- Cover Page -

Proceedings of the American Society for Composites 18th Technical Conference

Paper title: A comparison of an estimation-of-distribution algorithm and a stochastic hill-climber for composite optimization problems

Authors: Laurent Grosset, École des Mines de Saint Étienne, France
Rodolphe Le Riche, CNRS URA 1884, École des Mines de Saint Étienne, France
Raphael T. Haftka, University of Florida, USA

Paper number: # 168

A comparison of an estimation of distribution algorithm and a stochastic hill-climber for composite optimization problems

Laurent Grosset, Rodolphe Le Riche, Raphael T. Haftka

ABSTRACT

Evolutionary algorithms (EA) have become a standard tool for the optimization of complex composite structures because of their ability to solve combinatorial problems. However, several studies have shown that simpler algorithms, such as stochastic hill climbers (SHC) can be more efficient even on problems designed to demonstrate EAs superiority, such as the Royal Road problem. The present paper compares the performance of a variant of EA, the univariate marginal distribution algorithm (UMDA) with that of an SHC on different fitness landscapes found in laminate optimization problems and identifies factors that influence the algorithms' relative performance. In particular, it is found that mUMDA, a hybrid algorithm that combines UMDA's global distribution learning and SHC's local random search, outperforms SHC on large, highly constrained problems and on multimodal problems.

Keywords: Stacking sequence optimization, evolutionary computation, estimation of distribution algorithms, stochastic hill climber

1 INTRODUCTION

Stochastic hill climber (SHC) and univariate marginal distribution algorithms (UMDA) [1, 2] are two fundamentally different stochastic optimizers. SHC proceeds with local perturbations while UMDA uses a population to estimate the global probability density of promising designs. While the population allows UMDA to sample the design space by using information from several points, estimating the distribution accurately may require many evaluations at each generation, so that the computational cost incurred may neutralize the benefit of using populations.

Laurent Grosset, Centre SMS, École des Mines de Saint Étienne, France
Rodolphe Le Riche, CNRS URA 1884 / SMS, École des Mines de Saint Étienne, France
Raphael T. Haftka, Mechanical and Aerospace Engineering Department, University of Florida, USA

Previous contributions compared SHC to other population based EAs for specific objective functions. For example, SHC and genetic algorithms were compared in [3] on the Royal Road function. Comparisons between single point and population based evolution strategies on Long-Path problems are given in [4]. The benefits of using populations for deceptive or multimodal problems were studied in [5]. The convergence of estimation of distribution algorithms was investigated in [6]. These studies showed that SHC outperforms more sophisticated EAs on a number of non-trivial problems.

The present paper presents an experimental and theoretical comparison of the performance of SHC and UMDA for three composite laminate optimization problems. Each of these problems was chosen to exhibit particular characteristics: (1) the maximization of the in-plane longitudinal stiffness is a separable, unimodal problem that allowed us to isolate the effect of problem size. (2) The maximization of the first natural frequency of a simply supported laminated plate subject to a constraint on the effective Poisson's ratio reveals the effect of pseudo-equality constraints. It is a multimodal problem in which the sensitivity of the fitness function to the variables is not uniform. (3) The maximization of the strength of laminate is an example of a multimodal non separable problem.

Two performance measures were used in this paper: the optimization reliability, defined as the probability of finding the true optimum, and the expected maximum fitness function[1].

2 PRESENTATION OF THE ALGORITHMS

We consider the problem of maximizing a fitness function $\mathcal{F}$ over a design space $D = A^n$, where A is an alphabet of cardinality c. SHC searches the space by choosing an initial point $x = (x_1, x_2, \ldots, x_n)$ at random and applying random perturbations to it. A new point is accepted only if it improves the $\mathcal{F}$. Perturbations are applied until a predetermined number of evaluations have been performed.

Three types of perturbations (or mutations) are possible:

STEP: unit step in the Hamming space. The value of *one* variable chosen at random is changed to an adjacent value with probability 1.

LOC: local mutation. The value of each variable in the string is changed with probability p_m to one of its adjacent values.

GLOB: global mutation. The value of each variable in the string is changed with probability p_m to any value of the alphabet A.

