On success rates for controlled random search

ELIGIUS M.T. HENDRIX[1], P.M. ORTIGOSA[2] and I. GARCíA[2]
[1]*Group Operations Research and Logistics, Wageningen University, Hollandseweg 1, 6706 KN, Wageningen, The Netherlands;* [2]*Computer Architecture & Electronics Department, University of Almería Cta. Sacramento SN, 04120 Almería, Spain*

Abstract. Controlled Random Search (CRS) is a simple population based algorithm which despite its attractiveness for practical use, has never been very popular among researchers on Global Optimization due to the difficulties in analysing the algorithm. In this paper, a framework to study the behaviour of algorithms in general is presented and embedded into the context of our view on questions in Global Optimization. By using as a reference a theoretical ideal algorithm called N-points Pure Adaptive Search (NPAS) some new analytical results provide bounds on speed of convergence and the Success Rate of CRS in the limit once it has settled down into simple behaviour. To relate the performance of the algorithm to characteristics of functions to be optimized, constructed simple test functions, called extreme cases, are used.

Key words: stochastic global optimization, random methods, adaptive random search, speed of convergence.

1. Introduction, questions in GO

Research on Global Optimization (GO) mainly concerns the investigation of properties of algorithms and their interrelation with (mathematical structure of) Global Optimization problems. The final aim of GO research is to get a better understanding of which methods are best suited to be used on which type of practical optimization problems. Two aspects can be distinguished and classified further in research on GO:

(a) the methods

One can roughly distinguish two major approaches (Hendrix, 1998):

(1) Deterministic methods which aim at guarantee to approach the global optimum and therefore require a certain mathematical structure.

(2) Stochastic methods which are based on the random generation of feasible trial points and nonlinear local optimization procedures.

A profound discussion and classification of methods can be found in (Törn and Zilinskas, 1989). In this paper, focus is on a stochastic optimization procedure called Controlled Random Search.

(b) the structure of the optimization problem

Roughly two types of problems are discussed in literature on Global Optimization: The black box case, also called oracle case, in which no structure on the problem is assumed or available and the structured problems. In 'The Handbook of Global Optimization' (Horst and Pardalos, 1995), most chapters are classified with respect to a mathematical *structure*, assumed for the problem to be solved. Given a particular optimization problem, confronted with the question which method, or which particular algorithm is most appropriate to solve it, this distinction is important. One first should discover which useful mathematical structure and corresponding useful properties are hidden in the problem.

In this paper we consider the unstructured problem (black box case). One should notice, that despite information on structure is assumed not to be available, other information, so-called *value information*, becomes available when running algorithms, such as: number of local optima (better: results of a local optimization routine) found thus far, average number of function evaluations necessary for one local search, best function value found, etc. Such indicators can be measured empirically and on one hand can be used to get insight into what factors determine the success of a particular algorithm and perhaps on the other hand can be used to improve the performance of an algorithm.

Observing literature on GO, we can distinguish three types of *results*:

(1) A description of a practical problem, analysed and solved successfully by a particular method. Such successes illustrate the relevance of studying Global Optimization. Usually this leads to further questions on whether perhaps there are other methods more suited to solve the derived mathematical optimization problem. First of all, with this result the translation has been made from practical problem to a mathematical optimization problem.

Controlled Random Search has contributed to solve (and understand) practical problems resulting in many papers, among others García et al. (1997) and Klepper and Rouse (1991). It is one of the methods using a population of candidate points. It has been popular in parameter estimation, as the cloud of final points indicates the sizes of compartments around local optima providing insight into the uncertainty of estimated parameters (see, e.g., Klepper and Hendrix (1994)).

(2) Numerical results, sometimes called empirical or experimental results. They are often based on standard test problems for which a particular algorithm is run and various performance criteria are evaluated. Despite the fact that sometimes it looks like an *advertisement context* for a particular method, it is a useful way to illustrate tendencies, but even more a way to detect surprising counter intuitive results which require further analysis.

(3) Analytical results in the form of theorems and propositions. Assumptions are made either on algorithms or on problems to be solved and properties on the performance either on effectiveness or efficiency are derived. In contrast to

the attraction of CRS to be used to solve practical problems, CRS has never been very popular among researchers on the theory of global optimization algorithms; up to date there was no analysis on the performance of the algorithm. Some benchmark analytical results exist on stochastic methods related to CRS, which are discussed in Section 2.

In this paper we aim at an intermediate between analysis and experiments. Constructed test functions are used to analyse which measurable factors determine the performance of the algorithm. In Section 2, a description of the terminology in stochastic methods, its performance indicators, the relevance of Success Rates and a summary of some relevant results are given. In Section 3, a description of the algorithm under study is presented. The research question is refined towards CRS. In Section 4, some analytical results are presented. Finally in Section 5, the consequence of the results for the original research questions is discussed.

2. Success rate and benchmark results on stochastic methods

In this section, the relevant terminology of stochastic methods is introduced and the properties we can aim at, are embedded in existing results for stochastic methods. Furthermore, it is explained why Success Rates are interesting to investigate.

Stochastic methods are understood to contain some stochastic elements. Either the outcome of the method is a random variable or the objective function itself is considered a realisation of a stochastic process. For an overview on stochastic methods we refer to Törn and Zilinskas (1989) and Boender and Romeijn (1995). Apart from the continuity of objective function f, the methods in general require no structure on the optimization problem. Therefore the methods are generally applicable.

In literature on stochastic methods, the main property which is aimed at is that when effort increases to infinity, a point is generated in the ϵ-environment of a global minimum point with probability one. This defines the *effectiveness* of an algorithm. Analytically this main property can be derived whenever a designed algorithm is allowed to sample everywhere in the feasible area with a certain probability. Empirically, by means of test functions, one can measure how many times the global optimum has been reached by a certain algorithm. As in general with stochastic methods, we need many repetitions to measure performance criteria, as to average out the stochasticity. Moreover, for population based algorithms one can redefine the effectiveness criterion as to be able to locate all global minimum points and to cover all corresponding compartments of the $\delta-$level set (see Hendrix and Klepper, 2000). This is much harder and sometimes even impossible.

Efficiency can also easily be measured in an empirical way. As a performance criterion, in general the expected number of function evaluations necessary to reach convergence, is taken. As these empirical results depend on the test functions under consideration, it is good to have a better look at the structure of the functions: which factors determine the performance. Like in a design of experiments one should

construct extreme cases (best or worst cases) of the test functions as to investigate an expected relation as firmly as possible. From an analytical point of view, it is much harder to derive results on the speed of convergence.

In literature on nonlinear optimization (see, e.g., Gill et al., 1981), usually the rate of convergence of iterative sequences is derived as a limit measure when the iteration counter k approaches infinity. Population algorithms like CRS typically follow two stages. In a first stage the global minimum points and their environments should be found and the population has to leave the 'valleys' of the local minima. In a second stage the exact location of the local minimum points has to be identified. The function evaluations are dedicated to both tasks which complicates the analysis of the number of function evaluations as measure for efficiency. In Section 4, the analysis will be restricted to the limit behaviour in the final stages of the algorithm.

