Report from Dagstuhl Seminar 13271

Theory of Evolutionary Algorithms

Edited by

Benjamin Doerr[1], Nikolaus Hansen[2], Jonathan L. Shapiro[3], and L. Darrell Whitley[4]

1 **MPI für Informatik – Saarbrücken, DE,** doerr@mpi-inf.mpg.de
2 **INRIA Saclay – Orsay, FR,** hansen@lri.fr
3 **University of Manchester, GB,** jls@cs.man.ac.uk
4 **Colorado State University, US,** whitley@cs.colostate.edu

Abstract

This report documents the talks and discussions of Dagstuhl Seminar 13271 "Theory of Evolutionary Algorithms". This seminar, now in its 7th edition, is the main meeting point of the highly active theory of randomized search heuristics subcommunities in Australia, Asia, North America and Europe. Topics intensively discussed include a complexity theory for randomized search heuristics, evolutionary computation in noisy settings, the drift analysis technique, and parallel evolutionary computation.

Seminar 30. June to 05. July, 2013 – www.dagstuhl.de/13271

1998 ACM Subject Classification G.1.6 Optimization, F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases evolutionary algorithms, optimization, search heuristics, algorithms, artificial intelligence

Digital Object Identifier 10.4230/DagRep.3.7.1

Edited in cooperation with Rachael Morgan

1 Executive Summary

Benjamin Doerr
Nikolaus Hansen
Jonathan L. Shapiro
L. Darrell Whitley

Evolutionary algorithms (EAs) are stochastic optimization methods that are based on principles derived from natural evolution. Mutation, recombination, and selection are iterated with the goal of driving a population of candidate solutions toward better and better regions of the search space.

In recent years, new methods have been developed at a rapid pace. Some of the advancements for continuous optimization methods have been enabled by focusing on how evolutionary algorithms can be compared and contrasted to more traditional gradient based methods. Arguably, evolutionary algorithms are one of the best methods now available for derivative-free optimization (DFO) on higher dimensional problems.

Another area of rapid recent advancement is in the area of run-time analysis for evolutionary algorithms applied to discrete optimization problems. Here, some techniques could be successfully borrowed from traditional algorithm analysis, but many new techniques were necessary to understand the more complicated stochastic processes arising from nature-inspired algorithms.

EA theory has gained much momentum over the last few years and has made numerous valuable contributions to the field of evolutionary computation. Much of this momentum is due to the Dagstuhl seminars on "Theory of Evolutionary Algorithms", which has become the leading meeting for EA theorists in the world.

Specific Topics

This year, the following topics had the particular attention of organizers, speakers both of overview and specialized talks, and participants of the breakout sessions (also called "working parties" or "working groups" in other Dagstuhl seminars). A brief summary of the breakout sessions can be found in Section 4.

Advanced Runtime Analysis Methods. One difficulty common to the analysis of most randomized search heuristics is that, while in principle these are nothing more than randomized algorithms, their particular nature disallows the use of many methods used in the classical analysis of the randomized algorithms community. The particular difficulties include dealing with populations (instead of a single search point as in other local optimizers) or recombination (instead of mutation only, which creates a search point close to the parent). Both the fitness level method and various variants of the drift analysis method were greatly improved in the last three years to cope with these difficulties. Also, the fixed budget view on runtime analysis was recognized as an alternative way of analyzing the performance of randomized search heuristics, and may better reflect performance indicators used by practitioners.

Complexity Theory for Randomized Search Heuristics. Complexity theory is one of the corner stones of classical computer science. Informally speaking, the *black-box complexity* of a problem is the number of fitness evaluations needed to find its solution. Unfortunately, it turns out that some notoriously hard problems like the clique problem in graphs have a ridiculously small black-box complexity. In their 2010 GECCO award winning paper, Lehre and Witt presented a promising way out of this dilemma. They introduced a restricted version of black-box complexity that on the one hand still covers most known evolutionary approaches, but on the other hand forbids the counter-intuitive tricks that led to the undesired results in the first approach. Following up on this work, several variants of black-box complexity have been suggested. During the seminar, in particular during the breakout session on this topic, these were intensively discussed, new variations have been proposed, both from the theory perspective and from practitioners, and a new approach was presented explaining how to gain new and better evolutionary algorithms from black-box complexity results.

Theory of Natural Evolutionary Algorithms. Recently, the idea of conducting a natural gradient descent in the space of sampling probability distributions has been introduced in evolution strategies. The idea offers a very principled design technique for search algorithms that sample from a parameterized distribution. Comparable to classical deterministic optimization, an iterated gradient descent is performed on the distribution parameters. The remarkable difference is that the curvature information on this space is known a priori. A natural descent that is based on the inner product from the Fisher information matrix uses this curvature and is comparable to a Newton method. This new and promising idea is lesser-known and largely unexploited for evolutionary computation. This is a completely new topic for this seminar series, but it is related to previous work on Covariance Matrix Adaptation.

Theory for Multi-Objective Optimization. One of the most explosive areas of growth both within evolutionary algorithms and in derivative-free optimization is multi-objective optimization. This is because good evolutionary algorithms now exist that can cover complex Pareto fronts for 2 to 12 objectives. This gives practitioners a much more informative view of the trade-offs that are possible when facing a multi-objective decision, and can also reveal trade-offs that otherwise would never be seen: for example if we are wishing to minimize cost and maximize quality, there can be “knees” at specific locations on the Pareto front where one might dramatically improve quality while incurring only a slight increase in cost. This is why multi-objective optimization methods that “map” the Pareto front are exciting. Yet, there is not a great deal of work on the theory of multi-objective optimization. Evolutionary algorithms are the method of choice for derivative-free multi-objective optimization and there is a great need to bring together theoreticians who are interested in evolutionary algorithms and those practitioners who are developing multi-objective optimization methods. This was another new topic for this seminar series.

Landscape Analysis. Landscape Analysis is an old idea: one should be able to compute features of a search space that can be used to guide search. One of the problems is that the kinds of metrics that one might wish to know about usually take exponential time to compute exactly. However, recent work has shown that some NP-hard problems (TSP, Graph Coloring, MAXSAT) can be decomposed to the point that Fourier methods can be used to exactly compute statistic moments of the search space (and subspaces of the search space) in polynomial time; these computation normally require exponential time for arbitrary problems. How can this information be used to guide the search, and to potentially replace heuristics with more exact information? New results in this area open new opportunities for exploration at this seminar series.

2 Table of Contents

3 Overview of Talks

3.1 Linear Convergence of the Isotropic ES via a Continuous Time Approximation

Youhei Akimoto (Shinshu University, JP)

An evolution strategy with isotropic Gaussian distribution derived from the information-geometric optimization (IGO) framework is studied on a composite of any strictly increasing function and a convex quadratic function and on a composite of any strictly increasing function and a twice continuously differentiable function. By extending the so-called ordinary differential equation (ODE) method, which is usually employed in the theory of stochastic approximation, we prove the linear convergence of the evolution strategy toward the global or local optimum.

3.2 The Dynamical Systems Approach Applied to the Analysis of Evolution Strategies for Constrained Oprimization

Dirk V. Arnold (Dalhousie University, CA)

We present a concise review of the dynamical systems approach to the analysis of adaptive optimization algorithms. We then discuss results obtained using that approach to study the behaviour of evolution strategies for constrained optimization. As test problem classes we consider a linear problem with a single linear constraint as well as a linear problem with a conically constrained feasible region for which the optimal solution lies at the cone's apex. The interaction of cumulative step size adaptation with two simple constraint handling techniques is examined, and significant differences in the results obtained are analyzed.

3.3 Sampling from Discrete Distributions

Karl Bringmann (MPI für Informatik – Saarbrücken, DE)

In a classic sampling problem we are given probabilities $p_1, \dots, p_n$(summing up to 1) and want to sample from the input distribution. This problem comes up in evolutionary algorithms when we want to select random parents for crossover proportional to their fitness or rank. We review a classic solution by Walker [1] building in linear time a data structure that allows to sample in constant time. Moreover, we give an overview of recent improvements and generalizations.

References

1 A. J. Walker. New fast method for generating discrete random numbers with arbitrary distributions. *Electronic Letters*, volume 10, pages 127–128, 1974.

