The Annals of Statistics
1973, Vol. 1, No. 6, 1158–1174

TRANSFORMATION OF OBSERVATIONS IN STOCHASTIC APPROXIMATION[1]

BY SAMI NAGUIB ABDELHAMID

Alexandria University

The general stochastic approximation procedure:

$$X_{n+1} = X_n - a_n c_n^{-1} h(Y_n), \qquad n = 1, 2, \cdots$$

is considered, where h is a Borel measurable transformation on the random observations Y_n. Under some mild requirements on h and on the error random variables, the asymptotic properties, the a.s. convergence and the asymptotic normality are studied. The analysis is confined to the *case* where the error random variables are (conditionally) distributed according to a distribution function G which is symmetric around 0 and admits a density g. The optimal choices of the design sequences a_n and c_n as well as the transformation h are studied. The optimal transformation turned out to be equal to $-C(g'/g)$ (a.e. with respect to G) for a $C > 0$ and it is the transformation which minimizes the second moment of the asymptotic distribution of $n^\beta(X_n - \theta)$. The Robbins–Monro and the Kiefer–Wolfowitz situations are emphasized as special cases. With the optimal transformation, the new proposed generalized procedure is shown to yield asymptotically efficient estimators.

1. Introduction and summary. Consider the following general stochastic approximation procedure:

$$X_{n+1} = X_n - a_n c_n^{-1} h(Y_n), \qquad n = 1, 2, \cdots \tag{1}$$

where X_1 is an arbitrary random variable, Y_n are random observations, a_n and c_n are positive numbers and h is a Borel measurable transformation. With the choice $h =$ the identity, (1) includes both the Robbins–Monro (1951) procedure (RM) and the Kiefer–Wolfowitz (1952) procedure (KW). Fabian (1960, 1964) considered (1) with $h =$ sign; we shall call (1) with the choice $h =$ sign procedure (F).

In this paper we establish the asymptotic properties, the a.s. convergence and the asymptotic normality of this proposed generalized procedure. Then we characterize the optimal transformation h. Our analysis is general enough to include both the RM situation and the KW situation as special cases. To be more specific let f be a Borel measurable function defined on the real line. The exact analytic form of f may be unknown, but it is assumed that f belongs to a rather general family of functions. The only available information about f is that at any level

Received November 1971; revised April 1973.

[1] Parts of this paper are extracted from the author's Ph. D. thesis (1971) at Michigan State University.

Key words and phrases. Transformation of observations, optimal transformation, stochastic approximation procedure, optimal procedure, Robbins–Monro situation, asymptotically efficient estimators, Kiefer–Wolfowitz situation.

x, we can observe $f(x)$ subject to a random error; that is we can obtain unbiased observation of $f(x)$. In the RM situation, the goal is to estimate sequentially the unknown root of the equation $f(\theta) = 0$ (or, generally $f_n(\theta) = 0$, where f_n are Borel measurable functions); and in the KW situation the goal is to estimate sequentially the unknown point of minimum (or maximum), θ, of a function f.

The results on the RM situation were obtained independently by the author and by Anbar (1971) and will not be repeated here. The appropriate details can be found either in Anbar (1973) or in Abdelhamid (1971). As a special case of our analysis we shall emphasize the KW situation.

Let us denote

$$\mathscr{Z}_n = [X_1, X_2, \cdots, X_n]\,, \qquad M_n(\mathscr{Z}_n) = E_{\mathscr{Z}_n}[Y_n]\,, \qquad V_n = Y_n - M_n(\mathscr{Z}_n)\,.$$

We shall confine our analysis to the case where the random variables V_n are conditionally (given $\mathscr{Z}_n$) distributed according to distribution function G which is symmetric around 0 and admits a density function g. This requirement of symmetry is natural in the KW situation (see Section 3.4 below).

We shall state conditions on h under which both the almost sure convergence to θ and the asymptotic normality are preserved. Within the class $\mathscr{C}$ of such h's we consider the optimal transformation which minimizes the second moment of the asymptotic distribution of $n^\beta(X_n - \theta)$. This will be shown to be equivalent to finding $h \in \mathscr{C}$ which maximizes $H(h) = [d/dt \int h(t + v)G(dv)]_{t=0}$. This leads, under some regularity conditions, to $h = -(g'/g)$ (or any positive constant multiple of $-(g'/g)$).

The surprising fact is that with such an optimal h the stochastic approximation procedure is not only optimal within the class of stochastic approximation procedures considered, but also X_n is an asymptotically efficient estimator of θ in the sense that the variance of the asymptotic distribution of $n^\beta(X_n - \theta)$ corresponds to the Cramér–Rao lower bound for the variance of unbiased estimator of θ, based on the first n observations.

Knowing the optimal transformation, the KW procedure, as well as the RM procedure, is optimal if and only if the error random variables are normally distributed. As for procedure (F), we show it is optimal if and only if the error random variables have a double exponential distribution.

One of the regularity conditions is $0 < I(g) = \int (g'(v)/g(v))^2 G(dv) < \infty$; if it is not satisfied, it can be shown (cf. Abdelhamid (1971), Section 5.16) that one can design transformations which yield improved procedures.

Here we only note that our results would make it possible to compare different transformations and to study the asymptotic relative efficiency of procedure (F) relative to the optimal procedure. It may be of interest to report that the asymptotic relative efficiency of procedure (F) relative to the optimal procedure is the same as that of the sequential sign test relative to the sequential probability ratio test (cf. Groeneveld (1971)). For more details we refer the reader to Abdelhamid (1971).

Finally we give some examples of new optimal procedures.

2. Basic assumptions and notations. All random variables are supposed to be defined on a probability space $(\Omega, \mathcal{A}, P)$. Relations between random variables, including convergence, are meant with probability one. E denotes the expectation and E_T the conditional expectation given a random vector T. The real line is denoted by R and the indicator function of a set Λ by χ_Λ.

We shall write $N(\mu_1, \mu_2^2)$ to denote a normal random variable with mean $= \mu_1$ and variance $= \mu_2^2$ and we also write $T_n \to_{\mathcal{L}} \xi$ if T_n is asymptotically ξ-distributed.

For an easy reference, the KW situation is described below.

2.1 *Kiefer–Wolfowitz* (KW) *situation.* We assume that f is a Borel measurable function on R satisfying

$$(1) \qquad \sup_{-k<x-\theta<(1/k)} \bar{D}f(x) < 0 , \qquad \inf_{(1/k)<x-\theta<k} \underline{D}f(x) > 0 ,$$

for an unknown number θ, and every natural number k; where $\underline{D}f(x)$, and $\bar{D}f(x)$ denote the lower and upper derivative, respectively, of f at x. Furthermore, there exist constants A, B such that

$$(2) \qquad |f(x+1) - f(x)| < A|x-\theta| + B \qquad \text{for all} \quad x \in R .$$

The relation (1.1) holds with positive a_n, c_n satisfying

$$(3) \qquad c_n \to 0 , \qquad \sum_{n=1}^\infty a_n = \infty , \qquad \sum_{n=1}^\infty a_n^2 c_n^{-2} < \infty ;$$

and the random variables Y_n satisfy

$$(4) \qquad M_n(\mathcal{X}_n) = f(X_n + c_n) - f(X_n - c_n) ,$$

and

$$(5) \qquad E_{\mathcal{X}_n}[V_n^2] \leqq \sigma^2$$

for a number σ and every natural number n.