We introduce some formalisation here. We do not intend to give a full exposé on terminology on stochastic methods and leave out ideas, not necessary to understand CRS. Concepts like region of attraction and stopping criteria are not discussed. We introduce the generic problem formulation:

$$\min f(x), \quad x \in X \subset R^n, \tag{1}$$

in which $f(x)$ is a real valued continuous function and the feasible set X, is a compact body. The global minimum is denoted by f^* and the set of (global) minimum points by S^*. The level set belonging to level y is defined as those points in the feasible set with an objective function value less than or equal to y

$$S(y) = \{x \in X : f(x) \leqslant y\} \tag{2}$$

So $S^* = S(f^*)$ is the set of global minimum points. Important in the analysis is the notion of the relative size (volume) of a level set:

$$\mu(y) = V(S(y))/V(X) \tag{3}$$

in which the function V denotes the volume of a set. The problem is considered to be solved if one obtains a point in the neighbourhood $B_\epsilon(S^*)$ of one of the minimum points; the point is in the set

$$B_\epsilon(S^*) = \{x \in X : \|x - x^*\| \leqslant \epsilon \ \text{for some } x^* \in S^*\} \tag{4}$$

or the function value is in a δ-level set from the global minimum

$$x \in S(f^* + \delta) \tag{5}$$

In the comparison of algorithms, one benchmark is Pure Random Search (PRS). PRS only performs a global search by generating and evaluating independent and uniformly distributed random points x in the feasible set X. So the function values $y = f(x)$ are random variables of which the cumulative distribution function (c.d.f.) is given by $P(y \leqslant y) = P(f(x) \leqslant y) = \mu(y)$. The basis of reaching the

optimum with infinite effort is that after having generated N points, the probability that the problem is solved, i.e. the best point (record) found is in $B_\epsilon(S^*)$, is

$$1 - \left(1 - \frac{V(B_\epsilon(S^*))}{V(X)}\right)^N \tag{6}$$

or alternatively, the probability that the best point is in the $\delta-$level set

$$1 - (1 - \mu(f^* + \delta))^N \tag{7}$$

which converges to one when the number of generated points, N, grows to infinity.

Another benchmark algorithm is called Pure Adaptive Search (PAS). PAS is a hypothetical algorithm that can be used as a reference to study the speed of convergence. PAS is not a real implementable algorithm, but a tool for analysis of complexity and in some sense a theoretical ideal. The analysis in literature focuses on the question: given the current iterate x_k, $y_k = f(x_k)$; what will happen with the speed of convergence if we would be able to sample a point uniformly in the improving region, the level set $S_k = S(f(x_k))$. So, the most important assumption of PAS is that the next iterate x_{k+1} is a sample point from a uniform distribution over the interior of the current level set S_k. At every iteration the function value is improving* The most important property, shown in Patel et al. (1988) and Zabinsky and Smith (1992), is that in some sense the expected number of iterations grows less than exponential in the number of variables of the problem. In Hendrix and Klepper (2000), it has been shown why it is improbable that this ideal will be reached, as it implies the existence of a polynomial algorithm for GO. This means, as it is improbable to have a polynomial algorithm for GO, it is also unlikely that one can fulfill the assumption that sampling a point from a uniform distribution over the current level set can be performed in a time which grows only polynomially in the dimension n of the problem.

To make the way of thinking of PAS applicable for the analysis of population based algorithms such as CRS, the extension of PAS to N-points PAS (NPAS) was introduced. NPAS implies performing PAS with a population of N points simultaneously, in which the current level set $S_k = S(y_k)$ is defined by the worst point in the population. In Baritompa and Steel (1995) and Klepper and Hendrix (1994) was shown that the complexity is the same as that of PAS; NPAS is a hypothetical algorithm, an ideal, with a polynomial complexity. The algorithm to be investigated, CRS, is a realistic, implementable algorithm which deviates from the reference ideal of NPAS on the following points:

(1) A newly generated point is not a sample from a uniform distribution.
(2) Not every new point is an improving point, i.e. has a better function value than the worst point in the population.

With respect to 1, it is useful now to introduce the notion of Success Rate (or conversely Failure Rate). The *Success Rate* (SR_k) is defined in a theoretical

---

* To be more precise, strictly improving, i.e., $f(x_{k+1}) < y_k$ is required.

environment as the probability that the next generated point is an improving one, i.e. is in the interior of the current level set. To follow the line of reasoning of PAS, in NPAS it also should be strictly improving in the sense that $f(x_{k+1}) < y_k$, where y_k is the function value of the worst point of the population which is now discarded. This probability does not only depend on the algorithm, but also on the problem to be optimized, on the dynamics (current level set) and also on the current population (or points). One can define a measurable indicator such as *frequency with which one samples inside* S_k. This indicator can be used as an estimator of the Success Rate whenever SR_k is either fixed to or converges quickly to a 'steady-state' value SR^*. To analyse the behaviour, therefore in Section 4 we will focus on cases for which the Success Rate SR_k of the algorithm converges quickly to a limit value SR^*.

For the estimator, one can either calculate at the end of the algorithm the number of points being evaluated with an improvement compared to the total number of function evaluations, or one can update an estimate to get insight in the development during the course of the algorithm. In particular in the CRS algorithm, a success rate indicator sr_k is calculated and used explicitly to steer the algorithm in its choice between local and global search, as will be explained in Section 3.

An important result with respect to the notion of Success Rate is described in a paper called *Towards Pure Adaptive Search* (Baritompa et al., 1995). In this paper it was shown that PAS can be relaxed in the following way. The Success Rate defined as the probability that the next iteration point is uniform in the improving region should have a fixed (independent of the dimension) lower bound $\rho > 0$. This is called ρ-*adaptive* search and alternatively in (Bulger and Wood, 1998) called hesitant adaptive search; the algorithm hesitates in finding improvements. The same polynomial complexity bound holds as that of PAS. Again by reasoning reversely, we can conclude that it is unlikely that there exists an algorithm with a Success Rate which independently of the dimension is bounded from below by a positive constant ρ. On the other hand, it makes us curious how SR behaves for increasing dimension for realistic, implementable algorithms.

For the understanding of this concept, let us first note that PAS implies always improving, such that implicitly $\rho = 1$. Consider on the other hand, what happens with the Success Rate in Pure Random Search, another extreme. Let us assume that a level y_k has been reached. The probability of an improvement $P(y_{k+1} < y_k)$ (Success Rate for finding a point within $S(y_k)$) approaches zero when y_k approaches the minimum. So the Success Rate is not bounded from below by a positive constant, but converges to zero. When $\mu(y_k)$ is continuous between the minimum and maximum objective function value, (Karnopp, 1963) has shown that the probability that a new function evaluation leads to an improvement after nf evaluations have been performed, is equal to $1/(nf+1)$; it goes to zero. The surprising aspect of this result is, that it does not depend on the optimization problem under consideration.

For CRS no bounds or expressions have been derived in an analytical way, so far. In (Hendrix and Klepper, 2000), a derivation can be found for another population algorithm, similar to CRS, called Uniform Covering by Probabilistic Rejection

(UCPR). Essential in such a derivation is to make assumptions on the form of the level sets. Our framework implies the construction of test functions (extreme cases) in a systematic way such that a relation can be derived between performance (here SR) of an algorithm and factors such as the characteristics of the function to be optimized and the parameters of the algorithm. In the following section, this idea is elaborated for CRS.

