Pure Adaptive Search in Global Optimization

Zelda B. Zabinsky
Industrial Engineering Program, FU-20
University of Washington
Seattle,Washington 98195

Robert L. Smith
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117

Technical Report 89-22

June 1989

PURE ADAPTIVE SEARCH IN GLOBAL OPTIMIZATION

Zelda B. Zabinsky
Industrial Engineering Program, FU-20
University of Washington
Seattle, Washington 98195

Robert L. Smith
Department of Industrial & Operations Engineering
The University of Michigan
Ann Arbor, Michigan 48109

June 20, 1989

Abstract

Pure adaptive search iteratively constructs a sequence of interior points uniformly distributed within the corresponding sequence of nested improving regions of the feasible space. That is, at any iteration, the next point in the sequence is uniformly distributed over the region of feasible space containing all points that are equal or superior in value to the previous points in the sequence. The complexity of this algorithm is measured by the expected number of iterations required to achieve a given accuracy of solution. We show that for global mathematical programs satisfying the Lipschitz condition, its complexity increases at most *linearly* in the dimension of the problem.

Key Words: Random search, Monte Carlo optimization, global optimization.

1 Introduction

In this paper, we provide a theoretical analysis of pure adaptive search for general global optimization (see [18,19,20] for a survey of the field). The algorithm proceeds by generating a sequence of points uniformly distributed in a sequence of nested regions of the feasible space. At any iteration, the next point in the sequence is uniformly distributed over the region of feasible space containing all points that are equal or better in value to the previous points in the sequence. It has been shown [16] that for convex programs, the number of iterations required to achieve a given accuracy of solution increases at most linearly in the dimension of the problem. In this paper, we extend this linear complexity result for non-convex or *global* optimization problems satisfying the Lipschitz condition. We do this by modeling the sequence of values for points obtained by pure adaptive search as a nonhomogeneous Poisson process. A bound on the rate of this process is used to obtain an upper bound on the expected number of iterations which is a linear function of the dimension of the problem.

Although at this time there is no known efficient implementation of the pure adaptive search algorithm, the theoretical result of linear time complexity for global optimization is interesting in itself. Pure adaptive search is not practical because the principal computational effort of the algorithm lies in generating a point uniformly distributed in the improving region. At present, this is a challenging problem with no satisfactory solution. However, the linearity result suggests there is hope for an efficient random search method for global optimization. In fact, several random search algorithms have reported linearity in dimension [23,24,26], although only for convex programs. Pure adaptive search is analogous to a randomized method of centers [12] which, itself is not very practical, but has been the precursor to a class of extremely efficient projective and affine scaling methods [8,10,13,17] for linear programming. Our hope is that pure adaptive search can similarly inspire better random search methods for global programming.

2 Pure Adaptive Search

Consider the following global mathematical program,

$$(P) \qquad \min_{x \in S} f(x)$$

where x is an n-dimensional vector, S is a convex, compact subset of $\mathbf{R}^n$, and f is a real-valued continuous function defined over S. We will assume that f satisfies the *Lipschitz condition* with Lipschitz constant k_f, i.e.,

$$|f(x) - f(y)| \leq k_f \|x - y\|$$

for all x and $y \in S$, where $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^n$. To ensure a proper sampling distribution, we further assume that f has no flat spots, i.e., that $f'(x) \neq 0$ for almost all $x \in S$. Let the optimal solution to (P) be denoted by (x_*, y_*) where

$$x_* = \arg\min_{x \in S} f(x)$$

and

$$y_* = f(x_*) = \min_{x \in S} f(x).$$

It will also be convenient to define

$$y^* = \max_{x \in S} f(x).$$

Note that we do not require that a unique minimum exists at x_*. If there are multiple minima, let x_* be a arbitrary fixed minimum.

The pure adaptive search (PAS) procedure for solving (P) begins by generating a point X_1 uniformly distributed within the feasible region S. The associated objective function value is labeled $W_1 = f(X_1)$. The next point is generated from a uniform distribution over the region formed by the intersection of the feasible region with the level set of points with objective function values equal to or less than W_1. The procedure proceeds iteratively in this fashion until a stopping criterion is satisfied.

More formally,

Pure Adaptive Search (PAS)

Step 0. Set $k = 0$, $S_0 = S$, and $Y_0 \geq y^*$.

Step 1. Generate X_{k+1} uniformly distributed in $S_{k+1} = \{x : x \in S \text{ and } f(x) \leq W_k\}$.

Step 2. Set $W_{k+1} = f(X_{k+1})$. If a stopping criterion is met, stop. Otherwise set $k = k + 1$ and return to Step 1.