UMDA is a simple form of estimation of distribution algorithms (EDA). EDAs use populations of m points to infer the distribution $p(x)$ of good points. By a succession of sampling and selection steps, the distribution converges to identify regions of high fitness, eventually yielding the optimum. In UMDA, distributions are expressed as products of univariate marginal distributions, leading to the following update rule for the distributions:

$$p(x, t+1) = \prod_{i=1}^{n} p^s(x_i, t) \qquad (1)$$

[1] The two measures can conflict when an algorithm has a high probability of finding the optimum, but converges to extremely poor solutions otherwise, leading to mediocre average performance.

where $p(x, t+1)$ refers to the search distribution at time $t+1$ and $p^s(x_i, t)$ designates the univariate distribution of the variable x_i in the selected individuals at time t. In this work, truncation selection of ratio τ was used.

In this work, mutation was added to the original UMDA. The algorithm, called here mUMDA, for mutation UMDA, can be summarized as follows:

1. initialize time $t = 0$ and the search distribution $p(x, t = 0)$,
2. create m points by sampling from $p(x, t)$, and applying one of the perturbation operators,
3. select the τm best individuals based on the fitness function,
4. estimate the univariate marginal distributions $p^s(x_i, t)$ as the frequencies of all the values of x_i in the selected population, and $p(x, t+1)$ from Equation (1),
5. increment t and go to 3.

3 NUMERICAL EXPERIMENTS

Three laminate optimization problems are considered, all for symmetric balanced laminates. The first problem, "Max A_{11}", is a separable and unimodal function. The second, "Vibration/Poisson", is a highly constrained, non-separable, multimodal function and the third, "Strength", is a non-separable, multimodal problem.

A composite laminate is made of layers of fiber reinforced material (plies), and its response is determined by factors including the number of plies and the fiber and matrix properties. Here, we consider only laminates with a fixed total number of plies $4n$. The goal of laminate optimization is to determine the angles x_1 to x_n (symmetry and balance reduce the number of variables to n) that maximize some objective function. Even though fiber angles can take any value between 0° and 90°, the set of acceptable values is typically limited to a small number of discrete values. In the problems considered in this paper, the angles can take seven values from 0° to 90° in 15° steps. The laminate is assumed to be made of stacks of two plies. For example, $x_1 = 0°$, $x_2 = 45°$, $x_3 = 30°$ corresponds to the laminate $[0_2/\pm 45/\pm 30]_s$.

3.1 Max A_{11} problem

The first problem is maximizing the longitudinal in-plane stiffness A_{11} of a graphite-epoxy laminate, which is expressed by:

$$A_{11} = U_1\, h + 4U_2 \sum_{k=1}^{n} t_k \cos 2x_k + 4U_3 \sum_{k=1}^{n} t_k \cos 4x_k \tag{2}$$

where U_1, U_2 and U_3 are the material properties, t_k the thickness of the k^{th} ply, and h the total thickness of the laminate.

The fitness function A_{11} is a sum of functions of one variable only, so that they do not interact, and UMDA is expected to converge to the global optimum $x_i^* = 0°$, $i = 1, \ldots, n$. The function A_{11} is unimodal, so that SHC will also yield x^*. It allows us to compare the asymptotic performances of the two search mechanisms for local optimization.

The algorithms were applied to the problem for five different numbers of variables $n = 12, 20, 50, 100, 200$. The selection ratio τ of the UMDA was kept constant at $\tau = 0.3$ (as recommended in [6]). No mutation was applied for this problem ($p_m = 0$). Several population sizes (50, 100, 500 and 1000) were tried. The STEP mutation was used for SHC.

Two criteria were used to compare algorithm performance: the number of analyses required to reach 80% reliability (probability of finding the optimum, estimated over 50 independent runs), and the number of analyses needed until the average best fitness reaches 98% of the optimal fitness.

Figure 1 presents the number of evaluations to 98% of the maximum fitness against the number of variables for SHC and four different population sizes of UMDA. Clearly, SHC converges faster than UMDA for all these cases. The cost for SHC appears to be linear in the number of variables, which confirms results reported in Section 4 and in [7]. For UMDA with a given population size, the cost increases sub-linearly. Larger populations are more expensive, but smaller populations can fail to converge for large n, as happened for a population of 50 individuals. For $n \geq 50$, the average maximum fitness never reaches 98% of the maximum fitness. This can be explained by increased chance for smaller populations for particular values of the variables to disappear in the entire population, as will be discussed in section 4.2.