3.   Framework: questions on Controlled Random Search

The Controlled Random Search algorithm, proposed by Price, is a simple and direct procedure for global optimization, applicable both to unconstrained and constrained optimization problems (Price, 1978, 1979, 1983). In this work, it is assumed that the global optimization problem to be solved is that described by (1), where X is a hyper-rectangle.
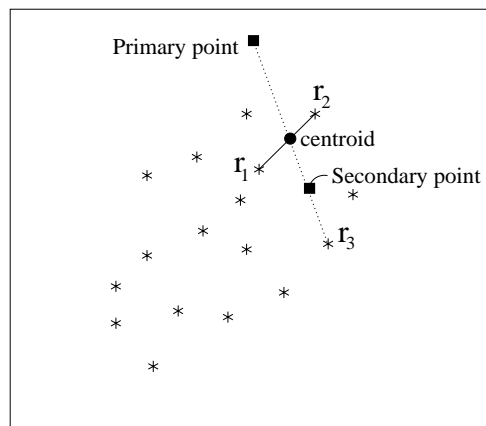

Figure 1. Controlled Random Search

Price described several variants of the algorithm. The version of CRS as investigated in this paper is detailed at Algorithm 1. The algorithm starts by filling the population R initially with a sample of N trial points uniformly distributed over the search space X. The algorithm iteratively updates the set of trial points (population) R until stopping conditions are reached. At every iterative step k, two different kinds of trial points may be computed; the so called primary and secondary trial points. Both kinds of trial points are generated by performing an operation on a subset $\{r_1, \ldots, r_{m+1}\} \subset R$ of $m + 1$ trial points. The points r_i, $i = 1, \ldots, m + 1$ of this generation subset are randomly selected from the current population R of N points. In the first versions of the algorithm, m was taken as the dimension n. Primary points are generated in a Nelder–Mead fashion (Nelder and Mead, 1965) by mirroring (reflecting) a point (r_{m+1}) over the centroid, $\overline{G}$, of the

Algorithm 1. **CRS(**f, X, N, m, α**)**

Set $k = 0, nf = 0$ and $ns = 0$
Generate and evaluate a set R, of N random points uniformly on X
 while (Stopping Conditions)
        select at random a subset $\{r_1, \ldots, r_{m+1}\}$ from R
        $\overline{G} = \frac{1}{m} \sum_{i=1}^{m} r_i$
        $\overline{P} = 2 \times \overline{G} - r_{m+1}$ **# Primary points**
        $k = k + 1, \; nf = nf + 1$
        determine $y_k = f(r^{max}) = max_i f(r_i), r_i \in R$
        if $\overline{P} \in X$ AND $f(\overline{P}) < f(r^{max})$

                replace r^{max} with $\overline{P}$ in R; $ns = ns + 1$
        else if $(sr_k = ns/nf) < \alpha$ **# Rate of Success Test**
                $\overline{P} = \frac{1}{2}(\overline{G} + r_{m+1})$ **# Secondary points**
                if $f(\overline{P}) < f(r^{max})$
                        replace r^{max} with $\overline{P}$ in R; $ns = ns + 1$
        $nf = nf + 1$
End **while**

remaining subset of points $r_1, \ldots, r_m$. A secondary point is located in the middle between r_{m+1} and the centroid $\overline{G}$ (see Figure 1). Note that a secondary point is a convex combination of a subset of the population.

Secondary points are only computed if the current primary trial point fails and the estimate (sr_k) of the success rate on finding better function values than the worst value y_k in the current population, is below the value of a parameter α (Price suggests $\alpha = 0.5$).

Note that whenever the population has reached a convex level set (or one convex compartment of a level set), the calculation of a secondary point implies an improvement (function value is not worse than y_k). However, this does not imply that the algorithm starts behaving as NPAS, in a constant improving manner due to the secondary points. Namely, when the Success Rate exceeds the value of α, only Primary points are generated. This means that in the analysis of the Success Rate we should take care when it starts approaching the level of α. Around this level, the algorithm fluctuates in its decision of generating secondary points.

3.1. **questions on values for the parameters of the algorithm**

Besides α, the population size N and the size m of the subset used to compute the centroid $\overline{G}$, are parameters which can be influenced by the user. In Ortigosa et al. (1999) a variant of the CRS algorithm namely $DCRS$ (Delayed CRS), which uses

a buffer, is described. This mutation permits the algorithm to be run in parallel and also introduces an additional parameter namely the buffer size b.
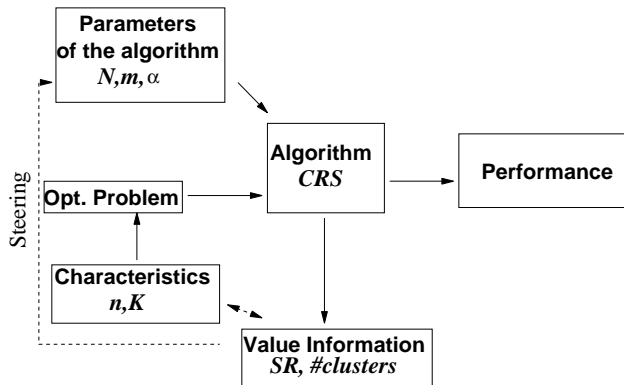

Figure 2. Questions on the performance of CRS, focus on Success Rate

The question is, how much the efficiency increases (or decreases), with varying dimension n of the problem to be solved and with a varying number of global or local minimum points. Which factors cause an increase or decrease in efficiency (expected number of function evaluations)? Some extreme cases are used to study relations between the performance and factors such as the characteristics of the function to be optimized and parameters of the algorithm. It is interesting to know how the Success (Failure) Rate of CRS evolves. In a certain sense it is a performance indicator which, given the analysis of ρ-adaptive search, gives an indication on the possible speed of convergence. The final aim is to get insight in how to improve the performance by adjusting parameters when value information becomes available. One could introduce a feedback choice rule in this way, thus making endogenous other parameters (see Figure 2). An example of such a rule is given in the CRS algorithm in which the estimate sr_k of the Success Rate can be considered value information to steer the choice between primary and secondary points thus introducing implicitly a new parameter α. Another possible source of information is to estimate the number of clusters the population consists of in an attempt to estimate the number of minimum points. This includes the local ones in the beginning, but converges to the global ones whenever the algorithm is sufficiently effective to detect all members of S^*. Let us assume that there is a finite number K of global minimum points in S^*. The algorithm is only able to detect the optima when the final level set consists of K compartments, i.e. there exists a level y for which the level set $S(y)$ consists of K compartments. In the next section, we are going to study the behaviour of CRS analytically.
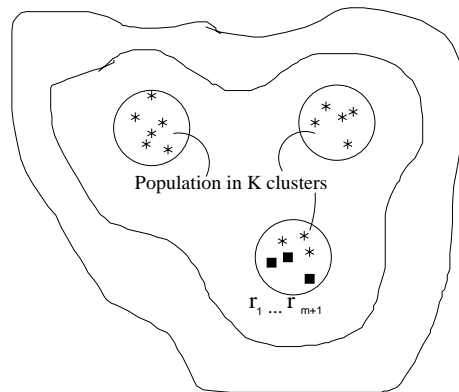
Figure 3. Population spread over several compartments

4. Analytical Results based on extreme cases

This section is intended to discuss the mechanisms involved in CRS and to derive some analytical results, which will be compared to the corresponding experimental results. CRS is a realistic algorithm which can be seen as an approximation of the theoretical ideal of NPAS (or a ρ-Adaptive search). We are interested in the behaviour of the Success Rate. As stated before, it is usual in Nonlinear Optimization to consider the limit behaviour of the algorithm, i.e. the tendency in the final iterations of the algorithm. We use test problems designed such that the Success Rate reaches a fixed value soon, which makes analysis of the limit situation possible and comparable with numerical (experimental) results. In limit, the algorithm encounters two possible situations:

(1) The problem has one global minimum point.

