MILITARY TECHNICAL COLLEGE

CAIRO - EGYPT

A NEWTON-RAPHSON VERSION OF THE MULTIVARIATE DYNAMIC ROBBINS-MONRO PROCEDURE

El Sayed Sorour

Military Technical College,
Department of Mathematics.

ABSTRACT

Let M be a function from R^k to R^k, let θ_n, $n = 1,2,\ldots$ be (unknown) vector numbers, the first θ_1 being the unique root of the equation $M(x) = 0$, set $M_1(x) = M(x)$, for $n = 1,2,\ldots$ set $M_n(x) = M(x - \theta_n - \theta_1)$ so that θ_n is the unique root of $M_n(x)=0$. Initially $M_n(x)$ is unknown, but for any x in R^k we can observe a random vector $Y_n(x)$ with conditional expectation $M_{n+1}(x)$. The unknown θ_n can be estimated recursively by the author (1978), that procedure requires the rather restrictive assumption that the infimum of the inner product $\langle x - \theta, M(x)\rangle$ over any compact set not containing θ be positive, i.e. along each line through θ_1, M(x) is unimodal with minimum θ_1. Unlike our previous method, the procedure introduced in this paper does not necessarily attempt to move in the direction of θ_n but except of that random fluctuations it moves in the direction which decreases $\|M_n(x)\|^2$, consequently it does not require that $\langle x - \theta_n, M_n(x)\rangle$ have a constant signum. This new procedure is a stochastic analog of the Newton-Raphson technique.

1. INTRODUCTION

This paper is concerned with the multivariate dynamic version of the Robbins-Monro [3] stochastic approximation procedure. This problem first studied by Dupac [1]; where the root of the regression function moves in a specified manner. He discussed in his papers [1]; [2] only the cases where the movement of the root (one-dimensional) or the maximum (multidimensional) can be expressed by a certain linear function of its present location, and where the trend is deterministic. Uosaki, K. [9] discussed some generalization of Dupac's work in the one-dimensional case where the movement of the root can be expressed by a specified non-linear function of its present location. Uosaki's result has been generalized to the multidimensional case by the author (1978). This version begins with an initial estimate X_1. Given X_1, $X_2, \ldots, X_n$ one observes Y_n, such that $E_n(Y_n) = M_{n+1}(\overset{*}{X}_n)$, where E_n denotes the conditional expectation given X_1, $X_2, \ldots, X_n$ and $\overset{*}{X}_n = g_n(X_n)$ for some function g_n from $R^k \longrightarrow R^k$. Then X_{n+1} is defined by

$$X_{n+1} = \overset{*}{X}_n - a_n Y_n \quad \ldots\ldots \tag{1.1}$$

where $\{a_n\}$ is a suitably choosen positive sequence converging to 0 as $n \longrightarrow \infty$. Let θ_n be the unique root of $M_n(x)$. Then $X_n - \theta_n \longrightarrow 0$, under the assumption for every $\varepsilon > 0$

$$\inf_{n\in N} \inf_{\|X-\theta_n\|>\varepsilon} \frac{\langle x - \theta_n, M_n(x)\rangle}{\|x - \theta_n\|} > 0. \tag{1.2}$$

In fact it can be proved that (1.2) can be replaced by a weaker one, for every $\varepsilon > 0$

$$\inf_{n\in N} \inf_{\varepsilon<\|X-\theta_n\|<\varepsilon} \langle x - \theta_n, M_n(x)\rangle > 0 \tag{1.2'}$$

The importance of (1.2)' can be easily seen as follows. Suppose that $\text{Sup}\left\{\text{Var } Y : x\in R^k\right\} < \infty$. Thus from (1.1) we have

$$E_n(\|X_{n+1} - \theta_{n+1}\|^2) \le \|X_n - \theta_n\|^2(1 + \mu_n)$$

$$- 2a_n(1 + O(1)) < \overset{*}{X}_n - \theta_{n+1}, M_{n+1}(\overset{*}{X}_n)> + v_n$$

