ON DVORETZKY STOCHASTIC APPROXIMATION THEOREMS[1]

BY J. H. VENTER

Potchefstroom University

1. Introduction and summary. Let H be a set and $\{T_n, n = 1, 2, \cdots\}$ a sequence of transformations of H into itself. Let X_1 and $\{U_n\}$ be random elements in H and generate the sequence $\{X_n\}$ by

$$X_{n+1} = T_n(X_n) + U_n .$$

Theorems giving conditions under which $\{X_n\}$ is "stochastically attracted" towards a given subset of H and will eventually be within or arbitrarily close to this set in an appropriate sense, are called Dvoretzky stochastic approximation theorems. The main results of this paper (Theorems 1, 2 and 3) are of this type. They generalize the work of Dvoretzky [6] and Wolfowitz [12] for the case H equal to the real line, of Derman and Sacks [4] for the case H a finite dimensional Euclidian space and Schmetterer [11] for the case H a Hilbert space.

2. Preliminaries. We assume throughout this paper that H is a real separable Hilbert space with inner product $(\cdot, \cdot)$ and norm $\|\cdot\|$.

Let $\mathcal{H}$ be the σ-field of subsets of H generated by the open sets. Let $(\Omega, \mathcal{A}, P)$ be a probability measure space; the elements of Ω are generically denoted by ω. A random element X (or $Y, Z, U, \cdots$) in H is a measurable mapping of $(\Omega, \mathcal{A})$ into $(H, \mathcal{H})$. For the theory of such random elements we refer to [9], [5], [7]. We state here a few facts needed below.

If X, Y are random elements and h a fixed element of H, then $\|X\|$, (X, Y), (h, X) are real-valued random variables in the usual sense. Denoting by E the expectation operator, if $E\|X\| < \infty$, then EX is defined by the requirement $E(h, X) = (h, EX)$ for all h in H. Similarly, if $\mathcal{B}$ is a sub-σ-field of $\mathcal{A}$, then the conditional expectation of X given $\mathcal{B}$, denoted by $E[X \mid \mathcal{B}]$, is defined by the requirement $E[(h, X) \mid \mathcal{B}] = (h, E[X \mid \mathcal{B}])$ a.s. (almost surely P), for all h in H. This conditional expectation operator has the usual properties. If Y is measurable with respect to $\mathcal{B}$, then it is true that $E[(Y, X) \mid \mathcal{B}] = (Y, E[X \mid \mathcal{B}])$ a.s. If $\mathcal{B}$ is induced by the random elements $X_1, \cdots, X_n$ then we will also write $E[X \mid X_1, \cdots, X_n]$ for $E[X \mid \mathcal{B}]$.

The transformations T_n will be more general than indicated in Section 1. For the purpose of formulating our conditions we will write $H^{(n)}$, $H^{(\infty)}$ for the n fold and denumerable cartesian products of H with itself and $H^{(n)} \times \Omega$, $H^{(\infty)} \times \Omega$ for the cartesian products of $H^{(n)}$, $H^{(\infty)}$ with Ω respectively. In order to avoid un-

Received 8 February 1966.

[1] Parts of this paper are taken from the author's Ph.D.-dissertation (1963) at the University of Chicago. This research was supported in part by Research Grant No. NSF-G21058 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

necessary repetition all transformations, functions, etc. to be introduced will be assumed measurable with respect to the appropriate σ-fields without this being explicitly stated.

We denote by D^c the complement of a set D in $\mathcal{A}$ and by I_D the indicator function of D. $K_1, K_2, \cdots$ will denote fixed positive constants. The following lemmas will be needed.

LEMMA 1. *Let* $\{b_n\}$, $\{c_n\}$, $\{d_n\}$ *be real sequences such that*

$$\sum b_n \quad \text{converges}, \qquad \sum b_n^2 < \infty \tag{1}$$

$$c_n \geqq 0 \quad \text{and} \quad \sum c_n = \infty \tag{2}$$

$$d_n \geqq 0 \quad \text{and} \quad \sum d_n < \infty. \tag{3}$$

(a) *If* $\{\xi_n\}$ *is a sequence of non-negative numbers such that, for some integer* n_0 *and for all* $n \geqq n_0$,

