Václav Dupač

Notes on stochastic approximation methods

_Czechoslovak Mathematical Journal_, Vol. 8 (1958), No. 1, 139–149

Persistent URL: [http://dml.cz/dmlcz/100283](http://dml.cz/dmlcz/100283)

Terms of use:

© Institute of Mathematics AS CR, 1958

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library [http://dml.cz](http://dml.cz)
NOTES ON STOCHASTIC APPROXIMATION METHODS

VÁCLAV Dupač, Praha

(Received March 11, 1957)

In section 1 and 2, asymptotic properties of the Robbins-Monro and the Kiefer-Wolfowitz stochastic approximation methods are studied under the assumption, that the solution lies in an a priori known finite interval. In section 3, a stochastic approximation method is considered for solving systems of linear equations with a symmetric matrix of coefficients.

0. Introduction and summary

Stochastic approximation methods deal with the following problems: 

\( M(x) \) is the (unknown) regression function of a family of random variables \( \{Y_x\} \); we have to solve the equation \( M(x) = \alpha \), or we have to find the value of \( x \) for which \( M(x) \) achieves its maximum, by means of an iterative process, using observations of \( Y \) on various levels of \( x \). The former problem has been solved by Robbins and Monro [1], the latter by Kiefer and Wolfowitz [2]; both problems — so as the methods of solution — have multidimensional analogs (Blum, [3]). The theoretical investigations of these methods go in two directions:

1° they are conditions studied, under which the approximations \( x_n \) converge to the solution \( \Theta \) with probability one;

2° the asymptotic order of second moments \( E[(x_n - \Theta)^2] \), or the asymptotic distribution of \( x_n \) are studied, and conclusions are drawn about the optimal choice of some eligible constants occurring in the approximation scheme.

In the second direction, the Chung's paper [4], concerning Robbins-Monro procedure, is the most advanced. Chung’s methods were adapted by Derman [5] and — independently — by the author [6] to derive asymptotic properties of Kiefer-Wolfowitz procedure.

The present paper contains two contributions to the investigations sub 2°. First it is shown, that the conditions under which the approximation procedure
has satisfactory asymptotic properties, can be considerably weakened, if the approximations $x_n$ are all restricted to a finite interval known as containing the solution $\Theta$. This is done in Section 1 for the Robbins-Monro procedure, and in Section 2 for the Kiefer-Wolfowitz method. Secondly, a multidimensional modification of the Robbins-Monro procedure is considered in a special case of linear regression with a symmetrical matrix. The upper bounds for the quantities $E[||x_n - \Theta||^2]$ are given (Section 3).

In the following, $K_0, K_1, \ldots, K_9$ are positive constants numbered in order of appearance. As $[f(x)]^2$ will be denoted the function

$$g(x) = \begin{cases} A, & \text{if } f(x) < A, \\ f(x), & \text{if } A \leq f(x) \leq B, \\ B, & \text{if } f(x) > B. \end{cases}$$

A lemma, due to Chung (Lemma 1 in [4]), will be used repeatedly:

Let $\{b_n\}, n \geq 1$, be a sequence of real numbers such that for $n \geq n_0$,

$$b_{n+1} \leq \left(1 - \frac{c}{n}\right)b_n + \frac{c_1}{n^{p+1}},$$

where $c > p > 0$, $c_1 > 0$. Then

$$b_n \leq \frac{c_1}{c - p} \cdot \frac{1}{n^p} + O \left(\frac{1}{n^{p+1}} + \frac{1}{n^c}\right).$$

1. The Robbins-Monro stochastic approximation method

Let to each value $x$ from a finite interval $\langle A, B \rangle$ correspond a distribution function $H(y \mid x)$, let $M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x)$ be a Borel measurable function bounded in $\langle A, B \rangle$. Suppose that the equation $M(x) = x$ has a unique root $x = \Theta$ in $(A, B)$, and that the inequality $(M(x) - x)(x - \Theta) > 0$ holds for all $x \neq \Theta, x \in \langle A, B \rangle$.

Let $a$ be a positive constant. Take $x_1 \in \langle A, B \rangle$ arbitrarily and for $n \geq 1$ set recursively

$$x_{n+1} = \left[x_n + \frac{a}{n} (x - y_n)\right]^B, \quad (1)$$

where $y_n$ is a random variable whose distribution function, for given $x_1, \ldots, x_n$, $y_1, \ldots, y_n$, is $H(y \mid x_n)$.

