LITERATURE CITED

1. V. S. Korolyuk and A. F. Turbin, Mathematical Principles of Phase Enlargement for Complex Systems [in Russian], Naukova Dumka, Kiev (1978).
2. V. V. Anisimov, "Limiting distributions for the functionals of a semi-Markov process given on a fixed set of states before the first exit," Dokl. Akad. Nauk SSSR, 193, No. 4, 743-745 (1970).
3. I. N. Kovalenko, Analysis of Rare Events in System Performance and Reliability Estimation [in Russian], Sovetskoe Radio, Moscow (1980).

STEP CONTROL FOR DIRECT STOCHASTIC-PROGRAMMING METHODS

S. P. Uryas'ev

UDC 512.98

There is considerable practical interest in efficient numerical methods of stochastic optimization, which can be used in planning under conditions of uncertainty and in optimizing stochastic schemes such as servicing systems, inventory management, and stock control.

A new method is described for step adjustment for stochastic algorithms of gradient type. The step is adjusted from the realizations, which does not require much computation, so the technique has been used with problems of large dimensions. The scheme is based on step completion, which provides an efficient criterion for halting the process.

The convergence is examined via an approach different from the traditional ones [2, 3] because the step size may increase during the process and is not necessarily required to tend to zero.

PRELIMINARY REMARKS

We present some informal arguments that explain the step adjustment. We consider a problem in stochastic programming without constraints [2], i.e., minimization of

$$f(x) = MF(x, \theta),$$

where θ is an elementary event in probability space $(\Theta, \mathfrak{F}, P)$; to minimize $f(x)$ we use the scheme $x^{s+1} = x^s - \rho_s \xi_H^s$, where $\xi_H^s = \xi^s / \| \xi^s \|$ for $\xi^s \neq 0$ and $M(\xi^s / x^0, \ldots, x^s) = \nabla f(x^s)$; if we know the function

$$\varphi_s(\rho) = f(x^s - \rho \xi_H^s) - f(x^s)$$

or $\nabla \varphi_s(\rho)$ then the step ρ_s can be chosen from the condition for a minimum in $\varphi_s(\rho)$ with respect to ρ. The exact values of $\varphi_s(\rho)$ are not known in this case. In a step s it is possible to calculate the scalar product (ξ^s, ξ_H^{s-1}), and it is readily seen that the following chain of equalities applies:

$$M((\xi^s, \xi_H^{s-1})/x^0, \ldots, x^s) = (\nabla f(x^s), \xi_H^{s-1}) = -\left.\frac{\partial f(x^{s-1} - \rho \xi_H^{s-1})}{\partial \rho}\right|_{\rho_{s-1}} = -\nabla \varphi_{s-1}(\rho_{s-1}),$$

i.e., in step s we know a random quantity whose mathematical expectation is the antigradient of $\varphi_{s-1}(\rho)$ at the point ρ_{s-1}. Then if $\varphi_s(\rho)$ and $\varphi_{s-1}(\rho)$ are similar one can approach the minimum in $\varphi_s(\rho)$ by using the gradient procedure $\rho_s = \rho_{s-1} + \lambda_s(\xi^s, \xi_H^{s-1})$, where $\lambda_s > 0$, $s = \overline{0, \infty}$.

Parameter λ_s must be chosen such that ρ_s is greater than zero, so the previous formula is better rewritten as $\rho_s = \rho_{s-1} \cdot a_{s-1}^{(\xi^s, \xi_H^{s-1})}$, where $a_s > 1$ for $s = \overline{0, \infty}$.

In fact, in step s a check is made whether a minimum has been passed through along the ξ_H^{s-1} direction in the preceding step. The sign of the scalar product (ξ^s, ξ_H^{s-1}) enables one to judge this with a certain probability, i.e., an attempt is made at adjusting the step adaptively.

Translated from Kibernetika, No. 6, pp. 85-87, 94, November-December, 1980. Original article submitted January 18, 1979.

0011-4235/80/1606-0886$07.50

CONVERGENCE PROOF

We thus have to minimize $f(x) = MF(x, \theta)$, where $f(x)$ is a smooth complex function that satisfies the Lipshits condition throughout the space R^n of the function.