$$\xi_{n+1} \leqq \max\,[a, (1 + b_n)\xi_n + d_n - c_n] \tag{4a}$$

where $a > 0$, *then*

$$\limsup_{n\to\infty} \xi_n \leqq a. \tag{5}$$

(b) *If instead of* (4a) $\{\xi_n\}$ *satisfies*

$$\xi_{n+1} \leqq \max\,[a, (1 + b_n)\xi_n + d_n], \tag{4b}$$

then we can still conclude that the sequence $\{\xi_n\}$ *is bounded.*

LEMMA 2. *Let* $\{\xi_n\}$ *be a sequence of non-negative numbers such that for all n large enough*

$$\xi_{n+1} \leqq (1 - n^{-1}c_n)\xi_n + dn^{-(1+p)} \tag{6}$$

where $d > 0$ *and* $c_n \to c$ *as* $n \to \infty$. *Then*

$$\xi_n = O(n^{-p}), \qquad \text{if} \quad c > p > 0, \tag{7}$$

$$\xi_n = O(n^{-c}\log n), \qquad \text{if} \quad c = p > 0, \tag{8}$$

$$\xi_n = O(n^{-c}), \qquad \text{if} \quad p > c > 0, \tag{9}$$

as $n \to \infty$.

LEMMA 3. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability measure space,* $\{V_n\}$ *a sequence of real-valued random variables and* $\{\mathcal{B}_n\}$ *a sequence of sub-*σ*-fields of* $\mathcal{A}$ *such that* $\{V_1, \cdots, V_{n-1}\}$ *is measurable with respect to* $\mathcal{B}_n$ *for* $n > 1$. *Then, if*

$$\sum EV_n^2 < \infty \tag{10}$$

and

$$\sum E[V_n \mid \mathcal{B}_n] \quad \text{converges a.s.}, \tag{11}$$

it follows that

$$\sum V_n \quad \text{converges a.s.} \tag{12}$$

Remarks. Lemma 1 is an extended version of a result in [4] and its proof is quite similar. Lemma 2 is similar to results in [3]. Lemma 3 is a slight extension of Theorem D, p. 387 of [8]. We will not prove these lemmas here.

3. Main results.

Theorem 1. *For each integer n, let T_n be a transformation of $H^{(n)} \times \Omega$ into itself. Let $\theta \varepsilon H$. Let N be a (finite) integer-valued random variable on Ω and suppose that for each sequence $\{x_n\}$ in H and for $\omega \varepsilon \Omega_0 \varepsilon \mathfrak{A}$ with $P(\Omega_0) = 1$, we have, for $n > N(\omega)$,*

$$(13) \quad \|T_n(x_1, \cdots, x_n, \omega) - \theta\|^2 \leqq \max[\alpha, (1 + \beta_n)\|x_n - \theta\|^2 - \gamma_n]$$

where

(i) *α is a positive constant;*

(ii) *β_n is a non-negative real-valued function on $H^{(n)} \times \Omega$ such that*

$$(14) \quad \beta_n(x_1, \cdots, x_n, \omega) \leqq K_1 \quad \text{and} \quad \sum \beta_n(x_1, \cdots, x_n, \omega) < \infty$$

for all sequence $\{x_n\}$ in H and for all $\omega \varepsilon \Omega_0$;

(iii) *γ_n is a real-valued function on $H^{(n)} \times \Omega$ such that for all $\{x_n\}$ in H,*

$$(15) \quad \gamma_n(x_1, \cdots, x_n, \omega) \geqq 0 \quad \text{if} \quad n > N(\omega)$$

and if $\omega \varepsilon \Omega_0$, while, if

$$(16) \quad \sup_n \|x_n\| < \infty,$$

then

$$(17) \quad \sum \gamma_n(x_1, \cdots, x_n, \omega) = \infty.$$

Let X_1 be an arbitrary random element in H and let the sequence $\{X_n\}$ satisfy

$$(18) \quad X_{n+1}(\omega) = T_n(X_1(\omega), \cdots, X_n(\omega), \omega) + U_n(\omega)$$

where $\{U_n\}$ is a sequence of random elements satisfying the conditions

$$(19) \quad \sum E\|U_n\|^2 < \infty$$

and

$$(20) \quad \sum \|E[U_n \mid \mathfrak{B}_n]\| < \infty \quad \text{a.s.}$$

where $\{\mathfrak{B}_n\}$ is an increasing sequence of sub-σ-fields of $\mathfrak{A}$ having the properties that the random elements $\{X_1, \cdots, X_n, T_1(X_1), \cdots, T_n(X_1, \cdots, X_n)\}$ are measurable with respect to $\mathfrak{B}_n$ for $n = 2, 3, \cdots$ and that

$$(21) \quad [\omega \varepsilon \Omega: n > N(\omega)] \varepsilon \mathfrak{B}_n.$$

