Proceedings of the 42nd IEEE
Conference on Decision and Control
Maui, Hawaii USA, December 2003

Probability Model for an Adaptive Random Search Algorithm

Rajeeva Kumar
Department of Aerospace Engineering
The University of Michigan
Ann Arbor, MI 48109, USA
rajeevak@umich.edu

David C. Hyland
Pierre T. Kabamba
Department of Aerospace Engineering
The University of Michigan
Ann Arbor, MI 48109, USA
[dhiland, kabamba]@umich.edu

Abstract— **One class of multi-modal optimization approaches that could overcome the curse of dimensionality is random search optimization. However, very little is known about how the parameters of such algorithms should be adjusted in order to achieve a desired convergence speed. In this paper we extend our previous work of developing the probability model for one-parameter systems to two parameter problems when the search domain is a circle. The analysis reveals the key parameters affecting convergence time and provides insight on ways to tune the algorithm for more rapid convergence.**

I. INTRODUCTION

Randomized algorithms provide feasible alternatives for finding a global optimum in multi-mode and higher dimensional problems. However, an issue in the implementation of such algorithms is the criterion for judging the degree of convergence and stopping the algorithm. In a deterministic local optimization problem, one can stop the algorithm and declare convergence when the successive "estimates" of the optimum do not lead to significant "refinement". However, such approach is not suitable for random search techniques. Another criterion for declaring "convergence" comes from the Chernoff Bound [1]. Using this bound one can compute the number of random samples needed in order to reach an acceptable solution within specified confidence level. This bound, however, does not take into account the exact probability distribution and hence normally gives a very conservative result (typically, orders of magnitude off when compared to actual simulation). Horst and Pardalos's handbook [2] contains some other terminating criteria for stochastic search algorithms. But again these criteria are quite conservative.

The current work is intended to remedy this situation by providing at the outset, a reasonable estimate of the number of samples one needs to evaluate for a particular random search algorithm, namely the Adaptive Random Search Algorithm (ARS). In this paper, we analyze the ARS algorithm of [3] for the two parameter case. The algorithm is very similar to the Adaptive Random Search technique first proposed by Bekey et.al. [5] and subsequently modified by Pronzato and Walter [4]. The main contribution of this paper is the development of a probability model for a two-parameter system that can easily be extended to higher dimensions. Using this model in conjunction with the stopping criterion proposed by Kumar and Hyland [3], one can estimate the maximum number of samples needed to reach an acceptable solution within the specified confidence level, for the proposed ARS algorithm.

The paper is organized as follows. In section 2, we present the Adaptive Random Search Algorithm. In section 3, we state the problem. In section 4, we develop an analytical model for the probability distribution of the number of iterations it takes the algorithm to converge for the two-parameter case. In section 5, we illustrate the accuracy of the model by verifying it for two example functions. We, then, discuss the tuning of the algorithm with respect to the different parameters involved. Section 6 concludes the paper with some recommendation for future work.

II. ARS ALGORITHM

The ARS algorithm, analyzed here, is a modified version of Pronzato's [4]. The algorithm is concerned with the global minimization of a continuous function $f(.)$, i.e.

$$f^* = \min_{\vec{x} \in \Omega} f(\vec{x}), \tag{1}$$

Let, the set of feasible input parameters $\Omega \subset R^n$, be compact and defined by $\|\vec{x} - \vec{x}_0\| \leq C_0$. Under these conditions the minimum solution value f^* exists and is attained on Ω. In other words, the set $X^* = \{\vec{x} \in \Omega : f(\vec{x}) = f^*\}$ is nonempty. Generally, we can not find a point in X* in finite time [2]. We consider the global optimization problem solved, if, we find a point in the level set

$$X_\varepsilon = \{\vec{x} \in \Omega : f(\vec{x}) \leq f^* + \varepsilon\} \tag{2}$$

for some $\varepsilon > 0$. We refer the level set X_ε as the acceptable set.

