Heuristics for Sampling Repetitions in Noisy Landscapes with Fitness Caching

Forrest Stonedahl
Electrical Eng. and Computer Science Dept.
Northwestern University
Evanston, Illinois, USA
forrest@northwestern.edu

Susa H. Stonedahl
Dept. of Civil and Environmental Engineering
Northwestern University
Evanston, Illinois, USA
s-stone2@northwestern.edu

ABSTRACT

For many large-scale combinatorial search/optimization problems, meta-heuristic algorithms face noisy objective functions, coupled with computationally expensive evaluation times. In this work, we consider the interaction between the technique of "fitness caching" and the straightforward noise reduction approach of "fitness averaging" by repeated sampling. Fitness caching changes how noise affects a fitness landscapes, as noisy values become frozen in the cache. Assuming the use of fitness caching, we seek to develop heuristic methods for predicting the optimal number of sampling replications for fitness averaging. We derive two analytic measures for quantifying the effects of noise on a cached fitness landscape (probabilities of creating "false switches" and "false optima"). We empirically confirm that these measures correlate well with observed probabilities on a set of four well-known test-bed functions (sphere, Rosenbrock, Rastrigin, Schwefel). We also present results from a preliminary experimental study on these landscapes, investigating four possible heuristic approaches for predicting the optimal sampling, using a random-mutation hill-climber with fitness caching.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*global optimization*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*heuristic methods*

General Terms

Algorithms, Measurement, Experimentation

Keywords

fitness caching, noise reduction, fitness landscapes, uncertainty, sampling, hill climbing, evolutionary algorithms

GECCO'10, July 7–11, 2010, Portland, Oregon, USA.

1. MOTIVATION

There are a number of problem features that universally pose challenges for all metaheuristic search/optimization processes: predominant among these are noise/uncertainty, and the slowness of fitness evaluation (i.e., the time necessary to evaluate the objective function for any point in the search space). The presence of noise in a fitness function impedes making accurate comparisons between candidate solutions, or knowing how close the search process is to reaching a certain performance objective. In many cases, it is possible to use an average of many independent fitness function evaluations in order to reduce the noise. The length of time required for a single fitness evaluation can be significant, as it expands the length of the search by a direct multiplicative factor, and limits the number of evaluations possible for the search. Sometimes it is possible to use a less accurate surrogate fitness function, which can be evaluated more quickly, but at the cost of additional noise in the fitness estimates (for a survey of fitness approximation, refer to [7]). In general, it is impossible to eliminate both of these problem features, although there are many problems where trade-offs can be made between the two.

When fitness evaluation is particularly computationally expensive (e.g., in large complex simulations), it is sometimes attractive to cache fitness values for re-use, to save the cost of re-evaluating them again later. At least in some non-noisy optimization problems, this has been shown to be an effective approach for reducing total computational cost [11, 12], and we believe there is potential for applying it to noisy search spaces as well. In this work, we apply a combination of formal and empirical methods to try to investigate the relationship between fitness caching and the noise reduction technique *fitness averaging by repeated sampling.* In noisy environments, too little sampling can make the search untenable, whereas too much sampling can be unacceptably slow. Somewhere in between, there exists an ideal number of sampling repetitions, or "sweet spot", where the search most efficiently reaches a desired fitness level. Assuming the use of fitness caching, and using only information that can be extracted from the fitness landscape with reasonable efficiency, we would like to be able to predict where this "sweet spot" will fall.

The basic intuition motivating this research is that some landscapes are much more sensitive to the effects of noise than others, with regard to movement through these landscapes. For instance, a landscape that contains large steep mountains may be easily traversed to values of high fitness, despite the presence of significant noise, whereas even a small amount of noise may cause a landscape comprised of gentle slopes to become unnavigable. It would be very useful to have an efficient method of assessing the robustness of a landscape with respect to noise, in order to choose an appropriate sampling rate when applying a meta-heuristic search technique to the problem. The current study investigates the correlation between the distribution of fitness gradients throughout the landscape and the deleterious effects of varying levels of noise on landscape traversal.

The paper begins by situating the present work in the context of related research in the field. We then propose two measures to quantify the impact of noise on search processes within fitness landscapes: the probability that noise generates false local optima in the landscape, and the probability that noise will result in an incorrect choice when comparing two neighboring locations in the space. We offer mathematical expressions for these two measures, which are numerically confirmed by Monte Carlo simulations of the two respective probabilities, on a set of four well-known test-bed functions (sphere, Rosenbrock, Rastrigin, Schwefel). Next, we discuss how these measures could be used in heuristics for choosing an optimal sampling number for noise reduction. We then present the results of an experimental study where we empirically determine optimal sampling rates on the four test landscapes, given a straightforward local search technique (stochastic hill climber) that uses fitness caching, and compare the potential of four heuristic approaches to predict the "sweet spot" for noise reduction. We conclude by presenting several avenues for possible future work.

