Proceedings of the 1995 Winter Simulation Conference
ed. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman

ON THE PERFORMANCE OF PURE ADAPTIVE SEARCH

Bruce W. Schmeiser

School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907–1287, U.S.A.

Jin Wang

Department of Mathematics and Computer Science
Valdosta State University
Valdosta, Georgia 31698–0040, U.S.A.

ABSTRACT

We study Pure Adaptive Search (PAS), an iterative optimization algorithm whose next solution is chosen to be uniformly distributed over the set of feasible solutions no worse than the current solution. We extend the results of Patel, Smith, and Zabinsky (1988) and Zabinsky and Smith (1992). In particular, we (1) show that PAS converges to the optimal solution almost surely, (2) show that each PAS iteration reduces the expected remaining feasible-region volume by 50%, and (3) improve the Patel, Smith, and Zabinsky (1988) complexity measure for convex problems.

1 INTRODUCTION

We consider the mathematical programming problem

$$\sup_{x \in S} z(x) \tag{1}$$

where x is a k dimensional vector of decision variables, the feasible region S is a Borel measurable subset of R^k, and z is a bounded measurable objective function on S. We assume that S is closed and bounded, and that z is continuous at its optimal points. Define

$$x^* = \arg \sup_{x \in S} z(x)$$

and

$$z^* = \sup_{x \in S} z(x) \ .$$

Then (x^*, z^*) is the optimal solution of (1).

Various random-search methods have been suggested for solving such problems (see, for early examples, Anderson 1953, Brooks 1958, Rastrigin 1963 and Karnopp 1963). Patel, Smith, and Zabinsky (1988) and Zabinsky and Smith (1992) study PAS, which moves from the current solution x_i to the next solution x_{i+1} that is generated randomly and uniformly from the set of all better feasible solutions.

We extend convergence-rate and complexity-measure results for Pure Adaptive Search (PAS). In Section 2 we review PAS, in Section 3 we show almost-sure convergence of PAS, in Section 4 we show that each iteration of PAS reduces the expected remaining-feasible-region volume by 50%, and in Section 5 we improve complexity bounds.

2 THE PAS METHOD

The PAS method for problem (1) is

Step 0. Set $n = 0$, $S_0 = S$. Select a point $X_0 \in S$ and set $Z_0 = z(X_0)$;
Step 1. Generate X_{n+1} uniformly distributed in $S_{n+1} = \{x : \ x \in S_n \text{ and } z(x) > z(X_n)\}$;
Step 2. Set $Z_{n+1} = z(X_{n+1})$, $n = n + 1$. Go to Step 1.

Here S_{n+1} is the set of feasible solutions better than the random current solution X_n, for $n = 0, 1, \ldots$. The sequence of objective-function values $z(X_n)$ is a Markov chain, as is any function of $z(X_n)$.

A difficult implementation question for PAS is how best to generate the random variable X_{n+1} uniformly distributed over S_{n+1} in Step 1. Patel, Smith and Zabinsky (1988) suggested using conventional approaches, such as the rejection and transformation techniques such as discussed in Devroye (1986) and Schmeiser (1980). But efficiently generating random points in high-dimensional regions is difficult. Therefore, we view PAS primarily as a theoretical benchmark against which other methods can be compared.

3 CONVERGENCE

We show in this subsection that the PAS Markov chain of objective-function values $\{Z_n, \ n \geq 1\}$ converges to the optimal value z^* almost surely.

Proposition 3.1 *The PAS Markov chain $\{Z_n,\ n \geq 1\}$ converges to the optimal objective-function value z^* almost surely.*

Solis and Wets (1981) show that a class of global-search algorithms converge in probability to the global optimum. Any such algorithm that provides a monotonic sequence of solutions then converges almost surely. We can not use their result directly, however, because their primary assumption H_2 — that the algorithm has a positive probability of returning to every positive-volume subset A of S — is not satisfied by PAS. But nothing is lost in their proof if H_2 is relaxed to require only that the algorithm can always return to the optimal region $R_{\epsilon,M}$, as defined in their paper. Their proof then implies that PAS converges in probability to the global optimum. Because $\{Z_n,\ n \geq 1\}$ is monotone, almost sure convergence is obtained.

