Aplikace matematiky

Václav Dupač; Ulrich Herkenrath

On integer stochastic approximation

Aplikace matematiky, Vol. 29 (1984), No. 5, 372–383

Persistent URL: http://dml.cz/dmlcz/104107

ON INTEGER STOCHASTIC APPROXIMATION

VÁCLAV DUPAČ, ULRICH HERKENRATH

(Received October 12, 1983)

1. NOTATION AND OVERALL ASSUMPTIONS

M will denote a Borel measurable function $M\colon \mathbb{R}^p \to \mathbb{R}^p$ or a function $M\colon \mathbb{Z}^p \to \mathbb{R}^p$, where $\mathbb{Z}^p$ is the set of all integer points in $\mathbb{R}^p$. M is assumed to be unknown; it is, nonetheless, observable at integer points. More precisely, at each point $x \in \mathbb{Z}^p$ and at any time-instant $n \in \mathbb{N}$, an observation can be made yielding $M(x) + e_n(x)$, where $e_n = (e_n(x), x \in \mathbb{Z}^p)$ is a p-vector valued random function on $\mathbb{Z}^p$ such that e_n, $n \in \mathbb{N}$, are independent, and $\mathsf{E}\, e_n(x) = 0$, $x \in \mathbb{Z}^p$, $n \in \mathbb{N}$. If all the random functions e_n, $n \in \mathbb{N}$, have the same distribution, we shall consider them as copies of a random function $e = (e(x), x \in \mathbb{Z}^p)$.

We assume that the equation $M(x) = 0$ has a solution θ, if $\mathbb{R}^p$ is the domain of M; our aim is to approximate θ by integers, i.e. to find a cube with integer vertices containing θ. If, however, $\mathbb{Z}^p$ is the domain of M, then it would not be realistic to postulate the existence of a solution of $M(x) = 0$ in $\mathbb{Z}^p$ (though we do not exclude this possibility). A more natural formulation of our goal is to find the point $\theta^* \in \mathbb{Z}^p$, that realizes $\min \{|M(x)| : x \in \mathbb{Z}^p\}$.

Iterative procedures, which materialize the just formulated aims and are nonparametric both with respect to M and to the distribution of the e_n's, will be called procedures of integer stochastic approximation (including related procedures, as approximating the point of the maximum of a function $M : \mathbb{R}^p \to \mathbb{R}^1$ by integers).

2. ONE-DIMENSIONAL PROCEDURES OF INTEGER STOCHASTIC APPROXIMATION

In this section, we give a review of various approaches to integer stochastic approximation and a few comments on them.

Derman-type procedure. Assume M is defined on $\mathbb{R}^1$, θ is its unique zero point; let e_n, $n \in \mathbb{N}$, be all distributed as $e = (e(x), x \in \mathbb{Z}^1)$, let $\mathsf{P}(M(x) + e(x) = 0) = 0$,

$x \in \mathbb{Z}^1$. Denote $p_x = \mathsf{P}(M(x) + e(x) > 0)$, $x \in \mathbb{Z}^1$; assume that p_x is nondecreasing on $\mathbb{Z}^1$ and such that

$$p_{[\theta]-1} < p_{[\theta]} \leqq \tfrac{1}{2} \leqq p_{[\theta]+1} < p_{[\theta]+2} ,$$

where $[\theta]$ denotes the integer part of θ. Choose X_1 as an arbitrary integer; for $n \in \mathbb{N}$ put

$$X_{n+1} = X_n - \operatorname{sign}(M(X_n) + e_n(X_n))$$

and define θ_n as the most frequent value among $X_1, X_2, \ldots, X_n$, if this is uniquely determined, or as the average of such values, if not. Then we have

$$\mathsf{P}([\theta] \leqq \theta_n \leqq [\theta] + 1 \text{ eventually}) = 1 . \tag{1}$$

Remark. Assume M is nondecreasing everywhere and strictly increasing in $[\theta - 1, \theta + 1]$; let the random variables $e(x)$ be identically distributed for all $x \in \mathbb{Z}^1$, with a symmetric positive probability density function. Then the convergence assertion (1) holds true.

Alternatively, assume that Z_n, $n \in \mathbb{N}$, are independent identically distributed random variables with the distribution function F strictly increasing in $[\theta - 1, \theta + 1]$, θ being the unique median of F. Let $1_{[Z_n \leqq x]}$, $x \in \mathbb{Z}^1$, be the indicator of the event in the brackets. Choose $X_1 \in \mathbb{Z}^1$ arbitrarily and put

$$X_{n+1} = X_n - \operatorname{sign}(1_{[Z_n \leqq X_n]} - \tfrac{1}{2}) , \quad n \in \mathbb{N} .$$

Define θ_n as above. Then the assertion (1) holds true. (To see that, we have only to identify $M(x)$ with $F(x) - \frac{1}{2}$ and $e_n(x)$ with $1_{[Z_n \leqq x]} - F(x)$. This is actually the case considered by Derman (1957). The proof of our version of his result is, however, identical.