2. RELATED WORK

The beneficial effects of fitness caching (specifically for genetic algorithms) have been discussed by Kratica [11], and also applied to a practical problem (plant location) in [12]. In [12], the authors note that one of the conditions for successfully applying fitness caching is a large evaluation time for the fitness function. One real-world example where fitness caching may be beneficial is the the optimization of simulation parameters, since complex simulations may require long running times. However, another aspect of many real-world optimization problems is the presence of noise or uncertainty. For example, a recent instance of fitness caching [18] used two simple meta-heuristic search algorithms (hill-climbing and genetic algorithms) to explore the parameter-space of several agent-based simulations of biologically-inspired flock formation. In this case, the multi-agent simulations were stochastic, resulting in noisy fitness evaluation [18]; however, the interaction of fitness caching with the noise was not explored. Moreover, while there may be a potential benefit for fitness caching, even in noisy environments, we are unaware of prior work discussing the use of fitness caching in noisy/uncertain optimization problems, or examining the potential repercussions for search performance in detail.

Considerable research has been done in the general area of meta-heuristic search and optimization in noisy fitness landscapes, and it remains a topic of considerable interest. For example, recent work spans from developing efficient techniques of determining the best individual from a noisy population [6], to defining standard sets of noisy functions for benchmarking different optimization techniques [4]. The breadth of work in this area is beyond the scope of this paper; for a comprehensive survey of noise/uncertainty in evolutionary algorithms, see [8].

It is worth highlighting some of the more closely-related research. One strand of research concerns the analysis of search spaces or fitness landscapes, such as the study of Kauffman's NK-landscapes [9, 10], similarly inspired tunable landscapes [16], as well as search performance on such landscapes (e.g., [14]). Particularly relevant is the work on adaptive walks through noisy fitness landscapes [13]. Our work also pertains to adaptive walks (or local neighborhood-based search algorithms in general) in noisy landscapes, but with fitness caching the noise becomes frozen, as we will discuss later. Also, as our application interests are focused more on simulation optimization rather than understanding of biological evolutionary processes, we chose to investigate landscapes based on optimization benchmarks (see Section 2 below). So-called "fitness evolvability portraits" [17] appear to be another promising direction for fitness landscape analysis. While [17] currently does not address issues of noise, several of the ideas about characterizing the landscape at different fitness levels might be productively incorporated into future work on the sampling with fitness caching problem we are addressing here.

Several prior works ([2], and more recently [1]) have discussed/debated the relative merits of repeated sampling for noise reduction versus alternative methods, such increasing population size. However, when fitness caching is used, separate individuals in a population-based search do not contribute independent fitness trials, so increasing the population offers no advantages in reducing the impact of noise. Rana et al. [15] examine the effects of noise on search landscapes, in particular discussing the creation of *false local optima* and the soft annealing of peaks (or "melting" effect, as referred to by Levitan and Kauffman [13]). Our current work is also interested in the creation of false local optima by noise, but the use of fitness caching changes both the character and consequences of such local optima (as we discuss in Section 3.1).

Our work also follows that of Hughes [5], which derived analytic expressions quantifying the probability of one individual having a higher true fitness than another, given noisy fitness evaluation, in the context of both single and multi-objective evolutionary algorithms. Though several of the derivations are mathematically related, the measures we derive attempt to characterize the fitness landscape as a whole, rather than a single comparison.

In conclusion, to our knowledge, this paper is the first to discuss and analyze the effect of fitness caching in noisy fitness landscapes, and to attempt to develop (preliminary) heuristics for choosing the number of sampling repetitions in this case.

3. THEORETICAL ANALYSIS

We will begin from a theoretical perspective, offering a formal description of the problem, and deriving several mathematical measures that may be useful, before we move on to more experimental methods.

In this paper, we will assume the presence of additive Gaussian (normally distributed) noise with mean 0. The situation we are concerned with is the repeated sampling of a noisy fitness function, and as a result of the Central Limit Theorem, the shape of the noise distribution will always approach a normal distribution when a reasonably large number of samples is used. However, the mathematical derivations we present below could equally be applied to other noise distributions, although the resulting expressions may be symbolically and/or computationally cumbersome. If the mean value of the noise is unknown (and nonzero), then regardless of any approach, it impossible to determine the true expected value of the fitness landscape at any point; thus we will only consider unbiased noise with zero mean. We will also assume that the variance of the additive noise is uniform across the search space – while this is not always the case, it serves as a reasonable first-order approximation to simplify the analysis. The extension of considering noise with location-dependent variance is left as future work.

We will also make the simplifying assumptions that there is ample memory such that all encountered fitness values will be cached and are never cleared, and that the computation time required for the caching is negligible compared to the time required for fitness evaluation. These assumptions are realistic when fitness evaluation is particularly time-consuming, such as when optimizing complex simulations with lengthy run-times. In this case, high-capacity disk-based caching becomes a feasible approach, when the disk-access time for reading a cached fitness value may be orders of magnitudes smaller than the run-time of the simulation.

