Report from Dagstuhl Seminar 15211

Theory of Evolutionary Algorithms

Edited by

Benjamin Doerr[1], Nikolaus Hansen[2], Christian Igel[3], and Lothar Thiele[4]

1 **École Polytechnique – Palaiseau, FR,** doerr@lix.polytechnique.fr
2 **INRIA Saclay – Île-de-France – Orsay, FR,** hansen@lri.fr
3 **University of Copenhagen, DK,** igel@diku.dk
4 **ETH Zürich, CH,** thiele@ethz.ch

Abstract

This report documents the talks and discussions at the Dagstuhl Seminar 15211 "Theory of Evolutionary Algorithms". This seminar, now in its 8th 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 rigorous runtime analysis and computational complexity theory for randomised search heuristics, information geometry of randomised search, and synergies between the theory of evolutionary algorithms and theories of natural evolution.

Seminar May 17–22, 2015- http://www.dagstuhl.de/15211
1998 ACM Subject Classification G.1.6 Optimization, F.2 Analysis of Algorithms and Problem Complexity
Keywords and phrases evolutionary algorithms, black-box optimization, randomized search heuristics, algorithms, artificial intelligence
Digital Object Identifier 10.4230/DagRep.5.5.57
Edited in cooperation with Carola Doerr

1 Executive Summary

Benjamin Doerr
Nikolaus Hansen
Christian Igel
Lothar Thiele

License Creative Commons BY 3.0 Unported license
© Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele

Evolutionary algorithms (EAs) are randomized search and optimization methods applicable to problems that may be non-continuous, multi-modal, noisy, multi-objective or dynamic. They have successfully been applied to a wide range of real-world applications and have demonstrated impressive performance in benchmarks for derivative-free optimization. The seminar was devoted to the theory underlying evolutionary algorithms and related methods, in order to gain a better understanding of their properties and to develop new powerful methods in a principled way. The highly international, interdisciplinary seminar brought together leading experts and young researchers in the field. The 45 participants came from 13 different countries, spread over 4 continents. Many additional researchers had expressed their interest to also attend the seminar, but could unfortunately not be considered.

Except where otherwise noted, content of this report is licensed under a Creative Commons BY 3.0 Unported license
Theory of Evolutionary Algorithms, *Dagstuhl Reports*, Vol. 5, Issue 5, pp. 57–91
Editors: Benjamin Doerr, Nikolaus Hansen, Christian Igel, and Lothar Thiele

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Topics

The following report covers all important streams of research in the theory of evolutionary algorithms with a focus on three topics of particular current interest:

Runtime and complexity. Rigorous runtime and analysis and computational complexity theory have become the most important tools in the theory of discrete evolutionary algorithms. The Dagstuhl seminar series "Theory of Evolutionary Algorithms" has sparked this development. The drastic increase in new results, new methods, and young researchers entering this field, but also the major unsolved problems naturally lead to keeping this a focus topic.

Information geometry. Using concepts from information geometry in evolutionary algorithms is one of the most promising new theoretical direction in evolutionary computing. The seminar provided a unique opportunity to discuss perspectives and limitations of this approach.

Natural evolution. Evolutionary computing is rooted in theories of natural evolution, and many early approaches to understand basic properties of evolutionary algorithms were inspired by biological evolution theory. Still, today these two research fields are almost completely separated. We invited experts from evolution biology to help better understanding the relations between both fields. We are particularly happy that we succeeded in bringing together researchers from evolution biology and computer science in a way that was stimulating and productive.

Organization

The seminar had three types of organized presentation and discussion formats to stimulate the free discussions among the participants. There were 20–30 minutes talks on current topics followed by discussions. These included a talk on potential industrial collaborations. In addition, we had a few longer talks, which combined recent work with an overview over the state-of-the-art in a certain domain: Thomas Jansen spoke on "Understanding Randomised Search Heuristics", Nick Barton on "Limits to Adaptation", Yann Olivier introduced "Information-geometric Optimization", and Timo Kötzing presented a talk on "Stochastic Fitness Functions and Drift". Furthermore, we continued with having "breakout sessions" for longer, parallel group discussions on timely, specialized topics. These were introduced in the last seminar on "Theory of Evolutionary Algorithms". This time, these session were even more productive than previously, both because the organizers and the participants were more used to this format of interaction. The talks and breakout sessions are summarized in Section 4 of this report.

We would like to thank the Dagstuhl team and the attendees for making seminar 15211 a great success and a pleasure to organize.

Benjamin Doerr (École Polytechnique – Palaiseau, FR)
Nikolaus Hansen (INRIA Saclay – Île-de-France – Orsay, FR)
Christian Igel (University of Copenhagen, DK)
Lothar Thiele (ETH Zürich, CH)

2 Table of Contents

3 Overview of Talks

3.1 Spectral Landscape Theory: Some Color from Infinite Population Analysis

Lee Altenberg (Konrad Lorenz Institute for Evolution & Cognitio, AT)

License Creative Commons BY 3.0 Unported license
© Lee Altenberg

To obtain analytical results on runtimes of evolutionary algorithms, tractability has required the use of very simple landscapes. Is there any hope of obtaining results for arbitrary landscapes? There may perhaps be some relevance from models of evolutionary dynamics with infinite populations where powerful spectral methods have been able to provide a variety of results for arbitrary landscapes. Karlin (1982) proved two theorems that systems which combine growth and state transitions exhibit smaller aggregate growth with increased rates of transition. These theorems have been extended to infinite dimensional spaces, and multiple independent transition events (Altenberg, 2011, 2012). Their application has been used to show that dispersal should reduce standing genetic variation, that self-adaptation in stationary populations should favor lowered mutation and crossover rates (Feldman's "Reduction Principle"), and that the fitness of a quasispecies decreases with increasing mutation rates up to m=1/2. In evolutionary computation, different representations and operators result in transmission matrices with different spectral properties. To illustrate, I show eight different ways to connect 64 genotypes by mutation, ranging from a simple path, to symbol sequences of length 6, 4, 3, and 2, to the complete graph, and show how the spectral gap of the associated mutation matrix increases for this series. Spectral landscape theory (Weinberger 1991, Stadler 1992) has produced a variety of results based on the relationships of 1) fitnesses to the 2) mutational eigenvalues and 3) mutational eigenvectors. Here I present new results showing that the asymptotic rate of growth of a quasispecies is an increasing function of the eigenvalues of its mutation matrix. In addition, a lower bound is found which increases with the alignment between fitnesses and the eigenvectors with largest eigenvalues. A principal open question most relevant to evolutionary computation is how the spectral gap of the quasispecies, which determines its rate of convergence from an initial state, depends on these three properties: fitnesses, mutational eigenvalues, and mutational eigenvectors.

References

1 Altenberg L (2011) An evolutionary reduction principle for mutation rates at multiple loci. *Bulletin of Mathematical Biology* 73:1227–1270.
2 Altenberg L (2012) Resolvent positive linear operators exhibit the reduction phenomenon. *Proceedings of the National Academy of Sciences U.S.A.* 109:3705–3710.
3 Feldman MW (1972) Selection for linkage modification: I. Random mating populations. *Theoretical Population Biology* 3:324–346.
4 Karlin S (1982) in *Evolutionary Biology*, eds Hecht MK, Wallace B, Prance GT (Plenum Publishing Corporation, New York) Vol. 14, pp 61–204.
5 Stadler PF (1992) Correlation in landscapes of combinatorial optimization problems. *Europhysics Letters* 20:479–482.
6 Weinberger ED (1991) Fourier and Taylor series on fitness landscapes. *Biological Cybernetics* 65:321–330.

3.2 Limits to Adaptation

Nick Barton (IST Austria – Klosterneuburg, AT)

License Creative Commons BY 3.0 Unported license
© Nick Barton

What limits the effectiveness of selection? We have a good theoretical understanding of simple forms of selection, which is used both to optimise animal breeding, and to explain the prevalence of sex and recombination. This theory carries over directly to evolutionary computation. Complex "fitness landscapes" are less well understood, but nevertheless, there are some general results that constrain the rate at which selection can accumulate information.

3.3 Fitness Information and Optimal Control of Mutation Rate

Roman V. Belavkin (Middlesex University, GB)

License Creative Commons BY 3.0 Unported license
© Roman V. Belavkin
Main reference R. V. Belavkin, "Dynamics of Information and Optimal Control of Mutation in Evolutionary Systems," in A. Sorokin, R. Murphey, M. T. Thai, P. M. Pardalos (Eds.), "Dynamics of Information Systems: Mathematical Foundations," Springer Proceedings in Mathematics & Statistics, Vol. 20, pp. 3–21, Springer, 2012.
URL http://dx.doi.org/10.1007/978-1-4614-3906-6_1

We consider genetic algorithms (GAs) as Markov chains, and then pose the problem of optimal control of their parameters. We give several formulations of this problem depending on additional constraints, such as time horizon and information constraints. In particular, we study the problem of optimal control of the mutation rate parameter. We show that solutions to the optimal control problems are control functions that can be derived from transition probabilities between level sets of the fitness function. Using our combinatorial result about the intersection of spheres in a Hamming space, we derive closed-form expressions for transition probabilities in the idealised case, when fitness is monotonic with respect to Hamming distance to an optimum [1]. The optimal mutation rate control functions are presented for several problems. We also discuss how these solutions can be applied to fitness functions that are only weakly monotonic. We also discuss the recent discovery of the mutation rate control in bacteria [2].

References

1 Belavkin, R. V. (2012). *Dynamics of Information and Optimal Control of Mutation in Evolutionary Systems*. In Sorokin, A.; Murphey, R.; Thai, M.T.; Pardalos, P.M. (Eds.) Dynamics of Information Systems: Mathematical Foundations. Springer Proceedings in Mathematics & Statistics, Vol. 20, pp 3–21. Springer.

2 Krasovec, R., Belavkin, R. V., Aston, J., Channon, A., Aston, E., Rash, B., Kadirvel, M., Forbes, S., Knight, C. (2014). *Mutation-rate-plasticity in rifampicin resistance depends on Escherichia coli cell-cell interactions*. Nature Communications, Vol. 5, No. 3742.

3.4 Information Geometric Optimization in $\mathbb{R}^N$: Open Problems and Misconceptions

Hans-Georg Beyer (Fachhochschule Vorarlberg, AT)

License Creative Commons BY 3.0 Unported license
© Hans-Georg Beyer
Main reference H.-G. Beyer, "Convergence Analysis of Evolutionary Algorithms That are Based on the Paradigm of Information Geometry, Evolutionary Computation," 22(4):679–709, 2014.
URL http://dx.doi.org/10.1162/EVCO_a_00132

After a short introduction into the philosophy and theory of Information Geometric Optimization (IGO) in real-valued search spaces, the dynamical theory of IGO on quadratic model functions will be reviewed. Recent results regarding the dynamics of the IGO flow [1] will be extended to provide convergence results for finite time steps in the case of expected value maximization. It will be shown that even in that case, the natural gradient definition leads to a disappointingly slow sublinear convergence order obeying an $1/t$ law. In order to get a faster approach (i.e., linear convergence order) to the optimizer one needs utility functionals that localize the IGO flow in time, such as local standardization of the sampling induced fitness values or truncation selection (i.e., local quantile based selection). Regarding the case of truncation selection, a special example will be considered where the fitness normality assumption [1] does not hold (therefore, the normality assumption is not the reason for the dynamical behaviors observed). The numerical solution of the resulting IGO ODE will be compared with the corresponding Evolution Strategy (ES). It will be shown that the classical mutation scale-invariance often observed in ES does not hold for a correctly working IGO algorithm unless one is willing to sacrifice the natural gradient definition and uses different time-step sizes for the mean and the covariance matrix update, respectively, thus departing from the IGO flow. The presentation will especially focus on open problems and misconceptions and is intended to initiate a vivid discussions regarding the concept of the natural gradient, information geometric optimization and its relation to Evolution Strategies.