Let $\vec{x}^k = \left(x_1^k, x_2^k, ...x_N^k\right)^T$ be the N-dimensional parameter vector at iteration k of the algorithm, and $f^k = f(\vec{x}^k)$, the corresponding value of the performance criterion. The ARS algorithm is :

- Set constants $f_1,\ f_3,\ f_5, \gamma$.
- Step 0 : assign $i = 0,\ \vec{x}_{best} = \vec{x}_0$.
- Step 1: $i = i + 1,\ C = \gamma^{(i-1)/N} C_0,\ \vec{x}_c = \vec{x}_{best}$, assign $k = 0$ and choose N_i and redefine the search space Ω_i as $\|\vec{x} - \vec{x}_c\| \leq C$.
- Step 2: $k = k + 1$, generate a new trial point $\vec{y}^k = \vec{x}_c + \vec{r}^k$, where the random displacement vector $\vec{r}^k$ is uniformly distributed random vector over the

sample space Ω_i. if $\vec{y}^k \notin \Omega_i$, it is discarded and replaced by an admissible one.

- Step 3: if $f(\vec{y}^k) < f(\vec{x}_{best})$ then assign $\vec{x}_{best} = \vec{y}^k$.
- Step 4: Stop if minimum is found and output the minimizer $\vec{x}_{min} = \vec{y}$, else if $k \leq N_i$ go to *step 2*, else, if $i \leq f_1$, go to *step 1*, else, go to *step 0*.

Some typical values for various parameters are, $f_1 = 5, N_i = f_3/i, f_3 = 100, f_5 = 5, \gamma = 0.1$. When applied to many benchmark problems from the literature [4], global convergence occurs over a wide range of test problems for the above nominal parameter (f_1 to γ) values. While the literature suggests that the convergence is faster or comparable to many other global optimization methods, there exists no *a priori* estimate of the number of iterations required for the desired degree of convergence.

III. PROBLEM STATEMENT

Given the ARS algorithm and the global minimization problem of the previous section, let $\vec{x}^m \in \Omega$, be the m^{th} sample drawn by the algorithm. Then,

- Characterize the statistical properties of the random variable m, such that

$$m = \min\left\{i : \vec{x}^i \in X_\varepsilon\right\}.$$

Assume, (A.1) the dimension of parameter space, $N = 2$. (A.2) the set X_ε is connected. (A.3) the function $f(.)$ is locally convex and (A.4) in the neighborhood of the acceptable set X_ε, the function $f(.)$ is such that $d(\vec{r}_1, X_\varepsilon) \leq d(\vec{r}_2, X_\varepsilon) \Rightarrow f(\vec{r}_1) \leq f(\vec{r}_2)$, where $d(\vec{r}_i, X_\varepsilon) =$ distance of the point $\vec{r}_i \notin X_\varepsilon$, from the boundary of X_ε. We further assume that (A.5) at each stage i, the number of trials, N_i , is large enough such that the algorithm results with probability one, in at least one point in the region where assumption (A.4) holds. Let, R and a be the Lebegue measures of set Ω and X_ε respectively. Let Pr(Γ) denote the probability measure of the event Γ .

IV. MAIN RESULT

For the problem as defined above, we have the following result

Theorem 1: Under assumptions (A.1)-(A.5), the probability of success at a given iteration, i.e. finding an $\vec{r}_m \in X_\varepsilon$, for the first time in the m^{th} iteration, of the ARS algorithm described earlier is given by:

$$\Pr(m) = \left(1 - \frac{a}{R}\right)^{m-1} \frac{a}{R}, \quad \forall m = 1, ..., N_1 \tag{3}$$

and, $\Pr(\sum_{j=1}^{i} N_j + m) = P_{not_i} F_{i+1}(m)$,

$$\forall i = 1, .., f_1; \quad \forall m = 1, ..., N_{i+1}, \tag{4}$$

where,

$P_{not_i(i=1...,f_1)}$=Probability of not finding a solution (i.e. $\vec{r} \notin X_\varepsilon$ till i^{th} stage)