3.4 What Information Can We Obtain Using Landscape Theory?

Francisco Chicano (University of Malaga, ES)

Joint work of Chicano, Francisco; Sutton, Andrew M.; Whitley, L. Darrell; Alba, Enrique

Landscape Theory is a mathematical framework with the purpose of getting some insight of Combinatorial Optimization Problems. Using Landscape Theory we can compute the Fitness-Distance Correlation [2], the expected value of the fitness function after a bit-flip mutation [4, 1] and the expected fitness after a uniform crossover [3]. In this talk we provide details on these expressions and present a new result: the computation of the probability distribution of the fitness after mutation. With this result we are able to compute the expected first hitting time of a (1+lambda) EA, which is probably the first connection between Landscape Theory and Runtime Analysis.

References

1 Francisco Chicano and Enrique Alba. Exact computation of the expectation curves of the bit-flip mutation using landscapes theory. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 2027–2034, 2011.

2 Francisco Chicano and Enrique Alba. Exact computation of the fitness-distance correlation for pseudoboolean functions with one global optimum. In *Proceedings of EvoCOP*, volume 7245 of *LNCS*, pages 111–123. Springer, 2012.

3 Francisco Chicano, Darrell Whitley, and Enrique Alba. Exact computation of the expectation curves for uniform crossover. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 1301–1308, 2012.

4 Andrew M. Sutton, Darrell Whitley, and Adele E. Howe. Mutation rates of the (1+1)-EA on pseudo-boolean functions of bounded epistasis. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 973–980, 2011.

3.5 Updates from the Black-Box

Carola Doerr (University Paris-Diderot, FR)

Joint work of Doerr, Benjamin; Doerr, Carola; Ebel, Franziska
Main reference B. Doerr, C. Doerr, F. Ebel, "Lessons From the Black-Box: Fast Crossover-Based Genetic Algorithms," in Proc. of 15th Annual Conf. on Genetic and Evolutionary Computation (GECCO'13), pp. 781–788, ACM, 2013.
URL http://dx.doi.org/10.1145/2463372.2463480

Black-box complexity aims at establishing a complexity theory for evolutionary algorithms and other randomized search heuristics. We often observe that black-box optimal algorithms have a smaller runtime than the search heuristics that we regard. In contrast to those algorithms, evolutionary algorithms typically do not benefit from search points that are inferior to the current-best solution ("best-so-far solution"). In this talk we present a class of new genetic algorithms that are inspired by this paradigm. We prove that, for suitable parameter choices, the expected runtime of these algorithms is $o(n \log n)$ on OneMax. This is the first time that an asymptotic improvement for OneMax can be achieved by using crossover operators. The talk is based on joint work with Benjamin Doerr and Franziska Ebel.

3.6 Non-Elitist Genetic Algorithm as a Local Search Method

Anton V. Eremeev (Sobolev Institute of Mathematics – Omsk, RU)

Main reference A.V. Eremeev, "Non-elitist genetic algorithm as a local search method," arXiv:1307.3463v2 [cs.NE], 2013.
URL http://arxiv.org/abs/1307.3463v2

In this paper, the non-elitist genetic algorithms with tournament selection (denoted by GA in what follows) are compared to the local search methods on the class of the NP optimization problems [1]. It is assumed that the fitness function of a GA is identical to the objective function on the set of feasible solutions and the infeasible solutions are penalized. The operators of crossover and mutation are supposed to be efficiently computable randomized routines. The population size and the tournament size may depend on problem instance. In many well-known NP optimization problems the set of feasible solutions is the whole search space. We show that the GA with suitable parameters tuning first reaches a local optimum to such problems in O(1) iterations on average. The population size and the tournament size are assumed to grow linearly in $1/s$ and $\log(m)$, where s is a lower bound on the probability to transform any given solution x into any given solution in the neighborhood of x, and m is the number of non-optimal levels of objective function. Often some termination condition is used to stop a GA when a solution of sufficient quality is obtained or because the population is trapped in some unpromising area. To incorporate the possibility of restarting the search, we also consider an Iterated GA, where the basic GA is terminated and independently initialized every time after a given number of iterations. It is shown that if the NP optimization problem and the value $1/s$ are polynomially bounded and the initial population always contains a feasible solution, then the Iterated GA with suitable choice of parameters first visits a local optimum on average in polynomially bounded time. Besides that, it is proven that any problem from the class of problems with guaranteed local optima (GLO) [1] may be approximated within a constant ratio by means of the Iterated GA with bitwise mutation operator in polynomial time on average.

References

1 G. Ausiello, M. Protasi. Local search, reducibility and approximability of NP-optimization problems.*Information Processing Letters*, volume 64, pages 73–79, 1995.

3.7 Parameterized Average-Case Complexity of the Hypervolume Indicator

Tobias Friedrich (Universität Jena, DE)

Joint work of Bringmann, Karl; Friedrich, Tobias;
Main reference K. Bringmann, T. Friedrich, "Parameterized Average-Case Complexity of the Hypervolume Indicator," in Proceedings of 15th Annual Conf. on Genetic and Evolutionary Computation (GECCO'13), pp. 575–582, ACM, 2013.
URL http://dx.doi.org/10.1145/2463372.2463450

The hypervolume indicator (HYP) is a popular measure for the quality of a set of n solutions in $\mathbb{R}^d$.We discuss its asymptotic worst-case runtimes and several lower bounds depending on different complexity-theoretic assumptions. Assuming that P $\neq$ NP, there is no algorithm with runtime poly(n, d). Assuming the exponential time hypothesis, there is no algorithm with runtime $n^{o(d)}$.In contrast to these worst-case lower bounds, we study the average-case complexity of HYP for points distributed i.i.d. at random on a d-dimensional simplex. We present a general framework which translates any algorithm for HYP with worst-case runtime $n^{f(d)}$ to an algorithm with worst-case runtime $n^{f(d)+1}$ and fixed-parameter-tractable (FPT) average-case runtime. This can be used to show that HYP can be solved in expected time $\mathcal{O}(d^{d^2/2}\, n + d\, n^2)$,which implies that HYP is FPT on average while it is W[1]-hard in the worst-case. For constant dimension d this gives an algorithm for HYP with runtime $\mathcal{O}(n^2)$ on average. This is the first result proving that HYP is asymptotically easier in the average case. It gives a theoretical explanation why most HYP algorithms perform much better on average than their theoretical worst-case runtime predicts.

3.8 Clustering as a Benchmark Optimization Problem

Marcus Gallagher (The University of Queensland, AU)

In recent years there has been increasing focus on improving the quality of experimental evaluations and comparisons of continuous evolutionary algorithms and metaheuristics. A good example of this is the Black-Box Optimization Benchmarking (BBOB) test problem suite and workshops. However, there is still alack of benchmark problems that are representative of real-world optimization problems. In this presentation, I discussed the data clustering problem in the 2-D Euclidean plane as potentially a very useful source of benchmark continuous optimization problems. Clustering is a fundamental problem in data mining and is also closely related to location-allocation problems from operations research. From a theoretical point of view clustering is an NP-hard problem and problem instances in practice are expected to present challenging features for algorithms. From a benchmarking point of view, clustering problems are unconstrained, simple to generate and have many properties that are desirable in benchmark functions. Random instances can be generated by generating datasets from a specified probability distribution. As an example, results were presented for clustering problems based on uniformly random data in the unit square. Using the standard k-means clustering algorithm, results confirm that problem instances have a large number of local optima and are challenging for black-box algorithms to solve. In addition,there appears to be a regime for the number of data points (n) and the number of cluster centers (k) where generated problems where k-means takes much longer to converge. An explanation of this observation is an open problem.

3.9 On the Analysis Approach of CSA on the General Ellipsoid Model

Michael Hellwig (Fachhochschule Vorarlberg, AT)

Joint work of Hellwig, Michael; Hans-Georg Beyer

The optimization behavior of evolution strategies (ES) applying cumulative step size adaptation (CSA) on convex-quadratic functions (referred to as ellipsoid model) is investigated.