We shall add the following assumptions:

Assumption (I1): There exists a constant $a^2$, such that

$$\int_{-\infty}^{\infty} (y - M(x))^2 \, dH(y \mid x) \leq a^2 \quad \text{for all } x \in \langle A, B \rangle.$$
**Assumption (IIj):** For every $\delta > 0$ we have
$$\inf_{|x - \Theta| > \delta, x \in \langle A, B \rangle} |M(x) - \lambda| = K_0(\delta) > 0.$$  

**Assumption (III1):** We have $M'(\Theta) > 0$.

We derive a simple consequence of the assumptions. For a given $\eta$, $0 < \eta < 1$, let $\delta_0(\eta)$ be the supremum of all $\delta$, such that
$$|x_n - \Theta| \leq \delta \Rightarrow |M(x_n) - \lambda| \geq \eta M'(\Theta).$$  
The existence of such $\delta$'s follows from (III1). From (IIj) it follows that
$$|M(x_n) - \lambda| \geq K_0(\delta_0(\eta)) \geq \frac{K_0(\delta_0(\eta))}{B - A} |x_n - \Theta| \quad \text{for } |x_n - \Theta| > \delta_0(\eta),$$
since $|x_n - \Theta| \leq B - A$, by the definition of $x_n$. Now set
$$g(\eta) = \min \left( \eta M'(\Theta), \frac{K_0(\delta_0(\eta))}{B - A} \right) \quad \text{and} \quad K_1 = \sup_{0 < \eta < 1} g(\eta).$$
Evidently,
$$|M(x_n) - \lambda| \geq K_1|x_n - \Theta| \quad \text{holds for all } n = 1, 2, \ldots \quad (2)$$
Similarly, from (III1) and from the boundedness of $M(x)$ in $\langle A, B \rangle$ it follows that
$$|M(x_n) - \lambda| \leq K_2|x_n - \Theta| \quad \text{for all } n = 1, 2, \ldots \quad (3)$$

We shall denote the second moment $E[(x_n - \Theta)^2]$ as $b_n$.

**Theorem I.** Suppose that the assumptions (I1), (IIj) and (III1) are satisfied, and that $a > \frac{1}{2K_1}$. Then
$$b_n = O\left( \frac{1}{n} \right).$$

**Remark.** The choice of $a$ depends on the unknown constant $K_1$. We can avoid this fact by replacing the factor $\frac{a}{n}$ in (1) through $\frac{a'}{n} \log n$, where now $a'$ is an arbitrary positive constant. Then — under the same assumptions — $b_n = o\left( \frac{1}{n^{1-\varepsilon}} \right)$ for every $\varepsilon > 0$, as could be easily shown.

**Proof of the Theorem 1:** From (1) it follows
$$(x_n + 1 - \Theta)^2 = \begin{cases} (A - \Theta)^2 \text{ for } x_n - \Theta + \frac{a}{n} (\lambda - y_n) < A - \Theta, \\
(B - \Theta)^2 \text{ for } x_n - \Theta + \frac{a}{n} (\lambda - y_n) > B - \Theta, \\
(x_n - \Theta)^2 + \frac{a^2}{n^2} (y_n - \lambda)^2 - \frac{2a}{n} (x_n - \Theta)(y_n - \lambda) \quad \text{otherwise}. \end{cases}$$
If we square the inequalities \( x_n - \Theta + \frac{a}{n} (x - y_n) < A - \Theta \), or \( B - \Theta \)
respectively, and note that \( A - \Theta \) is negative, \( B - \Theta \) positive, we get

\[
(x_{n+1} - \Theta)^2 \leq (x_n - \Theta)^2 + \frac{a^2}{n^2} (y_n - x)^2 - \frac{2a}{n} (x_n - \Theta)(y_n - x) \quad (4)
\]

for all three possibilities.