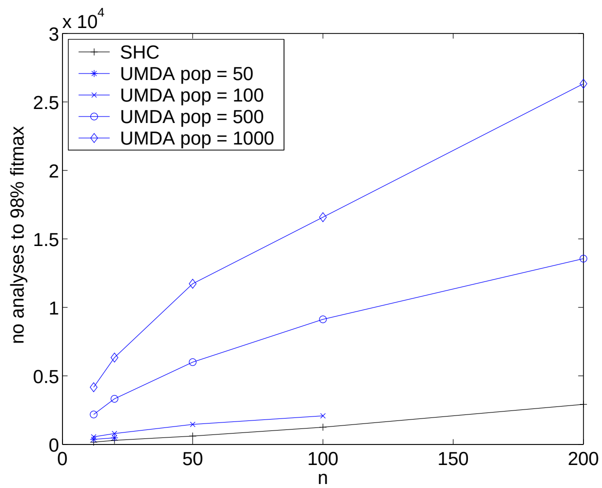


Figure 1: Number of analyses until the average maximum fitness reaches 98% of the optimal fitness, Max A_{11} problem.

The loss of variable values for small populations affects the reliability: for each problem size n, there exists a minimum population size m^* below which 80% reliability is never reached. This minimum population size was 100 for $n = 12$, 500 for $n = 20$, 50 and 100 and 1000 for $n = 200$. As a result, UMDA must work with large populations to preserve diversity. To prevent premature convergence of the distributions and allow smaller populations to be used, two mechanisms can be implemented: memory or perturbations. Memory was used in [2], with new probabilities obtained as the weighted average of the previous probabilities and the frequencies in the selected individuals. Memory led to an algorithm that converged with high reliability to the optimum, however this was achieved at the expense of convergence speed. To prevent definitive loss of variable values while allowing probabilities to quickly focus on high fitness regions, a perturbation approach was chosen in this work.

Figure 2 shows the number of evaluations until 98% of the optimal fitness and 80% reliability are reached, respectively for the case $n = 12$ with $m = 20$. A local mutation operator (LOC) was applied, with mutation probability varying from 0.01 to 0.3. The resulting algorithm, called here mUMDA, for mutation UMDA, is a hybrid between UMDA and SHC in the sense that is uses both distribution learning and random exploration to search the design space. A dramatic performance improvement is observed when mutation is used: while the original UMDA failed to reach 98% of the maximum fitness and to find the true optimum, mUMDA quickly converges to the optimum, with the best performance achieved for $p_m = 0.05$. On this problem, SHC remains the most efficient algorithm, requiring only 160 evaluations to reach 98% of the maximum fitness and 225 evaluations to achieve 80% reliability, versus 243 and 360 evaluations respectively for the best mUMDA scheme. Nonetheless, by allowing smaller populations to be used, mutation makes UMDA a viable option for solving laminate problems.

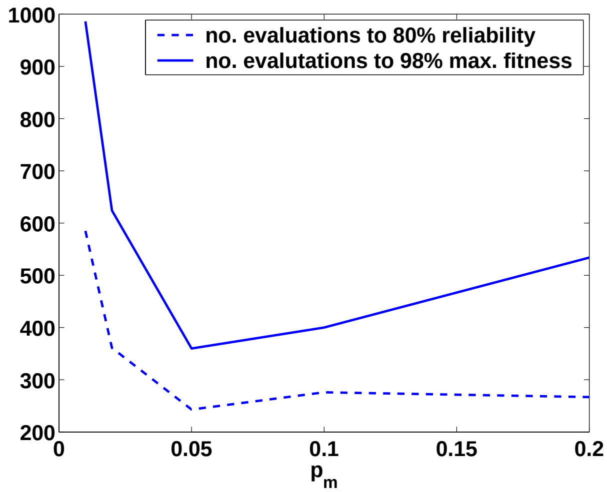


Figure 2: Influence of mutation on UMDA's performance ($n = 12$, $m = 20$).