Mukerjee's procedure. Assume M is defined on $\mathbb{R}^1$, $\sup_{x < \theta - \varepsilon} M(x) < 0$, $\inf_{x > \theta + \varepsilon} M(x) > 0$, $\forall \varepsilon > 0$. Let e_n, $n \in \mathbb{N}$, be all distributed as $e = (e(x), x \in \mathbb{Z}^1)$; define $G(t) = \sup_x \mathsf{P}(|e(x)| \geqq t)$, assume $G(t) \to 0$ for $t \to \infty$, $\int_0^{+\infty} t |\mathrm{d}G(t)| < \infty$. In the first step, choose integers $X_1, X_2, \ldots, X_{k_1}$ arbitrarily; observe M at these points, denote the observations by $Y_1, Y_2, \ldots, Y_{k_1}$ (i.e., $Y_i = M(X_i) + e(X_i)$). After n steps, let $(X_1, Y_1), \ldots, (X_{k_n}, Y_{k_n})$ be all pairs of observational points and observations obtained up to time n. Define $M_n(x)$ for $x \in \{X_1, \ldots, X_{k_n}\}$ as the isotonic regression of Y on X, i.e., as the least squares fit of the observed values subject to the constraint that $M_n(x)$ is nondecreasing. Extend M_n to a continuous polygonal nondecreasing function on $\mathbb{R}^1$. Denote by $x_{\min}$ and $x_{\max}$ the minimum and maximum of $\{X_1, X_2, \ldots, X_{k_n}\}$. Put $\theta_n = x_{\min} - 1$ or $\theta_n = x_{\max} + 1$, if $M(x)$ is positive or negative everywhere; put $\theta_n = (a + b)/2$, if $M_n^{-1}(0) \cap [x_{\min}, x_{\max}] = [a, b]$, possibly with $a = b$. In the $(n + 1)$ st step, take an observation at θ_n, if $\theta_n \in \mathbb{Z}^1$, or two observations, at $[\theta_n]$ and $[\theta_n] + 1$, if $\theta_n \notin \mathbb{Z}^1$. For this procedure, the assertion (1) holds true again, as proved by Mukerjee (1981).

Robbins-Monro procedure applied to the interpolated function and rounded off. Assume M is defined on $\mathbb{Z}^1$ and satisfies either

$$\exists\theta\colon M(\theta) = 0\,, \quad M(x) < 0\ \forall x < \theta\,, \quad M(x) > 0\ \forall x > \theta$$

or

$$\exists\theta'\colon M(x) < M(\theta') < 0 < M(\theta' + 1) < M(x')\ \forall x < \theta'\,,\ x' > \theta' + 1\,,$$
$$|M(\theta')| + |M(\theta' + 1)|\,.$$

θ^*, the point of the minimum of $|M|$, equals θ in the former case and equals θ' or $\theta' + 1$ in the latter case, according to whether $|M(\theta')| < |M(\theta' + 1)|$ or vice versa.

Assume $M^2(x) + \mathsf{E}e_n^2(x) \leqq K(1 + x^2)$, $x \in \mathbb{Z}^1$, K a positive constant. Let U_n, $n \in \mathbb{N}$, be random variables, uniformly distributed on $[0, 1]$, all U_n, e_n, $n \in \mathbb{N}$, independent. Define $\overline{M}(x)$ on $\mathbb{R}^1$ as the linear interpolation of M, i.e.

$$\overline{M}(x) = (1 - x + [x])\, M([x]) + (x - [x])\, M([x] + 1)\,, \quad x \in \mathbb{R}^1\,.$$

For each $x \in \mathbb{R}^1$ and $n \in \mathbb{N}$, define an observation $\overline{M}(x) + \bar{e}_n(x)$ in either of the following two ways:

(i) take observations at points $[x]$ and $[x] + 1$ and put

$$\overline{M}(x) + \bar{e}_n(x) =$$
$$= (1 - x + [x])\,(M([x]) + e_n([x])) + (x - [x])\,(M([x] + 1) + e_n([x] + 1))\,;$$

(ii) take one observation at point $[x]$ or at point $[x] + 1$ according to whether $U_n \geqq x - [x]$ or $U_n < x - [x]$, and put

$$\overline{M}(x) + \bar{e}_n(x) = 1_{[U_n \geq x - [x]]}(M([x]) + e_n([x])) +$$
$$+ 1_{[U_n < x - [x]]}(M([x] + 1) + e_n([x] + 1))\,.$$

Choose an arbitrary integer as X_1. For $n \in \mathbb{N}$ define

$$X_{n+1} = X_n - \frac{a}{n}(\overline{M}(X_n) + \bar{e}_n(X_n))\,, \quad a > 0 \text{ constant}\,.$$

Finally, define θ_n as that of the two points $[X_n]$ and $[X_n] + 1$ which is nearer to X_n. Then we have

$$\mathsf{P}(\theta_n = \theta^* \text{ eventually}) = 1\,.$$

See Dupač, Herkenrath (1982) for the proof (in a little different set up).

Remarks. In fact, all three procedures have been studied for a more general lattice of points $\{a + hn,\ n \in \mathbb{N}\}$ for some $a \in \mathbb{R}^1$, $h > 0$; the third method also for

a non-equidistant lattice. Here, we have confined ourselves to $\mathbb{Z}^1$ for convenience. Moreover, Mukerjee has proved his result under the assumption that $M^{-1}(0)$ is a finite interval, not necessarily a single point. However, both the Derman-type procedure and the interpolated and rounded off Robbins-Monro procedure can be modified so as to cover this situation as well.

Mukerjee also pointed out that in the Derman-type procedure some fraction of observations is necessarily taken far away from θ as $n \to \infty$, owing to the fact that $(X_n, n \in \mathbb{N})$ is a Markov chain on $\mathbb{Z}^1$ with all states recurrent and non-null, whereas in his method this loss of efficiency is not incurred. In fact, he proved not only $\mathsf{P}([\theta] \leqq \theta_n \leqq [\theta] + 1 \text{ eventually}) = 1$ but also $\mathsf{P}([\theta] \leqq X_k \leqq [\theta] + 1 \text{ eventually}) = 1$. Let us point out that the third procedure possesses the latter feature as well, with $[\theta]$ and $[\theta] + 1$ replaced by the two successive integers, in which M changes its sign.