References

1 H.-G. Beyer. Convergence Analysis of Evolutionary Algorithms That are Based on the Paradigm of Information Geometry. *Evolutionary Computation*, 22(4):679–709, 2014. DOI 10.1162/EVCO_a_00132.

3.5 Selection of Auxiliary Objectives with Reinforcement Learning: Overview of Theoretical Results

Arina Buzdalova (ITMO University – St. Petersburg, RU)

License Creative Commons BY 3.0 Unported license
© Arina Buzdalova

Efficiency of evolutionary algorithms may be enhanced by multi-objectivization [1, 2]. One of the corresponding approaches is to dynamically select auxiliary objectives from some predefined set during one run of evolutionary algorithm (EA) [3]. Dynamic selection of objectives is needed when it is efficient to optimize different objectives at different stages of optimization. It was proposed to use reinforcement learning (RL) for dynamic selection of auxiliary objectives. The corresponding approach was called EA+RL [4]. This approach was shown to be efficient during experimental studies. However, there is need for theoretical understanding of the underlying mechanisms that enable EA+RL to work efficiently. In this

overview, recent theoretical results are considered [4, 5, 6]. Efficiency of EA+RL is evaluated on some easy problems using runtime analysis. Some insights on proper design choices when implementing EA+RL are described as well. Based on the considered results, we propose directions for future improvement of dynamic selection of auxiliary objectives in evolutionary algorithms.

References

1 J. D. Knowles, R. A. Watson, D. W. Corne. *Reducing Local Optima in Single-Objective Problems by Multi-objectivization*. In Proc. of the 1st Int'l Conf. on Evolutionary Multi-Criterion Optimization EMO'01, Springer, 2001, pp. 269–283.

2 C. Segura, C. A. C. Coello, G. Miranda, C. Léon.*Using multi-objective evolutionary algorithms for single-objective optimization*. 4OR, Vol. 3, No. 11, 2013, pp. 201–228.

3 Mikkel T. Jensen. *Helper-Objectives: Using Multi-Objective Evolutionary Algorithms for Single-Objective Optimisation: Evolutionary Computation Combinatorial Optimization*. Journal of Mathematical Modelling and Algorithms, Vol. 3, No 4, 2004, p. 323–347.

4 M. Buzdalov, A. Buzdalova, A. Shalyto. *A First Step towards the Runtime Analysis of Evolutionary Algorithm Adjusted with Reinforcement Learning*. In Proc. of the Int'l Conf. on Machine Learning and Applications, Vol. 1, IEEE CS, 2013,pp. 203–208.

5 M. Buzdalov, A. Buzdalova. *OneMax Helps Optimizing XdivK: Theoretical Runtime Analysis for RLS and EA+RL*. In Proc. of Genetic and Evolutionary Computation Conf. Companion, ACM, 2014, pp. 201–202.

6 D. Antipov, M. Buzdalov, B. Doerr. *Runtime Analysis of(1+1) Evolutionary Algorithm Controlled with Q-learning using Greedy Exploration Strategy on OneMax+ZeroMax Problem*. Evolutionary Computation in Combinatorial Optimization, LNCS, 2015, Vol. 9026, pp. 160-1-72.

3.6 The Unrestricted Black-Box Complexity of Jump Functions

Maxim Buzdalov (ITMO University – St. Petersburg, RU)

License Creative Commons BY 3.0 Unported license
© Maxim Buzdalov
Joint work of Buzdalov, Maxim; Doerr, Benjamin; Kever, Mikhail
Main reference M. Buzdalov, M. Kever, B. Doerr, "Upper and Lower Bounds on Unrestricted Black-Box Complexity of $\text{Jump}_{n,\ell}$," in Proc. of the 15th Europ. Conf. on Evolutionary Computation in Combinatorial Optimization (EvoCOP'15), LNCS, Vol. 9026, pp. 209–221, Springer, 2015.
URL http://dx.doi.org/10.1007/978-3-319-16468-7_18

We analyse the unrestricted black-box complexity of the Jump function class. For upper bounds, we present three algorithms for small, medium and extreme jump sizes. We prove a matrix lower bound theorem which is capable of giving better lower bounds than the classic information theory approach. Using this theorem, we prove lower bounds for Jump separately for odd and even values of the bit string length. For several cases, notably for extreme jump sizes, lower and upper bounds coincide apart from lower order terms. Here is the complete list of proven upper bounds on $\text{Jump}_{n,\ell}$:

- for $\ell < n/2 - \sqrt{n}\log_2 n$: $\frac{2n(1+o(1))}{\log_2 n}$, where $o(1)$ is measured when $n \to \infty$;
- for $n/2 - \sqrt{n}\log_2 n \le \ell < \left\lfloor \frac{n}{2} \right\rfloor - 1$: $\frac{n(1+o(1))}{\log_2(n-2\ell)}$, where $o(1)$ is measured when $n - 2\ell \to \infty$;
- for $\ell = \left\lfloor \frac{n}{2} \right\rfloor - 1$: $n + \Theta(\sqrt{n})$.

The proven lower bounds are:

- for even n: $\left\lfloor \log_{\frac{n-2\ell+2}{2}} \left(1 + 2^{n-1}\frac{(n-2\ell)^2}{n-2\ell-1}\right) \right\rfloor - \frac{2}{n-2\ell} \ge \frac{n}{\log_2 \frac{n-2\ell+2}{2}} - 1$;
- for odd n: $\left\lfloor \log_{\frac{n-2\ell+1}{2}} \left(1 + 2^{n-2}(n-2\ell-1)\right) \right\rfloor - \frac{2}{n-2\ell-1} \ge \frac{n-1}{\log_2 \frac{n-2\ell+1}{2}} - 1$.

3.7 Understanding Simple Asynchronous Evolutionary Algorithms

Kenneth A. De Jong (George Mason University – Fairfax, US)

License Creative Commons BY 3.0 Unported license
© Kenneth A. De Jong
Joint work of De Jong, Kenneth A.; Scott, Eric
Main reference E. Scott, K. A. De Jong, "Understanding Simple Asynchronous Evolutionary Algorithms", in Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII (FOGA'15), pp. 85–98, ACM, 2015.
URL http://dx.doi.org/10.1145/2725494.2725509

When the cost of fitness evaluations is very high, parallel EAs are used to reduce the "wall clock" time required for an evolutionary run. In the case that there is high variance in the cost of these long running fitness evaluations, there can be a significant amount of multi-processor idle time when using synchronous parallel EAs. A simple asynchronous version is introduced that reduces idle time to zero, and analyzed it in several respects: 1) Does it have a built-in bias for individuals with faster running evaluation times? 2) For a fixed budget, how much improvement in wall clock time is obtained? 3) Is there also a speedup in finding better solutions faster?

3.8 Analyzing Self-Adjusting Parameter Choices in Discrete Search Spaces

Carola Doerr (CNRS and UPMC – Paris, FR)

License Creative Commons BY 3.0 Unported license
© Carola Doerr
Joint work of Doerr Benjamin; Doerr, Carola
Main reference B. Doerr, C. Doerr, "Optimal Parameter Choices Through Self-Adjustment: Applying the 1/5-th Rule in Discrete Settings," in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation (GECCO'15), pp. 1335–1342, ACM 2015; pre-print available as arXiv:1504.03212v1 [cs.NE], 2015.
URL http://dx.doi.org/10.1145/2739480.2754684
URL http://arxiv.org/abs/1504.03212v1

While evolutionary algorithms are known to be very successful for a broad range of applications, the algorithm designer is often left with many algorithmic choices, for example, the size of the population, the mutation rates, and the crossover rates of the algorithm. These parameters are known to have a crucial influence on the optimization time, and thus need to be chosen carefully, a task that often requires substantial efforts. Moreover, the optimal parameters can change during the optimization process. It is therefore of great interest to design mechanisms that dynamically choose best-possible parameters. An example for such an update mechanism is the one-fifth success rule for step-size adaption in evolutionary strategies. While in continuous domains this principle is well understood also from a mathematical point of view, no comparable theory is available for problems in discrete domains. In this work we show that the one-fifth success rule can be effective also in discrete settings. We regard the $(1 + (\lambda, \lambda))$ GA proposed in [1]. We prove that if its population size is chosen according to the one-fifth success rule then the expected optimization time on OneMax is linear. This is better than what *any* static population size λ can achieve and is asymptotically optimal also among all adaptive parameter choices.

References

1 B. Doerr, C. Doerr, and F. Ebel. *From black-box complexity to designing new genetic algorithms*. Theoretical Computer Science, 567:87–104, 2015.

3.9 On the Proportion of Fit Individuals in the Population of A Mutation-Based Genetic Algorithm

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

License Creative Commons BY 3.0 Unported license
© Anton V. Eremeev
Main reference A. V. Eremeev, "Modelling and Analysis of Genetic Algorithm with Tournament Selection," in Proc. of the 4th Europ. Conf. on Artificial Evolution (AE'99), LNCS, Vol. 1829, pp. 84–95, Springer, 2000.
URL http://dx.doi.org/10.1007/10721187_6

In this talk, a fitness-level model of non-elitist mutation-only genetic algorithm (GA) with tournament selection is considered. The model provides upper and lower bounds for the expected proportion of the individuals with fitness above given thresholds. In the case of GA with bit-flip mutation and OneMax fitness function, the lower bounds are tight when population size equals one, while the upper bounds are asymptotically tight when population size tends to infinity. The lower bounds on expected proportions of sufficiently fit individuals may be obtained from the probability distribution of an appropriate generalized random walk. This approach yields polynomial upper bounds on the run-time of the Iterated version of the GA on 2-SAT problem and on a family of symmetric set cover problems proposed by E. Balas. The research is supported by Russian Foundation for Basic Research grant 15-01-00785.

3.10 Lyapunov Stability Analysis of a Derandomized Self-Adaptive $(1, 2)$-ES

Carlos M. Fonseca (University of Coimbra, PT)

License Creative Commons BY 3.0 Unported license
© Carlos M. Fonseca
Joint work of Wanner, Elizabeth F.; Fonseca, Carlos M.; Cardoso, Rodrigo T. N.; Takahashi, Ricardo H. C.;

The convergence of a simple derandomized self-adaptive $(1, 2)$-ES is investigated on the class of strictly unimodal functions of one variable that are symmetric about the optimum. Using the theoretical framework proposed by Semenov and Terkel [1], not only can a stability region for the self-adaptation parameters be analytically determined, but also upper bounds on the rate of convergence can be established, allowing appropriate values for the self-adaptation parameters to be obtained numerically. Simulation results are in agreement with the theoretical conclusions.