3.2 Constrained vibration maximization problem

The second problem is maximizing the first natural frequency of a simply supported graphite-epoxy laminated plate of length $L = 50$" and width $W = 15$" subject to a constraint on the effective Poisson's ratio $\nu^l \leq \nu_{eff} \leq \nu^u$, with $\nu^l = 0.48$ and $\nu^u = 0.52$. The first natural frequency is given by:

$$F = \frac{\pi^2}{\sqrt{\rho h}} \sqrt{\frac{D_{11}}{L^4} + \frac{2(D_{12} + 2D_{66})}{L^2 W^2} + \frac{D_{22}}{W^4}} \qquad (3)$$

where ρ designates the mass density, and the D_{ij}'s are the bending stiffness coefficients.

The effective Poisson's ratio is given by:

$$\nu_{eff} = \frac{U_1 h - U_2 V_1 + U_3 V_3}{U_4 h - U_3 V_3} \qquad (4)$$

where the in-plane lamination parameters V_1 and V_3 are obtained by:

$$V_1 = 4\sum_{k=1}^{n} t_k \cos 2\theta_k, \quad V_3 = 4\sum_{k=1}^{n} t_k \cos 4\theta_k \tag{5}$$

Neither UMDA nor SHC accommodates constraints, therefore a penalty approach was used, where the fitness function $\mathcal{F}$ of infeasible designs is decreased in proportion to the constraint violation:

$$\mathcal{F}(x) = \begin{cases} F(x) & \text{if } g(x) \leq 0 \quad \text{(feasible)} \\ F(x) - pg(x) & \text{if } g(x) > 0 \quad \text{(infeasible)} \end{cases} \tag{6}$$

where p is the penalty parameter whose value is adjusted empirically to ensure that the algorithm yields feasible designs. The constraint term $g(x)$ was defined as

$$g(x) = \max\left(1 - \frac{\nu_{eff}(x)}{\nu^l}, \frac{\nu_{eff}(x)}{\nu^u} - 1\right). \tag{7}$$

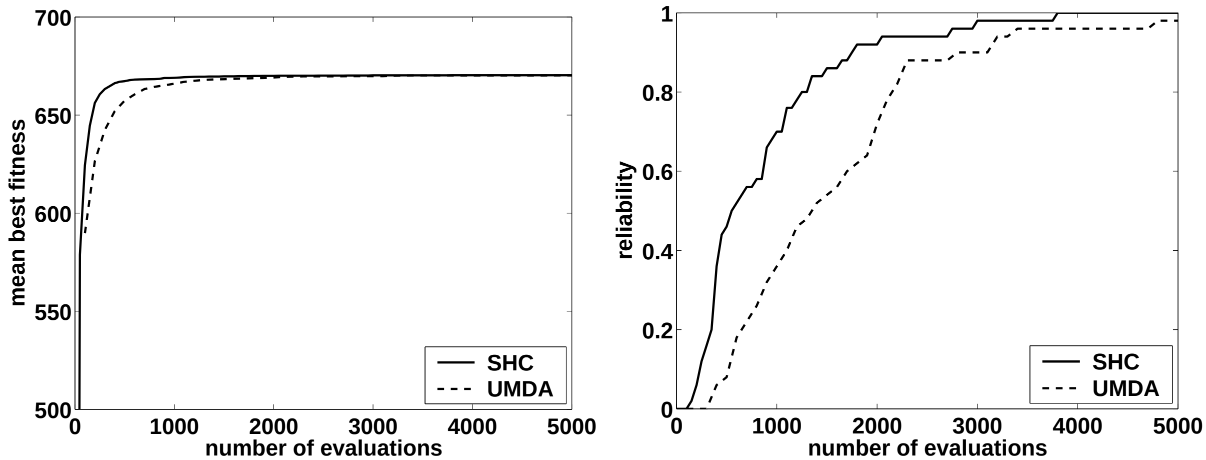
The constraint on the Poisson's ratio forces the points to remain in a narrow channel in the design space, which makes the problem particularly difficult for stochastic algorithms because many of the random perturbations result in infeasible designs.