Writing \( y_n - x = y_n - M(x_n) + M(x_n) - x \), taking conditional expectations
on both sides of (4), and using (I), we get

\[
E [(x_{n+1} - \Theta)^2 | x_n] \leq (x_n - \Theta)^2 + \frac{a^2}{n^2} \{ \sigma^2 + (M(x_n) - x)^2 \} - \frac{2a}{n} (x_n - \Theta)(M(x_n) - x);
\]

hence by (2) and (3)

\[
b_{n+1} \leq b_n + \frac{a^2}{n^2} (\sigma^2 + K_2^2 b_n) - \frac{2a}{n} K_1 b_n,
\]

i.e.,

\[
b_{n+1} \leq \left( 1 - \frac{2K_1 a + o(1)}{n} \right) b_n + \frac{\sigma^2 a^2}{n^2}.
\]

Hence by Chung’s lemma

\[
b_n \leq \frac{\sigma^2 a^2}{2K_1 a - 1} \cdot \frac{1}{n} + O \left( \frac{1}{n^2} + \frac{1}{n^2 K_1 a} \right).
\]

In order to prove the asymptotic normality of \( x_n \) we shall make further assumptions.

Assumption (IV): For every even integer \( p > 2 \) there exists a constant \( C_p \), such that

\[
\int_{-\infty}^{\infty} (y - M(x))^p \, dH(y \mid x) \leq C_p \quad \text{for all } x \in [A, B).
\]

Assumption (V): The function \( \sigma^2(x) = \int_{-\infty}^{\infty} (y - M(x))^2 \, dH(y \mid x) \)

is continuous and nonvanishing at \( x = \Theta \).

**Theorem 2.** Suppose that the assumptions (I), (II), (III), (IV) and (V) are satisfied, and that \( a > \frac{1}{2K_1} \). Then the random variable \( n^2 (x_n - \Theta) \) tends in distribution to the normal distribution with mean 0 and variance

\[
\frac{\sigma^2(\Theta) a^2}{M'(\Theta) a - 1}.
\]

Proof is very similar to that of an analogous theorem in [4], and will be only sketched here, with some differences pointed out.
1° Under the additional assumption (IV), the asymptotic order of the higher absolute moments $\beta_n^p = E[|x_n - \Theta|^p] = O(n^{-r})$ will be deduced by induction with respect to even $r$; (for odd $r$ it follows then by Lyapunov's inequality). As a consequence we get by Chebyshev's inequality,

$$\int_{|x_n - \Theta| > \delta} F_n(x) \, dP = O(n^{-q}) \quad (5)$$

for every Borel measurable function $F$ bounded in $\langle A, B \rangle$ and for every $\delta > 0$, $q > 0$ (i.e. for $\delta$ arbitrarily small and $q$ arbitrarily large).

2° We observe that

$$P\left(x_n + \frac{a}{n} (x - y_n) > B\right) \leq P\left(x_n - \Theta > \frac{B - \Theta}{2}\right) + P\left(\frac{a}{n} (x - y_n) > \frac{B - \Theta}{2}\right) \leq$$

$$\leq \frac{\beta_n^{2s}}{(B - \Theta)^{2s}} + \frac{E[|x - y_n|^q]}{(B - \Theta)^q} = O(n^{-q}),$$

and, similarly, $P\left(x_n + \frac{a}{n} (x - y_n) < A\right) = O(n^{-q})$, $q > 0$ arbitrary. Therefore

$$E[(x_{n+1} - \Theta)^r] = E\left[(x_{n+1} - \Theta)^r \mid x_n + \frac{a}{n} (x - y_n) \in \langle A, B \rangle\right] + O(n^{-q}) =$$

$$= E\left[(x_n - \Theta + \frac{a}{n} (x - y_n))^r \mid x_n + \frac{a}{n} (x - y_n) \in \langle A, B \rangle\right] + O(n^{-q}) =$$

$$= E\left[(x_n - \Theta + \frac{a}{n} (x - y_n))^r\right] + O(n^{-q}), \text{ for arbitrary } q > 0 .$$

Denoting $b_n^{(r)} = E[(x_n - \Theta)^r]$, we get

$$b_{n+1}^{(r)} = b_n^{(r)} + \sum_{t=1}^{r} (-1)^t \binom{r}{t} \frac{a^t}{n^t} E[(x_n - \Theta)^{r-t} (y_n - \Theta)^t] + O(n^{-q}) .$$

Evaluating expectations on the right side, we can by (5) reduce the integration to the interval $|x_n - \Theta| \leq \delta$, where by means of (III) and (V) more precise estimates are available; this enables us to prove (inductively) that

$$\lim_{n \to \infty} n^2 b_n^{(r)} = \begin{cases} 
0 & \text{for } r = 2s - 1 , \\
\left(\frac{\sigma^2(\Theta) a^2}{2M'(\Theta) a - 1}\right)^s (2s - 1)!! & \text{for } r = 2s ,
\end{cases}$$

which implies the statement of the theorem.

