Journal of the Nigerian Association of Mathematical Physics,
***Volume 8* (*November 2004*)**

A stochastic iteration method for the solution of finite dimensional variational inequalities

A. C. Okoroafor and **B. O. Osu**
Department of Mathematics
Abia State University, Uturu, Nigeria.

Abstract

Let A be a real $n \times n$ ***matrix, let b a real column n-vector and*** φ ***:*** $R^n \to R$ ***such that*** $Ax + \partial\varphi(x) \ni b$ ***where*** $\partial\varphi$ ***is the sub gradient of*** φ***. A computable stochastic iterative scheme is suggested; which is a modification of Robbins-Monroe procedure and studied in the context of the above concrete problem. This scheme is shown to converge strongly to the solution of the above problem.***

pp 301 - 304

1.0 Introduction

Let $A = (a_{ij})$ be a real positive definite $n \times n$ matrix and b a real column n-vector. For x, y in R^n, Euclidean n-dimensional space, let $(x, y) = \sum_{i=1}^{n} x_i y_i$ and $\|x\|^2 = (x, x) = x'x$ where x' denotes the transpose of x in R^n.

For a convex function φ, not necessarily differentiable, it is well known that if $D(\varphi) = \{x \in R^n : \varphi(x) \pi \infty\} \neq \phi$, then for $x \in D(\varphi)$ the sub gradient $\partial\varphi$ of $\varphi : R^n \to R$ at x is defined as $\partial\varphi(x) = \{g \in R^n : f(x+t) - f(x) \geq \pi\, g, t\, \phi\} \ \forall x + t \in D(\varphi)$ (1.1)

and it is a monotone. We consider the finite dimensional variational problem: Find $x \in D(\varphi)$ such that

$$Ax + \partial\varphi(x) \ni b \qquad (1.2)$$

This is a special case of a generalized equation consisting typically of a smooth part h_1 and a multi-valued non-smooth part h_2 as expressed in the form $h_1(x) + h_2(x) \ni b$ (1.3)

which has important applications in physical and engineering sciences and in many other fields (see for instance [3])?

When h_1 is the gradient of a real valued differentiable convex function H_1 and h_2, the sub-gradient of a proper lower semi continuous convex function H_2, the variational problem reduces to the search for the minimum of the non-smooth function $H_1(x) + H_2(x) + b'x$ (1.4)

so that the problem (1.2) is equivalent to minimizing the function f defined as

$$f(x) = \tfrac{1}{2} x'Ax + \varphi(x) - b'x \qquad (1.5)$$

A number of procedures are available for solving such problems (see for example [9])
The form of (1.5) suggests a reformation of the original multi-valued problem as a search for zero of a single-valued section of the non-smooth function ∂f. In this paper, a modified stochastic gradient type recursive sequence is suggested: $x_{j+1} = x_j - \rho_j d_j$ (1.6)

where d_j is the estimate of a single-valued section g of ∂f and $\{\rho_j\}$ is a sequence of positive scalars to be specified. This procedure is a way of stochastically solving the equation.

$$\{x^* : \partial f(x^*) = 0\} \qquad (1.7)$$

Stochastic approximation algorithms of different types have long been studied in many contexts (see for example [5]). The stochastic iteration method in this paper and some other stochastic algorithms differ mainly in the way the gradient vector and starting point of the algorithm are estimated to accelerate the convergence of the sequence.

2.0 **Mathematical formulation of the stochastic iteration method**

We associate with each random vector $u,v \in R^n$ and a fixed $a \in R^n$ the expectation operator E such that $E\,u$ is defined by the requirement that $E\langle a,u\rangle = \langle a, E\,u\rangle$ if $E\|u\| < \infty$ where $\|u\|, \langle u,v\rangle, \langle a,u\rangle$ are random

variables in the usual sense. For f defined in (1.5) we can see that $\lim_{t\to\infty} f(t\,x) = +\infty$ for any $x \in R^n$, $x \neq 0$ so that there exists a minimum of f in R^n and every minimizing sequence converges to the minimum of f.

