The Impact of Global Structure on Search

Monte Lunacek, Darrell Whitley, and Andrew Sutton

Colorado State University
Fort Collins, Colorado 80523, USA

Abstract. Population-based methods are often considered superior on multimodal functions because they tend to explore more of the fitness landscape before they converge. We show that the effectiveness of this strategy is highly dependent on a function's *global structure*. When the local optima are not structured in a predictable way, exploration can misguide search into sub-optimal regions. Limiting exploration can result in a better non-intuitive global search strategy.

Keywords: Funnel landscapes, test functions, exploration, dynamic populations.

Many artificial test functions have a "big valley" topology, where a decrease in fitness implies that, on average, search is getting closer to the global optimum. Although the search space is highly multi-modal, the local optima are structured such that there exists a global trend toward the best solution. Problems that exhibit this characteristic are sometimes referred to as *single-funnel* landscapes.

There are several real-world applications that do not have this simple structure. Wales [7] suggests that many optimization problems in computational biology are difficult because local optima often form in distinct, spatially separate clusters within the search space. Problems of this type have multiple funnels, resulting in a landscape that has a less predictable underlying global structure.

The way that global structure impacts evolutionary search is not well understood, in part, because many of the test functions used for evaluation have single-funnel landscapes. There are also a few test functions that have multiple funnels, but the number of funnels increases with dimensionality. This complexity makes it difficult to understand search behavior in high dimensions.

We have several objectives in this paper. First, we describe a method for creating landscapes that contain exactly two funnels, regardless of the problem size. Then, we empirically show that several *evolution algorithms* have an extremely low probability of success when the global optima is located in a proportionally smaller funnel. Finally, we demonstrate that limiting exploration can result in a performance gain.

1 Motivation and Background

The degree to which an algorithm will perform well on an application partly depends on how well the algorithm can deal with the features that make the problem difficult. Researchers within the computational chemistry community have started to pay attention to how global structure affects problem difficulty [4]. Much of their attention has

G. Rudolph et al. (Eds.): PPSN X, LNCS 5199, pp. 498–507, 2008.

been devoted to studying Lennard-Jones clusters, which are a class of configuration optimization problems where the goal is to find the spatial positions for a set of atoms that has the smallest potential energy.

The energy surface of the Lennard-Jones potential is highly multimodal, and the most difficult instances have a *double-funnel* landscape. Assuming that a search algorithm can escape local optima, the underlying global structure of a problem may have a greater impact on problem difficulty than the number of local optima [6].

The Rastrigin function is a classic single-funnel landscape. Kern *et al.* [3] point out that there are two potential strategies for solving this highly multimodal problem. The first is to exploit *separability*, which reduces its difficulty to N one-dimensional lines searches, where N is the number of parameters. The other strategy is to exploit the problems global structure; CMA-ES [2] and Basin-Hopping [8] avoid local optima by exploiting an underlying structure. The main question we are exploring in this paper is: how does this underlying structure impact evolutionary search?

2 Creating Double-Funnel Landscapes

The relative merit in any empirical study is limited by how well we understand the characteristics that make realistic parameter optimization problems difficult, and by our ability to embed these features into benchmark test functions. In this section, we describe two *double-funnel* test problems. First, we create a simple surface comprised of *only* two quadratic spheres. Then, we take this simple surface and add local optima to it. This creates a multi-funnel surface similar to Rastrigin's function.

2.1 The Double-Sphere

The landscape structure of our simple *double-sphere* test function is the minimum of two quadratic functions, where each sphere creates a single funnel in the search space. The placement of each sphere is critical because the barrier that divides them will be inconsequential if they are too close. We also want this barrier height to scale with dimensionality.

To address these concerns, we place each quadratic sphere along the positive diagonal of the search space, which is bounded on the interval $[-5, 5]^N$. The optimal sphere is located in the middle of the positive quadrant of the search space, at $\mu_1 = 2.5$ in each dimension. The sub-optimal sphere is centered at $\mu_2 = -2.5$ across all dimensions. The distance between each funnel increases proportionally with dimensionality, and this construction creates an underlying surface that is *globally* non-separable.

