

04081 Abstracts Collection
Theory of Evolutionary Algorithms
— **Dagstuhl Seminar** —

Hans-Georg Beyer¹, Thomas Jansen², Colin Reeves³ and Michael D. Vose⁴

¹ FH Vorarlberg, A

`hans-georg.beyer@fhv.at`

² Univ. Dortmund, DE

`jansen@ls2.cs.uni-dortmund.de`

³ Coventry Univ., GB

`c.reeves@coventry.ac.uk`

⁴ Univ. of Tennessee, US

`vose@cs.utk.edu`

Abstract. From 15.02.04 to 20.02.04, the Dagstuhl Seminar 04081 “Theory of Evolutionary Algorithms” was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Evolutionary algorithms, genetic programming, co-evolution, run time analysis, landscape analysis, Markov chains

04081 Executive Summary – Theory of Evolutionary Algorithms

The 2004 Dagstuhl seminar on the theory of evolutionary algorithm was the third following seminars in 2000 and 2002 which had the same title. These Dagstuhl seminars are recognized within the evolutionary algorithm community as a biannual series of high quality meetings concerned with theoretical aspects of evolutionary computing. Therefore, there was great interest in participating and presenting ideas. The seminar had 49 participants from twelve different countries, namely eighteen from Germany, eleven from the United States, eight from Great Britain, three from the Netherlands, two from Canada and one each from Austria, Australia, France, Japan, Mexico, Russia, and Switzerland.

From these 49 researchers, 39 were able to give a presentation. The topics of the talks were as diverse as the current status of evolutionary computation theory. There have been talks on general aspects of modeling and analyzing

evolutionary algorithms, in particular talks on the pros and cons of infinite population approaches. Also different ways of classifying and analyzing landscapes defined by different problems and genetic operators have been presented. Many talks presented concrete analytical results on specific evolutionary algorithms and specific problems. Such results have been presented for both, continuous and discrete search spaces, for artificial problems or problem classes as well as for a typical combinatorial optimization problem. Recent trends like optimization of noisy objective functions, estimation of distribution algorithms, and coevolution have been topic of several talks as well as the well-established genetic programming. Other randomized search heuristics used for optimization have also been subject of talks, namely estimation of distribution algorithms and ant colony optimization.

The talks have initiated lively and partially controverse discussions. The special Dagstuhl atmosphere has helped a lot to develop a constructive atmosphere that improved mutual understanding and inspiration. The traditional hike on Wednesday afternoon and time spent together in the evenings or late at night let room for more personal discussions, too.

During the seminar the participants were introduced to the new Dagstuhl online proceedings by Jutta Huhse. The idea and advanced possibilities of the new system compared to the old reports was well received. As a result, some of the abstracts here are extended abstracts which are way more detailed than the abstracts of the previous two seminars on the theory of evolutionary algorithms.

Keywords: Evolutionary algorithms, genetic programming, co-evolution, run time analysis, landscape analysis, Markov chains

Joint work of: Beyer, Hans-Georg; Jansen, Thomas; Reeves, Colin; Vose, Michael D.

Evolution strategies, outliers, and evolutionary gradient search

Dirk V. Arnold (Dalhousie University, CDN)

This talk focuses on two different aspects that arise in connection with evolutionary algorithms in local and noisy optimization. First, previous results obtained under the assumption of Gaussian noise are extended to the case of Cauchy-distributed noise. This is of interest as Cauchy-distributed noise is much more appropriate for modeling certain types of noise observed in practice that are characterized by the frequent occurrence of outliers. The analysis can be summarized in the conclusion that the differences between the effects of Gaussian noise and Cauchy noise on the performance of evolution strategies (ES) are mostly quantitative rather than qualitative, and that the recommendations with regard to the choice of strategy variants and parameters made earlier remain valid.

In the second part, an analysis of the performance of evolutionary gradient search (EGS) is presented. EGS is a hybrid strategy that combines features of both ES and gradient strategies. It is found that ideally, EGS operates with comparatively crude approximations to the gradient, and that while it benefits from an effect similar to genetic repair, it does not benefit to the same degree that evolution strategies do. In the absence of noise, as expected, its more complete use of the information available gives EGS a performance advantage compared to the ES. In the presence of noise, their ability to operate with higher mutation strengths make the ES the preferred strategy.

On the Analysis of the Convergence of Self-Adaptive Evolution Strategies using the Theory of ϕ -irreducible Markov Chains

Anne Auger (Université Paris Sud, F)

The purpose of this talk is to thoroughly investigate the convergence properties of the $(1, \lambda)$ -Self-Adaptive Evolution Strategies. The $(1, \lambda)$ -ES, introduced by Rechenberg and Schwefel, generates λ offsprings from one parent by normal mutation, and the best of those λ offsprings becomes the parent of the next generation. The critical issue is to adjust the parameters of the normal mutation: the SA-ES use (log-normal) mutation to tune the standard deviations of the mutation of the object variables.