Acknowledgment. The authors acknowledge Brazilian agencies CAPES, CNPq and FAPEMIG, the Portuguese agency FCT, and the iCIS project (CENTRO-07-ST24-FEDER-002003), which is co-financed by QREN in the scope of the Mais Centro Program and European Union's FEDER, for financial support.

References

1 Mikhail A. Semenov and Dmitri A. Terkel. *Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function.* Evolutionary Computation 11(4): 363–379, MIT Press, 2003.

3.11 Probabilistic Analysis of the (1+1)-Evolutionary Algorithm

Hsien-Kuei Hwang (Academica Sinica – Taipei, TW)

License Creative Commons BY 3.0 Unported license
© Hsien-Kuei Hwang
Joint work of Hwang, Hsien-Kuei; Panholzer, Alois; Rolin, Nicolas; Tsai, Tsung-Hsi; Chen, Wei-Mei
Main reference H.-K. Hwang, A. Panholzer, N. Rolin, T.-H. Tsai, W.-M. Chen, "Probabilistic analysis of the (1+1)-evolutionary algorithm," arXiv:1409.4955v1 [math.PR], 2014.
URL http://arxiv.org/abs/1409.4955v1

We give a detailed analysis of the cost used by the $(1 + 1)$-evolutionary algorithm. The problem has been approached in the evolutionary algorithm literature under various views, formulation and degree of rigor. Our asymptotic approximations for the mean and the variance represent the strongest of their kind. The approach we develop is also applicable to characterize the limit laws and is based on asymptotic resolution of the underlying recurrence. While most approximations have their simple formal nature, we elaborate on the delicate error analysis required for rigorous justifications. Our main results state that on the fitness function OneMax, the cost used by $(1 + 1)$-EA follows asymptotically a double-exponential distribution with mean and variance of order $n \log n$ and n^2, respectively. On the other hand, if the fitness function is changed to LeadingOnes, then the cost satisfies a central limit theorem with quadratic mean and cubic variance. These results are not new but our methods of proof are not only rigorous but provide stronger asymptotic approximations.

3.12 Understanding Randomised Search Heuristics

Thomas Jansen (Aberystwyth University, GB)

License Creative Commons BY 3.0 Unported license
© Thomas Jansen
Joint work of Corus, Dogan; He, Jun; Jansen, Thomas; Oliveto, Pietro S.; Sudholt, Dirk; Zarges, Christine
Main reference D. Corus, J. He, T. Jansen, P. S. Oliveto, D. Sudholt, C. Zarges, "On Easiest Functions for Somatic Contiguous Hypermutations And Standard Bit Mutations," in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation Conference (GECCO'15), pp. 1399–1406, AMC, 2015..
URL http://dx.doi.org/10.1145/2739480.2754799

Different theoretical perspectives can help us understand different aspects of randomised search heuristics like evolutionary algorithms, artificial immune systems and others. One particularly successful theoretical perspective is run time analysis. Recently, He, Chen and Yao have described a way of identifying and constructing easiest and hardest functions (which imply smallest and longest expected run time, respectively) for elitist randomised search heuristics that have population size one. We use this idea to investigate the easiest function of the $(1 + 1)$ CHM, a variant of the $(1 + 1)$ EA where the standard bit mutations from the $(1 + 1)$ EA are replaced with somatic contiguous hypermutations from the B-Cell algorithm, a popular artificial immune system for optimisation. An easiest function for the $(1 + 1)$ CHM is an asymptotically hardest function for the $(1 + 1)$ EA. Motivated by their very different strengths and weaknesses we consider hybrid algorithms that combine both algorithms by applying one of the two mutation operators with fixed probability. We prove that these hybrids are efficient on easiest functions of their components. We also discuss that easiest functions for such hybrids are more difficult to find than one might believe. We conclude with a number of open questions that could yield additional insight into the performance of hybrid heuristics.

3.13 Open Problems in Industry

Daniel Johannsen (SAP Innovation Center – Potsdam, DE)

License Creative Commons BY 3.0 Unported license
© Daniel Johannsen

SAP is one of the world's leading companies in the market of enterprise resource planning systems. Two focus areas in SAP's research strategy are Future of Knowledge Work and Personalized Medicine. Each of these areas is estimated to target a multi-trillion USD market by 2025. In the area of Future of Knowledge Work, the main challenge is to measure the business value of intangible assets of an organization such as employee skills and knowledge. Progress in this area is expected to be driven by self-adaptive and self-learning applications relying on algorithmic techniques such as evolutionary optimization and machine learning. Because of their versatility, the same techniques are also promising for the fundamentally different area of Personalized Medicine. We present six specific research challenges posed by different SAP business units. For each of these, SAP aims to establish long-term research collaborations with academic partners.

3.14 Stochastic Fitness Functions and Drift

Timo Kötzing (Hasso-Plattner-Institut – Potsdam, DE)

License Creative Commons BY 3.0 Unported license
© Timo Kötzing
Joint work of Gießen, Christian; Friedrich, Tobias; Kötzing, Timo; Krejca, Martin; Sutton, Andrew

In this talk we will see the powerful method of drift analysis as a tool to analyze random processes. The two main drift theorems are the additive and the multiplicative drift theorem, which we will apply in exemplary cases both from continuous and and discrete optimization. In the second part of the talk we will survey the current work on stochastic fitness functions. We will particularly discuss the disruptiveness of mutation operators and recent positive results on crossover operators.

3.15 Does the Natural Gradient Save Us?

Oswin Krause (University of Copenhagen, DK)

License Creative Commons BY 3.0 Unported license
© Oswin Krause

Recently, optimizing the expectation of an objective function over some probability distribution by following the natural gradient of the distribution parameters became a popular approach to describe and implement Evolutionary Strategies. In this framework, samples are drawn from the distribution to estimate the gradient which is multiplied by the inverse of the Fisher Information Matrix to form the natural gradient. The natural gradient is usually argued to be superior to the "vanilla" gradient as it is independent of the parameterisation and because the direction of the natural gradient better reflects the actual change of the distribution. The freedom of parameterisation is especially regarded as beneficial as it fits to the notion of black-box optimization. In this talk, we show that the natural gradient is not necessarily unique and that we have a degree of freedom to change it, even without changing

the expectation or its gradient, challenging the idea of a black-box metric. This spotlight talk is designed to spark interesting discussions regarding the use of the natural gradient in black-box optimization.

3.16 Level-based Analysis of Genetic Algorithms and Other Search Processes

Per Kristian Lehre (University of Nottingham, GB)

License Creative Commons BY 3.0 Unported license
© Per Kristian Lehre
Joint work of Corus, Dogan; Dang, Duc-Cuong; Eremeev, Anton
Main reference D. Corus, D.-C. Dang, A. V. Eremeev, P. K. Lehre, "Level-Based Analysis of Genetic Algorithms and Other Search Processes," in Proc. of the 13th Int'l Conf. on Parallel Problem Solving from Nature (PPSN'14), LNCS, Vol. 8672, pp. 912–921, Springer, 2014.
URL http://link.springer.com/chapter/10.1007/978-3-319-10762-2_90

We describe a very general technique for proving upper bounds on the expected optimisation time of a broad class of non-elitist population-based optimisation processes, ranging from simple (μ,λ) EAs to more complex genetic algorithms (GAs), and estimation of distribution algorithms (EDAs) applied in uncertain environments.

References

1 Dogan Corus, Duc-Cuong Dang, Anton V. Eremeev, and Per Kristian Lehre. Level-based analysis of genetic algorithms and other search processes. In *Proceedings of the 13th International Conference on Parallel Problem Solving from Nature*, PPSN'14, pages 912–921. Springer International Publishing, 2014.
2 Duc-Cuong Dang and Per Kristian Lehre. Evolution under partial information. In *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation*, GECCO '14, pages 1359–1366, New York, NY, USA, 2014. ACM.
3 Duc-Cuong Dang and Per Kristian Lehre. Efficient optimisation of noisy fitness functions with population-based evolutionary algorithms. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, FOGA '15, pages 62–68, New York, NY, USA, 2015. ACM.

3.17 Information Geometry of the Gaussian Distribution in View of Stochastic Optimization: Inverse Covariance Parameterization

Luigi Malago (Shinshu University, JP)

License Creative Commons BY 3.0 Unported license
© Luigi Malago
Joint work of Malagò, Luigi; Pistone, Giovanni
Main reference L. Malagò, G. Pistone, "Information Geometry of Gaussian Distributions in View of Stochastic Optimization," in Proc. of the 2015 ACM Conf. on Foundations of Genetic Algorithms XIII (FOGA'15), pp. 150–162, ACM, 2015; pre-print available as hal-01108986, 2015.
URL http://dx.doi.org/10.1145/2725494.2725510
URL https://hal.inria.fr/hal-01108986

We study the optimization of a continuous function by its stochastic relaxation, i.e., the optimization of the expected value of the function itself with respect to a density in a statistical model. In particular we focus on gradient descent techniques applied to the multivariate Gaussian distribution for the optimization of functions defined over continuous

domains. In the talk we present a parameterization for Gaussian distribution given by the mean and inverse covariance matrix based on the natural parameterization of the exponential family. We describe the advantages of this parameterization in the computation of natural gradient when a sub-model, based on conditional independence assumptions among variables, is employed in the relaxation. The use of restricted models instead of the full covariance plays an important role in the large scale setting, in particular when the interaction pattern between variables in the objective function is sparse. In the talk we will refer to standard results in the literature of the graphical models and in particular of Markov random fields. Based on a joint work with Giovanni Pistone, from Collegio Carlo Alberto.

3.18 On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling

Frank Neumann (University of Adelaide, AU)

License Creative Commons BY 3.0 Unported license
© Frank Neumann
Joint work of Neumann, Frank; Witt, Carsten
Main reference F. Neumann, C. Witt, "On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling," in Proc. of the 24th Int'l Joint Conf. on Artificial Intelligence (IJCAI'15), pp. 3742–3748, AAAI Press, 2015; pre-print available as arXiv:1504.06363v1 [cs.DS], 2015.
URL http://arxiv.org/abs/1504.06363v1

Evolutionary algorithms have been frequently used for dynamic optimization problems. With this paper, we contribute to the theoretical understanding of this research area. We present the first computational complexity analysis of evolutionary algorithms for a dynamic variant of a classical combinatorial optimization problem, namely makespan scheduling. We study the model of a strong adversary which is allowed to change one job at regular intervals. Furthermore, we investigate the setting of random changes. Our results show that randomized local search and a simple evolutionary algorithm are very effective in dynamically tracking changes made to the problem instance.