Lennard-Jones double-funnel problems are difficult when 1) the sub-optimal funnels is nearly as deep as the optima funnel, and 2) the basin of attraction to the optimal funnel is small. We simulate this by increasing the height of the sub-optimal funnel by a value of d. That way, the value of the optimal funnel is unchanged. In order to change the relative size of each funnel, we scaled the sub-optimal funnel by a constant factor, denoted s. This way, the optimal funnel retains it shape regardless of scaling, and therefore, has a more consistent level of difficulty. Multiplying the sub-optimal funnel by a number greater than one will create a more narrow sub-optimal funnel. The

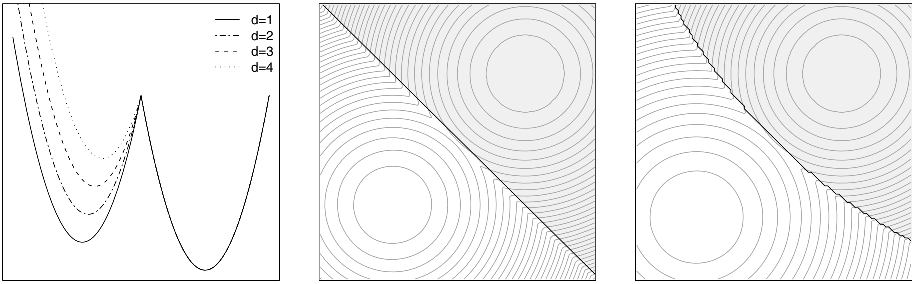

Fig. 1. The impact of d and s on the *double-sphere* function. Increasing d creates more distinction between the funnels (left). When $s = 0$ (middle), the two funnels are the same size. Decreasing s creates a larger sub-optimal funnel (right).

opposite is true when s is less than one. The overall form of our multi-funnel sphere function is:

$$f_{\text{double-sphere}}(\boldsymbol{x}) = \min\left(\sum_{i=1}^{N}(x_i - \mu_1)^2, \quad d \cdot N + s \cdot \sum_{i=1}^{N}(x_i - \mu_2)^2\right)$$

In order to make s the primary control characteristic for the size of each basin of attraction, we shifted the mean of the sub-optimal sphere such that the barrier between them, which is the point at which they intersect, is always located at the origin of the search space. This configuration requires $\mu_2 = -\sqrt{(\mu_1^2 - d)/s}$.

Values s and d control the size and depth of the sub-optimal funnel. The leftmost graph in Figure 1 is a diagonal slice showing how the different values of d impact the depth of the sub-optimal funnel. The middle and right-most contour plots illustrate the impact of s. The two funnels are the same size in the middle graph (e.g. $s = 1.0$), but the right-most graph creates a larger sub-optimal funnel (white) using $s = 0.7$. We use the quadratic penalty term described by Hansen and Kern [2] to enforce strong boundaries.

2.2 The Double-Rastrigin

We wanted a double-funnel test problem with properties similar to Rastrigin's function because it would isolate global structure as the main difference impacting problem difficulty on a problem that is well-understood. We create a double-funnel version of Rastrigin's function by adding local optima to the *double-sphere* function. We translate the cosine term used in Rastrigin's function by μ_1 so that the minimum of the local optima component is centered at the bottom of the optimal funnel. The overall form of the *double-Rastrigin* function is

$$f_{\text{double-Rastrigin}}(\boldsymbol{x}) = f_{\text{double-Sphere}}(\boldsymbol{x}) + 10\sum_{i=1}^{N}(1 - \cos 2\pi(x_i - \mu_1))$$

3 Understanding the Impact of Global Structure

In this section, we explore how the characteristics of the *double-sphere*, which we measure in terms of s and d, impact search. We compare a simple evolution strategy using *Cumulative Step-length Adaptation* (CSA-ES) [5,3], the *Covariance Matrix Adaptation Evolution Strategy* (CMA-ES)[2], and the CHC genetic algorithm [1]. Please see citations for descriptions and parameter settings.

We measure performance in terms of *success rate*, which we denote as ω, and define as the probability that an algorithm will converge to the global optimum. In each experiment, we estimate ω by running 1000 trials of each algorithm and counting the number of instances that find the global optimum.