Using (1.1), it is easy to see that for $g \in \partial f$, a single-valued section of ∂f,

$$f(x+t) - f(x) \geq \langle g,t\rangle \qquad (2.1)$$

for every $x + t \in D(f)$. We obtain an estimate d of a single-valued section g of ∂f by a noise corrupted measurement that adequately approximates g in the sense that $E\|d-g\| = 0$ (2.2)

and $E\|d-g\|^2$ is minimum. So that at each iteration of the stochastic sequence $\{x^k\}_{k=1}^{\infty}$, defined by (1.6), the estimated gradient vector is used to determine the direction of search, which provides the maximum rate of decrease in $f(x)$. In this connection, let, $t_j = (t_{1j}, t_{2j}, \ldots t_{nj}) \in R^n$ and

$$y_j = y(x_j) = f(x^k + t_j) - f(x^k), x^k \in D(f) \qquad (2.3)$$

for fixed k, and $j=1, \ldots, m$, $n+1 \leq m \leq \frac{1}{2}n(n+1) - 1$. Exploiting the fact that each point $x \in D(f)$ allows supporting hyper planes, so that if points x_1, x_2, Λ, x_m in R^n are chosen in the neighborhood of $x^k, t_j = x_j - x^k$ for a fixed k, then the relationship between y_j and t_j for $j = 1,\ldots,m$ is adequately approximated by

$$y_j = \langle g^k, t_j\rangle + e_j \qquad (2.4)$$

for some single-valued section $g^k \in \partial f(x^k)$ where y_j and $e_j = e(x_j)$ are respectively the independent observable random variables corresponding to the trial points $x_j \in R^n$ for fixed k and the random error of the j^{th} observation with $Ee(x_j) = 0$ and $E[e(x_i)e(x_j)] = \delta^2 \delta_{ij}$, $i, j = 1, \Lambda ., m$, $0 \,\pi\, \sigma^2 \,\pi\, \infty$

This idea was used in [6] to show that

Theorem 2.1

Let $\{\rho^k\}$ be a real sequence such that (i) $\rho^0 = 1, 0 < \rho^k < 1 \;\forall k < 1$ (ii) $\sum_{k=0}^{\infty} \rho^k = \infty$ (iii) $\sum_{k=0}^{\infty} \rho^{2k} < \infty$, then the sequence $\{x^k\}_{k=o}^{\infty}$ generated by $x^0 \in D(f)$ and defined iteratively by $x^k - \rho^k d^k$,

d^k, a least square estimate of the single-valued section $g^k \in \partial f(x^k)$ remains in $D(f)$ and converges strongly to $\{x^* : \partial f(x^*) = 0\}$

This approximation scheme turns out to be adequate since convex figures of small area are well approximated by an interval (see for example [4]).

For the case in which $\varphi \equiv 0$, it has been shown (see for instance [7]) that when x_1, x_2, Λ, x_m are chosen in the neighborhood x^k of such that $\sum_{j=1}^{m} t_{ij} = 0$ and $\sum_{j=1}^{m} t_{ij}^2 = 1$ (2.5)

Then this choice of t_j linearizes the function f so that the least squares approximation

$$d^k = M^{-1}\sum_{j=1}^{m} t_j y_j = 0,\ M = \sum_{j=1}^{m} t_j t_j' \qquad (2.6)$$

exists and is adequate for approximating g^k such that $E\| g^k - d^k \| = 0$ for such k and yields a minimum Euclidean distance between the true and the estimated gradient vector $E\| g^k - d^2 \|^2$. An easy calculation shows that $E\| g^k - d^2 \|^2 = 0$ and $E\| g^k - d^2 \|^2 = M^{-1}\sigma^2$ for each k. (2.7)

From the foregoing, it can be seen that the use of this scheme is justified. In the sequel, we assume without lost of generality, that $\sigma^2 = 1$.

3.0 Modification

In this section we attempt to improve on the convergence of the iteration in (1.6) by segmentation. This is a useful technique in accelerating the convergence of the algorithm (see for example [8]).