We consider the sequence of approximations

$$x^{s+1} = x^s - \rho_s \xi^s_H, \tag{1}$$

$$\rho_{s+1} = \rho_s \cdot a^{(\xi^{s+1}, \xi^s_H)} - \Delta, \quad \Delta > 0, \quad a > 1. \tag{2}$$

Vector ξ^s satisfies the following conditions:

a)

$$\left.\begin{aligned} &\xi^s = \nabla f(x^s) + b_1(x^s) + \eta_1(x^s), \\ &\sup_{x \in R^n} \| b_1(x) \| \leqslant b_1, \; M(\eta_1(x^s)/x^s) = 0; \\ &\xi^s_H = \alpha(\nabla f(x^s))_H + b_2(x^s) + \eta_2(x^s), \; \alpha > 0, \\ &\sup_{x \in R^n} \| b_2(x) \| < b_2, \quad M(\eta_2(x^s)/x^s) = 0; \end{aligned}\right\} \tag{3}$$

b)

$$\begin{gathered} |(b_1(x^{s+1}), \eta_2(x^s))| \leqslant \beta, \\ \sup_{x \in R^n} |\beta(x)| \leqslant \beta; \end{gathered} \tag{4}$$

c)

$$\sup_{x \in R^n,\; s=\overline{1,\infty}} \| \xi^s(x) \| < C. \tag{5}$$

We now make some comments on the conditions. It is readily seen that there always exist functions $b_1(x)$, $b_2(x)$, $\beta(x)$ and constants b_1, b_2, β, α such that (3) and (4) are met if (5) is obeyed.

We note also that (3) and (5) imply the following bound for the norm of the gradient $\| \nabla f(x) \| < c + b_1$ for all $x \in R^n$; if $b_1(x)$ is continuous, then condition (4) may be eliminated.

<u>THEOREM.</u> If the following relations apply for C, Δ, b_i, and a:

$$\ln a \cdot (C - \Delta) \leqslant 1, \tag{6}$$

$$\ln a \cdot (C + \Delta)^2 < \Delta - b_1, \tag{7}$$

then all the limiting points of the sequence $\{x^s\}$ generated by the process of (1) and (2) with probability 1 belong to the solution region X^*, where

$$X^* = \left\{ x : \| \nabla f(x) \| \leqslant \frac{b_1 \cdot b_2 + \alpha \cdot b_1 + (b_1 + C)\; b_2 + \Delta + \beta}{\alpha} \right\}. \tag{8}$$

It is always possible to choose Δ such as to meet the conditions of (6) and (7).

<u>LEMMA.</u> The following bound applies:

$$M\left(\sum_0^\infty \rho_s\right) \leqslant \frac{\dfrac{\rho_0}{\ln a} + f(x^0) - f(x^*)}{\Delta - \ln a \cdot (C + \Delta)^2 - b_1},$$

where x^* is the point at which the minimum in $f(x)$ occurs.

Note that the traditional condition $\sum_0^\infty \rho_s = \infty$ [2] for convergence of a stochastic process is not required.

<u>Proof.</u> We make some estimates. As $f(x)$ is a convex function we have

$$-(\nabla f(x^{s+1}), \xi^s_H) \geqslant \frac{f(x^{s+1}) - f(x^s)}{\rho_s}. \tag{9}$$

It is readily shown that the following inequality is obeyed if $p \leq 1$:

$$e^p \leqslant 1 + p + p^2. \tag{10}$$

We now derive a bound for ρ_s. From (2), (5), (6), and (10) we have

$$0 < \rho_{s+1} = \rho_s \cdot a^{(\xi^{s+1}, \xi^s_H) - \Delta} = \rho_s \cdot e^{\ln a((\xi^{s+1}, \xi^s_H) - \Delta)} \leqslant \rho_s (1 + \ln a((\xi^{s+1}, \xi^s_H) - \Delta) + (\ln a((\xi^{s+1}, \xi^s_H) - \Delta))^2) \leqslant \rho_s (1 + \ln a((\xi^{s+1}, \xi^s_H) - \Delta + \ln a(C + \Delta)^2)).$$

