MILITARY TECHNICAL COLLEGE

CAIRO - EGYPT

ON DYNAMIC KIEFER-WOLFOWITZ STOCHASTIC APPROXIMATION PROCEDURES

El Sayed Sorour

Military Technical College,
Department of Mathematics.

Abstarct

Let $M(x)$ be a function from $R^k \longrightarrow R$. Let θ_n, $n = 1,2,...$ be vectors in R^k, θ_1 being the value at which $M(x)$ achieves its unique minimum. Set $M(x) = M_1(x)$, for $n = 1,2,...,$ set $M_n(x) = M(x - \theta_n - \theta_1)$. Then θ_n is the unique minimum of $M_n(x)$, which is unknown and is to be estimated. In our case, we assume that θ_n moves in such a manner that $\theta_{n+1} = g_n(\theta_n) + v_n$ where $g_n(\theta_n)$ is general non-linear k-vector measurable function (known) defined for all $x \in R^k$ and v_n is an unknown k-vector function (random or non-random) independent of x. Let a_n, c_n, $n = 1,2,...$ be two sequences of positive numbers. Let x_1 be an arbitrary random variable. Define for $n = 1,2,...,x_{n+1} = \overset{*}{x}_n + a_n(\overset{*}{y}_{2n} - \overset{*}{y}_{2n-1})/c_n$ where $\overset{*}{x}_n = g_n(x_n)$, and $\overset{*}{y}_{2n}$, $\overset{*}{y}_{2n-1}$ are random variables such that their expectations given $x_1,x_2,....,x_n$ are $\sum_{i=1}^{k} M_{n+1}(x_n + e_i c_n)e_i$ and $\sum_{i=1}^{k} M_{n+1}(x_n - e_i c_n)e_i$ respectively and their conditional variance are bounded by a constant σ^2 and they are conditionally independent. Under conditions similar to those used by Dupac (1966), we show that $\|x_n - \theta_n\| \longrightarrow 0$ with probability one.

1. INTRODUCTION

This paper is concerned with the dynamic Kiefer-Wolfowitz (KW) stochastic approximation procedure. This problem has been firstly studied by Dupac [2], [3]. He discussed in his papers the cases where the movement of the maximum can be expressed by a certain linear function of its present location and determinestic trend, i.e. he assume that

$$\theta_{n+1} = (1 + \frac{1}{2})\theta_n + v_n \tag{1.1}$$

where θ_n is the unique maximum of the regression function $M_n(x)$. Uosaki, K. (1974) discussed the one-dimensional dynamic Robbins-Monro (RM) process. He considers the case where the movement of the root can be expressed by a specified non-linear function of its present location. Sorour (1978) geenralizes Uosaki's result to the multidimensional dynamic (R<). In this paper; we shall be concerned with the non-linear multidimensional dynamic systems. The convergence of the approximation to the moving minimum of a nonlinear regression function, with probability one, is proved. In sec. 4 we prove the convergence with probability one and in sec. 5, we show that under some regularity conditions on the noise, the process is asymtotoically normal.

2. DESCRIPTION OF THE PROCEDURE

Let R^k be a real k-dimensional vector space. If x and y are two vectors in R^k we denote their-inner product by (x,y) and their nroms by $\|x\|$ and $\|y\|$ respectively. Let $M_n(x)$ be a (unknown) function from $R^k \longrightarrow R$. Assume that θ_n is the unique minimum of $M_n(x)$. Our goal is to estimate θ_n. In our case we assume that θ_n moves in a manner that

$$\theta_{n+1} = g_n(\theta_n) + v_n \tag{2.1}$$

where $g_n(x)$ is in general a nonlinear measurable function (known) from $R^k \longrightarrow R^k$ and v_n is unknown k vector (non-random).

Let a_n, $n = 1,2,...$ be positive numbers. Let X_1 be an arbitrary random variable define for $n = 1,2,....$

$$X_{n+1} = \overset{*}{X}_n - a_n F_n \tag{2.2}$$

where

$$\overset{*}{X}_n = g_n(X_n) \; ;$$

$$F_n = \overline{D}_n + d_n \tag{2.3}$$

and the ith compenent of $\overline{D}_n$ is given by

$$\overline{D}_n^i = \left[M_{n+1}(x_n + e_i c_n) - M_{n+1}(x_n - e_i c_n)\right] / 2\, c_n .$$

where e_i is the ith column in the identity matrix and c_n is a positive sequence of real numbers. Let $\mathcal{F}_n$ be the σ-field generated by $X_1, X_2, ..., X_n$. For x a random vector in R^k let $E_n(x)$ and $Var_n(x)$ denote the conditional expectation and the conditional variance of x with respect to $\mathcal{F}_n$ respectively. Let