where $\Sigma \mu_n < \infty$; $\Sigma v_n < \infty$. Using (1.2)' and theorem 1 of Robbins and Sigmund [4] $\|X_n - \theta_n\|$ converges to 0. Unfortunately, (1.2)' is a rather restrictive assumption, amplying that for each x, $M_n(x)$ points a-way from θ_n. There are many practical examples for functions do not astisfy the condition (1.2)'. An alternative procedure would to apply the multivariate Kiefer-Wolfowitz (KW) to minize $\|M_n(x)\|^2$ a fact used by Ruppert (1985) for the ordinary (RM) procedure. We assume that $E_n(\|Y_n\|^2) = \|M_n(x)\|^2 + \text{const}$. Thus the (KW) procedure does not attempt to move towards θ_n; but in a direction of decreasing $\|M_n(x)\|^2$. If θ_n is the only local minimum of $M_n(x)$, we prove that $x_n - \theta_n \longrightarrow 0$ under mild conditions.

2. NOTATIONS AND ASSUMPTIONS

2.1 : Let R^k be the k-dimensional vector space. For x in R^k let x^i be the ith component of x. For x,y in R^k we define

$$\langle x,y\rangle = \sum_{i=1}^{k} x^i y^i \text{ and } \|x\|^2 = \langle x,x\rangle.$$

If A is a matrix of order $k\times\ell$ let A^{ij} be the entry of A. Also, let A^t be the transpose of A, and let

$$\|A\|^2 = \sum_{i=1}^{k} \sum_{i=1}^{\ell} |A^{ij}|^2.$$

2.2 : Let $M(x)$, $x \in R^k$ be an (unknown) twice differentiable function from $R^k \longrightarrow R^k$. Let $\theta_n (n = 1,2,\ldots)$ be (unknown) vectors in R^k. The first θ_1, being the unique root of the equation $M(x) = 0$,. Set $M_1(x) = M(x)$; for $n = 1,2,\ldots$ set $M_n(x) = M(x - \theta_n + \theta_1)$ so that θ_n is the unique root of $M_n(x)$.

2.3 : Let $D(x)$ be the derivative of $M(x)$, i.e., $D^{ij} = (\partial/\partial x_j) M^i(x)$ and assume the following assumptions on D

i) $D(\theta_1)$ is non singular

ii) For all $\varepsilon > 0$

$$\inf\left\{\|D^t(x)\ M(x)\| \ : \ \varepsilon < \|M(x)\| \leq \varepsilon^{-1}\right\} > 0$$

iii) $\sup\left\{\|D(x)\| \ : \ x \in R^k\right\} < \infty$.

2.4 : Let $H(x)$ be the Hessian of $\|M(x)\|^2$, i.e.

$H^{ij} = (\partial^2/\partial x^i\ \partial x^j \|M(x)\|^2$. and assume that

$$\sup\left\{\|H(x)\| \ : \ x \in R^k\right\} < \infty .$$

2.5 : Assume that θ_n moves in a such manner that

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

where $g_n(x)$ is in general a non-linear measurable function (known) from $R^k \longrightarrow R^k$ and v_n is an unknown (random or nonrandom) k-vector function independent of x and

$$\|v_n\| = 0(\delta_n), \quad \sum_{n=1}^{\infty} \delta_n < \infty . \qquad (2.2)$$

2.6 : For x and y in R^k, we assume that exists a sequence of positive numbers $\{\gamma_n\}$ independent of x and y and let

$$Z_n = x - y, \quad \text{and} \quad \overset{\star}{Z}_n = g_n(x) - g_n(y)$$

Then

$$\|M(\overset{\star}{Z}_n)\|^2 \leq \gamma_n \|M(Z)\|^2 \qquad (2.3)$$

$$\sum_{n=1}^{\infty} (\gamma_n - 1)^+ < \infty . \tag{2.4}$$

where Z^+ means $(Z + |Z|)/2$.

and

$$\lim_{n \to \infty} M(\overset{*}{Z}_n) = M(Z) \qquad \text{for } \|Z\| < \infty . \tag{2.5}$$

3. THE PROCEDURE

The dynamic Robbins-Monro procedure will be described formally by the following assumptions.