We introduce the symbol Ψ_s: $\Psi_s = (\xi^{s+1}, \xi^s_H) - (\nabla f(x^{s+1}), \xi^s_H)$ and use (9) with Ψ_s to get

$$0 \leqslant \rho_s (1 + \ln a(\Psi_s + (\nabla f(x^{s+1}), \xi^s_H) - \Delta + \ln a(C + \Delta)^2)) \leqslant \rho_s \left(1 + \ln a\left(\Psi_s - \frac{f(x^{s+1}) - f(x^s)}{\rho_s} - \Delta + \ln a \cdot (C + \Delta)^2\right)\right) = \rho_s + \ln a(\Psi_s \cdot \rho_s - f(x^{s+1}) + f(x^s) + \rho_s(-\Delta + \ln a(C + \Delta)^2))$$

$$\leq \rho_0 + \ln a\left(\sum_0^s \Psi_l \rho_l + \sum_0^s (-f(x^{l+1}) + f(x^l)) + (-\Delta + \ln a(C + \Delta)^2) \sum_0^s \rho_l\right) = \rho_0 + \ln a\left(\sum_0^s \Psi_l \rho_l - f(x^{s+1}) + f(x^0) + (-\Delta + \ln a(C + \Delta)^2) \sum_0^s \rho_l\right) \leqslant \rho_0 + \ln a \times \left(\sum_0^s \Psi_l \rho_l - f(x^*) + f(x^0) + (-\Delta + \ln a(C + \Delta)^2) \sum_0^s \rho_l\right).$$

From the latter inequality we get $0 \leqslant \rho_0 + \ln a\left(\sum_0^s M(\Psi_l \rho_l) - f(x^*) + f(x^0) + M\left(\sum_0^s \rho_l\right)(-\Delta + \ln a(C + \Delta)^2)\right)$.

We use (3) to get a bound for $\sum_0^s M(\Psi_l \rho_l)$:

$$\sum_0^s M(\rho_l \Psi_l) = \sum_0^s MM(\Psi_l \rho_l / x^{l+1}, x^l) = \sum_0^s MM(\rho_l ((\xi^{l+1}, \xi^l_H) - (\nabla f(x^{l+1}), \xi^l_H)) / x^{l+1}, x^l)$$

$$\sum_0^s MM(\rho_l (b_1(x^{l+1}), \xi^l_H) / x^{l+1}, x^l) \leqslant \sum_0^s MM(\rho_l \| b_1(x^{l+1}) \| / x^{l+1}, x^l) \leqslant \sum_0^s M(\rho_l \cdot b_1) = b_1 M\left(\sum_0^s \rho_l\right).$$

We substitute the last bound in the previous inequality to get $0 \leqslant \rho_0 + \ln a(f(x^0) - f(x^*) + M\left(\sum_0^s \rho_l\right) \cdot (b_1 - \Delta + \ln a(C + \Delta)^2))$.

We pass to the limit in the latter expression to get the assertion in the lemma.

Consequence. $\rho_s \underset{s \to \infty}{\to} 0$ with probability 1.

Subsequently for brevity we will omit the parentheses in $(\nabla f(x^s))_H$ and write $\nabla f_H(x^s)$.

By definition

$$\xi^{s+1} = \nabla f(x^{s+1}) + b_1(x^{s+1}) + \eta_1(x^{s+1}),$$
$$\xi^s_H = \alpha \nabla f_H(x^s) + b_2(x^s) + \eta_2(x^s),$$

where $M(\eta_1(x^{s+1}) / x^{s+1}) = 0$ and $M(\eta_2(x^s) / x^s) = 0$.

The consequence of the lemma implies that $\| x^{s+1} - x^s \| \underset{s \to \infty}{\to} 0$ with probability 1.

We introduce the symbol $T(x^s)$:

$$T(x^s) = (\eta_1(x^{s+1}), \alpha \nabla f_H(x^s) + b_2(x^s) + \eta_2(x^s)) + (\nabla f(x^s), \eta_2(x^s)) + (b_1(x^{s+1}), \eta_2(x^s)) - \beta(x^s).$$

We use the continuity of the vector function $\nabla f(x)$ for the scalar product (ξ^{s+1}, ξ^s_H) to make the following equivalent transformation with probability 1:

$$(\xi^{s+1}, \xi^s_H) = (\nabla f(x^{s+1}) + b_1(x^{s+1}) + \eta_1(x^{s+1}), \alpha \nabla f_H(x^s) + b_2(x^s) + \eta_2(x^s)) = (\nabla f(x^s) + b_1(x^{s+1}) + \eta_1(x^{s+1}), \alpha \nabla f_H(x^s) + b_2(x^s) + \eta_2(x^s)) + D(s).$$

The limit transition $D(s) \underset{s \to \infty}{\to} 0$ applies for $D(s)$.