The difficulties of implementing pure adaptive search are discussed in Patel, Smith, and Zabinsky [16]. In this paper we are however interested in its computational complexity as a model for other possible algorithms that share some of the same features as pure adaptive search. For example, if another algorithm can generate random points (uniform or not) with associated objective function values stochastically less than those of uniformly distributed points, then the performance of the new algorithm will be bounded by the performance of pure adaptive search. Thus a theoretical result of linear time complexity for PAS supports research in other random search algorithms for global optimization.

3 A Comparison with Pure Random Search

It is instructive to compare pure adaptive search with pure random search [4,6,7]. *Pure random search (PRS)* generates a sequence of independent, uniformly distributed points in the feasible region. When a stopping criterion is met, the best point of the sequence generated thus far is used as an approximation to the optimal solution. For the global optimization problem (P), consider the stochastic process $\{W_k,\ k = 0,1,2,\ldots\}$ of objective function values generated by PAS, and let $\{Y_k,\ k = 0,1,2,\ldots\}$ be the corresponding sequence of values for the points generated by PRS so that $Y_k = f(\mathsf{X}_k), k = 1,2,\ldots$ where $\mathsf{X}_1, \mathsf{X}_2, \ldots$ are independent and uniformly distributed over S. For convenience, define $W_0 = Y_0 = y^*$.

Let $p(y)$ be the probability that a point in the sequence generated by PRS has an objective function value less than or equal to y, that is,

$$p(y) = P(Y_k \leq y)$$

for $k = 1,2,\ldots$ and $y_* \leq y \leq y^*$. Note that $p(y)$ is the same for all k since PRS generates identically distributed points. Also, due to the uniform distribution employed in PRS, we have

$$p(y) = \nu(S(y))/\nu(S)$$

where $S(y) = \{x : x \in S \text{ and } f(x) \leq y\}$ and $\nu(\cdot)$ denotes Lebesgue measure. Incidentally, although a uniform distribution is used in both the PRS and PAS algorithms, much of the subsequent analysis holds for nonuniform absolutely continuous distributions. We have for pure adaptive search that

$$P(W_{k+1} \leq y | W_k = z) = \nu(S(y))/\nu(S(z)) = p(y)/p(z)$$

for any k, where $y_* \leq y \leq z \leq y^*$.

We now establish a fundamental relationship between the iterates of PAS and PRS. The lemma below states that the record values of pure random search are equal in distribution to the values generated by pure adaptive search.

Definition A *record* is said to occur at epoch i for 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 3.1 *For the global optimization problem (P), the stochastic process* $\{W_k,\ k = 0,1,2,\ldots\}$ *is equal in distribution to the process* $\left\{Y_{R(k)},\ k = 0,1,2,\ldots\right\}$ *where* $R(k)$ *is the* k^{th} *record value of the sequence* $\{Y_k,\ k = 0,1,2,\ldots\}$, *i.e.*

$$\{W_k,\ k = 0,1,2,\ldots\} \sim \left\{Y_{R(k)},\ k = 0,1,2,\ldots\right\}.$$

In particular,

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

Proof: See Appendix. ∎

An intuitive understanding of the previous lemma follows from the property that a point X uniformly distributed over a region S is conditionally uniform over the region $S' \subseteq S$ when given that X is in S'. It follows that a simple acceptance-rejection approach to generating $\{W_k\}$ would be to generate $\{Y_k\}$ and select the record values $\left\{Y_{R(k)}\right\}$.

Theorem 3.2 *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 $y_* \leq y \leq y^*$. Then*

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

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

Proof: We have by definition that

$$W_{k-1} > y \geq W_k.$$

However by the previous lemma, this holds for a given k with the same probability that

$$Y_{R(k-1)} > y \geq Y_{R(k)}$$

holds. From [9, p. 298], the record values $R(k)$ of a sequence of continuous independent and identically distributed random variables satisfy

$$\lim_{k\to\infty} \frac{\ln R(k)}{k} = 1 \qquad \text{with probability 1.}$$

This implies that

$$\ln R(k) = k + o(k) \qquad \text{w.p. 1}$$

and thus,

$$R(k) = e^{k+o(k)} \qquad \text{w.p. 1.} \quad ∎$$

The result in Theorem 3.2 states that the number of PRS iterations needed to reach the k^{th} minimum, $R(k)$, is exponentially growing in the number of iterations of PAS, k, needed to reach an equivalent minimum. Thus the complexity of PRS is exponentially greater than that of PAS. Of course, each iteration of PAS may be more difficult than an iteration of PRS since the search region changes with each iteration. It is nonetheless interesting that the simple device of forcing monotone value improvement on PRS achieves an exponential improvement in iterations required.

4 The Distribution of Improvement

We turn now to establishing the distribution of the values $\{W_k,\ k = 0, 1, 2, \ldots\}$ obtained by pure adaptive search.