3.19 Information-Geometric Optimization

Yann Ollivier (University Paris Sud, FR)

License Creative Commons BY 3.0 Unported license
© Yann Ollivier
Joint work of Ollivier, Yann; Akimoto, Youhei; Arnold, Ludovic; Auger, Anne; Hansen, Nikolaus
Main reference Y. Ollivier, L. Arnold, A. Auger, N. Hansen, "Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles," arXiv:1106.3708v3 [math.OC], 2014.
URL http://arxiv.org/abs/1106.3708v3

In this talk I gave an introduction to information-geometric optimization (IGO). I started with how gradient descent is quite unsatisfactory due to great sensitivity to problem encoding (e.g., scaling of each coordinate). Moving to the space of probability distributions over the original search space can help, by evolving a probability distribution in an intrinsic, coordinate-independent way. I presented the IGO-ML (maximum likelihood) family of algorithms, which update a probability distribution based on sampling and increasing the log-likelihood of the best sampled points, similarly to the cross-entropy method but modified as to obtain coordinate-invariance. This applies both to discrete and continuous search

spaces, and recovers known algorithms from basic principles, such as PBIL (starting with Bernoulli distributions on the hypercube) and CMA-ES (starting with Gaussian distributions). Interestingly, IGO-ML has a guarantee of improvement for finite learning rates, as opposed to usual gradient descents for which such a guarantee requires infinitesimal steps. Moreover, from its information-geometric construction, IGO-ML has a guarantee of minimal change in diversity, which could help for multimodal optimization, as is confirmed by preliminary experiments.

3.20 Neutrality in Fitness Landscapes

Peter F. Stadler (Universität Leipzig, DE)

License Creative Commons BY 3.0 Unported license
© Peter F. Stadler

Consider a fitness landscapes (X, f). For simplicity we take the configuration space X as a graph, and f as real-valued function on the vertices. Concepts such as adaptive walks, gradients, etc. are well-defined in an obvious way whenever f is injective. Local failures of injectivity, i.e., the presence of adjacent points in X with the same values are usually referred to as neutrality. Degeneracies and neutrality come in at least two different flavors: (1) Degeneracy may be a consequence of coarse grained "increments" in the fitness function itself as in the case of NK models or spin glass models with small integer coefficients. (2) Degeneracy may also arise from intrinsic symmetries in the structure of the of the cost function, as in the case of short-range spin glasses. The consequence of neutrality, and possibly more general types of degeneracies is very poorly understood in general and both positive and negative effects of optimization have been reported: In the case of RNA landscapes, neutral networks seem to improve adaptability and search space exploration. Work by van Nimwegen & Crutchfield a decade ago, on the other hand described "entropic barriers". Furthermore, ruggedness and neutrality are in principle independent properties and can be tuned in random landscape models independently of each other.

3.21 First Steps Towards a Runtime Comparison of Natural and Artificial Evolution

Dirk Sudholt (University of Sheffield, GB)

License Creative Commons BY 3.0 Unported license
© Dirk Sudholt
Joint work of Paixão, Tiago; Pérez Heredia, Jorge; Sudholt, Dirk; Trubenová, Barbora
Main reference T. Paixão, J. Pérez Heredia, D. Sudholt, B. Trubenová, "First Steps Towards a Runtime Comparison of Natural and Artificial Evolution," in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation (GECCO'15), ACM, pp. 1455–1462, 2015; pre-print available as arXiv:1504.06260v2 [cs.NE], 2015.
URL http://dx.doi.org/10.1145/2739480.2754758
URL http://arxiv.org/abs/1504.06260v2

Evolutionary algorithms (EAs) form a popular optimisation paradigm inspired by natural evolution. In recent years the field of evolutionary computation has developed a rigorous analytical theory to analyse their runtime on many illustrative problems. Here we apply this theory to a simple model of natural evolution. In the Strong Selection Weak Mutation (SSWM) evolutionary regime the time between occurrence of new mutations is much longer

than the time it takes for a new beneficial mutation to take over the population. In this situation, the population only contains copies of one genotype and evolution can be modelled as a (1+1)-type process where the probability of accepting a new genotype (improvements or worsenings) depends on the change in fitness. We present an initial runtime analysis of SSWM, quantifying its performance for various parameters and investigating differences to the (1+1) EA. We show that SSWM can have a moderate advantage over the (1+1) EA at crossing fitness valleys and study an example where SSWM outperforms the (1+1) EA by taking advantage of information on the fitness gradient.

3.22 The Potential of the Swarm's Potential: Convergence, Stopping, Runtime, Stagnation

Rolf Wanka (Universität Erlangen-Nürnberg, DE)

License Creative Commons BY 3.0 Unported license
© Rolf Wanka
Joint work of Bassimir, Bernd; Raß, Alexander; Schmitt, Manuel; Wanka, Rolf
Main reference B. I. Schmitt, "Particle Swarm Optimization Almost Surely Finds Local Optima," Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, FAU University Press, 2015.
URL http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:29-opus4-61621

After a short introduction into the continuous search heuristic "Particle Swarm Optimization" (PSO), the notion of the "potential" of the swarm is presented. The multifarious applications are shown: It can be shown that PSO in the 1-dimensional case converges towards a local optimum almost surely. As in the D-dimensional case ($D > 1$) the potential can be distributed unevenly among the dimensions, the movement equation is slightly changed to incorporate the potential. This can also be used to measure how often this rescue trick has to be used, in order to recognize that (presumably) the swarm is close to an optimum solution. Additionally, the potential can be used to define a measure of closeness to optimality such that, with the help of drift theory, a 1-dimensional runtime analysis can be performed. Finally, it can be explained in terms of the potential why sometimes the number of particles does not suffice to find a local optimum.

References

1 M. Schmitt. Convergence Analysis for Particle Swarm Optimization. Doctoral Thesis, University of Erlangen-Nuremberg, FAU University Press, 2015.[urn:nbn:de:bvb:29-opus4-61621]
2 M. Schmitt, R. Wanka. Particle Swarm Optimization Almost Surely Finds Local Optima. Theoretical Computer Science, 561A (2015) 57–72. [doi:10.1016/j.tcs.2014.05.017]
3 Rass, A.; Schmitt, M.; Wanka, R.: Explanation of Stagnation at Points that are not Local Optima in Particle Swarm Optimization by Potential Analysis. in: Companion of Proc. 17th Genetic and Evolutionary Computation Conference(GECCO); 2015. [doi:10.1145/2739482.2764654]
4 B. Bassimir, M. Schmitt, R. Wanka. How Much Forcing is Necessary to Let the Results of Particle Swarms Converge? in: Proc. Int. Conf. on Swarm Intelligence Based Optimization (ICSIBO), pp.98–105, 2014. [doi:10.1007/978-3-319-12970-9_11]

3.23 Population Size vs. Mutation Strength for the (1+λ) EA on OneMax

Carsten Witt (Technical University of Denmark – Lyngby, DK)

License Creative Commons BY 3.0 Unported license
© Carsten Witt
Joint work of Gießen, Christian; Witt, Carsten
Main reference C. Gießen, C. Witt, "Population Size vs. Mutation Strength for the (1+λ) EA on OneMax," in Proc. of the 2015 Annual Conf. on Genetic and Evolutionary Computation Conference (GECCO'15), pp. 1439–1446, ACM, 2015.
URL http://dx.doi.org/10.1145/2739480.2754738

The (1+λ) EA with mutation probability c/n, where $c > 0$ is an arbitrary constant, is studied for the classical OneMax function. Its expected optimization time is analyzed exactly (up to lower order terms) as a function of c and λ. It turns out that $1/n$ is the only optimal mutation probability if $\lambda = o(\ln n \ln\ln n / \ln\ln\ln n)$, which is the cut-off point for linear speed-up. However, if λ is above this cut-off point then the standard mutation probability $1/n$ is no longer the only optimal choice. Instead, the expected number of generations is (up to lower order terms) independent of c, irrespectively of it being less than 1 or greater. The results are obtained by a careful study of order statistics of the binomial distribution and variable drift theorems for upper and lower bounds.

4 Working Groups

4.1 Constrained Blackbox Optimization Benchmarking

Organized by Dimo Brockhoff (INRIA and University of Lille 1, FR)

Participants: Youhei Akimoto, Dirk Arnold, Anne Auger, Hans-Georg Beyer, Dimo Brockhoff, Carlos Fonseca, Nikolaus Hansen.

4.1.1 Summary

The breakout session on constrained blackbox optimization benchmarking was one among several breakout sessions held during the Dagstuhl seminar 15211 "Theory of Evolutionary Algorithms". Its main focus was to bring forward an extension of the well-established Blackbox Optimization Benchmarking exercise (BBOB) and its underlying Comparing Continuous Optimizers (Coco) platform towards constrained problems. The discussion covered in particular the concrete topics of designing constrained benchmark functions, performance assessment in the constrained case, and practical issues of benchmarking constrained blackbox optimization algorithms. This summary is expected to serve as a reference and the starting point to eventually provide an extension of the Coco platform towards constrained optimization to the research community.

4.1.2 Overview

The constrained blackbox optimization benchmarking breakout session took place in the evening (8:30pm–10:30pm) of the first day of the Dagstuhl seminar (May 18, 2015) and was attended by Youhei Akimoto, Dirk Arnold, Anne Auger, Hans-Georg Beyer, Dimo Brockhoff, Nikolaus Hansen, and Carlos Fonseca.

The breakout session started with a brief summary of known benchmarking exercises, the participants were aware of and used in their previous research (Sec. 4.1.3). Here, we will give a slightly extended overview of the state of the art in constrained optimization benchmarking. Later on, the discussion covered the topics of designing new benchmark functions, the interface between the test problem and the optimization algorithm (Sec. 4.1.4), and general aspects of performance assessment in the constrained case (Sec. 4.1.5). Throughout the discussion and this summary, we thereby assume the following constrained numerical optimization problem to be solved:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to } g_i(x) \geq 0 \text{ for all } 1 \leq i \leq m$$

in which we call f the objective function (with continuous domain) and the g_i's are called constraint functions.

4.1.3 State-of-the-Art Benchmarking

The state of the art in benchmarking numerical optimization algorithms on constrained problems centers around a few benchmark suites. Probably the most common benchmark suites in the context of blackbox optimization are related to the CUTE/CUTEr/CUTEst [1, 3, 4] framework and the competitions and special sessions organized at CEC 2006 [6] and CEC 2010 [7]. At CEC in 2008, 2010, 2015, additional competitions were held which covered bound constrained problems.[1]

The latest CUTEst benchmark suite provides a general interface to more than 1000 test functions (specified in the Standard Input Format (SIF)). About 300 of them are constrained and scalable in the number of variables. Their constraints range from bound constraints over linear and quadratic constraints to general non-linear constraints. Both problems with inequality and equality constraints are present and the objective and constraint functions are separately callable. If available, derivatives are provided as well as known bounds on objectives and constraints. Besides implementations of the problems' objective and constraint functions and interfaces to several commonly used numerical optimization algorithms such as BOBYQA, Direct Search, or NEWUOA, the CUTE/CUTEr/CUTEst framework does not prescribe or provide facilities for performance assessment.