4 CONVERGENCE RATES

We now show that each iteration of PAS reduces the expected remaining volume by one half. Let V_n denote the volume of $S_n = S \bigcap \{x : z(x) > z(X_{n-1})\}$, the random remaining feasible region. Because V_n is a function of $z(X_{n-1})$, then $\{V_n,\ n \geq 1\}$ is a Markov chain with initial value V_0, the volume of S.

Theorem 4.1 *The ratios of PAS volumes*

$$\frac{V_1}{V_0}, \frac{V_2}{V_1}, \frac{V_3}{V_2}, \ldots$$

are i.i.d. Uniform(0,1) random variables with

$$\mathrm{E}\left(\frac{V_n}{V_{n-1}}\right) = \frac{1}{2}$$

and

$$\mathrm{E}V_n = \left(\frac{1}{2}\right)^n V_0, \quad \forall\ n \in \{1, 2, \ldots\}.$$

Proof: We first show uniformity, then independence. For all $r \in [0,1]$,

$$\begin{aligned} \mathrm{P}\left(\frac{V_n}{V_{n-1}} \leq r\right) &= \mathrm{EP}(V_n \leq rV_{n-1} \mid V_{n-1}) \\ &= \mathrm{E}\left(\frac{rV_{n-1}}{V_{n-1}}\right) = \mathrm{E}r = r, \end{aligned}$$

and therefore

$$\mathrm{E}\left(\frac{V_n}{V_{n-1}}\right) = \frac{1}{2}.$$

To demonstrate independence, consider an arbitrary point $(r_1, r_2, \ldots, r_n) \in [0,1]^n$. Then by the Markov property

$$\begin{aligned} &\mathrm{P}\left(\frac{V_n}{V_{n-1}} \leq r_n \,\middle|\, \frac{V_{n-1}}{V_{n-2}} = r_{n-1}, \ldots, \frac{V_1}{V_0} = r_1\right) \\ = \ &\mathrm{P}(V_n \leq r_n r_{n-1} \cdots r_1 V_0 \mid \\ &\qquad V_{n-1} = r_{n-1} r_{n-2} \cdots r_1 V_0, \ldots, V_1 = r_1 V_0) \\ = \ &\mathrm{P}(V_n \leq r_n r_{n-1} \cdots r_1 V_0 \mid \\ &\qquad V_{n-1} = r_{n-1} r_{n-2} \cdots r_1 V_0). \end{aligned}$$

Because the ratios of the volumes are Uniform(0,1), this probability equals

$$\frac{r_n r_{n-1} \cdots r_1 V_0}{r_{n-1} r_{n-2} \cdots r_1 V_0} = r_n = \mathrm{P}\left(\frac{V_n}{V_{n-1}} \leq r_n\right).$$

Therefore $\frac{V_1}{V_0}, \frac{V_2}{V_1}, \frac{V_3}{V_2}, \ldots$ are independent Uniform (0,1) random variables. By the i.i.d. property, the expected volume after n iterations is

$$\begin{aligned} \mathrm{E}V_n &= \mathrm{E}\left(\frac{V_n}{V_{n-1}} \frac{V_{n-1}}{V_{n-2}} \cdots \frac{V_1}{V_0} V_0\right) \\ &= \left\{\mathrm{E}\left(\frac{V_1}{V_0}\right)\right\}^n V_0 = \left(\frac{1}{2}\right)^n V_0. \end{aligned}$$

□

5 COMPLEXITY

We now improve a complexity bound of Patel, Smith, and Zabinsky (1988). Let $K_{\alpha,m}$ denote the number of iterations to ensure an m-fold improvement with probability at least $1-\alpha$, i.e.,

$$K_{\alpha,m} = \min_{n \in \{1,2,\ldots\}} \left\{n : \mathrm{P}\left(D_n \leq \frac{1}{m}\right) \geq 1 - \alpha\right\},$$

where $D_n = (z^* - Z_n)/(z^* - Z_0)$ is the standardized remaining objective function after n iterations.