3. Almost sure convergence of the modified stochastic approximation procedures. To establish the a.s. convergence of the general procedure (1.1), we give the following theorem which can be proved under a somewhat weaker set of conditions than those of Theorem 1 in Burkholder (1956). By applying Lemma 10 in Dubins and Freedman (1965) (see also Robbins and Siegmund (1971)), the proof can be carried out in steps similar to those in Theorem 1 of Blum (1954), and we omit the proof here. This slightly more general theorem may be considered as a sharpening of the basic results in Blum (1954), and in Burkholder (1956).

3.1 *Almost Sure Convergence Theorem. Let* $\theta \in R$, *and* τ_n *be a nonnegative number sequence. Let* (1.1) *hold and* $\tilde{M}_n$ *be Borel measurable functions*; $\tilde{M}_n(X_n) = E_{\mathcal{X}_n}[h(Y_n)]$. *Suppose that*

(1) *if* $\varepsilon > 0$, $|x-\theta| > \varepsilon$ *and* $n > n_0(\varepsilon)$ *then* $(x-\theta)\tilde{M}_n(x) > 0$;

(2) *if* $0 < \delta_1 < \delta_2 < \infty$, *then* $\sum_{n=1}^\infty a_n c_n^{-1}[\inf_{\delta_1 \leqq |x-\theta| \leqq \delta_2} |\tilde{M}_n(x)|] = \infty$;

(3) *there exist positive constants* H_0 *and* H_1 *such that*

$$|\tilde{M}_n(x)| \leqq H_0 + H_1 \tau_n |x - \theta|$$

for all $x \in R$ *and* $n = 1, 2, \cdots$;

(4) $a_n c_n^{-1} \tau_n \to 0$ *as* $n \to \infty$ *if* τ_n *is an unbounded sequence*; *otherwise* $a_n c_n^{-1} \to 0$ *as* $n \to \infty$;

(5) $\sum_{n=1}^{\infty} a_n^2 c_n^{-2} E_{\mathscr{F}_n}[h(Y_n) - \tilde{M}_n(X_n)]^2 < \infty$.

Then

$$X_n \to \theta .$$

3.2 REMARK. Theorem 3.1 still holds if a_n, c_n are positive Borel measurable functions of $X_1, Y_1, Y_2, \cdots, Y_{n-1}$, τ_n is a nonnegative Borel measurable function of $X_1, Y_1, \cdots, Y_{n-1}$ with conditions on a_n, c_n, and τ_n holds for every sequence $X_1, Y_1, Y_2, \cdots$. In fact it is enough to assume that (3.1.2)—(3.1.4) hold for every sequence $X_1, Y_1, Y_2, \cdots$, such that $\sup |X_n| < \infty$.

3.3 REMARK. Given h, the a.s. convergence of the procedure (1.1) holds if the conditions of Theorem 3.1 are satisfied irrespective of whether they are also satisfied for h equal to the identity. But since we are interested in the optimal choice of h it seems useful to investigate conditions on h, which guarantee that conditions of Theorem 3.1 are satisfied with this h if they are satisfied for $h =$ the identity.

Henceforth the following two assumptions will be assumed to hold.

3.4 ASSUMPTION. In the general procedure (1.1), let

$$Y_n = M_n(X_n) + V_n$$

where V_n are random variables conditionally (given $\mathscr{F}_n$) distributed according to a distribution function G which is symmetric around 0 and admits a density g. The functions M_n are Borel measurable.

Here, we may remark that the requirement of symmetry is natural in the KW situation where Y_n is an unbiased estimator of $[f(X_n + c_n) - f(X_n - c_n)]$. The requirement of symmetry is then satisfied if the errors in estimating $f(X_n + c_n)$ and $f(X_n - c_n)$ are independent and identically distributed.

3.5 ASSUMPTION. We assume that h is an odd Borel measurable transformation defined on R and nonnegative on $[0, \infty)$. In addition we assume that $\Psi(t) = \int h(t + v)g(v)\, dv = \int h(v)g(v - t)\, dv$ exists for all $t \in R$.

3.6 LEMMA. *Assume that*

(1) (i) $\liminf_{t \downarrow 0} t^{-1}\Psi(t) > 0$,

and either

(ii) h *is non-decreasing*;

or

(iii) g *is continuous and non-increasing on* $[0, \infty)$, *and* h *is bounded and continuous; furthermore* $h(v) > 0$ *for all* $v > 0$.

Then (3.1.1) *and* (3.1.2) *hold for* h *if* (3.1.1), (3.1.2), *and* (3.1.3) *hold for the identity transformation.*

PROOF. From (i) we obtain, for some positive numbers Δ and ρ_0, that

$$t^{-1}\Psi(t) \geqq \rho_0 \qquad \text{for all} \quad 0 < t < \Delta . \tag{2}$$

If h is non-decreasing then so is Ψ, and thus (2) implies that $\inf\{\Psi(t), t > t_0\} > 0$ for every $t_0 > 0$. Therefore (i) and (ii) imply that if $0 < T_0 < T_1 < \infty$, then

$$\inf\{\Psi(t); t \in [T_0, T_1]\} > 0 . \tag{3}$$

Now suppose (iii) holds; we shall prove that (3) holds in this case too. Since h is odd and g is symmetric, $\Psi(t)$ can be written in the form:

$$\Psi(t) = \int_0^\infty h(v)[g(v - t) - g(v + t)]\, dv , \qquad t \in R . \tag{4}$$

The integrand is nonnegative for $t \geqq 0$, since $h(v) \geqq 0$ for $v \geqq 0$, by Assumption 3.5, and $g(v - t) - g(v + t) \geqq 0$ for $v \geqq 0$ and every $t \geqq 0$. The latter is obvious from (iii) if $0 \leqq t \leqq v$; if $0 \leqq v < t$ then $g(v - t) - g(v + t) = g(t - v) - g(v + t)$, which is again nonnegative by (iii). In particular $\Psi(t) \geqq 0$ for $t \geqq 0$. But suppose $\Psi(t) = 0$ for some $t > 0$. Then, since $h(v) > 0$ for all $v > 0$, $g(v - t) - g(v + t) = 0$ for almost all (Lebesgue) $v \geqq 0$. The function $F(v) = g(v - t) - g(v + t)$, $v \geqq 0$, is then continuous and thus identically zero. Moreover g is periodic with period $2t$; but since g is non-increasing for $v > 0$, then $g =$ constant on R. This is a contradiction to the fact that g is a density function. Hence $\Psi(t) > 0$ for all $t > 0$. Furthermore, since h is bounded and continuous, then by the dominated convergence theorem Ψ is continuous on $[T_0, T_1]$, $0 < T_0 < T_1 < \infty$. But $[T_0, T_1]$ is compact; then Ψ achieves its minimum on $[T_0, T_1]$ and thus (3) holds.

