarni, "Equivalent necessary and sufficient conditions on noise sequences for stochastic approximation algorithms," *Advances Appl. Probability,* vol. 23, pp. 784–801, 1996.

H. F. Chen (SM'94–F'97) received the Diploma from the Department of Mathematics and Mechanics, Leningrad (St. Petersburg) University, Russia.

He joined the Institute of Mathematics, Chinese Academy of Sciences, in 1961. Since 1979 he has been with the Institute of Systems Science, Chinese Academy of Sciences. He is a Professor of the Laboratory of System and Control and currently serves as the Director of the Institute. His research interests include stochastic systems, system identification, adaptive control, recursive estimation, and stochastic approximation. He has authored and coauthored more than 120 journal papers.

Mr. Chen is the Editor of *Systems Science and Mathematical Sciences* and is involved in the editorial boards of several international and domestic journals. He was elected to the Chinese Academy of Sciences in 1993. He now serves as the President of the Chinese Association of Automation and is a Standing Member of the Council of the Chinese Mathematics Society. He also serves on the Technical Board of IFAC, chairing the Coordinating Committee Systems and Signals. For the 14th IFAC Congress to be held in 1999, he serves as the IPC Chairman and NOC Cochairman.

T. E. Duncan (M'92–SM'96–F'98) received the B.E.E. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1963, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1964 and 1967, respectively.

He has held regular positions at the University of Michigan (1967–1971), the State University of New York, Stony Brook (1971–1974), and the University of Kansas, Lawrence (1974–present), where he is a Professor of Mathematics. He has held visiting positions at the University of California, Berkeley (1969–1970), the University of Bonn, Germany (1978–1979), and Harvard University, Cambridge, MA (1979–1980), and shorter visiting positions at numerous other institutions.

Dr. Duncan is an Associate Editor of *SIAM Journal on Control and Optimization* and is a member of AMS, MAA, and SIAM.

B. Pasik-Duncan (M'90–SM'95) received the M.S. degree from the Mathematics Department of Warsaw University in 1970 and the Ph.D. and Habilitation degrees from the Mathematics Department of Warsaw School of Economics in 1978 and 1986, respectively.

She was a staff member of the Mathematics Department of the Warsaw School of Economics from 1970 to 1983. In 1984 she joined the University of Kansas, where she is a Professor of Mathematics. Her research interests include adaptive control and its applications, stochastic theory and its applications, mathematics of finance, numberical methods in control, manufacturing systems, distributed parameter systems, mathematical statistics, and mathematical education.

Dr. Pasik-Duncan is currently serving as Program Director of the SIAM Activity Group on Control and Systems Theory and is a member of the SIAM Master Program Committee. She is an Associate Editor-at-Large of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and is currently serving as Chair of the Standing Committees on Assistance of Engineers at Risk and Women in Control. Her professional memberships include AMS, ASA, AWM, the Bernoulli Society, MAAA, the Polish Mathematical Society, SIAG/CST, and SIAM.