State-of-the-art theoretical results for Evolution Strategies deal with the particular case of the sphere function, and either derive estimations on the progress rate (but no convergence results) for SA-ES (H.-G. Beyer, 2001), or prove some convergence results for some variant of ES involving predefined step-size evolution (Yin et al. 1996) - whereas SA-ES are well-known by practitioners to be the best performing Evolution Strategies.

We prove, on the d -dimensional sphere function f , and for a class of Evolution Strategies that is more general than the usual $(1, \lambda)$ -SA-ES, a convergence result. Moreover, this convergence result is constructive, and allows us to derive an actual speed of convergence, that we are able to numerically estimate with theoretical bounds on the numerical estimates.

The mathematical techniques used for analyzing the convergence rely on the theory of Markov chains on continuous state space (see S.P. Meyn and R.L. Tweedie, 1993). Those mathematical techniques have been in particular developed for analyzing Markov Chains Monte Carlo algorithms, like Metropolis Algorithms or Gibbs sampler.

Using the powerful tools of Harris recurrent Markov chains, we are able to prove the existence of a critical size of population λ_0 , that can be computed from the different parameters of the algorithm, such that for any λ greater than λ_0 , $\frac{1}{n} \ln(f(\text{parent}_n))$ converges to some constant as n goes to infinity. This constant determines the convergence rate of the algorithm. It is also proved that an estimation of this constant along with error bars can be obtained with a Monte Carlo

method. Such Monte Carlo estimations allow us to verify well-known results by Schwefel or Beyer (see H.-G. Beyer, 2001) concerning the tuning of internal parameters of the algorithm - results that had been previously obtained through some estimations of the progress rate.

S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, New-York, 1993.

H.-G. Beyer, *The Theory of Evolution Strategies*, Springer, Heidelberg, 2001.

G. Yin, G. Rudolph and H.-P Schwefel, *Analysing $(1, \lambda)$ Evolution Strategy via Stochastic Approximation Methods*, *Evolutionary Computation*, 1996

Keywords: Convergence of real-valued Evolutionary Algorithms, Evolution Strategies, phi-irreducible Markov chains

Estimating the Steady State of CSA-ES on Ridge Functions

Hans-Georg Beyer (Universität Dortmund, D)

A simple method is presented for estimating the steady state distance R_∞ and the velocity φ_x by which the parental centroid travels along the ridge axis. The basic idea is to consider the ridge functions as a sphere model disturbed by noise, where the noise strength is equal to the mutation strength. Using a) the evolution criterion of the noisy sphere and b) the steady state progress rate formula of the CSA-ES, one obtains two equations which allow for the calculation of R_∞ and φ_x with the results $R_\infty \simeq \frac{1}{2d}$ and $\varphi_x \simeq \frac{\mu c}{2dN}$ for the parabolic ridge.

A similar calculation performed for the sharp ridge leads to the remarkable result that the CSA-ES exhibits premature convergence for $d > 1$, i.e., $R_\infty \rightarrow 0$, $x_{N+1} \rightarrow \text{const.}$; and for $d < 1$: $R_\infty \rightarrow \infty$, $x_{N+1} \rightarrow \infty$, i.e., the CSA-ES diverges as time goes to infinity. This behavior is actually observed in real CSA-ES runs. It indicates an application limit of Evolution Strategies with isotropic mutations and CSA step size adaptation.

Keywords: Evolutionary Algorithms Noisy Optimization Evolution Strategies Performance on Ridge Functions Progress Rate Analysis Dynamics of Evolution

Efficient sampling in the presence of noise

Jürgen Branke (Universität Karlsruhe, D)

Many real-world optimization problems are noisy in the sense that an individual's fitness is a random variable. This may cause the evolutionary algorithm to erroneously pick a bad individual which is perceived as superior due to the noise. Although the effect of noise can be reduced by averaging over multiple samples, sampling is computationally very expensive. We look at tournament selection in

the presence of Gaussian noise and propose to use sequential sampling techniques in order to reduce the effort required for sampling.

In sequential sampling, samples are drawn one at a time, until the observed fitness difference exceeds some threshold, while the threshold depends on the number of samples so far. We present and compare several ways to determine good sampling schedules and show that sequential sampling allows reducing the number of samples by between 30 and 80 percent compared to a simple schedule where each individual is sampled equally often.