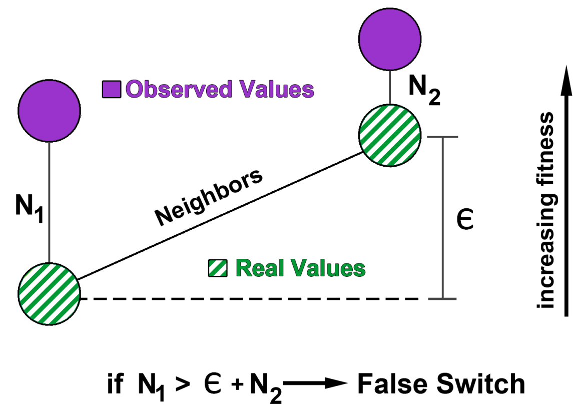

Figure 1: This figure illustrates variables used to determine the existence of a false switch. N_1 and N_2 represent the added noise to the original nodes, and ϵ represents the vertical distance between the two original neighbors. False switches occur whenever N_1 is greater than $\epsilon + N_2$.

3.1 Derivation of Measures

Let us consider a "true" (noiseless) landscape function L which has been obscured by some amount of additive noise (N), which is drawn from a normal distribution with mean 0 and standard deviation of σ ($N \sim \mathcal{N}(0, \sigma^2)$).[1] We will assume the neighborhood-based search, where the task is minimization (find x s.t. $L(x)$ is a minimum). Without fitness caching, each time a search algorithm evaluates a point x_1 in the search space S ($x_1 \in S$), a new fitness value $L(x_1)+N$ is returned, where N is independently drawn from $\mathcal{N}(0, \sigma^2)$. Let x_2 be a neighbor of x_1, such that $L(x_2)$ is greater than $L(x_1)$ by a positive amount ϵ ($L(x_2) = L(x_1) + \epsilon$). This means that if the search process was repeatedly choosing whether to move between x_1 and x_2, it would (probabilistically) end up at x_1. With fitness caching, this is not the case. Once fitnesses for x_2 and x_1 have been chosen, they are fixed, or *frozen*. This caching is effectively the same as reading values from a new "frozen" noisy landscape L_n, which is generated from L by adding N ($N \sim \mathcal{N}(0, \sigma^2)$ to every location in X. If the fitness value $L_n(x_2)$ turns out to be smaller than $L_n(x_1)$, then noise has caused a comparison between two points to now be wrong (we will denote this as a "false switch"). This freezing effect means that when fitness caching makes the impact of noise more serious. Furthermore, rather than noise having a positive "melting" effect that can help a search process escape local optima (as further discussed in [13, 15], and as is implicit in the design of simulated annealing), fitness caching causes any new local optima that are created by the noise to be "frozen" in place. We will denote local optima that are present in L_n, but not present in the original L as "false optima".

When faced with a new landscape to be searched, we do not know what the landscape looks like. However, it is possible to probe the landscape for some information, before starting a search process. Let us assume that we can obtain a reasonable estimate of the true ϵ-distribution within the landscape. That is, we would like to capture the distribution of fitness differences between neighboring points ($|L(x_i) - L(x_j)| \forall (x_i, x_j) \in S^2$ s.t. x_i and x_j are neighbors in the space). We will denote the probability density function (pdf) for this ϵ-distribution as $P(\epsilon)$. (Monte Carlo sampling from L_n will give an estimate of the noisy ϵ-distribution, which may be a tolerable approximation of the true ϵ-distribution, or may need to be corrected for noise.)

Given the pdf $P(\epsilon)$, we will now derive expressions for the likelihood of noise creating false switches and false optima, in terms of the standard deviation of the noise (σ).

For convenience, we will denote the pdf for the Gaussian distribution with mean value, μ, and standard deviation, σ by $f(x, \mu, \sigma)$, defined as follows:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} \tag{1}$$

3.1.1 False Switch Probability

In Equation 2 the inner integral represents the probability of the noise added to L(x_2), N_2 being less than the noise added to L(x_1), N_1. The inner two integrals (together) represent the probability of a false switch for a given difference between neighbors' real fitness values, ϵ. The outermost integral (integrating across all possible ϵs) computes the probability of a false switch for the given an ϵ-distribution P(ϵ).

$$2\int_0^\infty P(\epsilon) \left(\int_{-\infty}^\infty f(N_1, 0, \sigma) \left(\int_{-\infty}^{N_1} f(N_2, \epsilon, \sigma) dN_2 \right) dN_1 \right) d\epsilon \tag{2}$$

Equation 2 can be simplified to Equation 3, where Erf denotes the Gaussian error function. It has been remarked in certain contexts [5] that the Gaussian error function (Erf) is

[1] In the context of real-world problems, it may be confusing to think of there being a "true" fitness landscape with noise being added to it; alternatively, L may be viewed as the true *expected value* of the noisy function.

computationally very time-consuming to compute, and that more efficient (though slightly less accurate) approximations may be desirable. However, our approach is to derive a measure that will characterize the robustness of the fitness landscape as a whole. This is essentially an *offline* calculation which will be completed once before initiating a search process, rather than an *online* calculation that must be run repeatedly during the search process. Furthermore, since fitness caching is being used, there is an implicit assumption that evaluating a single point from the fitness landscape takes orders of magnitude longer than other operations, and the efficiency of numerical approximations is not a primary concern.

$$2\int_0^\infty P(\epsilon)\left(1-\frac{1+Erf\left[\frac{\epsilon}{2\sigma}\right]}{2}\right)d\epsilon \tag{3}$$

3.1.2 False Optima Probability

In order to obtain an analytic formula for the probability of creating false optima, we must make the additional simplifying assumption that the distribution $P(\epsilon)$ is the same throughout the space – i.e., at every x, $P(\epsilon)$ is the same regardless of $L(x)$.

