Pure Adaptive Search in Global Optimization by Z. B. Zabinsky and R. L. Smith

Presented by Michael Yee

November 3, 2003

This presentation is based on: Zabinsky, Zelda B., and Robert L. Smith. *Pure Adaptive Search in Global Optimization. Mathematical Programming* 55, 1992, pp. 323-338.

Outline

- Pure Random Search versus Pure Adaptive Search
- Relationship to Solis and Wetz (1981)
- Distribution of improvement in objective function value
- Performance bounds

Global Optimization Problem

- Problem (P):
  $$\min_{x \in S} f(x)$$
  where $x \in R^n$, S is convex, compact subset of R^n, and f continuous over S
- f satisfies *Lipschitz condition*, i.e., $|f(x) - f(y)| \leq k_f \|x - y\|, \ \forall x, y \in S$
- $x_* = \arg\min_{x \in S} f(x)$
- $y_* = f(x_*) = \min_{x \in S} f(x)$
- $y^* = \max_{x \in S} f(x)$

Pure Random Search (PRS)

- Generate sequence of independent, uniformly distributed points
  $$X_1, X_2, \ldots,$$
  in the feasible region S. Denote their associated objective function values by
  $$Y_1 = f(X_1), Y_2 = f(X_2), \ldots$$
- When stopping criterion met, best point generated so far is taken as approximation to true optimal solution

Pure Adaptive Search (PAS)

Step 0. Set $k = 0$, and $S_0 = S$

Step 1. Generate X_{k+1} uniformly distributed in S_k, and set $W_{k+1} = f(X_{k+1})$

Step 2. If stopping criterion met, STOP. Otherwise, set

$$S_{k+1} = \{x : x \in S \text{ and } f(x) < W_{k+1}\},$$

Increment k, Goto Step 1.

Solis and Wetz

- Give sufficient conditions for convergence of random global search methods
- Experimental support for linear relation between function evaluations and dimension
- PAS satisfies H1 since objective function values are increasing
- PAS satisfies H2 since the optimal solution is always in the restricted feasible region

Importance of Strict Improvement

- What if consecutive points were allowed to have *equal* objective function values?
- Let S be a unit hypersphere, with $f(x) = 1$ on S except for a depression on a hypersphere of radius ϵ, S_ϵ, where $f(x)$ drops to value 0 at the center of the ϵ-ball S_ϵ
- Then, $P(\text{random point is in } S_\epsilon) = \text{volume}(S_\epsilon)/\text{volume}(S) = \epsilon^n$
- Thus, PAS could have expected number of iterations that is exponential in dimension (if strict improvement were not enforced)

Some Notation

- Let $p(y) = P(Y_k \leq y)$, for $k = 1, 2, \ldots$ and $y_* \leq y \leq y^*$
- For PRS,
$$p(y) = v(S(y))/v(S),$$
where $S(y) = \{x : x \in S \text{ and } f(x) \leq y\}$ and $v(\cdot)$ is Lebesgue measure
- Note that for PAS,
$$P(W_{k+1} \leq y | W_k = z) = v(S(y))/v(S(z)) = p(y)/p(z),$$
for $k = 1, 2, \ldots$ and $y_* \leq y \leq z \leq y^*$

Connection Between PAS and PRS

Definition. Epoch i is said to be a *record* of the sequence $\{Y_k, k = 0, 1, 2, \ldots\}$ if $Y_i < \min(Y_0, Y_1, \ldots, Y_{i-1})$. The corresponding value Y_i is called a *record value*.

Lemma 1. *For the global optimization problem (P), the stochastic process* $\{W_k, k = 0, 1, 2, \ldots\} \sim \{Y_{R(k)}, k = 0, 1, 2, \ldots\}$, *where* $R(k)$ *is the kth record of the sequence* $\{Y_k, k = 0, 1, 2, \ldots\}$. *In particular,*

$$P(W_k \le y) = P(Y_{R(k)} \le y), \text{ for } k = 0, 1, 2, \ldots, \text{ and } y_* \le y \le y^*$$

Proof of Lemma 1

Proof. First, we show that the conditional distributions are equal.

$$\begin{aligned} P(Y_{R(k+1)} \le y | Y_{R(k)} = x) &= P(Y_{R(k)+1} \le y | Y_{R(k)} = x) \\ &\quad + P(Y_{R(k)+2} \le y, Y_{R(k)+1} \ge x | Y_{R(k)} = x) + \cdots \\ &= P(Y_{R(k)+1} \le y) \\ &\quad + P(Y_{R(k)+2} \le y) P(Y_{R(k)+1} \ge x) + \cdots \\ &= P(Y_1 \le y) \textstyle\sum_{i=0}^{\infty} P(Y_1 \ge x)^i \\ &= \tfrac{P(Y_1 \le y)}{1 - P(Y_1 \ge x)} \\ &= v(S(y))/v(S(x)) \\ &= P(W_{k+1} \le y | W_k = x). \end{aligned}$$

Next, we use induction to show that the unconditional distributions are equal. By definition, $R(0) = 0$ and $Y_0 = W_0 = y^*$, thus $Y_{R(0)} = W_0$.