(2) The problem has several (K) global minimum points.

In the end, during the last iterations of the algorithm, the number of clusters the population consists of, is determined by the number of global minimum points. The algorithm is only able to detect the K global minimum points, when the final level set $S(f^* + \delta)$ consists of K compartments. To avoid complications, in Nonlinear Optimization (see, e.g., Gill et al., 1981) usually assumptions on the objective function like sufficient smoothness is posed. For us the question is to study the behaviour when the population spreads out over K clusters. For this to happen, the final level set should consist of K compartments and correspondingly the function should at least have K global minimum points. When the algorithm effectively detects all global minimum points, the population is represented in all compartments of the final level set, as suggested in Figure 3.

For *secondary* points the following can be derived. When there is one global minimum, the final level set often consists of a convex region. Usually in analysis on nonlinear optimization (see, e.g., Gill et al., 1981) an elliptic region is assumed.

As the secondary point is a convex combination of the current population, an improvement (no failure) will be found. Without any further analysis we can derive from this reasoning the feedback rule that when the population during the course of the algorithm starts clustering in one small group, it might be profitable to increase α as more local search is performed.

When the population spreads out in limit over several clusters as suggested in Figure 3 (the clusters are much smaller than illustrated), the probability of a failure for a secondary point becomes high. The limit failure rate is high when the number of clusters (minima) K is high, when the points are more equally spread over the compartments of the level set and when m is big, because the probability that all $m + 1$ points are taken from the same cluster is lower. Of course one can construct exotic instances with such a symmetry that points taken from various compartments by coincidence lead to secondary points which fall into other compartments. The general tendency leads to a feedback rule to keep α low (not to generate many secondary points) whenever the population spreads out over many clusters.

The analysis up till now answers our questions on secondary points sufficiently. Therefore, we concentrate on primary points further. Two questions have our particular interest:

(1) What happens to the limit Success Rate with increasing dimension n?

(2) How is the limit Success Rate influenced by parameter m?

The general idea one can extend from the analysis above is to consider the case when there are K minimum points corresponding with K compartments of the limit level set. When the algorithm is effective, the final population is spread over K clusters corresponding to these compartments. Improvements like that of the *one minimum point* case, usually only take place when all $m + 1$ points are taken from the same compartment. The probability for all $m + 1$ points to be in one compartment when the points are equally spread is K^{-m}, which leads to low probabilities whenever the population is spread over many clusters, corresponding to a large number K of global minimum points. This idea leads to a tendency of concluding not to enlarge the number of points m which are used for building the centroid $\overline{G}$, unless the number of clusters is low. Before the limit has been reached, improvements may also occur by detecting new compartments, minimum points; the global search plays a role.

In the sequel, we will try to find a more quantitative base for the tendencies and to consider the case when the population is not equally spread over the compartments. This can only be done for constructed (extreme case) test functions.

4.1. CONSTRUCTION OF THE TEST CASES

In order to get estimates for a Success Rate, we should look for instances for which the Success Rate is relatively constant during the course of the algorithm. Again, we distinguish two interesting situations:

4.1.1. **The problem has one global minimum point**

In nonlinear optimization it is common to consider the level set around the minimum as an ellipsoidal region. By transformation with the Hessean this leads to spherical level sets. Spherical level sets appear for the so called spherical problem:

$$min\{f(x) = \|x\|\} \text{ on } X = \{x \in R^n : \|x\| \leqslant 1\} \tag{8}$$

The spherical problem is relevant for the general case for two reasons. Firstly, it can easily be transformed to the conical convex program which was used in Patel et al. (1988) to derive the polynomial result of PAS. The conical problem as such provides a bound on convergence speed of the general convex optimization problem. Secondly, as noted in Hendrix and Klepper (2000), the spherical problem provides an upper bound of the Success Rate on the general case of elliptic level sets as its volume-surface ratio is at a minimum. Let us repeat, that we are particularly interested in what happens to the Success Rate when the dimension n increases. To facilitate this analysis, focus is on the infinite norm variant of the spherical problem, which for clarity sake will be called the *squarical* problem. Squarical functions are defined by the infinite norm, $f^n(x) = max_j \mid x_j \mid$; $j = 1, \ldots, n$. Squarical functions behave similarly to spherical functions and in addition the shape of the level sets is similar to the initial search space; they consist of a hyper-square.

Apart from the bounding perspective, why making use of such extreme cases in the analysis? The spherical, conical and squarical functions have in common an essential aspect in the derivation of results on efficiency for (N)PAS, namely that the improvement is stationary. Stationarity means here that the distribution of the improvement during the course of the algorithm does not change. During the iterations, generating a new point x_{k+1} randomly in $S(y_k)$ gives a function value y_{k+1}. For this constructed class of optimization problems y_{k+1} has a probability distribution given by $F(y_{k+1}) = (y_{k+1}/y_k)^n$ for $0 \leqslant y_{k+1} \leqslant y_k$. This means that the relative improvement y_{k+1}/y_k has the same distribution in every iteration; it does not depend on the level y_k which has been reached. For CRS, when the points are uniformly distributed over the level set, the Success Rate SR_k has a constant value SR and can be estimated by a converging value (steady state) of sr_k.

4.1.2. **The problem has several global minimum points**

We are interested in the relation between the Success Rate of the one-minimum point case and the case when K minimum points exist; to be precise, when the limit level set contains K compartments. As argued before, in limit for *secondary points* a failure (no improvement) may occur when the m points are taken from various compartments. As will be argued in the sequel, the success of generating primary points may profit from geometrical structure in the location of the optima. In the usual test functions in GO, some structure exists in the location of the optima. This structure gives a special effect, which we study here from one of

the easiest modifications of the spherical problem and call a bispherical problem. The essential part of this problem is that (for low function value) the level sets of the global minimum contain two separated equal sized compartments. Although the bispherical function can be defined in any dimension, we give here the two dimensional variant:

$$f(x) = \begin{cases} (x_1 - 1)^2 + x_2^2 & if \ \ x_1 \geqslant 0 \\ (x_1 + 1)^2 + x_2^2 & if \ \ x_1 < 0 \end{cases} \tag{9}$$

The feasible set is defined by the box $x_1 \in [-4, 4]$ and $x_2 \in [-1, 1]$. For this extreme instance, an analysis is possible on the ratio between the Success Rate of this function and that of the spherical problem, thus giving insight in what can happen with an increasing number of global minimum points.

4.2. TEST ALGORITHM

The Success Rate is analysed for Primary Trial Points distributed on the search space $X = [-1, 1]^n$ for the squarical problem. Empirically it means to work with a version of CRS (CRS$(f, X, N, m, \alpha = 0)$) where Secondary Points are not computed.

As formulated above, it is of specific interest to know how the Success Rate of CRS evolves and how it depends on parameters such as N, m, α or on its own dynamics (current level set). The limit Success Rate SR is estimated with the steady-state value of the parameter sr_k used in the algorithmic definition of CRS. If the algorithm evolves in such a way that from a certain iteration k, sr_k reaches a steady state, then sr_k can be seen as a measure of the probability that the next sample point is in the S_k-level set $S(y_k)$. This happens (leaving out boundary effects) for spherical and bispherical problems when a level of 1, so $y_k < 1$, is reached. For the squarical problem this is the case from the start of the algorithm. The analysis starts now by first considering the Success Rate of the algorithm on a one dimensional function with one optimum. This leads to an estimate for higher dimensional functions with one optimum. Finally we will use the bispherical problem to analyse the Success Rates for cases with several optima.