For the standard convex program, Patel, Smith, and Zabinsky (1988) obtain the bound

$$K_{\alpha,m} \leq 2(k+1)\ln\left\{m\left(1 + \frac{1}{\sqrt{\alpha}}\right)\right\}, \quad \forall \alpha \in (0,1) \text{ and } \forall\ m \in (1, \infty),$$

via Chebyshev's inequality. We first obtain a minor improvement in the bound by using the Cantelli inequality in Billingsley (1986, p. 76).

$$\mathrm{P}(X - \mathrm{E}(X) \geq \alpha) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \alpha^2}, \quad \alpha \geq 0,$$

which implies

$$K_{\alpha,m} \leq 2(k+1)\ln\left\{m\left(1+\sqrt{\frac{1-\alpha}{\alpha}}\right)\right\},$$
$$\forall \alpha \in (0,1) \text{ and } \forall\, m \in (1,\infty).$$

We more substantially improve the bound by using the inequality $\mathrm{P}(D_n > \mathrm{E}(D_n)/\alpha) \leq \alpha$ in Billingsley (1986, p. 74).

Theorem 5.1 $\forall \alpha \in (0,1)$ *and* $\forall\, m \in (1,\infty)$,

$$K_{\alpha,m} \leq (k+1)\ln\left(\frac{m}{\alpha}\right).$$

Proof: Since $\alpha > 0$ and $\mathrm{E}(D_n) > 0$, then

$$\mathrm{P}\left(D_n \geq \frac{\mathrm{E}(D_n)}{\alpha}\right) = \mathrm{P}\left(\frac{\alpha D_n}{\mathrm{E}(D_n)} \geq 1\right)$$
$$= \int_{\{\alpha D_n/\mathrm{E}(D_n)\geq 1\}} d\mathrm{P} \leq \int_\Omega \frac{\alpha D_n}{\mathrm{E}(D_n)} d\mathrm{P}$$
$$= \frac{\alpha \mathrm{E}(D_n)}{\mathrm{E}(D_n)} = \alpha.$$

Therefore

$$\mathrm{P}\left(D_n < \frac{\mathrm{E}(D_n)}{\alpha}\right) > 1-\alpha.$$

Comparing this inequality to the definition of $K_{\alpha,m}$, we have $\mathrm{E}(D_n) = \alpha/m$. From Patel, Smith, and Zabinsky (1988) we have $\mathrm{E}(D_n) \leq (k/(k+1))^n$. Solving for n yields

$$n = \frac{\ln\left(\frac{\alpha}{m}\right)}{\ln\left(\frac{k}{k+1}\right)} = \frac{\ln\left(\frac{m}{\alpha}\right)}{\ln\left(1+\frac{1}{k}\right)}.$$

Because $\ln\left(1+\frac{1}{k}\right) \geq \frac{1}{k+1}$,

$$n \leq \frac{\ln\left(\frac{m}{\alpha}\right)}{\left(\frac{1}{k+1}\right)} = (k+1)\ln\left(\frac{m}{\alpha}\right).$$