Let R^n be partitioned into z exclusive segments $s_j,\ j = 1,.\Lambda\ ,z, n < z \le 2^n$. Let x_j be chosen randomly in s_j such that $f(x_j) > 0$ or $f(x_j) < 0\ \ \forall j$. Let $Pr(x_j = \alpha) = P_j$ be the probability that $x_j = \alpha,\ P_j \ge 0,\ \sum_{j=1}^{z} P_j = 1$. Put $P_j = \frac{f(x_j)}{\sum_{j=1}^{z} f(x_j)}$ so that $\bar{x} = \sum_{j=1}^{z} x_j P_j = \sum_{j=1}^{z} \frac{x_j f(x_j)}{\sum_{j=1}^{z} f(x_j)}$ (3.1)

Let

$$x^* = \bar{x} - \rho d, \rho \pi 0 \qquad (3.2)$$

where d is an estimate of $g \in \partial f$. But $f(x+t) - f(x) \ge \langle g,t \rangle$ by (1.1). Thus

$f(\bar{x}) - f(x^*) \ge \rho\langle g,d \rangle \ge 0$ and $f(\bar{x}) \ge f(x^*)$

(3.3)

Since f is convex, from (3.1) and (3.3) we have

$$f(\sum_{j=1}^{z} x_j P_j) \ge f(x^*) \forall j \qquad (3.4)$$

and

$$\sum P_j f(x_j) \ge f(x^*) \forall j \qquad (3.5)$$

Hence $\min_j \sum P_j f(x_j) = min\ f(x_j) \ge f(x^*)$, so that

$$f(x^*) = \min_j \{\ f(x_j) : x_j \in S_j\ \} \qquad (3.6)$$

Under the above conditions, we can make the following remark

***Remark* 3.1**

The segment S_T where $x^* \in S_T$ contains x for which $f(x)$ is minimum.

Thus we discard the other segments that do not contain x^*. Then x^* forms the starting point of our search.

Furthermore, the gradient direction estimated as a result of set of trial points in R^n differs from that of the true gradient due to experimental error.

However, the direction of search would be correspondingly uncertain and so may slow the rate of convergence of the sequence (see for example [1]).

Transforming the estimated gradient vector d is capable of reducing the Euclidean distance between the true and estimated gradient direction (see for example [8]). To this end, we state the following:

Lemma **3.1**

Let $T = diag\{\alpha_1, \alpha_2 \Lambda, \alpha_n\}$, $\alpha_1 = \alpha_2 = .\Lambda, \alpha_n$, $\sum_{i=1}^{n} \alpha_i = 1$ (3.7)

and let d be the least squares estimate of the single-valued section $g \in \partial f$ as in (2.6). Then $\| g - Td \| \le \| g - d \|$ for any d.

Proof

Let $M = \sum_{j=1}^{m} t_j t_j'$. But $E\|g-d\|^2 = M^{-1}$ as in (2.6). Hence, $E\|g - Td\|^2 = TM^{-1}T^1$

$= T^2 M^{-1} = \alpha M^{-1}, 0 \pi \alpha \pi 1$

$= E\|g^k - d^k\|^2$. It is easy to see that the choice of T minimizes the Euclidean distance so that

$\|g - Td\| \le \|g - d\|$ for all the least-squares approximation d of g. Thus, instead of search for the minimum of f in the direction of d, we consider an iterative scheme started at $\bar{x}$ defined in (3.1), which minimizes f successively in the direction of the stochastic independent vectors $\{Td^k\}_{k=1}^{\infty}$, $d^k = M^{-1} \sum_{j=1}^{m} t_j y_j$ along the line $x^k - \rho^k Td^k$ as follows:

1. Compute $g^k \approx Td^k$
2. Compute the corresponding ρ^k
3. Compute $x^{k+1} = x^k - \rho^k Td^k$
4. Has the process converged?

Is $\|x^{k+1} - x^k\| \pi \delta, \delta \pi 0$? If yes, then $x^{k+1} = x^n$. Where if no, we show that this sequence converges strongly to the solution of problem (1.2)

Theorem **3.1**

Let $\{\rho^k\}_{k=0}^{\infty}$ be a real sequence satisfying (i) $\rho^0 = 1, 0 < \rho^k < 1 \forall k > 1$ (ii) $\sum_{k=0}^{\infty} \rho^k = \infty$ (iii) $\sum_{k=0}^{\infty} \rho^{2k} < \infty$ then the stochastic sequence generated by $\bar{x}$ and defined iteratively by $x^{k+1} = x^k - \rho^k Td^k$, remains in $D(\partial f)$ and converges strongly to $\{\hat{x} : \partial f(\hat{x}) = 0\}$