The idea is to start by analysing the success rate $SR(n = 1)$ of the one-dimensional squarical function ($f^1(x)$) first. Thinking in a uniform distribution on a hyper-square leads to the idea that the $m + 1$ points to run the algorithm are determined by a sample which is independent for every dimension. The consequence of this, is to be tempted to estimate the Success Rate for the n-dimensional squarical function as $SR(n) = (SR(n = 1))^n$. If this would be true, the exponential behavior confirms the intuition that it is not possible to construct a $\rho - adaptive$ search algorithm. However, this straightforward conclusion is not completely justified as we will argue; $(SR(n = 1))^n$ defines a lower bound on Success Rate of squarical problem $f^n(x)$ for CRS. The observed Success Rate will be bigger.

4.3. SQUARICAL FUNCTION (ONE OPTIMUM) ONE DIMENSIONAL

The Success Rate is analysed for Primary trial Points distributed on the search space $X = [-1, 1]$. Note that the squarical function is the same as the spherical in one dimension and is equivalent to the conical problem. In the algorithm $y_k = f(r^{max})$ is the worst function value in the sample and defines the current level set $[-y_k, y_k]$. The Success Rate to be found is the probability that the reflected point is in the current level set, $SR(n = 1) = P(-y_k \leqslant 2 \times \overline{G} - r_{m+1} \leqslant y_k)$. An assumption in the analysis is that all points to build the Primary point are taken from a uniform distribution over $[-y_k, y_k]$. This is correct for the first iteration and in the other iterations does not deviate very much, as was shown in Hendrix and Klepper (2000) and also follows from the empirical results. The result of the derivation which can be found in Appendix A, is that for $m = 2$, the Success Rate $SR(n = 1, m = 2)$ is given by:

$$SR(n = 1, m = 2) = 2 \int_0^1 (1 - \overline{G}) \times (1 - \overline{G}) \mathrm{d}\overline{G} = \frac{2}{3} \tag{10}$$

In general ($m \geqslant 2$), $\overline{G}$ can be approximated by a random variable normally distributed according to $N(0, \frac{1}{3m})$. This leads to (Appendix B):

$$SR(n = 1, m) \approx 2\Phi(\sqrt{3m}) - 1 - \frac{2}{\sqrt{3m}\sqrt{2\pi}}(1 - e^{-\frac{3m}{2}}) \tag{11}$$

in which Φ is the cumulative distribution function of the standard normal distribution. The result of equation (11) illustrates that when m increases, the primary points of CRS start behaving more and more as a local search; SR tends to 1. We measured the resemblance of analytical approximation (11) of the Success Rate with running the algorithm. In Table 1, an analytical estimate of the Success Rate (SR) based on equation (11) and an experimental estimate (SR_e) being the average over 100 runs of the values of sr_k at the last iteration k of CRS, can be found. It can be seen from Table 1 that experimental averages do not significantly differ from the analytical ones. The value for $m = 2$ shows that approximating the distribution of $\overline{G}$ with the normal distribution gives nearly the same SR estimate as the real one of 2/3.

As can be observed from the idea of (11), an increasing number m of points to build the centroid, leads to stronger centralisation (variance of $\overline{G}$ becomes smaller), such that the probability of a point to be reflected out of the level set becomes smaller. From this idea, one can derive the feedback rule to increase parameter m, whenever one suspects from observations (value information) that the population clusters around one (global) minimum point.

Table 1. Success rate: analytical estimate and average over 100 runs ($n = 1$).

	Analytical estimate	Experimental estimate
m	$SR(n = 1, m)$	Average $SR_e(n = 1, m)$
2	0.676176	0.674582
3	0.734293	0.739865
4	0.769709	0.774382
5	0.793993	0.801280
6	0.811937	0.818653
7	0.825887	0.832684
8	0.837132	0.842565
9	0.846447	0.852221
10	0.854327	0.860099

4.4. SQUARICAL FUNCTION (ONE OPTIMUM) HIGHER DIMENSIONS

In the analysis, first focus was on behaviour of the algorithm with varying m starting from the point of view of a one-dimensional squarical (is equal to spherical) function. There are various extensions possible to consider the behaviour in higher dimensions. In the use of the algorithm for the choice of the parameter m often either $m = 2$ is taken or $m = n$, in which n is the dimension of the function under consideration. For the latter case a straightforward extension would be to approximate the Success Rate by the underestimate $SR(n, m = n) = (SR(1, m = n))^n$. The confrontation with an empirical estimate based on the average of 100 runs is given in Table 2. Where the analytical underestimate decreases significantly with increasing dimension n, we observe that reality is doing much better; the Success Rate of CRS on the squarical problem diminishes less problematic with the dimension.

   The estimated lower bound could be improved slightly by the following observation. The N points in the population are not equally spread over the hyper-square defined by y_k, as the maximum element value in every dimension is different; y_k is the maximum over them. Therefore the population is spread over a hyperrectangular set enclosed by the current level set. Analysis on this hyperrectangular set would lead to a reasonable estimate of the Success Rate, but is not directly possible. Instead we focus on a larger set on the basis of extreme order statistics which at least gives an improved lower bound in Appendix C for the case $m = 2$. This leads to the following (lower) estimate:

$$SR(n, m = 2) = \frac{2}{3}\left(\frac{2}{3} + \frac{1}{2(N-1)}\right)^{n-1} \tag{12}$$

Table 2. Success rate: analytical underestimate and average over 100 runs, $n = m$.

	Analytical results	Experimental results
m	$SR(n, m = n) =$ $(SR(n = 1, m))^n$	Average $SR_e(n, m = n)$
2	0.457215	0.464163
3	0.395921	0.429452
4	0.351000	0.404189
5	0.315562	0.384363
6	0.286508	0.369596
7	0.216455	0.355846
8	0.141493	0.345798
9	0.082035	0.336230
10	0.042902	0.328356

Table 3. Success rate: analytical underestimate (SR) and experimental estimate (SR_e), $m = 2$.

	Analytical results	Experimental results
n	$SR(n, m = 2)$	$SR_e(n, m = 2)$
1	0.666667	0.674582
2	0.446119	0.464163
3	0.297785	0.321358
4	0.198647	0.223850
5	0.132418	0.156067
6	0.088342	0.109405
7	0.058905	0.076559
8	0.039276	0.053930
9	0.026186	0.038192
10	0.017459	0.027264

It is interesting to see that another parameter of the algorithm, namely N has been included in the estimate. Table 3 gives the confrontation of the new lower bound with experimental estimates based on an average of 100 runs and population size $N = 100 \times n$. The choice of a big value for N does not promote the efficiency. However, the choice of N is usually determined by effective covering considerations; one tries to find all global minimum points. It is nice to observe that reality is performing better than what can be derived analytically. Comparing the experimental results in Table 3 with those in Table 2, shows again the improving behaviour when

larger values for parameter m are used in cases where the population of points is condensed around one global minimum point.