The error probabilities $\mathsf{P}(\theta_n \notin [[\theta], [\theta] + 1])$ or $\mathsf{P}(\theta_n \neq \theta^*)$ could provide a useful information about the performance of the listed procedures. However, the only result known to us is their exponential rate for the third procedure, i.e. $\mathsf{P}(\theta_n \neq \theta^*) \leqq \leqq \mathrm{e}^{-cn}$, $n \in \mathbb{N}$, under some additional assumptions; see [2]. From the algorithmic point of view, the third procedure is memoryless (in the one-dimensional case), while the first two procedures are not.

The idea of interpolating and rounding off has been applied also to the Kiefer-Wolfowitz procedure for approximating the point of the maximum of a function $\mathbb{R}^1 \to \mathbb{R}^1$, see Herkenrath (1983). A Derman-type procedure applied to the same problem has been investigated by Kirchen (1982).

3. A MULTIDIMENSIONAL PROCEDURE

In [2], Sect. 5, an attempt has been made to generalize the Robbins-Monro interpolated and rounded off procedure to the multidimensional case. An extension of this result will be given now, and at the same time, an error made at the cited place will be corrected. Some lemmas on minima of quadratic forms on integer points will be made use of; they are listed separately in Section 4.

We shall confine ourselves to the two-dimensional case for convenience; see also the remark at the end of this section.

We introduce the following notation: For $x \in \mathbb{R}^2$, $x = (x_1, x_2)^T$, the points $([x_1], [x_2])^T$, $([x_1], [x_2] + 1)^T$, $([x_1] + 1, [x_2])^T$, $([x_1] + 1, [x_2] + 1)^T$ will be denoted by x^1, x^2, x^3, x^4; the products $(1 - x_1 + [x_1])(1 - x_2 + [x_2])$, $(1 - x_1 + [x_1]) \cdot (x_2 - [x_2])$, $(x_1 - [x_1])(1 - x_2 + [x_2])$, $(x_1 - [x_1])(x_2 - [x_2])$ by $\varkappa_x^1$, $\varkappa_x^2$, $\varkappa_x^3$, $\varkappa_x^4$. The closed sphere with a center x and a radius r will be denoted by $S_r(x)$; the square with sides $[[x_1] - R, [x_1] + R + 1]$ and $[[x_2] - R, [x_2] + R + 1]$ by $Q_R(x)$. We shall drop the subscript if $R = 0$. Hence, $Q(x)$ is the unit square with vertices x^1, x^2, x^3, x^4.

Let M be either $M\colon \mathbb{R}^2 \to \mathbb{R}^2$ (observable, however, at the integer points only) or $M\colon \mathbb{Z}^2 \to \mathbb{R}^2$. Assume

$$|M(x)| \leqq K(1 + |x|), \quad \mathsf{E}|e_n(x)|^4 \leqq K_1(1 + |x|^4), \tag{1}$$

for all $x \in \mathbb{Z}^2$ and some $K > 0$, $K_1 > 0$. Define the interpolated function $\overline{M}$ by

$$\overline{M}(x) = \sum_{k=1}^{4} \varkappa_x^k M(x^k), \quad x \in \mathbb{R}^2 .$$

For each $x \in \mathbb{R}^2$ and $n \in \mathbb{N}$ define an observation $\overline{M}(x) + \bar{e}_n(x)$ by the following rule $\mathscr{R}$: Take observations at points x^1, x^2, x^3, x^4 and put

$$\overline{M}(x) + \bar{e}_n(x) = \sum_{k=1}^{4} \varkappa_x^k (M(x^k) + e_n(x^k)) .$$

Choose an arbitrary $X_1 \in \mathbb{Z}^2$ and put

$$X_{n+1} = X_n - \frac{a}{n} (\overline{M}(X_n) + \bar{e}_n(X_n)), \quad n \in \mathbb{N} ,$$

with $a > 0$ constant.

Assumption A(R). *There is $\bar{\theta} \in \mathbb{R}^2$ and an $r \geqq 0$ such that*

$$\inf \{\overline{M}(x)^T (x - \bar{\theta}) : x \notin S_{r+\varepsilon}(\bar{\theta})\} > 0, \quad \forall \varepsilon > 0 . \tag{2}$$

R denotes the smallest nonnegative integer such that $S_{r+\varepsilon}(\bar{\theta}) \subset Q_R(\bar{\theta})$ for some $\varepsilon > 0$.

Theorem 1. *Under* (1) *and* A(R), *we have*

$$\mathsf{P}(\bar{\theta} \in \operatorname{int} Q_R(X_n) \text{ eventually}) = 1 .$$

Define

$$\Theta_n = \arg \min_{1 \leqq k \leqq 4} \Big\{ \Big| \sum_{i=1}^{4n} Y_i 1_{[\xi_i = X_n + k]} \Big| \Big/ \sum_{i=1}^{4n} 1_{[\xi_i = X_n + k]} \Big\} , \tag{3}$$

where (ξ_i, Y_i), $1 \leqq i \leqq 4n$, are pairs of observational (integer) points and the corresponding observations (i.e., $Y_i = M(\xi_i) + e_j(\xi_i)$, $i = 4j - 3, \ldots, 4j$, $j = 1, \ldots, n$) made up to the time n. Any of its elements may be chosen for Θ_n, if the "arg" consists of more than one point; the minimized ratio is considered to be $+\infty$, if its denominator is 0.

Recall that $\theta^* = \arg \min_{x \in \mathbb{Z}^2} |M(x)|$; we will assume that θ^* is a single point.

Assumption B. $\theta^* \in \{\bar{\theta}^1, \bar{\theta}^2, \bar{\theta}^3, \bar{\theta}^4\}$.

Theorem 2. *Under* (1), A(0) *and* B, *we have*

$$\mathsf{P}(\Theta_n = \theta^* \text{ eventually}) = 1 .$$

First we give a lemma.