Keywords: Evolutionary algorithm, tournament selection, noise, sequential sampling

Joint work of: Branke, Jürgen; Schmidt, Christian

Towards a Geometric Theory of Interactive Domains

Anthony Bucci (Brandeis Univ. Waltham, USA)

Coevolutionary algorithms are often applied in interactive domains, namely domains in which two or more entities can come together with an observable outcome. Examples include board games, where players play a game and we can see who wins; classifier induction, where classifiers classify test cases and we see if they do so correctly; and function optimization, where candidate functions are compared against a target function and we can see the errors. We suggest that an interactive domain can be fruitfully viewed as a space, and the problem of search characterized as a problem of discovering and navigating that space. We can characterize several pitfalls and ideals of coevolution research, such as intransitive cycles, overspecialization, and arms races, in a uniform geometrical language. In general, as is the case with fitness landscapes, a geometrical viewpoint offers intuitions about interactive domains which might not be available otherwise. We will survey two lines of work in which we have made progress in elaborating this suggestion. This geometrical viewpoint offers new interpretations of known pitfalls of coevolutionary algorithms. Looking forward, we give suggestions for improving algorithms or algorithm assessment based on the geometry of interactive domains.

Keywords: Coevolution, arms race, cycling, intransitivity, overspecialization, geometry

Using Price's Theorem for EA Analysis

Kenneth A. De Jong (George Mason Univ. - Fairfax, USA)

Although we have well-developed theoretical analyses for specific types of evolutionary algorithms (EAs), the theory for more general EA formulations is quite weak. However, the evolutionary biology community has developed more general models, but with different purposes in mind. In this presentation I will review a well-known result from that community, namely Price's theorem, and show how it can be helpful in understanding EAs at a more general level.

Keywords: Evolutionary algorithms, evolutionary biology, selection, reproductive variation

Joint work of: De Jong, Kenneth A.; Bassett, Jeffrey; Potter, Mitchell

Analysis of the (1+1) EA on a Noisy OneMax

Stefan Droste (Universität Dortmund, D)

In practical applications evaluating a fitness function is frequently subject to noise, i.e., the "true fitness" is disturbed by some random variations.

Evolutionary algorithms (EAs) are often successfully applied to noisy problems, where they have turned out to be particularly robust. Theoretical results on the behavior of EAs for noisy functions are comparatively very rare, especially for discrete search spaces. Here we present an analysis of the (1+1) EA for a noisy variant of OneMax and compute the maximal noise strength allowing the (1+1) EA a polynomial runtime asymptotically exactly. The methods used in the proofs are presented in a general form with clearly stated conditions in order to simplify further applications.

Keywords: Runtime analysis, (1+1) EA, noise

Comparing Evolutionary Algorithms to the (1+1) EA by Means of Stochastic Ordering

Anton Eremeev (Sobolev Institute of Mathematics - Omsk, RUS)

In this paper, we study the conditions in which the (1+1)-EA compares favorably to other evolutionary algorithms in terms of fitness distribution function and average fitness at given iteration as well as the average optimization time. Our approach is applicable when the reproduction operator of an evolutionary algorithm is dominated by the mutation operator of the (1+1)-EA, where the notion of domination is defined analogously to the one used in the framework of stochastic ordering (see e.g. Kamae, Krengel, O'Brien, 1977).

In the case of such domination, one can extend the lower bounds obtained for the expected optimization time of the (1+1)-EA to other EAs based on the same mutation operator. This method is exemplified on the sorting problem with HAM landscape and exchange mutation operator which was analyzed with (1+1)-EA in (Scharnow, Tinnefeld, Wegener, 2003).

An interesting situation where it is easy to implement a mutation operator dominating the reproduction, is identified by the monotonicity condition. We consider several examples where this condition holds and discuss the ways to relax it on the basis of computational experiments with the maximum independent set problems.

Kamae, T., Krengel, U., O'Brien, G.L. (1977). Stochastic inequalities on partially ordered spaces, *The Annals of Probability*, 5 (6), 899-912.

Scharnow, J., Tinnefeld, K., Wegener, I. (2003). Private communication.

Keywords: Evolutionary Algorithms, Stochastic Ordering, Stochastic Monotonicity, Hitting Time

Joint work of: Eremeev, Anton; Borisovsky, Pavel

Runtime Analyses for a Simple Multi-objective Evolutionary Algorithm

Oliver Giel (Universität Dortmund, D)

Evolutionary algorithms are not only applied to optimization problems where a single objective is to be optimized but also to problems where several and often conflicting objectives are to be optimized simultaneously. Practical knowledge on the design and application of multi-objective evolutionary algorithms (MOEAs) is available but well-founded theoretical analyses of the runtime can hardly be found. Laumanns, Thiele, Zitzler, Welz, and Deb (2002) have started such an analysis for two simple multi-objective evolutionary algorithms (SEMO and FEMO).

In this talk, the (expected) runtime of a variant of SEMO that searches globally is investigated.

It is proven that the expected runtime is $O(n^n)$ for all objective functions $\{0, 1\}^n \rightarrow R^m$.

For each d in $\{2, \dots, n\}$, a bicriteria problem such that the expected runtime is $\Theta(n^d)$ is presented. Bounds on the runtime and the expected runtime of the algorithm when applied to the problems LOTZ (Leading Ones Trailing Zeroes) and MOCO (Multi Objective Counting Ones) are derived.