$$E_n(d_n) = 0 \tag{2.4}$$

and for constant k_1 let

$$E_n(\|d_n\|^2) \leq k_1\, c_n^{-2} \tag{2.5}$$

Let $D_n(x)$ denotes the vector of the partial derivatives of $M_n(x)$

Then

$$D_n^i(x) = \frac{\partial\, M_n(x)}{\partial\, x^i}$$

Assume that

$$\|D_n - D_{n+1}(\overset{\star}{x}_n)\| \leq k_2\, c_n \tag{2.6}$$

In what follows $(Q, \mathcal{F}, P)$ will be a probability space, relations and convergence of random variables, vectors and matrices will be meant with probability one. The indicator function of set A will be denoted $x^A = R^{k\times k}$ is the space of all real $k\times k$ natrices. The unit matrix in $R^{k\times k}$ is denoted by I and $\|\cdot\|$ is the Euclidean norm. With h_n a sequence of numbers let $0(h_n)$, $0(h_n)$ denote sequences g_n and G_n, say of elements in one of the sets R, R^k such that $h_n^{-1}\, g_n \longrightarrow 0$, $\|h_n^{-1}\, G_n\| \leq f$ for $f\in R$.

Remark : Let $Y_n(i,1)$, $Y_n(i,2)$ be two random variables such

$$E_n(Y_n(i,1)) = M_{n+1}(\overset{\star}{X}_n + e_i c_n)$$

$$E_n(Y_n(i,2)) = M_{n+1}(\overset{\star}{X}_n - e_i c_n)$$

Then $F_n = \left[Y_n(i,1) - Y_n(i,2)\right]/2c_n$, and $Y_n(i,1)$ and $Y_n(i,2)$ are conditionally independent. If $Var_n(Y_n(i,1))$ and $Var(Y_n(i,2))$ are bounded. Then (2.5) holed. Also let $H_n(x)$ be the Hessian of $M_n(x)$ i.e.

$$H_n^{ij}(x) = \frac{\partial^2 M_n(x)}{\partial x^i\, \partial x^j}$$

If $\sup_x \left\{\|H_n(x)\| < \infty\right\}$ for $n = 1,2,\ldots.$ Then (2.6) holds.

3. CONSITIONS

Conditions on the regression function $M_n(x)$

M1 : There existy two numbers A and B such that

$$\|D_n(x)\| \leq A\|x - \theta_n\| + B.$$

M2 : For all $\delta > 0$, we have

$$\text{if} \inf_{n\in N} \inf_{\delta<\|x-\theta_n\|<\delta^{-1}} \langle x - \theta_n, D_n(x)\rangle > 0$$

Conditions on the Sequences a_n and c_n

A1 : $\sum_{n=1}^{\infty} a_n c_n < \infty$; $\sum_{n=1}^{\infty} a_n = \infty$, $\sum a_n^2 < \infty$, $\sum_{n=1}^{\infty} a_n^2 c_n^{-2} < \infty$

Conditions on the functions $g_n(x)$ and v_n

G1 : There exists a sequence of positive number γ_n independent of x and y such that

$$\|g_n(x) - g_n(x)\| \le \gamma_n \|x - y\| \quad \text{for all } x,y\in R^k.$$

G2 : $\sum_{n=1}^{\infty} (\gamma_n^2 - 1)^+ < \infty$; where z^+ means $(x + |z|)/2$

G3 : $\lim_{n\to\infty} g_n(x) - g_n(y) - v_n$ exists for all $\|x - y\| < \infty$

G4 : $\sum_{n=1}^{\infty} \|v_n\| < \infty$.

4. ASYMPTOTIC CONVERGENCE

Theorem 4.1. If the conditions M1, M2, A1, G1-G4 hold. Then

$$\lim_{n\to\infty} \|X_n - \theta_n\| = 0$$

Proof :- From (2.1) and (2.2) we have

$$x_{n+1} - \theta_{n+1} = g_n(x_n) - g_n(\theta_n) - v_n - a_n F_n$$

Thus

$$\|x_{n+1} - \theta_{n+1}\|^2 = \|g_n(x_n) - g_n(\theta_n) - v_n\|^2 - 2a_n\langle g_n(x_n) - g_n(\theta_n) - v_n, F_n\rangle + a_n^2\|F_n\|^2,$$

$$= \|g_n(x_n) - g_n(\theta_n)\|^2 - 2\langle g_n(x_n) - g_n(\theta_n), v_n\rangle + + \|v\|^2 - 2a_n\langle \overset{*}{x}_n - \theta_{n+1}, F_n\rangle + a_n^2\|F_n\|^2$$