2. The Kiefer-Wolfowitz stochastic approximation method

Let again $\{H(y \mid x)\}$ be a family of distribution functions and $M(x) = = \int_{-\infty}^{\infty} y \, dH(y \mid x)$ the corresponding regression function. Suppose that $M(x)$ achieves its maximum for a value $x = \Theta$ from a (known) finite interval $\langle A, B \rangle$.
and that $M(x)$ is increasing or decreasing according to $x < \Theta$ or $x > \Theta$ in a larger interval $\langle A - c', B + c' \rangle$.

Let $a > 0$, $0 < c \leq c'$, $0 < \gamma < \frac{1}{2}$ be constants; denote $\frac{a}{n} = a_n$, $\frac{c}{n^\gamma} = c_n$.

Take $x_1 \in \langle A, B \rangle$ arbitrarily and for $n \geq 1$ set recursively

$$x_{n+1} = \left[ x_n + a_n \frac{y_{2n} - y_{2n-1}}{c_n} \right]_4,$$

where $y_{2n}, y_{2n-1}$ are random variables, which for given $x_1, \ldots, x_n, y_1, \ldots, y_{2n-2}$ have distribution functions $H(y \mid x_n + c_n), H(y \mid x_n - c_n)$ respectively, and are independent.

We shall still add the following assumptions.

**Assumption (I')**: There exists a constant $\sigma^2$, such that

$$\int_{-\infty}^\infty (y - M(x))^2 \, dH(y \mid x) \leq \sigma^2 \text{ for all } x \in \langle A - c', B + c' \rangle.$$

**Assumption (II')**: There exist $K_3 > 0$, $K_4 > 0$ such that

$$K_3 |x - \Theta| \leq |M'(x)| \leq K_4 |x - \Theta| \text{ in some neighbourhood of } \Theta.$$

**Assumption (III')**: There exists a $K_5 > 0$ and for every $\delta > 0$ a $K_6(\delta) > 0$, such that

$$|M'(x)| \leq K_5 \text{ for all } x \in \langle A - c', B + c' \rangle,$$

$$|M'(x)| \geq K_6(\delta) \text{ for all } |x - \Theta| > \delta, x \in \langle A - c', B + c' \rangle.$$

Remark. The assumption (II') is certainly satisfied, if $M''(\Theta) < 0$ exists. We deduce first some consequences of the assumptions. Denote $M_\varepsilon(x) = \frac{M(x + \varepsilon) - M(x - \varepsilon)}{2\varepsilon}$ for $x \in \langle A, B \rangle$, $0 < \varepsilon < c'$; we have

$$M_\varepsilon(x) = M'(x + \varepsilon) + M'(x - \varepsilon) \text{ with } 0 < \varepsilon < 1, \, i = 1, 2. \ (6)$$

Set $\kappa(x) = -\frac{M'(x)}{x - \Theta}$ for $x \neq \Theta$, $\kappa(\Theta) = K_3$; by (II') it holds $K_3 \leq \kappa(x) \leq K_4$ in some neighbourhood of $\Theta$, say for $|x - \Theta| \leq \delta$.

Suppose that $\varepsilon < \frac{1}{3} \delta$. We have

$$M_\varepsilon(x) = [\kappa(x + \varepsilon) + \kappa(x - \varepsilon)](x - \Theta) + \left[ \kappa(x + \varepsilon) - \kappa(x - \varepsilon) \right] \varepsilon,$$

hence

$$(x - \Theta) M_\varepsilon(x) \leq -2K_3(x - \Theta)\varepsilon + K_4\varepsilon |x - \Theta| \text{ for } |x - \Theta| \leq \delta - \varepsilon. \ (7)$$

On the other hand, by (III') and (6), we have

$$|M_\varepsilon(x)| \leq 2K_6 \left( \frac{\delta}{3} \right) \text{ for } |x - \Theta| > \delta - \varepsilon, \ (8)$$

$$M_\varepsilon^2(x) \leq 4K_5^2 \text{ for all } x \in \langle A, B \rangle. \ (9)$$
Returning to the approximation scheme, we see that $|x_n - \Theta| < B - A$ for all $n$, and $c_n < \delta$ for all $n > n_0(\delta)$; hence

$$|M_{c_n}(x_n)| > \frac{2K_6}{B - A} |x_n - \Theta| \text{ for } |x_n - \Theta| > \delta - c_n, \ n > n_0(\delta),$$
or, taking into account that $M(x)$ is increasing or decreasing as $x < \Theta$ or $x > \Theta$,

$$(x_n - \Theta) M_{c_n}(x_n) \leq -\frac{2K_6}{B - A} (x_n - \Theta)^2 \text{ for } |x_n - \Theta| > \delta - c_n, \ n > n_0(\delta).$$

Combining this with (7), we get

$$(x_n - \Theta) M_{c_n}(x_n) \leq -K_7 \cdot (x_n - \Theta)^2 + K_4 c_n |x_n - \Theta| \text{ for } n > n_0(\delta) \quad (10)$$
(without restriction on $x_n$).