Since Ψ is odd, (3) implies $\Psi(t)$ sign $(t) \geqq 0$. (3.1.1), (3.1.2) and (3.1.3) hold for M_n since (3.1.1)—(3.1.3) hold when $h =$ the identity. $\tilde{M}_n(X_n) = \Psi(M_n(X_n))$ and thus (3.1.1) for M_n implies (3.1.1) for $\tilde{M}_n$. Further (3) and (2) imply that for every $T > 0$ there is an $\eta > 0$ such that

$$|\Psi(t)| \geqq \eta|t| \qquad \text{for all} \quad |t| < T . \tag{5}$$

Suppose $0 < \delta_1 < \delta_2$. Then using (3.1.3) for M_n we obtain $|M_n(x)| < T$ for some $T > 0$ and all $|x| \leqq \delta_1$. Thus $|\tilde{M}_n(x)| = |\Psi(M_n(x))| \geqq \eta|M_n(x)|$ for some $\eta > 0$, and (3.1.2) for $\tilde{M}_n$ follows from (3.1.2) for M_n. □

3.7 LEMMA. *Assume there exist constants* K_1, K_2 *such that*

$$|\Psi(t)| \leqq K_1|t| + K_2 \qquad \textit{for all} \quad t \in R . \tag{1}$$

Then (3.1.3) *holds for* $\tilde{M}_n$ *if it does for* M_n.

PROOF. Since $\tilde{M}_n(x) = \Psi(M_n(x))$,

$$|\tilde{M}_n(x)| \leqq K_1|M_n(x)| + K_2 .$$

Thus (3.1.3) is satisfied, since M_n satisfies (3.1.3). □

3.8 REMARK. Let $\sum a_n^2 c_n^{-2} < \infty$ and $\sup \tau_n < \infty$. Now concerning (3.1.5) we notice that it is satisfied if h is a bounded transformation. We also add this remark: if h is bounded by a straight line, M_n is bounded, and G has a bounded second moment and it is then easy to verify that (3.1.5) holds; furthermore, (3.1.3) also holds provided that M_n satisfies (3.1.3).

Here are some examples.

3.9 EXAMPLE. Let $h(v) = \text{sign}(v)$, $v \in R$. If g is continuous at 0, $g(0) \neq 0$, and (3.1.1)—(3.1.3) hold for M_n, then (3.1.1)—(3.1.5) hold for $\tilde{M}_n$.

PROOF. h is a bounded function, hence (3.1.3)—(3.1.5) are easily verified because $\sum a_n^2 c_n^{-2} < \infty$ and τ_n could be taken $= 0$. Since h is non-decreasing, Lemma 3.6 will imply (3.1.1) and (3.1.2) if $\lim_{t \downarrow 0} t^{-1}\Psi(t) > 0$. But from the continuity of g at 0 we obtain

$$t^{-1}\Psi(t) = 2t^{-1} \int_{-t}^{0} g(v)\, dv \to 2g(0) > 0 ; \qquad g(0) \neq 0 . \; \square$$

3.10 EXAMPLE. Let T_0 be a positive number and $h(v) = v$ if $|v| < T_0$ and $|h(v)| = T_0$ if $|v| \geqq T_0$. Assume that (3.1.1)—(3.1.3) hold for M_n and let $\int_0^{T_0} g(v)\, dv > 0$. Then (3.1.1)—(3.1.5) hold for $\tilde{M}_n$.

PROOF. h is bounded and non-decreasing, so as in the previous example it is enough to show $\lim_{t \to 0} t^{-1}\Psi(t) > 0$. But one may easily check that

$$\Psi(t) = t \int_{-T_0-t}^{T_0-t} g(v)\, dv + \int_{T_0-t}^{T_0+t} (T_0 - v) g(v)\, dv ,$$

and then it follows that

$$\lim_{t \to 0} t^{-1}\Psi(t) = \int_{-T_0}^{T_0} g(v)\, dv = 2 \int_0^{T_0} g(v)\, dv > 0 . \qquad \square$$

3.11 EXAMPLE. Let h be a bounded odd function, h' be bounded and $\int h'(v) g(v)\, dv > 0$. In addition let g be continuous and non-increasing on $[0, \infty)$ and $h(v) > 0$ for all $v > 0$. If M_n satisfies (3.1.3)—(3.1.3) and $\sum a_n^2 c_n^{-2} < \infty$, then (3.1.1)—(3.1.5) hold for $\tilde{M}_n$.

PROOF. Since h is bounded, one can choose $\tau_n = 0$ and thus (3.1.3), (3.1.4) and (3.1.5) hold. Also since (3.1.1)—(3.1.3) hold for M_n and (3.6.1)-(iii) is satisfied, Lemma 3.6 will imply (3.1.1) and (3.1.2) if $\lim_{t \downarrow 0} t^{-1}\Psi(t) > 0$. But h' being bounded implies that

$$\Psi'(0) = \int h'(v) g(v)\, dv > 0 . \qquad \square$$

4. Asymptotic normality of the modified procedures. In this section we shall use the following 1-dimensional version of a theorem of Fabian (1968b).

4.1 THEOREM. *Let $\mathscr{F}_n$ be a non-decreasing sequence of σ-fields, $\mathscr{F}_n \subseteq \mathscr{F}$. Suppose U_n, $\mathscr{V}_n$, T_n, Γ_n, Φ_n are random variables, $\sigma_0, \Gamma, \Phi \in R$, and $\Gamma > 0$. Suppose Γ_n, Φ_{n-1}, $\mathscr{V}_{n-1}$ are $\mathscr{F}_n$-measurable, $C, \alpha, \beta \in R$, and*

(1) $$\Gamma_n \to \Gamma\,, \qquad \Phi_n \to \Phi\,, \qquad T_n \to T \quad \text{or} \quad E[|T_n - T|] \to 0\,.$$

(2) $$E_{\mathscr{F}_n}[\mathscr{V}_n] = 0\,, \qquad C > |E_{\mathscr{F}_n}[\mathscr{V}_n^2] - \sigma_0^2| \to 0\,,$$

and with

(3) $$\sigma_{j,r}^2 = E[\chi_{[|\mathscr{V}_j|^2 \geq r j^\alpha]} |\mathscr{V}_j|^2]\,,$$

let either

(4) $$\lim_{j\to\infty} \sigma_{j,r}^2 = 0 \qquad \text{for every} \quad r > 0\,,$$

or

(5) $$\alpha = 1\,, \qquad \lim_{n\to\infty} \frac{1}{N} \sum_{j=1}^n \sigma_{j,r}^2 = 0 \qquad \text{for every} \quad r > 0\,.$$

Suppose $\beta_+ = \beta$ if $\alpha = 1$, and $\beta_+ = 0$ if $\alpha \neq 1$.

