Report from Dagstuhl Seminar 17191

Theory of Randomized Optimization Heuristics

Edited by

Carola Doerr[1], Christian Igel[2], Lothar Thiele[3], and Xin Yao[4]

1 **CNRS & UPMC, Paris, FR,** carola.doerr@mpi-inf.mpg.de
2 **University of Copenhagen, DK,** igel@diku.dk
3 **ETH Zürich, CH,** thiele@ethz.ch
4 **University of Birmingham, GB,** x.yao@cs.bham.ac.uk

Abstract

This report summarizes the talks, breakout sessions, and discussions at the Dagstuhl Seminar 17191 on *Theory of Randomized Optimization Heuristics*, held during the week from May 08 until May 12, 2017, in Schloss Dagstuhl – Leibniz Center for Informatics. The meeting is the successor of the "Theory of Evolutionary Algorithm" seminar series, where the change in the title reflects the development of the research field toward a broader range of heuristics. The seminar has hosted 40 researchers from 15 countries. Topics that have been intensively discussed at the seminar include population-based heuristics, constrained optimization, non-static parameter choices as well as connections to research in machine learning.

Seminar May 7–12, 2017 – http://www.dagstuhl.de/17191
1998 ACM Subject Classification F.2
Keywords and phrases algorithms and complexity, evolutionary algorithms, machine learning, optimization, soft computing
Digital Object Identifier 10.4230/DagRep.7.5.22
Edited in cooperation with Martin S. Krejca

1 Executive Summary

Carola Doerr
Christian Igel
Lothar Thiele
Xin Yao

Randomized search and optimization heuristics such as evolutionary algorithms, ant colony optimization, particle swarm optimization, and simulated annealing, have become established problem solvers. They have successfully been applied to a wide range of real-world applications, and they are applicable to problems that are non-continuous, multi-modal, and/or noisy as well as to multi-objective and dynamic optimization tasks. Theory of randomized optimization heuristics aims at providing mathematically founded insights into the working principles of these general-purpose problem solvers, and at developing new and more powerful heuristic optimization methods in a principled way. The seminar has covered several important streams in this research discipline. Among several other topics, extended discussions have been held on the advantages of population-based heuristics and of non-static parameter choices, optimization problems with constrains, as well as existing and possible connections to research in machine learning.

The seminar continues to be one of the key stimulator for novel ideas, tools, and approaches in the theory of randomized optimization heuristics. Accordingly, the acceptance rate for the invitations has been staying at a very high level.

Topics

The research in theory of randomized optimization heuristics is as broad as the applicability of these methods. The seminar succeeded in covering the various theoretical approaches. There was a focus on important cross-cutting topics, which we briefly outline in the following.

One of the most prominent research areas in the theory of randomized optimization heuristics deals with *runtime* and *convergence analysis*, aiming at proving bounds on the speed of the convergence to an optimal solution. Typical questions concern the advantages of certain algorithmic choices, such as

- the size of the memory (*population*),
- the usage of different sampling strategies (*variation* of previously sampled search points, in particular via *mutation* of one previously evaluated solution candidate and *recombination* of two or more previous search points), and
- the selection strategies (e.g., *elitist selection* which never discards a best-so-far solution vs. the non-elitist Boltzmann strategies found in Simulated Annealing, SSWM, and the Metropolis algorithm).

One of the most relevant objectives in empirical and theoretical works on randomized optimization heuristics is to determine the best parameter settings for the above-described algorithmic components. Given the complex interactions between the parameter values, this *parameter tuning* task is a highly challenging one. It is further complicated by the observation that for most problems the optimal parameter settings change during the optimization process, thus asking for *parameter control* mechanisms that adjust the parameter value to the current state of the optimization. Identifying such reasonable (and possibly provably optimal) ways to update the parameter choices has been one of the intensively discussed topics of the seminar. Significant progress towards a better understanding of different parameter update schemes has been obtained in the last few years, as has been demonstrated by several talks, for example on self-adaptive and self-adjusting parameter updates as well as on estimation of distribution algorithms. Among other results, several connections to related questions in machine learning have been made, motivating the organizers to include machine learning as a focus topic of this seminar.

Randomized search heuristics are currently very popular in general machine learning[1] in the form of *Bayesian optimization*. However, there has been little connection between the research in Bayesian optimization and the established work on randomized search heuristics, and the seminar was a step to change this. The first talk of the seminar was an extended introduction to Bayesian optimization by Matthew W. Hoffman from Google DeepMind, a leading expert in the field. The talk set the stage for informed discussions on similarities and differences between methods—and potential synergies between the research fields. Thompson sampling, an important algorithm in Bayesian optimization, was revisited in the talk by Jonathan Shapiro on dueling bandit problems, which demonstrated randomized search heuristics in a scenario of high commercial relevance. A common application of randomized

[1] One may well argue that randomized search heuristics actually belong to the broader field of machine learning methods.

search heuristics in general machine learning is model selection, for example finding a tailored structure of a neural network. This was addressed in the talk by Olivier Teytaud from Google Brain, who discussed model selection heuristics for large-scale machine learning systems. Randomized search heuristics are also successfully used for reinforcement learning. Arina Buzdalova presented work in which the connection is the other way round: ideas from reinforcement learning are used to improve randomized optimization (by controlling the choice of objectives).

Another intensively discussed topic, highly relevant in both discrete and continuous optimization, was constrained optimization. Here the main research questions concern the different ways to model constrained problems in black-box settings, and suitable algorithmic approaches. In addition to a number of theoretical results on constrained optimization, the need for a well-designed benchmark suite has also been discussed. As a result of one of the breakout sessions of the previous Dagstuhl Seminar 15211 on *Theory of Evolutionary Computation*, Dimo Brockhoff presented the recent extension of the COCO benchmark set (http://coco.gforge.inria.fr/doku.php) to constrained optimization. Dirk Arnold presented some work indicating that this extension of COCO is very timely, and much needed in the randomized search heuristics community. Furthermore, another breakout session has been held this year on the topic of constrained optimization, organized by Frank Neumann, with a focus on the different ways to model soft and hard constraints in discrete black-box optimization.

Organization

The seminar schedule has offered a good flexibility for the participants to propose talks and discussions of different lengths. 29 talks of 10–30 minutes each have been held in total, in the plenary room. These plenary talks were complemented by a introductory tutorial on Bayesian Optimization by Matt Hoffman on Monday morning and by 7 breakout sessions on various topics, including methodology-oriented discussions on the applicability of drift analysis in continuous domains or how to interpret the CMA-ES in the framework of information geometry optimization as well as problem-driven brainstorming on constrained optimization, the role of diversity in heuristic search, preference-based selection, and the method of estimation of distribution algorithms. Another breakout session was devoted to discussing the importance and possible obstacles of bringing theory-and practice-driven research in heuristic optimization closer together. The breakout sessions have been held on Tuesday, Wednesday, and Thursday afternoon, respectively, and have all witnessed high attendance rates. All talks and breakout sessions are summarized in Sections 3 and 4 of the present report.

We would like to express our gratitude to the Dagstuhl staff and all participants for making this Dagstuhl Seminar 17191 on *Theory of Randomized Optimization Heuristics* such a successful event, which has been a pleasure to organize.

Carola Doerr (CNRS and Pierre et Marie Curie University Paris 6, FR)
Christian Igel (University of Copenhagen, DK)
Lothar Thiele (ETH Zürich, CH)
Xin Yao (University of Birmingham, GB and SUSTech Shenzhen, CH)

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Optimal Step-Size for the Weighted Recombination Evolution Strategy

Youhei Akimoto (Shinshu University – Nagano, JP)

Joint work of Youhei Akimoto, Anne Auger, Nikolaus Hansen

We focus on the optimality of the step-size for the weighted recombination evolution strategy. In the previous work by D. Arnold, the optimal step-size and the optimal weights are derived on the Sphere function in the limit of the dimension to infinity. In this talk, we address the optimal step-size for a finite dimension, especially when the population size is close to or greater than the dimension. We derive the optimal step-size under the limit of the learning rate for the mean vector to infinity. We show that the derived optimal step-size provides a good approximation on a finite dimensional Sphere function for the standard setup of the learning rate of the mean vector update, i.e., the learning rate equal one.

3.2 Evolutionary Computation with Constraints

Dirk V. Arnold (Dalhousie University – Halifax, CA)

I discuss several aspects of constrained black-box optimization, including a taxonomy of constraints in simulation based optimization recently proposed by Le Digabel and Wild, work on attempting to arrive at an analytically based understanding of the behaviour of evolution strategies on simple constrained problems, and a comparison of algorithms on a set of test problems.