Keywords: Runtime analysis, multi-objective evolutionary algorithms

Locality, Heritability and Heuristic Bias in Evolutionary Search

Jens Gottlieb (SAP - Walldorf, D)

The choice of representation and variation operators is essential for the successful application of evolutionary algorithms. We discuss three important aspects, which have strong impact on success or failure of evolutionary search: locality, heritability, and heuristic bias. Locality means that small genotypic changes cause small phenotypic changes. Without locality, evolutionary search becomes random search, which is typically ineffective.

Heritability covers properties of crossover, like the ability to preserve phenotypic properties that are identical in the parents, and the capability of creating innovative solutions differing from the parental phenotypes. Heuristic bias means that the genotype-phenotype mapping is biased by some heuristic. If chosen properly, such heuristic causes random sampling of the genotype space to produce phenotypes of high quality with higher probability. We introduce formal measures to characterize these three aspects of evolutionary search. Empirical results are presented for decoder-based evolutionary algorithms for the multidimensional knapsack problem. Locality, heritability and heuristic bias significantly affect the performance of evolutionary search and they explain success or failure of the chosen representation and variation operators.

Joint work of: Gottlieb, Jens; Raidl, Günther

The Covariance Matrix Adaptation (CMA) - Yet Another Estimation of Distribution Algorithm (EDA)?

Nikolaus Hansen (ETH Zürich, CH)

Consider a simplistic Estimation of Distribution Algorithm (EDA), where the search distribution is a multi-dimensional Gaussian distribution, parameterized by a mean and a covariance matrix. In each generation, from a sample of *selected* points, mean and empirical covariance matrix are calculated. These measures are used as mean and covariance matrix of the next generations distribution. In this simplistic EDA, two key issues can be identified:

(a) The population size is a sensitive parameter for the estimation of the covariance matrix.

(b) Premature convergence is intrinsic.

(a) is not surprising at all. For a small population size (namely smaller than 10 times the search space dimension) the estimation is deficient. For a larger population size the estimation becomes trustworthy, but the search progress (improvement of the objective function value per executed function evaluation) might be slow. We mention two measures to reduce the population size without

jeopardizing the reliability. 1) Restricting the model complexity (regularization) and 2) introducing a memory. While 1) is the usual method applied in EDAs, 2) is a key feature of the Covariance Matrix Adaptation (CMA).

(b) is not immediately obvious, but consider a linear fitness function.

Selection will reduce the expected variance in direction of the gradient (in directions orthogonal to the gradient selection is random and the expected variance stays constant). This result is independent of problem dimension and population size. Reducing the variance in gradient direction on a linear fitness function is clearly not desirable and brings about the danger of premature convergence.

Looking at this result in the light of optimal variances from well known evolution strategy theory, it is foreseeable that the population size must scale at least linearly with the problem dimension to circumvent premature convergence on non-linear fitness functions. In the CMA, the situation is slightly different, in that the covariance matrix of steps, not of points, is estimated. For selection ratios smaller than 0.5 the variance does not reduce in this case.

Nevertheless, to do reliable and fast search with small populations, the overall variance (expected step length) should be controlled separately from the covariance matrix.

Keywords: Estimation of distribution algorithm, CMA, covariance matrix adaptation

Selection-only Coevolution on NP-Hard Problems

Jeff Horn (Northern Michigan University - Marquette, USA)

It might be possible to use the complex, inter-species interactions of coevolution to attack NP-hard problems in an unusual, highly parallel manner. Recent work uses resource-based fitness sharing (RFS) to solve “shape nesting” problems, initially reported at Dagstuhl Seminar 02031. Selection, using the resource-based shared fitnesses, causes the population of shaped pieces to become distributed among non-overlapping niches. If a unique, maximum size set of non-overlapping (hence “cooperative”) niches exists in the initial population, selection under RFS should “find” it and drive the population distribution toward it. It seems that selection under RFS favors the maximum size clique in a “cooperation graph”, where each species is represented by a vertex, and each edge represents a pairwise cooperative relationship. Applying this approach to random graphs, such as 1993 DIMACS Max-Clique instances, we find performance comparable to that of other evolutionary algorithms. Extrapolating the basic scheme to other NP-hard domains, we have some ideas for formulating the optimization versions of 3SAT, Vertex Cover, X3C, Graph Coloring, Set Packing, 3DM, Bin Packing, and Set Cover. The essence of the approach to all of these problems is the use of a population to represent all members of a subset selection type of problem as distinct species, and then to use domain-specific information to create a measure of “species overlap”. RFS-based selection might then be used to select the largest

cooperative subset, thus acting as a heuristic for some NP-Hard problems. The nature of the computation, in which local, overlapping solutions compete and cooperate on a global level, seems to be similar to a few other approaches, e.g., PDI by Pramanick & Kuhl (1992) and Hopfield networks (Shrivastava & Reddy, 1992).