4.5. TWO MINIMUM POINTS, THE BISPHERICAL PROBLEM

We consider the relation between the limit Success Rate of the one-minimum point case of which the spherical problem is a pure case and the limit Success Rate when there are several global minimum points and corresponding compartments of the limit level set. The bispherical problem is an example of that. We focus on the ratio $\theta = SR^{bisph}/SR^{sph}$, which we expect to be smaller than one and call $\theta(m)$ for a value of parameter m. Two effects are distinguished:

(1) The probability $\theta_c(m)$ that all $m + 1$ points appear in the same compartment, sphere of the level set, will be called the *compartment component* of θ.
(2) The geometric structure of the location of the minimum points may add a so-called *geometric component* $\theta_g(m)$. This means that the reflected point may have the possibility to end up in another compartment of the current level set than it is taken from.

4.5.1. **Compartment component**

The probability that all $m+1$ points are taken from the same compartment depends on the distribution of the N points (population) over the two spheres. Given that a fraction p of the points is in one set and a fraction $(1 - p)$ in the other, leads to the probability:

$$\theta_c(m) = p^{m+1} + (1 - p)^{m+1} \tag{13}$$

If the algorithm covers well, p should have a value of 0.5 and equation (13) reduces to

$$\theta_c(m) = \left(\frac{1}{2}\right)^m \tag{14}$$

Note that whenever the algorithm does not cover completely (p deviates from 0.5), $\theta_c(m)$ is higher. In extreme cases when p becomes zero (or alternatively one), $\theta_c(m)$ takes a value of one, because the algorithm solves a spherical problem.

4.5.2. **Geometric component**

For a point to have the possibility to be reflected into another compartment, it is necessary for the centroid to be near the middle line separating the compartment the reflection point r_{m+1} is taken from and the compartment it is reflected into. For the bispherical problem, the geometric component $\theta_g(m)$ is estimated with the probability that the centroid is in the middle of the two (tiny in limit) spheres as indicated with the second possibility in Figure 4. Notice that this probability only is nonzero when m is an even number. Given that a fraction p of the N points is in
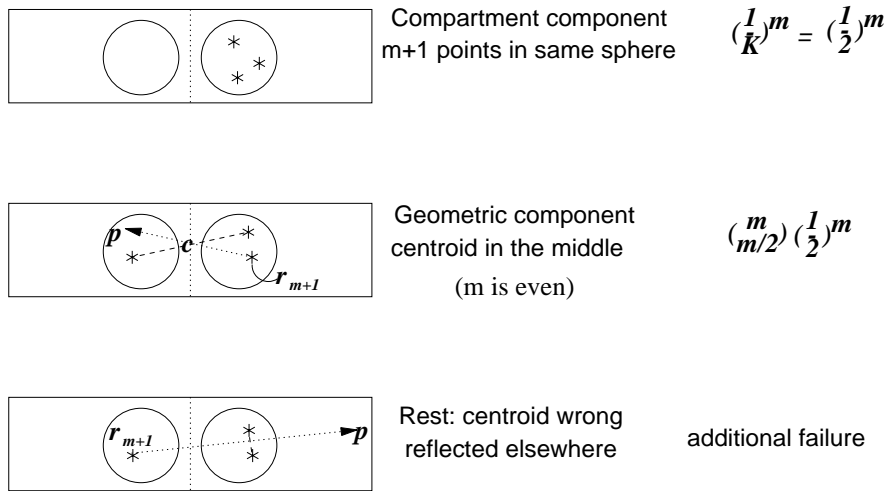
| | Compartment component m+1 points in same sphere | $(\frac{1}{K})^m = (\frac{1}{2})^m$ |

| | Geometric component centroid in the middle (m is even) | $\binom{m}{m/2}(\frac{1}{2})^m$ |

| | Rest: centroid wrong reflected elsewhere | additional failure |

Figure 4. Components in determining the Success Rate of the spherical versus the bispherical problem

one of the compartments, the probability that there are $\frac{m}{2}$ points as well as in the left sphere as in the right sphere (binomial distribution) is given by:

$$\theta_g(m) = \binom{m}{\frac{m}{2}} \times p^{\frac{m}{2}} \times (1-p)^{\frac{m}{2}} \tag{15}$$

For $p = 0.5$ this reduces to:

$$\theta_g(m) = \binom{m}{\frac{m}{2}} \times \left(\frac{1}{2}\right)^m \tag{16}$$

The idea of the two components of $\theta(m)$ is depicted in Figure 4 for $p = 0.5$.

We built an estimate of $\theta(m) = \theta_c(m) + \theta_g(m)$ based on Eqs. (14) and (16) and confronted this with empirical estimates based on average limit values for the spherical and bispherical problem. The geometric component may be hard to measure empirically, because it only can be observed clearly in the limit situations where the population forms two tiny clusters around the two global minimum points. Moreover, when for odd values of m the centroid is close to the middle, improvements are found in a non-limit situation such that the empirical estimate may deviate.

In Table 4.5.2 first a confrontation is given for $m = 2$ which leads to a fixed analytical ratio of $\theta(2) = 75\%$ for increasing dimension n. The empirical ratio is slightly bigger, which may be explained from deviations from equally covering. In Table 5 the value of m is varied and we can observe what happens with the Success Rate of spherical versus bispherical function when m is increased for a

Table 4. Empirical estimates of Success Rate $(n, m = 2)$ for bispherical (SR_e^{bisph}) and spherical (SR_e^{sph}) functions with corresponding experimental estimate of θ (E.E.).

n	SR_e^{bisph}	SR_e^{sph}	E.E.(%) of $\theta(2)$
2	0.363335	0.477999	76.01
3	0.262783	0.345604	76.04
4	0.191614	0.252941	75.75
5	0.141369	0.186863	75.65
6	0.105186	0.138794	75.79
7	0.078642	0.103492	75.99
8	0.059031	0.077592	76.08
9	0.044560	0.058359	76.35
10	0.033766	0.044173	76.44

two dimensional spherical and bispherical problem. Given the estimates in Table 4.5.2, there is no reason to expect the ratio to be different for $n > 2$. Two different types of deviations may be observed. For even values of m the measured values are slightly more optimistic than the analytical estimates. However, for odd values of m the gap is much bigger. Apparently the idea that the algorithm is effective and p has a value around 0.5, is not valid. We can check this for individual runs by following dynamically the attained values of sr_k. This is depicted in Figure

Table 5. Building an analytical estimate (A.E.) of $\theta(n = 2, m)$ and empirical estimates (E.E.).

m	θ_c	θ_g	A.E. (%)	SR_e^{bisph}	SR_e^{sph}	E.E. (%)
2	1/4	1/2	75.00	0.363445	0.478627	75.94
3	1/8	0	12.50	0.327768	0.577710	56.74
4	1/16	6/16	43.75	0.279986	0.636683	43.98
5	1/32	0	3.12	0.220582	0.678977	32.49
6	1/64	20/64	32.81	0.237113	0.707116	33.53
7	1/128	0	0.78	0.120877	0.730029	15.69
8	1/256	70/256	27.73	0.214815	0.747321	28.74
9	1/512	0	0.19	0.054925	0.762068	7.21
10	1/1024	252/1024	24.70	0.202490	0.773002	26.20

5. This figure shows the sharp contrast between m odd and m even. For odd values of m, after a while sr_k is increasing again which means that the algorithm is con-
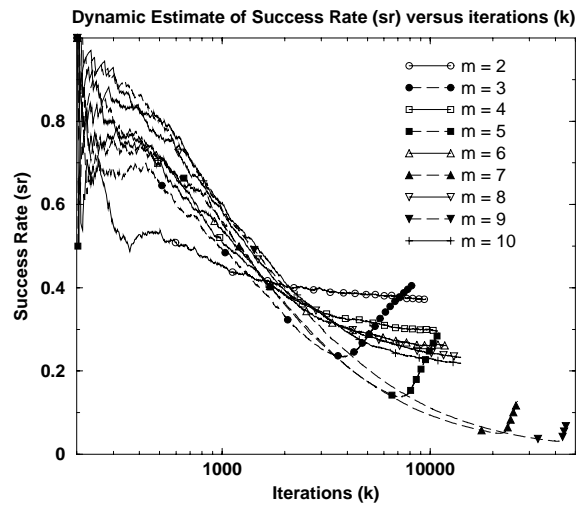
Figure 5. Dynamics of empirical estimate of Success Rate ($SR_e^{bisph}(n = 2, m)$): non-stationary for odd values of m.

verging to one of the global optimum points. The limit value of sr_k converges to that of the spherical problem. Notice again that the results in the figure represent one run, whereas the data in the table represent the average over 100 runs.