3.3 Connecting Stability of Markov Chains and Deterministic Control Models for Analyzing Randomized Algorithms

Anne Auger (INRIA RandOpf Team, FR)

Joint work of Alexandre Chotard, Anne Auger
Main reference A. Chotard, A. Auger, "Verifiable Conditions for the Irreducibility and Aperiodicity of Markov Chains by Analyzing Underlying Deterministic Models", arXiv:1508.01644v4 [math.PR], 2017.
URL https://arxiv.org/abs/1508.01644v4

Motivated by the analysis of randomized search optimization algorithms, this talk presents connections between Markov chain theory and stability of underlying deterministic control models. We consider a general model of Markov chain $\Phi_{t+1} = F(\Phi_t, \alpha(\Phi_t, U_{t+1}))$ where $\{U_t, t > 0\}$ are i.i.d. random vectors, F is typically C^1 and $\alpha(x, U_1)$ admits a lower semi continuous density. This model embeds Markov chains that arise in stochastic optimization where α models the selection which is discontinuous but where we can derive a lower semicontinuous density.

We show that there is equivalence between the existence of a globally attracting (GA) state for the deterministic control model and the φ-irreducibility of the Markov chain provided a controllability condition holds. We then show that the support of the irreducibility measure is the set of GA states. Last we show a practical condition for proving φ-irreducibility that consists in showing the existence of a GA state where the controllability condition is satisfied.

3.4 Towards a Theory of CMA-ES: But First, Simplify Your CMA-ES!

Hans-Georg Beyer (Fachhochschule Vorarlberg – Dornbirn, AT)

Joint work of Hans-Georg Beyer, Bernhard Sendhoff
Main reference H.-G. Beyer, B. Sendhoff, "Simplify Your Covariance Matrix Adaptation Evolution Strategy", IEEE Transactions on Evolutionary Computation, Vol. 21(5), pp. 746–759, IEEE, 2017.
URL http://dx.doi.org/10.1109/TEVC.2017.2680320

Before starting the endeavor of a theoretical convergence analysis of the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) it seems advisable to simplify this algorithm in such a manner that on the one hand it gets more amenable to such an analysis, but on the other hand without sacrificing the good optimization performance of the original CMA-ES.

In this talk it will be shown that one can remove one of the evolution path calculations from the CMA-ES keeping only the one used for the mutation strength control. In a second step it will be shown that one can also get rid of the covariance matrix update, thus removing the "C" from the CMA-ES resulting in the novel MA-ES that performs nearly equally well as the original strategy.

Besides the increased simplicity of the novel MA-ES, it also has a reduced algorithmic complexity of $O(N^2)$ compared to the original $O(N^3)$. Furthermore, the new M-matrix update rule derived has a special structure that allows for a direct interpretation of the M-matrix update. This update is driven by the departure of the actually selected (isotropically generated) z-vectors from isotropy. That is, the M-matrix changes its "shape" until the composition of the objective function $f(y + \sigma \cdot M \cdot z) =: g(z)$ has transformed the original problem $f(x)$ into a sphere function $g(z)$.

3.5 Progress Report: Towards a Constrained Test Suite for COCO

Dimo Brockhoff (INRIA Saclay – Île-de-France, FR)

Joint work of Anne Auger, Nikolaus Hansen, Asma Atamna, Olaf Mersmann, Tea Tusar, Dejan Tusar, Phillipe Sampaio

In the previous edition of this workshop series on Theory of Randomized Optimization Heuristics, a breakout session on "Constrained Blackbox Optimization Benchmarking" was held to discuss the first steps towards a constrained test suite within the well-known Comparing Continuous Optimizers platform (COCO, github.com/numbbo/coco/) and to identify (theoretical) questions related to this extension of the platform. In this talk, I quickly reminded us on how we benchmark optimization algorithms in COCO on unconstrained problems and then reported on the progress we made since the last breakout session towards a new constrained test suite in COCO.

3.6 How to Exploit Your Fitness-Distance Correlation: Runtime Analysis of the $(1 + (\lambda, \lambda))$ GA on Random Satisfiable 3-CNF Formulas

Maxim Buzdalov (ITMO University – St. Petersburg, RU) and Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

Joint work of Maxim Buzdalov, Benjamin Doerr
Main reference M. Buzdalov, B. Doerr, "Runtime Analysis of the $(1 + (\lambda, \lambda))$ Genetic Algorithm on Random Satisfiable 3-CNF Formulas", Proceedings of Genetic and Evolutionary Computation Conference, 2017.
An extended version is available at arxiv: https://arxiv.org/abs/1704.04366
URL http://dx.doi.org/10.1145/3071178.3071297

The $(1+(\lambda, \lambda))$ genetic algorithm, first proposed at GECCO 2013, showed a surprisingly good performance on some optimization problems. The theoretical analysis so far was restricted to the OneMax test function, where this GA profited from the perfect fitness-distance correlation. In this work, we conduct a rigorous runtime analysis of this GA on random 3-SAT instances in the planted solution model having at least logarithmic average degree, which are known to have a weaker fitness distance correlation.

We prove that this GA with fixed not too large population size again obtains runtimes better than $\Theta(n \log n)$, which is a lower bound for most evolutionary algorithms on pseudo-Boolean problems with unique optimum. However, the self-adjusting version of the GA risks reaching population sizes at which the intermediate selection of the GA, due to the weaker fitness-distance correlation, is not able to distinguish a profitable offspring from others. We show that this problem can be overcome by equipping the self-adjusting GA with an upper limit for the population size. Apart from sparse instances, this limit can be chosen in a way that the asymptotic performance does not worsen compared to the idealistic OneMax case. Overall, this work shows that the $(1 + (\lambda, \lambda))$ GA can provably have a good performance on combinatorial search and optimization problems also in the presence of a weaker fitness-distance correlation.

3.7 Is it necessary to perform multi-objective optimization when doing multiobjectivization?

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

Joint work of Irina Petrova, Maxim Buzdalov, Arina Buzdalova
Main reference A. Buzdalova, I. Petrova and M. Buzdalov, "Runtime analysis of different Approaches to select conflicting auxiliary objectives in the generalized OneMax problem," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1–7
URL https://doi.org/10.1109/SSCI.2016.7850140

It has been shown that single-objective optimization may be improved by introducing auxiliary objectives. In practice, the auxiliary objectives may be conflicting. This talk presents theoretical analysis of different approaches of using auxiliary objectives on the Generalized OneMax problem with conflicting auxiliary objectives OneMax and ZeroMax.

In most of the considered methods, the optimized objectives are selected dynamically. Particularly, the $O(n \log n)$ runtime is proven for a multi-objective algorithm that optimizes the target objective together with a dynamically selected auxiliary objective. At the same time, it is shown that asymptotically the same runtime holds for a single-objective algorithm with the preservation of the best found solution, where objectives are dynamically selected using reinforcement learning.

Acknowledgements
This work was supported by RFBR according to the research project No. 16-31-00380 mol_a.

3.8 On the Variable Interaction Graph in Gray-Box Optimization

Francisco Chicano (University of Málaga, ES)

Joint work of Darrell Whitley, Renato Tinós, Gabriela Ochoa, Andrew M. Sutton, Francisco Chicano
Main reference L. Darrell Whitley, Francisco Chicano, Brian W. Goldman: Gray Box Optimization for Mk Landscapes (NK Landscapes and MAX-kSAT). Evolutionary Computation 24(3): 491-519 (2016)
URL https://doi.org/10.1162/EVCO_a_00184

Given a pseudo-Boolean function, the Variable Interaction Graph is defined using the set of variables as node set and joining two variables with an edge when there is a nonzero Walsh coefficient in the Walsh expansion of the function whose index contains both variables. In the case of k-bounded pseudo-Boolean functions (like NK Landscapes or MAX-kSAT) the co-occurrence graph can be used as an approximation of the Variable Interaction Graph. Two operators and one search algorithm based on the Variable Interaction Graph are described in the talk. We also present some open questions regarding theoretical results related to the performance of the algorithms and the optimal parameter settings based on the Variable Interaction Graph.

3.9 Fast Genetic Algorithms

Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)

For genetic algorithms using a bit-string representation of length n, the general recommendation is to take $1/n$ as mutation rate. In this work, we discuss whether this is justified for multimodal functions. Taking jump functions and the $(1+1)$ evolutionary algorithm as the simplest example, we observe that larger mutation rates give significantly better runtimes. For the $\text{Jump}_{m,n}$ function, any mutation rate between $2/n$ and m/n leads to a speed-up at least exponential in m compared to the standard choice.

The asymptotically best runtime, obtained from using the mutation rate m/n and leading to a speed-up super-exponential in m, is very sensitive to small changes of the mutation rate. Any deviation by a small $(1 \pm \varepsilon)$ factor leads to a slow-down exponential in m. Consequently, any fixed mutation rate gives strongly sub-optimal results for most jump functions.