$$= 1 - \sum_{m=1}^{\sum_{j=1}^{i} N_j} p(\vec{r}_m \in X_\varepsilon), \tag{5}$$

$$F_i(m) = \int_0^{2\pi} d\hat{\theta} \left\{ \int_{(\hat{r},\hat{\theta}) \in B} p_{\min_i}(\hat{r}, \hat{\theta}) (1-\beta)^{m-1} \beta \, d\hat{r} + \int_{(\hat{r},\hat{\theta}) \in C} p_{\min_i}(\hat{r}, \hat{\theta}) (1-\kappa)^{m-1} \kappa \, d\hat{r} \right\}, \tag{6}$$

$$B = \{(r, \theta) : \Omega_i \cap X_\varepsilon \neq X_\varepsilon\}, \quad \beta = \frac{\alpha_i(\hat{r}, \hat{\theta})}{\gamma^{i-1} R}, \tag{7}$$

$$C = \{(r, \theta) : \Omega_i \cap X_\varepsilon = X_\varepsilon\}, \quad \kappa = \frac{a}{\gamma^{i-1} R}, \tag{8}$$

and

$$p_{\min_i}(\hat{r}, \hat{\theta}) = -\frac{\partial}{\partial \hat{r}} \left(\frac{1}{\gamma^{i-1} R - a} \int_{d(\hat{r},a) > r} dA \right)^{N_i}. \tag{9}$$

The integral in equation (9) denotes the area integral and $\alpha_i(\hat{r}, \hat{\theta})$ is the *Lebesgue measure* of set $(X_\varepsilon \cap \Omega_i)$. ■

Proof : We prove the above proposition by construction as follows. The algorithm has various distinct stages. For example, the first stage includes the first N_1 samples. For this stage, the probability of finding an acceptable candidate is trivial and is simply given by the expression (3). In the second stage of N_2 samples (i.e. $m = N_1 + i$, $i = 1, ..N_2$), the probability of finding an acceptable candidate sample is the product of probabilities that it has not been found in the first N_1 samples and that it will be found at the m^{th} additional sample as stated in (4), with the individual terms defined as

$$P_{not_1} = \left(1 - \frac{a}{R}\right)^{N_1}, \tag{10}$$

And

$F_2(m)$ = probability that $\vec{r}_{N_1+m} \in X_\varepsilon$ and $\vec{r}_{N_1+k} \notin X_\varepsilon$ for $k = 1, ..., m-1$.

In this stage, the samples are drawn from a modified search domain that is shrunk by a factor γ and relocated around the best point, i.e., the point having the smallest $f(.)$ value in the last N_1 samples. To compute $F_2(m)$, we need the information about the best point. Pertinent to this are assumptions (A.4) and (A.5). Thus the point $\vec{r}$, corresponding to the minimum value of $f(.)$ in the first N_1

trials, is that point which is closest to the boundary of the set X_ε. Since this point $\vec{r}$ itself is a random variable, we need its probability density function. For a general case, this density function, $p(\hat{r},\hat{\theta})$, is derived in appendix A. It is easy to see that for a given location of the center $(\hat{r}=r_0,\hat{\theta}=\theta_0)$ of the search domain Ω_2, we have for samples $\vec{r}_m,\ m=N_1+1,...,N_2$

$$\Pr\left(\vec{r}_m \in X_\varepsilon \text{ for the first time } |r_0,\theta_0\right) = \begin{cases} 0 & (r_0,\theta_0)\in A \\ \left(1-\frac{\alpha(r_0,\theta_0)}{\gamma R}\right)^{m-1}\frac{\alpha(r_0,\theta_0)}{\gamma R} & (r_0,\theta_0)\in B \\ \left(1-\frac{a}{\gamma R}\right)^{m-1}\frac{a}{\gamma R} & (r_0,\theta_0)\in C \end{cases} \quad (11)$$

Where *set* $A=\{(r,\theta):\Omega_2\cap X_\varepsilon=\Phi\}$, Set B and C are defined in (7) and (8 respectively; and $\alpha(r_0,\theta_0)$ is the Lebegue measure of set $(X_\varepsilon\cap\Omega_2)$. Averaging this over all possible minimum distances we get $F_2(m)$ as expression (6) with $i=2$.