Two numbers of dimensions were considered: $n = 8$ and $n = 15$. Without the constraint, the optimal orientation would be 90° for all the plies. The effective Poisson's ratio would then be $\nu_{eff} = \nu_{21} = 0.0165$. The Poisson's ratio constraint forces 30°, 45° 60°, and 75° plies into the inner layers of the laminate, where they are the least damaging to the frequency. The optimum for $n = 8$ is $[90_2/\pm75/\pm45_5/\pm30]_s$. When $n = 15$, the optimum is $[90_4/\pm75/\pm60_2/\pm45_5/\pm30_5]_s$. These optima were obtained as the best designs found over all the runs performed for this study.

Two mutation operators, LOC and GLOB, were used for this problem for both algorithms. The population size (for mUMDA) and the mutation type and rate were determined empirically to maximize performance.

For $n = 8$, two population sizes, $m = 20$ and $m = 50$, and three mutation rates, $p_m = 0.1$, $p_m = 0.2$ and $p_m = 0.3$ were tried for mUMDA. In order to select the best scheme, the reliability reached at 2000 evaluations was used as criterion. The highest reliability (88%) was achieved with $m = 20$ and LOC mutation with $p_m = 0.2$. A similar parameter study was conducted for SHC with p_m ranging from 0.1 to 0.5, and the highest reliability was achieved for $p_m = 0.4$. Figures 3a and 3b compare the mean best fitness and reliability of SHC and mUMDA, respectively. Clearly, SHC outperforms mUMDA on this problem. The two criteria show that SHC is able to converge faster to high quality solutions and to the optimum more reliably than mUMDA.

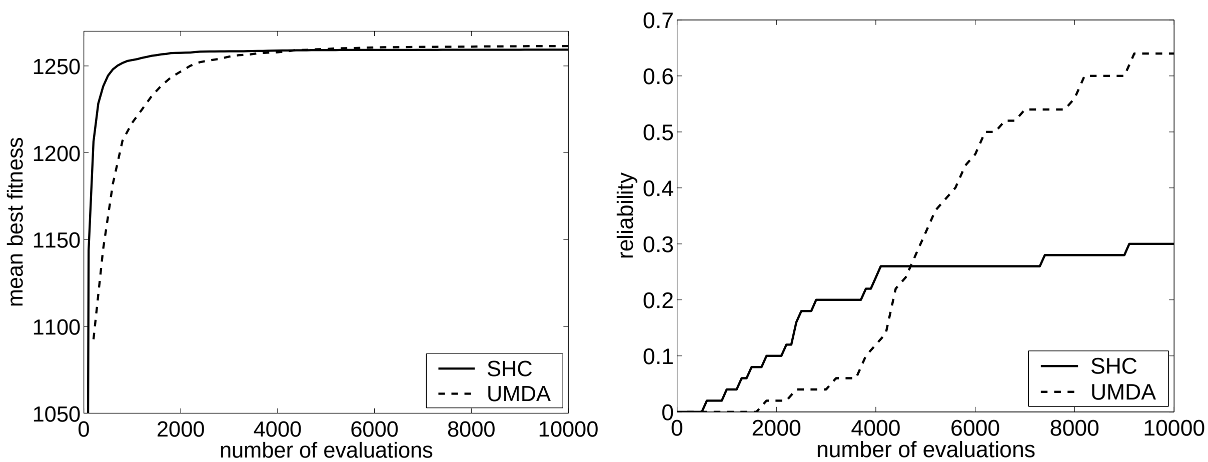
For $n = 15$, a population size of $m = 100$ individuals and a moderate mutation probability $p_m = 0.1$ were chosen. The larger population is justified because more individuals are needed to ensure that all variable values are present in the population. For SHC, the best mutation probability was found to be $p_m = 0.2$. Figures 4a and 4b compare the performance criteria for the two algorithms. This time, mUMDA seemed to benefit from the use of a global probabilistic model, which allows it to escape local minima and was able to reliably find the global optimum: after 10,000 evaluations, the reliability of the optimization had reached 64%, whereas the SHC only found the optimum in 30% of the runs. In 70% of the runs, SHC converged to a high quality solution but failed to yield the true optimum. For instance, one of the solutions was



(a) (b)

Figure 3: Mean best fitness (a) and reliability (b) for SHC and mUMDA for the constrained maximization of the first natural frequency of a laminated plate ($n = 8$).