Proof

Let $D_k = \rho^k \|g^k - Td^k\|$. Then $\{D_k\}$ is a sequence of independent random variables. From (2.6) $ED_k = 0$ for each k, thus the sequence of partial sums $\eta_k = \sum_{j=1}^{k} D_j$ is a Martingale. But $E\eta_k^2 = \sum_{j=1}^{k} ED_j^2 = \sum_{j=1}^{k} \rho^{2j} E\|g^j - Td^j\|^2 \le M^{-1} \sum_{j=1}^{k} \rho^{2j}$. Since $\sum \rho^{2j} < \infty$ hence, $\sum_{j=1}^{\infty} ED_j^2 < \infty$. So that by a version of Martingale convergence theorem [10], we have $\sum_{k=1}^{\infty} D_k < \infty$. Thus $\lim_{k \to \infty} p^k \|g^k - Td^k\| = 0$.

An earlier result in the theory of accretive operations, due to Chidume [2] shows that the sequence, $\{x^k\}_{k=0}^{\infty}$, generated by $x^0 \in D(f)$ and defined iteratively by $x^{k+1} = x^k - \rho^k g^k$, $g^k \in \partial f(x^k)$ a single section of ∂f, remains in $D(\partial f)$ and converges strongly to $(x^* : \partial f(x^*) x)$. It follows from his result that our sequence converges strongly to the solution of the problem (1.2). The convergent rate of this scheme is further improved if, as in Remark 3.1, the segment S_T for which f attains its minimum is further segmented into z disjoint sub segments $S_{T_2,j}, j=1,\Lambda,z$ and the point $x_j \in S_{T_{1,j}}$ is chosen such that $f(x^*) \geq f(x_j)$ for each j where the subscript, 1 denotes the first sub segmentation process. Then, define $\overline{x}_{T_1} = \sum_{j=1}^{z} \frac{x_j f(x_j)}{\sum_{j=1}^{z} f(x_j)}$, $x_j \in S_{T_{1,j}}$ so that by (3.2) and (3.6) $f(x^*{}_{T_1}) = min\{ f(x_j) : x_j \in S_{T_{1,j}} \}$

We discard the $j-1$ segments, which do not contain the point $x^*_{T_1}$, and denote the remaining segment, which contains $x^*_{T_1}$ by S_{T_2}. Then f attains its minimum on S_{T_2}. S_{T_2} is further segmented into z disjoint sub segments $S_{T_2,j}, j=1,\Lambda,z$ and the process repeated until $\left\| x^*{}_{T_i} - x^*{}_{T_{i+1}} \right\| \pi\ \varepsilon, \varepsilon\ \phi\ 0$ where $x^*T_0 = x^*$. This technique accelerates the convergence of the method indicated in [1] and extends to the solution of variational inequality.

References

[1] Brooks S.H and Mickey M.R (1961) "Optimum estimation in steepest ascent experiments" Biometrics vol.17 48-56

[2] Chidume C.E (1995) " Iterative solution of the Non linear Equations with strongly Accretive operators" J. Math and Appl. 192, 502-518

[3] Duruat .G and Lions J.C (1976)"Inequalities in physics and Mechanics" Springer Verlag, Berlin

[4] Levin A. Ju (1965) "On an algorithm for the minimization of convex functions" Soviet mathematics 6, 286-290

[5] Lying .L (1978) "Strong Convergence of a stochastic algorithm" Annal of statistics vol. 6 No.8 680-696

[6] Okoroafor A.C and Ekere A.E (1999) "A stochastic Approximation method for the Attractor of a Dynamical system"In Directions in mathematics (G.O.S Ekhagwere and O.O. Ugbebor, eds) Assoc. Books 131-141

[7] Okoroafor A.C. and Akpanta A.C (2003) "A stochastic iterative method for the solution of Reynold's Equation for a journal bearing".Global journal of pure and applied sciences vol. 9 No 1 157-162

[8] Onukogu I.B. and Ezele I.I. (1987) "Response surface Exploration with spline function: Biased Regressors" statistica, anno XLVII. n.2 287-298

[9] Uko L.U (1992) "The Numerical solution of finite Dimensional variational inequality" Journal of Nigerian Mathematical Society vol. II No.1 49-63

[10] Whittle P. "Probability" (1976) John Wiley and Sons