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 $\mathbb{R}^k$ to $\mathbb{R}^k$, let $\theta_n$, $n = 1, 2, \ldots$ be (unknown) vector numbers, the first $\theta_1$ 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 $\theta_n$ is the unique root of $M_n(x) = 0$. Initially $M_n(x)$ is unknown, but for any $x$ in $\mathbb{R}^k$ we can observe a random vector $Y_n(x)$ with conditional expectation $M_{n+1}(x)$. The unknown $\theta_n$ 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 $\theta$ be positive, i.e. along each line through $\theta_1$, $M(x)$ is unimodal with minimum $\theta_1$. Unlike our previous method, the procedure introduced in this paper does not necessarily attempt to move in the direction of $\theta_n$ 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-Henro [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 $X_1, X_2, ..., X_n$ one observes $Y_n$, such that $E_n(Y_n) = M_{n+1}(X_n)$, where $E_n$ denotes the conditional expectation given $X_1, X_2, ..., X_n$ and $\hat{X}_n = g_n(X_n)$ for some function $g_n$ from $R^k \rightarrow R^k$. Then $X_{n+1}$ is defined by

$$X_{n+1} = \hat{X}_n - a_n Y_n \quad \ldots \ldots \quad (1.1)$$

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

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

In fact it can be proved that (1.2) can be replaced by a weaker one, for every $\varepsilon > 0$.
The importance of (1.2)' can be easily seen as follows. Suppose that \( \text{Sup} \{\text{Var } Y : x \in \mathbb{R}^k\} < \infty \). Thus from (1.1) we have

\[
E_n(\|X_{n+1} - \theta_{n+1}\|^2) \leq \|X_n - \theta_n\|^2 (1 + \mu_n) - 2a_n(1 + O(1)) \langle X_n - \theta_{n+1}, M_{n+1}(X_n) \rangle + v_n
\]

where \( \sum \mu_n < \infty \); \( \sum v_n < \infty \). 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, implying that for each \( x \), \( M_n(x) \) points away from \( \theta_n \). There are many practical examples for functions do not satisfy the condition (1.2)'. An alternative procedure would be to apply the multivariate Kiefer-Wolfowitz (KW) to minimize \( \|M_n(x)\|^2 \) a fact used by Ruppert (1985) for the ordinary (PM) 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 \( \theta_n \); but in a direction of decreasing \( \|M_n(x)\|^2 \). If \( \theta_n \) is the only local minimum of \( M_n(x) \), we prove that \( x_n - \theta_n \to 0 \) under mild conditions.

2. NOTATIONS AND ASSUMPTIONS

2.1: Let \( \mathbb{R}^k \) be the \( k \)-dimensional vector space. For \( x \) in \( \mathbb{R}^k \) let \( x^i \) be the \( i \)-th component of \( x \). For \( x,y \) in \( \mathbb{R}^k \) we define

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

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

\[
\|A\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} |A_{ij}|^2.
\]
2.2: Let $M(x), x \in \mathbb{R}^k$ be an (unknown) twice differentiable function from $\mathbb{R}^k \to \mathbb{R}^k$. Let $\theta_n (n = 1, 2, \ldots)$ be (unknown) vectors in $\mathbb{R}^k$. The first $\theta_1$, 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 $\theta_n$ is the unique root of $M_n(x)$.

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

i) $D(\theta_1)$ is nonsingular

ii) For all $\varepsilon > 0$

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

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

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

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

2.5: Assume that $\theta_n$ moves in a such manner that

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

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

$$\|v_n\| = O(\delta_n), \sum_{n=1}^{\infty} \delta_n < \infty.$$

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

$$Z_n = x - y, \quad \text{and} \quad Z_n^* = g_n(x) - g_n(y)$$

Then

$$\|M(Z_n^*)\|^2 \leq \gamma_n \|M(Z)\|^2$$

(2.3)
\[ \sum_{n=1}^{\infty} (\gamma_n - 1)^+ < \infty. \]  
(2.4)

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

and

\[ \lim_{n \to \infty} M(Z_n) = M(Z) \quad \text{for} \quad \|Z\| < \infty. \]  
(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

\[ X_{n+1} = X_n - a_n F_n^T Y_n \]  
(3.1)

where

\[ X_n = G_n(X_n); \quad a > 0; \quad X_n \in \mathbb{R}^k \] and \( F_n \) is \( k \times k \) random matrix, is used to estimate \( D_{n+1}(x_n) \) and \( Y_n \) is the observation with conditional expectation equal to \( M_{n+1}(x_n) \).

The \( i \)th column of \( F_n \) is constructed as follows: Let \( e(i) \) be the \( i \)th column of the \( k \times k \) 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 \( M_{n+1}(x_n + c_n e(i)) \) and \( M_{n+1}(x_n - c_n e(i)) \), respectively. Then, the \( i \)th column of \( F_n \) is

\[ F_n^i = \frac{Y_n(i,2) - Y_n(i,1)}{2 c_n} \]  
(3.2)

Let \( \mathcal{F}_n \) be the \( \sigma \)-algebra generated by \( X_1, X_2, \ldots, X_n \). For any random vector \( X \) in \( \mathbb{R}^k \), let \( F_n(X) \) and \( \text{Var}_n(X) \) be respectively the conditional mean and the variance of \( X \) given \( \mathcal{F}_n \).
Given $\xi_n$, $Y_n$, $D_n$ and $\nu_n$ are conditionally independent. Let