The analysis considers $(\mu/\mu_I, \lambda)$-CSA-ES using intermediate recombination and isotropic mutations. Introducing the asymptotically exact quadratic progress rate the talk gives insight into a new analysis approach for the dynamical ES system. The dynamical system is modelled by a set of non-linear difference equations which describe the component-wise progress towards the optimizer, the mutation strength control as well as the change in the search path vector within the CSA. In the steady state the system can be transformed into an eigenvalue problem. Solving the eigenvalue problem allows for predictions of the ES systems behavior in the vicinity of the steady state. While the analysis is not yet complete the approach is validated by the good agreement of first theoretical predictions with experimental results.

3.10 Fixed Budget Computations: Why, How and What?

Thomas Jansen (Aberystwyth University, GB)

Run time analysis of evolutionary algorithms (EAs) is currently a dominating topic in the theory of EAs. It can be traced back 20 years, the last 15 years have seen a large number of publications that jointly have developed a wealth of analytical methods, surprising insights, and a better understanding of many aspects of EAs. However, these results are hardly appreciated outside the theory community. We argue that one reason is the lacking connection to the way EAs are applied in practice. Run time analysis is about the time needed to find an optimal or approximately optimal solution, something that is often not achieved in practice. In applications, evolutionary algorithms are more often stopped after some time. The assessment of the performance of EAs based on the expected function value after a fixed budget of function evaluations is closer to this way of applying EAs. We suggest this method as an alternative, discuss why we believe it offers a more useful perspective, how results in this framework can be obtained, and what the current state of the art is. The talk is mainly an open invitation to join us in the development of this novel research direction. This talks presents joint work with Benjamin Doerr, Carsten Witt, and Christine Zarges.

3.11 Some Runtime Analyses on the (1+λ) EA

Marvin Künnemann (MPI für Informatik – Saarbrücken, DE)

Joint work of Doerr, Benjamin; Künnemann, Marvin;
Main reference B. Doerr, M. Künnemann, "Royal road functions and the (1 + λ) evolutionary algorithm: Almost no speed-up from larger offspring populations," in Proc. of the 2013 IEEE Congress on Evolutionary Computation (CEC'1313), pp. 424–431, IEEE, 2013.
URL http://dx.doi.org/10.1109/CEC.2013.6557600

We study properties of Evolutionary Algorithms (EAs) with non-trivial offspring populations in the context of the (1+λ) EA. Already on linear functions, we observe that there are choices for the population size λ that yield, surprisingly, different asymptotic runtimes for the (1+λ) EA, more precisely for the prominent functions OneMax and BinVal. On the natural test function class of Royal Road functions, we show that a polynomial-sized offspring population has a very restricted speed-up effect. We give some ideas of how to prove upper and lower bounds for the (1+λ) EA on these function classes. This talk summarizes joint work with Benjamin Doerr.

References

1 B. Doerr and M. Künnemann. How the $(1 + \lambda)$ Evolutionary Algorithm optimizes linear functions. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2013)*, pages 1589–1596. ACM, 2013.

2 B. Doerr and M. Künnemann. Royal road functions and the $(1+\lambda)$ Evolutionary Algorithm: Almost no speed-up from larger offspring populations. In *IEEE Congress on Evolutionary Computation (CEC 2013)*, pages 424–431. IEEE, 2013.

3.12 Optimising Existing Software with Genetic Programming

William B. Langdon (University College London, GB)

Main reference W. B. Langdon, M. Harman, "Optimising Existing Software with Genetic Programming," accepted for publication in IEEE Trans. of Evolutionary Computation; authors' submitted version available.
URL http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/papers/Langdon_2013_ieeeTEC.pdf

We show genetic improvement of programs (GIP) can scale by evolving increased performance in a widely-used and highly complex 50000 line system. GISMOE found code that is 70 times faster (on average) and yet is at least as good functionally. Indeed it even gives a small semantic gain.

References

1 William B. Langdon and Mark Harman. Optimising existing software with genetic programming. *IEEE Transactions on Evolutionary Computation.* Accepted.

3.13 Random Declines and Drift Analysis

Johannes Lengler (ETH Zürich, CH)

Drift analysis has been a major tool for runtime analysis of Evolutionary Algorithms. For finite search spaces, they often come in pairs of matching upper and lower bounds. While the upper bounds hold in general, I present in my talk a class of Random Walks for which the standard lower bounds fail as the search space becomes infinite. In particular, the lower bounds known as "Variable Drift Theorems" fail although drift analysis still gives the correct runtime after a suitable transformation of the search space. Hence, the known Variable Drift Theorems need further improvement.

3.14 Runtime Analysis of Ant Colony Optimization on Dynamic Shortest Path Problems

Andrei Lissovoi (Technical University of Denmark, DK)

Joint work of Lissovoi, Andrei; Witt, Carsten;
Main reference A. Lissovoi, C. Witt, "Runtime analysis of ant colony optimization on dynamic shortest path problems," in Proc. of the 15th Annual Conf. on Genetic and Evolutionary Computation (GECCO'13), pp. 1605–1612, ACM, 2013.
URL http://dx.doi.org/10.1145/2463372.2463567

We consider the behaviour of a simple ACO algorithm called λ-MMAS on instances of dynamic shortest path problems where the optimum rapidly oscillates between two similar solutions. It is shown that reducing the evaporation rate ρ slows, but does not prevent, pheromone freezing when a single ant is started at each vertex. Starting a constant number of ants with a not-too-high evaporation rate allows the optimum solution to be constructed with constant probability in each of any polynomial number of iterations of certain easy problem instances. Additionally, the limitations of pheromone memory are demonstrated by considering an oscillation between two optimums that are not similar: while uncertainty about a specific choice can be represented in pheromone values, the memory cannot store distinct solutions.

3.15 Quantifying Optimization Problem Similarity Using Information Distance

Rachael Morgan (The University of Queensland, AU)

Quantifying the similarity between black-box optimization problems is important but non-trivial. In practice, there are two general methodologies employed: 1) problems are compared via measurements of particular problem properties/features (e.g. correlation length) and 2) problems are compared via the performance of algorithm instances. Ideally, a problem similarity measure should be a metric, estimated easily from data (with minimal loss of information) and able to compare problems of varying size and dimensionality. In addition, in the context of black-box optimization, such a measure is restricted to using only the information available in the black-box setting.

This talk proposed using the information distance – a universal distance – to measure the (dis)similarity between continuous optimization problems. While mainly a theoretical notion, the information distance can be approximated via the Normalised Compression Distance (NCD). NCD utilises standard compression algorithms to compute an approximation of the information distance between objects. Hence, by simply sampling the solutions and their respective objective function values of two optimization problems, the NCD between the samples can be used as a measure of distance between the problems.

To illustrate the proposed methodology, the NCD between circle packing problems and problems within the Black-Box Optimization Benchmark (BBOB) set were presented. The results showed that the similarity of the structures known to be within these problems is preserved by NCD. The results also demonstrated the use of Multi-Dimensional Scaling as a technique to visualise the resulting NCD between problems, and hence, visualise where problems are located within the 'problem space'.

3.16 How to Bridge Theory and Practice in Supply Chain Management

Frank Neumann (The University of Adelaide, AU)

Joint work of Bonyadi, Mohammad Reza; Michalewicz, Zbigniew; Barone, Luigi; Ghandar, Adam; Neumann, Frank

We discuss the current gap in the area of evolutionary computation methods for large scale optimization problems arising in the area of supply chain management. Evolutionary algorithms have been widely applied to a wide range of classical combinatorial optimization problems such as the traveling salesman problem or the knapsack problem. We argue that real world problems differ from these classical NP-hard optimization problems significantly in complexity as real-world problems are composed of different (NP-hard) sub-problems and the increase in complexity is due to the combination of these problems. Hence, it is important to understand how different problems that interact can be solved by evolutionary computing methods. We argue that there is a strong need for new benchmarks addressing this current gap and point out by an example problem, based on the classical traveling salesperson problem and the knapsack problem, how such interactions look like. For theoretical research on evolutionary algorithms, we pose it as an open challenge to understand the interactions of problems by means of theoretical investigations. We point out open topics for research in the areas of runtime analysis and fitness landscape analysis. In the case of rigorous runtime analysis it would be desirable to have results that show how the runtime behaviour changes when combining sub-problems. Investigations in the area of fitness landscape analysis could reveal on how the fitness landscape changes when combining subproblems such as the traveling salesperson problem and the knapsack problem.