**Theorem 3.** Suppose that the assumptions (I$_2$), (II$_2$) and (III$_2$) are satisfied, and that $a > \frac{1}{2K_7}$. Then

$$b_n = \begin{cases} O\left(\frac{1}{n^{1-\gamma}}\right) & \text{for } \gamma \geq \frac{1}{4}, \\ O\left(\frac{1}{n^{2\gamma}}\right) & \text{for } \gamma < \frac{1}{4}. \end{cases}$$

**Remark.** These upper bounds for $b_n$ cannot be lowered in general; therefore the choice $\gamma = \frac{1}{4}$, giving $b_n = O\left(\frac{1}{n^{1/4}}\right)$, is the optimal one (under the assumptions made above).

In order to prove the statement in the remark, we use a family $\{H(y \mid x)\}$ with $M(x) = \begin{cases} - (x - \Theta)^2 & \text{for } x \leq \Theta \\ - \frac{1}{2} (x - \Theta)^2 & \text{for } x > \Theta \end{cases}$ and with $\sigma^2(x) = \sigma^2 > 0$. This special case leads — for every choice of $\gamma$ — to $b_n$ of exactly that order which is given as upper bound in Theorem 3. (Cf. [6]!)

**Proof of Theor. 3.** As in Sect. 1, it is easily seen that

$$(x_{n+1} - \Theta)^2 \leq (x_n - \Theta)^2 + a_n^2 \frac{(y_{2n} - y_{2n-1})^2}{c_n^2} + 2a_n(x_n - \Theta) \frac{y_{2n} - y_{2n-1}}{c_n},$$
hence

$$E[(x_{n+1} - \Theta)^2 \mid x_n] \leq (x_n - \Theta)^2 + 2a_n^2 c_n^{-2} + a_n^2 M_{c_n}(x_n) + 2a_n(x_n - \Theta) M_{c_n}(x_n).$$

Taking once more expectations and using (9) and (10) we get

$$b_{n+1} \leq b_n + 2a_n^2 c_n^{-2} + \frac{4K_6 K_4}{n^{2-2\gamma}} b_n + 2K_4 a c \frac{1}{n^{1+\gamma}} E[|x_n - \Theta|].$$
By means of the inequality $E[|x_n - \Theta|] \leq \varepsilon_n + \frac{1}{\varepsilon_n} b_n$ we obtain

$$b_{n+1} \leq \left( 1 - \frac{(2 - \varepsilon) K_7 a}{n} \right) b_n + \frac{2 \sigma^2 a^2 e^{-2} + o(1)}{n^{2-2\gamma}} + \frac{4 K_7^2 K_7^2 e^{-1} a c^2}{n^{1+2\gamma}}.$$

The application of Chung’s lemma completes the proof.

The proofs of the following three theorems will be omitted; they are entirely analogous to the proofs of corresponding theorems in [6].

Assumption (IV$\_2$): The bounded third derivative $M''(x)$ exists in some neighbourhood of $\Theta$.

Theorem 4. Suppose that the assumptions (I$\_2$), (II$\_2$), (III$\_2$) and (IV$\_2$) are satisfied, and that $\alpha > \frac{1}{2 K_7}$. Then

$$b_n = \begin{cases} O\left(\frac{1}{n^{1-\gamma}}\right) & \text{for } \gamma \geq \frac{1}{6}, \\ O\left(\frac{1}{n^3}\right) & \text{for } \gamma < \frac{1}{6}. \end{cases}$$

Remark. These bounds for $b_n$ cannot be lowered without adding further restrictive assumptions; therefore the choice $\gamma = \frac{1}{6}$, giving $b_n = O\left(\frac{1}{n^3}\right)$, is the optimal one.

