$$A=\begin{vmatrix} 10 & 6 & 8 \\ 8 & 30 & 13 \\ 9 & 30 & 50 \end{vmatrix}, \quad P_1=14, \quad P_2=25, \quad P_3=39, \quad T_1=9, \quad T_2=36, \quad T_3=17.$$

This matrix does not satisfy Hadamard's test of non-degeneracy, nor does it satisfy Brauer's test, yet it satisfies the conditions of Theorem 17.

Note. A matrix that satisfies Hadamard's non-degeneracy test will obviously satisfy the conditions of Theorem 17.

Theorem 17 gives an additive analogue of Brauer's non-degeneracy test. It is easy to construct examples that satisfy Brauer's test yet do not satisfy Theorem 17, and vice versa.

REFERENCES

1. PARODI M., Localization of eigenvalues of matrices and applications /Russian translation/, IIL, Moscow, 1960.
2. FAM VAN AT, Isolation of Gershgorin discs, Zh. vych. Mat. i mat. Fiz., 17, No.3, 756-759, 1977.
3. PORCHING T.A., Diagonal similarity transformations which isolate Gershgorin disks, Numer. Math., 8, 437-443, 1966.
4. LANCASTER P., Theory of matrices /Russian translation/, Nauka, Moscow, 1978.

Translated by D.E.B.

U.S.S.R. Comput.Maths.Math.Phys.,Vol.23,No.6,pp.20-27,1983 0041-5553/83 $10.00+0.00
Printed in Great Britain © 1985 Pergamon Press Ltd.

ADAPTIVE STEP ADJUSTMENT FOR A STOCHASTIC OPTIMIZATION ALGORITHM*

F. MIRZOAKHMEDOV and S.P. URYAS'EV

The quasigradient algorithm of stochastic optimization is considered. The conditions to be imposed on the step multiplier, for Cesaro convergence of the algorithms with probability 1, are studied. Adaptive step adjustment is proposed, and the convergence of the corresponding algorithm is proved. A numerical algorithm containing heuristic elements is described. The results of numerical experiments are quoted.

Our paper studies the gradient method of stochastic optimization of unsmooth functionals. We give a number of sufficient conditions for Cesaro convergence of the algorithm with probability 1, on the same lines as in /1/. When proving the convergence, we impose less rigid than usual /2/ conditions on the step multiplier of the algorithm. We propose an adaptive step adjustment for the method, such as was earlier used for methods of stochastic optimization with normalized quasigradient /3/.

The algorithm is considerably improved compared with /3/, and allowance is made for projection onto a convex compactum when solving constrained problems. The results of a numerical check of the algorithm are quoted, and recommendations are made for choosing the parameters.

An important merit of the algorithm is that the computer only needs a small working memory, so that it can be used to solve problems of high dimensionality. The text of a program in FORTRAN, realizing our approach, may be found in /4/.

1. Informal justification of adaptive step adjustment. We shall give a few loose arguments to explain the idea underlying our iterative scheme.

We wish to solve the problem

$$f(x)\to \min.$$

For simplicity, we shall assume here that $f(x)$ is smooth and convex in R^n. We shall use the scheme of successive approximations /2/

$$x^{s+1}=x^s-\rho_s\xi^s, \qquad \rho_s>0, \quad s=0,1,\ldots, \tag{1.1}$$

where ρ_s is the step multiplier, ξ^s is the stochastic quasigradient, i.e.,

$$M(\xi^s|x^0,\ldots,x^s)=M(\xi^s|B^s)=\nabla f(x^s), \qquad s=0,1,\ldots$$

(the σ-algebra B^s is induced by the random quantities $(x^0,\ldots,x^s)$). It is natural to choose the step ρ_s from the condition for minimum of the function $\varphi_s(\rho)$ with respect to ρ, where

$$\varphi_s(\rho)=M(f(x^s-\rho\xi^s)\,|B^s), \qquad s=0,1,\ldots.$$

It is easily seen that

**Zh.vȳchisl.Mat.mat.Fiz.*,23,6,1314-1325, 1983

$$\frac{\partial}{\partial\rho}\varphi_s(\rho)\Big|_{\rho_s}=\frac{\partial}{\partial\rho}M(f(x^s-\rho\xi^s)|B^s)|_{\rho_s}=M\left(\frac{\partial}{\partial\rho}f(x^s-\rho\xi^s)|_{\rho_s}B^s\right)=$$
$$-M((\nabla f(x^s-\rho_s\xi^s),\xi^s)|B^s)=-M((\nabla f(x^{s+1}),\xi^s)|B^s)=$$
$$-M((\xi^{s+1},\xi^s)|B^s),\quad s=0,1,\ldots.$$

At the $s+1$ -th iteration, therefore, we have the stochastic quasigradient $-(\xi^{s+1},\xi^s)$ of the function $\varphi_s(\rho)$ at the point ρ_s. To modify the step ρ_s we use the gradient procedure

$$\rho_{s+1}=\rho_s+\lambda_s(\xi^{s+1},\xi^s),\quad \lambda_s>0,\quad s=0,1,\ldots. \tag{1.2}$$

The quantity (ξ^{s+1},ξ^s) gives some information as to whether or not the minimum of the function $\varphi_s(\rho)$ with respect to ρ has been passed at the s-th iteration. If $(\xi^{s+1},\xi^s)>0$, then it is very probable that the minimum has not been passed at the s-th iteration. This information is used when varying the step ρ_s.