$[90_6/\pm 60/\pm 45_{10}/\pm 30]_s$, which had a fitness of $\mathcal{F} = 1,279.3$ ($F = 1,257.9$, $\nu = 0.481$). In order to obtain the global optimum $[90_4/\pm 75/\pm 50_2/\pm 45_5/\pm 30_5]_s$ ($\mathcal{F} = 1,262.6$, $F = 1,262.6$, $\nu = 0.481$), six variables have to be mutated. However, all single mutations lead to a reduction in the fitness function, either because they make the design infeasible (variable 5) or because they decrease the vibration frequency. Consequently, multiple mutations must occur simultaneously for the fitness function to improve. The probability of that event decreases as n increases, thus making further progress of SHC unlikely. These results agree with [5] in which an SHC was shown that have an exponential time complexity for a multimodal function, while the cost of a population based EA was only polynomial.



(a) (b)

Figure 4: Mean best fitness (a) and reliability (b) for SHC and mUMDA for the constrained maximization of the first natural frequency of a laminated plate ($n = 15$).

3.3 Strength maximization: a multimodal problem

Many practical laminate optimization problems exhibit a multimodal relation between the variables and the fitness function. A typical example of such behavior is the maximization of the strength of a laminate. In this section, we maximize the strength of a laminate for the maximum strain criterion.

The problem can be formulated as follows:

$$\text{maximize } \lambda_s = \min_{k=1,\ldots,n} \left(\min \left(\frac{\epsilon_1^{ult}}{|\epsilon_1(k)|}, \frac{\epsilon_2^{ult}}{|\epsilon_2(k)|}, \frac{\gamma_{12}^{ult}}{|\gamma_{12}(k)|} \right) \right) \tag{8}$$

where $\epsilon_1(k)$, $\epsilon_2(k)$ and $\gamma_{12}(k)$ are the strains in the principal directions for the k^{th} ply, and ϵ_1^{ult}, ϵ_2^{ult} and γ_{12}^{ult} are the ultimate strains for the material considered.

We considered the case $n = 8$ design variables. The laminate is subjected to an in-plane loading: $N_x = -1000$ kN/m, $N_y = 200$ kN/m, $N_{xy} = 100$ kN/m. For this problem, the optimum laminate is $[0_6/90_{10}]_s$ or its permutations, and the optimum load factor is $\lambda_s = 5.39$. Depending on the orientation of the fibers, one of the three possible failure modes (fiber failure, matrix cracking, shear failure) becomes critical. The combination of these three failure modes results in a multimodal fitness function[2].

For mUMDA the population size was 50 and a moderate mutation rate of $p_m = 0.1$ was implemented to prevent premature convergence. Several variants of SHC were compared to obtain the best competitor to mUMDA. For this problem, local mutation only (LOC) was tried and the best rate was $p_m = 0.2$. Both the reliability (Figure 5a) and the mean best fitness (Figure 5b) show clearly the superiority of UMDA for this problem. The reliability of SHC increases faster than that of UMDA, but culminates at 36%, while mUMDA was able to converge reliably to the optimum. SHC often converged to local optima ($[0_8/90_8]_s$, $\lambda_s = 4.97$, $[0_6/\pm 30/90_8]_s$, $\lambda_s = 4.47$ or $[0_8/\pm 30/90_6]_s$, $\lambda_s = 4.02$) and was not able to escape even when large mutation rates were used because unlikely coordinated mutations were needed to reach the basin of attraction of the global optimum.