For the base case $k = 1$,

$$\begin{aligned} P(Y_{R(1)} \le y) &= P(Y_{R(1)} \le y | Y_0 = y^*) \\ &= P(W_1 \le y | W_0 = y^*) \\ &= P(W_1 \le y), \quad \text{for all } y_* \le y \le y^* \end{aligned}$$

Thus, $Y_{R(1)} \sim W_1$.

For $k > 1$, suppose that $Y_{R(i)} \sim W_i$ for $i = 1, 2, \ldots, k$. Then,

$$\begin{aligned} P(Y_{R(k+1)} \le y) &= E[P(Y_{R(k+1)} \le y | Y_{R(k)})] \\ &= \textstyle\int_0^x P(Y_{R(k+1)} \le y | Y_{R(k)} = x) \, \mathrm{d}F_{Y_{R(k)}}(x) \\ &= \textstyle\int_0^x P(W_{k+1} \le y | W_k = x) \, \mathrm{d}F_{W_k}(x) \\ &= E[P(W_{k+1} \le y | W_k)] \\ &= P(W_k \le y), \quad \text{for all } y_* \le y \le y^* \end{aligned}$$

Thus, $Y_{R(k+1)} \sim W_{k+1}$.

Finally, since the two sequences are equal in conditional and marginal distribution, they are equal in joint distribution. $\square$

Linear versus Exponential

Theorem 1. *Let k and $R(k)$ be respectively the number of PAS and PRS iterations needed to attain an objective function value of y or better, for $y_* \le y \le y^*$. Then*

$$R(k) = e^{k+o(k)}, \text{ with probability } 1,$$

where $\lim_{k\to\infty} o(k)/k = 0$, with probability 1.

Proof. Use general fact about records that $\lim_{k\to\infty} \frac{\ln R(k)}{k} = 1$, with probability 1... □

Relative Improvement

Definition. Let $Z_k = (y^* - Y_k)/(Y_k - y_*)$ be the *relative improvement* obtained by the kth iteration of PRS.

Then, the cumulative distribution function F of Z_k is given by

$$\begin{aligned} F(z) &= P(Z_k \le z) \\ &= P(Y_k \ge (y^* + zy_*)/(1+z)) \\ &= \begin{cases} 0 & \text{if } z < 0, \\ 1 - p((y^* + zy_*)/(1+z)) & \text{if } 0 \le z < \infty. \end{cases} \end{aligned}$$

Note also that the random variables Z_k are iid and nonnegative.

Relative Improvement Process

Lemma 2. *Let $Z_1, Z_2, \ldots$ denote a sequence of iid nonnegative continuous random variables with density f and cdf F. Let $M(z)$ denote the number of record values (in the max sense) of $\{Z_i, i = 1, 2, \ldots\}$ less than or equal to z.*

Then $\{M(z), z \ge 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(z) = f(z)/(1 - F(z))$ and mean value function $m(z) = \int_0^z \lambda(z)\ ds$.

Theorem 2. *Let $N(z)$ be the number of PAS iterations achieving a relative improvement at most z for $z \ge 0$. Then $\{N(z), z \ge 0\}$ is a nonhomogeneous Poisson process with mean value function*

$$m(z) = \ln(1/p((y^* + zy_*)/(1+z))), \text{ for } 0 \le z < \infty.$$

Distribution of Objective Function Values

Theorem 3. $P(W_k \le y) = \sum_{i=0}^{k-1} \frac{p(y)(\ln(1/p(y)))^i}{i!}$

Proof. The events $\{W_k < y\}$ and $\{N((y^* - y)/(y - y_*)) < k\}$ are equivalent, so

$$P(W_k \le y) = P(W_k < y) = P(N((y^* - y)/(y - y_*)) < k),$$

and by previous theorem $N(z)$ is a Poisson random variable with mean

$$m(z) = \ln(1/p((y^* + zy_*)/(1+z))),$$

etc. □

Performance Bounds

Let $N^*(y)$ be the number of iterations require by PAS to achieve a value of y or better. Then

$$N^*(y) = N((y^* - y)/(y - y_*)) + 1$$

Corollary 1. *The cumulative distribution of $N^*(y)$ is given by*

$$P(N^*(y) \le k) = \sum_{i=0}^{k-1} \frac{p(y)(\ln(1/p(y)))^i}{i!},$$

with

$$E[N^*(y)] = 1 + \ln(1/p(y)), \quad Var(N^*(y)) = \ln(1/p(y))$$

Bounds for Lipschitz Functions

Lemma 3. *For global optimization problem (P) over a convex feasible region S in n dimensions with diameter $d_S = \max\{\|w - v\|, w, v \in S\}$ and Lipschitz constant k_f,*

$$p(y) \ge ((y - y_*)/k_f d_S)^n, \text{ for } y_* \le y \le y^*.$$

Theorem 4. *For any global optimization problem (P) over a convex feasible region in n dimensions with diameter at most d and Lipschitz constant at most k,*

$$E[N^*(y)] \le 1 + [\ln(kd/(y - y_*))]n$$

and

$$Var(N^*(y)) \le [\ln(kd/(y - y_*))]n$$

for $y_* \le y \le y^*$.

Conclusions

- Complexity of PRS is exponentially worse than that of PAS
- General performance bounds using theory from stochastic processes
- Specific performance bounds for Lipschitz functions : linear in dimension!
- But is this too good to be true?!