For an arbitrary noise distribution ($P(N)$), the probability of being a local optimum in L_n is given by Equation 4.

$$\int_{-\infty}^{\infty} P(N_1)\left(\int_{-\infty}^{\infty} P(\epsilon)\left[\int_{-\infty}^{-\epsilon+N_1} P(N_2)dN_2\right]d\epsilon\right)^n dN_1 \tag{4}$$

Similarly, the probability of a given point being a local optimum in both L and L_n is given by Equation 5.

$$\int_{-\infty}^{\infty} P(N_1)\left(\int_{-\infty}^{0} P(\epsilon)\left[\int_{-\infty}^{-\epsilon+N_1} P(N_2)dN_2\right]d\epsilon\right)^n dN_1 \tag{5}$$

False optima are points that appear as local optima after noise is applied, but were not local optima before noise, thus the probability of being a false optimum is calculated by subtracting Equation 5 from Equation 4. Equations 4 and 5 were for arbitrary noise distributions, but since we are assuming all noise is additive Gaussian noise, we can transform them into Equations 6 and 7 respectively.

$$\frac{1}{2}\int_{-\infty}^{\infty} f(N_1,0,\sigma)\left(\int_{-\infty}^{\infty} P(\epsilon)\left[1+Erf\left[\frac{-\epsilon+N_1}{\sigma\sqrt{2}}\right]\right]d\epsilon\right)^n dN_1 \tag{6}$$

$$\frac{1}{2}\int_{-\infty}^{\infty} f(N_1,0,\sigma)\left(\int_{-\infty}^{0} P(\epsilon)\left[1+Erf\left[\frac{-\epsilon+N_1}{\sigma\sqrt{2}}\right]\right]d\epsilon\right)^n dN_1 \tag{7}$$

Given $P(\epsilon)$ (the probability density function for the ϵ-distribution of a fitness landscape), we now have closed-form expressions for the probabilities of a "false switch" occurring between any two neighboring points, and the probability of any given point becoming a "false optimum."[2]

[2]Despite being closed-form mathematical expressions, numerical integration approaches will generally be required, especially since $P(\epsilon)$ may be any arbitrary pdf.

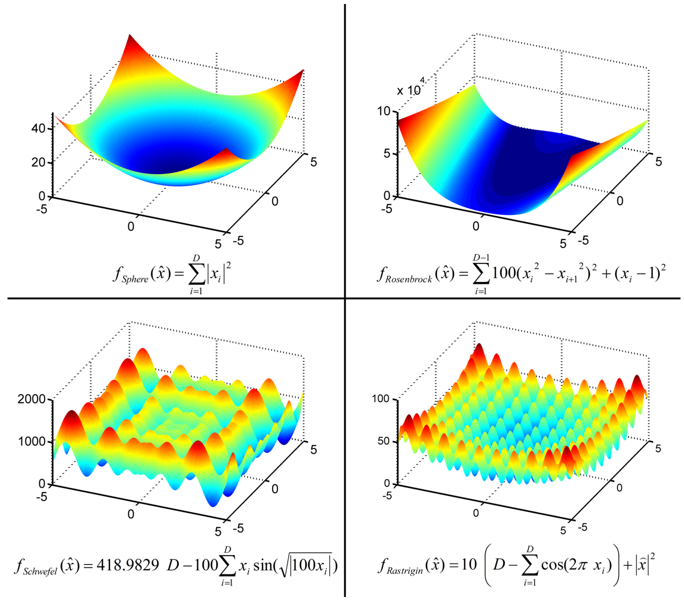

Figure 2: This figure shows 2-D versions of the sphere, Rosenbrock, Schwefel, and Rastrigin functions we used as our fitness landscapes. The equations are shown below each plot.

3.2 Fitness Landscapes

The abstract elegance of formally deriving mathematical measures or descriptive statistics about fitness landscapes must be grounded by the study of concrete fitness landscapes. This partially serves to validate the derivations, but more importantly it helps us judge the appropriateness of any simplifying assumptions that were made in order to make the mathematics tractable.