The subsequent stages of the algorithm are similar to stage 2, except for the number of samples drawn and the size of the search domain. So, we can obtain the probability model just by replacing the number of samples N_1 by N_i , the size of the search domain γR by $\gamma^{i-1}R$. By doing this we will obtain the expression (3)-(4). ■

The term $F_i(m)$ is very difficult to compute in its complete generality. However, many arbitrarily shaped X_ε of Lebegue measure, $\mu(X_\varepsilon)=a$, can be approximated to be a circle with radius $r_d=\sqrt{a/\pi}$ under certain conditions (refer to appendix A). This approximation leads to the following Lemma.

Lemma 1: If the acceptable set X_ε , is bounded from the inside by circle D_1 of radius r_a and from the outside by circle D_2 of radius r_b $(r_b\geq r_a)$, then the probability density function of the minimum distance ((9), above) can be written as

$$p_{\min_i}(r)=p(r)|_{D_1}+\Psi(r_a,r_b,r), \quad (12)$$

where,

$$|\Psi(r_a,r_b,r)|\leq\frac{r_b-r_a}{r+r_a}\,p(r)|_{D_1}, \quad (13)$$

$$p(r)|_{D_1}=2\pi N_i\left(\frac{r}{\gamma^{i-1}R-\alpha}\right)\left(\frac{\gamma^{i-1}R-\pi r^2}{\gamma^{i-1}R-\alpha}\right)^{N_i-1}, \quad (14)$$

and α is Lebegue measure of $D_1=\pi r_a^2$

Proof : Appendix A. ■

If X_ε is a circle, then $r_a=r_b=r_d$(say). In this case the correction term $\Psi(r_a,r_b,r)$ vanishes and expression (6) for $F_i(m)$ simplifies greatly as

$$F_{i+1}(m)= \int_{\max(\gamma^{i/2}C_0-r_d,r_d)}^{r_d+\gamma^{i/2}C_0} dr\,p_{\min_i}(r)\left(1-\frac{\alpha(r)}{\gamma^iR}\right)^{m-1}\frac{\alpha(r)}{\gamma^iR} + \int_{r_d}^{\max(\gamma^{i/2}C_0-r_d,r_d)} dr\,p_{\min_i}(r)\left(1-\frac{a}{\gamma^iR}\right)^{m-1}\left(\frac{a}{\gamma^iR}\right). \quad (15)$$

The probability function derived in this section can be used in conjunction with the stopping criterion proposed in [3] to estimate, at the outset, the maximum number of independent trials needed using the ARS algorithm for an acceptable solution. In essence, the stopping criterion in [3] states that, for a given desired confidence level,δ , one should find the number of iterations such that the cumulative probability (for each iteration) adds up to at least $1-\delta$. In next section we compare our probability model with results obtained from simulation.

V. EXAMPLES

Example-1 : Two Dimensional Negative Sinc Function $f(\vec{r})=-\sin(|\vec{r}|)/|\vec{r}|$, where $|\vec{r}|\leq 30$.

This function conforms to the assumption (A1-A4) of the theorem. The ARS algorithm is used to find the global minimum of this function with parameters $\gamma=0.1, N_1=300$ and $N_{i(i=2,..5)}=N_1/i$. To obtain an estimate of the probability of the number of iterations required to find a solution, we run the algorithm 100000 times with the tolerance $\varepsilon=0.0036$. With each run, we record the number of iterations the algorithm takes to find an acceptable solution. In Fig. 1, we plot the histogram of these number of iterations with the probability mass function computed as per (3) and (4) with the correction term of (12), $\Psi=0$. For the theoretical model, we have used $r_d=0.0219$ which corresponds to $\varepsilon=0.0036$. As can be seen, the theoretical model matches very well with the simulation results.

