Mathematical Programming 69 (1995) 443–448

Pure adaptive search for finite global optimization ☆

Z.B. Zabinsky [a,*], G.R. Wood [b], M.A. Steel [c], W.P. Baritompa [c]

[a] *Industrial Engineering Program, FU-20, University of Washington, Seattle, WA 98195, USA*
[b] *Department of Mathematics and Computing, Central Queensland University, Rockhampton, Australia*
[c] *Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*

Received 13 December 1993; revised manuscript received 28 June 1994

Abstract

Pure Adaptive Search is a stochastic algorithm which has been analyzed for continuous global optimization. When a uniform distribution is used in PAS, it has been shown to have complexity which is linear in dimension. We define strong and weak variations of PAS in the setting of finite global optimization and prove analogous results. In particular, for the n-dimensional lattice $\{1,\ldots,k\}^n$, the expected number of iterations to find the global optimum is linear in n. Many discrete combinatorial optimization problems, although having intractably large domains, have quite small ranges. The strong version of PAS for all problems, and the weak version of PAS for a limited class of problems, has complexity the order of the size of the range.

Keywords: Global optimization; Discrete optimization; Algorithm complexity; Random search; Markov chains

1. Introduction

Pure adaptive search (PAS) is a random search method which has been defined and analyzed for continuous global optimization [5,6]. Pure adaptive search generates a sequence of feasible points according to a probability distribution, with the stipulation that the points always have strictly improving objective function values. It has been shown [6] that pure adaptive search has an encouraging feature: for continuous functions satisfying a Lipschitz condition the expected number of PAS iterations to reach convergence is proportional to the dimension.

In this paper we examine PAS for the global optimization problem in which the domain is a finite set of points. We begin by introducing two variations of finite PAS,

☆ The authors would like to thank the Department of Mathematics and Statistics at the University of Canterbury for support of this research.

* Corresponding author. E-mail: zelda@u.washington.edu.

SSDI 0025-5610(94)00072-7

weak and *strong*. For each variation, we derive an expression for the expected number of iterations to reach the global optimum, and a simple upper bound. In the special case where the domain is the n-dimensional lattice $\{1,\dots,k\}^n$, and the vertices are sampled according to a uniform distribution, the expected number of iterations to exact convergence is proportional to n, the "dimension."

As in the continuous case, strong PAS is inefficient to implement for general discrete functions. Weak PAS appears to be related closely to discrete optimization applications, for example, genetic reconstruction via the "Great Deluge" algorithm [2]. Our analysis of finite PAS suggests that such algorithms may have reasonable complexity.

2. Analysis of finite PAS

2.1. Terminology

In this paper we consider the following finite global optimization problem:

$$\begin{aligned}&\text{minimize } f(x)\\ &\text{subject to } x \in S,\end{aligned}$$

where $f(x)$ is a real-valued function on a finite set S.

The following algorithms for finite optimization are considered in this paper. *Pure random search* (PRS) [1] samples the domain at each iteration according to a fixed distribution. *Strong pure adaptive search* samples from that part of the domain that gives a strictly improving objective function value at each iteration. This is a translation of PAS from the continuous to the finite problem. By relaxing the strictness, we can define *weak pure adaptive search*. This algorithm samples from that part of the domain which gives an equal or improving objective function value.

Let $y_1 < y_2 < \cdots < y_K$ be the distinct objective function values. Notice that there may be more than K points in S. For $m = 0, 1, \dots$, let the random variable Y_m be the objective function value on the mth iteration of PRS. Note that $Y_0, Y_1, \dots$ are independent and identically distributed. Given a probability measure μ on S, we define a probability measure $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ on the range of f as follows. Let π_j be the probability that any iteration of pure random search attains a value of y_j. That is $\pi_j = P(Y_0 = y_j) = \mu(f^{-1}(y_j))$ for $j = 1, 2, \dots, K$. Throughout this paper p_j denotes $\sum_{i=1}^{j} \pi_i$ the probability that PRS attains a value of y_j or less.

We now describe the link between PRS and the two versions of finite PAS. Epoch $i > 0$ is said to be a *record* of the sequence $\{Y_m\}$ for strong PAS if $Y_i < \min\{Y_0, \dots, Y_{i-1}\}$ and for weak PAS if $Y_i \leqslant \min\{Y_0, \dots, Y_{i-1}\}$. Epoch $i = 0$ is always considered to be a record. The corresponding value Y_i is called a *record value*. Let the random variable W_m be the objective function value on the mth iteration of PAS, and let $R(m)$ be the epoch of the mth record of PRS. Then as in the continuous problem, PAS consists of the record values of pure random search. That is, W_m is stochastically equivalent to $Y_{R(m)}$. This is directly analogous to [6, Lemma 3.1] and is proved identically.