Assumption (V$\_2$): The function $M(x)$ is analytical and symmetrical about $\Theta$ in some neighbourhood of $\Theta$.

Theorem 5. Suppose that the assumptions (I$\_2$), (II$\_2$), (III$\_2$) and (V$\_2$) are satisfied, and that $\alpha > \frac{1}{2 K_7}$. Then

$$b_n = O\left(\frac{1}{n^{1-\gamma}}\right) \text{ for all } 0 < \gamma < \frac{1}{2}.$$

Assumption (VI$\_2$): For every even integer $p > 2$ there exists a constant $C_p$ such that

$$\int_{-\infty}^{\infty} (y - M(x))^p \, dH(y \mid x) \leq C_p \text{ for all } x \in (A, B).$$

Assumption (VII$\_2$): The function

$$\sigma^2(x) = \int_{-\infty}^{\infty} (y - M(x))^2 \, dH(y \mid x)$$

is continuous and nonvanishing at $x = \Theta$. 

146
Theorem 6. Suppose that the assumptions (I_2), (II_2), (III_2), (VI_2) and (VII_2) are satisfied, and that \( a > -\frac{1}{2K} \). Then in each of the following three cases

1° \( \gamma \geq \frac{1}{2} \) and the continuous \( M''(x) < 0 \) exists in some neighbourhood of \( \Theta \),

2° \( \gamma > \frac{1}{2} \) and the assumption (IV_2) is satisfied,

3° the assumption (V_2) is satisfied,

the random variable \( n^{\frac{1}{2}} - r(x_n - \Theta) \) tends in distribution to the normal distribution with mean 0 and variance \( \sigma^2(\Theta) a^2 \left( -2M''(\Theta) a - \frac{1}{2} + \gamma \right) e^2 \).

3. Solving systems of linear equations by a stochastic approximation method

In this section, the (column) vectors will be denoted by \( x, x_n, \ldots, \Theta \), and by \( \xi_i, \xi_{ni}, \ldots, \Theta_i \) (\( i = 1, \ldots, r \)) — their coordinates. The rows of a matrix \( M \), considered as row vectors, will be denoted by \( M_i \). The Euclidean norm \( \|x\| = (\sum_{i=1}^{r} \xi_i^2)^{1/2} \) and the corresponding norm of a matrix \( \|M\| = (\text{Max}_{i} \lambda_i(M'M))^{1/2} \) will be used.

Let to each \( x \in E_r \) correspond a distribution function of \( r \) variables \( H(y \mid x) \). Suppose that the regression of \( y \) on \( x \) (i.e. the vector function, whose \( i \)-th coordinate is given by the integral \( \int_{-\infty}^{\eta} dH_i(\eta \mid x) \), where \( H_i(\eta \mid x) = H(-\infty, \ldots, \eta, \ldots, +\infty, \eta \text{ on the } i \text{-th place}) \) is of the type \( Mx \), where \( M \) is a matrix with constant elements. Suppose that \( M \) is nonsingular, so that, for given \( \lambda \), the system of linear equations \( Mx = \lambda \) has a unique solution \( \Theta \).

Let \( a \) be a positive constant. Define the approximation procedure by taking \( x_1 \) arbitrarily, and for \( n \geq 1 \) setting recursively

\[
x_{n+1} = x_n - \frac{a}{n^2} y_n, \tag{11}
\]

where \( y_n \) is a random vector whose distribution function, for given \( x_1, \ldots, x_n, y_1^*, \ldots, y_n^*, y_1, \ldots, y_{n-1}, \) is \( H(y \mid y_n^* - \lambda) \), and \( y_n^* \) is a random vector whose distribution function, for given \( x_1, \ldots, x_n, y_1^*, \ldots, y_{n-1}^*, y_1, \ldots, y_{n-1} \), is \( H(y \mid x_n) \).

Remark. The realization of this approximation scheme is the following: Given \( x_n \), we get first \( y_n^* \) as a result of an observation on the level \( x_n \); then we get \( y_n \) as result of an observation on the level \( y_n^* - \lambda \), and construct \( x_{n+1} \) according to (11).

We make further following two assumptions:
Assumption (I3): There exists a constant $S^2$ such that
\[
\int_{E_r} \|y - Mx\|^2 \, dP_x \leq S^2 \quad \text{for all } x \in E_r,
\]
where $P_x$ denotes the probability measure in $E_r$ induced by $H(y \mid x)$.

Assumption (II3): The matrix $M$ is symmetrical.