Our results show that population-based methods are vulnerable to the size of each funnel, as controlled by s, when the depth of the two funnels are relatively close. That is, there exists some funnel characteristics where the exploration process will misguide search into the biggest funnel, not the deepest.

This section is organized in the following way. First, we measure the performance of local search in order to get a rough estimate of the size of each basin of attraction over a range of s and d values. Then we investigate how CMA-ES and CHC perform on the double-sphere. Although this problem only has two local optima, we still find both algorithms can fail even when the size of the optimal basin of attraction is fairly large. Finally, we discuss why this is important from a global optimization perspective by evaluating CMA-ES and CSA-ES on the double-Rastrigin function.

3.1 Local Search Properties of the Double-Sphere

As a baseline, we use the success rate of a local search method, where the probability of finding the global solution is proportional to the size of the basin of attraction to the optimum. We start by considering the double-sphere with dimension $N = 30$. We vary s between $[0.2, 1.4]$ by increments of 0.1, and evaluate different sub-optimal depths of $d = 1, 2$, and 3.

When we estimate $\hat{\omega}$ for local search, we find a positive and approximately linear relationship between $\hat{\omega}_{\text{LS}}$ and s. That is, as we decrease s, we also decrease the probability of finding the global optimum using local search. This makes sense because a small s value increases the size of the sub-optimal funnel, making the optimal funnel proportionally smaller (e.g. the basin of attraction to the optimum is smaller). The left graph in Figure 2 shows the relationship between $\hat{\omega}_{\text{LS}}$ and s. Notice that when $s = 1$, each funnel occupies $\approx 50\%$ of the search space (black dot).

We use this estimate of the size of each basin of attraction as a baseline for interpreting our results. That is, instead of graphing $\hat{\omega}$ for each algorithm as a function of s, we plot the $\hat{\omega}$ values as a function of $\hat{\omega}_{\text{LS}}$, the estimate size of basin of attraction to the global optimum. This makes it easier to observe when the evolutionary search is under- or over-performing with respect to what we would expect from local search.

For example, Figure 2 also shows the success rates of CMA-ES using the default population size of $\lambda = 14$ (for $N = 30$). Since CMA-ES is always above the gray line, we can observe that the success rates for CMA-ES are greater than that of local search. However, there is still a strong *linear* relationship between $\hat{\omega}$ and the size of the optimal funnel.

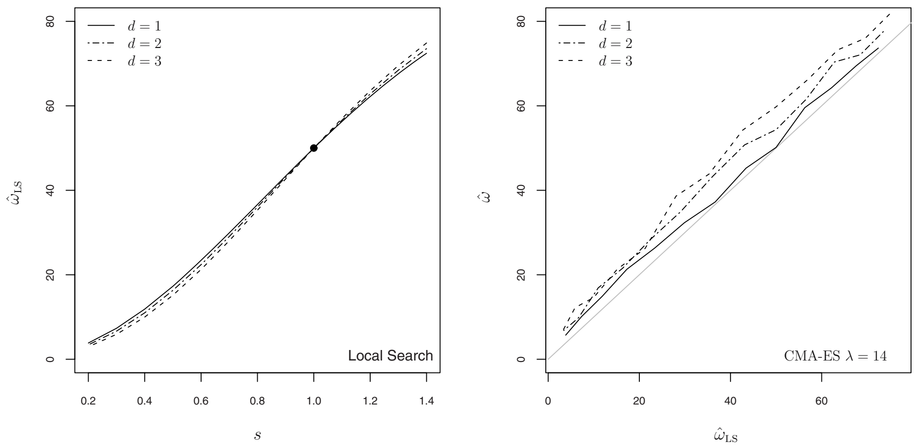

Fig. 2. Local search on the double-sphere: There is an approximately linear relationship between the size of the optimal funnel and success rate of local search, $\hat{\omega}_{LS}$ (left). Notice that the depth of the sub-optimal funnel does not greatly impact $\hat{\omega}_{LS}$. The right plot shows success rate for CMA-ES using the default population size. The success rates for CMA-ES are greater than that of local search, but still strongly tied to the size of the optimal funnel ($\approx \hat{\omega}_{LS}$).