The stochastic process $\{W_m \mid m = 0, 1, \ldots\}$ for either weak or strong PAS can be modeled as a Markov chain with states $y_1, \ldots, y_K$, where state y_1 represents the global optimum. The initial probability distribution for W_0 is given by $\boldsymbol{\pi}$. In standard Markov chain terminology [4], y_1 is the absorbing state of this chain and all other states are transient. Finite PAS converges when the chain reaches the absorbing state. The expected number of iterations to convergence can be expressed in terms of the transition matrix of the Markov chain. The expected number of iterations to solve the problem used in this paper does not include a stopping rule, and thus indicates the average computational effort to sample the global optimum but not necessarily to confirm it. We present a simple direct probabilistic argument here.

2.2. General analysis

Theorem 1. *The expected number of iterations to solve the finite optimization problem is*

(i) $1 + \sum_{j=2}^{K} \pi_j / p_j$ *for strong PAS and*

(ii) $1 + \sum_{j=2}^{K} \pi_j / p_{j-1}$ *for weak PAS*

where $p_j = \sum_{i=1}^{j} \pi_i$.

Proof. Let the random variable X be the number of iterations required to solve the finite optimization problem. Then $X = 1 + X_2 + \cdots + X_K$, where X_j is the number of iterations spent in state y_j. Thus,

$$\begin{aligned} E(X) &= 1 + E(X_2) + \cdots + E(X_K) \\ &= 1 + E(X_2 \mid V_2)P(V_2) + \cdots + E(X_K \mid V_K)P(V_K), \end{aligned}$$

where V_j is the event that state y_j is visited. Now,

$$\begin{aligned} P(V_j) &= P(V_j \cap \{W_0 > y_j\}) + P(V_j \cap \{W_0 = y_j\}) \\ &= \sum_{i=j+1}^{K} P(V_j \cap B_{i,j} \cap \{W_0 > y_j\}) + P(V_j \cap \{W_0 = y_j\}), \end{aligned}$$

where $B_{i,j}$ is the event that state y_i is visited immediately before a state less than or equal to y_j is visited. Since $P(V_j \mid B_{i,j} \cap \{W_0 > y_j\}) = \pi_j / p_j$ and $P(W_0 = y_j) = \pi_j$, we have

$$\begin{aligned} P(V_j) &= \sum_{i=j+1}^{K} P(V_j \mid B_{i,j} \cap \{W_0 > y_j\}) P(B_{i,j} \mid W_0 > y_j) P(W_0 > y_j) \\ &\quad + P(V_j \mid W_0 = y_j) P(W_0 = y_j) \\ &= (1 - p_j) \frac{\pi_j}{p_j} \sum_{i=j+1}^{K} P(B_{i,j} \mid W_0 > y_j) + \pi_j \\ &= (1 - p_j) \frac{\pi_j}{p_j} + \pi_j = \frac{\pi_j}{p_j}. \end{aligned}$$

For strong PAS, $E(X_j \mid V_j) = 1$, and $E(X) = 1 + \sum_{j=2}^{K} \pi_j/p_j$ as required. For weak PAS, once state y_j is left, it is never re-entered. As with a geometric distribution, $E(X_j \mid V_j)$ then equals the inverse of the probability of leaving state y_j:

$$E(X_j \mid V_j) = (p_{j-1}/p_j)^{-1} = p_j/p_{j-1}.$$

This yields $E(X) = 1 + \sum_{j=2}^{K} \pi_j/p_{j-1}$ as required. □

Note that the *hazard rate* π_j/p_{j-1} appears in the formula for the expected time of weak PAS.

For completeness, the $K \times K$ transition matrix P, in standard form, having the (i,j)th element $P[W_m = y_j \mid W_{m-1} = y_i]$, for strong PAS consists of the first K rows of the matrix below. For weak PAS, it consists of the last K rows:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \pi_1/p_1 & 0 & 0 & \cdots & 0 & 0 \\ \pi_1/p_2 & \pi_2/p_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \pi_1/p_{K-1} & \pi_2/p_{K-1} & \pi_3/p_{K-1} & \cdots & \pi_{K-1}/p_{K-1} & 0 \\ \pi_1/p_K & \pi_2/p_K & \pi_3/p_K & \cdots & \pi_{K-1}/p_K & \pi_K/p_K \end{bmatrix}.$$

It follows that $P(W_m = y_i)$, the probability of objective function value y_i for weak or strong PAS on the mth iteration, is the ith entry of πP^m, with the appropriate P.

The exact expressions given in Theorem 1 are valid for an arbitrary sampling distribution reflected by π. We turn to obtaining bounds on these expressions.

3. Bounds on performance

3.1. General bounds

An upper bound on the expected number of strong PAS iterations to solve the finite optimization problem can be stated simply in terms of π_1, the probability of sampling the global optimum with pure random search.

Corollary 2. *The expected number of strong PAS iterations to solve the finite optimization problem is bounded above by* $1 + \log(1/\pi_1)$.