Then

$$(22) \quad \limsup_{n\to\infty} \|X_n - \theta\|^2 \leqq \alpha \quad \text{a.s.}$$

Proof. We reduce the theorem to an a.s. pointwise application of Lemma 1. It involves no loss of generality to take $\theta = 0$. We abbreviate $T_n(X_1(\omega), \cdots, X_n(\omega), \omega)$ to $T_n(\omega)$ and drop the ω where convenient. Define

$$A_n = [\omega \varepsilon \Omega: \|T_n(\omega)\|^2 \leqq \alpha]. \tag{23}$$

From (18), for $n > N$ and on A_n,

$$\|X_{n+1}\| \leqq \alpha^{\frac{1}{2}} + \|U_n\|. \tag{24}$$

Also from (18)

$$\|X_{n+1}\|^2 = \|T_n\|^2 + \|U_n\|^2 + 2(T_n, U_n). \tag{25}$$

From (13), for $n > N$ and on $\Omega_0 \cap A_n^c$,

$$\alpha < \|T_n\|^2 \leqq (1 + \beta_n)\|X_n\|^2 - \gamma_n. \tag{26}$$

Substituting in (25), we have for $n > N$ and on $\Omega_0 \cap A_n^c$,

$$\|X_{n+1}\|^2 \leqq (1 + \beta_n)\|X_n\|^2 - \gamma_n + \|U_n\|^2 + 2(T_n, U_n). \tag{27}$$

Define

$$\begin{aligned} V_n &= 2(T_n, U_n)\|X_n\|^{-2} \qquad \text{on} \quad A_n^c \cap [n > N] \\ &= 0 \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned} \tag{28}$$

We note that V_n is well-defined a.s. since from (26) for $n > N$ and on A_n^c,

$$\|X_n\|^2 \geqq (\alpha + \gamma_n)(1 + \beta_n)^{-1} \geqq \alpha(1 + K_1)^{-1} > 0 \tag{29}$$

where we have used (14) and (15).

Substituting (28) into (27) and taking the result together with (24), we have for $n > N$ and on Ω_0

$$\|X_{n+1}\|^2 \leqq \max [(\alpha^{\frac{1}{2}} + \|U_n\|)^2, (1 + \beta_n + V_n)\|X_n\|^2 - \gamma_n + \|U_n\|^2]. \tag{30}$$

Now (19) implies that

$$\sum \|U_n\|^2 < \infty \quad \text{a.s.} \tag{31}$$

We also show that

$$\sum V_n \quad \text{converges} \quad \text{a.s.} \tag{32}$$

From (28) and Schwarz's inequality

$$|V_n|^2 \leqq 4 \|U_n\|^2 \|T_n\|^2 \|X_n\|^{-4} I_{A_n^c} I_{[n>N]}.$$

From (26) and (15), (29) and (14)

$$\|T_n\|^2 \|X_n\|^{-4} I_{A_n^c} I_{[n>N]} \leqq (1 + K_1)^2 \alpha^{-1} = K_2. \tag{33}$$

Hence $E|V_n|^2 \leqq 4K_2 E\|U_n\|^2$ and by (19)

$$\sum E|V_n|^2 < \infty. \tag{34}$$

Further, substituting for U_{n-1} from (18) into (28) we see that V_{n-1} is a function of X_n, $T_{n-1}(X_1, \cdots, X_{n-1})$ and $I_{[n-1>N]}$ and hence measurable with respect to $\mathcal{B}_n$. Varying the index n, it follows that $V_1, \cdots, V_{n-1}$ are measurable with respect to $\mathcal{B}_n$. From (28), (21), the Schwarz inequality and (33),

$$\begin{aligned} |E[V_n \mid \mathcal{B}_n]| &= 2\,|(T_n, E[U_n \mid \mathcal{B}_n])|\, \|X_n\|^{-2} I_{A_n^c} I_{[n>N]} \\ &\leqq 2\,\|T_n\|\,\|X_n\|^{-2} I_{A_n^c} I_{[n>N]}\,\|E[U_n \mid \mathcal{B}_n]\| \\ &\leqq 2K_2^{\frac{1}{2}}\,\|E[U_n \mid \mathcal{B}_n]\|. \end{aligned}$$