To facilitate the study of the convergence of scheme (1.1), it is convenient to rewrite (1.2) as

$$\rho_{s+1}=\rho_s\exp\{a_s(\xi^{s+1},\xi^s)\},\quad a_s>0,\quad s=0,1,\ldots. \tag{1.3}$$

To guarantee the satisfaction of one of the traditional conditions for stochastic programming algorithms to converge /2/

$$\sum_{s=0}^{\infty}\rho_s=\infty,\quad \rho_s>0,\quad s=0,1,\ldots,$$

we choose the quantity a_s as

$$a_s=a\rho_s,\quad a>0,\quad s=0,1,\ldots. \tag{1.4}$$

In the determinate case, if $f(x)$ is a strongly convex function, scheme (1.1), (1.3), (1.4) can be shown to converge to the minimum of $f(x)$.

In the stochastic case, we modify relation (1.3) as follows:

$$\rho_{s+1}=\rho_s\exp\{a[\rho_s(\xi^{s+1},\xi^s)-\delta\rho_s]\}=\rho_s\exp\{a[(\xi^{s+1},x^s-x^{s+1})-\delta\rho_s]\},\quad \delta>0,\quad s=0,1,\ldots.$$

This change guarantees satisfaction of the following convergence condition for algorithm (1.1):

$$\rho_s\to 0 \text{ a.s.},\quad s\to\infty.$$

(Here and henceforth, a.s. means convergence almost surely.)

We give below a general theorem on the Cesaro convergence of algorithm (1.1) with probability 1. The algorithm allows for projection onto non-stochastic constraints.

2. Sufficient conditions for Cesaro convergence. Consider the problem of minimizing the convex function $f(x)$ in the convex closed bounded set $X\subset R^n$:

$$f(x)\to\min_{x\in X}.$$

Notice that $f(x)$ may be an unsmooth function.

Let ω be an elementary event of probability space (Ω, F, P). The sequence of random points $\{x^s(\omega)\}$ is specified by the recurrence relation

$$x^{s+1}=\Pi_X(x^s-\rho_s\xi^s),\quad s=0,1,\ldots, \tag{2.1}$$

where $\Pi_X(y)$ is the operator projecting point y onto X, and $\xi^s(\omega)$ is a random vector for which

$$M(\xi^s|x^0,\ldots,x^s)=\hat f_x(x^s)+b^s,\quad s=0,1,\ldots.$$

Here, $\hat f_x(x^s)$ is the generalized gradient of the function $f(x)$ at the point x^s, the vector b^s depends on $(x^0,\ldots,x^s)$ (measurable with respect to the σ-algebra B^s induced by the random quantities $(x^0,\ldots,x^s)$), and step ρ_s is a positive random quantity which is non-zero.

Theorem 1. Let the function $f(x)$ be convex and continuous in the convex compactum $X\subset R^n$, let

$$\max_{x,y\in X}\|x-y\|=c_1, \tag{2.2}$$

$$\|b^s\|+\|\hat f_x(x^s)\|+\|\xi^s\|\le c_2 \text{ a.s.}, \tag{2.3}$$

$$\limsup_{s\to\infty}\|b^s\|\le\bar b \text{ a.s.}, \tag{2.4}$$

$$\rho_s\to 0 \text{ a.s.},\quad s\to\infty, \tag{2.5}$$

$$\sum_{s=0}^{\infty}\rho_s=\infty \text{ a.s.} \tag{2.6}$$

and let one of the following two relations be satisfied:

1) ρ_s depends only on $(x^0,\ldots,x^s)$ (is measurable with respect to the σ-algebra B^s induced by $(x^0,\ldots,x^s)$);

2) $\rho_s\rho_{s-1}^{-1}\xrightarrow[s\to\infty]{}1$ a.s., ρ_s depends only on $(x^0,\ldots,x^s,\xi^0,\ldots,\xi^s)$ (is measurable with respect to the σ-algebra induced by $(x^0,\ldots,x^s,\xi^0,\ldots,\xi^s)$).

Then,

$$\limsup_{s\to\infty} f(\bar{x}^s)-f(x^*)\leqslant \bar{b}c_1 \text{ a.s.}, \tag{2.7}$$

where

$$\bar{x}^s=\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=0}^{s}\rho_l x^l, \quad s=0,1,\dots, \qquad x^*\in X^*=\{x^*: f(x^*)=\min_{x\in X} f(x)\}.$$

Proof. We need a preliminary lemma on convergence of a series.