The latest CEC 2010 benchmark suite of constrained problems evolved from the earlier CEC 2006 testbed. It now contains 18 problems overall with a variety of constraint difficulties (bound constraints, inequality constraints, equality constraints). The objective functions are thereby often simple (e.g., the maximum over the variables; one third of the functions are separable) and the constraints can be relatively difficult (there are for example 18 equality constraints but only 21 inequality constraints present). Equality constraints are transformed into inequality constraints with a precision of $\epsilon = 10^{-4}$ by changing $g_i(x) = 0$ into $|g_i(x)| - \varepsilon \leq 0$. The number of resulting inequality constraints per problem ranges from one to four. Regarding the performance assessment of optimization algorithms, the CEC 2010 competition prescribes to report statistics about the objective function values and the number of constraint violations over 25 algorithm runs at three predefined budgets in two dimensions (10-D and 30-D). Constraint function evaluations are thereby to be counted as any other objective function evaluation. Source code for the performance assessment or the creation of LaTeX tables is not provided.

[1] See http://www3.ntu.edu.sg/home/epnsugan/index_files/cec-benchmarking.htm for further details.

There are several other constrained benchmark functions, e.g., in the GLOBAL Library, the Hock/Schittkowski problems, Mittelmann's collection of AMPL problems, or in the COPS library, which, however, have not been discussed during the breakout session. For a more detailed list, we refer the interested reader to the web page of Arnold Neumeier at http://www.mat.univie.ac.at/~neum/glopt/test.html#test_constr.

All mentioned test suites and benchmarking exercises for constrained optimization have in common that they only provide implementations of the test functions and a rudimentary description of how to assess performance of algorithms (e.g., prescribed budgets of function evaluations at which the objective function values and constraint violations are supposed to be reported). Two exceptions are the (independently developed) frameworks PAVER2 [2] and optimizationBenchmarking.org [8], which are not specifically designed for constrained problems. Both frameworks allow to read in data from algorithm runs (in a standardized format) and to display the data in tables and data/performance profiles. The former framework, in addition, allows to detect inconsistencies in the data by running pre-defined tests. None of the available frameworks, however, allow for a "fully automated" benchmarking in which *all* the tedious and recurring steps of data acquisition, processing, and visualization are provided to the user.

As to *unconstrained* optimization, however, the state-of-the-art in automated algorithm benchmarking is more developed. A recent effort towards the goal of automated benchmarking is the Comparing Continuous Optimizers platform (Coco, see coco.gforge.inria.fr) and the corresponding Blackbox Optimization Benchmarking testbed and workshop series (BBOB, see the same link). The Coco platform currently offers two testbeds: one with 24 noiseless and one with 30 noisy test functions – all of them are unconstrained. The performance assessment is based on the well-established concept of target-based run lengths where the expected running time (ERT) to reach a certain target precision of the objective function is used as main performance criterion. The Coco platform is available in five different languages (C/C++, python, MATLAB/Octave, Java, and R). Algorithms, written in any of the provided languages, can be easily plugged to Coco by calling a generic objective function which both provides the function evaluation and collects the data for the performance assessment itself. A postprocessing script, written in python, then allows to display various plots and tables about the algorithm performance automatically up to the compilation of pre-prepared LaTeX templates.

Six BBOB workshops and special sessions have been organized around the Coco platform so far and the data sets of about 150 algorithm variants are available online. They can also all be used easily within the Coco postprocessing to compare the performance of algorithms and algorithm variants, see for example [5].

The participants of the breakout session had no doubt that a constrained BBOB testbed with the automated data gathering, performance assessment, and plotting functionality of the Coco platform will trigger a new interest in constrained blackbox optimization – especially in the communities of numerical blackbox optimizers and evolutionary computation. The resulting open access collection of easily comparable algorithm data will furthermore, as in the unconstrained BBOB case, allow eventually for recommendations on which algorithms to use for certain problem classes.

4.1.4 Benchmark Functions

In terms of the definition of a benchmark function suite, all participants quickly agreed on the following essential points:

- Functions shall be scalable with the search space dimension.
- Constraints in practice are often linear and/or of a blackbox type.
- There exist functions for which knowing the constraint function(s) $g(x)$ is crucial to find the optimum quickly (for example when the objective function's optimum lies in a very small basin of attraction but the constraint function(s) give(s) hints towards the optimum.
- The extension of the Coco platform should not distinguish between finding the feasible region and optimizing within it. An algorithm for constrained blackbox optimization should be able to do both simultaneously.

The discussion then focused on the general aspects of defining a concrete set of benchmark functions, especially with respect to which objective and constraint functions should be used when extending the BBOB testbed to the constrained case.

Given that already the sphere function with a single linear constraint seems difficult for many common evolutionary algorithms, the participants agreed on aiming to have three main function groups/testbeds in an initial constrained blackbox optimization exercise:

- linear orthogonal constrained problems (which includes the special case of bound constraints)
- problems with linear constraints which are not orthogonal
- problems with true blackbox constraints (some of the already established test problems of the BBOB testbed could be used as constraints)

Related to the previous topic was a discussion about the concrete interface between the constrained testbed and the optimization algorithm. In the cases of linear constraints, the breakout session's participants could imagine to allow the algorithm knowledge about the constraints' normal vectors and a support vector. Another option, working in all three above cases, is to expose all values $g_i(x)$ directly. The third option is to provide only a binary response about feasibility/infeasibility of a search point – even if this setting might be too restrictive for some search algorithms. In all cases, allowing the algorithm to make recommendations about its best estimate of the optimum seems to make sense like in the noisy case. A recommended point will thereby replace the last evaluated point for the performance assessment.

Though the discussion stayed focused around the mentioned topics and several concrete assumptions have been fixed throughout the breakout session, some important questions remain open: Probably the most important ones are related to the usage of the optimum in the performance assessment: in general, we do not always know the optimum analytically when using arbitrary constraint functions – even if we know the optimum of the (unconstrained) objective function. One open question is then, for example, how to use "the best function value known so far" as a reference in the performance assessment: it raises the question of how to approximate the best known value well and how to deal with potentially improving values over time. Note here that the constrained CEC benchmarks, mentioned earlier, report the best known value so far. However, the approach is not based on target-based run lengths as in the BBOB/Coco framework for which a good knowledge of the optimal function value seems more critical. Another open question is the relation between certain assumptions on the objective and constraint functions and existence guarantees of analytical descriptions of the optima.

As the first step towards a general constrained blackbox optimization benchmarking exercise, the participants finally thought that a first test suite should be defined, which does not include all existing BBOB functions but simpler ones for which the optima are

known also in the constrained case. This should be feasible under regularity and convexity assumptions.

4.1.5 Performance Assessment

The main difference between benchmarking unconstrained and constrained optimization algorithms is the way how the information about sampled infeasible solutions are used in the performance measures. Inherently, assessing performance in the constrained case can be defined as a multiobjective problem, with a trade off between the minimization of the objective function f and the minimization of the m constraint violations.

To not complicate the performance evaluation, the participants agreed that looking only at feasible search points in the performance measures seems to make sense as a first step when generalizing the Coco platform. This allows to use the same standard performance measures such as the expected running time (ERT) to hit a certain f-value target precision as in the unconstrained case. However, the number of used constraint evaluations should also be taken into account in addition to the number of function evaluations: in practice, an algorithm for constrained optimization might not even evaluate the objective function if the evaluation of the constraint functions already renders a search point infeasible. This is especially the case if both the objective and (some of) the constraint functions are of a blackbox type and expensive to compute.

Since it is important to know how the algorithm performs over time, in such a case, we need to know how the algorithm performance scales with the number of objective function evaluations *and* constraint function evaluations. We therefore suggest to have the axes in standard BBOB/Coco plots which are related to time (such as in the data profiles) as well as the definition of the expected running time to be defined according to either the number of objective function evaluations, the number of constraint function evaluations, or according to the sum of both. On the other hand, there might be cases, where both the computation of the objective function and the computation of the constraint functions are by themselves cheap but rely on some (joint) computations which are expensive (e.g., a numerical simulation). In this case, it seems reasonable to measure the search costs by the max of f-calls and g-calls, i.e., the number of the expensive simulations. All in all, we can expect that good algorithms for one case might not at all perform well in other cases. This is another good reason for pursuing the benchmarking efforts of the Coco framework to find out quantitative results. In order to do so, the benchmark suite might want to potentially offer the information about which objective function/constraint function combination is expensive to evaluate to the algorithms.

4.1.6 Conclusion

The breakout session on constrained blackbox optimization benchmarking covered the topics of designing new benchmark functions, the interface between the test problem and the optimization algorithm, and general fundamental aspects of performance assessment in the context of an extension of the well-established BBOB/Coco platform towards constrained optimization problems.

Beside making progress towards the concrete definition of a new constrained blackbox optimization benchmarking testbed based on the BBOB/Coco benchmarking exercise, the breakout session's discussion resulted in several new research questions – both of theoretical and practical nature. The breakout session has been very fruitful and has seen a lively and efficient discussion – as such, breakout sessions should therefore be kept in the program of future editions of the Dagstuhl seminar series.

Acknowledgments

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 I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint. CUTE: Constrained and Unconstrained Testing Environment. *ACM Transactions on Mathematical Software (TOMS)*, 21(1):123–160, 1995.
2 M. R. Bussieck, S. P. Dirkse, and S. Vigerske. PAVER 2.0: An Open Source Environment for Automated Performance Analysis of Benchmarking Data. *Journal of Global Optimization*, 59(2–3):259–275, 2014.
3 N. I. M. Gould, D. Orban, and P. L. Toint. CUTEr and SifDec: A Constrained and Unconstrained Testing Environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2005.
4 N. I. M. Gould, D. Orban, and P. L. Toint. CUTEst: a Constrained and Unconstrained Testing Environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 2015.
5 N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posík. Comparing Results of 31 Algorithms from the Black-box Optimization Benchmarking BBOB-2009. In J. Branke et al., editors, *GECCO workshop on Black-Box Optimization Benchmarking (BBOB'2010)*, pages 1689–1696. ACM, July 2010.
6 J. J. Liang, T. P. Runarsson, E. Mezura-Montes, M. Clerc, P. N. Suganthan, C. A. Coello Coello, and K. Deb. Problem Definitions and Evaluation Criteria for the CEC 2006 Special Session on Constrained Real-Parameter Optimization. Technical report, Nanyang Technological University, 2006.
7 R. Mallipeddi and P. N. Suganthan. Problem Definitions and Evaluation Criteria for the CEC 2010 Competition on Constrained Real-Parameter Optimization. Technical report, Nanyang Technological University, 2010.
8 T. Weise. optimizationbenchmarking.org, 2015. last accessed on June 17, 2015.

4.2 Bridging the Gap Between Experiments and Theory Using Feature-Based Run-Time Analysis

Organized by Frank Neumann (University of Adelaide, AU) and Heike Trautmann (Universität Münster, DE)

Feature-based analysis has been used to explore the performance of evolutionary algorithms based on features of given instances. Experimental studies show that problem instances may be classified as easy or difficult for a given EA based on instance features. Such classifications can be used for algorithm selection and/or parameter configuration. The goal of this breakout session was to discuss how mathematical analyses can contribute to this experimentally driven research area. A natural candidate is the runtime analysis of EAs taking into account problem features but any theoretical approach that supports these experimental investigations is highly welcome.

The session started with a brief introduction on feature-based analysis such that participants who were not familiar with this area could contribute to the discussion. The breakout session was targeted as a brainstorm session which should discuss approaches that may be applicable and there was the goal to identify some concrete problems that can become a topic of new research work.