From (20) it follows that $\sum E[V_n \mid \mathcal{B}_n]$ converges a.s.. Applying Lemma 3 we conclude that (32) holds. It also follows from (34) that

$$\sum V_n^2 < \infty \quad \text{a.s.}. \tag{35}$$

Now fix a point $\omega \varepsilon \Omega_0$ for which (31), (32) and (35) hold simultaneously. Let $\epsilon > 0$. Since $\|U_n(\omega)\| \to 0$ as $n \to \infty$, there exists an integer $N_1(\omega) \geqq N(\omega)$ such that for all $n > N_1(\omega)$, $\|U_n(\omega)\|^2 + 2\alpha^{\frac{1}{2}}\|U_n(\omega)\| < \epsilon$. From (30), for $n > N_1(\omega)$,

$$\|X_{n+1}(\omega)\|^2 \leqq \max\,[\alpha + \epsilon,\ (1 + \beta_n(\omega) + V_n(\omega))\|X_n(\omega)\|^2 - \gamma_n(\omega) + \|U_n(\omega)\|^2]. \tag{36}$$

Applying Part (b) of Lemma 1 to (36), we find that the sequence $\{\|X_n(\omega)\|^2\}$ is bounded and by (17) then

$$\sum \gamma_n(\omega) \equiv \sum \gamma_n(X_1(\omega), \cdots, X_n(\omega), \omega) = \infty.$$

Applying Part (a) of Lemma 1 to (36), it follows that

$$\limsup_{n\to\infty} \|X_n(\omega)\|^2 \leqq \alpha + \epsilon$$

and since ϵ is arbitrary and the set of ω points under consideration has probability one, the theorem follows.

REMARKS. (i) If (20) is changed to $\sum \|E[U_n I_{[n>N]} \mid \mathcal{B}_n]\| < \infty$ a.s. then (21) can be dropped.

(ii) The theorem gives conditions under which $\{X_n\}$ is stochastically attracted towards the sphere with centre θ and radius $\alpha^{\frac{1}{2}}$ and will eventually a.s. be within or arbitrarily close to this sphere. Under somewhat stronger conditions we also prove that $\{X_n\}$ will eventually be within or close to this sphere in a mean square sense.

THEOREM 2. *Consider the set-up in Theorem* 1 *and let the conditions be strengthened as follows: N is a fixed finite integer and β_n a fixed sequence of non-negative numbers such that*

$$\sum \beta_n < \infty. \tag{40}$$

$\{U_n\}$ *satisfies* (19) *and instead of* (20),

$$\sum (E\,\|E[U_n \mid \mathcal{B}_n]\|^2)^{\frac{1}{2}} < \infty. \tag{41}$$

Other conditions remain unchanged except in so far as they are changed by the new conditions just introduced. Thus e.g. (21) *becomes redundant. If*

$$E\,\|X_N\|^2 < \infty \tag{42}$$

then

$$\limsup_{n\to\infty} E\,\|X_n - \theta\|^2 \leqq \alpha. \tag{43}$$

PROOF. (41) does imply (20), for $E\,\|E[U_n \mid \mathcal{B}_n]\| \leqq (E\,\|E[U_n \mid \mathcal{B}_n]\|^2)^{\frac{1}{2}}$; hence (41) implies

$$\sum E\,\|E[U_n \mid \mathcal{B}_n]\| < \infty \tag{44}$$

which implies (20). The conclusion of Theorem 1 therefore holds in the present case. We also note that

$$E\,\|X_n\|^2 < \infty \tag{45}$$

for each n, since, from (18), (13) and (15) we have

$$\begin{aligned}\|X_{n+1}\|^2 &\leqq 2\,\|T_n\|^2 + 2\,\|U_n\|^2\\ &\leqq 2(1+\beta_n)\|X_n\|^2 + 2\alpha + 2\,\|U_n\|^2,\end{aligned}$$

a.s., and hence taking expectations, noting (42) and applying induction on n, (45) follows.

Now, there is no loss of generality in supposing that $\theta = 0$ and $N = 1$. We will write $r^+ = \max\,[0, r]$ for any real number r. Let $a > \alpha^{\frac{1}{2}}$. Then $\|X_n\| \leqq a + (\|X_n\| - a)^+$. Hence writing

$$Y_n = (\|X_n\| - a)^+ \tag{46}$$

we have

$$\|X_n\|^2 \leqq a^2 + 2aY_n + Y_n^2. \tag{47}$$

We will show that

$$EY_n^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{48}$$

Since $EY_n \leqq (EY_n^2)^{\frac{1}{2}}$, (47) and (48) will imply $\limsup_{n\to\infty} E\,\|X_n\|^2 \leqq a^2$ and since a^2 is arbitrarily larger than α, the theorem will follow.