Lemma 4.1 *Let $Z_1, Z_2, \ldots$ denote a sequence of independent and identically distributed nonnegative continuous random variables, whose hazard rate function is given by $\lambda(z)$, $\lambda(z) = f(z)/(1-F(z))$ where f and F are respectively the density and cumulative distribution function of Z. Let $M(z)$ denote the number of record values (maximum) of $\{Z_i,\ i = 1, 2, \ldots\}$ less than or equal to z.*

Then, $\{M(z), z \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(z)$ and mean value function $m(z) = \int_0^z \lambda(s)ds$.

Proof: See [22, p. 47]. ∎

Applying the lemma above, let Z_k be the *relative improvement* obtained on the k^{th} iteration of PRS where $Z_k = (y^* - Y_k)/(Y_k - y_*)$. Then, the assumptions of Lemma 4.1 are satisfied since $Z_0 = 0$, and $\{Z_k,\ k = 1, 2, \ldots\}$ are independent identically distributed nonnegative continuous random variables. The cumulative distribution function F of Z_k, $k = 1, 2, \ldots$ can be written in terms of $p(y)$, as follows. For $z \geq 0$,

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

Since $M(z)$ counts the number of records of $\{Z_k,\ k = 0, 1, 2, \ldots\}$ with values less than or equal to z, $M(z)$ by Lemma 3.1 is equal in distribution to the number $N(z)$ of PAS iterations achieving a relative improvement of z or less.

Theorem 4.2 *Let $N(z)$ equal the number of PAS iterations achieving a relative improvement at most z for $z \geq 0$. Then $\{N(z), z \geq 0\}$ is a nonhomogeneous Poisson process with mean value function*

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

for $0 \leq z < \infty$.

Proof: By definition,

$$m(z) = \int_0^z \lambda(s)ds$$

where $\lambda(s) = f(s)/(1 - F(s))$. Making the substitution $t = 1 - F(s)$ yields

$$\begin{aligned} m(z) &= \int_1^{1-F(z)} -(1/t)dt \\ &= -\ln t|_1^{1-F(z)} \\ &= -\ln(1 - F(z)) \end{aligned}$$

for $z \geq 0$. Now,

$$1 - F(z) = p((y^* + zy_*)/(1 + z))$$

and hence,

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

It is now an easy matter to obtain the distribution of the objective function values obtained through pure adaptive search. In particular, since

$$W_k < y \text{ if and only if } N((y^* - y)/(y - y_*)) < k$$

and W_k is a continuous random variable,

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

where by Theorem 4.2, $N(z)$ is a Poisson distributed random variable with mean $m(z) = \ln(1/p((y^* + zy_*)/(1 + z)))$. We therefore have

Theorem 4.3

$$P(W_k \leq y) = \sum_{i=0}^{k-1} \frac{p(y)(\ln(1/p(y)))^i}{i!} \qquad \textit{for } k = 1, 2, \ldots \textit{ and } y_* \leq y \leq y^*.$$

Proof: See Appendix. $\blacksquare$

There are several problem classes in the literature where the asymptotic distribution for large sample sizes of PRS and PAS have been obtained (see for example [1,5,11,15,16]). The result in Theorem 4.3 is particularly striking in that it provides the exact distribution of values generated by PAS for all sample sizes and all global optimization problems.

5 Performance Bounds

A simple measure of the performance of pure adaptive search is the number of iterations $N^*(y)$ required to achieve a value of y or better. Since an objective function value of y corresponds to a relative improvement of $z = (y^* - y)/(y - y_*)$, we have that

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

The distribution of $N^*(y)$ then follows from Theorem 4.2.

Corollary 5.1 *The cumulative distribution function of $N^*(y)$, the number of iterations of PAS needed to achieve a value of y or better, is given by*

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

for $k = 1, 2, \ldots$ and $y_* \leq y \leq y^*$.
The expected value of $N^*(y)$ is given by

$$E(N^*(y)) = 1 + \ln(1/p(y))$$

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

As seen in Theorem 4.2 and Corollary 5.1, performance measures of PAS depend on the function $p(y)$ for $y_* \leq y \leq y^*$, where $p(y)$ is the probability of obtaining an objective function value between y and y_* when selecting a feasible point at random according to a uniform distribution. We now derive a bound on $p(y)$ for the class of global optimization problems with objective functions that satisfy the Lipschitz condition over a convex feasible region. The bound is a function of the dimension, n of the problem; the Lipschitz constant, k_f of the objective function; and the maximum diameter, d_S of the feasible region, where $d_S = \max\{\|w - v\|, w, v \in S\}$.

Lemma 5.2 *For the global optimization problem (P) over a convex feasible region S in n dimensions with diameter d_S and Lipschitz constant k_f for objective function f,*

$$p(y) \geq \left((y - y_*)/k_f d_S\right)^n \qquad \text{for } y_* \leq y \leq y^*.$$

Proof: See Appendix. ∎

From the above lemma together with Corollary 5.1, we get the main result of this paper.