Lemma 1. *Let* $\tilde{M} : \mathbb{R}^p \to \mathbb{R}^p$ *be* $\mathfrak{B}^p$*-measurable, let* $\tilde{e}_n = (\tilde{e}_n(x), x \in \mathbb{R}^p)$, $n \in \mathbb{N}$, *be* $\mathfrak{B}^p \times \mathfrak{F}_n$*-measurable p-vector valued random functions, where* $\mathfrak{F}_n = \sigma\{\tilde{e}_k(x), x \in \mathbb{R}^p, 1 \leqq k \leqq n\}$, $\tilde{e}_n$ *independent of* $\mathfrak{F}_{n-1}$, $\mathsf{E}\tilde{e}_n(x) = 0$, $x \in \mathbb{R}^p$, $n \in \mathbb{N}$. *Assume there is a* $\tilde{\theta} \in \mathbb{R}^p$ *and an* $r \geqq 0$ *such that*

$$\inf\{\tilde{M}(x)^T(x - \tilde{\theta}) : x \notin S_{r+\varepsilon}(\tilde{\theta})\} > 0\,, \quad \forall \varepsilon < 0\,;$$

further assume

$$|\tilde{M}(x)| \leqq K(1 + |x|)\,, \quad \mathsf{E}|\tilde{e}_n(x)|^4 \leqq K_1(1 + |x|^4)\,, \quad x \in \mathbb{R}^p\,, \quad n \in \mathbb{N}\,.$$

Let X_1 *be arbitrary,* $X_{n+1} = X_n - a_n(\tilde{M}(X_n) + \tilde{e}_n(X_n))$, $n \in \mathbb{N}$, *with constants* $a_n > 0$, $\sum_{n=1}^{\infty} a_n = +\infty$, $\sum_{n=1}^{\infty} a_n^2 < +\infty$. *Then we have*

$$\mathsf{P}(X_n \in S_{r+\varepsilon}(\tilde{\theta}) \text{ eventually}) = 1\,, \quad \forall \varepsilon > 0\,.$$

Proof of Lemma 1. Assume $\tilde{\theta} = 0$ without loss of generality.
Put $V(x) = (|x|^2 - r^2)^2 1_{[|x|>r]}$. Find $L_n V(x) = \mathsf{E}(V(X_{n+1}) - V(X_n)\,|X_n = x)$, the generating operator of the Markov sequence $(X_n, n \in \mathbb{N})$. We easily get

$$\begin{aligned} L_n V(x) \leqq\ & a_n^2 K(1 + V(x)) - 4a_n(|x|^2 - r^2)\,\tilde{M}(x)^T x\, 1_{[|x|>r]} + \\ & + (|x|^2 - r^2)^2\,\mathsf{E}(|1_{[|x - a_n\tilde{M}(x) - a_n\tilde{e}_n(x)|>r]} - 1_{[|x|>r]}|) - \\ & - 4a_n(|x|^2 - r^2)\,\mathsf{E}(\tilde{e}_n(x)^T x\, 1_{[|x - a_n\tilde{M}(x) - a_n\tilde{e}_n(x)|>r]})\,, \end{aligned}$$

and, after some calculations, we verify that both the two last terms on the right are bounded by $a_n^2 K(1 + V(x))$. Then the assertion of the lemma immediately follows from Theorem 2.7.1 in Nevel'son, Has'minskij (1972).

Proof of Theorem 1. As follows from their definitions, $\bar{M}$, $\bar{\theta}$ and $\bar{e}$, $n \in \mathbb{N}$, satisfy all assumptions of Lemma 1 (playing the role of $\tilde{M}$, $\tilde{\theta}$ and $\tilde{e}_n$). Hence, $\mathsf{P}(X_n \in \in S_{r+\varepsilon}(\bar{\theta})$ eventually) $= 1$, $\forall_{\varepsilon>0}$, which entails $\mathsf{P}(X_n \in \operatorname{int} Q_R(\bar{\theta})$ eventually) $= 1$, and, in turn, $\mathsf{P}(\bar{\theta} \in \operatorname{int} Q_R(X_n)$ eventually) $= 1$.

Proof of Theorem 2. From Theorem 1, we have $\mathsf{P}(X_n \in \operatorname{int} Q(\bar{\theta})$ eventually) $= 1$. Thus, observations are taken only at points $\bar{\theta}^k$, $1 \leqq k \leqq 4$, eventually, the arithmetic averages of their outcomes tend to the corresponding values $M(\bar{\theta}^k)$, $1 \leqq k \leqq 4$, respectively, and this enables us to find θ^* in a finite (though random) number of steps.

Remark 1. Let $\bar{\theta}$ have the same meaning as in (2). Let there be a $\delta > 0$, known to us, such that for each $j = 1, 2$, the following implications hold: $\bar{\theta}_j - [\bar{\theta}_j] < \delta$ or $\bar{\theta}_j - [\bar{\theta}_j] > 1 - \delta$ implies $\theta_j^* = [\bar{\theta}_j]$ or $\theta_j^* = [\bar{\theta}_j] + 1$, respectively. Modify the definition (3) of Θ_n as follows: If X_n is nearer than δ to a side [to two sides] of the square $Q(X_n)$, then take the minimum in (3) only over the vertices of that side [the common

vertex of those two sides]; otherwise let the definition (3) unchanged. Then the assertion of Theorem 2 remains valid, this time under (1), $\mathrm{A}(\delta, -\delta)$ and B, where $\mathrm{A}(\delta, -\delta)$ stands for "either $S_{r+\varepsilon}(\bar{\theta}) \subset Q_{-\delta}(\bar{\theta})$ or $S_{r+\varepsilon}(\bar{\theta}) \subset Q_{\delta}(\bar{\theta}) - Q_{-\delta}(\bar{\theta})$".