For our fitness landscapes, we selected four noiseless fitness functions that are often studied in the context of real-valued black-box optimization, and which exhibit differing landscape features (such as multi-modality/nonconvexity). Specifically, we chose the sphere, Rosenbrock, Schwefel, and Rastrigin functions (adapted from [3]). These noiseless landscapes are assumed to be the "true" underlying functions, which we will combine with varying levels of additive Gaussian noise to create the "obscured" noisy fitness landscapes that must be searched. Surface plots of the 2-dimensional versions of these fitness landscapes are shown in Figure 2, shown for illustrative purposes to communicate the general shape of these spaces. All results presented in the paper used the 10-dimensional version of these functions, where each dimension was discretized on the domain $[-5, 5]$ at a resolution of 0.05, creating a discrete search space of size $201^{10} \approx 1.1 \times 10^{23}$. The general mathematical function to generate the N-dimensional case for each landscape is displayed below the graphics in Figure 2.

In Figure 3, kernel density distribution plots[3] show the ϵ-distributions (distribution of differences between the "real" fitness values at neighboring locations in the fitness space)

[3]Kernel density distribution plots provide a way to visualize distributional information that avoids the artifacts caused by bin-size choices in histograms.

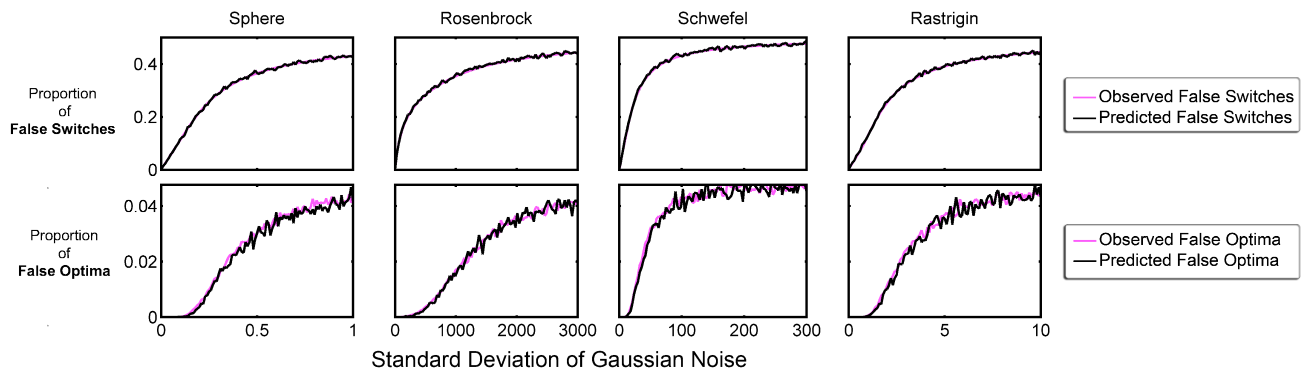

Figure 4: We predicted the probabilities of false switches and false optima occurring using the measures presented in Section 3 and observed the actual probabilities that each occurred by adding various amounts of noise to each function and evaluating the resulting proportions of false switches and false optima.

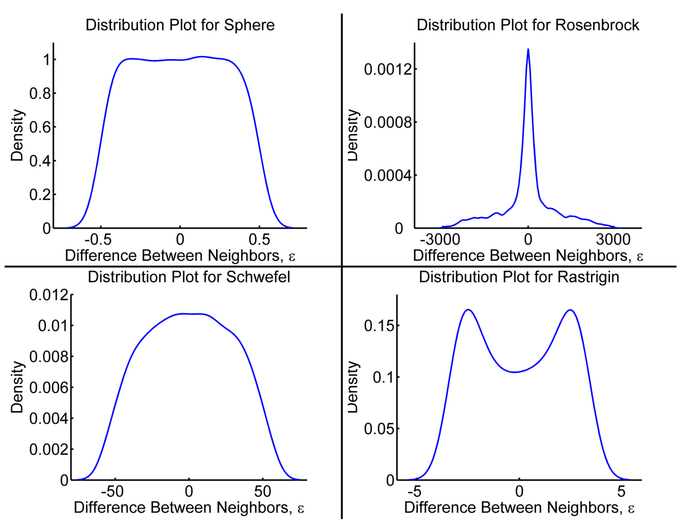

Figure 3: This figure shows the ϵ-distribution (fitness differences between neighboring locations) for each fitness landscape.

for each of these landscapes. Note that the different distributions vary significantly in shape and range of values.

3.3 Empirical Measure Validation

We predicted the number of false switches and false optima in each fitness landscape using the measures defined in Section 3 above and an approximate ϵ-distribution defined by sampling 5000 differences between neighbors' real fitness values. Then we observed the real probability of false switches being created by noise by testing 10,000 pairs of neighboring points, which were evaluated before and after varying amounts of Gaussian noise was added. Similarly, we used a Monte Carlo method (testing 10,000 points) to estimate the real probability that a point becomes a false optimum as a result of differing magnitudes of Gaussian noise. As shown in Figure 4, the formulas we derived for these two measures closely approximate the directly observed measures.

4. EXPERIMENTS

We are further interested in whether these or other simple measures can be useful in predicting the performance of an evolutionary search technique on a noisy landscape. In particular, it would be most useful to be able to choose the number of times a noisy function should be evaluated and averaged, to enable a search mechanism to reach very good locations in the space with as few function evaluations as possible. Specifically, we ran experiments at varying noise levels to determine the number of evaluations required by a stochastic hill climber to reach an average fitness value that is in the best 0.0001% of the landscape. These numbers of evaluations are then scaled by the number of times the function would need to be evaluated to reach their respective noise levels. The pseudocode for the simple random-mutation hill climbing algorithm is given in Table 1.