Building on this observation, we propose to use a random mutation rate α/n, where α is chosen from a power-law distribution. We prove that the $(1+1)$ EA with this heavy-tailed mutation rate optimizes any $\text{Jump}_{m,n}$ function in a time that is only a small polynomial (in m) factor above the one stemming from the optimal rate for this m.

Our heavy-tailed mutation operator yields similar speed-ups (over the best known performance guarantees) for the vertex cover problem in bipartite graphs and the matching problem in general graphs.

Following the example of fast simulated annealing, fast evolution strategies, and fast evolutionary programming, we propose to call genetic algorithms using a heavy-tailed mutation operator *fast genetic algorithms*.

3.10 Optimal Recombination for the Asymmetric TSP: Theory and Experiment

Anton V. Eremeev (Sobolev Institute of Mathematics – Novosibirsk)

We consider two approaches to formulation and solution of optimal recombination problems arising as supplementary problems in genetic algorithms for the Asymmetric Traveling Salesman Problem. The first approach uses a representation of solutions where a genotype is the sequence of vertices visited in the traveling salesman tour (position-based encoding). The second approach uses a representation where each gene defines an arc of the tour (adjacency-based representation). Both optimal recombination problems under consideration are NP-hard but relatively fast worst-case exponential-time algorithms are presented for solving them [2]. Besides that, in the case of position-based encoding, the optimal recombination problem is shown to be solvable in linear time for "almost all" pairs of parent solutions.

As a proof of concept we develop a genetic algorithm with a crossover operator which solves an optimal recombination problem. The algorithm also incorporates problem-specific mutation operator, local search and initialization method. A computational experiment on TSPLIB instances shows that the proposed genetic algorithm yields competitive results to other state-of-the-art genetic algorithms [1].

The research is supported by Russian Science Foundation grant 15-11-10009.

References

1 Anton V. Eremeev and Yulia V. Kovalenko.*Genetic algorithm with optimal recombination for the asymmetric travelling salesman problem*. To appear in Pros. of LSSC-17, Sozopol, Bulgaria, 2017
2 Anton V. Eremeev and Yulia V. Kovalenko.*Experimental evaluation of two approaches to optimal recombination for permutation problems*. In Proc. of EvoCOP-16, Porto, Portugal, Springer, Switzerland, LNCS vol. 9595. pp. 138–153 (2016)

3.11 Mathematical Models of Artificial Genetic Representations with Neutrality

Carlos M. Fonseca (University of Coimbra, PT)

Joint work of Carlos M. Fonseca, Vida Vukašinović, Nino Bašić

In this work, a mathematical framework for the study and characterisation of a family of uniformly redundant binary representations based on error-control codes proposed previously [1] is developed. Such representations can exhibit various degrees of redundancy, neutrality, and other properties believed to influence the performance of evolutionary algorithms, such as connectivity, locality, and synonymity [2], and have allowed this influence to be studied experimentally to some extent [3]. The definition of suitable equivalence classes leads to a partitioning of the representation space with respect to neutral network structure and connectivity, which should allow the effect of locality on search performance to be studied while other properties are kept fixed. The practical implications of the proposed framework are also discussed.

Acknowledgements. This talk is based upon work from COST Action CA15140 on Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO), supported by COST (European Cooperation in Science and Technology). Partial support by national funds through the Portuguese Foundation for Science and Technology (FCT) and by the European Regional Development Fund (FEDER) through COMPETE 2020 – Operational Programme for Competitiveness and Internationalisation (POCI) is also acknowledged.

References

1 C. M. Fonseca and M. B. Correia. "Developing redundant binary representations for genetic search," in *The 2005 IEEE Congress on Evolutionary Computation*, vol. 2, (Edinburgh, U.K.), pp. 1675–1682, Sept. 2005.
2 F. Rothlauf. *Representations for Genetic and Evolutionary Algorithms*. Springer, 2nd Edition, 2006.
3 M. B. Correia. "A study of redundancy and neutrality in evolutionary optimization," *Evolutionary Computation*, vol. 21, pp. 413–443, 2013.

3.12 Monotone Functions on Bitstrings – Some Structural Notes

Christian Gießen (Technical University of Denmark – Lyngby, DK)

Joint work of Tobias Friedrich, Timo Kötzing, Martin Schirneck, Christian Gießen

Analyzing the (1+1) EA with mutation probability c/n on monotone functions for a constant c is a challenging problem. Only for $c \leq 1$ and $c > 2.2$ upper bounds on the expected runtime have been established, but the whole truth is unkown. We present unfinished work that led to two curious and surprising structural observations. First, the linear function BinVal is a special monotone function: globally, it is an optimal adversarial function in the sense that it maximizes the possibility to increase the distance to the optimum while increasing the fitness at the same time. This result can be shown using techniques that stem from extremal set theory and general order theory and is related to the Shadow Minimization Problem. However, this notion of hardness is not sufficient to describe hardness for the (1+1) EA. The second surprising observation is the distribution of monotone functions that are structurally equivalent for small n. OneMax-like functions occur most often, while BinVal-like functions occur the least. It is however unclear if this behaviour holds for all n. It remains an open question how to structurally define hardness for monotone functions.

References

1 The on-line encyclopedia of integer sequences. published electronically at https://oeis.org. https://oeis.org/A000372.

2 The on-line encyclopedia of integer sequences. published electronically at https://oeis.org. https://oeis.org/A046873.
3 Graham R. Brightwell and Prasad Tetali. The number of linear extensions of the boolean lattice. *Order*, 20(4):333–345, 2003.
4 Sylvain Colin, Benjamin Doerr, and Gaspard Férey. Monotonic functions in EC: anything but monotone! In *Genetic and Evolutionary Computation Conference, GECCO '14, Vancouver, BC, Canada, July 12-16, 2014*, pages 753–760, 2014.
5 Benjamin Doerr, Thomas Jansen, Dirk Sudholt, Carola Winzen, and Christine Zarges. Mutation rate matters even when optimizing monotonic functions. *Evolutionary Computation*, 21(1):1–27, 2013.
6 K. Engel. *Sperner Theory.* Cambridge Solid State Science Series. Cambridge University Press, 1997.
7 S.R. Finch. *Mathematical Constants.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2003.
8 Graham. The number of linear extensions of ranked posets. Technical report, London School of Economics, 2003.
9 Thomas Jansen. On the brittleness of evolutionary algorithms. In *FOGA*, pages 54–69, 2007.
10 Jichang Sha and Daniel J. Kleitman. The number of linear extensions of subset ordering. *Discrete Mathematics*, 63(2-3):271–278, 1987.

3.13 Global Convergence of the (1+1)-ES

Tobias Glasmachers (Ruhr-Universität Bochum, DE)

We prove several global convergence theorems for the (1+1)-ES algorithm. We refrain from inserting nursing mechanisms into the algorithms, and at no point we resort to asymptotic or otherwise approximate analysis. Therefore the theorems and their proofs reflect the actual behavior of the algorithm.

The analysis is based on two ingredients. We start with a generic sufficient decrease condition for elitist rank-based evolution strategies, formulated for an essentially monotonically transformed variant of the objective function. Then we show that the algorithm state is found infinitely often in a regime where step size and success rate are simultaneously bounded away from zero, with full probability.

The main result is proven by combining these statements. More powerful variants are derived based on additional regularity conditions. The statements ensure under minimal technical preconditions that the sequence of iterates has a limit point in a critical point of some sort.

Based on our theorems we analyze the behavior of the (1+1)-ES on a number of problems ranging from the smooth (non-convex) cases over various types of ridge functions to several discontinuous problems.

3.14 How to Guaranty Positive Definiteness in Active CMA-ES

Nikolaus Hansen (INRIA Saclay – Île-de-France, FR)

Main reference Nikolaus Hansen, The CMA Evolution Strategy: A Tutorial. CoRR abs/1604.00772 (2016)
URL http://arxiv.org/abs/1604.00772

In active CMA-ES, a positive definite covariance matrix undergoes an additive update using also negative weights. Depending on the chosen weights and learning rate, the covariance matrix may become negative definite with a small probability. We can prove a condition on the weights which guaranties positive definiteness with probability one, if also the length of the random vectors is upper bounded. The condition limits the decrement of the smallest eigenvalue to a factor of about one over dimension in each iteration. We investigate the population size when the constraint becomes effective as a function of dimension.

3.15 An overview of Bayesian Optimization

Matthew W. Hoffman (Google DeepMind – London, GB)

Joint work of Bobak Shariari, Matthew W. Hoffman
Main reference Matthew W. Hoffman, Zoubin Ghahramani. "Output-Space Predictive Entropy Search for Flexible Global Optimization." NIPS Workshop on Bayesian Optimization, 2015.

In this talk I give a high-level overview of Bayesian optimization methods for global optimization. While typically used for continuous, black-box optimization I also briefly touch on relations to the broader optimization community.