Remark 2. Both Theorems 1 and 2 as well as Remark 1 remain valid, if the rule $\mathscr{R}$ is replaced by the following rule $\mathscr{R}_1$: Take one observation at only one of the points x^k, $1 \leqq k \leqq 4$, chosen with the respective probabilities $\varkappa_x^k$; i.e., define $\bar{M}(x) + \bar{e}_n(x)$ as $M(x^k) + e_n(x^k)$ with the probabilities $\varkappa_x^k$, $1 \leqq k \leqq 4$.

In what follows, F will always denote a symmetric positive matrix (not depending on x), $\lambda_{\min}$ and $\lambda_{\max}$ its eigenvalues, H a nonlinear mapping $\mathbb{R}^2 \to \mathbb{R}^2$, ϱ the Euclidean distance in $\mathbb{R}^2$, ∂Q the boundary of Q.

We will give sufficient conditions for $\mathrm{A}(\mathrm{R})$, $\mathrm{A}(0)$ and **B** to be fulfilled.

Lemma 2. *Let $M : \mathbb{R}^2 \to \mathbb{R}^2$ or $M : \mathbb{Z}^2 \to \mathbb{R}^2$, $\theta \in \mathbb{R}^2$. For an $r > 0$ and each $x \notin S_r(\theta)$ let be either*

$$M(x^k) = F(x^k - \theta) + H(x^k), \tag{4_r}$$

where

$$|H(x^k)| < \lambda_{\min}(\max\{r, \varrho(\theta, Q(x))\} - \varepsilon_1), \quad 1 \leqq k \leqq 4,$$

or

$$M(x^k)^T (x^k - \theta) > \varepsilon_1, \tag{5}$$

$$\max_{1 \leqq k' \leqq 4} |M(x^{k'}) - M(x^k)| \, |x^k - \theta| < M(x^k)^T (x^k - \theta) - \frac{\varepsilon_1}{2}, \quad 1 \leqq k \leqq 4,$$

for some $\varepsilon_1 < 0$. Then $\mathrm{A}(\mathrm{R})$ is fulfilled with $\bar{\theta} = \theta$.

Note that (4_r) may be valid for some points and (5) for the others. (It is, however, impossible to satisfy (5) for $x = \theta$.) The inequality in (4_r) says how small the nonlinear part of M should be; (5) requires the projection of $M(x^k)$ onto $x^k - \theta$ to be positive and the differences between $M(x^k)$ and the values at the other vertices of $Q(x)$ to be less than the length of that projection. Notice that, in general, the restrictiveness of both conditions (4_r) and (5) decreases for points more remote from θ.

Let us specialize for $\mathrm{A}(0)$.

Lemma 2'. *Let $M : \mathbb{R}^2 \to \mathbb{R}^2$ or $M : \mathbb{Z}^2 \to \mathbb{R}^2$, $\theta \in \mathbb{R}^2$. Let $M(\theta^k) = F(\theta^k - \theta) + H(\theta^k)$, $|H(\theta^k)| < \lambda_{\min} \varrho(\theta, \partial Q(\theta))$, $1 \leqq k \leqq 4$. Further, for each $x \in \mathbb{Z}^2$, $x \neq \theta^1$, let either (4_0) or (5) hold true. Then $\mathrm{A}(0)$ is fulfilled with $\bar{\theta} = \theta$.*

Remark 3. For all $x \in \mathbb{Z}^2 \cap Q_1(\theta)$ let $M(x) = F(x - \theta) + H(x)$, $|H(x)| < \lambda_{\min} |\delta - \varrho(\theta, \partial Q(\theta))|$ for some $\delta > 0$. Further, for all $x \in \mathbb{Z}^2$ except those which are left lower vertices of unit squares whose union is $Q_1(\theta)$, let either (4_0) or (5) hold true. Then $\mathrm{A}(\delta, -\delta)$ is fulfilled with $\bar{\theta} = \theta$.

Lemma 3. *Let* $M : \mathbb{R}^2 \to \mathbb{R}^2$ *or* $M : \mathbb{Z}^2 \to \mathbb{R}^2$, $\theta \in \mathbb{R}^2$. *Let* $M(x) = F(x - \theta) + H(x)$, $x \in \mathbb{Z}^2$; $\lambda_{\max}/\lambda_{\min} < 1 + \sqrt{2}$;

$$\arg \min_{x \in \mathbb{Z}^2} |F(x - \theta)| = \arg \min_{x \in \mathbb{Z}^2} |F(x - \theta) + H(x)| . \tag{6}$$

Then B *is fulfiled with* $\bar{\theta} = \theta$.

The condition (6) means that the presence or absence of the nonlinear term in M does not influence the location of the minimum of M over the integer points. Note that no other condition is imposed on H; thus, Lemma 3 can be combined in an obvious manner with Lemma 2′ (or with Remark 3).