Lemma 1. For any ρ_s, satisfying $0<\rho_s\leqslant\bar{\rho}$, we have

$$\sum_{s=0}^{\infty}\Big[\rho_s^2\Big(\sum_{l=0}^{s}\rho_l\Big)^{-2}\Big]<\infty.$$

The proof is simple and may be omitted.

Using (2.1) and the properties of the projection operation, we estimate $\|x^{s+1}-x^*\|^2$:

$$\begin{aligned}\|x^{s+1}-x^*\|^2&=\|\Pi_X(x^s-\rho_s\xi^s)-x^*\|^2\leqslant\|x^s-x^*-\rho_s\xi^s\|^2=\\ &\|x^s-x^*\|^2-2\rho_s(\xi^s,x^s-x^*)+\rho_s^2\|\xi^s\|^2\leqslant\|x^0-x^*\|^2-\\ &2\sum_{l=0}^{s}\rho_l(\xi^l,x^l-x^*)+\sum_{l=0}^{s}\rho_l^2\|\xi^l\|^2,\quad s=0,1,\dots.\end{aligned} \tag{2.8}$$

We introduce the notation

$$d^l=(\xi^l-\hat{f}_x(x^l)-b^l,\ x^l-x^*),\qquad l=0,1,\dots.$$

Recalling (2.2)–(2.4), we obtain from (2.8):

$$\begin{aligned}2\sum_{l=0}^{s}\rho_l(\hat{f}_x(x^l),x^l-x^*)&\leqslant\|x^0-x^*\|^2-2\sum_{l=0}^{s}\rho_l d_l-2\sum_{l=0}^{s}\rho_l(b^l,x^l-x^*)+\\ &\sum_{l=0}^{s}\rho_l^2\|\xi^l\|^2\leqslant c_1^2-2\sum_{l=0}^{s}\rho_l d_l+2c_1\sum_{l=0}^{s}\rho_l\|b^l\|+c_2^2\sum_{l=0}^{s}\rho_l^2.\end{aligned} \tag{2.9}$$

Since the function $f(x)$ is convex, we have

$$\sum_{l=0}^{s}\rho_l(\hat{f}_x(x^l),x^l-x^*)\geqslant\sum_{l=0}^{s}\rho_l[f(x^l)-f(x^*)]\geqslant[f(\bar{x}^s)-f(x^*)]\sum_{l=0}^{s}\rho_l. \tag{2.10}$$

From (2.10) and (2.9) we obtain the estimate

$$0\leqslant f(\bar{x}^s)-f(x^*)\leqslant\Big(2\sum_{l=0}^{s}\rho_l\Big)^{-1}\Big(c_1^2+c_2\sum_{l=0}^{s}\rho_l^2\Big)-\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=0}^{s}\rho_l d_l+\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}c_1\sum_{l=0}^{s}\rho_l\|b^l\|. \tag{2.11}$$

Since properties (2.5), (2.6) hold, we have

$$\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\Big(c_1^2+c_2^2\sum_{l=0}^{s}\rho_l^2\Big)\to0 \text{ a.s.},\quad s\to\infty. \tag{2.12}$$

From conditions (2.4) and (2.6) we have

$$\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\lim_{s\to\infty}c_1\sum_{l=0}^{s}\rho_l\|b^l\|\leqslant c_1\bar{b} \text{ a.s.} \tag{2.13}$$

To prove the theorem, it is now sufficient, in the light of (2.11)–(2.13), to show that

$$\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=0}^{s}\rho_l d_l\to0 \text{ a.s.},\quad s\to\infty. \tag{2.14}$$

We first take the case when condition 1) of Theorem 1 holds. In accordance with Theorem 4 of /5, p.507/, since $(\Sigma_{l=0}^{s}\rho_l d_l, B^{s+1})$ is a square summable martingale, and $(\Sigma_{l=0}^{s}\rho_l, B^s)$ is a predictable increasing sequence, then (2.14) follows from the inequality

$$\sum_{s=0}^{\infty}\mathrm{M}((\rho_s d_s)^2|B^s)\Big(\sum_{l=0}^{s}\rho_l\Big)^{-2}<\infty \text{ a.s.} \tag{2.15}$$

Using (2.2), (2.3), (2.5) and Lemma 1, it is easily seen that

$$\sum_{s=0}^{\infty}\mathrm{M}((\rho_s d_s)^2|B^s)\Big(\sum_{l=0}^{s}\rho_l\Big)^{-2}\leqslant c_2^2c_1^2\sum_{s=0}^{\infty}\rho_s^2\Big(\sum_{l=0}^{s}\rho_l\Big)^{-2}<\infty \text{ a.s.},$$

whence (2.15) follows.