Further, while the design of methods for Bayesian optimization involves a great number of choices that are often implicit in the overall algorithm design, in this work I argue for a modular approach to Bayesian optimization. In particular this includes selection of the acquisition function, kernel, and hyper-priors as well as less-discussed components such as the recommendation and initialization approaches. In this work I also argue for an information-theoretic approach to the design of acquisition strategies. Finally, in this work I present a Python implementation, pybo, that allows us to easily vary these choices. Ultimately this approach provides us a straightforward mechanism to examine the effect of each choice both individually and in combination.

3.16 Estimation of Distribution Algorithms

Martin S. Krejca (Hasso-Plattner-Institut – Potsdam, DE)

Joint work of Tobias Friedrich, Timo Kötzing, Andrew M. Sutton, Martin S. Krejca

Evolutionary algorithms (EAs) are optimization techniques inspired by nature. They are a popular choice if the problem at hand is, for example, noisy or highly complex and cannot be well formalized but the quality of a single solution can be easily measured. Typically, EAs maintain a set of samples from the solution space, which is iteratively updated, keeping better solutions and discarding bad ones. An alternative and more direct approach that is also commonly used is to not store samples but a probability distribution over the search space that generates these samples. Such algorithms are called estimation of distribution algorithms (EDAs).

In practice, EDAs are widely applied and perform very well. However, theoretical results on EDAs explaining this success are very scarce so far. We introduce an EDA framework we proposed, which subsumes many EDAs used for discrete domains, and we present our theoretical results for this framework. This includes robustness of EDAs to noise, restrictions on the way an EDA can update its distribution, and unbiasedness.

3.17 Convergence in Genetic Programming

William B. Langdon (University College London, GB)

Main reference William B. Langdon, "Long-Term Evolution of Genetic Programming Populations," GECCO 2017: The Genetic and Evolutionary Computation Conference", 2017.
URL https://arxiv.org/abs/1703.08481

In the future we hope to use evolution to solve challangeing problems. This may require it to be run for many generations. Therefore we investigate what happens in long runs of a current evolutionary algorithms. Specifically we look at an easy problem 6-mux when solved by genetic programming when the GP system is run on long past the point where GP solves the six multiplexor problem.

We evolve binary mux-6 trees for up to 100000 generations evolving some programs with more than a hundred million nodes. Initially tree growth is O(generations squared), [GP+EM (1)1 pp95-119] but existing theory could be made more formal. Long after the first time when everyone in the finit population solves the problem, and so has identical fitness, the tree size *appears* to execute a random walk, albeit with a lower bound. However, even in this region, the distribution of tree sizes within the population is not a gammar distribution as predicted for large populations and no selection (http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/poli_2007_eurogp.html). Our unbounded Long-Term Evolution Experiment LTEE GP appears not to evolve building blocks but does suggests a limit to bloat. We do see periods of tens even hundreds of generations where the population is 100 percent functionally converged.

3.18 Landscape of the Triangle Program

William B. Langdon (University College London, GB)

Joint work of William B. Langdon, Nadarajen Veerapen, Gabriela Ochoa
Main reference William B. Langdon, Nadarajen Veerapen, Gabriela Ochoa, "Visualising the Search Landscape of the Triangle Program," Genetic Programming – 20th European Conference, 96–113, 2017.
URL http://dx.doi.org/doi:10.1007/978-3-319-55696-3_7

the triangle program is a small software engineering benchmark recently analysed in terms of global search, genetic algorithms schema, iterated local search and local optima networks. Results presented at EuroGP-2017 (http://dx.doi.org/doi:10.1007/978-3-319-55696-3_7).

I build on these results to support the thesis that real software is not as fragile as is often assumed and then consider in detail the simplest of the test cases – the scalene triangle, looking at the variable interaction graph for the test case when only comparison operators are to be mutated. I propose the variable interaction graph in real software may lead to theoretical insights to the improvement of sizable programs using evolutionary improvement methods such as genetic programming.

The triangle benchmark is available via http://www.cs.ucl.ac.uk/staff/w.langdon/gggp/#triangle_dataset.

3.19 Runtime Analysis Evolutionary Algorithms with Self-adaptive Mutation Rates

Per Kristian Lehre (University of Birmingham, GB)

Joint work of Duc-Cuong Dang, Per Kristian Lehre
Main reference Duc-Cuong Dang, Per Kristian Lehre, "Self-adaptation of Mutation Rates in Non-elitist Populations," Parallel Problem Solving from Nature – PPSN XIV, 803–813, 2016.
URL http://www.cs.nott.ac.uk/~pszpl/selfadapt/

The runtime of evolutionary algorithms (EAs) depends critically on their parameter settings, which are often problem-specific. Automated schemes for parameter tuning have been developed to alleviate the high costs of manual parameter tuning.

Experimental results indicate that self-adaptation, where parameter settings are encoded in the genomes of individuals, can be effective in continuous optimisation. However, results in discrete optimisation have been less conclusive. Furthermore, a rigorous runtime analysis that explains how self-adaptation can lead to asymptotic speedups has been missing.

This talk presents the first runtime analysis of self-adaptation for discrete, population-based EAs. We apply the level-based theorem to show how a self-adaptive EA is capable of fine-tuning its mutation rate, leading to exponential speedups over EAs using fixed mutation rates.

For a simulation and a link to the paper, please see http://www.cs.nott.ac.uk/~pszpl/selfadapt/.

3.20 Noise models for comparison-based evolutionary algorithms

Johannes Lengler (ETH Zürich, CH)

Noise models for evolutionary algorithms typically assume that the fitness function evaluation is distorted. However, for comparison-based evolutionary algorithms this makes only limited sense, since often such algorithm do not actually evaluate a fitness function. I have discussed three examples where fitness functions are not evaluated, and their implications for noise models: swap-based sorting, Schöning's algorithms, and the evolution of game engines.

References

1 Tomáš Gavenčiak, Barbara Geissmann, and Johannes Lengler. *Sorting by Swaps with Noisy Comparisons*. Proceedings of the 2017 conference on Genetic and Evolutionary Computation (GECCO '17)

3.21 Features, Diversity, Random Walks and Digital Art

Frank Neumann (University of Adelaide, AU)

Joint work of Frank Neumann, Wanru Gao, Samadhi Nallaperuma, Aneta Neumann, Bradley Alexander, James Kortman, Zygmunt Szpak, Wojciech Chojnacki

We consider diversity with respect to features of a given problem and introduce an evolutionary algorithm that maximizes the feature diversity under a constraint given by a lower bound on the desired objective function value (in the case of maximization) [1]. We show how this increases feature diversity for the problem of evolving hard TSP instances (in terms of approximation ratio) for the classical 2-opt local search algorithm. Furthermore, we discuss how this feature-based diversity approach can be used to create variation of images (based on a given image) with respect to various features [2]. Afterwards, we introduce evolutionary approaches for image transition and composition based on random walks [3]. The evolutionary image transition approach allows to create artistic sequences of images in form of a video by transforming a given starting image into a target image. Our evolutionary image composition approach combines two artistic images taken into account interesting parts of the given images [4]. The key element is a fitness function based on covariance matrix image descriptors taking into account a set of features.

References

1 W. Gao, T. Friedrich, F. Neumann: Fixed-parameter single objective search heuristics for minimum vertex cover. In: Parallel Problem Solving from Nature XIII, PPSN 2016, Springer, 2016.

2 B. Alexander, J. Kortman, A. Neumann: Evolution of artistic image variants through feature based diversity optimisation. In: Genetic and Evolutionary Computation Conference, GECCO 2017, ACM Press, 2017.

3 A. Neumann, B. Alexander, F. Neumann: Evolutionary image transition using random walks. In: International Conference on Computational Intelligence in Music, Sound, Art and Design, EVOMUSART 2017, Springer, 2017.

4 A. Neumann, Z. L. Szpak, W. Chojnacki, F. Neumann: Evolutionary image composition using feature covariance matrices. In: Genetic and Evolutionary Computation Conference, GECCO 2017, ACM Press, 2017.

3.22 Standard Steady State Genetic Algorithms can Hillclimb Faster than Mutation-only Evolutionary Algorithms

Pietro S. Oliveto (University of Sheffield, GB)

Joint work of Dogan Corus, Pietro S. Oliveto

Explaining to what extent the real power of genetic algorithms lies in the ability of crossover to recombine individuals into higher quality solutions is an important problem in evolutionary computation. Recently it has been shown how the interplay of mutation and crossover may create the necessary diversity to efficiently escape local optima. In this talk we discuss how such an interplay can also make genetic algorithms hillclimb faster than their mutation-only counterparts. We devise a Markov Chain framework that allows to rigorously prove an upper bound on the runtime of standard steady state genetic algorithms to hillclimb the ONEMAX function. The bound establishes that the steady-state genetic algorithms are 25% faster than all unbiased mutation-only evolutionary algorithms with static mutation rate up to lower order terms for moderate population sizes. The analysis also suggests that larger populations may be faster than populations of size 2. We present a lower bound for a greedy (2+1) GA that matches the upper bound for populations larger than 2, rigorously proving that 2 individuals cannot outperform larger population sizes under greedy selection and greedy crossover up to lower order terms. In complementary experiments the best population size is greater than 2 and the greedy genetic algorithms are faster than standard ones, further suggesting that the derived lower bound also holds for the standard steady state (2+1) GA.