Set $K_8 = \min \lambda_i^2$, $K_9 = 1 + \max \lambda_i^2 = 1 + \|M\|^2$, where $\lambda_i$ are the latent roots of the (symmetrical!) matrix $M$. Denote $b_n = E[\|x_n - \Theta\|^2]$.

Theorem 7. Suppose that the assumptions (I3) and (II3) are satisfied, and that $a > \frac{1}{2K_8}$. Then
\[
b_n = O\left(\frac{1}{n}\right), \
\text{or, more precisely, } b_n \leq \frac{K_8S^2a^2}{2K_8a - 1} \cdot \frac{1}{n} + O\left(\frac{1}{n^2} + \frac{1}{n^2K_8a}\right).
\]

Proof. First we observe that
\[
E[\eta ni \mid y_n^* \mid x_n] = M_i(y_n^* - \alpha),
\]
\[
E[\eta ni \mid x_n] = E[M_i(y_n^* - \alpha) \mid x_n] = M_i M(x_n - \Theta).
\]

Then we shall find an upper bound for $E[\|y_n\|^2 \mid x_n]$:
\[
\|y_n\|^2 = \sum_{i=1}^{r} \eta ni = \sum_{i=1}^{r} [M_i(y_n^* - \alpha) - M_i M(x_n - \Theta)] + [M_i M(x_n - \Theta)]^2,
\]
hence by (12)
\[
E[\|y_n\|^2 \mid y_n^*, x_n] = E[\|y_n - M(y_n^* - \alpha)\|^2 \mid y_n^*, x_n] +
+ \|M(y_n^* - Mx_n)\|^2 + \|M^2(x_n - \Theta)\|^2 + 2 \sum_{i=1}^{r} [M_i(y_n^* - Mx_n)] \cdot [M_i M(x_n - \Theta)];
\]
further, by the definition of $y_n^*$, by (I3) and by (13),
\[
E[\|y_n\|^2 \mid x_n] \leq S^2 + E[\|M(y_n^* - Mx_n)\|^2 \mid x_n] + \|M^2(x_n - \Theta)\|^2 \leq
\leq (1 + \|M\|^2) S^2 + \|M\|^4 \|x_n - \Theta\|^2.
\]

Now, by (11) and in the next row by (13),
\[
\|x_{n+1} - \Theta\|^2 = \|x_n - \Theta\|^2 - \frac{2\alpha}{n} (x_n - \Theta)' y_n + \frac{\alpha^2}{n^2} \|y_n\|^2,
\]
\[
E[\|x_{n+1} - \Theta\|^2 \mid x_n] = \|x_n - \Theta\|^2 - \frac{2\alpha}{n} (x_n - \Theta)' M^2(x_n - \Theta) + \frac{\alpha^2}{n^2} E[\|y_n\|^2 \mid x_n].
\]

Since $M$ is nonsingular and symmetrical, the matrix $M^2$ is positive definite, so that
\[
(x_n - \Theta)' M^2(x_n - \Theta) \geq K_8 \|x_n - \Theta\|^2
\]
with $K_8$ equal to the smallest latent root of $M^2$. 

148
Inserting (14) and (16) into (15), we obtain

\[ E[||x_{n+1} - \Theta||^2 | x_n] \leq ||x_n - \Theta||^2 - \frac{2K_n a}{n} ||x_n - \Theta||^2 + \frac{a^2}{n^2} [(1 + ||M||^2) S^2 + \\
+ ||M||^2 ||x_n - \Theta||^2] , \]

and finally,

\[ b_{n+1} \leq \left( 1 - \frac{2K_n a + o(1)}{n} \right) b_n + \frac{K_n S^2 a^2}{n^2} . \]

Applying Chung's lemma, we get the statement of the theorem.

REFERENCES

(AMS — Annals of Mathematical Statistics)


Резюме

ЗАМЕТКИ К СТОХАСТИЧЕСКИМ АППРОКСИМАЦИОННЫМ МЕТОДАМ

ВАЦЛАВ ДУПАЧ, (Václav Dupáč), Прага

(Поступило в редакцию 11/III, 1957 г.)


В §§ 1-ом и 2-ом настоящей статьи показано, что эти условия можно значительно ослабить, если предположить, что искомое решение лежит в некотором заранее известном конечном промежутке.

В § 3-ем исследуется стохастический аппроксимационный метод для решения систем линейных уравнений с симметрической матрицей коэффициентов.