Using (2.4) we obtain

$$E_n(\|x_{n+1} - \theta_{n+1}\|^2) = \|g_n(x_n) - g_n(\theta_n)\|^2 - 2\langle g_n(x_n) - g_n(\theta_n), v_n\rangle$$

$$+ \|v_n\|^2 - 2a_n\langle \overset{*}{x}_n - \theta_{n+1}; \overline{D}_n\rangle + a_n^2 E_n(\|F_n\|^2)$$

$$\leq \|g_n(x_n) - g_n(\theta_n)\|^2 - 2a_n\langle \overset{*}{x}_n - \theta_{n+1}; \overline{D}_{n+1}(\overset{*}{x}_n)\rangle$$

$$+ 0(\|v_n\|\|g_n(x_n) - g_n(\theta_n)\|) + 0(\|v_n\|^2$$

$$+ 0(a_n\|g_n(x_n) - g_n(\theta_n) - v_n\|\|\overline{D}_n - D_{n+1}(\overset{*}{x}_n)\|^2)$$

$$+ 0(a_n^2\|D_{n+1}(\overset{*}{x}_n)\|^2) + 0(a_n^2\|\overline{D}_n - D_{n+1}(\overset{*}{x}_n)\|^2)$$

$$+ 0(a_n^2\|E_n(d_n^2)\|) \qquad (4.1)$$

By (2.5), 92.6) and M1 and using the identity $\|X\| \leq \|x\|^2 + 1$, we obtain

$$E_n(\|x_{n+1} - \theta_{n+1}\|^2) = \|g_n(x_n) - g_n(\theta_n)\|^2 - (1 + 0\|v_n\|) + 0(a_n c_n)$$

$$+ 0(a_n^2)) - 2a_n\langle \overset{*}{x}_n - \theta_{n+1}; D_{n+1}(\overset{*}{x}_n)\rangle$$

$$+ 0(\|v_n\| + \|v_n\|^2 + a_n c_n + \|v_n\| a_n c_n + a_n^2)$$

$$= \|g_n(x_n) - g_n(\theta_n)\|^2 (1 + \mu_n)$$

$$- 2a_n\langle \overset{*}{x}_n - \theta_{n+1}; D_{n+1}(\overset{*}{x}_n)\rangle + \delta_n \qquad (4.2)$$

where

$$\mu_n = 0(\|v_n\| + a_n c_n + a_n^2)$$

$$\delta_n = 0(\mu_n + a_n^2 c_n^{-2})$$

From A1 and G4 it follows that

$$\sum_{n=1}^{\infty} \mu_n < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \delta_n < \infty \qquad (4.3)$$

Using G1, (4.2) can be written as

$$E_n(\|x_{n+1} - \theta_{n+1}\|^2) \le \|x_n - \theta_n\|^2 - (1 + \mu_n + (\gamma_n^2 - 1^+)$$

$$+ 2a_n\langle \overset{*}{x}_n - \theta_{n+1};\ D_{n+1}(\overset{*}{x}_n)\rangle + \delta_n \qquad (4.4)$$

From M2, G2, (4.3) and (4.4), it follows from Theorem 1 Robbins and Siegmund [8] that

$$\lim_{n \to \infty} \|x_n - \theta_n\| \text{ exists and finite}$$

and

$$\sum_{n=1}^{\infty} a_n\langle \overset{*}{x}_n - \theta_{n+1};\ D_{n+1}(\overset{*}{x}_n)\rangle < \infty . \qquad (4.5)$$

Using M2, A1 and G3 obtain that

$$\lim_{n \to \infty} \|x_n - \theta_n\| = 0,$$ which completes the proof of theorem.

5. ASYMPTOTIC NORMALITY OF THE PROCEDURE

Theorem 5.1. Let $A, P \in R^{k\times k}$, A positive definite, P orthogonal,

$P'AP = \Lambda$ diagonal, $\lambda = \min \Lambda^{(ii)}$, $0 < \beta < 2\lambda a$,

$$a_n = a\, n^{-1}, \quad c_n = cn^{-\gamma}, \quad \gamma = \frac{1}{2}(1 - \beta); \qquad (5.1)$$

$$X_n - \theta_n \longrightarrow 0 \quad C > \|E_n(d_n d'_n) - \Sigma\| \longrightarrow 0,$$

$$n^{-1}\sum_{j=1}^{n} \sigma_{j,r}^2 \longrightarrow 0 \text{ for every } r > 0 \text{ with}$$

$$\sigma_{j,r}^2 = E_x\left\{\|d_j\|^2 \ge r\, j^{\alpha}\right\} \|d_j\|^2 \text{ and}$$

let for X_n in a neighborhood of θ_n

$$\|\bar{D}_n - A(X_n - \theta_{n+1}) - n^{-\beta/2} m\| \le 0(1)\left[n^{-\beta/2} + \|\overset{*}{X}_n - \theta_{n+1}\|\right] \qquad (5.2)$$

$$\|g_n(x) - g_n(\theta_n) - (X_n - \theta_n)\| \le 0(1)\left[n^{-\beta/2} + n^{-1}\|X_n - \theta_n\|\right] \qquad (5.3)$$