Keywords: Coevolution, cooperation evolution, NP, NP-hard, niching, selection, fitness sharing, resource sharing, optimization, species, niches, maximum clique, independent set, shape nesting

Recent Results on No-Free-Lunch for Optimization

Christian Igel (Ruhr-Universität Bochum, D)

The sharpened No-Free-Lunch-theorem (NFL-theorem) states that, regardless of the performance measure, the performance of all optimization algorithms averaged uniformly over any finite set F of functions is equal if and only if F is closed under permutation (c.u.p.). In this talk, some consequences of this theorem are summarized: The number of subsets c.u.p. can be neglected compared to the total number of possible subsets. In particular, problem classes relevant in practice are not likely to be c.u.p. The average number of evaluations needed to find a desirable (e.g., optimal) solution can be calculated independent of the optimization algorithm in certain scenarios. As the main result, the NFL-theorem is extended.

Necessary and sufficient conditions for NFL-results to hold are given for arbitrary distributions of target functions.

Keywords: No-Free-Lunch theorem

Rethinking Multi-Objective Evolutionary Algorithms

Yaochu Jin (HONDA Research Institute Europe - Offenbach, D)

This talk identifies a few general problems in the existing multi-objective evolutionary algorithms. 1) Multi-objectivity gives rise to problems to many adaptation and self-adaptation methods developed for single objective evolutionary algorithms. 2) It is not straightforward to maintain the non-dominated solutions in the population as it is often believed. 3) Many performance indices are misleading. Besides, it is also argued that adaptive representation and structure learning in multi-objective evolutionary algorithms are more important. This include: 1) Representing non-dominated solutions with a model rather than a set. 2) Using crossover methods that are able to learn the problem structure at hand. 3) Estimation of distribution algorithms are more suitable for solving multi-objective optimization problems. 4) Adaptive representation, such as a hybrid of binary and real-coded representations is very effective in improving the performance of multi-objective evolutionary algorithms.

Keywords: Multiobjective, Evolutionary algorithms, adaptive representation

Joint work of: Jin, Yaochu; Okabe, Tatsuya

Reversible functions have normal fitness

William Langdon (University College London, GB)

The distribution of reversible programs tends to a limit as their size increases.

For problems with a Hamming distance fitness function the limiting distribution is binomial with an exponentially small chance (but non zero) chance of perfect solution. Sufficiently good reversible circuits are more common. Expected RMS error is also calculated. Random unitary matrices may suggest possible extension to quantum computing. Using the genetic programming (GP) benchmark, the six multiplexor, circuits of Toffoli gates are shown to give a fitness landscape amenable to evolutionary search. Minimal CCNOT solutions to the six multiplexer are found but larger circuits are more evolvable.

Keywords: Genetic algorithms, genetic programming, fitness landscape, evolutionary computation, quantum computing, CCNOT, Toffoli, low power consumption

Genetic Drift in Finite Models of Evolutionary Games

Anthony Liekens (TU Eindhoven, NL)

For the analysis of the dynamics of populations that evolve strategies for games, it is common practice to assume infinitely large populations.

Infinite population models yield predictions of fixed points and their stability properties.

However, these models are not suited to anticipate the influence of genetic drift, caused by stochastic sampling in small populations. Instead, we propose Markov models of finite populations for the analysis of genetic drift in games. With these exact models, we can study the stability of evolutionary stable strategies, and measure the influence of genetic drift in the long run. We show that genetic drift can introduce significant differences in the expectations of long term behavior.

Keywords: Finite population models, infinite population models, evolutionary game theory, genetic drift, population size, mutation rate

On the strength of size limits in linear genetic programming

Nicholas Freitag McPhee (University of Minnesota - Morris, USA)

Bloat is a common and well studied problem in genetic programming. Size and depth limits are often used to combat bloat, but to date there has been little detailed exploration of the effects and biases of such limits.

In this paper we present statistical and theoretical analyses of the effect of size limits on variable length linear structures. Specifically, we examine the relationship between size limits and the average size of individuals in a population and define the notion of size limit strength. When a size limit is strong, the average size of a population converges to a relatively stable value which has a well defined relationship to the limit. When a size limit is weak, no such convergence occurs. The average size instead appears to perform a random walk within a bounded region. We use schema theory to show that limit weakness is likely a result of sampling error.

Joint work of: McPhee, Nicholas Freitag; Jarvis, Alex; Fussell Crane, Ellery

The $(1, \lambda)$ -Quality Gain in Noisy Environments

Silja Meyer-Nieberg (Universität Dortmund, D)

Noise is a common phenomenon in optimization tasks that have to deal with real-world applications.

Evolution Strategies and other Evolutionary Algorithms are supposed to be especially apt to cope with noisy information due to the use of a population of candidate solutions.

However, since noise does deceive the information obtained from the objective function, even population based algorithms are degraded in their performance and the convergence velocity is reduced.