In the above example, if we choose the threshold $\delta=1\text{x}10^{-5}$, the above stopping criterion gives the value of m as 1271. This means that if we stop the algorithm after 1271 iterations, the minimum point attained by the ARS lies in the acceptable set X_ε with probability $1-\delta$. Actually, for the 100000 times the algorithm runs, none of the runs had to exceed our estimate of 1271 iterations and the maximum recorded necessary number of iterations (667) was comparable to this estimate.

Example -2: The Rastrigin function[6]

$$f(x_1,x_2)=2+x_1^2-\cos(18x_1)-\cos(18x_2),$$

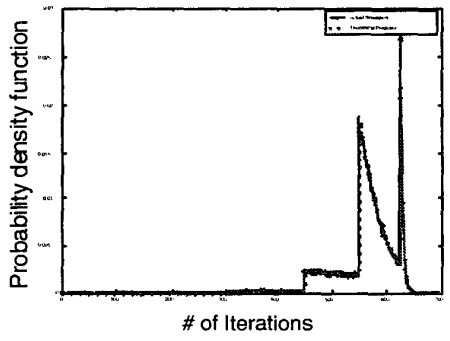

Fig. 1. Comparison of the theoretical pdf with simulation for 2-D Sinc Function

where, $x_1^2 + x_2^2 \leq 1$.
This function has a global minimum at the origin. This function does not satisfy assumption (A.4). Here again the algorithm is run 100000 times with parameters $\gamma = 0.1, N_1 = 300$, $N_{i(i=2,..,5)} = N_1/i$ and the tolerance $\varepsilon = 0.01$. As before, we record the number of iterations the algorithm takes to find an acceptable solution, in each run. In Fig. 2 we plot the histogram of these numbers of iterations with the probability mass function computed as per (3)-(4). For the theoretical model, we have used $r_d = 0.0065$ computed as $\sqrt{X_\varepsilon/\pi}$ where X_ε corresponds to the size of acceptable set for $\varepsilon = 0.01$. As can be seen, the theoretical model matches reasonably well with the experimental results.

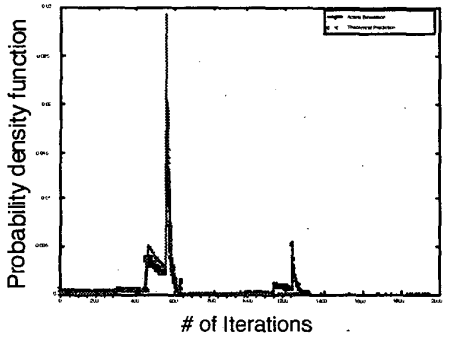

Fig. 2. Comparison of the theoretical pdf with simulation for Rastrigin function

In this case for $\delta = 1\text{x}10^{-5}$, the stopping criterion predict 1297 iterations. Out of 100000 simulations 224 exceeded this predicted maximum number of iterations.

The above examples show that for cases where the optimizing function satisfies the assumptions listed in section III, the probability model given in (3)-(4) characterizes the randomness of the numbers of iteration for the algorithm to converge, extremely well.

A. Usefulness of the Probability Model

The real use of this probability model lies in estimating the maximum number of trials one needs to perform in order to obtain the minimum value as per the stopping criterion proposed by Kumar and Hyland [3]. Even though the model assumes the knowledge of r_d, $(= \sqrt{\mu(X_\varepsilon)/\pi})$, it is not very restrictive. This is because, one can always use a very small value of r_d and obtain a conservative estimate of the number of trials required. Further, the probability model for the number of iterations, as given in (3) and (4) in conjunction with the stopping criterion, can be used to choose various parameters of the algorithm for faster convergence. Assuming we have fixed all other parameters of the algorithm but one, we can obtain the most suitable value of this parameter by plotting it against the number of iterations obtained from the stopping criterion [3]. In Fig. 3, we present one such curve as a function of the reduction parameter γ of the algorithm for a search space of dimension 2.