Our first step in proving (48) is to obtain a bound for Y_{n+1}^2 in terms of Y_n^2. Let

$$\begin{aligned} W_n &= T_n && \text{if} \quad \|T_n\| \leqq a\\ &= aT_n\,\|T_n\|^{-1} && \text{if} \quad \|T_n\| > a. \end{aligned} \tag{49}$$

Then

$$\|T_n - W_n\| = (\|T_n\| - a)^+ \tag{50}$$

and $\|W_n\| \leqq a$. Hence, from (18),

$$\begin{aligned}\|X_{n+1}\| - a &\leqq \|X_{n+1}\| - \|W_n\| \\ &\leqq \|X_{n+1} - W_n\| \\ &= \|T_n - W_n + U_n\|.\end{aligned}$$

Using (50),

$$Y_{n+1}^2 \leqq [(\|T_n\| - a)^+]^2 + \|U_n\|^2 + 2(T_n - W_n, U_n). \tag{51}$$

From (13) and (15), $\|T_n\|^2 \leqq \max[\alpha, (1 + \beta_n)\|X_n\|^2]$ a.s.. On the set $\|T_n\|^2 \leqq \alpha$, we have

$$(\|T_n\| - a)^+ = 0 \tag{52}$$

and on the set $\|T_n\|^2 > \alpha$, $\|T_n\| \leqq (1 + \frac{1}{2}\beta_n)\|X_n\|$ a.s. and hence

$$(\|T_n\| - a)^+ \leqq (1 + \tfrac{1}{2}\beta_n)Y_n + \tfrac{1}{2}a\beta_n \quad \text{a.s.}. \tag{53}$$

Using the inequality $(p + q)^2 \leqq (1 + q)p^2 + q(1 + q)$ which holds for $q \geqq 0$, (53) yields

$$[(\|T_n\| - a)^+]^2 \leqq (1 + \beta_n')Y_n^2 + \delta_n' \quad \text{a.s.}, \tag{54}$$

where $\beta_n' = (1 + \frac{1}{2}a\beta_n)(1 + \frac{1}{2}\beta_n)^2 - 1$, $\delta_n' = \frac{1}{2}a\beta_n(1 + \frac{1}{2}a\beta_n)$ so that

$$\beta_n', \delta_n' \geqq 0 \quad \text{and} \quad \sum \beta_n' < \infty, \qquad \sum \delta_n' < \infty. \tag{55}$$

Substituting (54) into (51) we obtain the required bound, viz.

$$Y_{n+1}^2 \leqq (1 + \beta_n')Y_n^2 + \delta_n' + \|U_n\|^2 + 2(T_n - W_n, U_n), \quad \text{a.s.}. \tag{56}$$

Now, let b be a positive number and M an integer to be specified further below. Define sets B_n in Ω by

$$B_n = [\inf_{M \leqq j \leqq n} Y_j > b], \qquad n \geqq M. \tag{57}$$

Since I_{B_n} and $T_n - W_n$ are measurable with respect to $\mathcal{B}_n$, we get from the Schwarz inequality, (50) and (53)

$$\begin{aligned}|EI_{B_n}(T_n - W_n, U_n)| &= |E\{I_{B_n}E[(T_n - W_n, U_n) \mid \mathcal{B}_n]\}| \\ &= |E\{I_{B_n}(T_n - W_n, E[U_n \mid \mathcal{B}_n])\}| \\ &\leqq E\{I_{B_n}\|T_n - W_n\| \, \|E[U_n \mid \mathcal{B}_n]\|\} \\ &\leqq E\{I_{B_n}(\|T_n\| - a)^+ \|E[U_n \mid \mathcal{B}_n]\|\} \\ &\leqq (1 + \tfrac{1}{2}\beta_n)EI_{B_n}Y_n \|E[U_n \mid \mathcal{B}_n]\| \\ &\quad + \tfrac{1}{2}a\beta_n E\|E[U_n \mid \mathcal{B}_n]\|.\end{aligned}$$

Using the Schwarz inequality in the first expectation here together with $(EI_{B_n}Y_n^2)^{\frac{1}{2}} \leqq 1 + EI_{B_n}Y_n^2$, we get

$$|EI_{B_n}(T_n - W_n, U_n)| \leqq \beta_n'' EI_{B_n}Y_n^2 + \delta_n'' \tag{58}$$

where

$$\beta_n'' = (1 + \tfrac{1}{2}\beta_n)(E\,\|E[U_n \mid \mathcal{B}_n]\|^2)^{\frac{1}{2}}$$
$$\delta_n'' = \beta_n'' + \tfrac{1}{2}a\beta_n E\,\|E[U_n \mid \mathcal{B}_n]\|$$

so that

$$(59) \qquad \beta_n'', \delta_n'' \geqq 0 \quad \text{and} \quad \sum \beta_n'' < \infty, \qquad \sum \delta_n'' < \infty$$

in view of (41) and (44).