Theorem 5.3 *For all global optimization problems (P) over a convex feasible region in n dimensions with diameter at most d, and with Lipschitz constant at most k,*

$$E(N^*(y)) \leq 1 + [\ln(kd/(y - y_*))]n \qquad for\ y_* \leq y \leq y^*.$$

Proof: This follows immediately from Lemma 5.2, where

$$p(y) \geq ((y - y_*)/kd)^n$$

implies that

$$1/p(y) \leq (kd/(y - y_*))^n$$

and from Corollary 5.1,

$$\begin{aligned} E(N^*(y)) &= 1 + \ln(1/p(y)) \\ &\leq 1 + [\ln(kd/(y - y_*))]\, n. \quad \blacksquare \end{aligned}$$

From the above theorem, we conclude that the expected number of PAS iterations grows linearly in dimension for a class of problems with finite Lipschitz constant k and feasible region diameter d. This is in dramatic contrast to PRS where from Theorem 3.2 we know that the expected number of iterations will be an exponential function of dimension n. The logarithmic term, $kd/(y - y_*)$ can be viewed as a bound on the "length" of the graph of f expressed in units of the specified error from the optimal. Clearly, an exponential increase in the Lipschitz constant or the diameter gives rise to a linear increase in the correponding number of iterations of PAS required to achieve the same value error.

Although several researchers have empirically reported linear behavior in dimension for a variety of other random search algorithms including Schumer and Steiglitz [24], Schrack and Borowski [23], and Solis and Wets [26], PAS is a difficult algorithm to implement. The principal reason is that there is no known efficient procedure for generating a point uniformly distributed in a general region. Although the problem of efficiently generating *many* points uniformly distributed within a single bounded region has met with some success [2,3,21,25], the problem of efficiently generating a *single* point uniformly distributed in each of many bounded regions is still unresolved.

An alternative is to design an improving algorithm that generates points that in value stochastically dominate the uniform distribution. That is, if an algorithm generates random points (uniform or not) with associated objective function values stochastically better than those of uniformly distributed points, then the same linear bound on performance will apply. Natural candidates include interior point methods that have displayed similar dimensional linearity for linear programming problems, such as Karmarkar's projective scaling algorithm [13] and its affine scaling variants [8,10,17]. Several of these algorithms [17] are similar

in spirit to the method of centers due to Huard [12]. PAS can in fact be viewed as a randomized method of centers. Just as the method of centers has given rise to a class of extremely practical interior point methods for linear programming, perhaps PAS can be similarly employed to inspire a class of practical methods for global programming.

References

[1] F. Archetti, B. Betro and S. Steffe, "A Theoretical Framewok for Global Optimization Via Random Sampling," Working Paper, (Quaderno dei gruppi di recerca matematica del C.N.R., University of Pisa, 1975)

[2] H.C.P. Berbee, C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith, and J. Telgen, "Hit-and-Run Algorithms for the Identificaion of Nonredundant Linear Inequalities," *Mathematical Programming 37* (1987) 184-207.

[3] A. Boneh, "A Probabilistic Algorithm for Identifying Redundancy by a Random Feasible Point Generator (RFPG)," in: M.H. Karwan, Bl Lotfi, J. Telgen and S. Zionts, eds., *Redundancy in Mathematical Programming* (Springer-Verlag, Berlin, 1983)

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

[5] D.J. Clough, "An Asymptotic Extreme-Value Sampling Theory for Estimation of a Global Maximum," *CORS Journal 7* (1969) 102-115.

[6] L.C.W. Dixon and G.P. Szegö, eds., *Towards Global Optimization* (North-Holland, Amsterdam, 1975).

[7] L.C.W. Dixon and G.P. Szegö, eds., *Towards Global Optimization 2* (North-Holland, Amsterdam, 1978).

[8] R.M. Freund, "Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function," MIT working paper OR-182-88 (Massachusetts Institute Of Technology, Cambridge, 1988).

[9] J. Galambos, *The Asymptotic Theory Of Extreme Order Statistics* (John Wiley and Sons, New York, 1978).

[10] C.C. Gonzaga, "Polynomial affine algorithms for linear programming," Report ES-139/88 (Federal University of Rio de Janeiro, Rio de Janeiro, 1988).

[11] L. De Haan, "Estimation of the Minimum of a Function Using Order Statistics," *Journal of the American Statistical Association 76* (1981) 467-469.

[12] P. Huard, "Resolution of mathematical programming with non-linear constraints by the method of centers," in J. Abadie, ed., *Non-Linear Programming* (North-Holland, Amsterdam, 1967) pp. 207-219.

[13] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica 4* (1984) 373-395.

[14] M.G. Kendall, *A Course In The Geometry Of n-Dimensions* (Hafner Publishing, New York, 1961).