(6) $$0 < \alpha \leqq 1\,, \qquad 0 \leqq \beta\,, \qquad \beta_+ < 2\Gamma\,,$$

and

(7) $$U_{n+1} = (1 - n^{-\alpha}\Gamma_n)U_n + n^{-(\alpha+\beta)/2}\Phi_n \mathscr{V}_n + n^{-\alpha-\beta/2}T_n\,.$$

Then the asymptotic distribution of $n^{\beta/2}U_n$ is normal with

(8) $$\text{mean} = 2T(2\Gamma - \beta_+)^{-1} \qquad \text{and} \qquad \text{variance} = \sigma_0^2\Phi^2(2\Gamma - \beta_+)^{-1}\,.$$

(For the proof see Fabian (1968 b).)

4.2 *Asymptotic Normality Theorem. Let α_0, a, c and β be positive numbers, γ be a nonnegative number and $\zeta_0 \in R$. Consider the modified procedure* (1.1) *with*

$$a_n = a/n\,, \qquad c_n = c/n^\gamma\,, \qquad 0 \leqq \gamma < \tfrac{1}{2},\ n = 1, 2, \cdots;$$

suppose that $X_n \to \theta$ (this will be satisfied if $\tilde{M}_n$ satisfies (3.1.1)—(3.1.4), *since* (3.1.5) *will be implied by* (4.2.4) (*cf. Theorem* 3.1)).

(1) *Let h be continuous* a.e. *with respect to G; Ψ' exist at 0 and $\Psi'(0) = H(h) > 0$;*

(2) $$\beta = 1 - 2\gamma \qquad \text{and} \qquad a > \frac{\beta}{2\alpha_0 H(h)}\,.$$

Set

(3) $$S^2(t) = \int [h(t + v) - \Psi(t)]^2 g(v)\, dv\,, \qquad S_0^2(h) = S^2(0)\,;$$

and assume that

(4) *the function S^2 is bounded by a number σ^2 and is continuous at 0.*

In addition, let α_n, ζ_n be $\mathscr{X}_n$-measurable random variables, and with $s = \beta/(2\gamma)$ if

$\gamma \neq 0$; $s = 0$ *otherwise, assume that*

$$c_n^{-1}M_n(X_n) = \alpha_n(X_n - \theta) + \zeta_n c_n^s ; \tag{5}$$

$$\alpha_n \to \alpha_0 \quad \text{and} \quad \zeta_n \to \zeta_0 . \tag{6}$$

Then the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ *is normal with*

$$\text{mean} = -2ac^s H(h)\zeta_0[2a\alpha_0 H(h) - \beta]^{-1} \tag{7}$$

and

$$\text{variance} = a^2c^{-2}S_0^2(h)[2a\alpha_0 H(h) - \beta]^{-1} . \tag{8}$$

PROOF. The proof will be established by verifying the conditions of Theorem 4.1. We have

$$(X_{n+1} - \theta) = (X_n - \theta) - a_n c_n^{-1}\tilde{M}_n(X_n) + a_n c_n^{-1} Z_n ; \tag{9}$$

where

$$Z_n = -(h(Y_n) - \tilde{M}_n(X_n)) . \tag{10}$$

Define the following $\mathscr{X}_n$-measurable random variables

$$\begin{aligned} H_n &= H(h) && \text{if } M_n(X_n) = 0 , \\ &= M_n^{-1}(X_n)\tilde{M}_n(X_n) && \text{if } M_n(X_n) \neq 0 ; \qquad n = 1, 2, \cdots . \end{aligned} \tag{11}$$

Then the term $a_n c_n^{-1}\tilde{M}_n(X_n)$ in (9) can be written as $a_n H_n c_n^{-1} M_n(X_n)$; and using (5) we obtain

$$a_n c_n^{-1}\tilde{M}_n(X_n) = a_n \alpha_n H_n(X_n - \theta) + a_n c_n^s H_n \zeta_n .$$

Since $a_n c_n^s = ac^s n^{-1-\beta/2}$ and $a_n c_n^{-1} = ac^{-1}n^{-\frac{1}{2}-\beta/2}$, we can rewrite (9) as

$$\begin{aligned} (X_{n+1} - \theta) = (X_n - \theta)&(1 - a\alpha_n H_n n^{-1}) \\ &+ ac^{-1}n^{-(1+\beta)/2}Z_n - ac^s n^{-1-\beta/2}H_n\zeta_n . \end{aligned} \tag{12}$$

Apply Theorem 4.1 with

$$\alpha = 1 , \quad \mathscr{F}_n = \mathscr{X}_n , \quad \Gamma_n = a\alpha_n H_n , \quad U_n = (X_n - \theta) ,$$
$$\Phi_n = a/c , \quad \mathscr{V}_n = Z_n \quad \text{and} \quad T_n = -ac^s H_n \zeta_n .$$

Now $c_n^{-1}M_n(X_n) \to 0$ by (5), since $X_n \to \theta$, and this with (1) implies that $H_n \to H(h)$. Thus

$$\Gamma_n \to a\alpha_0 H(h) , \qquad T_n \to -ac^s H(h)\zeta_0 , \qquad \Phi_n \to \Phi = a/c .$$

Also from (3) and the continuity of S^2 at 0 we obtain

$$E_{\mathscr{X}_n}[Z_n] = S^2(M_n(X_n)) \to S_0^2(h) .$$

Thus we have shown that (4.1.1) and (4.1.2) of Theorem 4.1 are satisfied with $\Gamma = a\alpha_0 H(h)$, $\Phi = a/c$, $T = -ac^s H(h)\zeta_0$, $C = \sigma^2$ and $\sigma_0^2 = S^2(0) = S_0^2(h)$.