Hence, taking expectations over B_n on both sides of (56), we have

$$(60) \qquad EI_{B_n}Y_{n+1}^2 \leqq (1 + b_n)EI_{B_n}Y_n^2 + d_n$$
$$(61) \qquad \leqq (1 + b_n)EI_{B_{n-1}}Y_n^2 + d_n, \qquad n > M,$$

where

$$b_n = \beta_n' + \beta_n''$$
$$d_n = \delta_n' + \delta_n'' + E\,\|U_n\|^2$$

so that

$$(62) \qquad \sum b_n < \infty, \qquad \sum d_n < \infty .$$

Iterating (61) back to $n = M + 1$ and using (60) for $n = M$, we get

$$(63) \qquad EI_{B_n}Y_{n+1}^2 \leqq [\textstyle\prod_{j \geqq M}(1 + b_j)][EI_{[Y_M > b]}Y_M^2 + \sum_{j \geqq M} d_j].$$

Now we turn to B_n^c. Define $C_M = [Y_M \leqq b]$ and for $n > M$

$$C_n = [Y_M > b,\, Y_{M+1} > b, \cdots, Y_{n-1} > b,\, Y_n \leqq b].$$

Then, agreeing that $B_{M-1}^c = \emptyset$, the empty set, for $n \geqq M$

$$(64) \qquad B_n^c = C_M + C_{M+1} + \cdots + C_n = B_{n-1}^c + C_n .$$

Also, let $D_M = [0 < Y_M \leqq b]$ and for $n > M$

$$(65) \qquad D_n = [Y_M > b,\, Y_{M+1} > b, \cdots, Y_{n-1} > b,\, 0 < Y_n \leqq b].$$

Then

$$C_n = D_n + [Y_M > b,\, Y_{M+1} > b, \cdots, Y_{n-1} > b,\, Y_n = 0].$$

Hence

$$(66) \qquad EI_{C_n}Y_n^2 = EI_{D_n}Y_n^2 \leqq b^2P(D_n).$$

Taking expectation over B_n^c on both sides of (56) and using (58) with B_n replaced by B_n^c (which does not invalidate it), we get

$$EI_{B_n^c}Y_{n+1}^2 \leqq (1 + b_n)EI_{B_n^c}Y_n^2 + d_n$$
$$\leqq (1 + b_n)[EI_{B_{n-1}^c}Y_n^2 + b^2P(D_n)] + d_n ,$$

having used (64) and (66). Iterating this inequality back to $n = M$ we get

$$(67) \quad EI_{B_n^c}Y_{n+1}^2 \leqq [\prod_{j\geqq M}(1+b_j)][b^2P(D_M) + \cdots + b^2P(D_n) + \sum_{j\geqq M} d_j].$$

Noting that the sets $D_M, \cdots, D_n$ are disjoint and that $D_M + \cdots + D_n \subset [\inf_{M\leqq j\leqq n} Y_j > 0] \equiv F_n$, say, it follows from taking (63) and (67) together that

$$(68) \quad EY_{n+1}^2 \leqq [\prod_{j\geqq M}(1+b_j)][EI_{[Y_M>b]}Y_M^2 + b^2P(F_n) + 2\sum_{j\geqq M} d_j].$$

Now let $\epsilon > 0$. Then choose and fix M so large that $\prod_{j\geqq M}(1+b_j) < 2$, $\sum_{j\geqq M} d_j < \epsilon$. This is possible in view of (62). Also, from (45) and (46) it follows that Y_M^2 is integrable so that b can be chosen and fixed large enough so that $EI_{[Y_M>b]}Y_M^2 < \epsilon$. Finally, from Theorem 1, $P(F_n) \to 0$ as $n \to \infty$. Hence there exists an integer M_1 such that for all $n > M_1$ we have $b^2P(F_n) \leqq \epsilon$. Thus, from (68), for $n > M_1$,

$$EY_{n+1}^2 \leqq 2[\epsilon + \epsilon + 2\epsilon] = 8\epsilon.$$

Since ϵ is arbitrary (48) and the theorem is proved.