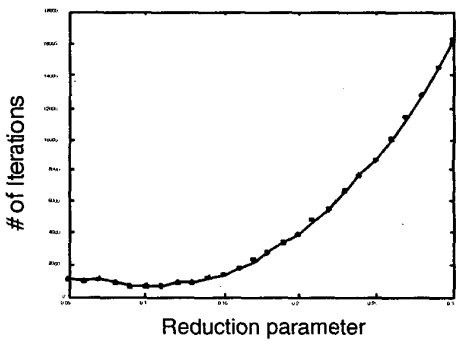

Fig. 3. Tuning of ARS: Selecting γ

In obtaining this curve, we fixed the other parameters as $N_1 = 100$, $C_0 = 1.5$ and $r_d = 0.0074$. It is interesting that unlike the 1-D case [3], here $\gamma = 0.1$ appears to be the best value. This is also verified through simulation. Similar studies can be undertaken for other parameters too.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we considered a version of the Adaptive Random Search Algorithm that is similar to the algorithm proposed by Pronzato. We developed the probabilistic model for the number of iterations the algorithm takes to find an acceptable solution in a two parameter system. The key here is to assume a particular shape for the set search domain.

We verified the model through numerical simulation. This model can be used in conjunction with the stopping criterion proposed by Kumar and Hyland [3] to predict the maximum number of samples one would require in order to achieve an acceptable solution with the desired confidence. We also illustrated the usefulness of the probability model through examples. This formulation assumes that the function is monotonically increasing with distance from the boundary of acceptable set X_ε. If the deviation from this assumption is not very large, the model still represents the actual algorithm reasonably well.

The model developed here is valid for the two parameter case and where the acceptable set is connected. However, the formulation is such that this can easily be extended to a larger number of parameters. Additional extensions could include generalization of this model to a more general form of acceptable set and for functions that do not monotonically increase with distance from the boundary of acceptable set X_ε.

VII. REFERENCES

[1] Vidyasagar, M., *A Theory of Learning and Generalization: with Application to Neural Networks and Control Systems*, Springer-Verlag, London, 1997.

[2] Horst, R.,ed., Pardalos, P. M., e.d., *Handbook of Global optimization*, Kluwer Academic Publishers, 1995.

[3] Kumar, Rajeeva, Hyland, D., "Tuning of a Random Search Algorithm for Controller Design", *AIAA Paper 2002-4762,AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2002.

[4] Pronzato, L., Walter, E., Venor, A., Lebruchec, J.-F, "A General Purpose Global Optimizer: Implementation and Application", *Mathematics in Computers and Simulation*, Vol. 26, pp. 412-422, 1984.

[5] Bekey, G.A., Masri, S.F., "Random Search Techniques for Optimization of Non-linear Systems with many parameters", *Mathematics and Computers in Simulation*, Vol.25, pp. 210-213, 1983.

[6] Solomatine, D. P. "Random Search Methods in Model Calibration and Pipe Network Design", *Water Industry Systems: Modelling and Optimization Applications*, Vol.2, D.Savic, G. Walters(eds). Research Studies Press Ltd., Baldock, UK, 1999, pp317-332.

Appendix A

PROBABILITY DENSITY FUNCTION OF MINIMUM DISTANCE FOR AN ARBITRARY SHAPED ACCEPTABLE DOMAIN

Consider the search domain S in R^2, sampled by uniformly distributed random vectors, $\vec{r}_1, ..., \vec{r}_N$. We are interested in finding the probability distribution of the minimum distance of these N samples from the arbitrary shaped boundary of domain X_ε provided none of the samples fall in set X_ε. The distances of points represented by vectors $\vec{r}_1, ..., \vec{r}_N$, from the boundary of X_ε are denoted by $r_1, ..., r_N$ respectively(Fig. 4). Let α and β be the Lebegue measures of X_ε and S respectively. The probability density of the terminus of vector $\vec{r}_i$, given that it is not contained in X_ε is $(\beta - \alpha)^{-1}$.