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

Let

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

$$d_n = F_n - \overline{F}_n \quad (3.5)$$

and assume that

$$E_n(\|\xi_n\|^2) \leq \delta^2 < \infty, \quad (3.6)$$

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

and

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

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

$$\sum_{n=1}^{\infty} n^{-1} c_n < \infty, \quad \sum_{n=1}^{\infty} n^{-2} c_n^{-2} < \infty, \quad (3.9)$$

4. THE MAIN RESULT

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

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

$$X_{n+1} - \Theta_{n+1} = g_n(X_n) - g_n(\Theta_n) - \nu_n - \alpha_n \overline{F}_n Y_n. \quad (4.1)$$

Let

$$Z_{n+1} = X_{n+1} - \Theta_{n+1}, \quad \hat{z}_n = g_n(X_n) - g_n(\Theta_n) \quad \text{and} \quad \Delta_n = \alpha_n \overline{F}_n Y_n$$

Then (4.1) can be written as

$$Z_{n+1} = \hat{z}_n - \nu_n \Delta_n. \quad (4.2)$$

and there is $\eta$ in $(0,1)$ such that

$$\|M(Z_{n+1})\|^2 = \|\hat{z}_n\|^2 - 2\nu_n + \Delta_n, \quad D^t(\hat{z}_n)M(\hat{z}_n) + \frac{1}{2} \nu_n + \Delta_n,$$

Using 2.3 and (3.3)-(3.5) we get
\[ E_n(\|M(Z_{n+1})\|^2) = \|M(Z_n^*)\|^2 - 2E_n(\Delta_n) , \quad D^t(Z_n^*)M(Z_n^*) > -2E_n(V_n) , \]
\[ D^t(Z_n^*)M(Z_n^*) > + \frac{1}{2} E_n[<V_n + \Delta_n, H(Z_n - \eta(V_n + \Delta_n))(V_n + \Delta_n)>] \]
\[ \leq \|M(Z_n^*)\|^2 - 2an^{-1}\|D^t(Z_n^*)M(Z_n^*)\|^2 - 2an^{-1}<[F_n - D^t(Z_n^*)]\]
\[ D^t(Z_n^*)M(Z_n^*) + D^t(Z_n^*)M(Z_n^*) > + O(\|V_n\|)\|D^t(Z_n^*)M(Z_n^*)\| + O(n^{-2}E_n(\|F_n V_n\|^2)) + O(\|V_n\|^2) + O(\|V_n\|E_n\|F_n^t V_n\|). \]

Hence
\[ E_n(\|M(Z_{n+1})\|^2) = \|M(Z_n^*)\|^2 - 2an^{-1}\|D^t(Z_n^*)M(Z_n^*)\|^2 + \sum_{i=1}^{5} T_i \quad (4.4) \]

where \( T_i \quad i = 1, 2, \ldots, 5 \) are the corresponding terms respectively in (4.3). By (3.8) we have
\[ |T_1| = O(n^{-1}c_n\|M(Z_n^*)\|^2) \quad (4.5) \]
\[ |T_2| = O(\|V_n\|\|D^t(Z_n^*)M(Z_n^*)\|^2) + \|V_n\| \quad (4.6) \]

By (3.3) - (3.8) we have
\[ E_n(\|F_n^t V_n\|^2) = \|D^t(Z_n^*)M(Z_n^*)\|^2 + M^t(Z_n^*)[F_n F_n^t - D(Z_n^*)D^t(Z_n^*)]M(Z_n^*) \]
\[ + E_n[\|d_n t \xi_n\|^2] + E_n[\|F_n^t \xi_n\|^2] \]
\[ \leq \|D^t(Z_n^*)M(Z_n^*)\|^2 + 0\left[(c_n + c_n^{-2})\|M(Z_n^*)\|^2\right] \]
\[ + O(1 + c_n + c_n^{-2}). \]

Thus
\[ |T_3| = O(n^{-2}\|D^t(Z_n^*)M(Z_n^*)\|^2) + O(n^{-2}c_n + n^{-2}c_n^{-2})\|M(Z_n^*)\|^2 \]
\[ + O(n^{-2} + n^{-2}c_n + n^{-2}c_n^{-2}) \]

From which it follows also that
\[ |T_5| = O(\|V_n\|E_n\|\Delta_n\|) = O(\|V_n\| + \|V_n\| |T_3|) \quad (4.8) \]
Substituting (4.5)-(4.8) in (4.5) and using (2.3) we get

\[ E_n\left(\|Z_{n+1}\|^2\right) \leq \|M(Z_n)\|^2(1 + \mu_n) - 2an^{-1}(1 + O(1))\|D^t(Z_n^*)M(Z_n^*)\|^2 \epsilon_n \]

where

\[ \mu_n = O(n^{-1}c_n + n^{-2}c_n^{-2} + \gamma_n^{-1}) \]

and

\[ \epsilon_n = O(\|V_n\| + n^{-2}c_n^{-2}) \]

From (2.2), (2.4) and (3.9) it follows that \( \sum \mu_n < \infty \) and \( \sum \epsilon_n < \infty \).

Therefore by Theorem 1 of Robbins and Siegmund (1971) \( \lim_{n \to \infty} M(Z_n^*) \) exists and is finite and

\[ \sum_{n=1}^{\infty} n^{-1}\|D^t(Z_n^*)M(Z_n^*)\|^2 < \infty \]

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

REFERENCES