3.2 Global Search Properties of the Double-Sphere

Most evolutionary algorithms perform better on multimodal surfaces when they use a larger population size. This is especially true for CMA-ES [2,3] and CSA-ES [3]. CHC is probably the exception, as it was designed to use smaller populations [9].

In this section, we would like to understand how the double-sphere impacts global search. In the next section we will consider a range of population sizes for CSA-ES and CMA-ES, but for now, we fix the population size of CMA-ES to $\lambda = 500$. For CHC we use the default of population size of 50 with 10-bits of precision. Our results did not change dramatically with increased population sizes for CMA-ES or CHC. Changing the precision on CHC to $20-$bits also had little impact. A maximum of $100,000$ evaluations were allocated and *no random restarts* were used (expect for the soft-restarts used by CHC). We discuss the role of restarts in the next section.

We observe a similar probability distribution for each algorithm. Instead of a linear trend, as observed for the local search methods, the distribution is pulled into a sigmoid. When the optimal funnel is proportionally larger than the sub-optimal funnel, success rates are extremely high. However, when the optimal funnel is proportionally smaller, the success rates for CMA-ES and CHC drop dramatically. Figure 3 show the probability of success for CMA-ES and CHC as a function of the basin of attraction size.

Consider the extreme cases. When the relative size of the optimal funnel is $\approx 70\%$, evolutionary search is highly successful ($\hat{\omega} \approx 100\%$). This means that when local search finds the optimal solution $\approx 70\%$ of the time, CMA-ES and CHC will almost always find the optimal solution. On the other hand, when the relative size of the optimal

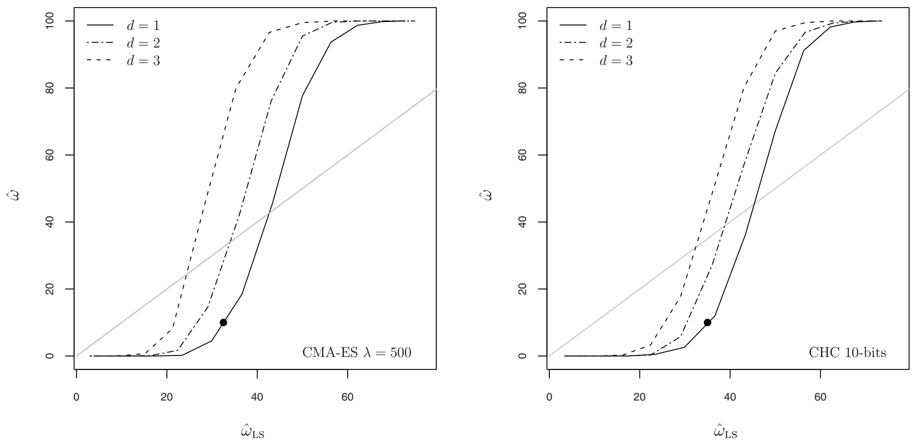

Fig. 3. CMA-ES and CHC on the double-sphere: The probability of success as a function of the size of the basin of attraction to the optimal funnel, as estimated with local search ($\hat{\omega}_{\mathrm{LS}}$). The gray line indicates the success probability of local search. For each algorithm, the trend is similar; when the optimal funnel is relatively large, the success rates for evolutionary algorithms are high. When the relative size of the optimal funnel is low, evolutionary search is more likely to fail.

funnel is only $\approx 10\%$, CMA-ES and CHC fail to find the global optimum. This is true for all the d values we considered.

As we increase d, we increase the height of the sub-optimal funnel. Figure 3 show that larger values of d shift the $\hat{\omega}$ distribution to the left, meaning that a smaller s value, and therefore, a smaller basin of attraction to the global optimal, is required to observe failure. The hardest problems for CHC and CMA-ES are those where the depths of the two funnels are close ($d = 1$) and the basin of attraction to the optimal funnel is comparatively small (s is small).

The black dots in Figure 3 represent a success rate of 10% for each algorithm when $d = 1$. This means that CMA-ES will succeed less than 10% of the time even when the relative size of the optimal funnel is $\approx 33\%$. A similar problem occurs with CHC. Even when the basin of attraction to the global optima is $\approx 35\%$, the success rate for CHC is about 10%. In general, when local search finds the optimal solution $\approx 1/3$ of the time, the evolutionary algorithms we tested are likely to fail when the depths of the two funnels are relatively close.