We will consider the quality gain of on arbitrary fitness functions which describes the expected change of the fitness function of the parent population from one generation to the next.

In the case of an $(1, \lambda)$ -ES, the quality gain in noisy environments is given as $\overline{\Delta Q}_{1, \lambda} \simeq \frac{S_Q^2}{\sqrt{S_Q^2 + \sigma_\epsilon^2}} c_{1, \lambda} + M_Q$, where M_Q and S_Q^2 are the expected value and the standard deviation of the local quality change Q and σ_ϵ is the variance of the noise. The constant $c_{1, \lambda}$ is the so-called progress coefficient (see, Beyer 2001, p.74).

The expression obtained for the quality gain was applied to several fitness functions leading to good results as long as the mutation strength is sufficiently small and the dimensionality of the search space sufficiently high.

Joint work of: Beyer, Hans-Georg; Meyer-Nieberg, Silja

Evolutionary Algorithms on Complex Fitness Landscapes

Adam Prügél-Bennett (Univ. of Southampton, GB)

In this talk I pose the question: What features of hard optimisation problems determine the performance of evolutionary algorithms? I describe the landscape of two NP-Hard problems, the Ising perceptron and MAX-K-SAT.

I explore some simple models that try to capture some of the features of hard problems. In particular, I concentrate on the problem of symmetry breaking in finite populations. I present some simulations and mathematical analysis showing how evolutionary algorithms behave on these types of landscapes.

Partitioning Landscapes

Colin Reeves (University of Coventry, GB)

Fitness landscapes are induced by the use of a neighbourhood operator or function on a search space for the points of which a 'fitness function' is defined. Methods of measuring the smoothness or ruggedness of a landscape have been proposed, but are typically difficult to calibrate.

In this talk, I discuss the idea of 'neighbourhood ranks' (NR) as a means of partitioning the points on a landscape. This provides a way of categorizing landscape types, and also (by using some number-theoretic concepts) of counting them. For example, we can estimate the number of possible unimodal landscapes for a particular search space size. Finally, by measuring empirical properties (such as moments) of the NR-distribution, calibration of landscape properties becomes a more feasible proposition.

Keywords: Fitness landscape, neighbourhood rank

Evolution in Abstract Domains

Jonathan E. Rowe (University of Birmingham, GB)

A population over a finite search space can be described by a vector representing a distribution over the search space. We can think of the action of a single generation of an evolutionary algorithm as creating a new distribution from the population, from which the new population is created by sampling. In the case of a finite search space, we have the nice property that the set of distributions forms a compact convex subset of real Euclidean space. Thus continuous maps from this set to itself are guaranteed to have fixed-points.

When moving to infinite search spaces, the set of distributions is no longer compact. Thus sequences of populations can diverge, as seen for example, when the evolution of a GP run bloats. I argue that, in some cases, it may be interesting to examine where such a divergent sequence is going. For example, the evolution of a program to generate the digits of π (using Koza style representation) is likely to bloat, producing bigger and bigger trees. But as these trees get bigger, they may be producing slightly more accurate approximations to π . It would therefore be of interest to be able to prove that in some sense the sequence of outputs of such programs is in fact converging to π , and that therefore we can get as close as we like to our target by running the evolution for long enough. The problem, of course, is that real numbers like pi can require infinite size trees

to produce, and these cannot actually exist in our computers. We are therefore in the following position: we have a set of finite objects, which our populations can contain, and we have a set of ideal, infinite objects which our population cannot contain, and yet sequences of finite objects seem to be converging to them.

Mathematically, we can describe this situation as follows. Consider the search space of finite objects as having the discrete topology. We now add our “points at infinity”. Now define a partial order on the whole set by defining, for any two objects, x and y , the concept of the “largest object that they both have in common”, which we denote $x?y$.

We then define our partial order by: $x?y$ if $x = x?y$.

For example, if the set were strings of binary numbers (and we have added the infinite length strings), then one string comes before another in the partial order if it is a prefix of the latter. Now we add some more open sets to our topology to include the infinite objects. These are the so-called Scott-open sets, which are generated by all the upper sets of the partial order. The whole topology is known as the Lawson topology. It is Hausdorff and compact. By considering distributions over this set as the space of Borel measures on this topology, it is known that this set is itself compact. In effect, sequences of populations that diverge (in terms of the finite objects they contain) now have somewhere to “diverge to”: they converge to distributions containing infinite objects.

Using this framework, we can start to expand our ideas about what search spaces we can apply evolutionary algorithms to. Program trees and variable length strings are just two examples. Others include:

- intervals of real numbers (ordered by reverse inclusion)
- functions
- shapes (approximated, for example, by polygons)

The partial order construction may even give us a natural way to think about crossover in such domains, since “respectful” crossovers are supposed to preserve elements that the parents have in common. Thus crossover between two variable length strings would maintain in the offspring the longest prefix common to both parents.