The noise level (standard deviation of noise) for which the search progresses most rapidly is denoted σ_{ideal} (which will vary for each landscape). See Figure 5 for an illustration of this process.

We considered four heuristic methods for using a landscape's ϵ-distribution to predict σ_{ideal} and compared the

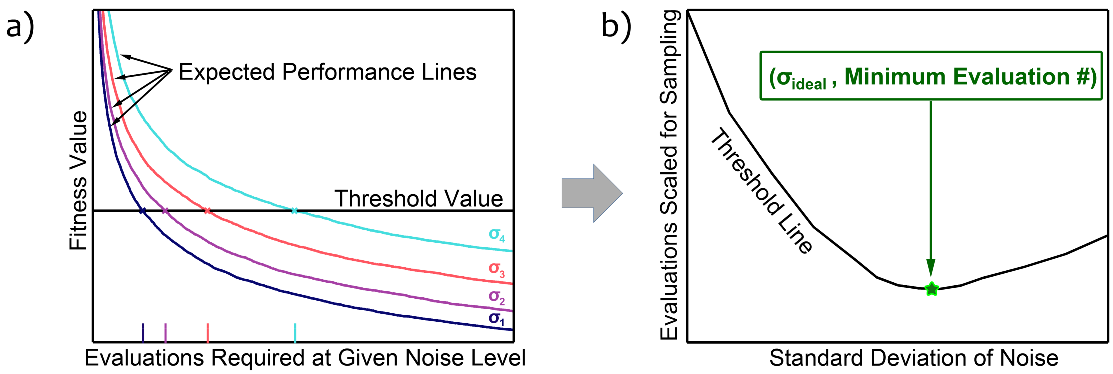

Figure 5: a) Each shaded line shows fitness values reached after some number of evaluations, for a given noise level, σ_x. Using this information we calculated the number of evaluations it took to reach a threshold value, and scaled this by the number of replicate evaluations required to reduce noise to the specified level (σ_x). b) This scaled number of evaluations is plotted at each noise level. We denote the noise level corresponding to the minimum number of evaluations as σ_{ideal}, which is the "sweet spot" target for noise reduction.

Given a (memoizing) noisy landscape function L_n, and a function *neighbor*(x) which returns a new location by increasing or decreasing x along a single randomly chosen dimension:

1. Let $x_{best} = \varnothing$
2. Choose x randomly from S (the search space)
3. If $x_{best} = \varnothing$ or $L_n(x) < L_n(x_{best})$: Set $x_{best} = x$
4. If evaluation limit exceeded: Return x_{best}.
5. If x has been compared to all of its neighbors and is a local minimum: Go to Step 2.
6. Let $x' = neighbor(x)$
7. If $L_n(x') < L_n(x)$: Set $x = x'$
8. Go to Step 3

Table 1: Pseudocode for a random-mutation hill climber, which restarts when stalled.

number of evaluations required by the hill climber at each method's predicted σ_{ideal} to those required at the true σ_{ideal}.

The four heuristics for predicting σ_{ideal} are listed below. In order to calibrate the heuristics, it was necessary to use scaling factors based on the true σ_{ideal} for the landscape. We then tested the heuristics by applying them to each landscape in turn, in order to evaluate whether they could capture the differences between the landscapes.

- *Fixed Noise Level:* The geometric mean of the σ_{ideal} for each landscape is 1.91 and this constant noise value was used as the $\sigma_{Fixed\ Noise\ Level}$. This is the most naïve heuristic, as it treats all landscapes the same, without making use of the ϵ- distribution information at all. It is included mainly as a baseline for comparison.

- *Direct Ratio:* The geometric mean of the ratio of the median of each ϵ-distribution to the σ_{ideal} for each landscape is 3.97. We calculated $\sigma_{Direct\ Ratio}$ by dividing the median of each landscape's ϵ-distribution by this ratio.

- *False Switch:* The geometric mean of the proportion of false switch values corresponding to the σ_{ideal} for each landscape is 0.084. The standard deviation of noise which predicts a proportion of false switch value of 0.084 is the $\sigma_{False\ Switch}$

- *False Optima:* The geometric mean of the proportion of false optima values corresponding to σ_{ideal} for each landscape is 5.16×10^{-5}. The standard deviation which predicts this value is the $\sigma_{False\ Optima}$.

5. RESULTS AND DISCUSSION

To compare these methods on each of the four landscapes, we calculate the *inefficiency ratio* as the number of evaluations required by each method's prediction for σ_{ideal} (i.e., $\sigma_{Fixed\ Noise\ Level}, \sigma_{Direct\ Ratio}, \sigma_{False\ Switch}, \sigma_{False\ Optima}$) divided by the number required at the true σ_{ideal}. Note that an inefficiency ratio of 1.0 would be a perfect prediction, and also that ratios higher than 20 have been cut off, due to computational constraints.