3.23 Linear multiobjective drift analysis

Jonathan E. Rowe (University of Birmingham, GB)

Drift analysis is now a standard tool for analysing the run-time of stochastic optimisation algorithms. The expected progress in the “distance” to the target state is used to derive a first hitting time of that state. We consider the situation where the process is best described by more than distance function, and derive a generalisation of multiplicative drift to this situation. Example applications include: an evolutionary algorithm solving a multi-objective optimisation problem; a parallel island model with probabilistic migration.

3.24 Theoretical Aspects of the Averaged Hausdorff Indicator in Biobjective Optimization

Günter Rudolph (TU Dortmund, DE)

Joint work of Oliver Schütze, Heike Trautmann, Günter Rudolph
Main reference Günter Rudolph, Oliver Schütze, Christian Grimme, Christian Domínguez-Medina, Heike Trautmann, "Optimal averaged Hausdorff archives for bi-objective problems: theoretical and numerical results." Comp. Opt. and Appl., 64(2): 589-618, 2016.

The averaged Hausdorff indicator is an alternative to the dominated hypervolume indicator for assessing the quality of Pareto front approximations if equispaced solutions on the Pareto front are desired. This may happen frequently in dynamic control applications. Recently there have been several proposals how to integrate this indicator in evolutionary search for up to 4 objectives. Empirical evaluations have shown that this approach is promising. Therefore a theoretical analysis of the indicator is desireable. We show results in case of two objectives. Point sets that have minimal averaged Hausdorff distance to the Pareto front are called optimal archives. If the Pareto front is concave, then the optimal archives are on the Pareto front. If the Pareto front is linear, the optimal archive consists of equispaced solutions on the Pareto front. If the Pareto front is circularly concave, then the optimal archive consists of equispaced points. If the Pareto front is circularly concave, then the optimal archive consists of equispaced points but no element is on the Pareto front. But it can be shown that the averaged Hausdorff distance of the archive to the Pareto front decreases to zero with order $1/m^2$ for increasing archive size m. For practical purposes with archive sizes $m \geq 100$ the accuracy, i.e. the closeness to the Pareto front, is sufficient.

3.25 Max-Min Thompson Sampling for the K-Armed Dueling Bandit Problem

Jonathan L. Shapiro (University of Manchester, GB) and Joseph Mellor

Joint work of Joseph Mellor, Jonathan Shapiro

The K-armed dueling bandit problem is a variation on the classic K-armed bandit problem. The distinguishing feature of the problem is that only relative preference between pairs of arms is given as feedback. This paper proposes a new algorithm, Max-Min Thompson Sampling, to solve the problem. The algorithm uses a method derived from game theory to choose appropriate pairs of arms, and Thompson Sampling to learn the preferences from observations. We derive an $O(K \log T)$ problem-dependent and finite-time regret bound for the strategy, where T is the time. Our bound is as low as others in the literature. We provide empirical results of the method on a variety of simulations including the Komiyama et.al. (2015) benchmarks, and a real-world information retrieval task. These results show very strong performance on the simulations investigated. The use of game theory as a principle suggests other applications and extensions of the method.

References

1 Komiyama, Junpei and Honda, Junya and Kashima, Hisashi and Nakagawa, Hiroshi *Regret Lower Bound and Optimal Algorithm in Dueling Bandit Problem*. COLT 2015,

http://jmlr.org/proceedings/papers/v40/Komiyama15.html, JMLR Proceedings, Vol. 40, pp 1141-1154, 2015.

3.26 Fundamentals of ESs' Statistical Learning

Ofer M. Shir (Tel-Hai College – Upper Galilee, IL)

Joint work of Ofer M. Shir, A. Yehudayoff
Main reference Shir, O. M. and Yehudayoff, A. (2017). On the statistical learning ability of evolution strategies. In Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, FOGA'17, pages 127–138, New York, NY, USA. ACM.
URL http://dx.doi.org/10.1145/3040718.3040722

We consider Evolution Strategies operating only with isotropic Gaussian mutations on positive quadratic objective functions and investigate the Covariance matrix when constructed out of selected individuals by truncation. We prove that the statistically constructed Covariance matrix over such selected decision vectors becomes proportional to the inverse of the landscape Hessian as the population size increases. This generalizes a previous result that proved an equivalent phenomenon when sampling is carried out in the vicinity of the optimum [FOGA'17]. It further confirms the classical hypothesis that statistical learning of the landscape is an inherent characteristic of standard ESs, and that this distinguishing capability stems only from the usage of isotropic Gaussian mutations and rank-based selection.

References

1 Shir, O. M. and Yehudayoff, A. On the statistical learning ability of evolution strategies. In Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, FOGA'17, pages 127–138, New York, NY, USA. ACM, 2017

3.27 Low discrepancy for one-shot optimization

Olivier Teytaud (Google Switzerland – Zürich, CH)

Joint work of Google Brain, Zurich

The selection of hyper-parameters is critical in Deep Learning. Because of the long training time of complex models and the availability of compute resources in the cloud, "one-shot" optimization schemes – where the sets of hyperparameters are selected in advance (either on a grid or in a random manner) and the training is executed in parallel – are commonly used. (Bergstra & Bengio, 2012) show that grid search is sub-optimal, especially when only a few parameters matter, and suggest to use random search instead. Yet, random search can be "unlucky" and produce sets of values that leave some part of the domain unexplored. Quasi-random methods, such as Low Discrepancy Sequences (LDS) avoid these issues. We show that such methods have theoretical properties that make them appealing for performing hyperparameter search, and demonstrate that, when applied to the selection of hyperparameters of complex Deep Learning models (such as state-of-the-art LSTM language models), they yield suitable hyperparameters values with much fewer runs than random search. We propose a particularly simple LDS method which can be used as a drop-in replacement for grid/random search in any Deep Learning pipeline.

3.28 Recent Advances in Runtime Analysis of Estimation-of-Distribution Algorithms

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

We consider three simple estimation-of-distribution algorithms (EDAs) on the OneMax benchmark function. The runtime of these algorithms depends on the number of bits n and the precision λ of the model it builds. Exponential runtimes as well as polynomial runtimes of the kind $\Theta(\lambda n)$ and $\Theta(\lambda\sqrt{n})$ are obtained in different regimes for λ. Two phase transitions in the runtime behavior are identified, and open problems are discussed.

Literature:

References

1 Tianshi Chen, Ke Tang, Guoliang Chen, and Xin Yao. On the analysis of average time complexity of estimation of distribution algorithms. In *Proc. of CEC '07*, pages 453–460, 2007.

2 Tianshi Chen, Per Kristian Lehre, Ke Tang, and Xin Yao. When is an estimation of distribution algorithm better than an evolutionary algorithm? In *Proc. of CEC '09*, pages 1470–1477, 2009a.

3 Tianshi Chen, Ke Tang, Guoliang Chen, and Xin Yao. Rigorous time complexity analysis of univariate marginal distribution algorithm with margins. In *Proc. of CEC '09*, pages 2157–2164, 2009b.

4 Tianshi Chen, Ke Tang, Guoliang Chen, and Xin Yao. Analysis of computational time of simple estimation of distribution algorithms. *IEEE Transactions on Evolutionary Computation*, 14 (1): 1–22, 2010.

5 Duc-Cuong Dang and Per Kristian Lehre. Simplified runtime analysis of estimation of distribution algorithms. In *Proc. of GECCO '15*, pages 513–518, 2015.

6 Stefan Droste. A rigorous analysis of the compact genetic algorithm for linear functions. *Natural Computing*, 5(3): 257–283, 2006.

7 Tobias Friedrich, Timo Kötzing, Martin S. Krejca, and Andrew M. Sutton. The benefit of recombination in noisy evolutionary search. In *Proc. of ISAAC '15*, pages 140–150, 2015.

8 Tobias Friedrich, Timo Kötzing, and Martin S. Krejca. EDAs cannot be balanced and stable. In *Proc. of GECCO '16*, pages 1139–1146, 2016.

9 Martin S. Krejca and Carsten Witt. Lower bounds on the run time of the univariate marginal distribution algorithm on OneMax. In *Proc. of FOGA 2017*, pages 65–79, 2017.

10 Per Kristian Lehre and Phan Trung Hai Nguyen. Improved runtime bounds for the univariate marginal distribution algorithm via anti-concentration. In *Proc. of GECCO '17*, 2017. To appear.