The key observation we make in this section is this: when the depths of two funnels are close (e.g. $d = 1$, about 17% different from the barrier that divides them), the global search parameter settings employed by the evolutionary algorithms we tested are likely to cause failure, even when the optimal basin of attraction is relatively large, $\approx 30\%$. As we increase the depth of the sub-optimal funnel, evolutionary search is more successful.

3.3 Implications for Global Search: Double-Rastrigin

Considering that CMA-ES using the default population size has probabilities of success that are similar to, or even better than, that of local search, why should we care about the

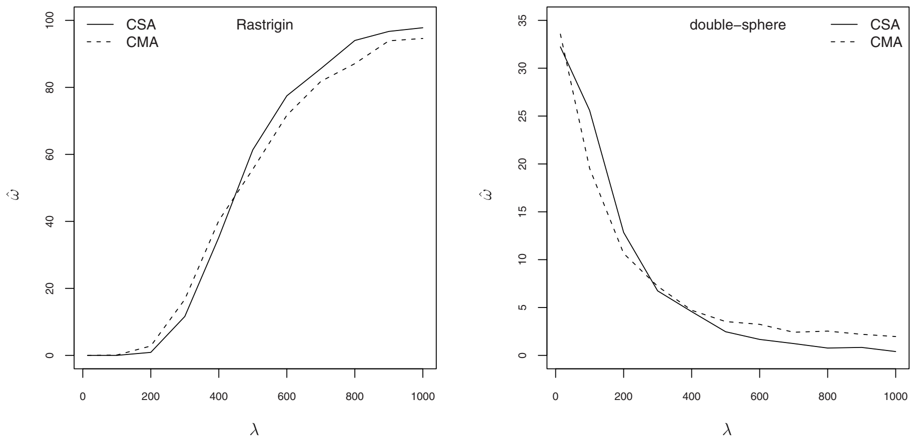

Fig. 4. Increasing the population size (λ) increases the probability that each evolution strategy will find the optimal solution on Rastrigin's' function (left), but decreases the probability of success on the double-Sphere function

bias of larger populations? The main reason this matters is that if an algorithm cannot cope with the simple structure of the double-sphere, it will also not be successful on more complex multimodal surfaces, like the double-Rastrigin, where the double-sphere dictates the underlying global structure.

We consider three 30-dimensional functions: Rastrigin, double-sphere, and double-Rastrigin. For the double-sphere and the double-Rastrigin, we created instances that are intentionally difficult for CMA-ES by choosing $d = 1$ and $s = 0.7$, which corresponds to an $\hat{\omega}_{\mathrm{LS}} \approx 30\%$. We only consider CSA-ES and CMA-ES because they have strong termination criteria and can solve the 30-dimensional Rastrigin function with large populations. This simplifies the interpretation of our results.

The leftmost graph in Figure 4 shows the estimated success rates for the ES algorithms, *without restarts*, on Rastrigin's function as the population varies from $[100, 1000]$ by increments of 100. We have also included the default population size of $\lambda = 14$. These results are consistent with previously reported success rates [2,3]. The noticeable trend is that larger populations are more able to exploit the underlying sphere structure of the Rastrigin function and locate the best solution. Smaller population sizes tend to get stuck in one of the many local optima. For example, CMA-ES with a population of $\lambda = 14$ never finds the global solution. Using a population size of $\lambda = 100$, CMA-ES only finds the optimal about 10 out of 1000 times.

As we vary the population size for the ES algorithms on the double-sphere, we find the opposite is true. High success rates are realized with low population sizes, but larger values of λ cause CSA-ES and CMA-ES to exhibit extremely low success rates. The right-most graph in Figure 4 shows these results.

This presents an interesting trade-off for the double-Rastrigin function: find a population size that balances the difficult characteristics of both the *modality* of the Rastrigin function and the *structure* of the double-sphere. Unfortunately, this balance is disappointing. When we run both algorithms on the double-Rastrigin function, we find that the success rates are lower than 3%, regardless of population size. This is because the success rates for the double-Rastrigin function can be decomposed into the success rates of its components. That is, the probability that an algorithm will be successful on the double-Rastrigin is approximately the joint probability that it is successful on the Rastrigin function *and* the probability that it will succeed on the double-sphere.