Now consider the case when condition 2) is satisfied. We shall show that, here again, (2.14) holds. We perform the equivalent transformation

$$\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=1}^{s}\rho_l d_l=\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=1}^{s}(\rho_l-\rho_{l-1})d_l+\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=1}^{s}\rho_{l-1}d_l. \tag{2.16}$$

Since $\|d_l\|\leqslant c_2c_1$ and $\rho_l\rho_{l-1}^{-1}\to 1$ a.s. as $l\to\infty$, then, in accordance with (2.6),

$$\Big(\sum_{l=0}^{s}\rho_l\Big)^{-1}\sum_{l=1}^{s}(\rho_l-\rho_{l-1})d_l\to 0 \text{ a.s.},\quad s\to\infty. \tag{2.17}$$

Denote by D^s the σ-algebra induced by the random quantities $(x^0,\ldots,x^s,\xi^1,\ldots,\xi^{s-1})$.

Obviously, $(\Sigma_{l=1}^{s}\rho_{l-1}d_l,D^{s+1})$ is a square summable martingale and $(\Sigma_{l=0}^{s-1}\rho_l,D^s)$ is a predictable increasing sequence, and hence, in the same way as when condition 1) holds, using Theorem 4 of /5, p.507/, we can show that

$$\Big(\sum_{l=0}^{s-1}\rho_l\Big)^{-1}\sum_{l=1}^{s}\rho_{l-1}d_l\to 0 \text{ a.s.},\qquad s\to\infty,$$

which, in view of (2.16), (2.17), implies (2.14). This proves the theorem.

Corollary. If the function $f(x)$ has a unique extremum x^* and $\bar{b}=0$, then $\bar{x}^s\to x^*$, $s\to\infty$, with probability 1.

3. Convergence of the algorithm with adaptive step adjustment in the determinate case. Consider the algorithm for solving the problem

$$f(x)\to\min_{x\in X},$$

where the function $f(x)$ is strongly convex.

We shall state a theorem on the convergence of the sequence $\{x^s\}$, specified by relation (2.1), to the minimum point of the function $f(x)$ if

$$\xi^s=\hat{f}_x(x^s),\quad \rho_{s+1}=\rho_s\exp\{a(\hat{f}_x(x^{s+1}),x^s-x^{s+1})\},\quad a>0. \tag{3.1}$$

Theorem 2. Let the function $f(x)$ be strongly convex in the compactum X, i.e.,

$$(\hat{f}_x(x^{s+1}),x^s-x^{s+1})\leqslant f(x^s)-f(x^{s+1})-t\|x^s-x^{s+1}\|^2,\quad t>0,$$

while we also have

$$\|\hat{f}_x(x^s)\|\leqslant c_1,\qquad \max_{x,y\in X}\|x-y\|\leqslant c_2;$$

then sequence $\{x^s\}$, given by algorithm (2.1), (3.1), converges to the minimum point x^* of $f(x)$.

The proof is obtained by verifying the sufficient conditions for the algorithms of nonlinear programming to converge 6, p.17/. A detailed proof may be found in /4/.

4. Convergence of the algorithm with adaptive step adjustment in the stochastic case. We shall now prove a theorem on Cesaro convergence with probability 1 of algorithm (2.1) with step adjustment

$$\rho_{s+1}=\rho_s\exp\{a[(\xi^{s+1},x^s-x^{s+1})-\delta\rho_s]\},\quad a>0,\quad s=0,1,\ldots. \tag{4.1}$$

Theorem 3. Let the function $f(x)$ be convex in a convex compactum X; if conditions (2.2)—(2.4) hold and $\delta\geqslant 2\bar{b}c_2$, relation (2.7) holds for sequence $\{x^s\}$ given by relations (2.1), (4.1).

Proof. We shall first show that condition (2.6) holds for step ρ_s.

Lemma 2.

$$\sum_{s=0}^{\infty}\rho_s=\infty.$$

This is easily proved by reductio ad absurdum.

Lemma 3. We have the convergence relation $\rho_s\to 0$ a.s as $s\to\infty$.

Proof. Put $d_s=(\xi^{s+1}-\hat{f}_x(x^{s+1})-b^{s+1},\ x^s-x^{s+1})$. Since the function $f(x)$ is convex, we have $(\hat{f}_x(x^{s+1}),x^s-x^{s+1})\leqslant f(x^s)-f(x^{s+1})$.