Then the asymptotic distribution of $n^{\beta/2}(X_n - \theta_n)$ is normal with mean $-a(aA - (\beta/2)\ I)^{-1}m$ and covariance matrix PMP' with $M^{(ij)} = a^2\ c^{-2}[P'\ \Sigma\ P]^{(ij)} / (a\ \Lambda^{(ij)} + a\ \Lambda^{(ij)} - \beta)$

Proof : From (2.1) (2.2) and (2.3) we have

$$X_{n+1} - \theta_{n+1} = \overset{*}{X}_n - \theta_{n+1} - an^{-1}\ D_n - an^{-1}d_n \tag{5.4}$$

From (2.5) it follows that

$$\overline{D}_n = A_n(\overset{*}{X}_n - \theta_{n+1})n^{-\beta/2}\ m_n$$

with A_n, m_n are $\mathcal{F}_n$-measurable and $A_n \longrightarrow A$, $m_n \longrightarrow m$ uniformally. Thus (5.4) can be written as

$$X_{n+1} - \theta_{n+1} = (\overset{*}{X}_n - \theta_{n+1})(I - an^{-1}\ A_n) - an^{-1}\ n^{-\beta/2}\ m_n - an^{-1}d_n \tag{5.5}$$

Also it follows from (5.3) that

$$\overset{*}{X}_n - \theta_{n+1} = G_n(X_n - \theta_n) + n^{-\beta/2}\ g_n \tag{5.6}$$

where

$$G_n \longrightarrow I$$
$$g_n \longrightarrow 0$$

uniformally, $\|G_n - I\| = 0(n^{-1})$.

Substituting (5.6) in (5.5) we get

$$X_{n+1} - \theta_{n+1} = (I - an^{-1}(A_n + 0(1))(X_n - \theta_n) - an^{-1}\ n^{-\beta/2}(m_n + 0(1)) - an^{-1}d_n \tag{5.7}$$

From (5.7) and Theorem (2.2) Fabian [5], giving the desired result.

REFERENCES

[1] Doob, J.L. (1953), Stochastic processes. Wiley, New York.

[2] Dupac, V. (1965), A dynamic stochastic approximation method. Annal. Math. Statist., 1695-1702.

[3] Dupac, V. (1966), Stochastic aproximation in the presence of the trend. Czechosloak Mathematical Jounal,16, 454-461.

[4] Fabian, V. (1968 a), On the choice of design in stochastic approximation methods. Ann. Math. Statist. 39(457-465).

[5] Fabian, V. (1968 b), On asymtotic normality in stochastic approximation. Ann. Math. Statist. 39 1327-1332.

[6] Kiefer, J. and Wolfowitz, J. (1952), Stochastic estimation of the maximum of a regression function, Ann. Math. Statist. 23, 462-466.

[7] Robbins, H. and Monro (1951), A stochastic approximation method. Ann, Math. Statist. 22, 400-407.

[8] Robbins, H. and Siegmund, D. (1971), A convergence theorem for non negative almost super martingles and some applications. In optimizing Methods in statistics (J.S. Rustagi, ed.) 233-257.

[9] Rupper, D. (1985), A Newton-Raphson Version of the multivariate Robbins-Monro procedure. Ann. Statist. 13, 236-245.

[10] Sorour, E. (1978), On the convergence of the dynamic stochastic approximation method for stochastic non-linear multidimensional dynamic dystems. Kybernetika, Volume 14, 28-37.

[11] Sorour, E. (1991), A dynamic Kiefer-Wolfowitz stochastic approximation procedure. Statistics & Decision 9, 201-211.

[12] Sorour,E. (1993), Dynamic multivariate stochastic approximation using simultaneous perturbation gradient approximation. Proceedings of Eighteenth (18) international conference for statistics and computer science and applications (17-22 April), pages ( )

[13] Uosaki, K. (1974), Some generalization of the dynamic stochastic approximation process. Ann. Statist. 2, 1042-1048.