Using $T(x^s)$ with (3), (4), and (5) together with the previous equality for (ξ^{s+1}, ξ_H^s) we can estimate the latter:

$$(\xi^{s+1}, \xi_H^s) = \alpha (\nabla f(x^s), \nabla f_H(x^s)) + (\nabla f(x^s), b_2(x^s)) + \alpha (b_1(x^{s+1}), \nabla f_H(x^s)) + (b_1(x^{s+1}), b_2(x^s))$$

$$+ \beta + T(x^s) + D(s) \geqslant \alpha \| \nabla f(x^s) \| - (C + b_1)\cdot b_2 - \alpha b_1 - b_1 b_2 - \beta + T(x^s) + D(s).$$

From (2) and the bound for (ξ^{s+1}, ξ_H^s) we have

$$\rho_{s+1} = \rho_0 \cdot a^{\sum_0^s ((\xi^{l+1}, \xi_H^l) - \Delta)} \geqslant \rho_0 \cdot a^{s^{\frac{2}{3}}\left(\frac{1}{s^{\frac{2}{3}}} \sum_0^s (\alpha \| \nabla f(x^l) \| - (C + b_1)\cdot b_2 - \alpha \cdot b_1 - b_1 b_2 - \Delta - \beta + D(l)) + \frac{1}{s^{\frac{2}{3}}} \sum_0^s T(x^l)\right)}. \quad (11)$$

The random quantity $\frac{1}{s^{\frac{2}{3}}} \sum_0^s T(x^l)$ converges in probability to zero, since

$$M\left(\frac{1}{s^{\frac{2}{3}}} \sum_0^s T(x^l)\right)^2 = \left(\frac{1}{s^{\frac{4}{3}}} \sum_0^s M(T(x^l))^2 + \frac{2}{s^{\frac{4}{3}}} \sum_1^s M(T(x^l)\cdot T(x^{l-1}))\right) \underset{s\to\infty}{\to} 0.$$

Using (11), we obtain the assertion in the theorem from the converse. The lemma implies that the sequence $\{x^s\}$ converges with probability 1. The function $\| \nabla f(x) \|$ is continuous, and therefore if $x^s \to x$ is not obeyed with probability 1, where $x \in X^*$, the probability that

$$s^{\frac{2}{3}}\left(\frac{1}{s^{\frac{2}{3}}} \sum_0^s (\alpha \| \nabla f(x^l) \| - (C + b_1) b_2 - \alpha \cdot b_1 - b_1 \cdot b_2 - \Delta - \beta + D(l)) + \frac{1}{s^{\frac{2}{3}}} \sum_0^s T(x^l)\right) \underset{s\to\infty}{\to} +\infty$$

is not zero. Then it follows from (11) that the limit $\rho_{s+1} \underset{s\to\infty}{\to} 0$ is not obeyed with probability 1. The latter conflicts with the lemma.

SOME PRACTICAL RECOMMENDATIONS

We make some comments on the practical implementation of (1) and (2). The criterion for halting the process may be $\rho_s < \varepsilon$, where ε is some sufficiently small positive quantity. In the implementation, it is better not to reduce the step size continuously via parameter Δ but instead to vary a, i.e., one can use the formula

$$\rho_{s+1} = \rho_s \cdot \begin{cases} a_1^{(\xi^s, \xi_H^{s+1})}, & \text{if} \quad (\xi^s, \xi_H^{s+1}) > 0, \\ a_2^{(\xi^s, \xi_H^{s+1})} & \text{if} \quad (\xi^s, \xi_H^{s+1}) \leqslant 0. \end{cases} \quad a_1 < a_2,$$

The coefficients a_1 and a_2 are chosen adaptively in the process in such a way that ρ_s does not vary too sharply. In the deterministic case, it has been demonstrated [4] that this adjustment converges for unsmooth convex functions. If the dispersion of the random quantity is not large, one can use the following scheme, which is highly recommended by practice although few theoretical evaluations have been made:

$$\rho_{s+1} = \rho_s \cdot \begin{cases} \gamma_1, & \text{if} \quad (\xi^s, \xi^{s+1}) > 0, \\ \gamma_2, & \text{if} \quad (\xi^s, \xi^{s+1}) \leqslant 0, \end{cases}$$

$$1 < \gamma_1, \quad 0 < \gamma_2 < 1, \quad \gamma_1 \cdot \gamma_2 < 1.$$

LITERATURE CITED

1. A. M. Gupal, Stochastic Methods of Solving Unsmooth Turning-Point Problems [in Russian], Naukova Dumka, Kiev (1979).
2. Yu. M. Ermol'ev, Stochastic-Programming Methods [in Russian], Nauka, Moscow (1976).
3. E. A. Nurminskii, Numerical Methods of Solving Deterministic and Stochastic Minimax Problems [in Russian], Naukova Dumka, Kiev (1979).