3.17 Parameterized Complexity Analysis and More Effective Construction Methods for ACO Algorithms and the Euclidean Travelling Salesperson Problem

Samadhi Nallaperuma (The University of Adelaide, AU)

Joint work of Nallaperuma, Samadhi; Sutton, Andrew M.; Neumann, Frank

We propose a new construction procedure for ant colony optimization (ACO) algorithms working on the Euclidean traveling salesperson problem (TSP) that preserves the ordering on the convex hull of the points in the instance. The procedure is inspired by theoretical analyses for simple evolutionary algorithms that are provably more efficient on instances where the number of inner points of the instance is not too large. We integrate the construction procedure into the well-known Max-Min Ant System (MMAS) and empirically show that it leads to more efficient optimization on instances where the number of inner points is not too high.

3.18 Run-time Analysis to Elucidate the Benefits of Crossover

Adam Prugel-Bennett (University of Southampton, GB)

Joint work of Prugel-Bennett, Adam; Rowe, Jonathan; Shapiro, Jonathan

In this talk we describe current work on obtaining run time bounds on a set of problems which demonstrate the benefit of crossover. We start by showing that an EA using a population of size $P = \Omega(n)$ can solve onemax in no more than $\sqrt{n}\log(n)$ generations with overwhelming probability. The EA works by pairing the parents and producing two complementary children using uniform crossover. The fitter of the two children is chosen. The pairing is done so that each parent is chosen twice, and paired with a different partner. We show that at each site there is a sufficient drift towards increasing the number of ones that fixation is exponentially unlikely and the expected run time is no more than $\sqrt{n}\log(n)$ generations. We then consider four problems to illustrate the different benefits of the use of a population with crossover. The first of these examines the ability to put together building blocks. We consider a concatenated trap function consisting of m traps of size k (so that $n = m \times k$). The trap functions have a fitness function which is a symmetric V-shaped function of the number of ones in each trap except for the all ones state which has an additional fitness of one. To ease the analysis we assume k is odd, We consider the EA hybridised with a local hill-climber. The hill-climber is always run until completion so the traps are either in the all ones or all zeros state. We run the hill-climber on the two children produced by crossover and choose the fitter of the two children after the hill-climber. The children therefore have strings where each trap is either in the local or global optimum. Uniform crossover does not change the traps value if the two parents are in the same local optimum. If one parent is in the all 1's optimum and the other in the all 0's for a particular trap, then after crossover one child will have more than half 1's and the other less than half 1's. After hill-climbing one child will be in the all 0's state and the other in the all 1's state for that trap function. This means the problem is identical to the onemax except the state of the trap plays the role of the binary variable. We can thus reuse the analysis for onemax. If we choose the traps sufficiently large (e.g. $k = \Theta(\sqrt{n})$) then a (1+1)-EA would take super-polynomial time while our algorithm solves this problem in $O(n^{7/4}\log(n))$. The second mechanism we study is the concentration of search by crossover. We consider a modified onemax with a plateau of size $g\sqrt{n}$ when there is an ones in a binary string ($a > 1/2$). We show that because it is quite likely for two children to differ in the number of ones by $\Omega(\sqrt{n})$, then with reasonable frequency the two children will span the gap and there will be a drift towards increasing the total number of ones. The drift is sufficient to ensure that the population will rapidly cross the gap. Once again we show that a (1+1)-EA will take super-polynomial time to cross the gap. The third mechanism we consider is the tolerance to noise of a population based EA using crossover. To illustrate this we consider onemax, but with a noise at each fitness evaluation drawn from a normal distribution, $N(0, n)$. We show that despite this noise the EA still has a significant positive drift towards increasing the number of ones, thus, with a very small modification we show that we can solve onemax with noisy fitness in $O(n\log(n))$. In contrast a (1+1)-EA will take super-polynomial time. The final mechanism (somewhat related to the previous), is the tolerance of our EA to local randomness. Here, we consider an objective function which is the sum of the onemax problem and a random function. The random function at each configuration is just chosen form a normal distribution, $N(0, n)$. The difference to the problem discussed above is that the noise is static or quenched in this model rather than varying at each fitness evaluation. The objective is to find a solution with $(1-\epsilon)n$ ones (note that we have no guarantee as to where the actual global optimum is). We have no proof for our EA, however, we make the observation that if the minimum Hamming distance between strings were above δn for some $\delta > 0$, then as the number of potential children at each crossover is $\Omega(2^{\delta n})$, it is exponentially unlikely that a child will visit a configuration we have already visited (given that we have run for a polynomial number of generations). Thus with overwhelming probability this case will be identical to the dynamic noise case and we can solve it in $O(\sqrt{n})$ generations (we no longer need the $\log(n)$ as we only require to get within ϵ of the optimum). For a hill-climber and (1+1)-EA we can show that the expected run time is super-exponential. The $\log(n)$ in the bound can be dropped if the fitnesses were concentrated. We show this is highly plausible. For the case of quenched noise, we need to show that the Hamming distances are concentrated away from zero. The concentration is complicated, due to incest (children that share parents, grandparents, etc. will be more concentrated that those that don't). On the other hand, crossover rapidly mixes the members of the population, making it unlikely that the Hamming distances become very small. It remains an open problem to prove these concentration results.

3.19 Mathematical Landscape Theory

Peter F. Stadler (Universität Leipzig, DE)

Combinatorial landscapes are investigated with a focus on at least three different types of features: correlation measures, systems of adaptive or gradient walks, and the distribution of local optima. While all three approaches eventually provide insights into landscape ruggedness, their interconnections remain poorly understood. A connection between local optima and path systems is provided by coarse grained abstractions of the landscape, in particular its barrier trees representing local optima and their connecting saddle points [1, 2].

Typically, landscapes are viewed over a graph or metric space. This captures well search by mutation operators. Generalized closure operators allow to describe the topology of the search space induced by crossover-based search and may provide a meaningful way to study differences in landscape geometry arising from differences in search operators. An interesting variation on the idea of varying landscape geometry in algorithm design is to find biased embeddings in larger search spaces that are constructed in such a way that good solutions are enriched in the redundant representation [3].

References

1 B. M. R. Stadler and P. F. Stadler. Combinatorial Vector Fields and the Valley Structure of Fitness Landscapes. *J. Math. Biol.*, volume 61, pages 877-898. Springer, 2010.
2 K. Klemm, J. Qin and P. F. Stadler. Geometry and coarse-grained representations of landscapes. *Emergence, Complexity, and Computation*, Springer, 2012. In Press.
3 K. Klemm, A. Mehta and P. F. Stadler. Landscape Encodings Enhance Optimization. *PLoS ONE*, volume 7, pages e34780. 2012.

3.20 Analysis of Parallel Evolutionary Algorithms for Combinatorial Optimization

Dirk Sudholt (University of Sheffield, GB)

Joint work of Sudholt, Dirk; Lässig, Jörg; Neumann, Frank; Oliveto, Pietro; Rudolph, Günter; Mambrini, Andrea; Yao, Xin

We consider the speedup gained by island models in terms of the parallel running time for problems from combinatorial optimization: sorting (as maximization of sortedness), shortest paths and Eulerian cycles. The results show in which settings and up to what degree evolutionary algorithms can be parallelized efficiently. Potential speedups depend on many design choices such as the search operators, representations and fitness functions used on the islands, and also the parameters of the island model. In particular, we show that a natural instance for Eulerian cycles leads to exponential vs. logarithmic speedups, depending on the frequency of migration. We also discuss the use of crossover in island models for an instance of the VertexCover problem.

References

1. J. Lässig and D. Sudholt. General upper bounds on the running time of parallel evolutionary algorithms. *CoRR*, abs/1206.3522, 2012.
2. J. Lässig and D. Sudholt. Analysis of speedups in parallel evolutionary algorithms for combinatorial optimization. In *22nd International Symposium on Algorithms and Computation (ISAAC 2011)*, volume 7074 of *LNCS*, pages 405–414. Springer, 2011.
3. F. Neumann, P. S. Oliveto, G. Rudolph, and D. Sudholt. On the effectiveness of crossover for migration in parallel evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 1587–1594. ACM Press, 2011.
4. A. Mambrini, D. Sudholt, and X. Yao. Homogeneous and heterogeneous island models for the set cover problem. In *Parallel Problem Solving from Nature (PPSN 2012)*, volume 7491 of *LNCS*, pages 11–20. Springer, 2012.