3.1 : Let X_1 be arbitrary; define

$$X_{n+1} = \overset{*}{X}_n - an^{-1} F_n^t Y_n \tag{3.1}$$

where

$\overset{*}{X}_n = G_n(X_n)$; $a > 0$; X_n in R^k and F_n is k×k random matrix, is used to estimate $D_{n+1}(\overset{*}{x}_n)$ and Y_n is the observation with conditional expectation equal to $M_{n+1}(\overset{*}{x}_n)$. The ith column of F_n is constructed as follows : Let $e(i)$ be the ith column of the k×k identity matrix. Let $c_n > 0$ be constant and let $Y_n(i,2)$ and $Y_n(i,1)$ each be the observation with conditional expectation equal to $M_{n+1}(\overset{*}{x}_n + c_n\, e(i))$ and $M_{n+1}(\overset{*}{x}_n - c_n\, e(i))$, respectively. Then, the ith column of F_n is

$$F_n^i = \left[Y_n(i,2) - Y_n(i,1)\right]/2\, c_n \tag{3.2}$$

Let $\mathcal{F}_n$ be the σ-algebra generated by $X_1, X_2, \ldots, X_n$. For any random vector X in R^k, let $E_n(X)$ and $Var_n(X)$ be respectively the conditional mean and the variance of X given $\mathcal{F}_n$.

Given $\mathcal{F}_n$, Y_n, D_n and v_n are conditionally independent. Let

$$\xi_n = Y_n - M_{n+1}(x_n); \quad E_n(\xi_n) = 0 \tag{3.3}$$

Let

$$\overline{F}_n = E_n(F_n) \tag{3.4}$$

$$d_n = F_n - \overline{F}_n \tag{3.5}$$

and assume that

$$E_n(\|\xi_n\|^2) \leq \delta^2 < \infty \,, \tag{3.6}$$

$$E_n(\|d_n\|^2) \leq k\ c_n^{-2} \tag{3.7}$$

and

$$\|\overline{F}_n - D_{n+1}(x_n)\| \leq k\ c_n \tag{3.8}$$

3.2 : $c_n > 0$; $c_n \downarrow 0$ and assume tha

$$\sum_{n=1}^{\infty} n^{-1}\ c_n < \infty \,, \qquad \sum_{n=1}^{\infty} n^{-2}\ c_n^{-2} < \infty \,, \tag{3.9}$$

4. THE MAIN RESULT

Theorem 4.1. : If the assumptions 2.1-3.2 hold. Then $X_n - \theta_n \longrightarrow 0$.

Proof : Using (2.1) and (3.1) we obtain

$$X_{n+1} - \theta_{n+1} = g_n(X_n) - g_n(\theta_n) - V_n - an^{-1}\ F_n^t\ Y_n. \tag{4.1}$$

Let

$$Z_{n+1} = X_{n+1} - \theta_{n+1}, \quad \overset{*}{Z}_n = g_n(X_n) - g_n(\theta_n) \quad \text{and} \quad \Delta_n = an^{-1}\ F_n^t\ Y_n$$

Then (4.1) can be written as

$$Z_{n+1} = \overset{*}{Z}_n - V_n\ \Delta_n \tag{4.2}$$

and there is η in (0,1) such that

$$\|M(Z_{n+1})\|^2 = \|M(\overset{*}{Z}_n)\|^2 - 2\langle V_n + \Delta_n,\ D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\rangle + \frac{1}{2}\langle V_n + \Delta_n,$$
$$H(\overset{*}{Z}_n - \eta(V_n + \Delta_n))(V_n + \Delta_n)\rangle.$$

Using 2.3 and (3.3)-(3.5) we get

$$E_n(\|M_2(Z_{n+1})\|^2) = \|M(\overset{*}{Z}_n)\|^2 - 2E_n(\Delta_n), D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)> -2<E_n(V_n), D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)> + \frac{1}{2} E_n\left[<V_n + \Delta_n, H(\overset{*}{Z}_n - \eta(V_n + \Delta_n))(V_n + \Delta_n)>\right]$$

$$\leq \|M(\overset{*}{Z}_n)\|^2 - 2an^{-1}\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2 - 2an^{1-}<\left[F_n^{-t} - D^t(Z_n)\right]M(\overset{*}{Z}_n), D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)> + 0(\|V_n\|)\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|)+ 0\,(n^{-2}E_n(\|F_n^t\ Y_n\|^2)) + 0(\|V_n\|^2) + 0(\|V_n\|E_n\|F_n^t\ Y_n\|). \quad (4.3)$$