We will estimate the step ρ_{s+1} by using (1.5) and (2.3) and the last inequality:

$$\rho_{s+1}=\rho_s\exp\{a[(\hat{f}_x(x^{s+1}),x^s-x^{s+1})+(b^{s+1},x^s-x^{s+1})+d_s-\delta\rho_s]\}\leqslant \tag{4.2}$$

$$\rho_s\exp\{a[f(x^s)-f(x^{s+1})+\|b^{s+1}\|c_2\rho_s+d_s-\delta\rho_s]\}=\rho_s\exp\Big\{a\Big[f(x^s)-f(x^{s+1})+\Big(\|b^{s+1}\|c_2-\frac{2}{3}\delta\Big)\rho_s+d_s-\frac{\delta}{3}\rho_s\Big]\Big\}=$$

$$\rho_0\exp\Big\{a\Big[\sum_{l=0}^{s}[f(x^l)-f(x^{l+1})]+\sum_{l=0}^{s}\Big(\|b^{l+1}\|c_2-\frac{2}{3}\delta\Big)\rho_l+\sum_{l=0}^{s}\Big(d_l-\frac{\delta}{3}\rho_l\Big)\Big]\Big\}.$$

We will estimate the exponent in relation (4.2). For the first sum we have

$$\sum_{l=0}^{s}[f(x^l)-f(x^{l+1})]=f(x^0)-f(x^{s+1})\leq f(x^0)-f(x^*)=\text{const}. \qquad (4.3)$$

For the second sum, noting the hypothesis concerning $\bar{b}$, we have

$$\limsup_{s\to\infty}\sum_{l=0}^{s}\left(\|b^{l+1}\|c_2-\frac{2\delta}{3}\right)\rho_l<\infty \text{ a.s.} \qquad (4.4)$$

For the third sum we have

$$\sum_{l=0}^{s}\left(d_l-\frac{\delta}{3}\rho_l\right)=\frac{\delta}{3}\sum_{l=0}^{s}\rho_l\left[\left(\frac{\delta}{3}\sum_{l=0}^{s}\rho_l\right)^{-1}\sum_{l=0}^{s}d_l-1\right],\qquad s=0,1,\dots. \qquad (4.5)$$

It is easily seen that $(\Sigma_0^s d_l, D^{s+2})$ is a square summable martingale, and $(\Sigma_0^s\rho_l, D^{s+1})$ is a predictable increasing sequence, whence, using Theorem 4 of /5, p.507/ and Lemma 1, we find that

$$\left(\sum_{l=0}^{s}\rho_l\right)^{-1}\sum_{l=0}^{s}d_l\to 0 \text{ a.s.},\quad s\to\infty, \qquad (4.6)$$

since

$$\sum_{s=0}^{\infty}\left(\sum_{l=0}^{s}\rho_l\right)^{-2}\mathrm{M}(d_s^2|D^{s+1})\leq c_2^4\sum_{s=0}^{\infty}\left(\sum_{l=0}^{s}\rho_l\right)^{-2}\rho_s^2<\infty.$$

From (4.5), (4.6) and Lemma 2, we have

$$\sum_{l=0}^{s}\left(d_l-\frac{\delta}{3}\rho_l\right)\to-\infty \text{ a.s.},$$

whence, in view of (4.2)—(4.4), the lemma follows.

It is easily seen from Lemma 3 and the relation

$$\rho_{s+1}\rho_s^{-1}=\exp\{a[(\xi^{s+1}, x^s-x^{s+1})-\delta\rho_s]\},\qquad s=0, 1,\dots,$$

for the step ρ_s, that condition 2) of Theorem 1 holds.

Thus it only remains to observe that Theorem 3 follows from Theorem 1.

We shall next consider the numerical realization of the stochastic algorithm based on scheme (2.1), (4.1).

5. Numerical algorithm.

Certain heuristic elements are usually needed for the numerical realization of the algorithm in practice.

If the target function is unsmooth, convergence of the recurrence sequence to the minimum is not guaranteed in practice. In this situation, however, the method may be used with fast discovery of the initial approximation for the slower algorithm with programmed step adjustment.

The computational scheme has the drawback inherent in any method with full-step adjustment of the step, namely, slow convergence at a ravine. In spite of these drawbacks, the version of the algorithm given below works much better than the method with programmed step adjustment.

Algorithm. At the start of the computation $s=0$.

Step 1. Computation of the stochastic quasigradient ξ^s.

Step 2. Averaging of the norm of the stochastic quasigradient ξ^s: $G_s=G_{s-1}+(\|\xi^s\|-G_{s-1})D$. At the start of the computation $G_{-1}=0$.

Step 3. Computation of the mean shift in coordinate space (the mean distance between successive approximations) $Q_s=G_s\rho_s$.

Step 4. Check end of computation: if $Q_s<Q_*$ or $s>s_*$, then stop the computation; otherwise proceed to the next step.

Step 5. Compute the scalar product $T_s: T_s=(\xi^s, x^{s-1}-x^s)$.

Step 6. Average the modulus $T_s: z_s=z_{s-1}+(|T_s|-z_{s-1})D$. At the start of the computation $z_{-1}=0$.

Step 7. Adjust the step ρ_s:

$$\rho_s=\rho_{s-1}R^{T_s/z_s}\times\begin{cases}1, & \text{if } T_s>0,\\ u, & \text{if } T_s\leq 0.\end{cases}$$

Step 8. Check the step variations:

$$\rho_s=\begin{cases}3\rho_{s-1}, & \text{if } \rho_{s-1}^{-1}\rho_s>3,\\ 4^{-1}\rho_{s-1}, & \text{if } \rho_{s-1}^{-1}\rho_s<4^{-1},\\ \rho_s & \text{otherwise.}\end{cases}$$

Step 9. Find the next approximation $x^{s+1}=x^s-\rho_s\xi^s$.