Proof of Lemmas 2, 2′ and 3. Take an $x^* \notin S_r(\theta)$. Assume (4_r) holds true for this x^* and hence for all $x \in Q'(x^*)$, where the prime means "without the upper and right sides". The linear map $F(x - \theta)$ remains unaffected by linear interpolation; that is, $\overline{M}(x) = F(x - \theta) + \overline{H}(x)$, $\forall x \in Q'(x^*)$, where $\overline{H}(x) = \sum_{k=1}^{4} \varkappa_x^k H(x^k)$. Hence

$$\overline{M}(x)^T (x - \theta) \geqq (\lambda_{\min}|x - \theta| - \max_{1 \leqq k \leqq 4} |H(x^k)|)\, |x - \theta| , \quad \forall x \in Q'(x^*) ,$$

which entails

$$\overline{M}(x)^T (x - \theta) > \eta \quad \text{for all} \quad x \in Q'(x^*) - S_{r+\varepsilon}(\theta) , \quad \varepsilon > 0 ,$$

$$\eta = \eta(\varepsilon) > 0 .$$

Now, assume that (5) holds true for x^*. Then we have, for all $x \in Q'(x^*) - S_r(\theta)$,

$$\overline{M}(x)^T (x - \theta) = \sum_{k'=1}^{4} \varkappa_x^{k'} M(x^{k'})^T \sum_{k=1}^{4} \varkappa_x^k (x^k - \theta) \geqq \sum_{k=1}^{4} \varkappa_x^k \{M(x^k)^T (x^k - \theta) -$$

$$- \max_{1 \leqq k' \leqq 4} |M(x^{k'}) - M(x^k)|\, |x^k - \theta|\} > \frac{\varepsilon_1}{2} .$$

Lemma 2′ and Remark 3 follow similarly. Lemma 3 is a consequence of Lemma 2 of Section 4.

Remark 4. All the results of this section hold true for $p \geqq 2$ as well, with obvious notational changes only; Lemma 3, however, only in case that our Conjecture (of Section 4) is true.

4. MINIMA OF QUADRATIC FORMS ON INTEGER POINTS

Let $\mathfrak{G} \subset \mathbb{R}^{2\times 2}$ denote the class of all symmetric positive matrices; let $G = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$ be an element of $\mathfrak{G}$, let $\Lambda = \lambda_{\max}/\lambda_{\min}$ be the ratio of its eigenvalues. Let $f_{G,\theta}(x) = (x - \theta)^T G(x - \theta)$ be the corresponding quadratic form, centred at θ. Let $|x|_\infty$

denote the maximum-norm in $\mathbb{R}^2$, i.e., $|x|_\infty = \max\{|x_1|, |x_2|\}$. For an $f\colon \mathbb{R}^2 \to \mathbb{R}^1$ possessing a unique minimum (and possibly a nonunique one over $\mathbb{Z}^2$), consider the inequality

$$(1) \qquad \left|\arg\min_{x\in\mathbb{R}^2} f(x) - \arg\min_{x\in\mathbb{Z}^2} f(x)\right|_\infty < 1 .$$

We say that (1) holds true, if it holds true for each $x \in \arg\min_{x\in\mathbb{Z}^2} f(x)$.

Lemma 1. (1) *holds true for every* $f \in \{f_{G,\theta} : \theta \in \mathbb{R}^2\}$ *if and only if* $|c| < \min\{a, b\}$.

Lemma 2. (1) *holds true for every* $f \in \{f_{G,\theta} : G \in \mathfrak{G},\ \Lambda < \Lambda_0,\ \theta \in \mathbb{R}^2\}$ *if and only if* $\Lambda_0 \leqq (1 + \sqrt{2})^2$.

Proof of Lemma 1. Denote $I^2 = [0, 1] \times [0, 1]$, $\mathbb{I}^2 = \{(0, 0)^T, (0, 1)^T (1.0)^T, (1, 1)^T\}$. The property (1) is obviously invariant under the transformations $x - \theta \mapsto T(x - \theta)$, $T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, $T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, i.e. under replacing G by TGT, i.e. under replacing c by $-c$ and/or interchanging a and b. It is also invariant under a shift $x \mapsto x + \xi$ by an integer vector ξ. Hence, for the proof it is sufficient to consider $a \geqq b$, $c \geqq 0$, $\theta \in I^2$.

(i) Assume $(0 \leqq)\ c < b (\leqq a)$. Divide $\mathbb{Z}^2$ into four parts, $Z_1 = \{x: x_1 \geqq 1, x_2 \geqq 1\}$, $Z_2 = \{x: x_1 \geqq 1, x_2 \leqq 0\}$, $Z_3 = \{x: x_1 \leqq 0, x_2 \leqq 0\}$, $Z_4 = \{x: x_1 \leqq 0, x_2 \geqq 1\}$. Obviously, $f_{G,\theta}$ is increasing in both variables x_1, x_2 in Z_1 and decreasing in Z_3; hence, $\min_{x\in Z_1} f_{G,\theta}(x) = f_{G,\theta}(1, 1)$ and $\min_{x\in Z_3} f_{G,\theta}(x) = f_{G,\theta}(0, 0)$. In Z_2, the function $f_{G,\theta}$ is everywhere increasing in the direction of $(1, -1)^T$. In fact, $f_{G,\theta}(x_1 + 1, x_2 - 1) - f_{G,\theta}(x_1, x_2) = 2(a - c)(x_1 - \theta_1) - 2(b - c)(x_2 - \theta_2) + (a - c) + (b - c) > 0$, $\forall x \in Z_2$. Hence, it remains to investigate the points $x \in Z_2$ with $x_1 = 1$ or $x_2 = 0$. For $x_1 = 1$, $f_{G,\theta}$ is increasing in the direction $(0, -1)^T$ for all $x_2 \leqq -1$ and any θ, and also at the point $x_1 = 1$, $x_2 = 0$, if $\theta_1 \geqq 1/2$ or $\theta_2 \geqq 1/2$ (as is easily seen from $f_{G,\theta}(1, x_2 - 1) - f_{G,\theta}(1, x_2) = b - 2b(x_2 - \theta_2) - 2c(1 - \theta_1)$). In the remaining case ($\theta_i < \frac{1}{2}$ for both $i = 1, 2$), however, $f_{G,\theta}(1, -1) < f_{G,\theta}(0, 0)$. Similarly for $x_2 = 0$, $f_{G,\theta}$ is increasing in the direction $(1, 0)^T$ for all $x_1 \geqq 2$ and any θ, and also at the point $x_1 = 1$, $x_2 = 0$, if $\theta_1 \leqq \frac{1}{2}$ or $\theta_2 \leqq \frac{1}{2}$. In the remaining case ($\theta_i > \frac{1}{2}$ for both $i = 1, 2$), however, $f_{G,\theta}(2, 0) < f_{G,\theta}(1, 1)$. Hence, $\min_{x\in Z_2} f_{G,\theta}(x) \geqq \min_{x\in\mathbb{I}^2} f_{G,\theta}(x)$.