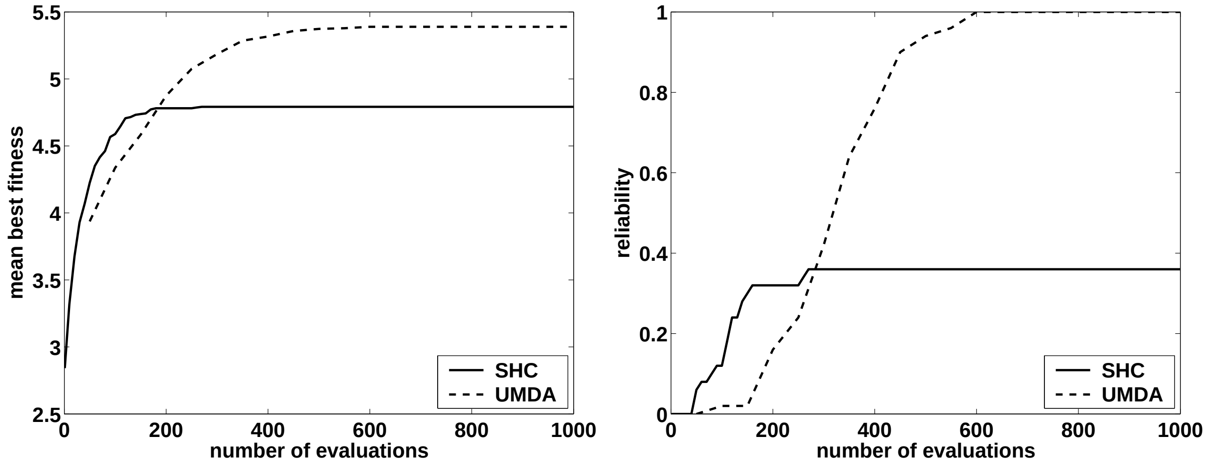
4 ELEMENTS OF THEORETICAL EXPLANATIONS FOR THE MAX A_{11} PROBLEM

The numerical experiments that were just described are now theoretically analyzed by calculating expected convergence times. There are n variables that can take c discrete values.

4.1 Convergence time of the SHC

The following analysis considers an SHC with STEP mutation, operating on a unimodal function. If the SHC is at a point where k out of the n variables are correctly set, the expected time before one of the non optimal variables is perturbed is $n/(n-k)$. The random perturbation can then take the variable closer to the optimum or not, with probabilities 1/2 (neglecting distortions due to limits on the values). The expected time for one beneficial step is then $2n/(n-k)$. Let d_i denote the average distance between the i-th variables of a random point

[2] Note that its is not necessary to consider all the failure modes for the response to be multimodal, as individual failure modes can be multimodal for specific combinations of material properties and loading.



(a) (b)

Figure 5: Reliability (a) and mean best fitness (b) of mUMDA and SHC for the strength problem. Unlike SHC, which was trapped in local optima, mUMDA was able to find the global optimum reliably.

and the optimum. If the c discrete values that can be taken by variables are mapped onto the set $1, 2, \ldots, c$, d_i is given by:

$$d_i \;=\; \frac{1}{c}\sum_{j=1}^{c} |x_i^j - x_i^*| \;, \tag{9}$$

where x_i^j is the j-th possible value of the i-th variable. In the $\max A_{11}$ problem, all variables values at the optimum are the same ($x_i^* \;=\; 0$), therefore the average distance to the optimum is the same for all variables i, $d_i \;=\; d$ (but it varies with the problem). For each variable that is not correctly set, an average of d steps in the right direction is needed to reach the optimum. By summing the expected times of each beneficial step, one obtains the expected time to locate the optimum from a random starting point that has k optimal variables

$$T_k \;=\; \sum_{i=k}^{n-1} \frac{2dn}{n-i} \;. \tag{10}$$

T_k can now be averaged over all random starting points, which yields the expected convergence time of an SHC on a unimodal function,

$$T_{SHC} \;=\; \frac{1}{2^n}\sum_{k=0}^{n} C_n^k T_k \;=\; \frac{nd}{2^{n-1}}\sum_{k=0}^{n} C_n^k \sum_{i=k}^{n-1} \frac{1}{n-i} \;. \tag{11}$$

Estimated and measured convergence times are compared for in Figure 4.1. The linear dependency of the number of function evaluations in terms of the dimension n is correctly predicted.

4.2 Convergence time of the UMDA

In [6], the behavior of a UMDA with truncation selection is studied on the *Onemax* function, which maximizes the number of 1's in a binary string. Like in $\max A_{11}$, the function is separable,