The following challenges for a theoretical/mathematical analysis have been identified:

- for NP-hard problems: features will never capture all information about the problem
- which are the most important (ideally deterministic) features?
- features interact in practice

As a consequence, we think that an analysis should start using a single feature combined with a simple algorithm concept.

The break-out group identified the following topics for future research:

- start with classical polynomially solvable combinatorial problems
- examine structural instance features and influence on algorithm performance
- use empirical results in order to get hints to important features
- either runtime estimation directly or configuration of the algorithm to optimize runtime
- candidate problems: shortest path, spanning trees, matchings, graph colouring, k-Sat, finding subgraphs, knapsack, modified versions of OneMax

More specifically:

- Graph Colouring: Runtime (Phase transition) based on the density of the graph
- k-Sat: probability of failure based on ratio of variables and clauses
- knapsack: runtime based on profits and weights

Candidate algorithms for the investigations of these problems are simple algorithms such as the (1+1)-EA and the (1+λ)-EA. Based on the features, it would be interesting to optimize important parameters such as the mutation probability or offspring population size. It is highly recommended that the theoretical investigations utilize empirical studies for validating the alignment of theoretical findings with practical observations.

4.3 Bringing together Evolutionary Computation and Population Genetics

Organized by Tobias Friedrich (Hasso-Plattner-Institut, Potsdam, DE)

Participants: Lee Altenberg, Nick Barton, Arina Buzdalova, Carlos M. Fonseca, Tobias Friedrich, Kenneth A. De Jong, Joshua D. Knowles, William B. Langdon, Per Kristian Lehre, Jonathan L. Shapiro, Dirk Sudholt, Peter F. Stadler.

One of the three themes of this Dagstuhl meeting was "theory of natural evolution". The meeting was joined not only by computer scientists working in evolutionary computation, but also by a number theoretical biologists and other related fields.

At the beginning of evolutionary computation, a lot of inspiration was drawn from biology. However, we discussed that for example the infinite population of Michael Vose is not considered useful for real GAs. It was argued that also in theoretical biology simplified models are essential. Some operators like crossover are very interesting for both communities. The differences often come from the different aims (explaining nature vs. optimization). There are still surprisingly many similarities. Another joint aspect are island models. Even evolutionary algorithms which appear to be rather different from nature, like genetic programming, have similar features: The common problem of bloat in genetic programming is also well known for genes in nature.

4.4 Theory in Multimodal Optimisation

Organized by Christine Zarges (University of Birmingham, GB)

Participants: Hans-Georg Beyer, Carola Doerr, Tobias Glasmachers, Thomas Jansen, Oswin Krause, Pietro S. Oliveto, Günter Rudolph, Christine Zarges.

4.4.1 Summary

The main aim of this breakout session was to discuss potential routes for theoretical research in multimodal optimisation and how theory could help to promote this important application area. The discussion centered around three different aspects: the goal of multimodal optimisation, benchmark functions for theory and the evaluation and comparison of different algorithms. Given the composition of the group, aspects relevant for both discrete and continuous domains were discussed.

4.4.2 What is the goal of Multimodal Optimisation?

There has been some discussion about the appropriateness of the term *multimodal optimisation* and what it incorporates. Multimodal optimisation is concerned with problems that have a number of different local and global optima. However, there are different optimisation goals one could be interested in. On one hand, one can take a global perspective and consider the goal of finding a single optimum of the given problem. For this case there are already a couple of theoretical results available, in particular in the context of pseudo-Boolean and combinatorial optimisation. On the other hand, one could be interested in finding several different optima, either in a simultaneous or sequential fashion. It was agreed that this latter perspective should be called *multi-local optimisation*. Different evaluation criteria in the context of multi-local optimisation are discussed in Section 4.4.4.

4.4.3 How does an appropriate benchmark look like?

In the context of discrete optimisation some analysis has been performed on the TwoMax function, which can be defined as follows:

$$\text{TwoMax} : \{0,1\}^n \to \mathbb{R} \text{ with}$$
$$\text{TwoMax}(x) = \max(|x|_0, |x|_1),$$

where $|x|_0$ and $|x|_1$ denote the number of zeros and ones in x, respectively. This function has two global optima and different algorithms including diversity preserving mechanisms have been considered in its context. However, it has been agreed that a good benchmark function should have a much larger number of optima. After some discussion, we came up with a generalisation of the TwoMax function, which we call k-Max. The function has k different optima which can be selected randomly or purposefully. All optima have two parameters: slope ($a_i > 0$) and offset ($b_i > 0$). All other search points have a fitness value depending on the distance to the closest optimum such that each search point is in the basin of attraction of its closest optimum. k-Max can formally be defined as follows:

$$k\text{-Max} : \{0,1\}^n \to \mathbb{R} \text{ with}$$
$$k\text{-Max}(x) = n - \left(a_i \cdot H\left(x, x^{(i)}\right) + b_i\right) \quad \text{with } i = \text{argmin}_{j \in \{1,\dots,k\}} H\left(x, x^{(j)}\right)$$

where $H(x, x^{(i)})$ denotes the Hamming distance of search point x to optimum $x^{(i)}$. It is noted that this function exhibits a similar 'constructive pattern' (namely random locations of local optima and scaling) as Shekel's foxholes in continuous optimisation.

4.4.4 How should different algorithms be compared?

Based on the k-Max function, several algorithms were discussed. As usual algorithms can be compared under a runtime (*How long does your algorithm need to achieve a given goal?*) or fixed budget (*Given a fixed budget, how well can your algorithm perform?*) perspective. However, with respect to the evaluation of algorithms in a multimodal context there are other important aspects to consider:

- How many optima did the algorithm find?
- What quality do the optima have?
- What is the distribution of the optima?

A crucial insight is that there is not a single solution to the problem of algorithm evaluation since an appropriate metric depends on the concrete application at hand. The group agreed that evaluation criteria are an important question for future research. For some of these criteria and for several optimisation problems, the state-of-the-art technique in real-world optimisation challenges are *(random or deterministic) restarts*. The group agreed that it is an interesting unsolved question to classify problems for which population-based algorithms are superior to such restart strategies.

4.5 Theory of Genetic Programming

Organized by Pietro S. Oliveto (University of Sheffield, GB)

Participants: Lee Altenberg, Maxim Buzdalov, Arina Buzdalova, Benjamin Doerr, Thomas Jansen, Timo Kötzing, William B. Langdon, Alberto Moraglio, Frank Neumann, Pietro S. Oliveto, Dirk Sudholt, Carsten Witt, Christine Zarges.

4.5.1 Summary

The goal of Genetic Programming (GP) is to evolve computer programs i.e., executable functions with desired functionality. While numerous successful applications of GP have been reported, there is limited and fragmented theoretical understanding of the working principles of GP. Compared to the theoretical analysis of genetic algorithms, the study of GP poses several extra challenges, eg., tree representations of variable length. Some preliminary work on the theory of GP has been undertaken by various participants of this Dagstuhl seminar. The purpose of this breakout session was to discuss potential feasible and promising routes to build a theoretical foundation of GP, taking advantage of the progresses in the runtime analyses of evolutionary algorithms for function optimisation.

4.5.2 How to approach the theoretical analysis of GP?

The theoretical understanding of Evolutionary Algorithms (EAs) has grown considerably in the past 20 years with the development of mathematical techniques that have allowed runtime analyses of more and more sophisticated EAs for problems bearing increasing similarities with those tackled in real-world applications. The achievement of such results has been possible by gradually building upon techniques constructed through the analysis of extremely

simplified EAs, such as the notorious (1+1) EA, for artificially constructed test problems with particular characteristics. Popular examples of such problems are the Onemax and Trap functions that have characteristics that make the optimisation respectively easy and hard for EAs. Given the successful progress made over the years by the EA community, a promising direction to build a theoretical foundation of GP would be to follow a similar route to that taken by the field for the runtime analysis of EAs for function optimisation. The first steps in following such a strategy would be to identify the analogues for GP of the (1+1) EA and of useful test problems such as Onemax and Trap.

4.5.3 What should our GP algorithms evolve?

While the original motivation behind Genetic Programming (GP) was to introduce a tool that would allow the evolution of computer programs (i.e., executable functions with a desired functionality), GP techniques have also been used to address various kinds of other problems. Apart from evolving functions depending on input/output relationships, other perspectives have been to use GP to address problems that are typical to the field of machine learning or to evolve *designs* with a particular behaviour. Also, several different ways to represent a computer program have been proposed in the literature. Apart from the syntax trees used in the original GP, various linear representations have been proposed in subclasses of GP such as linear genetic programming or cartesian genetic programming. Since the use of syntax trees is the most wide spread, it was agreed that it is sensible to consider this kind of representation in the early foundational steps of GP theory.

Recent work concerning geometric semantic GP (GSGP), a non-standard form of GP, was also considered. Here the search operators are designed to induce a Onemax fitness landscape for any Boolean function, thus facilitating the obtainment of runtime analysis results [5]. It was remarked though that GSGP, even if efficient in the training phase, cannot possibly generalise well (at least on Boolean functions). Another research direction that was proposed was the design of refined search operators for GSGP that are provably efficient in the training phase and provably good at generalisation (in the PAC-learning sense) on some PAC-learnable sub-classes of all Boolean functions.

4.5.4 Test problems for GP

Since the goal of GP is to evolve functions with a desired functionality, an essential characteristic for a GP test problem is that the solution should be an executable function and qualities of candidate solutions should depend on how well they map inputs to outputs. Initial foundational steps should identify test problems that are particularly easy and particularly hard to evolve for a GP system. Such problems should have characteristics that will facilitate the progressive development of mathematical techniques to allow the analysis of more sophisticated algorithms for the same problems and of simple GP algorithms for more complicated problems. Several GP benchmarks were considered including the Max problem, even-n-parity, multiplexor problems [4], Royal trees [6, 7], tree-string problems [2] and K landscapes [8].

A promising domain is that of Boolean functions, especially because they are often used as benchmark problems in experimental GP research. Parity is well known to be a hard problem in the evolvability and GP communities, while other Boolean functions such as AND or ON (i.e., the Boolean function that always returns 1 as output) are known to be easy. A desirable characteristic for easy problems might be the lack of epistasis – that the change in fitness by flipping any bit is independent of the state of the bits at other loci, a

characteristic present in the Onemax function. No final conclusions were made concerning what the analogue of the Onemax test problem for GP is.

4.5.5 The Simple-GP Algorithm

Genetic Programming algorithms behave similarly to standard genetic algorithms (GAs) in the sense that they both have high crossover rates and very small mutation rates. A significant difference though is that GP trees vary in size and depth during the evolutionary process while GA genotypes remain of fixed size during the entire evolution. Hence, while the operators of a (1+1) GP may be simplified in a similar fashion to those of the (1+1) EA and random local search, they should allow trees to shrink and grow in size. The operators used in the few runtime analyses available in the literature [1, 3] (i.e., Insert, Substitute and Delete) seem suitable for the purpose. In these works symptoms of bloat have appeared even though no crossover was used.