[15] N.R. Patel and R.L. Smith, "The Asymptotic Extreme Value Distribution of the Sample Minimum of a Concave Function under Linear Constraints", *Operations Research 31* (1983) 789-794.

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

[17] J. Renegar, "A polynomial-time algorithm, based on Newton's method, for linear programming," *Mathematical Programming 40* (1988) 59-93.

[18] A.H.G. Rinnooy Kan and G.T. Timmer, "Stochastic methods for global optimization," *American Journal of Mathematical and Management Sciences 4* (1984) 7-40.

[19] A.H.G. Rinnooy Kan and G.T. Timmer, "Stochastic global optimization methods part I: clustering methods," *Mathematical Programming 39* (1987) 27-56.

[20] A.H.G. Rinnooy Kan and G.T. Timmer, "Stochastic global optimization methods part II: multi-level methods," *Mathematical Programming 39* (1987) 57-78.

[21] R.Y. Rubinstein, "Generating Random Vectors Uniformly Distributed Inside and on the Surface of Different Regions", *European Journal of Operations Research 10* (1982) 205-209.

[22] S.M. Ross, *Stochastic Processes* (John Wiley and Sons, New York, 1983).

[23] G. Schrack and N. Borowski, "An experimental comparison of three random searches," in: F. Lootsma, ed., *Numerical Methods For Nonlinear Optimization* (Academic Press, London, 1972) pp. 137-147.

[24] M.A. Schumer and K. Steiglitz, "Adaptive step size random search," *IEEE Transactions On Automatic Control AC-13* (1968) 270-276.

[25] R.L. Smith, "Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions," *Operations Research 32* (1984) 1296-1308.

[26] F.J. Solis and R.J.-B. Wets, "Minimization by random search techniques," *Mathematics Of Operations Research 6* (1981) 19-30.

Appendix

Lemma 3.1 *For the global optimization problem (P), the stochastic process* $\{W_k,\ k = 0,1,2,\ldots\}$ *is equal in distribution to the process* $\left\{Y_{R(k)},\ k = 0,1,2,\ldots\right\}$ *where* $R(k)$ *is the* k^{th} *record value of the sequence* $\{Y_k,\ k = 0,1,2,\ldots\}$, *i.e.*

$$\{W_k,\ k = 0,1,2,\ldots\} \sim \left\{Y_{R(k)},\ k = 0,1,2,\ldots\right\}.$$

In particular,

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

Proof: First, we show that the conditional distributions are equal. Let k be any iteration, $k = 1,2,\ldots$, and x, y be such that $y_* \leq y \leq x \leq y^*$. Now,

$$\begin{aligned} P\left(Y_{R(k+1)} \leq y | Y_{R(k)} = x\right) &= P\left(Y_{R(k)+1} \leq y | Y_{R(k)} = x\right) \\ &\quad + P\left(Y_{R(k)+2} \leq y, Y_{R(k)+1} \geq x | Y_{R(k)} = x\right) + \cdots \end{aligned}$$

and since the Y_k are independent and $R(k)$ is a stopping time for all k,

$$= P\left(Y_{R(k)+1} \leq y\right) + P\left(Y_{R(k)+2} \leq y\right) \cdot P\left(Y_{R(k)+1} \geq x\right) + \cdots$$

and since the Y_k are identically distributed, i.e., $Y_k \sim Y_1$ for $k \geq 1$,

$$\begin{aligned} &= P\left(Y_1 \leq y\right) \sum_{i=0}^{\infty} P\left(Y_1 \geq x\right)^i \\ &= \frac{P\left(Y_1 \leq y\right)}{1 - P\left(Y_1 \geq x\right)} \\ &= \nu(S(y))/\nu(S(x)), \end{aligned}$$

which is the conditional distribution of improvement for pure adaptive search,

$$= P\left(W_{k+1} \leq y | W_k = x\right).$$

We have

$$P\left(Y_{R(k+1)} \leq y | Y_{R(k)} = x\right) = P\left(W_{k+1} \leq y | W_k = x\right).$$

Next, induction is used to show that the unconditional distributions are equal. By definition, $Y_0 = W_0 = y^*$. Thus,

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

and hence,

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

Let k be an integer greater than 1, and suppose $Y_{R(i)} \sim W_i$ for $i = 1, 2, \ldots, k$. Then,

$$\begin{aligned} P\left(Y_{R(k+1)} \leq y\right) &= E\left[P\left(Y_{R(k+1)} \leq y | Y_{R(k)}\right)\right] \\ &= \int_0^x P\left(Y_{R(k+1)} \leq y | Y_{R(k)} = x\right) dF_{Y_{R(k)}}(x) \end{aligned}$$

and using the equality of the conditional distributions and the induction hypothesis

$$\begin{aligned} &= \int_0^x P\left(W_{k+1} \leq y | W_k = x\right) dF_{W_k}(x) \\ &= E\left[P\left(W_{k+1} \leq y | W_k\right)\right] \\ &= P\left(W_{k+1} \leq y\right) \qquad \textit{for all } y_* \leq y \leq y^*. \end{aligned}$$