Keywords: Populations dynamics compactness infinite search spaces

The Crossover Landscape for Subset Selection Problems

J. David Schaffer (Philips Labs. - Briarcliff Manor, USA)

The subset selection problem is to choose a subset of size S from a pool of P objects that is optimal in some sense. Prior work on GA’s for this problem by Radcliffe has explored crossover operators for their properties of respect and assortment. We provide a set of crossover operators including variations on Radcliffe’s that differ in systematic ways on the dimensions of positive and negative respect (a new term we introduce). Having thus defined the crossover landscape,

we present empirical results using a subset selection problem generator that permits controlled variation of the degree of epistasis in the problems. The results show that for low epistasis problems, positive respect is of key importance, but that as epistasis increases, a degree of mutation (violation of negative respect) is needed to achieve superior performance. Furthermore, the mutation schedule is important. The best schedule among our alternatives was a convergence constrained mutation with a fixed probability of 5

Keywords: Genetic algorithms, crossover, positive respect, negative respect, subset selection, landscape

Joint work of: Schaffer, J. David, Larry Eshelman, Keith Mathias

Population-based Iterated Local Search

Dirk Thierens (Utrecht University, NL)

Iterated local search (ILS) is a powerful meta-heuristic algorithm applied to a large variety of combinatorial optimization problems.

Contrary to evolutionary algorithms (EAs) ILS focuses only on a single solution during its search. EAs have shown however that there can be a substantial gain in search quality when exploiting the information present in a population of solutions. In this paper we explore the use of a population for ILS. First, we define the general form of the resulting meta-heuristic, called population-based iterated local search (PILS). Then we discuss a specific instantiation of the PILS algorithm that uses a 2-ary crossover operator to reduce the dimensionality of the neighbourhood explored by the ILS.

Structured Search, Crossover, and EDAs

Marc Toussaint (Ruhr-Universität Bochum, D)

Evolution can be understood as a process in the space of distributions over search space. In this process, selection is a source of information (=decrease of entropy) and mutation and crossover a source of entropy (=exploration). Besides entropy, one can analyze the decrease and increase of mutual information in this process, which captures the structural dependencies between solution components.

Depending on the problem, selection might be a natural source of mutual information. One can prove that mutation and crossover (in the case of a direct encoding) can only destroy such mutual information – thus selection is the only source of mutual information. In contrast, a Simple Heuristic Search scheme comparable to EDAs realizes an exploration that can increase mutual information – roughly by estimating the structure present in the parent population and adopting this structure for the search distribution.

A Discipline of Evolutionary Programming: Genetic Fitness Optimization Using Rapidly Mixing Markov Chains

Paul Vitanyi (CWI - Amsterdam, NL)

Genetic fitness optimization using small populations or small population updates across generations generally suffers from randomly diverging evolutions.

We propose a notion of highly probable fitness optimization through feasible evolutionary computing runs on small size populations. Based on rapidly mixing Markov chains, the approach pertains to most types of evolutionary genetic algorithms, genetic programming and the like. We establish that for systems having associated rapidly mixing Markov chains and appropriate stationary distributions the new method finds optimal programs (individuals) with probability almost 1. To make the method useful would require a structured design methodology where the development of the program and the guarantee of the rapidly mixing property go hand in hand. We analyze a simple example to show that the method is implementable. More significant examples require theoretical advances, for example with respect to the Metropolis filter.

Note: This preliminary version may deviate with respect to typos, minor corrections, from the corrected final published version in:

Theoret. Comp. Sci., 241:1-2 (2000), 3–23.

Keywords: Genetic Algorithms, bounded populations, average run versus real runs, almost sure convergence, rapidly mixing Markov Chains, Metropolis filter, Kolmogorov complexity

Coarse Graining

Michael D. Vose (University of Tennessee, USA)

The simple genetic algorithm's behavior is determined by its transition matrix, which in turn is determined by its infinite population model. Given an equivalence relation over the search space, there is a corresponding aggregated Markov chain which faithfully captures stochastic behavior (of the original chain) if its infinite population model makes a particular diagram commute. This motivates the investigation of commutative diagrams involving infinite population models, or, more generally, involving differentiable maps.

A necessary and sufficient condition is given for the commutativity of the relevant kinds of diagrams. These conditions—and their corresponding proofs—establish a link between differential calculus and the coarse graining of Markov chains.

Most of the motivational results are contained in my book, so the uploaded draft paper deals only with the necessary and sufficient condition and some simple examples.

Keywords: Coarse graining

Evolutionary Algorithms, Randomized Local Search, and the Maximum Matching Problem

Ingo Wegener (Universität Dortmund, D)

There is a growing interest in the analysis of the behavior of randomized search heuristics like randomized local search, simulated annealing and evolutionary algorithms on combinatorial optimization problems. This research direction is motivated by the many applications of these heuristics in real-life optimization.