11 Frank Neumann, Dirk Sudholt, and Carsten Witt. A few ants are enough: ACO with iteration-best update. In *Proc. of GECCO '10*, pages 63–70, 2010.

12 Dirk Sudholt and Carsten Witt. Update strength in EDAs and ACO: How to avoid genetic drift. In *Proc. of GECCO '16*, pages 61–68, 2016.

13 Carsten Witt. Upper Bounds on the Runtime of the Univariate Marginal Distribution Algorithm on OneMax. In *Proc. of GECCO '17*, 2017. To appear. Preprint: http://arxiv.org/abs/1704.00026

3.29 (1+λ) Evolutionary Algorithm with Self-Adjusting Mutation Rate

Jing Yang (Ecole Polytechnique – Palaiseau, FR)

Joint work of Benjamin Doerr, Christian Gießen, Carsten Witt, Jing Yang

We propose a self-adapting strategy for the population-based evolutionary algorithms. It creates half the offspring with a mutation rate that is twice the current mutation rate and the other half with half the current rate. The mutation rate is then updated to the rate which generates the best offspring. We prove that the self-adapting (1+λ) EA solves OneMax in an expected generations of $O(n/\log(\lambda) + n\log(n)/\lambda)$. According to previous work, this is best-possible among all lambda-parallel mutation-based unbiased black-box algorithms.

4 Working groups

4.1 Breakout Session: Information Geometric Optimization

Youhei Akimoto (Shinshu University – Nagano, JP)

Joint work of Youhei Akimoto, Dirk Arnold, Anne Auger, Hans-Georg Beyer, Dimo Brockhoff, Tobias Glasmachers, Nikolaus Hansen, Christian Igel

Information geometric optimization (IGO) [1] is a generic framework for probability model based search algorithms for arbitrary search space. Given a family of probability distributions over the search space, a probability model based search algorithm like an estimation of distribution algorithm (EDA) is derived. The population based incremental learning (PBIL) for binary optimization, compact genetic algorithm (cGA), and some components of the covariance matrix adaptation evolution strategy (CMA-ES) can be derivedfrom the IGO framework, while they have been proposed independently from it. The development of the IGO framework has so far contributed to analyze a simplified CMA-ES [2] and to derive a novel variant of the CMA-ES [3]. However, as the outcome of these results, we find that the analyzed model of the CMA-ES behaves differently from what we observe in practice. Moreover, the components of the CMA-ES that are not included in the IGO framework such as the step-size adaptation and the rank-one covariance matrix update are often critical to the performance, hence we need to incorporate them when deriving a novel algorithm. The objective of the breakout session was to share our understanding of the algorithms and the IGO framework and to develop the framework so that it provides more reasonable mathematical models for analysis and more practical algorithms. In this breakout session, we started with discussing how the rank-one covariance matrix update can be interpreted. Then, we discussed why the ordinary differential equation that the IGO framework provides as a mathematical model to analyze behaves differently from the behavior of the real algorithm, and how we can correct it. The use of stochastic differential equations was proposed. In relation to this, we also discussed the dependency of the algorithm performance on the choice of the parameters of the probability distribution. Finally, we shared recent developments on the IGO and related algorithms such as the population size adaptation [4].

References

1 Yann Ollivier, Ludovic Arnold, Anne Auger, Nikolaus Hansen. Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles, Journal of Machine Learning Research 18(18), pages 1-65, 2017.

2 Hans-Georg Beyer. Convergence Analysis of Evolutionary Algorithms That are Based on the Paradigm of Information Geometry, Evolutionary Computation 22(4), pages 679-709, 2014.

3 Youhei Akimoto, Anne Auger, Nikolaus Hansen. Comparison-Based Natural Gradient Optimization in High Dimension, In Proceedings of the Genetic and Evolutionary Computation Conference, pages 373-380, 2014.

4 Kouhei Nishida, Youhei Akimoto. Population Size Adaptation for the CMA-ES Based on the Estimation Accuracy of the Natural Gradient, In Proceedings of the Genetic and Evolutionary Computation Conference, pages 237-244, 2016.

4.2 Breakout Session: Preference-based Selection in Evolutionary Multiobjective Optimisation

Carlos M. Fonseca (University of Coimbra, PT)

Participants: Dimo Brockoff, Arina Buzdalova, Francisco Chicano, Carlos M. Fonseca, Jonathan L. Shapiro

Multiobjective optimisation consists in simultaneously optimising two or more incommensurable objective functions over the same domain. In practice, however, the optima of all such objective functions seldom coincide, and there is no ideal solution. Instead, there are usually multiple incomparable solutions that are optimal (or *efficient*) in the sense that no other solution is at least as good in all objectives and strictly better in at least one of them. Therefore, a compromise solution is often sought among all efficient solutions. The image of the set of efficient solutions in objective space is known as the Pareto front.

Selecting a single compromise solution involves a decision making process, and requires additional information, known as preference information, which may not be explicitly available when optimisation is performed. One approach to this situation consists in searching for a diverse set of efficient solutions, in the hope that it will contain a suitable compromise solution. Traditionally, preference information has been considered to pertain to individual candidate solutions, but this notion has meanwhile been extended to sets of candidate solutions, including the preference for diverse sets of solutions.

The aim of this breakout session was to discuss how solution-oriented preferences and set-oriented preferences relate to each other, and to what extent they can be meaningfully combined. The starting point for the discussion was the observation that diversity appears to be in contradiction with solution-oriented preferences. In particular, if the preferences of the Decision Maker (DM) are fully known in advance, why should optimisation provide a diverse set of non-preferred solutions?

In Evolutionary Multiobjective Optimisation (EMO), three main approaches to diversity promotion can be identified. Initially, selection was guided by the solution-oriented preferences, and diversity was promoted by other means, such as niching techniques [1]. In indicator-based approaches [2] and decomposition-based approaches [3], on the other hand, diversity is promoted directly via selection, but in different ways. Whereas quality indicators directly specify what a good set is, in decomposition-based approaches multiple solution-oriented preferences are considered, leading to multiple preferred solutions. In can be argued that both of these approaches attempt to approximate the Pareto front in some way, where approximation quality is tied to set-oriented preferences that must be expressed by the DM.

An alternative view embraces the fact that DM preferences are seldom well understood in advance, and adopts a probabilistic view of the associated uncertainty. As a result, the quality of a single solution is itself a random variable, and selection can be reinterpreted as a portfolio optimisation problem [4]. This entails modelling the uncertainty about how good each solution is and the dependence between the quality of pairs of solutions, rather than expressing preferences about sets. Solving the portfolio optimisation problem consists of maximising expected return (i.e., some measure of DM satisfaction) and balancing it against deviations from this expectation (risk), leading to diverse solutions in the portfolio. In contrast to other approaches, diversity emerges as a risk-balancing strategy due to preference uncertainty, rather than being preferred as such.

From the discussion, the formulation of alternative preference uncertainty models, ideally supporting DM interaction, the study of alternative portfolio optimisation formulations leading to simpler computational problems (e.g., linear versus quadratic programming) and the characterisation of the resulting quality indicators were identified as current research challenges.

References

1 C. M. Fonseca and P. J. Fleming. "Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization," in *Genetic Algorithms: Proceedings of the Fifth International Conference*, pp. 416–423, San Mateo, CA: Morgan Kaufmann, 1993.

2 E. Zitzler and S. Künzli. "Indicator-based selection in multiobjective search," in *Parallel Problem Solving from Nature – PPSN VIII*, vol. 3242 of LNCS, pp. 832–842, Springer, 2004.

3 Q. Zhang and H. Li. "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 712–731, 2007.

4 I. Yevseyeva, A. P. Guerreiro, M. T. M. Emmerich, and C. M. Fonseca. "A portfolio optimization approach to selection in multiobjective evolutionary algorithms." in *Parallel Problem Solving from Nature – PPSN XIII*, vol. 8672 of LNCS, pp. 672–681, Springer, 2014.

4.3 Breakout Session: Drift Theorems for Continuous Optimization

Tobias Glasmachers (Ruhr-Universität Bochum, DE)

Participants: Nikolaus Hansen, Anne Auger, Per Kristian Lehre, Günter Rudolph, Johannes Lengler, Carlos Fonseca, Youhei Akimoto, Carsten Witt, Hans-Georg Beyer, Tobias Glasmachers

Concept of the session:

- inform everyone on current state-of-the-art drift theorems
- collect requirements to make drift analysis applicable to optimization and online parameter control in continuous optimization
- see what can already be done, check discrepancies, check what can be improved and adapted in the theorems to make them more directly applicable to continuous problems

We started with a short presentation given by Per Kristian. It turns out that even very classic results are directly applicable to continuous problems. The seeming discrepancy was mostly caused by specific adaptations of general drift theorems (coming from supermartingale theory) to the specific needs of discrete optimization. The only relevant limitation seems be that spaces need to be bounded in order to obtain lower bounds, but expected runtime and upper bounds can be derived in surprising generality. Johannes took over, explaining a simple and insightful proof, demonstrating why prerequisites are more strict for lower bounds.