Step 10. Project onto the admissible domain: $x^{s+1}=\Pi_X(x^{s+1})$.

Step 11. Pass to step 1, increasing s by unity.

Two stopping criteria are realized in the method. The first is based on the number of iterations. The second is based on the mean shift, which is equal to the product of the mean norm of the quasigradient ξ^s with step ρ_s. When the shift becomes less than the threshold Q_*, the method ceases to operate (steps 3, 4). The step adjustment differs from the theoretical (4.1) in several ways (step 7). First, T_s in the power of R is normalized to a mean value of the modulus T_s. As a result, if say $R=2$, then step ρ_s will vary (either increasing or decreasing) due to the factor R^{T_s/z_s} on average by the factor 2. Instead of the step reduction parameter δ in (4.1), we introduce a supplementary step reduction by means of the coefficient, u, $0<u\leqslant 1$. An additional decrease only occurs when $T_s=(\xi^s, x^{s-1}-x^s)\leqslant 0$.

Since T_s/z_s in the power of R is random, the step ρ_s may in general increase or decrease by an extremely large factor (step 7). To avoid the difference between ρ_s and the previous step ρ_{s-1} being excessive, thresholds are set for the increase or decrease of the step (step 8).

Notice that the projection operation at step 10 is equivalent to the solution of a nonlinear programming problem. In particular, if $X=\{x: Ax\leqslant B\}$ (A is a matrix and B a vector), then realization of the projection amounts to a problem of quadratic programming. We wrote a program in FORTRAN of the Gildret d'Esopo method (deposit No.4689 in the Republican bank of algorithms and programs, 1977), realizing the projection operation for certain types of linear constraints.

6. Recommendations on the choice of algorithm parameters.

The following recommendations are based on the results of numerical test experiments.

1. The mean step variation R, $1<R<3$, is usually put equal to $R=2$.
2. The initial value of the step multiplier ρ_0 is of the order of $\|x^0-x^*\|[\mathrm{M}(\|\xi^0\|)]^{-1}$.
3. The parameter k defines the averaging factor $D=1/k$ in the relations of the method of averaging (steps 2, 6). Usually, we choose k in the range $4\leqslant k\leqslant 5$, when the target function is specified with noise, or $k=1$ if it is specified without noise.
4. The auxiliary step reduction factor u is taken in the range $0.5\leqslant u\leqslant 1$. If the target function is specified with noise, u is mostly taken in the range $0.8\leqslant u<1$. Notice that, if $k>1$, then u may be equal to 1, since the step decreases rapidly even without auxiliary reduction.
5. The mean shift Q_* in the stopping criterion is made of the order of the required accuracy of solution with respect to the x components.
6. The stopping parameter s_* allows the algorithm to stop after a preassigned number of iterations.

7. Results of computing experiments.

Notice that, when solving stochastic problems, as the result of optimization it is desirable to take the averaged values of the coordinates and function after a certain number of successive iterations. Henceforth the averaged value of the coordinate x^s is denoted by $\bar{x}$, and the averaged value of $f(x^s)$, specified in noise, by $\bar{f}(x)$.

Problem I. To minimize $f(x)=\mathrm{M}_\theta f(x, \theta)$, $x=(x_1, x_2)\in R^2$, under condition $x_2\geqslant 1$, where $f(x, \theta)=|x_1|+|x_2|+\theta(x_1+x_2)$.

Here, θ is a random quantity, uniformly distributed in the interval $[-0.5, 0.5]$. The function $f(x)$ is unsmooth. For $f(x)$, the stochastic quasigradient is evaluated by the relation

$$\xi_i^s=\operatorname{sign} x_i^s+\theta^s, \qquad i=1, 2.$$

Here, θ^s is the sample of random θ at the s-th iteration. It is easily seen that, for $\xi^s=(\xi_1^s, \xi_2^s)$, we have the condition

$$\mathrm{M}(\xi^s|x^s)=f_x(x^s).$$

The initial parameters in the algorithm are $R=2, k=5, u=0.9, \rho_0=1$. The initial approximation is $x_1^0=100$, $x_2^0=100$. At the 60-th iteration the step $\rho_{60}=0.00075$. The results for the averaged values of the coordinates and function from the 51-st to 60-th iteration are

$$\bar{x}_1=\frac{1}{10}\sum_{s=51}^{60} x_1^s=-0.00031, \qquad \bar{x}_2=\frac{1}{10}\sum_{s=51}^{60} x_2^s=1.00000,$$

$$\bar{f}(x)=\frac{1}{10}\sum_{s=51}^{60} f(x^s, \theta^s)=1.15580.$$

The exact solution is $x_1^*=0.00000$, $x_2^*=1.00000$, $f(x^*)=1.00000$.