3.21 Noisy Optimization in Continuous Domains

Olivier Teytaud (University of Paris South XI, FR)

Black-box noisy optimization is a great useful research area. Simple regret, cumulative regret, uniform rates, are various relevant criteria. We show that using second order information can improve the convergence rate of simple regret when noise is moderate; that sampling only around the optimum reduces the optimal convergence rates [1]; that ES are reasonably good in the case of cumulative regret [2].

References

1. Jérémie Decock and Olivier Teytaud. Noisy optimization complexity under locality assumption. In *Foundations of Genetic Algorithms (FOGA 2013)*, pages 183-190, 2013.
2. Sandra Astete Morales, Jialin Liu and Olivier Teytaud. Noisy optimization convergence rates. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2013) Companion*, pages 223–224, 2013.

4 Breakout Sessions

After having successfully experimented with breakout sessions during the last edition of this Dagstuhl seminar, this time we made breakouts a substantial part of the seminar. After a public vote, both collecting preferences and scheduling conflicts, we decided to have the following breakout sessions, each lasting for around two hours, partially scheduled in parallel. The breakouts differed greatly in style and size, ranging from semi-plenary open problem sessions to half-a-dozen people discussing in-depth a specialized topic. The following notes, collected by session organizers or participants, try to collect the main points of each session.

4.1 Evolutionary Multiobjective Optimization

Dimo Brockhoff (INRIA Nord Europe – Lille, FR, dimo.brockhoff@inria.fr)
Joshua D. Knowles (University of Manchester, GB, j.knowles@manchester.ac.uk)
William B. Langdon (University College London, GB)

The breakout session on EMO during the Dagstuhl seminar 13271 "Theory of Evolutionary Algorithms" brought together people familiar with EMO and researchers who have not yet worked on multiobjective optimization. The main discussion topics included: open problems, Information Geometric Optimisation (IGO), interactive evolution (IE) in a Pareto optimal context, especially the practical role of not necessarily rational human decision makers subject to fatigue and business pressures. We also discuss the role of: external finite archives and elitism, crossover and mutation particularly in fixed binary and non-binary representations, drift analysis, reducing run time, LOTZ, two-sphere and other benchmarks and the SEMO algorithm. Our goal is to provide the basis from which concrete theoretical studies can rapidly emerge.

4.1.1 Introduction

Multiobjective optimization deals with optimization problems where two or more objective functions are to be optimized simultaneously and for which, typically, no single solution results in optimal values for all objectives. Instead, the search space exhibits a natural partial order ("Pareto dominance relation") with the so-called set of Pareto-optimal solutions as its minima. For tackling multiobjective optimization problems with the use of evolutionary computation approaches, the term Evolutionary Multiobjective Optimization (EMO) has been coined. Although several theoretical studies in the field of EMO have emerged in recent years, fundamental questions are still open.

4.1.2 Overview

The EMO breakout session took place in the afternoon (4pm–6pm) of the second day of the Dagstuhl seminar (July 2, 2013) and was attended by Karl Bringmann, Dimo Brockhoff, Ken De Jong, Benjamin Doerr, Carola Doerr, Josephine Doerr, Joshua D. Knowles, Marvin Künnemann, W. B. Langdon, Jörg Lässig, Samadhi Nethmini Nallaperuma, Günter Rudolph, Manuel Schmitt and Lothar Thiele.

Most of the topics discussed during the breakout session originated from theoretical studies by some of the participants or from plenary presentations given during the first two days of the Dagstuhl seminar. The discussions related to Information Geometric Optimization, interactive EMO, and runtime analyses were the longest and most lively and so will be described in detail particularly focusing on the open questions which were identified.

4.1.3 Information Geometric Optimization and Multiobjective Optimization

Information Geometric Optimization (IGO) [12] is a recently proposed canonical and systematic way of looking at single-objective black-box optimization problems. The so-called *IGO flow*, a general continuous-time model of a black-box optimization algorithm, thereby optimizes the parameters of a given family of probability distributions by following the direction of the natural gradient of a joint optimization criterion posed on the manifold of probability distributions. It inherently possesses several invariance properties such as "invariance under reparametrization of the search space, under a change of parameters of the probability distribution, and under increasing transformation of the function to be optimized" (see [12]). Discretizations of the IGO flow for specific families of probability distributions result in known algorithms such as the CMA-ES [13], xNES [4] and PBIL [1].

As the IGO flow formulation is very general, the natural question is whether and how this would be formulated for multiobjective optimization problems and most importantly, which algorithms would emerge by the discretization of the (multiobjective) IGO flow.

One first step towards an answer, discussed during the breakout session, was to start with the simplest possible probability distribution for sampling *sets* of solutions: define μ linked probability distributions, one per member of the EMO algorithm's population, with μ the population size. In order to transform the set problem back to the IGO framework, one could use for example the hypervolume indicator of the set as the scalar objective function. As underlying family of probability distributions, we should probably start with the simplest ones (either Gaussian distributions in the case of continuous search spaces or Bernoulli distributions in case of pseudo-Boolean search spaces). It was suggested that it might help to start with a concrete problem as this might help gain intuition into possible future algorithms, e.g. the double sphere ($f_1(x) = |x|^2, f_2(x) = |x-a|^2$ with $x, a \in \mathbb{R}^n$). In this problem, the goal is to minimise both f_1 and f_2. This naturally leads to the need to trade off f_1 and f_2 when $a \neq \vec{0}$.

4.1.4 Analyses of Interactive EMO Algorithms

Evolutionary multiobjective optimization algorithms typically aim at finding a set of solutions that approximates the set of Pareto-optimal solutions. However, in the real world not all of the Pareto-optimal solutions are equally interesting. Often, where there is a single customer, a single solution has to be picked eventually by a human decision maker (or a group of decision makers)—corresponding to the *most preferred alternative*. Even where there are many customers, they may not pick solutions from the whole range of Pareto-optimal solutions. Since there is no information about the acceptable trade-offs between the objectives the definition of the preferred solution before the search is often too difficult. Similarly after the search, and making one final choice within an overwhelmingly large number of solutions can be equally difficult. *Interactive EMO algorithms* have been recently proposed to reduce the user load inherent in EMO algorithms presenting many potential Pareto-optimal solutions to the users (see for example [7] for an overview). These interactive EMO algorithms ask the decision maker during the optimization to provide certain information about their preferences, e.g., to say which of two solutions is preferred. The algorithm tries to focus the search on search space regions "preferred" by the decision maker instead of finding an approximation of the entire Pareto front.

The first theoretical analysis of such an interactive EMO algorithm appeared last year in [3] in which not only the runtime of simple interactive EMO algorithms on pseudo-Boolean test functions have been investigated but also the expected number of times, the decision maker is required to interact until the most preferred solution is found. After a brief informal presentation of the paper's content, the discussion centered on how to extend this preliminary study towards more realistic algorithms and decision making scenarios.

One line of suggestions focused on the decision maker who typically is not always deciding rationally. To investigate more realistic cases, analysis of the decision maker behavior might involve noisy decisions, additional (hidden) objective functions, current decisions that depend on previous ones (i.e., a "learning" decision maker), or the unavailability of the decision maker at certain times. Another interesting aspect that came up in the discussion was to allow the decision maker to also say "I don't know" or "I don't care" for certain questions asked. (NB. these are not identical replies.) Developing algorithmic strategies for dealing with such scenarios and analyzing their impact on the search abilities theoretically is definitely an interesting open problem for the near future.

Another suggestion was to analyse interactive EMO algorithms on more complicated problems. A possible next step is the bi-objective spanning tree problem [14].