Concerning (4.1.4) we have

$$\sigma_{j,r}^2 = E[\chi_{[Z_j^2 \geq rj]} Z_j^2] = E[E_{\mathscr{X}_j}[\chi_{[Z_j^2 \geq rj]} Z_j^2]] ; \tag{13}$$

the conditional expectations form a uniformly integrable sequence since they are dominated by $E_{\mathscr{F}_j} Z_j^2 = S^2(M_j(X_j)) \leqq \sigma^2$, by (4). Thus (4.1.4) will be verified if we show that the conditional expectations converge to 0. But the jth conditional expectation is equal to $Q_j(M_j(X_j))$ where

$$Q_j(t) = \int \xi(t, v)\chi_{[\xi(t,v)\geqq rj]}G(dv) , \tag{14}$$

with

$$\xi(t, v) = [h(t + v) - \Psi(t)]^2 \qquad \text{for all} \quad t, v \in R .$$

The integrands in (14) are uniformly integrable as $t \to 0$; since they are dominated by $\xi(t, v)$ for all $v \in R$ and $\int \xi(t, v)G(dv) = S^2(t) \to S_0^2(h)$ as $t \to 0$. Thus to show $Q_j(t) \to 0$ as $t \to 0$ and $j \to \infty$, it is enough to show for almost all (with respect to G) $v \in R$, $\xi(t, v)\chi_{[\xi(t,v)\geqq rj]} \to 0$ as $t \to 0$ and $j \to \infty$; and for this it is enough to show for almost all (with respect to G) $v \in R$, $\xi(t, v) \to h^2(v)$ as $t \to 0$. Let Λ be the set of points at which h is discontinuous. Λ has a probability zero under G. By the continuity of h on $R - \Lambda$, and since by (1) Ψ is continuous at 0, then $\xi(t, v) \to h^2(v)$ as $t \to 0$ for all $v \in R - \Lambda$. This completes the proof of (4.1.4). The measurability condition follows from the definition of Γ_n, Φ_{n-1} and $\mathscr{V}_{n-1}$. Also, by (2), $2a\alpha_0 H(h) > \beta$. Hence the conclusion of the theorem follows by Theorem 4.1. □

4.3 Remark. Because of our interest in the question of optimality of the transformation h in the modified procedure (1.1), the preceding theorem is stated in terms of the behavior of M_n, (see conditions (4.2.5) and (4.2.6)), with sufficient conditions on h in order to guarantee the asymptotic normality of (1.1). But the theorem can also be applied to the modified procedure directly, because this procedure can be written as the original procedure with a change of the meaning of Y_n's and with h equal to the identity transformation.

4.4 Theorem. (KW situation). *Let f, Y_n satisfy the conditions of the* KW *situation in Section* 2.1 *with possibly the exception of* (2.1.5). *Let f'' exist and be continuous in a neighborhood of θ with $f''(\theta) = M > 0$. Consider the modified procedure with a transformation h which is continuous* a.e. *with respect to G; further assume that* (3.1.1)—(3.1.3) *are satisfied for $\tilde{M}_n$ if they are for M_n. Let Ψ' exist at* 0 *with $\Psi'(0) = H(h) > 0$ and*

$$a_n = \frac{a}{n} , \qquad c_n = \frac{c}{n^\gamma} , \qquad \gamma = \tfrac{1}{4} , \qquad a > [8MH(h)]^{-1} . \tag{1}$$

In addition let $S^2(t) = \int [h(t + v) - \Psi(t)]^2 g(v)\, dv$, $t \in R$, be bounded by a constant σ^2 and continuous at 0. *Then $X_n \to \theta$, and the asymptotic distribution of $n^{\frac{1}{4}}(X_n - \theta)$ is normal with*

$$\textit{mean} = 0 , \qquad \textit{and} \qquad \textit{variance} = a^2c^{-2}S_0^2(h)[4Ma(h) - \tfrac{1}{2}]^{-1} . \tag{2}$$

Proof. It can be easily shown that $M_n(X_n) = f(X_n + c_n) - f(X_n - c_n)$ satisfy conditions (3.1.1)—(3.1.3), and thus by Theorem 3.1 we conclude that $X_n \to \theta$, since conditions (3.1.4) and (3.1.5) are satisfied too. The rest of the conclusion of the theorem follows simply by verifying the conditions of Theorem 4.2. The only conditions of Theorem 4.2 which are not directly assumed in our theorem are (4.2.5), (4.2.6) and (4.2.2). Let $\Delta(x, c_n) = f(x + c_n) - f(x - c_n)$, $x \in R$. Let $I = [\theta - \varepsilon, \theta + \varepsilon]$ be an interval on which f'' exists and is continuous. Define

$$(3) \qquad \varphi(x) = (x-\theta)^{-1}f'(x) - f''(\theta) \quad \text{for } x \in I \text{ with } \varphi(\theta) = 0\,, \qquad = 0 \quad \text{otherwise.}$$

Then by expanding $\Delta(x, c_n)$ as a function of c_n, and substituting for $f'(x)$ from (3), it follows that

$$(4) \qquad c_n^{-1}\Delta(x, c_n) = 2(x-\theta)[f''(\theta) + \varphi(x)] + \eta(x, c_n)c_n$$

where $\varphi(x) \to 0$ as $x \to \theta$ and $\eta(x, c_n) \to 0$ if $c_n \to 0$ and $x \to \theta$. Obviously φ and $\eta(\cdot, c_n)$, the latter as defined by (4), are Borel measurable functions. Then

$$c_n^{-1}M_n(X_n) = \alpha_n(X_n - \theta) + \zeta_n c_n$$

with $\mathscr{Z}_n$-measurable $\alpha_n = 2[f''(\theta) + \varphi(X_n)]$ and $\zeta_n = \eta(X_n, c_n)$; further, $\alpha_n \to 2f''(\theta) = 2M$ and $\zeta_n \to 0$. Thus (4.2.5) and (4.2.6) hold with $\alpha_0 = 2M$, $\zeta_0 = 0$ and $s = 1$, since $\beta = \frac{1}{2}$. Condition (4.2.2) then follows from (1). Hence the conditions of Theorem 4.2 are satisfied. This completes the proof. □

4.5 THEOREM. (KW situation). *In Theorem* 4.4 *let* (4.4.1) *be replaced by*

$$(1) \qquad a_n = \frac{a}{n}\,, \qquad c_n = \frac{c}{n^\gamma}\,, \qquad \gamma = \tfrac{1}{6} \qquad \text{and} \qquad a > [6MH(h)]^{-1}\,.$$

Moreover, let f''' exist and be continuous in a neighborhood of θ. Then $X_n \to \theta$, and the asymptotic distribution of $n^{\frac{1}{3}}(X_n - \theta)$ is normal with

$$(2) \qquad \text{mean} = -\tfrac{2}{3}ac^2H(h)f'''(\theta)[4MaH(h) - \tfrac{2}{3}]^{-1}\,, \qquad \text{and}$$
$$\text{variance} = a^2c^{-2}S_0^2(h)[4MaH(h) - \tfrac{2}{3}]^{-1}\,.$$

PROOF. It is again easy to conclude, by Theorem 3.1, that $X_n \to \theta$, and then it remains to verify (4.2.5), (4.2.6) and (4.2.2) of Theorem 4.2 in order to complete the proof of the theorem. Let $\Delta(x, c_n) = f(x + c_n) - f(x - c_n)$, $x \in R$. Let $I = [\theta - \varepsilon, \theta + \varepsilon]$ be an interval on which f''' exists and is continuous. Similarly, as in the proof of the preceding theorem and with the same φ, we obtain that

$$(3) \qquad c_n^{-1}M_n(X_n) = \alpha_n(X_n - \theta) + \zeta_n c_n^2$$

with $\mathscr{Z}_n$-measurable $\alpha_n = 2[f''(\theta) + \varphi(X_n)]$ and $\zeta_n = \frac{1}{3}[f'''(\theta) + \eta(X_n c_n)]$; further $\alpha_n \to 2f''(\theta) = 2M$ and $\zeta_0 = \frac{1}{3}f'''(\theta)$. Thus (4.2.5) and (4.2.6) hold with $\alpha_0 = 2M$, $\zeta_0 = \frac{1}{3}f'''(\theta)$, $\beta = \frac{2}{3}$ and $s = 2$. Condition (4.2.2) follows from (1). Hence the conditions of Theorem 4.2 are satisfied. This completes the proof. □

The following theorem is presented here to cover a case of the RM situation which is treated in Albert and Gardner (1967).