An increase in the parameter value m leads to improvements in the Success Rate for the spherical (one minimum point) case, due to intensification of the local search behaviour. We can see from Table 5, that when two minimum points appear, using a higher value for m does not reduce the Success Rate dramatically as long as it is an even number. The apocalyptic prediction of the compartment component disappears due to the additional probability that the centroid may end up in the middle.

5. Conclusions and discussion

In this paper a framework has been presented to investigate the behaviour of Global Optimization algorithms. Within this framework the algorithm of Controlled Random Search has been investigated and the following elements have been discussed:

- Criteria on effectiveness and efficiency. The relevance of the idea of Success Rate SR for speed of convergence (in limit) has been brought forward and a measurable estimator sr_k has been defined.
- Parameters of the algorithm which can be used to steer the algorithm are the population size N, the secondary point calculation determined by α and the parameter m which defines the number of points to build the centroid.

   — As characteristics of instances the number of global minimum points K and the dimension of the problem n have been distinguished.

The investigation lead to some qualitative statements and quantitative analytical results. Qualitatively the following can be said.

   — A value for the population size N is usually determined by effectiveness considerations; one tries to find all global optimum points. Although its value directly influences the number of function evaluations, it does not have a large impact on the final Success Rate

   — The value of parameter α (secondary points) can be made high whenever the population finds itself in one convex level set (one optimum point case). As a secondary point is a convex combination of the population, its function value is at least not worse than the function value determining the current level set. On the other hand, whenever the population is spread out over several clusters, secondary points tend to lead to failures (lowering the Success Rate). This means that it is worthwhile to look into possible value information consisting of indicators that measure the clustering behaviour.

   — Increasing parameter m leads to local search behaviour whenever the population clusters around one global minimum point.

To study the dependency of the Success Rate SR on the parameter m as well as dimension n, an attempt has been made to come to analytical estimates for cases where the Success Rate is constant, such that empirical estimates can be made.

   — For a one-dimensional function equivalent to the spherical, squarical and conical problem we obtained analytical estimates for varying m which match empirical estimates 'reasonably'. Qualitatively it supports the above mentioned local search behaviour that SR improves with increasing m.

   — Extending the analysis for increasing dimension n with the help of the squarical problem is not easy. The straightforward extension provides analytical underestimates that predict an exponentially bad behaviour. The underestimate shows an increasing (in n) gap with what can be observed experimentally. Empirically seen, the limit Success Rate when having reached the level set of one global minimum point, does not decrease as much as expected in the dimension n as long as the choice of m is adapted to n, e.g., $m = n$.

   — Estimates on how the Success Rate is reduced when the algorithm is confronted with two equally sized compartments of the level set, were made on the base of the bispherical versus spherical problem. Both quantitative results, analytical and empirical, matched quite fine for even values of parameter m. In contrary, for odd values of m, effectiveness considerations started playing a role; it appeared that for odd values of m the algorithm converges to only one of the global minimum points.

With respect to extending the analysis in the future for cases with more than two global minimum points the following remark can be made. The bilateral help between any pairs of compartments which was described by the geometric effect θ_g, does also appear in cases where three (or more) global minimum points exist. When three final clusters have been detected, using a value for m which is a multiple of three does not necessarily increase the geometric component. It depends on the geometric structure in the location of the optimal points, whether the reflection of a point r_{m+1} out of one of the compartments has any chance to end up in one of the other two compartments.

6. Acknowledgement

Appendix A. Success Rate squarical, m=2, n=1

The Success Rate is the probability that the reflected point is in the current level set, $SR(n = 1) = P(-y_k \leqslant 2 \times \overline{G} - r_{m+1} \leqslant y_k)$. The start of the analysis takes place from stretching the interval $[-y_k, y_k]$ and considering the points as taken from $[-1, 1]$, such that $SR(n = 1) = P(-1 \leqslant 2 \times \overline{G} - r_{m+1} \leqslant 1)$. In this probability two stochastic variables are involved, namely the centroid $\overline{G}$ and the reflection point r_{m+1}. First the conditional probability $\pi(\overline{G})$ of the Primary point being in the interval $[-1, 1]$ given the centroid $\overline{G}$ is considered:

$$\pi(\overline{G}) = P(2 \times \overline{G} - 1 \leqslant \mathbf{r}_{m+1} \leqslant 2 \times \overline{G} + 1) \tag{17}$$

Then we calculate the unconditional Success Rate by integrating over the probability distribution $F(\overline{G})$ of $\overline{G}$

$$SR = \int_{-1}^{1} \pi(\overline{G}) dF(\overline{G}) \tag{18}$$

The conditional probability given $\overline{G}$ that $\mathbf{r}_{m+1}$ is reflected into $[-1, 1]$ is

$$\pi(\overline{G}) = \begin{cases} \int_{2 \times \overline{G} - 1}^{1} \frac{1}{2} dx = 1 - \overline{G} & \text{if } \overline{G} \geqslant 0 \\ \int_{-1}^{2 \times \overline{G} + 1} \frac{1}{2} dx = 1 + \overline{G} & \text{if } \overline{G} < 0 \end{cases} \tag{19}$$

So the probability is highest when $\overline{G}$ is close to the middle of the interval. Now we should determine the distribution of $\overline{G}$ which is an average of m points from a uniform distribution over $[-1, 1]$. For $m = 2$ the probability density function $F'(\overline{G})$ corresponds to a triangular distribution and is given by:

$$F'(\overline{G}) = 1 - |\overline{G}| \tag{20}$$

So for $m = 2$, the Success Rate $SR(n = 1, m = 2)$ is given by (using symmetry around 0):

$$SR(n = 1, m = 2) = 2 \int_0^1 (1 - \overline{G}) \times (1 - \overline{G}) d\overline{G} = \frac{2}{3} \tag{21}$$

Appendix B. Success Rate squarical, m, $n=1$

In general ($m \geqslant 2$), $\overline{G}$ can be approximated by a random variable normally distributed according to $N(0, \frac{1}{3m})$. Using a transformation $s = \sqrt{3m} \times \overline{G}$, (so the variance of s is $V(s) = 3m \times V(\overline{G}) = 1$) towards the standard normal distribution $s \sim N(0, 1)$ with cumulative distribution function Φ and density function ϕ we obtain:

$$SR(n = 1, m) \approx 2 \times \int_0^1 (1 - \overline{G}) dF(\overline{G}) = \\ 2 \times \int_0^{\sqrt{3m}} (1 - \frac{s}{\sqrt{3m}}) d\Phi(s) \tag{22}$$