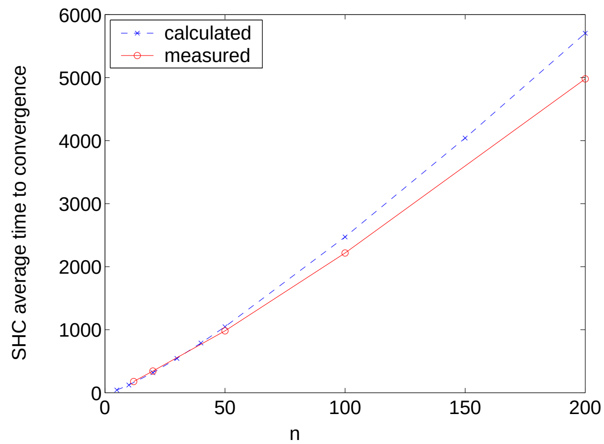


Figure 6: Estimated and measured convergence times for the max A_{11} problem.

and each variable has the same contribution to the objective function. If the population size m, is larger than a critical value m^*, it is shown in [6] that the expected number of generations to convergence[3], N_g, is

$$N_g \approx \mathcal{O}(\sqrt{n}) \tag{12}$$

The expected number of objective function evaluations to convergence is

$$N_f = N_g m^* \ . \tag{13}$$

No analytical expression for m^* was given in [6]. An approximation to m^*, $\widehat{m^*}$ is now proposed based on the initial random population sampling, and neglecting variable values lost during selection. The probability that a given variable value is not represented in the population is $((c-1)/c)^m$. The probability that the values making up the optimum, x^*, have at least a sample in the initial population is

$$P_{pop} = \left(1 - \left(\frac{c-1}{c}\right)^m\right)^n \ . \tag{14}$$

For a given P_{pop} (typically close to 1), the critical population size is estimated from Equation (14),

$$\widehat{m^*} = \frac{\ln(n/(1-P_{pop}))}{\ln(c/(c-1))} \approx \mathcal{O}(\ln(n)) \ . \tag{15}$$

From Equations (12) to (15), the order of magnitude of the number of evaluations to convergence is

$$N_f \approx \mathcal{O}(\sqrt{n}\ln(n)) \tag{16}$$

This order of magnitude agrees well with the UMDA convergence tests for the max A_{11} in Section 3.

[3]Following [6], convergence time is defined here as the time when $p(x^*, N_g) = 1$.

5 CONCLUDING REMARKS

In this paper, the performance of two stochastic algorithms, a stochastic hill climber and a univariate marginal distribution algorithm was compared. The influence of several factors: problem dimension, presence of constraints and multimodality was investigated in order to understand the conditions under which the sophistication of UMDA becomes advantageous. A study on a unimodal, separable problem revealed that the mechanisms underlying UMDA are asymptotically more efficient than random perturbations, but on a simple problem such as $\max A_{11}$, the superiority would only be apparent in very high dimensions ($n > 200$). However, if additional complexity is present, for example via constraints (constrained frequency maximization problem), or multimodality (strength maximization problem), mUMDA, which uses global distribution learning and local random search, proves more effective than SHC in finding the optimum.

Acknowledgment — This research was supported in part by Air Force grant 4502620.

REFERENCES

[1] Mühlenbein, H., Paaß, G.: From recombination of genes to the estimation of distributions, I. Binary parameters. Lecture Notes in Computer Science 1141: Parallel Problem Solving from Nature IV (1996) 178–187

[2] Baluja, S.: Population based incremental learning: a method for integrating genetic search based function optimization and competitive learning. Technical Report CMU-CS-94-163, Carnegie Mellon Univ., (1994)

[3] Mitchell, M., Holland, J.H., Forrest, S.: When will a Genetic Algorithm Outperform Hill Climbing. Advances in Neural Information Processing Systems **6**, Morgan Kaufmann Publishers, (1994) 51–58

[4] Garnier, J., Kallel, L.: Statistical distribution of the convergence time of evolutionary algorithms for Long-Path problems. IEEE-EC, **4**(1) (2000)

[5] He, J., Xin, Y.: From an Individual to a Population: An Analysis of the First Hitting Time of Population-Based Evolutionary Algorithms. IEEE Transactions on Evolutionary Computation **6** (2002) 495–511

[6] Mühlenbein, H., Mahnig, T., Rodrigez, A.O.: Schemata, distributions and graphical models in evolutionary optimization. J. of Heuristics **5** (1999) 215–247

[7] Pelikan, M., Goldberg, D.E., Sastry, K.: Bayesian Optimization Algorithm, Decision Graphs, and Occam's Razor. IlliGAL Report No. 2000020 (2000)