4. S. P. Uryas'ev, "A method of adjusting the step size for the E-subgradient method," in: Behavior of Systems in Random Environments [in Russian], Inst. Kibern. Akad. Nauk UkrSSR, Kiev (1979), pp. 76-84.

A METHOD OF MINIMIZING AN UNDIFFERENTIABLE FUNCTION WITH GENERALIZED-GRADIENT AVERAGING

V. I. Norkin

UDC 519.853.6

A method is given here for unconditional minimization of an undifferentiable function (generalized differential function), in which the direction of descent is chosen from a convex shell of generalized (anti)gradients taken on a fixed number of preceding iterations, while the step size is adjusted by the software. This method resembles the method of [1] taking a position intermediate between relaxation and nonrelaxation methods.

Definition [2, 3]. The function $F(x)$, $x \in R^m$ is called generalized differentiable if there exists a semicontinuous from above point-set mapping $G(F): x \in R^m \to G(F, x) \subset R^m$ such that the sets $G(F, x)$ are bounded, convex, and closed, and at each point $y \in R^m$ the following applies:

$$F(x) = F(y) + (g(x), y - x) + o(y, x, g), \tag{1}$$

where $o(y, x, g)/|x - y| \to 0$ uniformly for $x \to y$ and $g \in G(F, x)$. The elements of set $G(F, x)$ are called the generalized gradients of F at point x.

The class of generalized differentiable functions contains continuously differentiable ones, convex functions, and concave functions, and it is closed under the finite operations of maximum, minimum, and superposition. The gradients of continuously differentiable functions and the subgradients of convex functions are generalized gradients of these. To calculate the generalized gradients of complicated functions one has rules analogous to the rules for calculating ordinary gradients. A generalized differentiable function satisfies the local Lipshits condition. A necessary condition for a turning point in F at x is $0 \in G(F, x)$ [2, 3].

To minimize F we use an algorithm:

$$x^0, x^1, \ldots, x^q \in R^m, \tag{2}$$

$$x^{k+1} = x^k - \rho_k \cdot P^k(x^k, g^k, \ldots, x^{k-n_k}, g^{k-n_k}), \quad k \geq q, \tag{3}$$

$$\rho_k \geq 0, \quad \rho_k \to 0, \quad \sum_{k=0}^{\infty} \rho_k = \infty, \tag{4}$$

$$P^k = \sum_{r=k-n_k}^{k} \lambda_{kr}(x^k, g^k, \ldots, x^{k-n_k}, g^{k-n_k})\, g^r, \tag{5}$$

$$g^r \in G(F, x^r), \tag{6}$$

$$0 \leq n_k \leq \min(n, k), \quad n = \text{const}, \tag{7}$$

$$\sum_{r=k-n_k}^{k} \lambda_{kr}(x^k, g^k, x^{k-1}, g^{k-1}, \ldots, x^{k-n_k}, g^{k-n_k}) = 1, \tag{8}$$

$$|P^k| \leq M < \infty. \tag{9}$$

The following minimizing property of the algorithm of (2)-(9) applies.

THEOREM 1. Let $y_s \to y$, $0 \in \widetilde{G}(F, y)$; for each s we consider the sequence $\{x_s^k, k \geq \max(0, s - n)\}$ formed in accordance with the following rules:

$$x_s^k = y_k, \quad k \leq s; \tag{10}$$

$$x_s^{k+1} = x_s^k - \rho_k \cdot P_s^k, \quad k > s, \tag{11}$$

$$P_s^k = P^k(x_s^k, g_s^k, \ldots, x_s^{k-n_k}, g_s^{k-n_k}), \quad g_s^k \in G(F, x_s^k). \tag{12}$$

Translated from Kibernetika, No. 6, pp. 88-89, 102, November-December, 1980. Original article submitted September 13, 1978.

 0011-4235/80/1606-0890$07.50