REMARKS. If α can be chosen arbitrarily small, Theorems 1 and 2 give conditions for a.s. and mean square convergence of $\{X_n\}$ to θ. Most applications of these results are of this type. In addition the transformations $\{T_n\}$ are often of the following special type:

$$(69) \quad T_n(x_1, \cdots, x_n, \omega) = x_n - S_n(x_n, \omega)$$

where S_n is a transformation of $H \times \Omega$ into H. Our next theorem gives conditions under which $\{X_n\}$ will converge to θ in this case.

THEOREM 3. *Let T_n be specified by (69) and suppose that S_n satisfies the following conditions. For each $x \varepsilon H$ and $\omega \varepsilon \Omega_0 \varepsilon \mathfrak{A}$ where $P(\Omega_0) = 1$,*

$$(70) \quad \|S_n(x, \omega)\|^2 \leqq \beta_n \|x - \theta\|^2 + \delta_n$$

for all n, where $\{\beta_n\}$, $\{\delta_n\}$ are non-negative real sequences such that

$$(71) \quad \sum \beta_n < \infty, \qquad \sum \delta_n < \infty.$$

Also, for each $\epsilon > 0$, define

$$(72) \quad c_n(\epsilon, \omega) = \inf_{\epsilon\leqq\|x-\theta\|\leqq\epsilon^{-1}} 2(x - \theta, S_n(x, \omega))$$

and suppose that there is a finite integer-valued random variable N_ϵ such that for all $n > N_\epsilon(\omega)$ and for all $\omega \varepsilon \Omega_0$,

$$(73) \quad c_n(\epsilon, \omega) \geqq \delta_n$$

while also

$$(74) \quad \sum c_n(\epsilon, \omega) = \infty.$$

Define $\{X_n\}$ by (18) with $\{U_n\}$ as given in Theorem 1 and suppose also that (21) holds with N replaced by N_ϵ. Then

$$(75) \quad X_n \to \theta \quad \text{a.s. as} \quad n \to \infty.$$

In addition, if N_ϵ is degenerate and $\{U_n\}$ is as given in Theorem 2, then

$$E\,\|X_n - \theta\|^2 \to 0 \quad as \quad n \to \infty. \tag{76}$$

PROOF. As before we may take $\theta = 0$. Let $\alpha > 0$. If $\|x\|^2 \leqq \alpha$, then

$$\begin{aligned} \|T_n(x, \omega)\|^2 &\leqq \|x - S_n(x, \omega)\|^2 \\ &\leqq 2\,\|x\|^2 + 2\,\|S_n(x, \omega)\|^2 \\ &\leqq 2(\alpha + \alpha\beta_n + \delta_n) \\ &\leqq 4\alpha \end{aligned} \tag{77}$$

for all $n > n_0$, a certain integer. For $\omega\ \varepsilon\ \Omega_0$ we also have

$$\begin{aligned} \|T_n(x, \omega)\|^2 &= \|x\|^2 + \|S_n(x, \omega)\|^2 - 2(x, S_n(x, \omega)) \\ &\leqq (1 + \beta_n)\|x\|^2 + \delta_n - 2(x, S_n(x, \omega)). \end{aligned} \tag{78}$$

Define

$$\begin{aligned} \gamma_n(x, \omega) &= -\delta_n + 2(x, S_n(x, \omega)) \qquad \text{if} \quad \|x\|^2 > \alpha \\ &= 1 \qquad\qquad\qquad\qquad\qquad\quad \text{if} \quad \|x\|^2 \leqq \alpha. \end{aligned} \tag{79}$$

Substituting into (78) and taking the result together with (77) we have, for all $n > n_0$ and $\omega\ \varepsilon\ \Omega_0$

$$\|T_n(x, \omega)\|^2 \leqq \max\,[4\alpha, (1 + \beta_n)\|x\|^2 - \gamma_n]. \tag{80}$$

Now, let $\{x_n\}$ be any sequence in H such that $\sup_n \|x_n\| = K_4 < \infty$. Put $\epsilon = \min\,[\alpha^{\frac{1}{2}}, K_4^{-1}]$. Then, for $n > N_\epsilon(\omega)$, from (79), (72) and (73),

$$\gamma_n(x_n, \omega) \geqq \min\,[1, -\delta_n + c_n(\epsilon, \omega)] \geqq 0, \tag{81}$$

while from (79), (72), (74) and (71)

$$\sum \gamma_n(x_n, \omega) \geqq \sum \min\,[1, -\delta_n + c_n(\epsilon, \omega)] = \infty.$$