The common idea is to understand why these heuristics are often quite successful.

This is done by the analysis of the expected optimization time either for the worst-case instance, for random instances, or for semi-random instances. Here, the behavior of randomized local search and a simple and fundamental evolutionary algorithm on the well-known maximum matching problem is analyzed.

Both are polynomial-time randomized approximation schemes although they do not employ the idea of augmenting paths. They find a maximum matching in expected polynomial time for simple graphs but their exist bipartite graphs with a degree bounded by 3 where the expected optimization time grows exponentially.

Keywords: Evolutionary Algorithms, Randomized Local Search, Maximum Matchings, Polynomial-time Randomized Approximation Scheme

Tracking problems and the population size

Karsten Weicker (Universität Stuttgart, D)

$(1, \lambda)$ -evolutionary algorithms are investigated when applied to tracking problems with linear dynamics and low dimensionality. There are other results available for ES by Arnold and Beyer and bitflipping on a hypercube by Droste. On a $Z \times Z$ search space the search process is modeled using a Markov chain. This is done both exact as well as for a worst-case by collapsing all states with the same distance to the optimum. The first model is used for computations of the probability distribution concerning the position of the best individual (within a fixed number of generations) and the second for the computation of fixpoint distributions as lower bounds. The framework enables the investigation of several pdf's for the mutation operator (including the step-size parameter) as well as the population size. Since we assume that the time resource is very limited, the strength of the dynamics $s = \alpha\lambda + \beta$ depends on the number of offspring individuals (parameterized by α and β). Where in general an increase in the population size reduces the distance to the optimum, in this scenario there is an optimal population size in the range of 10–15. In the bottom line, high severity values are accepted if there is a strong influence of the population size on the severity. This is sensible since – with respect to the tracking accuracy – the positive influence

of an increasing population size decreases exponentially where the negative influence of increasing dynamics is linear. In fact, experiments using an evolution strategy with fixed, non-adaptive mutation affirms that the optimal population size is independent of α . Altogether the study of the abstract problem has proven to be useful and the results appear to be transferable quantitatively to other representations and operators.

Analyzing Cross-Population Epistasis in Cooperative Coevolutionary Algorithms

R. Paul Wiegand (George Mason Univ. - Fairfax, USA)

Increasingly, cooperative coevolutionary algorithms (CCEAs) are being applied to complex optimization tasks. Historically, research investigating collaboration methodology and problem difficulty for these systems have focused on the existence or absence of non-linear relationships between the populations. However, recent research has suggested that the issues surrounding cross-population epistasis, problem difficulty, and collaboration are more complicated than simply whether or not such relationships exist. To understand what makes problems challenging for the CCEA, and what demands more sophisticated collaboration methods, researchers will need to understand the nature of the cross-population epistatic linkages themselves. But what theoretical tools can be harnessed to answer such questions?

I propose two different methods for understanding these relationships: an analysis of the game-theoretic properties of certain types of problems and a multi-dimensional Walsh analysis of the cooperative rewards matrix. Early observations and results will be discussed in addition to the advantages and disadvantages of these techniques. Ideas for extensions and alternatives to these methods are encouraged.

Keywords: Cooperative coevolution, cross-population epistasis

An Analysis of the $(\mu+1)$ EA on Simple Pseudo-Boolean Functions

Carsten Witt (Universität Dortmund, D)

In this talk, a rigorous analysis of the $(\mu+1)$ EA on pseudo-Boolean functions is presented. For three example functions well-known from the analysis of the $(1+1)$ EA, bounds on the expected runtime and success probability are derived. For two of these functions, upper and lower bounds on the expected runtime are tight, and the $(\mu+1)$ EA is never more efficient than the $(1+1)$ EA. Moreover, all lower bounds grow with μ . On a more complicated function, however, a small increase of μ provably decreases the expected runtime drastically.

For the lower bounds, a novel proof technique is developed. The stochastic process creating family trees of individuals is investigated and relationships with well-known models of random trees, e. g., uniform random recursive trees, are established. Thereby, a known theory on random trees is transferred to the analysis of EAs.

Moreover, generalizations of the technique are applicable to more complex population-based EAs.

Discovering the structure of real-valued functions on binary strings

Alden Wright (Univ. of Montana - Missoula, USA)

This work addresses the problem of discovering the structure of a function from fixed length binary strings to the nonnegative reals when the function is given as a black box. The function is assumed to be a sum of component functions, where each component function depends on at most k bits (where k is less than the string length). An algorithm is given that finds the complete structure of the given function by sampling function values. Under the assumption that k is constant and that the number of component functions grows linearly with the string length, the complexity of this algorithm is shown to be $O(L^2 \log L)$ function evaluations where L is the string length.

Keywords: Epistasis "genetic algorithm" Walsh "black box"

Joint work of: Wright, Alden; Heckendorn, Robert