4.1.5 Runtime Analyses of EMO Algorithms

Though theoretical runtime analyses of single-objective optimization algorithms is now established, runtime analyses of multiobjective algorithms remain comparatively sparse. Therefore, more basic questions remain open than in the single-objective case. Open questions collected and topics discussed during the session include:

- the fundamental question of when crossover is helpful before reaching the Pareto front,
- whether drift analysis has been used for multiobjective optimization (yes, for example in [5]),
- whether problems exist where the known lower bounds on the expected runtime match the upper bounds when global mutations are allowed (so far no, to the best of our knowledge), or
- which other algorithms could "improve" the runtime on the LOTZ problem [9] when compared to the SEMO algorithm [8].

4.1.6 Archiving and Non-Standard Search Spaces

Finally, two other aspects of EMO were discussed during the breakout session: archiving and non-standard search spaces.

In terms of archiving and performance guarantees, the question came up of how non-elitist and comma-strategies can be analysed in the multiobjective case. For example, in terms of the quality of an external archive of pair wise non-dominated solutions. Similarly what is the impact of a finite limit versus allowing the archive to grow indefinitely? The general question of how to deal with finite size archives was discussed briefly and it was acknowledged that although there is work on convergence and quality [2, 10, 6], there is no consensus model or framework for analysis as yet.

In terms of non-standard search spaces (e.g. discrete, but not bit string $\{0,1\}^n$), the general question came up which properties of mutation operators are desired (cf. for example [11]), which is especially open for *mutation operators on solution sets*. Also theoretical studies comparing different mutation operators might be interesting to practitioners.

4.1.7 Conclusion

Several theoretical studies of evolutionary multiobjective optimization (EMO) algorithms and techniques have been published recently, but various fundamental questions remain open. The breakout session on EMO during the Dagstuhl seminar "Theory of Evolutionary Algorithms" (July 2013) brought together researchers interested in theory and multiobjective optimization and provided a good opportunity to collect several of those open questions and start a discussion about how to tackle them. The brief description of those discussions, provided here, can be used as a basis for future (joint) theoretical research in the growing field of Evolutionary Multiobjective Optimization.

4.1.8 Acknowledgements

Special thanks go to the organizers of the Dagstuhl seminar and the entire Dagstuhl team for providing such a stimulating and welcoming atmosphere.

References

1 S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithm. pages 38-46. Morgan Kaufmann Publishers, 1995.

2 R. Berghammer, T. Friedrich, and F. Neumann. Set-based Multi-objective Optimization, Indicators, and Deteriorative Cycles. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 495-502, 2010.

3 D. Brockhoff, M. López Ibáñez, B. Naujoks and G. Rudolph. Runtime Analysis of Simple Interactive Evolutionary Biobjective Optimization Algorithms. In *Parallel Problem Solving from Nature (PPSN 2012)*, volume 7491 of LNCS, pages 123-132. Springer, 2012.

4 T. Glasmachers, T. Schaul, Y. Sun, D. Wierstra, and J. Schmidhuber. Exponential natural evolution strategies. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 393-400. ACM, 2010.

5 J. Handl, S. C. Lovell, and J. Knowles. Multiobjectivization by Decomposition of Scalar Cost Functions. In *Parallel Problem Solving from Nature (PPSN 2008)*, volume 5199 of LNCS, pages 31-40. Springer, 2008.

6 T. Hanne. On the convergence of multiobjective evolutionary algorithms. *European Journal of Operational Research*, 117(3):553-564, 1999.

7 A. Jaszkiewicz and J. Branke. Interactive Multiobjective Evolutionary Algorithms. *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 179-193. Springer, 2008.

8 M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, volume 10, pages 263–282, 2002.

9 M. Laumanns, L. Thiele, and E. Zitzler. Running Time Analysis of Multiobjective Evolutionary Algorithms on Pseudo-Boolean Functions. *IEEE Transactions on Evolutionary Computation*, volume 8, pages 170–182, 2004.

10 M. López-Ibáñez, J. Knowles, and M. Laumanns. On sequential online archiving of objective vectors. In *Evolutionary Multi-Criterion Optimization*, volume 6576 of LNCS, pages 46-60. Springer, 2011.

11 R. Moritz, T. Ulrich, L. Thiele, and S. Bürklen. Mutation operator characterization: Exhaustiveness, locality, and bias. In *IEEE Congress on Evolutionary Computation (CEC 2011)*, pages 1396-1403. IEEE Press, 2011.

12 Y. Ollivier, L. Arnold, A. Auger, and N. Hansen. Information-geometric optimization algorithms: A unifying picture via invariance principles, 2011. arXiv:1106.3708.

13 A. Ostermeier, A. Gawelczyk, and N. Hansen. A derandomized approach to selfadaptation of evolution strategies. *Evolutionary Computation*, volume 2, pages 369–380, 1995.
14 R.M. Ramos, S. Alonso, J. Sicilia, and C. Gonzalez. The problem of the optimal biobjective spanning tree. *European Journal of Operational Research*, volume 111, pages 617–628, 1998.

4.2 Black-Box Complexity

Carola Doerr (University Paris-Diderot, FR)

The breakout session on black-box complexity brought together researchers with much different interest in this topic. We discussed in the first 30 minutes our motivation for studying black-box complexity. While black-box complexity is mainly interesting as a tool to derive fairly general lower bounds for the optimization time of evolutionary algorithms and other randomized search heuristics for some, its main benefit is seen to be a source for inspiration in algorithmic design for others.

The participants in the discussion agreed that it would be nice to see as a future development black-box complexity results that go beyond the studies of the OneMax test functions. While the latter is an interesting source for inspiration and has imposed several beautiful questions and results in the last 50 years, it is important also to explore the black-box complexities of more intricate test beds, e.g., combinatorial optimization problems.

Johannes Lengler presented a new black-box algorithm that is built on the idea that offspring should be sampled not too far from the optimum. He explains that many optimal black-box algorithms gather a lot of knowledge about the problem at hand by consciously sampling non-optimal search points. This raises the question whether or not the black-box complexity necessarily increases when algorithms are allowed to sample only in a suitably chosen neighborhood of the search points that have been evaluated so far. While very restrictive neighborhood definitions will clearly lead to an increase in black-box complexity, recent progresses in algorithmic design seem to suggest that this is not the case for a more relaxed neighborhood structure. Johannes presented a new algorithm that has running time $o(n \log n)$ on the OneMax test functions. Also the algorithm by Benjamin Doerr, Carola Doerr, and Franziska Ebel [1] that was presented in a talk preceding this session achieves this. Benefits and drawbacks of both algorithms were discussed.

The participants agreed that transferring results from optimal black-box algorithms to the design of new evolutionary algorithms is a promising research direction. At the same time, the quest for a suitable black-box model continues. Several recent attempts, including memory-restricted, ranking-based, and unbiased versions of the black-box model seem to be a step in the right direction. However, in particular for algorithms using operators of arity strictly larger than one, a suitable black-box model is still missing. One potential model that was discussed during this breakout session is a "blind" version, where algorithms are given the fitness of the search points but no hint whatsoever about the search point itself. This version is related but not identical to the unbiased black-box model of Lehre and Witt [2].

References

1 Benjamin Doerr, Carola Doerr and Franziska Ebel. Lessons From the Black-Box: Fast Crossover-Based Genetic Algorithms. In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2013)*, pages 781–788, ACM, 2013.
2 Per Kristian Lehre and Carsten Witt. Black-Box Search by Unbiased Variation. *Algorithmica*, volume 64, pages 623–642, Springer, 2012.

4.3 Natural Gradient

Nikolaus Hansen (INRIA Saclay – Orsay, FR)

An evolutionary search and optimization process can be described by an iterative update of a probability distribution over the space of candidate solutions. The fact that search and adaptation is accordingly conducted in the space of distributions is often not considered, and the algorithms and their analysis do not usually exploit the properties of this rather well-defined space. They use, for example, strategies rooted in standard Euclidean geometry, however, the space of probability distribution calls for a different mathematical treatment, which is investigated in the research field of information geometry. Information geometry is a rather young mathematical research field pioneered by S.-I. Amari. It has been successfully applied to computational learning problems. In particular, this way of thinking suggests replacing the gradient in Euclidean space by the gradient defined in the space of probability distributions – the so called natural gradient.

Independently from the work in the learning community, information geometric randomized search algorithms have been developed just recently. First, the natural evolution strategy was proposed. Then it was shown that the popular covariance matrix evolution strategy (CMA-ES) can be derived from an information geometric perspective. These new developments open the door to new theoretical insights into randomized search in general.