This is also an incomplete picture because the results presented so far have not used *random restarts*. From a practical point of view, restarts can increase performance because the success probabilities will add. That is, each restart represents an independent event. So, we are not just forced to find a population that balances the characteristics of both the Rastrigin and double-sphere function, we also need to account for the general observation that smaller populations will use fewer evaluations and restart more often.

When we include restarts and allow each algorithm to use $1e7$ evaluations, we still observe low success rates. For example, CSA-ES peaks at $\hat{\omega} \approx 11\%$ with a population of $\lambda = 400$. CMA-ES operating with $\lambda = 300$ yields an expected best of $\hat{\omega} \approx 5\%$.

The results of this section reinforce the notion that an algorithm's success or failure largely depends on its ability to cope with the features of a function. A population size suitable for Rastrigin's function is a poor choice for the double-sphere and vise-versa.

4 Limiting Exploration with Dynamic Populations in CSA-ES

On the double-sphere function, larger populations in CSA-ES (and CMA-ES) tend to pull the mean towards the funnel with the most samples. When the funnels are close in depth, a larger sub-optimal funnel is more likely to have more samples. Smaller populations are less vulnerable to this because less information being sampled. On the double-Rastrigin, we need the best of both worlds: a small population size to drop into a funnel without being pulled towards a larger basin of attraction, and then a large population size to exploit the underlying structure of that particular funnel.

As a proof of concept, we implemented CSA-ES with a dynamic population size that increased as the global step-size decreased. A decrease in step-size indicates a higher level of exploitation. When search is first exploring, it is utilizing a small population size. As it begins to exploit a promising region, increasing the population size will help exploit the underlying funnel structure. The algorithm is identical to CSA-ES in every way except at the end of each generation, we compute a new population size based on a function of the global step-size σ, the initial step-size σ_0, and an upper bound of the population size, λ_{MAX}.

$$\lambda = \lambda_{\mathrm{MAX}} \left(\frac{\sigma}{\sigma_0} - 1 \right)^2$$

We ensure that λ never falls below the default population size, $\lambda_d = 14$, or exceeds the maximum λ_{MAX}, which is an input parameter.

We ran this strategy, which we denote D-CSA-ES, on the 30-dimensional Rastrigin, double-sphere, and double-Rastrigin functions for the same values of λ used in the previous section, except that D-CSA-ES interprets this value as λ_{MAX}. The resulting search strategy is less effective on the Rastrigin function, but operates at a consistent

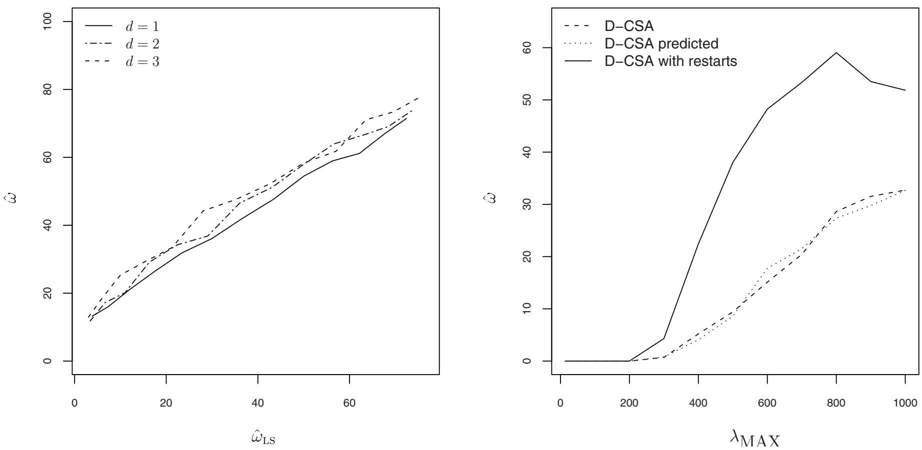

Fig. 5. D-CSA-ES on the double-sphere (left) and on the double-Rastrigin (right). The relationship between success rate and the size of the optimal funnel remains linear. This results in a much higher success rate on the double-Rastrigin function.

level on the double-sphere function that is proportional to the size of the optimal funnel, regardless of the population size. The left graph in Figure 5 shows D-CSA-ES on the double-sphere as a function of optimal funnel size for $d = 1, 2$, and 3 using $\lambda_{\text{MAX}} = 500$. The most striking feature is the approximately linear relationship between the size of the optimal funnel and the success rate of D-CSA-ES. This resembles the relationship of CMA-ES using a default population size on the double-sphere, but with a setting for λ that is more appropriate for global optimization.