Therefore,

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

and by induction the two sequences are equal in marginal distribution. Finally, by the equality of conditional and marginal distributions, the two sequences are equal in joint distribution. ∎

Theorem 4.3

$$P\left(W_k \leq y\right) = \sum_{i=0}^{k-1} \frac{p(y)\left(\ln\left(1/p(y)\right)\right)^i}{i!} \qquad \textit{for } k = 1, 2, \ldots \textit{ and } y_* \leq y \leq y^*.$$

Proof: As noted in the text,

$$\begin{aligned} P\left(W_k \leq y\right) &= P\left(N((y^* - y)/(y - y_*)) < k\right) \\ &= \sum_{i=0}^{k-1} P\left(N((y^* - y)/(y - y_*)) = i\right) \end{aligned}$$

and because $N(z)$ has a Poisson distribution with mean $m(z)$ (Theorem 4.2)

$$= \sum_{i=0}^{k-1} \mathrm{e}^{-m((y^*-y)/(y-y_*))} m((y^* - y)/(y - y_*))^i / i!$$

and since algebraically $m((y^* - y)/(y - y_*)) = \ln\left(1/p(y)\right)$ for $y_* \leq y \leq y^*$, we have

$$\begin{aligned} &= \sum_{i=0}^{k-1} \mathrm{e}^{-\ln\left(1/p(y)\right)} \left(\ln\left(1/p(y)\right)\right)^i / i! \\ &= \sum_{i=0}^{k-1} p(y)\left(\ln\left(1/p(y)\right)\right)^i / i!. \quad ∎ \end{aligned}$$

Lemma 5.2 *For the global optimization problem (P) over a convex feasible region S in n dimensions with diameter d and Lipschitz constant k,*

$$p(y) \geq \left((y - y_*)/kd\right)^n \qquad for\ y_* \leq y \leq y^*.$$

Proof: To obtain a lower bound on $p(y)$, we construct two intermediate bounds. Let

$$\begin{array}{lcll}
r & = & (y^* - y_*)/k & \\
B_r & = & n\text{-dimensional hypersphere centered at } x_* \text{ with radius } r & \\
B_d & = & n\text{-dimensional hypersphere centered at } x_* \text{ with radius } d & \\
g(x) & = & \begin{cases} k\|x - x_*\| + y_* & \text{if } x \in B_r \\ y^* & \text{otherwise} \end{cases} & \\
 & = & \begin{cases} \inf\left[y : (x,y) \in \text{convex hull of } (x_*, y_*) \text{ and } (B_r, y^*)\right] & \text{if } x \in B_r \\ y^* & \text{otherwise} \end{cases} & \\
G(y) & = & \{x : x \in S \text{ and } g(x) \leq y\} & \\
S_r & = & B_r \bigcap S & \\
h(x) & = & \begin{cases} \inf\left[y : (x,y) \in \text{convex hull of } (x_*, y_*) \text{ and } (S_r, y^*)\right] & \text{if } x \in S_r \\ y^* & \text{otherwise} \end{cases} & \\
H(y) & = & \{x : x \in S \text{ and } h(x) \leq y\}. &
\end{array}$$

For a geometrical interpretation of these, refer to figure 1. Notice that two expressions are given for $g(x)$. These expressions are equivalent in defining $g(x)$, which can be easily verified by comparing the ratio $\|x - x_*\|/r$ to the ratio $y/(y^* - y_*)$. The sets $G(y)$ and $H(y)$ are the level sets of the functions $g(x)$ and $h(x)$ respectively.

We first show that the level sets are nested,

$$H(y) \subseteq G(y) \subseteq S(y)$$

for any $y_* \leq y \leq y^*$. We begin by considering any $x \in H(y)$ for any $y_* \leq y \leq y^*$. By definition of $H(y)$, $x \in S$ and $h(x) \leq y$. There are two cases, either $x \in S_r$ or $x \notin S_r$.