As the situation is quite analogous in Z_4, we finally get $\min_{x\in\mathbb{Z}^2} f_{G,\theta}(x) = \min_{x\in\mathbb{I}^2} f_{G,\theta}(x)$. At the same time, the strict monotonicity (fulfilled in all the above cases) implies $\min_{x\in\mathbb{Z}^2-\mathbb{I}^2} f_{G,\theta}(x) > \min_{x\in\mathbb{I}^2} f_{G,\theta}(x)$. Hence, (1) is proved.

(ii) Assume $(0 <)\ b \leqq c (< a)$. To find a θ for which (1) fails to hold true, it suffices to put $\theta = (\frac{1}{2}, 0)^T$. The values of $f_{G,\theta}$ at $(0, 0)^T$ and $(1, 0)^T$ are then equal to $\frac{1}{4}a$, at

$(0, 1)^T$ and $(1, -1)^T$ equal to $\frac{1}{4}a + b - c$ and at $(1, 1)^T$ equal to $\frac{1}{2}a + b + c$. Hence, $\min_{x \in \mathbb{I}^2} f_{G,\theta}(x) = f_{G,\theta}(0, 1) = f_{G,\theta}(1, -1) \geqq \min_{x \in \mathbb{Z}^2 - \mathbb{I}^2} f_{G,\theta}(x)$, i.e., (1) does not hold true.

Remark. The strict inequality $\min_{x \in \mathbb{I}^2} f_{G,\theta}(x) > \min_{x \in \mathbb{Z}^2 - \mathbb{I}^2} f_{G,\theta}(x)$ holds true, if $\max\{a, b\} > |c| > \min\{a, b\}$ and if, at the same time, $\max\{a, b\} \neq 2|c|$.

To verify that, assume again $a \geqq b$, $c \geqq 0$ without loss of generality and put $\theta = (\frac{1}{2} + \varepsilon, 0)^T$ if $a > 2c$ or $\theta = (\frac{1}{2} - \varepsilon, 0)^T$ if $a < 2c$, with $\varepsilon > 0$ small enough to avoid a minimum at $(1, 0)^T$ or at $(0, 0)^T$, respectively. Then, in either case, we have $\min_{x \in \mathbb{I}^2} f_{G,\theta}(x) = f_{G,\theta}(0, 1) < f_{G,\theta}(1, -1) \geqq \min_{x \in \mathbb{Z}^2 - \mathbb{I}^2} f_{G,\theta}(x)$.

Proof of Lemma 2. The property (1) is invariant also under the mapping $x - \theta \mapsto k(x - \theta)$, $k > 0$, i.e., under replacing G by $k^2 G$; this does not change the ratio Λ. None of the maps listed at the beginning of the proof of Lemma 1 changes it, either. Hence, we can confine ourselves to $a \geqq b$, $c = 1$, $\theta \in I^2$, without loss of generality. We have to find the maximal Λ_0 such that $\Lambda < \Lambda_0 \Rightarrow 1 < b$; that is, we have to find $\min\{\Lambda: G \in \mathfrak{G}, c = 1, a > 1 \geqq b\}$, or explicitly

$$\min\left\{\frac{a + b + \sqrt{[(a - b)^2 + 4]}}{a + b - \sqrt{[(a - b)^2 + 4]}} : b \leqq 1 < ab\right\}.$$

An easy calculation shows that this is realized by $a = 3$, $b = 1$, and is equal to $(1 + \sqrt{2})^2$. (The number 6 has been erroneously given as this minimum in [2].) For the "only if" part of the assertion, the same example can be made use of as in the proof of Lemma 1.

Lemma 3. *Consider the class* $\{f_{G,\theta}: G \in \mathfrak{G}, \Lambda < (1 + \sqrt{2})^2, \theta \in \mathbb{R}^2\}$. *For every* $\delta > 0$, *there exists a maximal number* Λ_δ *with the following property: If* $\Lambda < \Lambda_\delta$ *and the distance of* θ *from a side* [*two sides*] *of the square* $Q(\theta)$ *is less than* δ, *then* $\arg\min_{x \in \mathbb{Z}^2} f_{G,\theta}(x)$ *is equal to one or both endpoints of that side* [*to the common endpoint of those two sides*].

Proof. Again, assume $a \geqq b > c = 1$, $\theta \in I^2$, without loss of generality. Introduce $D = \frac{1}{2}(a - b)$, $E = \frac{1}{2}(b - 1)$. Consider the set $\{\theta: f_{G,\theta}(0, 0) = f_{G,\theta}(0, 1)\}$; it turns out to be the line $\theta_2 = \frac{1}{2} - \theta_1/(1 + 2E)$. Similarly, $\{\theta: f_{G,\theta}(0, 0) = f_{G,\theta}(1, 0)\}$ is the line $\theta_2 = \frac{1}{2} + D + E - (1 + 2D + 2E)\theta_1$, etc. Denote the intersection of both the lines mentioned as $(\xi, \eta)^T$; our assumptions imply that $0 < \eta < \xi$. As the situation is quite symmetrical with respect to $(0, 0)^T$ and $(1, 1)^T$, the following holds true: If the center θ of the quadratic form $f_{G,\theta}$ lies at a distance smaller than η from a side of the square I^2, then $\arg\min_{x \in \mathbb{Z}^2} f_{G,\theta}(x)$ equals one or both of the endpoints of that side. Now, it remains to find for $0 < \delta < \frac{1}{2}$ a Λ_δ such that $\Lambda < \Lambda_\delta \Rightarrow$