Letting $N = N_\epsilon + n_0$, then (80) and (81) are certainly satisfied for all $n > N$ and we also have

$$[n > N] \equiv [n - n_0 > N_\epsilon]\ \varepsilon\ \mathcal{B}_{n-n_0} \subset \mathcal{B}_n\,.$$

Hence the conditions of Theorem 1 are satisfied completely and we conclude that $\limsup_{n\to\infty} \|X_n - \theta\|^2 \leqq 4\alpha$ a.s. and since α is arbitrary, (75) follows. (76) follows from a similar application of Theorem 2.

REMARKS. More explicit results on the order of magnitude of $E\,\|X_n - \theta\|^2$ can be obtained under stronger conditions than those specified in Theorem 3. For the sake of completeness we indicate some of these. If (72)–(74) is replaced by

$$2(x - \theta, S_n(x, \omega)) \geqq c_n\,\|x - \theta\|^2 \tag{82}$$

for each n and all $x\ \varepsilon\ H$, $\omega\ \varepsilon\ \Omega_0$, where $\{c_n\}$ is a sequence of constants such that

$$c_n \geqq 0 \quad \text{and} \quad \sum c_n = \infty, \tag{83}$$

while (41) is strengthened to

$$E[U_n \mid \mathcal{B}_n] = 0 \quad \text{a.s.}, \tag{84}$$

then it follows readily from (18), (69), (70), (82) and (84) that

$$E\|X_{n+1} - \theta\|^2 \leqq (1 + \beta_n - c_n)E\|X_n - \theta\|^2 + \delta_n + E\|U_n\|^2. \tag{85}$$

Iteration of this inequality yields a bound for $E\|X_{n+1} - \theta\|^2$ in terms of $\{\beta_n\}$, $\{c_n\}$, $\{\delta_n\}$ and $E\|U_n\|^2$. We mention some asymptotic results in this connection. Suppose that $c_n - \beta_n \sim cn^{-1}$ and $\delta_n + E\|U_n\|^2 \sim O(n^{-(1+p)})$ as $n \to \infty$, then

$$E\|X_n - \theta\|^2 = O(n^{-p}), \qquad \text{if } c > p > 0,$$

$$E\|X_n - \theta\|^2 = O(n^{-p} \log n), \qquad \text{if } c = p > 0,$$

$$E\|X_n - \theta\|^2 = O(n^{-c}), \qquad \text{if } p > c > 0.$$

These results follow immediately from Lemma 2 and (85). We refer to [11] for further results under these assumptions.

4. Concluding remarks. Theorem 3 is at the same time sufficiently general and simple to make applications to the standard stochastic approximation procedures such as the Robbins-Monro [6], [1] and Kiefer-Wolfowitz [6] procedures and especially their multi-dimensional extensions [2], [10] routine. These applications do not use the generality of our theorems; applications to more complicated procedures requiring this generality will be given in a forthcoming paper by the author.

REFERENCES

[1] Blum, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 382–386.

[2] Blum, J. R. (1954). Multidimensional stochastic approximation methods. *Ann. Math. Statist.* **25** 737–744.

[3] Chung, K. L. (1954). On a stochastic approximation method. *Ann. Math. Statist.* **25** 463–483.

[4] Derman, C. and Sacks, J. (1959). On Dvoretzky's stochastic approximation theorem. *Ann. Math. Statist.* **30** 601–605.

[5] Driml, M. and Hans, O. (1959). Conditional expectations for generalized random variables. *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions and Random Processes.* 123–143.

[6] Dvoretzky, A. (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–55.

[7] Grenander, U. (1963). *Probabilities on Algebraic Structures.* Almqvist and Wiksell, Stockholm.

[8] Loève, M. (1960). *Probability Theory* (2nd Ed.) Van Nostrand, Princeton.

[9] Mourier, E. (1952–53). Élèments alèatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* **13** 161–244.

[10] Sacks, J. (1958). Asymptotic distributions of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373–405.

[11] Schmetterer, L. (1958). Sur l'itération stochastique. *Le Calcul des Probabilités et ses Applications.* **87** 55–63.

[12] Wolfowitz, J. (1956). On stochastic approximation methods. *Ann. Math. Statist.* **27** 1151–1156.