What does this mean for the double-Rastrigin function? The right graph in Figure 5 show D-CSA-ES on the double-Rastrigin function for $s = 0.7$ and $d = 1$ as a function of population size. Without restarts (dash), D-CSA-ES has a success rate the is about 10 times higher than either CMA-ES or CSA-ES. When D-CSA-ES runs with restarts (solid line) until $1e7$ evaluations, it success rates are as high as $\approx 60\%$.

The dotted line in this graph represents the predicted performance obtained by multiplying the $\hat{\omega}$ from Rastrigin with $\hat{\omega}$ from the double-sphere. The prediction is very close to the empirical results and reinforces the notion that successful search must cope with both modality and global structure.

5 Summary

Global structure can clearly impact the performance of evolutionary optimization. When the optimal funnel is proportionally smaller, the success rates for CHC and CMA-ES decrease dramatically on the double-sphere, especially when the depths of the two funnels are close. Exploration is not able to distinguish between funnel quality, and is pulled into the larger funnel. We believe these results generalize to other algorithms.

This presents a problem for CMA-ES and CSA-ES on the double-Rastrigin function because, although larger population sizes are necessary to exploit the underlying structure of the Rastrigin, they are also more bias towards funnel size. The population size that is best for Rastrigin is the least effective on the double-sphere. A compromise that works on both is disappointing.

By dynamically adapting the population size, D-CSA-ES is less biased toward funnel size while exploring the search space. However, as it descends into a particular funnel, and it begins to exploit the search space, increasing the population size aids D-CSA-ES in detecting the underlying structure of the funnel and avoiding local optima. This results in a strategy whose success rate is dependent on funnel size; when the optimal funnel is large, the success rates for D-CSA-ES are not a good as CHC or CMA-ES. But when the optimal funnel is small, D-CSA-ES will still find the global solution with a probability proportional to relative funnel size. The highs are not as high, but the lows are still acceptable.

Exploring the search space to gain a global perspective before exploiting a particular region may be an effective strategy for “big valley”, single-funnel problems. But on multi-funnel landscapes, the effectiveness of exploration comes into question as a global search strategy. This work supports an ongoing awareness that, if an algorithm is going to be successful, then it must be able to deal with the features in the landscape.

Acknowledgments. This research was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number FA9550-07-1-0403. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

References

1. Eshelman, L.J.: The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional Genetic Recombination. In: FOGA (1991)
2. Hansen, N., Kern, S.: Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In: PPSN 2004. Springer, Heidelberg (2004)
3. Kern, S., Muller, S., Hansen, N., Buche, D., Ocenasek, J., Koumoustakos, P.: Learning Probability Distributions in Continous Evolutinary Algorithms—a Comparative Review. Natural Computing 3, 77–112 (2004)
4. Leary, R.H.: Global Optimization on Funneling Landscapes. Journal of Global Optimization 18 (2000)
5. Ostermeier, A., Gawelczyk, A., Hansen, N.: Step-Size Adaptation Based on Non-local use of Selection Information. In: PPSN 1994, pp. 189–198. Springer, Heidelberg (1994)
6. Pardalos, P.M., Schoen, F.: Recent Advances and Trends in Global Optimization: Deterministic and Stochastic Methods. In: CAPD (2004)
7. Wales, D.J.: Energy Landscapes and Properties of Biomolecules. Physical Biology (2005)
8. Wales, D.J., Doye, J.P.: Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. Journal of Chemical Physics 101(28) (April 1997)
9. Whitley, D., Beveridge, R., Graves, C., Mathias, K.: Test Driving Three 1995 Genetic Algorithms: New Test Functions and Geometric Matching. Journal of Heuristics (1995)