4.5.6 Runtime Analysis

In order to determine how the computational resources depend on the size of the problem a notion of scalability has to be decided. Several ways could be used to represent the size of the problem. In some works the number of nodes in the optimal tree was used. The number of input variables of the benchmark function seems to be the most general and appropriate measure. Generalisation is an important issue to be considered in the analysis. The target function may not necessarily be the exact function (i.e., one of the optimal solutions) but one that approximates it nicely.

References

1 G. Durrett, F. Neumann, and U.-M O'Reilly. Computational complexity analysis of simple genetic programming on two problems modeling isolated program semantics. In *Proc. of the 11th Workshop on Foundations of Genetic Algorithms (FOGA'11)*, pages 69–80. ACM Press, 2011.

2 S. Gustafson, E. Burke, and N. Krasnogor. The tree-string problem: An artificial domain for structure and content search. In *Proceedings of EuroGP 2005*, pages 215–226. Springer, 2005.

3 T. Kötzing, A. M. Sutton, F. Neumann, and U.-M. O'Reilly. The max problem revisited: the importance of mutation in genetic programming. *Theoretical computer science*, 545:91–107, 2014.

4 W. B Langdon and R. Poli. *Foundations of Genetic Programming*. Springer, 2001.

5 A. Moraglio, A. Mambrini, and L. Manzoni. Runtime analysis of mutation-based geometric semantic genetic programming on boolean functions. In *Proceedings of the 12th workshop on foundations of genetic algorithms (FOGA '13)*, pages 119–131. ACM Press, 2013.

6 B. Punch, D. Zongker, and E. Goodman. The royal tree problem, a benchmark for single and multiple population genetic programming. In P. Angeline and K. Kinnear, editors, *Advances in Genetic Programming 2*, pages 299–316. MIT Press, 1996.

7 Walter Alden Tackett. *Recombination, selection, and the genetic construction of computer programs*. PhD thesis, University of Southern California, 1994.

8 L. Vanneschi, M. Castelli, and L. Manzoni. The k landscapes: a tunably difficult benchmark for genetic programming. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation (Gecco 2011)*, pages 1467–1474. ACM Press, 2011.

4.6 Continuous Optimization: Fitness vs. Ranks

Organized by Oswin Krause (University of Copenhagen, DK)

The current state of the art of continuous evolutionary optimization is to perform steps based on value comparisons and ranking of points, in contrast to of using the function values directly. This strategy, also known as "function value free" optimization, is very successful in practice as it gives invariance to monotone transformations of the objective function. This is in contrast to other randomized search techniques like stochastic gradient descent.

During the session we focused on the issue of noise handling. We concluded that proper ranking becomes hard once we are close to the optimum, especially when the noise is additive. In this case, ranking will always introduce a bias into the gradient estimate, which can (depending on the algorithm) slow down or prevent convergence to the optimum.

Therefore we discussed whether it is better to increase the population size or to reevaluate points in order to decrease the bias. Algorithmically, there is a clear trade-off between number of re-evaluations of an point and the number of steps taken, given a fixed budget of function evaluations, and it is in general unclear whether the bias induced by noise will prevent convergence or just slow it down (this depends on many aspects, e.g., how gradients are used, whether the noise is additive or multiplicative or something else, etc). In contrast, fitness-based algorithms (in contrast to rank-based, during the debate we borrowed the term "fitness proportional selection" to make the distinction explicit, although this changed the original meaning of the term) can easily be implemented to be unbiased and to converge (see standard results on (stochastic) gradient descent, SGD), while the trade-offs are not so important due to the fact that increasing the population size always decreases the noise on the gradient estimate, and noise can also average out over multiple steps (e.g., in standard SGD, due to the cooling schedule).

Still, removing ranks and working with plain fitness values in this case leads to a number of new problems as we loose the invariance against monotone transformations. Thus the algorithm needs to adapt to the curvature of the target function.

We discussed, whether there are useful transformations of the objective function other than ranking, however the easily computable ones are inferior and also need to be estimated from the noisy samples. It also has to be taken care that the transformation does not cause the vector field to contain circles, which can prevent convergence.

4.7 Neutrality

Organized by Peter F. Stadler (Universität Leipzig, DE)

Neutrality is clearly still a poorly understood topic. In biological landscapes, neutrality and redundancy does appear at sometimes extensive scales. It remains open under which conditions neutrality, or more precisely neutral networks, are beneficial to the performanace of (EC-type) optimization algorithms. While examples are known where this is the case – using a non-uniform representation of phenotypes that favours good solutions, it is unclear whether uniform representations, in which each phenotype is represented by the same number of genotypes, can also have an advantage. Computational studies by Fonseca, for example, suggest that this is at least hard to achieve. An related question in whether neutrality can effectively be replaced by suitably expanded definitions of neighborhood. Conversely, there are combinatorial optimization problems, such as certain scheduling problems, which do not seem to have a essentially non-redundant or even non-neutral representations. So far, no

convincing strategies to construct efficient EC-type algorithms for the class of problems seems to available. A point made repeatedly is the distinction between redundancy (degeneracy) of the fitness function and neutrality (i.e., local redundancy). This also related to the issue of accessibility among the (phenotypic) representations. These can be effectively non-symmetric, for instance in the case that the redundancy of the encoding is very different for different phenotypes. To what extent this asymmetry is relevant to optimization algorithms remains open. The importance of neutrality in continuous settings remains unclear.

4.8 Issues with Optimization for Machine Learning Using Variational Inference

Organized by Yann Ollivier (CNRS and University Paris Sud, FR)

In this session we discussed specific technical problems that arise in some optimization problems from machine learning. In machine learning, typical optimization problem require finding the parameters of (probabilistic) models to best fit available data. However, there is a risk of overfitting, namely, finding a parameter value that is too precise given the available data. This is especially relevant in high dimension.

Avoiding the overfitting problem can be done by working with a modified objective, in which one is looking, not only for a single optimal parameter value, but for a whole region in parameter space where performance is good. In the "variational inference" approach, one works with probability distributions over the original parameter space, and the modified objective is the average of the original objective under this distribution, minus the entropy of the distribution. The latter term prevents collapsing to a single point. The whole modified objective has an interpretation as a compressed length of the data.

This setting is very similar to ones appearing in evolutionary strategies such as CMA-ES or NES. However, the specific form of the objective (an average of the original objective function) makes it impossible to directly use these strategies; especially, rank reweighting is impossible, which makes it technically much more difficult to update, e.g., a Gaussian distribution in a stable way.

Moreover, in high dimension, it is tempting to include a model selection aspect such as keeping or removing each individual component of the parameter, which puts a discrete optimization problem on top of the continuous parameter search. Once more, one must deal with an entropy term describing the information cost of selecting/deselecting components.

In the session we discussed possible approaches for these problems, but no fully satisfying solution was found during the time of the session, so that more work is clearly needed.

4.9 Effects of Initialisation Process for Random Search Heuristics

Organized by Jonathan E. Rowe (University of Birmingham, GB)

It is known that in some situations (e.g., (1+1)EA on OneMax) that the exact method of initialisation does not have a significant effect on the run time (in this case, at most affecting low order terms). Empirically, however, it has been observed that for some problems the initialisation makes a big difference. One example is Set Cover, where it seems better to start with the empty set than with a random set. One reason for this is that a random set is likely to be feasible (at least on typical benchmark problems). The initial choice then

limits the area of the search space explored. Moreover, the fitness function gives no guidance as to which are the best elements to exclude – all choices improve fitness by one. On the other hand, the empty set is infeasible, and there will be selective pressure towards choosing sets which reduce the constraint violation. It therefore seems that the dependancy on initial conditions arises through asymmetries in the information the fitness function provides for feasible and infeasible solutions. It would be an interesting research goal to analyse this effect theoretically. A number of other examples were discussed where a similar effect might be observed (e.g., TSP, Vertex Cover). One example where the analysis may be tractable is Minimum Spanning Tree, since much is already known about the run-time – although initialisation here is likely to only affect low order terms.

4.10 Theory to Practice in Evolutionary Computation

Organized by Joshua D. Knowles (University of Manchester, GB)

Participants: Dimo Brockhoff, Arina Buzdalova, Kenneth A. De Jong, Carlos M. Fonseca, Tobias Friedrich, Hsien-Kuei Hwang, Joshua D. Knowles, W. B. Langdon, Frank Neumann, Jonathan L. Shapiro.

Our question was: "How can we continue to bridge the gap between the practical application of EAs and the Theory of EAs?" In practical applications, algorithms are designed and tested to obtain maximum performance on complicated real-world optimization, learning and control problems; in theoretical work, the algorithms and problems are often simple in order to admit analysis and focus on core issues. Establishing a bridge from one end of this spectrum to the other is a central challenge of the Theory field (see e.g. [12]) and a core aim of the Theory of EAs series at Dagstuhl. In this breakout session we focused the discussion more particularly on identifying what kinds of practice work make a good target for theoretical analysis, and on the group's personal experiences trying to bring practice and theory closer together.

We began by noting that in the wider Dagstuhl meeting a continuing bridging activity is apparent, and in diverse contexts. For example, we have seen rigorous theoretical work about hybrid algorithms consisting of more than one base algorithm[2] [11], use of auxiliary objective functions with adaptive switching between them [1, 3], asynchronous EAs for handling variability in the objective function latency [10], and several new analyses of sophisticated algorithms based on information geometry [5, 15]. Thus, while theoretical tools are still being sharpened in simpler settings, they are also, it seems, being used 'in anger' to obtain new results of more practical import.

Joshua Knowles next presented joint work on evolutionary algorithms for problems embedded in a resource-constrained context where evaluations of solutions are done by real (e.g., physical / chemical) experiments. In this general setting, certain parts of the feasible solution space are not available for some periods during the optimization run [2] because resources specific to the solution are required for its evaluation and may be depleted (e.g., in a problem to optimize combinatory drug therapies [18], certain drugs may be used up). Knowles suggested this sort of setting would be ripe for theoretical analysis because the optimization function itself can be simple (e.g., OneMax and TwoMax have been used in [2]), allowing a transfer of known results and techniques, yet, at the same time, empirical

[2] Also referred to as Algorithm Portfolios in some works.

results show that the effects of the resource constraints are interesting and can depend on a number of factors. Thus there is plenty in the area to challenge existing theoretical tools.

This raises the question to what extent do practical problems help theory? We did not reach a general answer to this but it was noted by Carlos Fonseca that sometimes practical problems do reinvigorate interest in a latent theoretical result in the literature. He gave the example that a known complexity bound for computing the hypervolume of the union of axis-aligned polytopes from computational geometry [14, 16] had been improved ([8, 19, 9, 7]) as a direct result of interest in evolutionary algorithms using hypervolume computations for assessing the population 'fitness' in multiobjective problems [13, 6, 4]. In general, however, it can be hard to get theoreticians interested in a practical problem and they may be unwilling to help even if they would be capable of doing so!

Kenneth A. De Jong made the point that it is crucial to find the right level of abstraction in describing problems (to make them accessible and interesting to theory) – very much an art rather than a science. Even this may not be enough to simulate immediate interest. Frank Neumann related recent experience in presenting a new benchmark, the Traveling Thief Problem (TTP) [17], which models an important aspect of many real-world problems that they are composed of two or more different combinatorial optimization subproblems (each of which may be relatively easy to solve in isolation). Even though TTP has this practical relevance, it seems difficult, at the present time, to make progress on it from a theoretical perspective and interest has not been piqued. Nevertheless, the existence of the challenge in a well-defined form, with much unimportant detail abstracted away, will hopefully provide impetus for theoretical attacks on it in time.