Notice that, in spite of the poor initial approximation and the small initial step, a tolerable solution is obtained by the algorithm after a relatively small number of iterations.

Problem II. To minimize

$$f(x)=\mathrm{M}_\theta f(x, \theta), \quad x\in R^1, \qquad f(x, \theta)=\max\{a(x-\theta),\ b(\theta-x)\}.$$

Here, $a=2, b=4$, θ is a random quantity, uniformly distributed in the interval $[0, 30]$. The function $f(x)$ has the analytical form

$$f(x)=\frac{1}{10}x^2-4x+60.$$

This problem arises when solving single-nomenclature problems of stock control /7/.

The stochastic quasigradient ξ^s is computed from

$$\xi^s=\begin{cases} a, & \text{if } x^s\geqslant\theta^s,\\ -b, & \text{if } x^s<\theta^s,\end{cases}$$

where θ^s is the realization of random θ at the s-th iteration. The exact value of the extremum is $x^*=20.00000$, $f(x^*)=20.00000$. The initial parameters are $R=3$, $k=5$, $u=1$, $\rho_0=1$. The initial approximation is $x^0=-100.00000$, $f(x^0)=1460$. At the 69-th iteration, step $\rho_{69}=0.063$, and at the 131-th, $\rho_{131}=0.0031$. The results for the averaged values of the variable x^s and the function $f(x^s,\theta^s)$ from the 131-th to the 140-th iterations are

$$\bar{x}=\frac{1}{10}\sum_{s=131}^{140}x^s=20.48,\qquad \bar{f}(x)=\frac{1}{10}\sum_{s=131}^{140}f(x^s,\theta^s)=17.74.$$

Notice that, in the neighbourhood of the extremum, the variance of the quasigradient is large, so that it is difficult to obtain a more exact approximation even with programmed step control, even if a good choice is made of the initial step ρ_0 .

We shall quote for comparison the results of solving this problem by the method with programmed step control.

In /8/, for quadratic functions, the following asymptotically optimal programmed step control is proposed: $\rho_s=1/l(s+a)$. The parameter l has to be equal to the least eigenvalue of the Hessian of the target function $f(x)$, i.e., $l=0.2$. We choose $a=1$. At the 240-th iteration the step $\rho_{240}=0.0207469$, the averaged values of the coordinates and function from the 240-th to 249-th iteration are

$$\bar{x}=\frac{1}{10}\sum_{s=240}^{249}x^s=5.37,\qquad \bar{f}(x)=\frac{1}{10}\sum_{s=240}^{249}f(x^s,\theta^s)=45.02.$$

Since the initial approximation is badly chosen, the algorithm with programmed step control does not reach the extremum.

Problem III. The following problem arises when solving multinomenclature problems of stock control /7/.

To minimize

$$f(x)=\mathrm{M}\sum_{i=1}^{5}\max\{a_i(x_i-\theta_i),\ b_i(\theta_i-x_i)\}$$

under the constraints

$$\begin{aligned} x_1+x_2+2x_3+3x_4+x_5&=200,\\ x_1&\leqslant 50,\\ x_2&\leqslant 7,\\ x_3&\leqslant 7,\\ x_4&\leqslant 8,\\ x_5&\leqslant 25,\end{aligned}$$

$$x_i\geqslant 0,\qquad i=1,\dots,5.$$

Here, θ_i are random quantities, uniformly distributed in the intervals $[A_i,B_i]$, $i=1,\dots,5$. The vectors $a=(a_1,\dots,a_5)$, $b=(b_1,\dots,b_5)$, $A=(A_1,\dots,A_5)$, $B=(B_1,\dots,B_5)$ are given:

$$A=(0,0,0,0,0),\qquad B=(60,15,17,90,40),\qquad a=(1,0,3,1,2),\qquad b=(3,4,1,2,3).$$

The analytic expression for the function is

$$f(x)=\frac{1}{3}x_1^2+\frac{2}{15}x_2^2+\frac{2}{17}x_3^2+\frac{1}{60}x_4^2+\frac{1}{16}x_5^2-3x_1-4x_2-x_3-2x_4-3x_5+278.5.$$

The stochastic quasigradient is computed from

$$\xi^s=(\xi_1^s,\dots,\xi_5^s),\qquad \xi_i^s=\begin{cases} a_i, & \text{if } x_i^s\geqslant\theta_i^s,\\ -b_i, & \text{if } x_i^s<\theta_i^s,\quad i=1,\dots,5.\end{cases}$$

The following exact solution was obtained by quadratic programming:

$$f(x^*)=98.10089,$$
$$x^*=(41.88057,\ 7.00000,\ 2.48092,\ 41.27456,\ 22.33456).$$