Therefore,

$$K_{\alpha,m} \leq (k+1)\ln\left(\frac{m}{\alpha}\right).$$

□

Table 1 gives upper bounds for the number of iterations $K_{\alpha,m}$ versus dimension k required to obtain a million-fold improvement with 99% certainty, the case considered in Patel, Smith and Zabinsky (1988). Three integer bounds are shown; the original Patel, Smith and Zabinsky (1988) bound K_1, the Cantelli-inequality bound K_2, and the Theorem 5.1 bound K_3. Specifically

$$K_1 = \left\lceil 2(k+1)\ln\left\{m\left(1+\frac{1}{\sqrt{\alpha}}\right)\right\}\right\rceil,$$
$$K_2 = \left\lceil 2(k+1)\ln\left\{m\left(1+\sqrt{\frac{1-\alpha}{\alpha}}\right)\right\}\right\rceil,$$
$$K_3 = \left\lceil (k+1)\ln\left(\frac{m}{\alpha}\right)\right\rceil,$$

where $\lceil s \rceil$ is the smallest integer not less than s.

Table 1: Upper Bounds for $K_{\alpha,m}$ versus Dimension k for $\alpha = 0.01$ and $m = 10^6$

Dimension	Number of iterations		
k	K_1	K_2	K_3
1	65	65	37
2	98	98	56
10	357	357	203
100	3276	3275	1861
1000	32460	32451	18440
10000	324301	324210	184226

For these values of α and m, the difference between K_1 and K_2 is negligible; the difference between K_1 and K_3 is substantial, a reduction of about 43 percent.

The ratios of these numbers of iterations are not functions of the dimensionality k. For any value of α, in the limit as $m \to \infty$, the ratio $K_1/K_3 = 2$; in Table 1 the ratio is 1.76. Proving that $K_3 < K_1$ for all values of α and m is straightforward.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grants DMS-8717799 and DDM-93-00058. The work of the second author was also supported in part by the Faculty Development and Research Grant from Valdosta State University.

REFERENCES

Anderson, R. L. 1953. Recent advances in finding best operating conditions. *Journal of the American Statistical Association* 48:789–798.

Billingsley, P. 1986. *Probability and Measure*. 2d ed. New York: John Wiley & Sons.

Brooks, S. H. 1958. A discussion of random methods for seeking maxima. *Operations Research* 6:244–251.

Devroye, L. 1986. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Karnopp, D. C. 1963. Random search techniques for optimization problems. *Automatica* 1:111–121.

Patel, N. R., R. L. Smith, and Z. B. Zabinsky. 1988. Pure adaptive search in Monte Carlo optimization. *Mathematical Programming* 43:317–328.

Rastrigin, L. A. 1963. The convergence of the random search method in the extremal control of many-parameter systems. *Automation and Remote Control* 24:1337–1342.

Schmeiser, B. W. 1980. Random variate generation: A survey. In *Simulation with Discrete Methods: A State-of-the-Art View*, ed. T. I. Ören, C. M. Shub, and P. F. Roth, 79–104, New York: IEEE.

Solis, F. J., and R. J. Wets. 1981. Minimization by random search techniques. *Mathematics of Operations Research* 6:19–30.

Zabinsky, Z. B., and R. L. Smith. 1992. Pure adaptive search in global optimization. *Mathematical Programming* 53:323–338.

AUTHOR BIOGRAPHIES

BRUCE W. SCHMEISER is a Professor in the School of Industrial Engineering at Purdue University. He received his Ph.D. from the School of Industrial and Systems Engineering at the Georgia Institute of Technology in 1975. His primary research interests are the probabilistic and statistical aspects of digital-computer stochastic simulation. He is an active participant in the Winter Simulation Conference, including being Program Chair in 1983 and chairing the Board of Directors during 1988–1990.

JIN WANG is an Assistant Professor in the Department of Mathematics and Computer Science at Valdosta State University. He received his Ph.D. degree in Industrial Engineering from Purdue University. He holds a M.S. degree in Probability from Peking University, P.R. China, and a M.S. degree in Statistics from Purdue University. His research interests include stochastic modeling and optimization, applied probability and statistics, and stochastic simulation. He is a member of INFORMS, the INFORMS College on Simulation, IIE, Omega Rho, and Pi Mu Epsilon.