$\Rightarrow \eta > \delta$, i.e., $\eta \leqq \delta \Rightarrow \Lambda \geqq \Lambda_\delta$. In other words, we have to find

$$\Lambda_\delta = \min \{\Lambda: D \geqq 0, E \geqq 0, \eta \leqq \delta\}, \tag{2}$$

where – as immediately follows from their definitions –

$$\Lambda = \Lambda(D, E) = 1 + \frac{2\sqrt{(1 + D^2)}}{1 + D + 2E - \sqrt{(1 + D^2)}},$$

$$\eta = \eta(D, E) = \frac{E(1 + 2D + 2E)}{2(D + 2E + 2DE + 2E^2)}.$$

The set defined by the constraints in (2) can be rewritten as

$$\left\{D \geqq 0,\ E \geqq 0;\ 0 \leqq E \leqq \frac{1}{2}\left[-\left(\frac{1 - 4\delta}{2 - 4\delta} + D\right) + J_\delta(D)\right]\right\}, \tag{3}$$

where

$$J_\delta(D) = \sqrt{\left[\left(\frac{1 - 4\delta}{2 - 4\delta} + D\right)^2 + \frac{4\delta}{1 - 2\delta} D\right]}.$$

As $(\partial/\partial E)\, \Lambda(D, E) < 0, \forall_{D,E}$, the minimum (2) must lie on the boundary of the set (3), that is, $\Lambda_\delta = \min_{D \geqq 0} L_\delta(D)$, where

$$L_\delta(D) = \Lambda\left(D, \frac{1}{2}\left[-\left(\frac{1 - 4\delta}{2 - 4\delta} + D\right) + J_\delta(D)\right]\right).$$

The equation

$$\frac{\mathrm{d}}{\mathrm{d}D} L_\delta(D) = 0 \quad \text{reads} \quad DJ_\delta = 1 + \frac{3 - 8\delta}{2 - 4\delta} D - D^2;$$

it has a unique solution D_δ on the set $\{D \geqq 0\}$, which can be found as the unique real positive root of the cubic equation

$$4(1 - 2\delta) D^3 + 2\delta D^2 - (3 - 8\delta) D - (1 - 2\delta) = 0.$$

Then $\Lambda_\delta = L_\delta(D_\delta)$.

For a few values of δ, Λ_δ are tabulated in Table 1, together with their square roots (as the matrix F of Section 3 plays the role of $G^{1/2}$ of Section 4). With a little bit of inconsistency in notation, we have used the symbol Λ_0 for $(1 + \sqrt{2})^2$ here.

Table 1

δ	Λ_δ	$\Lambda_\delta^{1/2}$
0	5·8284	2·4142
0·01	5·7185	2·3913
0·05	5·2794	2·2977
0·1	4·7318	2·1753
0·2	3·6435	1·9088

The following assertion seems plausible to us, although we do not have a formal proof.

Conjecture. *Lemma 2 remains valid for any integer* $p \geqq 2$ *instead of* $p = 2$ *(*Λ *denoting again the ratio of the maximal and minimal eigenvalues of* G*).*

Acknowledgement. The authors acknowledge the support of their research by the Deutsche Forschungsgemeinschaft, SFB 72. They also want to express their thanks to the referee, Dr. T. Fiala, for his helpful comments.

References

[1] *C. Derman:* Non-parametric up-and-down experimentation. Ann. Math. Statist. 28 (1957), 795—797.

[2] *V. Dupač, U. Herkenrath:* Stochastic approximation on a discrete set and the multiarmed bandit problem. Comm. Statist.-Sequential Analysis 1 (1982), 1—26.

[3] *U. Herkenrath:* The N-armed bandit with unimodal structure. Metrika 30 (1983), 195—210.

[4] *A. Kirchen:* Überlegungen zur eindimensionalen stochastischen Approximation. Diploma work. University of Bonn, 1982.

[5] *H. G. Mukerjee:* A stochastic approximation by observations on a discrete lattice using isotonic regression. Ann. Statist. 9 (1981), 1020—1025.

[6] *M. B. Nevel'son, R. Z. Has'minskij:* Stochastic Approximation and Recursive Estimation. Translation of Mathem. Monographs, vol. 47, Amer. Mathem. Soc., Providence, 1976. (Russian original, Nauka, Moskva 1982.)

Souhrn

O CELOČÍSELNÉ STOCHASTICKÉ APROXIMACI

VÁCLAV DUPAČ, ULRICH HERKENRATH

Funkce $M : \mathbb{R} \to \mathbb{R}$ nechť je pozorovatelná, s experimentální chybou, pouze v celočíselných bodech; jinak není známa. Iterační neparametrické metody pro hledání nulového bodu funkce M se nazývají metodami celočíselné stochastické aproximace. Jsou popsány a vzájemně porovnány tři takové metody: Dermanova, Mukerjeeho a autorů článku. Je navržena a vyšetřována dvojrozměrná analogie třetího z těchto přístupů; je vyslovena domněnka o jeho vícerozměrném zobecnění.

Authors' addresses: Dr. *Václav Dupač*, Matematicko-fyzikální fakulta Univerzity Karlovy, Sokolovská 83, 186 00 Praha 8; Dr. *Ulrich Herkenrath*, Institut für Angewandte Mathematik der Universität Bonn, Wegelerstrasse 6, D-5300 Bonn 1, BRD.