We concluded by confirming that the aim of bringing together theory and practice was essential to the continuing development of evolutionary methods, and one that should be pursued by practitioners as well as theorists. It is difficult to get the alignment to work – for the practical problem to be one that is ripe and interesting for theory – but we must not give up on this project. Indeed the only way to make this happen is to redouble our efforts to engage with and include practitioners (be it industrial or academic) in theory meetings in future.

References

1. Denis Antipov, Maxim Buzdalov, and Benjamin Doerr. Runtime Analysis of $(1 + 1)$ Evolutionary Algorithm Controlled with Q-learning Using Greedy Exploration Strategy on OneMax+ ZeroMax Problem. In *Evolutionary Computation in Combinatorial Optimization*, pages 160–172. Springer, 2015.
2. Richard Allmendinger and Joshua Knowles. On handling ephemeral resource constraints in evolutionary search. *Evolutionary Computation*, 21(3):497–531, 2013.
3. Arina Buzdalova and Maxim Buzdalov. Increasing efficiency of evolutionary algorithms by choosing between auxiliary fitness functions with reinforcement learning. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 1, pages 150–155. IEEE, 2012.
4. N. Beume. $\mathcal{S}$-Metric Calculation by Considering Dominated Hypervolume as Klee's Measure Problem. *Evolutionary Computation*, 17(4):477–492, 2009.
5. Hans-Georg Beyer. Information geometry optimization in $\mathbb{R}^n$: Open problems and misconceptions. In *Dagstuhl Reports (this issue)*. 2015.
6. N. Beume and G. Rudolph. Faster S-Metric Calculation by Considering Dominated Hypervolume as Klee's Measure Problem. Technical Report CI-216/06, Sonderforschungsbereich 531 Computational Intelligence, Universität Dortmund, 2006. shorter version published at IASTED International Conference on Computational Intelligence (CI 2006).

7 K. Bringmann. Bringing order to special cases of klee's measure problem. In *Mathematical Foundations of Computer Science (MFCS 2013)*, pages 207–218. Springer, 2012.

8 T. M. Chan. A (Slightly) Faster Algorithm for Klee's Measure Problem. *Computational Geometry*, 43(3):243–250, 2010.

9 T. M. Chan. Klee's Measure Problem Made Easy. In *Foundations of Computer Science (FOCS 2013)*, pages 410–419, 2013.

10 Kenneth A. De Jong. Understanding asynchronous evolutionary algorithms. In *Dagstuhl Reports (this issue)*. 2015.

11 Thomas Jansen. Understanding randomised search heuristics. In *Dagstuhl Reports (this issue)*. 2015.

12 Thomas Jansen and R. Paul Wiegand. Bridging the gap between theory and practice. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 61–71. Springer, 2004.

13 Joshua Knowles and David Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *Evolutionary Computation, IEEE Transactions on*, 7(2):100–116, 2003.

14 V. Klee. Can the Measure of $\bigcup_i^n \{a_i, b_i\}$ be Computed in Less Than $O(n \log n)$ Steps? *The American Mathematical Monthly*, 84(4):284–285, 1977.

15 Yann Ollivier. Laplace's rule of succession in information geometry. *arXiv preprint arXiv:1503.04304*, 2015.

16 M. H. Overmars and C. K. Yap. New Upper Bounds in Klee's Measure Problem. *SIAM Journal on Computing*, 20(6):1034–1045, 1991. conference version is oy1988a.

17 Sergey Polyakovskiy, Mohammad Reza Bonyadi, Markus Wagner, Zbigniew Michalewicz, and Frank Neumann. A comprehensive benchmark set and heuristics for the traveling thief problem. In *Proceedings of the 2014 conference on Genetic and evolutionary computation*, pages 477–484. ACM, 2014.

18 Ben G. Small, Barry W. McColl, Richard Allmendinger, Jürgen Pahle, Gloria López-Castejón, Nancy J. Rothwell, Joshua Knowles, Pedro Mendes, David Brough, and Douglas B. Kell. Efficient discovery of anti-inflammatory small-molecule combinations using evolutionary computing. *Nature Chemical Biology*, 7(12):902–908, 2011.

19 Hakan Yildiz and Subhash Suri. On Klee's measure problem for grounded boxes. In *Symposium on Computational Geometry (SoGC 2012)*, pages 111–120. ACM, 2012.

5 Seminar Schedule

Monday

9.00–10.00	Introduction
	Thomas Jansen: Understanding Randomised Search Heuristics (Part I)
10.00–10.30	Coffee Break
10.30–12.00	**Thomas Jansen:** Understanding Randomised Search Heuristics (Part II)
	Hsien-Kuei Hwang: Probabilistic analysis of the (1+1)-evolutionary algorithm
12.15–14.00	Lunch Break
14.00–15.30	**Kenneth A. De Jong:** Understanding Asynchronous Evolutionary Algorithms
	Proposals for Breakout Sessions
15.30–16.00	Coffee Break
16.00–18.00	**Timo Kötzing:** Introduction to Drift Analysis and Stochastic Optimization with Randomized Search Heuristics
	Carsten Witt: Population Size vs. Mutation Strength for the (1+λ) EA on OneMax
20.30	Breakout Session **Dimo Brockhoff:** Benchmarking & Constraints

Tuesday

9.00–10.00	**Nick Barton:** Limits to adaptation
10.00–10.30	Coffee Break
10.30–12.00	**Peter F. Stadler:** Neutrality in Fitness Landscapes
	Dirk Sudholt: First Steps Towards a Runtime Comparison of Natural and Artificial Evolution
12.15–14.00	Lunch Break
14.00–15.30	Breakout Session I
	Frank Neumann and Heike Trautmann: Bridging the Gap Between Experiments and Theory Using Feature-Based Run-Time Analysis
	Tobias Friedrich: Bringing together Evolutionary Computation and Population Genetics
	Christine Zarges: Theory in Multimodal Optimisation
15.30–16.00	Coffee Break
16.00–18.00	**Frank Neumann:** On the Runtime of Randomized Local Search and Simple Evolutionary Algorithms for Dynamic Makespan Scheduling
	Carola Doerr: Analyzing Self-Adaptive Parameter Choices in Discrete Search Spaces
	Maxim Buzdalov: The Unrestricted Black-Box Complexity of Jump Functions

Wednesday

9.00–10.00	**Yann Olivier:** Information-geometric optimization with application to machine learning
10.00–10.30	Coffee Break
10.30–12.00	**Hans-Georg Beyer:** Information Geometry Optimization in $\mathbb{R}^N$ Open Problems and Misconceptions
	Oswin Krause: Does the natural gradient save us?
12.15–13.30	Lunch Break
13.30–15.30	Hike
15.30–16.30	Coffee Break
16.30–18.00	Breakout Session II
	Peter F. Stadler: Neutrality
	Pietro S. Oliveto: Theory of Genetic Programming
	Oswin Krause: Continuous Optimization: Fitness vs. Ranks

Thursday

9.00–10.00	**Arina Buzdalova:** Selection of Auxiliary Objectives with Reinforcement Learning: Overview of Theoretical Results
	Carlos Fonseca: Lyapunov Stability Analysis of a Derandomized Self-Adaptive (1,2)-ES
10.00–10.30	Group Photo and Coffee Break
10.30–12.00	**Anton Eremeev:** On the proportion of fit individuals in the population of a mutation-based genetic algorithm
	Per Kristian Lehre: Level-based Analysis of Genetic Algorithms and Other Search Processes
12.15–14.00	Lunch Break
14.00–15.30	Breakout Session III
	Yann Olivier: Issues with Optimization for Machine Learning Using Variational Inference
	Jonathan Rowe: Effects of Initialisation Process for Random Search Heuristics
	Joshua Knowles: Theory to Practice in Evolutionary Computation
15.30–16.00	Coffee Break
16.00–18.00	**Daniel Johansen:** Open Problems in Industry
	Summary and Discussion of Breakout Sessions

Friday

9.15–10.15	**Lee Altenberg:** Spectral Landscape Theory: Some Color from Infinite Population Analysis
	Rolf Wanka: The Potential of the Swarm's Potential
10.15–10.45	Coffee Break
10.45–12.00	**Roman V. Belavkin:** Fitness Information and Optimal Control of Mutation Rate
	Luigi Malagó: Information Geometry of the Gaussian Distribution in View of Stochastic Optimization: Inverse Covariance Parameterization
12.15	Lunch

Participants

- Youhei Akimoto
Shinshu University, JP
- Lee Altenberg
Konrad Lorenz Institute for Evolution & Cognitio, AT
- Dirk V. Arnold
Dalhousie University, CA
- Anne Auger
INRIA Saclay – Île-de-France, FR
- Nick Barton
IST Austria – Klosterneuburg, AT
- Roman V. Belavkin
Middlesex University, GB
- Hans-Georg Beyer
Fachhochschule Vorarlberg, AT
- Dimo Brockhoff
INRIA – University of Lille 1, FR
- Maxim Buzdalov
ITMO University – St. Petersburg, RU
- Arina Buzdalova
ITMO University – St. Petersburg, RU
- Kenneth A. De Jong
George Mason University – Fairfax, US
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Carola Doerr
UPMC – Paris, FR
- Anton V. Eremeev
Sobolev Inst. of Mathematics – Omsk, RU
- Carlos M. Fonseca
University of Coimbra, PT
- Tobias Friedrich
Hasso-Plattner-Institut – Potsdam, DE
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Nikolaus Hansen
INRIA Saclay – Île-de-France – Orsay, FR
- Hsien-Kuei Hwang
Academica Sinica – Taipei, TW
- Christian Igel
University of Copenhagen, DK
- Thomas Jansen
Aberystwyth University, GB
- Daniel Johannsen
SAP Innovation Center – Potsdam, DE
- Joshua D. Knowles
University of Manchester, GB
- Timo Kötzing
Hasso-Plattner-Institut – Potsdam, DE
- Oswin Krause
University of Copenhagen, DK
- Marvin Künnemann
MPI für Informatik – Saarbrücken, DE
- William B. Langdon
University College London, GB
- Per Kristian Lehre
University of Nottingham, GB
- Luigi Malago
Shinshu University – Nagano & INRIA Saclay
- Silja Meyer-Nieberg
Universität der Bundeswehr – München, DE
- Alberto Moraglio
University of Exeter, GB
- Frank Neumann
University of Adelaide, AU
- Pietro S. Oliveto
University of Sheffield, GB
- Yann Ollivier
University Paris Sud, FR
- Jonathan E. Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Peter F. Stadler
Universität Leipzig, DE
- Dirk Sudholt
University of Sheffield, GB
- Lothar Thiele
ETH Zürich, CH
- Heike Trautmann
Universität Münster, DE
- Rolf Wanka
Univ. Erlangen-Nürnberg, DE
- Carsten Witt
Technical University of Denmark – Lyngby, DK
- Xin Yao
University of Birmingham, GB
- Christine Zarges
University of Birmingham, GB