4.6 THEOREM. *Let f_n be a sequence of Borel measurable functions defined on R such that* (3.1.1)—(3.1.3) *are satisfied for $M_n = f_n$ with $\tau_n = 1$, $c_n = 1$ and $f_n(\theta) = 0$. Let $d > 0$ and*

$$(1) \qquad D_n(x) = (x-\theta)^{-1} f_n(x) \quad \text{if } x \neq \theta , \qquad = d \quad \text{if } x = \theta ; \qquad n = 1, 2, \cdots$$

be continuously convergent at θ to d. Consider the modified procedure (1.1) *with a transformation h which is continuous* a.e. *with respect to G, and let h be such that* (3.1.1)—(3.1.3) *are satisfied for $\tilde{M}_n$ if they are for M_n. Also let Ψ' exist at* 0, *$\Psi'(0) = H(h) > 0$ and*

$$(2) \qquad a_n = \frac{a}{n} , \qquad a > [2dH(h)]^{-1} .$$

In addition let $S^2(t) = \int [h(t+v) - \Psi(t)]^2 g(v)\, dv$, $t \in R$, be bounded by a constant σ^2 and continuous at 0. *Then $X_n \to \theta$ and the asymptotic distribution of $n^{\frac{1}{2}}(X_n - \theta)$ is normal with*

$$(3) \qquad \textit{mean} = 0 \qquad \textit{and} \qquad \textit{variance} = a^2 S_0^2(h)[2adH(h) - 1]^{-1} .$$

PROOF. Under the given conditions we obtain, by applying Theorem 3.1, that $X_n \to \theta$. To obtain the rest of the conclusion of the theorem, we apply Theorem 4.2, for which we need only to verify (4.2.5), (4.2.6) and (4.2.2). We have

$$M_n(X_n) = f_n(X_n) = D_n(X_n)(X_n - \theta) .$$

Since $\gamma = 0$, then $\beta = 1$ and $s = 0$. Thus (4.2.5) and (4.2.6) hold with $\mathscr{F}_n$-measurable $\alpha_n = D_n(X_n)$ and $\zeta_n = \zeta_0 = 0$; further $\alpha_n \to d_0 = d$, since $(D_n)_{n=1}^{\infty}$ is continuously convergent at θ to d. Condition (4.2.2) follows from (2). Hence the conclusion follows by Theorem 4.2.

4.7 *The optimal choice of* (a, c). Let ξ be a normal random variable with mean and variance given by (4.2.7) and (4.2.8) respectively. Then

$$(1) \qquad E\xi^2 = \frac{4a^2 H^2(h)\zeta_0^2}{(2a\alpha_0 H(h) - \beta)^2} c^{2s} + \frac{a^2 S_0^2(h)}{(2a\alpha_0 H(h) - \beta)} c^{-2} .$$

By elementary manipulation it can be shown (cf. Abdelhamid (1971)) if $\zeta_0 \neq 0$, the optimal values of (a, c), i.e. values for which $E\xi^2$ is minimized with the other quantities being fixed, are given by

$$(2) \qquad (a, c) = \left(\frac{1}{\alpha_0 H(h)} , \left[\frac{(s+2)}{4s(s+1)} (S_0^2(h)/\zeta_0^2 H^2(h))\right]^{1/(2s+2)}\right) .$$

With this optimal choice of (a, c), (1) becomes

$$(3) \qquad E\xi^2 = \frac{(1+s)^2}{s(s+2)^2 \alpha_0^2} [4s\zeta_0^2 (S_0^2(h)/H^2(h))^s]^{1/(s+1)} .$$

On the other hand, if $\zeta_0 = 0$, then the optimal value of a, i.e. value of a which minimizes $E\xi^2$ with the other quantities being fixed, is given by:

$$a = \frac{1}{\alpha_0 H(h)} . \tag{4}$$

With this value of a, (1) reduces to

$$E\xi^2 = \left(\frac{s+1}{s+2}\right)[S_0^2(h)/\alpha_0^2 H^2(h)]c^{-2} . \tag{5}$$

In particular, in the KW situation under the conditions of Theorem 4.5 and if $f'''(\theta) \neq 0$, we have

$$(a, c) = \left(\frac{1}{2MH(h)}, [\tfrac{3}{2}(S_0^2(h)/f'''^2(\theta)H^2(h))]^{\frac{1}{6}}\right) . \tag{6}$$

4.8 Remark. We have investigated the optimal values of a and c for cases covered by Theorem 4.2. There are results on convergence and asymptotic normality where a_n, c_n are not necessarily of the form an^{-1} and $cn^{-\gamma}$ (Burkholder (1956), Sacks (1958), Schmetterer (1968)), but they do not lead to improvement in speed and so are not considered here. The determination of the optimal values of a and c in cases where asymptotic normality does not hold have not been considered and an investigation of this point is welcomed.

We also notice that if the mean of the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ is zero (see, for example, Theorem 4.4), then (unless c is known) there is no optimal value of c. In such cases c may be designed according to the physical nature of each problem considered.

4.9 Remark. We notice that the unpleasant feature of the optimal values of a and c is that they depend on values, $(f'(\theta), f''(\theta), f'''(\theta))$, which are, in general, unknown. But the value of a in the RM situation and the value of (a, c) in the KW situation can be estimated during the approximation process and fed back into the procedure.

For the original RM procedure Venter (1967) used a procedure (later generalized by Fabian (1968 b)), which estimates the optimal value of a. Recently in Fabian (1971) a procedure was described which itself estimates the optimal value of a for a modified version of the original KW procedure. The same ideas can be used to obtain a procedure which estimates the optimal value of a or both (a, c) for the modified procedure.