So,

$$SR(n = 1, m) \approx 2\Phi(\sqrt{3m}) - 1 - \frac{2}{\sqrt{3m}} \int_0^{\sqrt{3m}} s d\Phi(s) = \\ 2\Phi(\sqrt{3m}) - 1 - \frac{2}{\sqrt{3m}} \int_0^{\sqrt{3m}} s\phi(s) ds \tag{23}$$

where $\phi(s) = \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}}$. So,

$$SR(n = 1, m) \approx 2\Phi(\sqrt{3m}) - 1 - \frac{2}{\sqrt{3m}\sqrt{2\pi}} (1 - e^{-\frac{3m}{2}}) \tag{24}$$

Appendix C. Underestimate Success Rate squarical, $m=2$, n on the base of extreme order statistics

The current point set (population) $R = \{r_1, \dots, r_N\}$ defines the function value $y_k = max_{ij} \mid r_{ij} \mid$ over $i = 1, \dots, N$; $j = 1, \dots, n$. In every dimension j, a maximum coordinate value $z_j = max_i \mid r_{ij} \mid \leqslant y_k$ exists. For one coordinate l the absolute maximum is attained, $z_l = y_k$. In the other dimensions we assume z_j to be a maximum of N uniform points in $[0, y_k]$ and the points $\{r_{1,j}, \dots, r_{m+1,j}\}$ which determine the primary point to be taken uniformly from $[-z_j, z_j]$. To facilitate the notation, index j is left out in the sequel. Now we base the probability of the primary point to be reflected into $[-y_k, y_k]$ for dimensions other than l on these assumptions, which increases the lower bound. This is elaborated for $m = 2$ only.

Analogously to the derivation for the one-dimensional case, first the conditional probability of $2 \times \overline{G} - r_{m+1}$ being in the set $[-y_k, y_k]$ is determined given a value for z. Again without loss of generality set $[-y_k, y_k]$ is stretched to $[-1, 1]$. After

this, the conditional probability is integrated over the distribution of z. So we consider z to be the maximum of N points uniform on $[0, 1]$. The c.d.f. of z is given by

$$F_z(z) = z^N \tag{25}$$

Now $\overline{G}$ is the average of m points from $[-z, z]$ and r_{m+1} is uniform over $[-z, z]$. Analogous to equation (19) the probability of $2 \times \overline{G} - r_{m+1}$ to be in $[-1, 1]$ given the value of $\overline{G}$ is

$$\pi(\overline{G}) = \begin{cases} \int_{2\times\overline{G}-1}^{z} \frac{1}{2z}dx = \frac{1}{2} + \frac{1}{2z} - \frac{\overline{G}}{z} & \text{if } \overline{G} \geqslant 0 \\ \int_{-z}^{2\times\overline{G}+1} \frac{1}{2z}dx = \frac{1}{2} + \frac{1}{2z} + \frac{\overline{G}}{z} & \text{if } \overline{G} < 0 \end{cases} \tag{26}$$

The p.d.f. of $\overline{G}$ being the average of $m = 2$ uniform points from $[-z, z]$ is given by:

$$F'_{\overline{G}}(\overline{G}) = \frac{1}{z} \cdot \left(1 - \frac{|\overline{G}|}{z}\right) \quad, -z \leqslant \overline{G} \leqslant z \tag{27}$$

Integration over the distribution of $\overline{G}$ now results into a conditional Success Rate $P(-1 \leqslant 2 \times \overline{G} - r_{m+1} \leqslant 1)$ given z for $m = 2$, in the dimension $j \neq l$ of:

$$2 \int_0^z \frac{1}{z} \left(\frac{1}{2} + \frac{1}{2z} - \frac{\overline{G}}{z}\right) \times \left(1 - \frac{\overline{G}}{z}\right) d\overline{G} = \frac{1}{6} + \frac{1}{2z} \tag{28}$$

The final Success Rate in a dimension $j \neq l$ is determined by the integral over all possible values of z:

$$\int_0^1 \left(\frac{1}{6} + \frac{1}{2z}\right) dF_z(z) = \int_0^1 \left(\frac{1}{6} + \frac{1}{2z}\right) dz^N = \frac{2}{3} + \frac{1}{2(N-1)} \tag{29}$$

For big values of N, this does not deviate much from the value of $\frac{2}{3}$ which is found for dimension l which defines the level set $S(y_k)$. The final underestimate has been slightly improved towards

$$SR(n, m = 2) = \frac{2}{3} \left(\frac{2}{3} + \frac{1}{2(N-1)}\right)^{n-1} \tag{30}$$

References

W. P. Baritompa, R.H. Mladineo, G. R. Wood, Z. B. Zabinsky and Zhang Baoping (1995). Towards pure adaptive search. *Journal of Global Optimization* 7, 73–110.

W. P. Baritompa and M. A. Steel (1996). Bounds on absorption times of directionally biased random sequences. *Random Structures and Algorithms*, 9, 279–293.

C. G. E. Boender and H. E. Romeijn (1995). Stochastic methods, in *Handbook of Global Optimization* (Horst, R. and Pardalos, P.M. eds.), Kluwer, Dordrecht, 829–871.

D. W. Bulger. and G. R. Wood (1998). Hesitant adaptive search for global optimisation. *Mathematical Programming* 81, 89–102.

I. García, P. M. Ortigosa, L. G. Casado, G. T. Herman and S. Matej (1997). Multidimensional optimization in image reconstruction from projections. in *Developments in Global Optimization* (I. M. Bomze, T. Csendes, R. Horst and P. M. Pardalos Eds), Kluwer, Dordrecht, 289–300.

P. E. Gill, W. Murray and M. H. Wright (1981). *Practical Optimization*, Academic Press, London

E. M. T. Hendrix and O. Klepper (2000). On uniform covering, adaptive random search and raspberries. Journal of Global Optimization, 18, 143–163.

E. M. T. Hendrix (1998) Global Optimization at Work. PhD thesis, Wageningen Agricultural University.

R. Horst and P. M. Pardalos, eds. (1995). *Handbook of Global Optimization*, Kluwer, Dordrecht.

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

O. Klepper and D. I. Rouse (1991). A procedure to reduce parameter uncertainty for complex models by comparison with real system output illustrated on a potato growth model. *Agricultural systems*, 36, 375–395.

O. Klepper and E. M. T. Hendrix (1994). A method for robust calibration of ecological models under different types of uncertainty *Ecological Modelling*, 74, 161–182.

J. A. Nelder and R. Mead (1965). A Simplex method for function minimization. *The Computer Journal*, 8, 308–313.

P. M. Ortigosa, J. Balogh and I. García (1999). A Parallelized sequential random search global optimization algorithm. *Acta Cibernetica* 14 (2), 403–418.

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

W. L. Price (1978). A controlled random search procedure for global optimization. in *Towards Global Optimization 2*; L. C. W. Dixon and G. P. Szegö (eds), pp. 71–84, North Holland, Amsterdam.

W. L. Price (1979). A controlled random search procedure for global optimization. *The Computer Journal* 20, 367–370.

W. L. Price (1983). Global optimization algorithms by controlled random search. *Journal of Optimization Theory and Applications* 40, pp. 333–348.

A. Törn and A. Zilinskas (1989). *Global Optimization. Lecture Notes in Computer Science 350*. Springer, Berlin,

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