To summarize the performance results from Figure 6:

1. None of the methods performed well on the Rosenbrock landscape. The Rosenbrock function is sometimes referred to as a "banana function" due to its long bending valley which must be followed to reach the global optimum. The failure to predict an optimal level of noise may be due in large part to the importance of traversing this valley, where the fitness gradient is not very strong. In other words, the initial sampling of the whole space to determine the ϵ-distribution is misleading, since a particular region of the space (the valley floor) is much more important for search performance than the space at large, and requires lower noise values to traverse.

2. The *fixed noise level* method performed quite poorly on all but one landscape. In general, this is not too

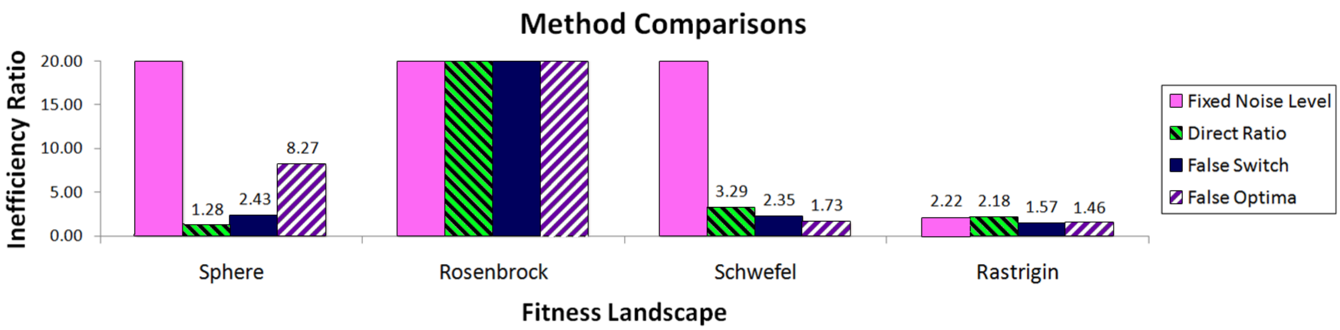

Figure 6: This figure shows how inefficient the standard deviation chosen by each method is by calculating the ratio of evaluations to that required at optimal noise level, σ_{ideal}. A perfect solution would have an inefficiency ratio of 1.0.

surprising. We expect that different landscapes will require different optimal noise levels, and choosing a fixed level value to apply to all landscapes is unlikely to perform well.

3. There is no clear winner among the other three methods: the *false optima* and *direct ratio* methods were each best on certain landscapes, but the *false switch* method also generally performed well. This result is somewhat disappointing, in that heuristics using our derived metrics (*false switches* and *false optima*) do not have a strong advantage over the simpler approach (*direct ratio*) of scaling by the median value from the ϵ-distribution.

While these results are not decisive, it is somewhat encouraging that the three methods using information from the ϵ-distribution serve as better predictors than the most naive approach. This shows that the heuristics used are at least partially correlated with choices for σ_{ideal}, and perhaps improved mappings may be developed along similar lines, in order to offer prescriptive guidelines for choice of sampling repetitions based on this information.

6. FUTURE WORK AND CONCLUSIONS

The experimental results we have presented are based only on an examination of four fitness landscapes, which is too small to be a good representation of the types of fitness landscapes encountered in real problems. Furthermore, it has been argued that some of these particular test landscapes may not be the most appropriate choice for benchmark functions for evolutionary algorithms [19]. Accordingly, further studies along similar lines are called for, involving a greater diversity of noisy fitness landscapes.

However, perhaps a more fundamental issue with our current approach is that the search performance on these landscapes appear to be significantly different enough that none of the heuristics approaches we investigated served as a good predictor for all four landscapes. In particular, the failure to predict a good noise level for the Rosenbrock landscape merits further investigation. This may also suggest that a fundamentally different approach is needed. One hypothesis is that knowledge of the global ϵ-distribution for a landscape is insufficient to make a good prediction of what the optimal noise level would be, and thus additional knowledge is required. This may be because intelligent search techniques find relatively good solution areas quickly, and thus spend very little time in the large poor-performance areas of the space, in which case a more biased approach for sampling ϵ-distributions might be fruitful (e.g., taking inspiration from [17]). For instance, one could imagine running a sequence of searches, bootstrapping the ϵ-distribution from the points that were encountered by the previous search on the noisy landscape, thus refining the estimates for optimal sampling choices in later searches.

In addition to their role in meta-heuristic search processes, fitness landscapes also play an important role in the study of many complex systems, and may provide a lens for viewing adaptive or evolving systems in new and enlightening ways (c.f. Kauffman's work on evolutionary landscapes [9]). It would be interesting to investigate whether there are interdisciplinary implications for studying frozen noisy landscapes, in relation to processes that occur in real biological systems.