4.10 *Effect of taking m observations at each stage.* Suppose that an experimenter observes m random variables $Y_{n,1}, \cdots, Y_{n,m}$ instead of one, Y_n, at stage n such that these m random variables are conditionally, given $\mathscr{F}_n$, independently distributed according to G. Suppose he then uses $(1/m)\sum_{j=1}^m h(Y_{n,j})$ instead of $h(Y_n)$ in the modified procedure (1.1). The conditional expectation of the average will be the same as that of $h(Y_n)$ and the conditional variance will only be changed by a factor of $(1/m)$, and it is easy to see, under the conditions of Theorem 4.2, that this will result in changing the variance of the asymptotic distribution by the factor $(1/m)$. Thus, in the KW situation, under the conditions of Theorem 4.5 we obtain

$$(1) \qquad n^{\frac{1}{3}}(X_n - \theta) \to_{\mathscr{L}} N\left(-\frac{\frac{2}{3}ac^2 f'''(\theta)H(h)}{4MaH(h) - \frac{2}{3}},\ \frac{a^2c^{-2}S_0^2(h)}{m(4MaH(h) - \frac{2}{3})}\right).$$

On the other hand, if the experimenter wishes to continue the modified procedure for nm stages, rather than using averages, then by Theorem 4.5 we have

$$(2) \qquad (nm)^{\frac{1}{3}}(X_{nm} - \theta) \to_{\mathscr{L}} N\left(-\frac{\frac{2}{3}ac^2 f'''(\theta)H(h)}{4MaH(h) - \frac{2}{3}},\ \frac{a^2c^{-2}S_0^2(h)}{4MaH(h) - \frac{2}{3}}\right).$$

Let $f'''(\theta) \neq 0$. Then in (2) the optimal choice of (a, c) (see (4.7.6)), is given by

$$(3) \qquad (a, c) = \left(\frac{1}{2MH(h)},\ [\tfrac{3}{2}(S_0^2(h)/f'''^2(\theta)H^2(h))]^{\frac{1}{6}}\right)$$

while the optimal choice of (a, c) in (1) is given by

$$(4) \qquad (a, c) = \left(\frac{1}{2MH(h)},\ [\tfrac{3}{2}(S_0^2(h)/mf'''^2(\theta)H^2(h))]^{\frac{1}{6}}\right).$$

Let ξ_1, ξ_2 be the normal random variables on the R.H.S. of (1) and (2), respectively. Then with the corresponding optimal choice of (a, c), one can easily check that

$$E\xi_1^2 = E\xi_2^2 .$$

Therefore using an average of m independent observations at each stage is asymptotically equivalent to continuation of the modified procedure for nm stages. The only effect is in decreasing the optimal value of c by a factor of $(1/m)^{\frac{1}{6}}$.

5. Optimal transformations. We have seen, in Section 4, that the asymptotic results for the modified procedure (1.1) in *both the* RM *situation and the* KW *situation* are special cases of the situation described in Theorem 4.2. The second moment of the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ can be written (see (4.2.7) and (4.2.8)) as

$$(1) \qquad \frac{4a^2c^{2s}\zeta_0^2}{(2a\alpha_0 - (\beta/H(h)))^2} + \frac{a^2c^{-2}S_0^2(h)}{(2a\alpha_0 H(h) - \beta)}.$$

If $\zeta_0 \neq 0$, then with the optimal choice of (a, c), (1) becomes

$$(2) \qquad \frac{\beta^2(1 + s)^2}{\alpha_0^2 s(s + 2)^2\alpha_0^2}[4s\zeta_0^2(S_0^2(h)/H^2(h))^s]^{1/(s+1)}.$$

It can be easily shown (cf. Abdelhamid (1971)) that it is enough to consider transformations with $S_0^2(h) = 1$. Then the above expressions are minimized by the choice h which maximizes $H(h)$.

Let $\mathscr{H}$ be the family of all Borel measurable transformations h such that h satisfies Assumption 3.5, $S_0^2(h) = 1$, and $H(h)$ can be computed by differentiating under the integral sign; that is

$$0 < H(h) = \Psi'(0) = \int h(v)(-g'(v))\,dv\,.$$

5.1 LEMMA. *Let the density g have a derivative* a.e. *with respect to G. In addition let*

$$(1) \qquad 0 < I(g) = \int (g'(v)/g(v))^2\,dG(v) < \infty\,,$$

and set $\Gamma = [I(g)]^{\frac{1}{2}}$. *Suppose that* $h_0 = -1/\Gamma(g'/g)$ a.e. *with respect to G and* $h_0 \in \mathscr{H}$. *Then within* $\mathscr{H}$, $H(h)$ *is maximized by* h^* *if and only if* $h^* = h_0$ a.e. *with respect to G.*

The proof is the same as that of Theorem 1 in Anbar (1973).

5.2 DEFINITION. Suppose that $0 < I(g) < \infty$, $h_0 = -1/\Gamma(g'/g)$ a.e. with respect to G, and $h_0 \in \mathscr{H}$. In addition suppose that the modified procedure is used with $h = h_0$, for which $X_n \to \theta$, and $n^{\beta/2}(X_n - \theta)$ has asymptotic distribution as given in Theorem 4.2. Then we shall call h_0 the *optimal transformation*. A modified procedure with such an optimal transformation will be called *optimal procedure*.

5.3 *Asymptotic efficiency of optimal stochastic approximation procedures.* The surprising fact is that the optimal stochastic approximation procedures are not only optimal within the class of approximation procedures considered but also they are asymptotically as efficient as the best unbiased estimators of θ, the parameter to be estimated. This is true despite the very simple recurrence relation that generates the approximation sequence X_n.

In more detail, we show that the variance of the asymptotic distribution of $n^{\beta/2}(X_n - \theta)$ corresponds to the Cramér–Rao lower bound for the variance of an unbiased estimator based on the first n observations.

As an application of Theorem 4.6, let

$$(1) \qquad Y_n = f_n(\theta) + V_n\,, \qquad n = 1, 2, \cdots$$

be observations on known functions f_n except for θ which is assumed to belong to some interval, Θ. Let the error random variables $V_n = Y_n - f_n(\theta)$ be independent and distributed according to G, which satisfies the conditions of Lemma 5.1. Furthermore, for each n let f_n have the same unique root $\theta \in \Theta$, and f_n exist at θ and $f_n'(\theta) \to d$ where d is positive and known. Also let f_n satisfy the conditions stated in Theorem 4.6. Then by using the optimal procedure one can show (cf. Abdelhamid (1971)) that $n^{\frac{1}{2}}(X_n - \theta) \to_{\mathscr{L}} N(0, (d\Gamma)^{-2})$. That means (see Theorem 5.2 in Albert and Gardner (1967), page 68) our optimal procedure is asymptotically efficient. The case $f_n = f$ has been treated independently by Anbar (1973).

Albert and Gardner (1967; see Chapter 5 there) tried to increase the efficiency of the RM type procedure which they used in their monograph by making transformation of the parameter space Θ. Their procedure stayed less efficient except when the error random variables are normally distributed.

The optimal procedure applied to the case $f(x) = d(x - \theta)^2$ can also be used to generate asymptotically efficient estimators by applying Theorem 4.2.