If $x \in S_r$, then

$$h(x) = \inf\left[y : (x,y) \in \text{convex hull of } (x_*, y_*) \text{ and } (S_r, y^*)\right].$$

Since $S_r = B_r \bigcap S$, $x \in B_r$, so

$$g(x) = \inf\left[y : (x,y) \in \text{convex hull of } (x_*, y_*) \text{ and } (B_r, y^*)\right].$$

Also, $S_r \subseteq B_r$ implies that $h(x) \geq g(x)$, thus $x \in G(y)$. Now, because $f(x)$ satisfies the Lipschitz condition, we have

$$|f(x) - f(x_*)| \leq k\|x - x_*\|$$

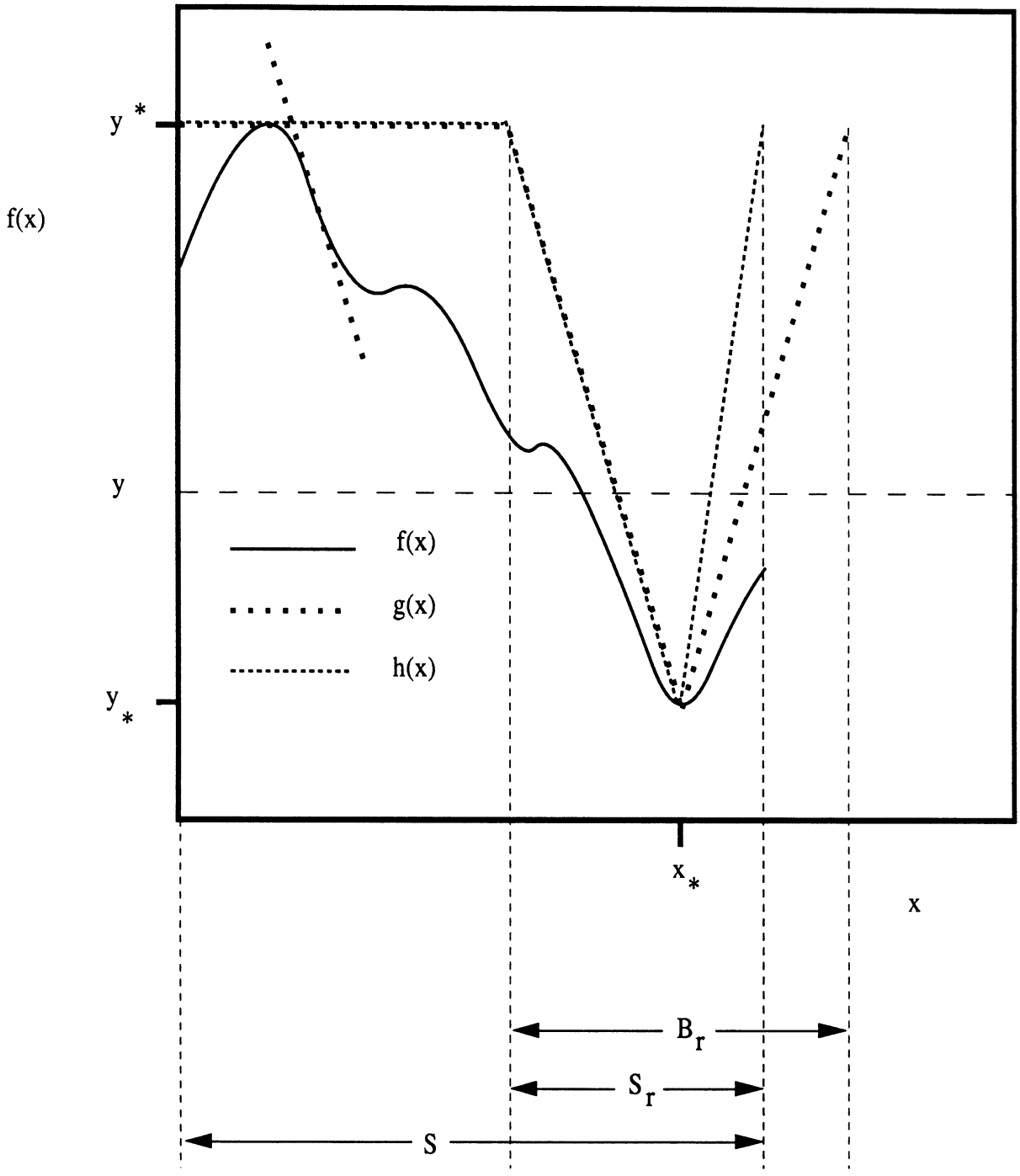

Figure 1: A Geometric Interpretation in One Dimension ($n = 1$)

for any $x \in S$, or equivalently

$$f(x) \leq k\|x - x_*\| + y_*$$

which equals $g(x)$ using the first expression for defining $g(x)$. Thus $f(x) \leq g(x)$ and $x \in S(y)$.