In the breakout session, different types of evolutionary algorithms using information geometric concepts were discussed. Similarities and differences between the approaches were outlined, ideas between people specialized in discrete and continuous search spaces, respectively, were exchanged. The open questions and next steps in this new research direction were discussed. Important problems were identified: first, the observation that actual behavior of the iterative algorithms often differs from the behavior predicted by the analysis of the corresponding time-continuous flow (along the natural gradient). Second, the fact that the algorithms most successful in practice update their parameters describing the probability distribution in a non-uniform way in the sense that, for example, the CMA-ES has different leaning rates for the mean and the covariance – although the information theoretic interpretation (in terms of following the natural gradient) suggests that all parameters should be updated in the same way. Furthermore, the question of how to extend the existing theoretical frameworks to multi-objective optimization was discussed.

4.4 Fixed Budget Computations

Thomas Jansen (Aberystwyth University, GB)

The breakout session about fixed budget computations was a well attended and lively session that turned out to concentrate mostly on collecting open problems and questions that we would like to see answered. Participants in the discussion were Dimo Brockhoff, Kenneth A. De Jong, Benjamin Doerr, Carola Doerr, Thomas Jansen, Daniel Johannsen, Timo Kötzing, Joshua D. Knowles, Johannes Lengler, Frank Neumann, Jonathan E. Rowe, Dirk Sudholt and Christine Zarges. Notes have been taken by Dimo Brockhoff and Dirk Sudholt. These notes are the basis of this summary.

Currently, fixed budget results concentrate on very simple algorithms like random local search and the (1+1) EA [1, 2]. However, there is consensus that a wide range of algorithms should be analysable and be analysed. It is remarked that hybrid algorithms (e. g., memetic algorithms) may be of particular interest.

There is a wide range of different results that are desirable. Among them are results for large ranges of budgets; average case results; relative results for the comparison of algorithms; results for algorithms with high variance in performance (a kind of algorithms for which currently there are no good methods to deal with them analytically); results that highlight the influence of the selection pressure for a fixed budget; results on optimal algorithms (in the terms of black-box complexity); results for multiple runs, in particular for parallel runs; results for NP- hard combinatorial optimisation problems (similar to optimisation time results for PARTITION [3]) where results about structural properties of the problem instances would be of particular interest; results for FPT problems with fixed n depending on k; precise results for fixed n in general; results for dynamic fitness functions; and results for functions defined over continuous domains. It is pointed out that one should concentrate on proving lower bounds and that it is probably advisable to be more generous with respect to changes of the budget.

There is an open debate if making statements about the expected function values is actually the most useful thing to do. It may be more interesting to make statements about the median or quantiles instead of the expectation. Also rank-based results may be more meaningful than statements about the function value. An important aspect is the robustness (or brittleness) of results with respect to changes of the considered algorithm or problem.

For future analyses it is desirable to develop more analytical tools. Coming up with a method based on fitness levels appears to be a reasonable first step. It is open how fixed budget results can be obtained in general without revisiting optimisation time results. It is also open if and how results from landscape theory can help.

Another more general aspect that was also discussed was concerned with the goal of this kind of research, about the things that we want to see achieved. One obvious positive outcome would be novel insights that lead to the design of better search heuristics. Specific aspects are the design of stopping or restart criteria and the question of resampling when dealing with noisy fitness functions. A perhaps obvious application would be in algorithm portfolios where algorithms could be combined based on the relative strength in different epochs of the search. Less concrete but still useful would be a better understanding of the balance of exploitation and exploration in heuristic search. Another fruitful goal would be to address contradictory results between theory (in the sense of optimisation time) and practice, hopefully leading to more meaningful theoretical results. It can also be hoped that more useful notions of when one algorithm is better than another can be obtained.

References

1 B. Doerr, T. Jansen, C. Witt, and C. Zarges (2013): A method to derive fixed budget results from expected optimisation times. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2013)*, pages 1581–1588.

2 T. Jansen and C. Zarges (2012): Fixed budget computations: A different perspective on run time analysis. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2012)*, pages 1325–1332.

3 C. Witt (2005): Carsten Witt (2005): Worst-case and average-case approximations by simple randomized search heuristics In *Proceedings of the 22nd Annual Symposium on Theoretical Aspects of Computer Science (STACS 2005)*, pages 44–56.

4.5 Theory and Practice and Co-Evolution

William B. Langdon (University College London, GB)

Present: Christine Zarges, Manuel Schmitt, Rachael Morgan, Samadhi Nethmini Nallaperuma, Frank Neumann (FN), Ken de Jong (KDJ), Thomas Jansen, Jon Shapiro, Lothar Thiele, W B Langdon (WBL), Joshua D Knowles, Marcus Gallagher, Darrell Whitley (DW), Tom Schaul, Francisco Chicano.

The session was held outside with Frank Neumann interested in how theory helps with applications, particularly with respect to the Travelling Thief Problem TTP proposed at CEC 2013 by Zbigniew Michalewicz. KDJ suggests Michalewicz should provide abstraction of his real world problem that theory community can deal with.

FN described Travelling Thief Problem and KDJ suggested it be posed (and solved?) as a coevolutionary problem in which Travelling Salesman Problem (TSP) be viewed as one problem to be dealt with by one population and a cooperative population deal with second part. The second part being treated as knapsack problem. The two populations would each solve their part with the cooperation of members of the other population.

FN discussion of whether TTP is multi-objective problem or single objective. FN might propose a variant of TTP as a competition at GECCO or CEC.

Thomas Jansen said exact theorems are available for 1+1 coevolutionary algorithms

There was a general discussion of how (if at all) two known problems can be combined in such away that knowledge helps solution of new problems.

Jon Shapiro provided a counter example: two polynomial problems on graphs (one uniform colouring the other majority? colouring) can be combined to create an NP hard problem.

KDJ suggested what was needed was way of combining two NP hard problems to get a polynomial problem.

KDJ said Elena Popovici's work on coevolutionary landscapes suggests combining two problems can lead to vary different behaviour depending upon details of parameter settings. In coevolution very important to retain diversity. A good strategy is to retain memory of trial solutions off to the side, ie, in addition to the evolving populations.

Lothar Thiele suggested coal supply sub-problems interact via shared variables.

Bill Langdon (WBL) suggested that by taking the coevolutionary route we were mandating the two sub-problems be represented by sticking their two representations side by side. This may work, it may be easy, but by doing so we rule out all other possibilities.

Joshua D Knowles suggested reconsidering Richard Watson's HIFF problems. And also the biological literature on symbiotic evolutions.

Marcus Gallagher asked whether solving first instance give you an instance of the second problem. FN suggested if we fix a picking list (ie contents of the knapsack) it determines the TSP to be solved.

WBL said the picking list changes to the TTP "velocity" can be thought of as increasing the distance between cities in the TSP. Redrawing the TSP map. Darrell Whitley said also equivalent to needing more petrol between cities when have place heavy items in the knapsack.

DW said if we look at OR community, first rate results are obtained only by domain experts with say 20 years experience. These guys are not doing black-box optimisation, they are using their domain knowledge.

KDJ mentioned Christine Mumford's (http://users.cs.cf.ac.uk/C.L.Mumford/) work on collecting time tabling examples for use as useful benchmarks.

KDJ wondered if a useful problem might be to consider a direct graph (DAG) where the nodes were labelled with (randomly generated) problem instances. Solutions of each node's problem become inputs to the problems located at the node's children nodes.

DW recounted problem with academic job shop scheduling (JSS) benchmarks. Industry disregards these (and hence academic research) because they are tailored uniform randomly generated, whereas industrial problems are not uniform. DW said these benchmarks are worse than useless. If replaced by non-random problems the techniques that do best now do the worst. Introduction of tailored benchmarks has harmed the field.

WBL described his experiments on getting genetic programming to create benchmarks (IEEE TEC doi:10.1109/TEVC.2006.886448) in which a state of the art EA was defeated by a no-frills particle swarm optimiser (PSO). Also GP generated a search problem where the EA beat the PSO.

Lothar Thiele was interested in what theoretical analysis tells us.