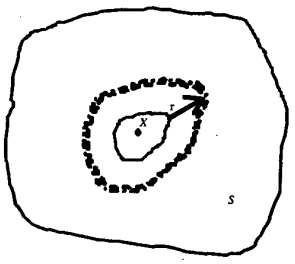

Fig. 4. Calculating probability of minimum distance

Assume that $r = r_1$ is the minimum distance amongst all N samples. This means $r_1 < r_2, r_3, ...r_N$. This implies that the probability $Pr(\vec{r}_1 \in [\vec{r} + d\vec{r}])$ is the product of probabilities of sample $Pr(\vec{r}_1 \in [\vec{r} + d\vec{r}])$ and the probability that all other samples lie beyond this strip, i.e.

$$\Pr(r_1 \in [r + dr]) = \frac{dr}{\beta - \alpha} \left(\int_{d(\hat{r},a)>r} d\hat{r} \frac{1}{\beta - \alpha} \right)^{N-1} \tag{16}$$

Here the integral denotes the area of domain S for which $\hat{r} > r$ from the boundary of X_ε . Since sample $\vec{r}_1$ is no special sample and any samples amongst $\vec{r}_1, ..., \vec{r}_N$ can correspond to the minimum distance, we have the probability of the minimum distance as

$$\Pr(r \in [r + dr]) = \frac{N\, dr}{\beta - \alpha} \left(\int_{d(\vec{r}_2,a)>r} d\vec{r}_2 \frac{1}{\beta - \alpha} \right)^{N-1} . \tag{17}$$

$$= -N \left(\frac{\partial}{\partial r} \int_{d(\hat{r},a)>r} d\hat{r} \frac{1}{\beta - \alpha} \right) \left(\int_{d(\hat{r},a)>r} d\hat{r} \frac{1}{\beta - \alpha} \right)^{N-1} . \tag{18}$$

$$\Rightarrow p(r) = -\frac{\partial}{\partial r} \left(\int_{d(\hat{r},a)>r} d\hat{r} \frac{1}{\beta - \alpha} \right)^{N} \tag{19}$$

If the region X_ε is a circle, then the above expression reduces to

$$p(r_0) = N2\pi \frac{r_o}{R-\alpha} \left(\frac{R - \pi r_0^2}{R-\alpha} \right)^{N-1} . \quad (20)$$

Here, r_0 is measured from the center of domain X_ε.

Let r_a be the radius of largest circle (say D_1) contained in X_ε. Similarly, let $r_b (\geq r_a)$ be the radius of smallest circle (say D_2) containing X_ε. Further, let r_S be the radius of the initial search domain. Then using (20), the probability density function of the minimum distance r from the boundary of D_1.

$$p(r)|_{D_1} = \left(\frac{1}{R-a} \right)^N N \left(R - \pi (r + r_a)^2 \right)^{N-1} 2\pi (r + r_a) . \quad (21)$$

Hence,

$$\frac{p(r)|_{D_2} - p(r)|_{D_1}}{p(r)|_{D_1}} \leq \left(\frac{r_b - r_a}{r + r_a} \right) . \quad (22)$$

Boundaries D_1 and D_2 represent two extreme cases concerning the actual situation. For arbitrary shaped X_ε, the probability density of the minimum distance will fall between these two extreme cases. In other words,

$$p(r)|_{X_\varepsilon} \leq p(r)|_{D_1} + \frac{r_b - r_a}{r + r_a} p(r)|_{D_1} . \quad (23)$$

If set X_ε is such that r_a and r_b are of comparable order but both are very small as compared to r_s, $p(r)|_{X_\varepsilon}$ can be approximated by $p(r)|_{D_1}$.