If $x \notin S_r$, then

$$h(x) = y^*.$$

Also, because $x \in S$, we have $x \notin B_r$, and thus

$$g(x) = y^*.$$

And $f(x) \leq y^*$, so again, $x \in H(y)$, $x \in G(y)$, and $x \in S(y)$. Combining both cases gives us

$$H(y) \subseteq G(y) \subseteq S(y).$$

Now, using this result yields,

$$\begin{aligned} p(y) &= \nu(S(y))/\nu(S) \\ &\geq \nu(G(y))/\nu(S) \\ &\geq \nu(H(y))/\nu(S) \\ &= (\nu(H(y))/\nu(S_r))\,(\nu(S_r)/\nu(S)) \end{aligned}$$

and due to the similarity of the convex conical sets,

$$= ((y - y_*)\,/\,(y^* - y_*))^n\,(\nu(S_r)/\nu(S))\,.$$

Thus the first intermediate bound is:

$$p(y) \geq ((y - y_*)\,/\,(y^* - y_*))^n\,(\nu(S_r)/\nu(S))\,. \tag{1}$$

It remains to develop a lower bound on $\nu(S_r)/\nu(S)$. Before we continue, notice that for some mathematical programs, the set S_r equals the set S, and thus the bound simplifies to

$$p(y) \geq ((y - y_*)\,/\,(y^* - y_*))^n$$

For example, let $f(x) = |x|$, with $S = [-2, 2]$. In this simple example, $k = 1$, $y^* - y_* = 2$, and $r = 2$. Thus we have an example with $B_r = S = S_r$, and the bound is tight. In fact, the bound in equation 1 is tight for the class of convex programs, even though S_r is not necessarily equal to S for all convex programs. A proof is given in Patel, Smith and Zabinsky [16].

We now show that $\nu(S_r)/\nu(S) \geq \nu(B_r)/\nu(B_d)$. In order to continue the proof, we need to introduce a similarity transformation. Let $\lambda(x) : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be the affine function defined by

$$\lambda(x) = x_* + c\,(x - x_*) \qquad \text{for any } x \in S \text{ with } c = \frac{r}{d} = \frac{(y^* - y_*)}{kd}$$

which takes a point x in S and moves it towards x_* by a factor of c. Also, let

$$\tilde{B}_d = \{\tilde{x} : \tilde{x} = \lambda(x), x \in B_d\}$$

and similarly,

$$\tilde{S} = \{\tilde{x} : \tilde{x} = \lambda(x), x \in S\}.$$

The first step is to establish that

$$\tilde{B}_d = B_r. \tag{2}$$

This follows directly from applying the similarity transformation, λ, to B_d, which shrinks a ball of radius d centered at x_* by a factor of $c = r/d$. This yields a ball of radius r centered at x_*, or B_r.

The second step is to establish that

$$\tilde{S} \subseteq S_r. \tag{3}$$

To prove this, consider any $\tilde{x} \in \tilde{S}$. We show that $\tilde{x} \in S$, and $\tilde{x} \in B_r$. Now, $\tilde{x} \in S$ because

$$\begin{aligned} \tilde{x} &= \lambda(x) \qquad \text{for some } x \in S \\ &= x_* + c(x - x_*) \\ &= cx + (1 - c)x_* \end{aligned}$$

and since $x, x_* \in S$ and $0 \leq c \leq 1$, and S is convex, we conclude

$$\tilde{x} \in S.$$

Also, $\tilde{x} \in B_r$ because

$$\begin{aligned} \|\tilde{x} - x_*\| &= \|\lambda(x) - x_*\| \qquad \text{for some } x \in S \\ &= \|x_* + c(x - x_*) - x_*\| \\ &= \|c(x - x_*)\| \\ &= c\|(x - x_*)\| \end{aligned}$$

and by the definition of $c = r/d$

$$= (r/d)\|(x - x_*)\|$$

and because $x \in S$ and d is the diameter of S,

$$\begin{aligned} &\leq (r/d)d \\ &\leq r, \end{aligned}$$

UNIVERSITY OF MICHIGAN
3 9015 04732 6759

hence,

$$\tilde{x} \in B_r.$$

Therefore, $\tilde{x} \in S_r$, since $S_r = S \bigcap B_r$.

Now, using equation 3 yields,

$$\begin{aligned} \nu(S_r)/\nu(S) &\geq \nu(\tilde{S})/\nu(S) \\ &= \nu(\tilde{B}_d)/\nu(B_d) \end{aligned}$$

because the ratio of the contents of sets is preserved under the similarity transformation, and

$$= \nu(B_r)/\nu(B_d)$$

from equation 2. Therefore,

$$\nu(S_r)/\nu(S) \geq \nu(B_r)/\nu(B_d).$$

Substituting the above inequality into equation 1 yields the second intermediate bound:

$$p(y) \geq \left((y - y_*) / (y^* - y_*)\right)^n \left(\nu(B_r)/\nu(B_d)\right). \tag{4}$$

But the ratio of the volumes of two n-dimensional hyperspheres of radii a and b is $(a/b)^n$ [14], thus

$$\begin{aligned} \nu(B_r)/\nu(B_d) &= (r/d)^n \\ &= \left((y^* - y_*)/kd\right)^n. \end{aligned}$$

Hence, the final bound on $p(y)$ is:

$$p(y) \geq \left((y - y_*)/kd\right)^n \tag{5}$$

and the proof is concluded. ∎