Hence

$$E_n(\|M(Z_{n+1})\|^2) = \|M(\overset{*}{Z}_n)\|^2 - 2an^{-1}\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2 + \sum_{i=1}^{5} T_i \quad (4.4)$$

where T_i $i = 1,2,\ldots,5$ are the corresponding, terms respectively in (4.3) By (3.8) we have

$$|T_1| = 0(n^{-1}\ c_n\|M(\overset{*}{Z}_n)\|^2) \quad (4.5)$$

$$|T_2| = 0(\|V_n\|\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2) + \|V_n\| \quad (4.6)$$

By (3.3) - (3.8) we have

$$E_n(\|F_2{}^tY_n)\|^2) = \|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2 + M^t(\overset{*}{Z}_n)\left[F_n\ F_n^{-t} - D(Z_n)D^t(Z_n)\right]M(\overset{*}{Z}_n) + E_n\left[\|d_n^t\ \xi_n\|^2\right] + E_n\left[\|F_n^{-t}\ \xi_n\|^2\right]$$

$$\leq \|D^t(Z_n)M(Z_n)\|^2 + 0\left[(c_n + c_n^{-2})\|M(\overset{*}{Z}_n)\|^2\right] + 0(1 + c_n + c_n^{-2}).$$

Thus

$$|T_3| = 0(n^{-2}\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2) + 0(n^{-2}c_n + n^{-2}c_n^{-2})\|M(\overset{*}{Z}_n)\|^2 + 0(n^{-2} + n^{-2}c_n + n^{-2}c_n^{-2}) \quad (4.7)$$

From which it follows also that

$$|T_5| = 0(\|V_n\|\ E_n\|\Delta_n\|) = 0(\|V_n\| + \|V_n\|\ |T_3|) \quad (4.8)$$

Substituting (4.5)-(4.8) in (4.5) and using (2.3) we get

$$E_n(\|Z_{n+1}\|^2) \leq \|M(Z_n)\|^2(1+\mu_n) - 2an^{-1}(1+0(1))\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2+\varepsilon_n \tag{4.9}$$

where

$$\mu_n = 0(n^{-1}c_n + n^{-2}c_n^{-2} + \gamma_n - 1)^+ \tag{4.10}$$

and

$$\varepsilon_n = 0(\|V_n\| + n^{-2}c_n^{-2}) \;\ldots\ldots \tag{4.11}$$

From (2.2), (2.4) and (3.9) it follows that $\sum \mu_n < \infty$ and $\sum \varepsilon_n < \infty$. Therefore by Therem 1 of Robbins and siegmund (1971) $\lim\limits_{n\to\infty} M(\overset{*}{Z}_n)$ exists and is finite and

$$\sum_{n=1}^{\infty} n^{-1}\|D^t(\overset{*}{Z}_n)M(\overset{*}{Z}_n)\|^2 < \infty$$

By 2.2(ii) and (2.5) $X_n - \theta_n \longrightarrow 0$, which completes the proof of the theorem.

REFERENCES

[1] Dupac, V. (1965), Adynamic stochastic approximation method. The Anal. Math. Statist. 36, 1695-1702.
[2] Dupac, V. (1966), Stochastic aproximation in the presence of the trend. Czechosloak Mathematical Jounal,16, 454-461.
[3] Robbins, H. and Monoro, S. (1951), A stochastic approximation method. Ann. Math. Statist. 22, 400-407.
[4] 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, Academic, New York.
[5] Ruppert, D. (1985), A Newton-Raphson version of the multivariate Robbins-Monoro procedure. Ann Statist. 13, 236-245.
[6] Sorour, E. (1978), On the convergence of the dynamic stochastic approximation method for stochastic non-linear multidimensional dynamic dystems. Kybernetika, Volume 14, 28-37.
[7] Sorour, E. (1991), A dynamic Kiefer-Wolfowitz stochastic approximation procedure. Statistics & Decision 9, 201-211.
[8] Sorour, E. (1993), Dynamic multivariate stochastic approximation using a simultaneous perturbation gradient approximation.(appear).
[9] Uosaki, K. (1974), Some generalization of the dynamic stochastic approximation process. Ann. Statist. 2, 1042-1048.