It is worth noting that the KW procedure, as well as the RM procedure, is optimal if and only if the error random variables are normally distributed. As for procedure (F), it is optimal if and only if $(-g/g)(v) = C \operatorname{sign}(v)$ with a constant $C > 0$, and this is true if and only if G is a double exponential distribution.

5.4 *Some examples of new optimal procedures.* In the following we give examples of new optimal procedures which are different from the original RM and KW procedures (see also Anbar (1973)). The first two examples fall under the case of Example 3.11 (for more details see Abdelhamid (1971)).

(a) *Student's type distribution.* Let G have a density function given by:

$$g(t) = \frac{1}{(\nu\pi)^{\frac{1}{2}}} \frac{\Gamma((1+\nu)/2)}{\Gamma(\nu/2)} (1 + t^2/\nu)^{-(1+\nu)/2}, \qquad t \in R,\ \nu > 0.$$

The case $\nu = 1$ gives the Cauchy density.

Recall that for $\nu \leqq 2$ the variance of G does not exist and such types of densities (with $\nu \leqq 2$) are not allowed by either the RM procedure or the KW procedure, since both stipulate the existence of the variance of G.

With some manipulation and application of Theorem 4.2 (see also Section 3.11), one can check that the optimal transformation is given by:

$$h_0(t) = C_\nu \frac{t}{\nu + t^2}, \qquad t \in R,$$

where C_ν is a positive constant satisfying $\int_{-\infty}^{\infty} h_0^2(v) g(v)\, dv = 1$.

(b) *Logistic distribution.* Let G have a logistic density function given by:

$$g(v) = \frac{1}{2(1 + \cosh(v))}, \qquad v \in R.$$

It can also be checked (see Section 3.11) that the optimal transformation is given by

$$h_0(v) = C \frac{\sinh(v)}{1 + \cosh(v)}, \qquad v \in R,$$

with $C = 3^{\frac{1}{2}}/2$.

(c) Let G have a density function given by

$$\begin{aligned} g(v) &= \frac{C_0}{(2\pi)^{\frac{1}{2}}} e^{-v^2/2}, && \text{if } |v| < T \\ &= \frac{C_0}{(2\pi)^{\frac{1}{2}}} e^{-T|v| + (T^2/2)}, && \text{if } |v| \geqq T, \end{aligned}$$

where C_0 and T are positive constants.

This g behaves like a normal density for small v, and then like a double exponential for large v. (This is what some authors call Huber's density.) It follows that

$$-\frac{1}{\Gamma}(g'(v)/g(v)) = Kv \qquad \text{if} \quad |v| < T$$
$$= KT \operatorname{sign}(v) \qquad \text{if} \quad |v| \geqq T,$$

where K also depends on T. Denoting this transformation by h_0 we see that h_0 is an odd, bounded and non-decreasing transformation which satisfies Assumption 3.5. Thus $h_0 \in \mathscr{H}$ and $h_0 = -1/\Gamma(g'/g)$ a.e. maximizes $H(h)$. Also h_0 preserves the a.s. convergence (see Lemmas 3.6 and 3.7) and it satisfies the conditions of Theorem 4.2. Hence h_0 is an optimal transformation.

Acknowledgment. The author wishes to express his sincere gratitude and indebtedness to his advisor, Professor Václav Fabian, who patiently read the manuscript and whose valuable comments and careful criticism helped in improving the presentation of the results in this paper.

REFERENCES

ABDELHAMID, SAMI N. (1971). Transformation of observations in stochastic approximation. Ph. D. thesis, Michigan State Univ.

ALBERT, ARTHUR E. and GARDNER, LELAND A. (1967). Stochastic approximation and nonlinear regression. Research Monograph No. 42, M.I.T. Press.

ANBAR, D. (1973). On optimal estimation methods using stochastic approximation procedures. *Ann. Statist.* **1** 1175–1184.

BLUM, J. R. (1954). Approximation methods which converge with probability one. *Ann. Math. Statist.* **25** 737–744.

BURKHOLDER, D. L. (1956). On a class of stochastic approximation processes. *Ann. Math. Statist.* **27** 1044–1059.

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

CHUNG, KAI LAI (1968). *A Course in Probability Theory.* Harcourt, Brace and World, New York.

DERMAN, C. and SACKS, J. (1954). On Dvoretzky's stochastic approximation theorem. *Ann. Math. Statist.* **30** 601–606.

DUBINS, LESTER E. and FREEDMAN, DAVID (1965). A sharper form of the Borel-Cantelli lemma, and the strong law. *Ann. Math. Statist.* **36** 800–818.

DUPAČ, V. (1957). On Kiefer-Wolfowitz approximation method. *Časōpis Pěst. Mat.* **82** 47–75.

DVORETZKY, ARYEH (1956). On stochastic approximation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 39–59. Univ. of California Press.

FABIAN, V. (1960). Stochastic approximation methods. *Czech. Math. J.* **10** 123–159.

FABIAN, V. (1964). A new one-dimensional stochastic approximation method for finding a local minimum of a function. *Trans. Third Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*, Czech. Acad. Sci. Prague. 85–105.

FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.* **38** 191–200.

FABIAN, V. (1968a). On the choice of design in stochastic approximation methods. *Ann. Math. Statist.* **39** 457–465.

FABIAN, V. (1968b). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327–1332.

FABIAN, V. (1971). Stochastic approximation. *Proc. Symp. Optimizing Methods in Statistics.* Ohio State Univ. June 1971. J. S. Rustagi, ed., Academic Press, New York.

GROENEVELD, R. A. (1971). A note on the sequential sign test. *Amer. Statistician.* **25** 2.

HODGES, J. L., JR. and LEHMANN, E. L. (1956). Two approximations to the Robbins-Monro process. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 95-104. Univ. of California Press.

KESTEN, HARRY (1958). Accelerated stochastic approximation. *Ann. Math. Statist.* **29** 41-59.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* **23** 462-466.

LOÈVE, MICHEL (1963). *Probability Theory*, (3rd ed.). Van Nostrand, Princeton.

PONTRYAGIN, L. S. *et al.* (1963). *The Mathematical Theory of Optimal Process.* Interscience, New York.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400-407.

ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for non-negative almost sure supermartingale and some applications. *Proc. Symp. Optimizing Methods in Statistics*, Ohio State Univ. June 1971. J. S. Rustagi, ed. Academic Press, New York.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* **29** 373-405.

SCHMETTERER, L. (1968). Multidimensional stochastic approximation, Multivariate analysis, II. *Proc. 2nd Int. Symp.*, Dayton, Ohio. 443-460. Academic Press, New York.

VENTER, J. H. (1967). An extension of the Robbins-Monro procedure. *Ann. Math. Statist.* **38** 181-190.

WOLFOWITZ, J. (1956). On stochastic approximation methods. *Ann. Math. Statist.* **27** 1151-1155.

ZACKS, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.

FACULTY OF ENGINEERING
ALEXANDRIA UNIVERSITY
ALEXANDRIA, EGYPT