Results are generally of the following form. Given a stochastic process $(Y_k)_{k\in\mathbb{N}}$ and a drift condition $E[Y_{k+1}|Y_k] \leq d$, the expected runtime for hitting a lower bound a is simply $(Y_0 - a)/d$. Statements for quantiles look similar and hold with overwhelming probability (one minus exponentially small exception). Besides this additive drift there is also multiplicative drift and variable drift.

These results seem to be generally well suited for the analysis of evolution strategies, for all state variables (optimization progress, step size and covariance matrix adaptation, and evolution paths). However, finding good potential functions may be challenging. Combined drift (presented earlier by Jonathan Rowe) seems to be highly relevant.

As proposed by Anne, the next step is to try this out in simple instructive cases, like the 1/5 success rule.

We closed with the question of good resources presenting these drift theorems in the form of a review paper or a concise collection. Per Kristian and Carsten will update their arXiv papers to reflect the state-of-the-art.

It turns out that Jens Jägerküpper had already used such techniques in his convergence analysis of the (1+1)-ES, although it was not recognized at the time that the technique is of very generic value, and it was not called drift.

The session went surprisingly smoothly. In contrast to the expectations, we did not encounter any serious obstacles. It seems that all prerequisites are in place for applying drift to the analysis of evolution strategies in the near future.

4.4 Breakout Session: Discrete Estimation of Distribution Algorithms

Martin S. Krejca (Hasso-Plattner-Institut – Potsdam, DE)

Participants: Youhei Akimoto, Tobias Friedrich, Christian Gießen, Nikolaus Hansen, Martin S. Krejca, Per Kristian Lehre, Johannes Lengler, Frank Neumann, Dirk Sudholt, Andrew M. Sutton, Carsten Witt

This breakout session was held to discuss the next steps to take in the analysis of estimation of distribution algorithms (EDAs). Many of the active researchers in that area were present.

We started by discussing open problems and searching for topics that everyone was interested in. The discussed topics included optimization under constraints, more general run time analyses, and dependencies. Our group was most interested in dependencies. One of the selling points of EDAs generally is that they can adapt very well to the underlying structure of a problem, which often includes dependencies. Since most of the EDAs considered so far in theory use a univariate model (assuming independence of the problem variables) [1, 2, 3, 4], it seemed important and necessary to also analyze multivariate EDAs.

After discussing different multivariate EDAs, our group agreed to start with MIMIC [5]: a bivariate EDA whose probabilistic model generates a permutation of the problem variables. A natural first problem to analyze for such an algorithm is LeadingOnes, which is the problem of finding a hidden permutation. However, we were not convinced that MIMIC will outperform univariate EDAs on that function. This lead to a discussion of MIMIC's performance on other functions like OneMax. Overall, we decided to first run some experiments to see how MIMIC performs on standard benchmarks used in theory. After that, we want to prove the observed behavior.

References

1 T. Friedrich, T. Kötzing, and M. S. Krejca. EDAs Cannot Be Balanced and Stable. In *Proc. of GECCO '16*, pp. 1139–1146, 2016.
2 P. K. Lehre and P. T. H. Nguyen. Tight Bounds on Runtime of the Univariate Marginal Distribution Algorithm via Anti-Concentration. In *Proc. of GECCO '17*, 2017.
3 D. Sudholt and C. Witt. Update Strength in EDAs and ACO: How to Avoid Genetic Drift. In *Proc. of GECCO '16*, pp. 61–68, 2016.
4 C. Witt. Upper Bounds on the Runtime of the Univariate Marginal Distribution Algorithm on OneMax. In *Proc. of GECCO '17*, 2017.
5 J. S. De Bonet, C. L. Isbell, and P. Viola. MIMIC: Finding Optima by Estimating Probability Densities. In *Proc. of NIPS '97*, pp. 424–431, 1997.

4.5 Breakout Session: Theory of Evolutionary Algorithms for Problems With (Dynamic) Constraints

Frank Neumann (University of Adelaide, AU)

Participants: Tobias Friedrich Carola Doerr Andrew M. Sutton William B. Langdon Dirk Sudholt Jing Yang Francisco Chicano Arina Buzdalova Thomas Jansen Martin Krejca Christine Zarges Christian Giessen Anton Eremeev Pietro Oliveto Dimo Brockhoff Frank Neumann

Most of the results in the area of runtime analysis are for unconstrained problems. However, in practice constraints play a crucial role. The aim of the breakout session has been to explore new research directions to provide meaningful theoretical insights into the working behavior of evolutionary computing techniques for problems with dynamically changing constraints. Starting the discussion, it has been observed that understanding evolutionary computing techniques for problems with constraints is a rather unexplored area (apart from a few publications so far). So the focus of the discussion has been on static constraints.

Linear functions play a key role in the area of rigorous runtime analysis of evolutionary computation. In [1], the runtime of evolutionary algorithms for linear functions with a linear constraint have been examined.

$$\max f(x) = \sum_{i=1}^{n} w_i x_i \text{ subject to } g(x) = \textstyle\sum_{i=1}^{n} g_i x_i \leq B.$$

In the general case, this is equivalent to the classical knapsack problem. However, it has been pointed out that the (1+1) EA which optimizes unconstrained linear functions in time $\Theta(n \log n)$ requires 2-bit flips even if the constrained is uniform, i.e. $g_i = 1$, $1 \leq i \leq n$. Understanding how even this simple case and provide tight upper and lower bounds would be of interest.

The participants mentioned different constrained handling mechanisms that could be examined such as a weightening of the given objective function f and the constrained g by considering the fitness given by

$$f(x) - \alpha \cdot \max\{0, g(x) - B\}.$$

An important question would be how to choose α (possibly adapting it over the run of the algorithm). Furthermore, multi-objective formulations (taking the constrained as an additional objective) may be examined.

Another topic discussed has been different black box models. For example, is the EA able to get information on the amount of constrained violation for an infeasible solutions or does it only get the information that a solution is infeasible. It's worthwhile comparing such different black box settings for prominent examples and analyze how evolutionary algorithms perform in the different settings.

Later there has been a discussion on dynamic changes to the constraints. One important question is what is the reoptimization time of evolutionary algorithms, i.e. the time to recompute a good or optimal solution after a change to the constraints has happened. Other questions involved the benefit of a population to cater for changes or reoptimize quicker. Furthermore, constrained handling mechanisms in relation to dynamic changes could be examined.

The breakout session has shown that there are a lot of open questions and interesting research directions for understanding how and why evolutionary algorithms can deal with constrained optimization problems.

References

1 T. Friedrich, T. Kötzing, G. Lagodzinski, F. Neumann, M. Schirneck: Analysis of the (1+1) EA on Subclasses of Linear Functions under Uniform and Linear Constraints. FOGA 2017: 45-54

4.6 Breakout Session: Diversity

Lothar Thiele (ETH Zürich, CH)

The breakout session was devoted to the concept of diversity. At the beginning, we discussed various areas in randomized search algorithms where the concept of diversity plays a major role. In particular, we agreed on the following classification:

Genotype: Diversity is a fundamental issue when discussing variation operators like crossover. In addition, it is used when arguing about adaptation and control, as well discussing about diversity in genotypes in order to deal with multi-modal optimization problems.

Phenotype: A novel and interesting optimization objective in many practical applications is the diversity of solutions, i.e. the goal to determine not only one solution but a set with different characteristics.

Objective Function: In multi-objective optimization, one major issue is to achieve diversity in the objective space. In particular, many methods are available to achieve a diverse set of solutions on or close to the Pareto Front.

Search Behavior: In complex search spaces, it is important to achieve a diverse search behavior, i.e. by using meta-search approaches. Moreover, other aspects of search behavior related to diversity are exploration vs. exploitation, initial solutions to randomized search algorithms, as well as examples for machine learning.

After discussing in some detail the above mention classification, possible definitions of diversity have been discussed that are suitable in several of the above instances:

- How well does a set cover a universe?
- Using a suitable information-theoretic measure, like Kolmogorov complexity.
- Following the concept of diversity as used in biology: Given is a completely connected undirected graph with edge weights. Adding a duplicate does not change the diversity, adding a distinct node increases the diversity, and increasing one of the edge weights also increases it.

Some of the interesting findings that should be considered further are the link to discrepancy theory and the associated concept of dispersion, and the fact that the desire of diversity is often a sign of uncertainty. The need for a general definition of diversity was questioned as (a) ad-hoc definitions worked well so far, and (b) diversity can also be controlled implicitly (selection pressure, mutation rate).