The initial parameters in the procedure were $R=1.5$, $k=4$, $u=0.9$, $\rho_0=1.0$. The initial approximation was $x^0=(0,0,0,0,0)$, $f(x^0)=278.5$. At the 91-st iteration the step $\rho_{91}=0.1532$. The results for the averaged values of the coordinates and function at the 91-st to 100-th iterations were

$$\bar{x}_i=\frac{1}{10}\sum_{s=91}^{100}x_i^s,\quad i=1,\ldots,5,\qquad \bar{x}_1=40.5485,\quad \bar{x}_2=6.9981,$$

$$\bar{x}_3=2.4381,\qquad \bar{x}_4=42.2561,\qquad \bar{x}_5=20.3561,$$

$$\bar{f}(x)=\frac{1}{10}\sum_{s=91}^{100}f(x^s,\theta^s)=97.4185.$$

For comparison we give below the results of solving the same problem by the method with programmed step control. The initial approximation was the same. In the asymptotically optimal /8/ programmed step control $\rho_s=1/l(s+a)$, the parameter l must be equal to the least eigenvalue of the Hessian of the function $f(x)$, i.e., $l={}^1/_{30}$; we select $a=10$. The results for the averaged values of the coordinates and function at the 161-st to 170-th iterations are: $\bar{x}_1=39.9574$, $\bar{x}_2=6.9554$, $\bar{x}_3=1.5152$, $\bar{x}_4=42.3507$, $\bar{x}_5=24.0324$, $\bar{f}(x)=98.5438$.

If, in the programmed step control, we take different values of the parameters l and a, say, $l=1, a=1$, then, averaging the data from the 161-st to 170-th iterations, we have $\bar{x}_1=22.9804$, $\bar{x}_2=6.9882$, $\bar{x}_3=3.629$, $\bar{x}_4=47.2279$, $\bar{x}_5=21.0516$, $\bar{f}(x)=110.8414$.

The results of the numerical experiment show that, if the target function is sufficiently smooth, the choice of initial approximation is good, and there is detailed information about the target function, then the method with programmed control works as well as that with adaptive control, and in some special cases, better. However, if there is no detailed information about the target function and initial approximation, then an error in the choice of programmed step control parameters may lead to a serious deterioration in the operation of the method. Adaptive control does not have this drawback.

REFERENCES

1. NEMIROVSKII A.S. and YUDIN D.B., Complexity of problems and the efficiency of optimization methods (Clozhnost' zadach i effektivnost' metodov optimizatsii), Nauka, Moscow, 1979.
2. ERMOL'EV YU.M., Methods of stochastic programming (Metody stokhasticheskogo programmirovaniya), Nauka, Moscow, 1976.
3. URAS'EV S.P., Step control for direct methods of stochastic programming, Kibernetika, No. 6, 96-98, 1980.
4. MIRZOAKHMEDOV F. and URAS'EV S.P., Methods of unsmooth optimization with adaptive step control in determinate and stochastic cases, Preprint IK Akad. NAUK UkSSR, Kiev, 1981.
5. SHIRYAEV A.N., Probability (Veroyatnost'), Nauka, Moscow, 1980.
6. NURMINSKII E.A., Numerical methods of solving determinate and stochastic min-max problems (Chislennye metody resheniya determinirovannykh i stokhasticheskikh minimaksnykh zadach), Naukova dumka, Kiev, 1979.
7. MIRZOAKHMEDOV F., On some numerical experiments in solving problems of stock control, in: Methods of operations research and reliability theory in systems analysis (Metody issl. operatsii i teroii nadezhnosti v analize sistem), IK Akad. Nauk AN UkSSR, Kiev, 1977.
8. POLYAK B.T., Convergence and convergence rate of iterative stochastic algorithms, II, The linear case, Avtomatika i telemekhan., No.4, 101-107, 1977.

Translated by D.E.B.

U.S.S.R. Comput.Maths.Math.Phys.,Vol.23,No.6,pp.27-36,1983 0041-5553/83 $10.00+0.00
Printed in Great Britain ©1985 Pergamon Press Ltd.

ON REGULARIZATION OF LEXICOGRAPHIC PROBLEMS BY THE METHOD OF MINIMUM CONCESSIONS*

V.N. NEFEDOV

A method of regularizing multistage lexicographic problems, based on the introduction of "concessions," is studied. The class of lexicographic problems is determined, for which the method of concessions is regularizing with minimal tolerable concessions with respect to the criteria. For convex piecewise polynomial lexicographic problems, an estimate is obtained for the deviation of the regularized from the true solution.

Introduction. The method of regularizing multistage lexicographic problems (l.g.p.), based on the introduction of concessions /1/, was first used for regularizing, l.g.p. in /2, 3/. A multistage l.g.p. consists of the following (see /4/). Let X be a compactum of m-dimensional Euclidean space E^m, and let the functions (criteria) $f_1,\ldots,f_k$ (where $k\geqslant 2$) be defined and continuous in X. We put

**Zh.vychisl.Mat.mat.Fiz.*,23,6,1326-1340,1983