An improved understanding of the extent to which noise can be present in a fitness landscape without seriously inhibiting successful search and adaptation in that space is a broad but desirable goal, which would significantly advance the field of search/optimization when dealing with uncertain problems. Our present research provides some progress toward this goal in the specific context of fitness caching, but the path is far from clear, and significant work remains to be done in this direction. The lack of prior literature on fitness caching with noise may suggest either that the combination has not been given serious consideration, or possibly that fitness caching is not an advisable approach when dealing with noisy search problems. While we believe that in many cases it would still prove beneficial, this is ultimately an empirical question, and one that we hope will be resolved by future work using fitness caching in noisy environments.

In conclusion, we have offered a preliminary foray into the study of the interactions between noisy landscape sampling and fitness caching. We presented and verified analytic formulas for two measures that could be useful for predicting the impact of noise on the performance of fitness-caching neighborhood based meta-heuristic search processes in discrete fitness landscapes. We also explored several heuristics for choosing an optimal sampling level under these conditions, and while none of these heuristics offer perfect solutions to this problem, they could provide reasonable initial choices when there is no *a priori* information about what sampling level to use for an unknown fitness landscape. Additionally, they provide a starting place for developing better heuristics for this problem. However, further research is required before we can offer prescriptive recommendations for noise level reduction methodology. Similar investigations on additional fitness landscapes using other meta-heuristic search methods (simulated annealing, GAs, tabu search, PSO, etc.) will likely offer further insight into the effects of noise on landscape structure.

Acknowledgments: We gratefully acknowledge support from NSF grant IIS-0713619, and also thank Dr. Uri Wilensky for his oversight of this project.

7. REFERENCES

[1] Balaji, P., Srinivasan, D., and Tham, C. Uncertainties reducing Techniques in evolutionary computation. In *IEEE Congress on Evolutionary Computation, 2007. CEC 2007* (2007), pp. 556–563.

[2] Fitzpatrick, J. M., and Grefenstette, J. J. Genetic algorithms in noisy environments. *Machine Learning 3*, 2 (1988), 101–120.

[3] Hansen, N., Finck, S., Ros, R., and Auger, A. Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions. Research Report RR-6829, INRIA, 2009.

[4] Hansen, N., Finck, S., Ros, R., and Auger, A. Real-Parameter Black-Box Optimization Benchmarking 2009: Noisy Functions Definitions. Research Report RR-6869, INRIA, 2009.

[5] Hughes, E. Multi-objective probabilistic selection evolutionary algorithm. Tech. Rep. DAPS/EJH/56/2000, Department of Aerospace, Power, & Sensors, Cranfield University, RMCS, Shrivenham, UK, September 2000.

[6] Jaskowski, W., and Kotlowski, W. On selecting the best individual in noisy environments. In *GECCO '08: Proceedings of the 10th annual conference on Genetic and evolutionary computation* (New York, NY, USA, 2008), ACM, pp. 961–968.

[7] Jin, Y. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing-A Fusion of Foundations, Methodologies and Applications 9*, 1 (2005), 3–12.

[8] Jin, Y., and Branke, J. Evolutionary optimization in uncertain environments-a survey. *Evolutionary Computation, IEEE Transactions on 9*, 3 (2005), 303–317.

[9] Kauffman, S. *The origins of order: Self organization and selection in evolution*. Oxford University Press, USA, 1993.

[10] Kauffman, S., and Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *Journal of theoretical Biology 128*, 1 (1987), 11–45.

[11] Kratica, J. Improving performances of the genetic algorithm by caching. *Computers and artificial intelligence 18*, 3 (1999), 271–283.

[12] Kratica, J., Tosic, D., Filipovic, V., Ljubic, I., et al. Solving the simple plant location problem by genetic algorithm. *RAIRO Operations Research 35*, 1 (2001), 127–142.

[13] Levitan, B., and Kauffman, S. Adaptive walks with noisy fitness measurements. *Molecular Diversity 1*, 1 (1995), 53–68.

[14] Merz, P., and Freisleben, B. On the effectiveness of evolutionary search in high-dimensional nk-landscapes. In *In Proceedings of the 1998 IEEE International Conference on Evolutionary Computation* (1998), IEEE Press, pp. 741–745.

[15] Rana, S., Whitley, L., and Cogswell, R. Searching in the presence of noise. *Lecture Notes in Computer Science 1141* (1996), 198–207.

[16] Smith, R., and Smith, J. New methods for tunable, random landscapes. *Foundations of Genetic Algorithms 6* (2001), 47–69.

[17] Smith, T., Husbands, P., Layzell, P., and O'Shea, M. Fitness landscapes and evolvability. *Evolutionary Computation 10*, 1 (2002), 1–34.

[18] Stonedahl, F., and Wilensky, U. Finding forms of flocking: Evolutionary search in abm parameter-spaces. In *Proceedings of the MABS workshop at the 9th international conference on Autonomous Agents and Multi-Agent Systems* (Toronto, Canada, 2010).

[19] Whitley, D., Mathias, K., Rana, S., and Dzubera, J. Building better test functions. In *Proceedings of the Sixth International Conference on Genetic Algorithms* (1995), Morgan Kaufmann, pp. 239–246.