Proof. For $0 < x < 1$, $x < -\log(1-x)$, so for $j = 2, \ldots, K$,

$$\frac{\pi_j}{p_j} < -\log\left(1 - \frac{\pi_j}{p_j}\right) = \log\left(\frac{p_j}{p_{j-1}}\right).$$

Therefore from the theorem, the expected number of iterations is less than $1+\log(p_2/p_1) + \log(p_3/p_2) + \cdots + \log(p_K/p_{K-1}) = 1 + \log(1/\pi_1)$. □

For weak PAS the maximum hazard rate is involved in the bound. Note the term π_j/p_{j-1} appearing in Theorem 1(ii) can be written as $(1+\pi_j/p_{j-1})\pi_j/p_j$, so a similar argument as Corollary 2 gives the following corollary.

Corollary 3. *The expected number of weak PAS iterations to solve the finite optimization problem is bounded above by* $1+(M_{\text{hazard}}+1)\log(1/\pi_1)$ *where* $M_{\text{hazard}} = \max_{j=2,\dots,K}\pi_j/p_{j-1}$.

3.2. Special bounds

Many combinatorical optimization problems of interest have extremely large domains (e.g., 2^{100} points) but much smaller ranges (e.g., 100 values). In fact, most purely discrete problems which are NP-hard fall into this category. A good example is the MAXCLIQUE problem; given a graph G, find the largest number of vertices which induces a clique in the graph. In this case the range (the number of vertices in the graph) is small, but the domain is the set of all graphs on these vertices and is highly exponential in the range [3].

Clearly, strong PAS has complexity of order K, the size of the range, as it never requires more than K iterations, and thus is fast for these problems. This cannot be said for weak PAS (which is closer to practical algorithms). Weak PAS requires a large number of iterations when π_{i+1}/π_i is large for some i in the problem. However, an analogous result for weak PAS holds on problems where these ratios are bounded. The bounding factor r becomes the constant of proportionality.

Corollary 4. *If* $\pi_{i+1}/\pi_i \leqslant r$, *then weak PAS has complexity of order* K. *The expected number of weak PAS iterations is bounded above by* $1+(K-1)r$.

The following special case shows the upper limit of Corollary 4 is approached.

Corollary 5. *Given that* $\pi_i \propto r^i$ *for* $r>1$, *the expected number of iterations is bounded below by*

(i) $K(r-1)/r$ *for strong PAS and*

(ii) $1+(K-1)(r-1)$ *for weak PAS.*

Proof. For strong PAS, Theorem 1(i) gives the required value of $(r-1)/r\sum_{j=1}^{K} r^j/(r^j-1)$. This is at least $K(r-1)/r$. For weak PAS, Theorem 1(ii) gives the required value of $1+\sum_{j=2}^{K}\pi_j/p_{j-1} = 1+r\sum_{j=1}^{K-1}\pi_j/p_j$. This is at least $1+(K-1)(r-1)$. □

3.3. Uniform sampling

In order to compare the performance of pure adaptive search on a finite optimization problem with that on a continuous problem, we consider the special case in which the distribution on the objective function values is uniform (i.e., $\pi_j = 1/K$ for all j).

Corollary 6. *The expected number of iterations for finite global optimization, given a uniform distribution on the objective function values, is*

(i) $\sum_{j=1}^{K} 1/j$, *bounded above by* $1 + \log K$ *for strong PAS and*

(ii) $1 + \sum_{j=1}^{K-1} 1/j$, *bounded above by* $2 + \log(K - 1)$ *for weak PAS.*

To get the analogous linear complexity result, consider the vertices of an n-dimensional lattice, $\{1, \ldots, k\}^n$, with distinct objective function values at the vertices. Here $K = k^n$ and Corollary 6 (for either weak or strong PAS) gives a bound of $2 + \log k^n = 2 + n \log k$, an expression of order n.

4. Summary

For an n-dimensional lattice, the expected number of iterations for weak or strong PAS given a uniform distribution is of order n, consistent with the results for PAS on a continuous optimization problem. For strong PAS, the expected number of iterations on a finite optimization problem is bounded by a simple expression involving π_1, the probability of randomly sampling the global optimum. A similar bound for weak PAS is presented which involves the hazard rate function. This bound may be closer to that experienced by practical algorithms, and may inspire use of random search methods for finite optimization.

References

[1] S.H. Brooks, "A discussion of random methods for seeking maxima," *Operations Research* 6 (1958) 244–251.

[2] D. Dueck and T. Scheuer, "Threshold Accepting: a general purpose optimization algorithm appearing superior to simulated annealing," *Journal of Computational Physics* 90 (1990) 161–175.

[3] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness* (Freeman, San Francisco, CA, 1979).

[4] J.G. Kemeny and J.L. Snell, *Finite Markov Chains* (Springer, New York, 1976).

[5] N.R. Patel, R.L. Smith and Z.B. Zabinsky, "Pure adaptive search in Monte Carlo optimization," *Mathematical Programming* 43 (1988) 317-328.

[6] Z.B. Zabinsky and R.L. Smith, "Pure adaptive search in global optimization," *Mathematical Programming* 53 (1992) 323–338.