KDJ recalled Toby Walsh's keynote at FOGA 2013: http://www.sigevo.org/foga-2013/keynote.html.

DW talked about "backbone" of solutions to SAT like problems and said TSP does not have such a backbone. DW then briefly described his Walsh analysis of k-SAT and how if leads in $O(n)$ time to the problem's backbone and hence rapidly to solutions. WBL this is an instance of theory driving algorithm design (rather than the reverse).

Francisco Chicano spoke in favour of theory helping to design algorithms. DW says applies to all pseudo Boolean problems (pretty big class).

4.6 Landscape Analysis

Adam Prugel-Bennett (University of Southampton, GB)
L. Darrell Whitley (Colorado State University, US)

The session discussed new methods for characterising problem hardness. Over the last 20 years one focus of landscape analysis has centred on elementary landscapes. However, these properties, such as the autocorrelation function, are identical for the same problem class for easy instances and instances which are found in practice to be very hard to solve. Thus, there is a need to find new measures. There is some promise in looking at the "superposition" of elementary landscapes. This works from domains such as pseudo-Boolean functions, and provides new statistic information about subregions of the search space. But there is still a challenge to leverage this information in useful ways. For k-bounded pseudo-Boolean functions, Walsh analysis still holds promise as a means of tracking the nonlinear interactions between variables; understanding these nonlinear interactions is key to understanding how to exploit improving moves on neighborhood structures.

It is possible that landscapes are so varied that finding a single property which captures problem hardness is a fruitless exercise, nevertheless there are many regularities found in a large class of problems when viewed at a large scales. In particular they tend to show similar distribution of local optima weakly correlated with a global optimum. Often there were regularities which were not immediately obvious. One new line of possible research is to use parameterised complexity and possibly some approximation to that idea to isolate the true complexity. Another idea is to search for structure in the distribution of local optima using principal component analysis to capture the major directions in which the local optima vary. Early investigations show that there is a lot of regularity which is easily missed. This raised the question of how these regularities could be exploited. In this regards we discussed the great success of crossover on many classic optimisation problems and tried to understand how this could be understood in terms of the landscape structure. We also discussed some recent ideas on modelling problems in terms of Gaussian Process models with the same mean fitness and two point correlation function as a real world problem.

4.7 General Drift Analysis with Tail Bounds

Jonathan E. Rowe (University of Birmingham, GB)

The discussion centred around two open questions in drift analysis. Firstly, the recently discovered flaw in the "negative drift" theorem. Secondly, whether or not tail bounds could be derived in the general variable drift case.

For the first part, it was noted that a fix for the problem has been proposed – namely disallowing large moves even away from the target. While this works (i.e. we get a proof), in some ways it seems not necessarily helpful, since there may be situations where there are large moves away from the target in which we would like to apply the theorem. A range of potential alternative conditions were suggested, and it was left for further work to pursue these for those interested.

For the second part, we looked at the problem of how tail bounds in the constant or multiplicative drift scenario might be transformed with the change of variable given by the variable drift theorem. It seemed that some progress could be made along these lines. Per Kristian Lehre told the group that indeed this could be done and some new results were shortly to be published on arxiv.

4.8 Parallel Evolutionary Algorithms

Dirk Sudholt (University of Sheffield, GB)

The aim of this breakout session was to discuss the theory of parallel variants of evolutionary algorithms. A previous talk by Dirk Sudholt (Section 3.20) about the analysis of parallel EAs for combinatorial optimization served as a primer.

At first, Bill Langdon gave a short introduction to parallel hardware: from multiple CPU cores with shared memory to parallel implementations on GPUs using CUDA and clusters of machines using MPI and message passing for communication. This led into a discussion about what is the right performance measure regarding the communication costs between parallel populations in island models – a question that arose during the previous talk. The importance of this criterion varies strongly between architectures, but it was generally agreed that the number of packets sent between islands was a sensible measure, and that typically migrants sent out would fit into one packet. Kenneth De Jong pointed out that in some architectures setting up the parallel system can be very costly.

In the remainder of the session the participants engaged in an open, inspiring, and sometimes controversial discussion around various aspects in the design and theoretical understanding of parallel EAs. Questions discussed include the following.

When do island models behave differently from panmictic populations? (Kenneth De Jong made the point that with frequent migrations island models take on characteristics of panmictic populations. Bill Langdon contributed a rule of thumb saying that there is a threshold behaviour at sending one individual per generation on average. Dirk Sudholt explained a theoretical result [2] on a problem where island models with migration outperform panmictic populations as well as independent runs.)

What is the right way of comparing a panmictic EA (e.g. a (100+100)EA) against a parallel EA with several subpopulations (e.g. 10 islands reach running a (10+10)EA)? (One conclusion was to keep the number of function evaluations the same for both settings.)

During migration, should individuals be removed from their original population, or should copies thereof be sent? How to choose migrants to be sent?

Should islands send migrants to all neighbouring islands probabilistically, or pick one island chosen uniformly at random? (The following discussion showed that the analysis method from [1] is applicable to various such settings.)

How can we deal with settings where nodes might fail and not respond, or have random response times?

Can we derive fixed-budget results for parallel evolutionary algorithms to show that parallelisation can lead to better solution quality?

How can diversity in parallel EAs be increased?

A final question was what theoretical results everyone would like to see in the future. Frank Neumann's answer was theoretical results that are closer to practice.

References

1 Jörg Lässig and Dirk Sudholt. General scheme for analyzing running times of parallel evolutionary algorithms. In *Parallel Problem Solving from Nature (PPSN 2010)*, pages 234–243, 2010.

2 Jörg Lässig and Dirk Sudholt. The benefit of migration in parallel evolutionary algorithms. In In *Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 1105–1112, 2010.

Participants

- Youhei Akimoto
Shinshu University, JP
- Khulood Alyahya
University of Birmingham, GB
- Dirk V. Arnold
Dalhousie University, CA
- Anne Auger
INRIA Saclay – Île-de-France – Orsay, FR
- Karl Bringmann
MPI für Informatik – Saarbrücken, DE
- Dimo Brockhoff
INRIA Nord Europe – Lille, FR
- Francisco Chicano
University of Malaga, ES
- Kenneth A. De Jong
George Mason University – Fairfax, US
- Benjamin Doerr
MPI für Informatik – Saarbrücken, DE
- Carola Doerr
University Paris-Diderot, FR
- Anton V. Eremeev
Sobolev Institute of Mathematics – Omsk, RU
- Tobias Friedrich
Universität Jena, DE
- Marcus Gallagher
University of Queensland, AU
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Nikolaus Hansen
INRIA Saclay – Île-de-France – Orsay, FR
- Michael Hellwig
Fachhochschule Vorarlberg, AT
- Christian Igel
University of Copenhagen, DK
- Thomas Jansen
Aberystwyth University, GB
- Daniel Johannsen
Berlin, DE
- Joshua D. Knowles
University of Manchester, GB
- Timo Kötzing
Universität Jena, DE
- Marvin Künnemann
MPI für Informatik – Saarbrücken, DE
- Jörg Lässig
Hochschule Zittau/Gorlitz, DE
- William B. Langdon
University College London, GB
- Per Kristian Lehre
University of Nottingham, GB
- Johannes Lengler
ETH Zürich, CH
- Andrei Lissovoi
Technical Univ. of Denmark, DK
- Luigi Malago
University of Milan, IT
- Rachael Morgan
University of Queensland, AU
- Samadhi Nethmini Nallaperuma
University of Adelaide, AU
- Frank Neumann
University of Adelaide, AU
- Pietro S. Oliveto
University of Birmingham, GB
- Adam Prugel-Bennett
University of Southampton, GB
- Jonathan E. Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Tom Schaul
New York University, US
- Manuel Schmitt
Univ. Erlangen-Nürnberg, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Peter F. Stadler
Universität Leipzig, DE
- Sebastian U. Stich
ETH Zürich, CH
- Dirk Sudholt
University of Sheffield, GB
- Andrew M. Sutton
Colorado State University, US
- Olivier Teytaud
University of Paris South XI, FR
- Lothar Thiele
ETH Zürich, CH
- L. Darrell Whitley
Colorado State University, US
- Christine Zarges
University of Birmingham, GB