4.7 Breakout Session: COST Action CA15140

Christine Zarges (Aberystwyth University, GB)

Participants: Maxim Buzdalov, Arina Buzdalova, Francisco Chicano, Carola Doerr, Anton V. Eremeev, Carlos M. Fonseca, Thomas Jansen, William B. Langdon, Pietro S. Oliveto, Ofer M. Shir, Christine Zarges

The main aim of this breakout session was to discuss potential future collaborations within COST Action CA15140 "Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO)".

4.7.1 Overview of the Action

COST is a European framework that provides funding for networking activities with an emphasis on Early Career Investigators (researchers with less than 8 years between the date of the PhD/doctorate and the date of involvement in the COST Action), inclusiveness and widening participation. Since most of the participants were not involved in the COST Action, yet, the first part of the meeting mainly aimed at introducing different activities and upcoming events and how new people can join.

4.7.1.1 Working Groups

Christine Zarges introduced the main aim of COST Action CA15140, which is chaired by Thomas Jansen: Build a platform where theoreticians and practitioners in nature-inspired optimisation can meet and exchange insights, ideas and needs. To achieve this, activities are planned along four different working groups:

- Working Group 1 (Theory-Driven Applications), led by Tobias Friedrich, is tasked with the development of novel theory-driven practical paradigms, thus pushing from theory to practice. Starting point are existing theoretical results and insights and the task is to use those to create practical guidelines, tangible advice and help for the application of nature-inspired search and optimisation heuristics.
- Working Group 2 (Practice-Driven Theory), led by Christine Zarges and Bosko Blagojevic, is tasked with the development of novel practice-driven theoretical frameworks and paradigms, thus pushing from practice to theory. Starting point are needs and unanswered questions as they arise in applications and the task is to create theoretical perspectives and novel results that directly address these needs.
- Working Group 3 (Benchmarks), led by Günther Raidl and Borys Wrobel, is tasked with the development of useful benchmarks for nature-inspired search and optimisation heuristics with a strong focus on discrete search spaces and discrete optimisation problems making sure that the developed benchmarks are relevant from a practical perspective and accessible from a theoretical perspective.
- Working Group 4 (Software), led by Carlos Fonseca and Florin Pop, is tasked with support for software development with a focus on the development of useful rules for the development of software and the adaptation of nature-inspired search and optimisation heuristics that are based on and guided by theoretical insights in their functioning.

More detailed information and current developments can be found on the COST action website: http://imappnio.dcs.aber.ac.uk

4.7.1.2 Training School

Carola Doerr introduced the upcoming COST training school, which will be centred around bridging the gap between theory and practice and making nature-inspired search and optimisation heuristics more applicable. It will take place from 18 to 24 October 2017 in Paris, France, right before the Biennial International Conference on Artificial Evolution (EA 2017), https://ea2017.inria.fr. Participation will be free and limited funding will be available for trainees from participating COST countries (see below). It is expected to be sufficient to pay for accommodation, subsistence and a significant contribution towards the cost of travel. For more details on the application process contact the action chair.

4.7.1.3 Short-Term Scientific Missions

Carola Doerr and Christine Zarges introduced the tool of Short-Term Scientific Missions (STSMs), exchange visits between researchers from two different countries involved in a COST Action. Applications for STSMs are invited at any time and need to be made via the e-COST system. More information about the process can be found on the COST action website: http://imappnio.dcs.aber.ac.uk/stsms

4.7.1.4 How to Participate?

The best first point of contact is the action chair, Thomas Jansen. To join a working group an additional email to the working group leader can be useful.

In a nutshell, COST distinguishes between three different types of countries: Member states (http://www.cost.eu/about_cost/cost_member_states), COST Near Neighbour Countries (http://www.cost.eu/about_cost/strategy/international_cooperation/nnc) and COST International Partner Countries (any other country). Any researcher affiliated with an institution in a member state already participating in the action is eligible for all activities. Researchers from non-participating member states need to contact the action chair and their national coordinator (http://www.cost.eu/about_cost/who/(type)/3) first to discuss how the state can join the action. The current list of participants can be found here: http://www.cost.eu/COST_Actions/ca/CA15140?parties. Institutions from COST Near Neighbour and International Partner Countries can join on a case by case basis and should discuss this with the action chair.

4.7.2 Discussion of Future Directions

In the remainder of the meeting, ideas for the training school including potential speakers were discussed in more detail. It was suggested to have ThRaSH-like talks towards the end of each day to make the school more attractive for senior researchers in the field who are not directly involved in the training.

A second discussion was centred around more concrete research ideas and potential routes to make working groups more effective. As a starting point participants shared their personal experiences including concrete collaborations and benchmarks. It was argued that issues such as language barriers or data protection could be overcome by first concentrating on problem modelling and presenting benchmarks as a black-box.

It was also suggested that the perceived gap between theory and practice is not symmetric as theory cannot expect to make significant progress in the available timeframe. Thus, an idea would be to ask practitioners to pose very simple questions to obtain a useful starting point and gain clarity about open questions. Here, it might be more promising to concentrate on collaborations with partners who are interested in developing methods to solve a specific *kind* of problem, rather than solutions for a very specific problem. However, in any case getting different researchers interested in theory and practice and collaborating on each others problems is a good starting point. A success story that solves a concrete problem and poses new questions would be desirable.

Finally, Carlos Fonseca gave a summary of the panel discussion at the COST industry workshop in Copenhagen earlier this year discussing requirements for evolutionary algorithms (e. g., speed, scalability, good default parameterisation), obstacles to adoption of evolutionary algorithms (e. g., lack of integration, trust and education) and aspects in favour of evolutionary algorithms (e. g., cost effectiveness, availability). It is hoped that wide collaboration in the COST Action can help to address some of these obstacles and generally improve the applicability of such methods.

5 Seminar Schedule

Monday, May 8

9.00 - 10.15	Welcome
	Introduction of Participants (40 participants, each 1–2 minutes)
10.15 - 10.45	Coffee Break
10.45 - 12.00	*Matt Hoffman*: Introductory Talk on Bayesian Optimization
12.15 - 14.00	Lunch Break
14.00 - 15.30	*Carsten Witt*: Recent Advances in Runtime Analysis of Estimation-of-Distribution Algorithms
	Olivier Teytaud: Randomized one-shot optimization
	J*on Rowe*: Linear multi-objective drift analysis
15.30 - 16.00	Coffee Break
16.00 - 18.00	*Benjamin Doerr*: Fast Genetic Algorithms
	Johannes Lengler: Noise Models for Comparison-Based EAs
	Bill Langdon: The fitness landscape of genetic improvement
	Thomas Jansen: COST Action

Tuesday, May 9

9.00 - 10.15	*Carlos M. Fonseca*: Mathematical Models of Artificial Genetic Representations with Neutrality
	Planning of the Breakout Sessions
10.15 - 10.45	Coffee Break
10.45 - 12.00	*Dirk Arnold*: Evolutionary optimization with constraints
	Dimo Brockhoff: Towards a Constrained Test Suite for COCO
	Tobias Glasmachers: Global Convergence of the (1+1)-ES
12.15 - 14.00	Lunch Break
14.00 - 15.30	Breakout Session I
	◦ Theory of evolutionary algorithms for problems with (dynamic) constraints
	◦ Drift theorems for continuous optimization - what we have vs. what we would need
	◦ Preference-based selection in evolutionary multiobjective optimization
15.30 - 16.00	Coffee Break
16.00 - 18.00	*Per Kristian Lehre*: Self-adaptation
	Maxim Buzdalov: How to Exploit Your Fitness-Distance Correlation: Runtime Analysis of the (1+(?,?))-GA on Random Satisfiable 3-CNF Formulas
	Christian Giessen: Monotone Functions on Bitstrings - Some Structural Notes

Wednesday, May 10

9.00 - 10.15	*Nikolaus Hansen*: How to Guarantee Positive Definiteness in Active CMA-ES
	Hans-Georg Beyer: Towards a Theory of CMA-ES: But first Simplify your CMA-ES
10.15 - 10.45	Coffee Break
10.45 - 12.00	*Anton Eremeev*: Optimal Recombination for the TSP: Theory and Experiments
	Frank Neumann: Features, Diversity, Random Walks and Images
12.15	Lunch
13.30 - 15.30	Social Activity: Walk and Talk
15.30 - 16.30	Coffee Break
16.30 - 18.00	Breakout Session II
	◦ COST: Theory vs. Practice
	◦ Discrete estimation of distribution algorithms

Thursday, May 11

9.00 - 10.00	*Ofer Shir*: Fundamentals of evolution strategies' statistical learning
	Anne Auger: Connecting stability of Markov chains and deterministic control models for analyzing randomized algorithms
10.00 - 10.30	Coffee Break
10.30 - 12.00	*Pietro Oliveto*: Standard Steady State Genetic Algorithms can Hillclimb Faster than Mutation-Only Evolutionary Algorithms
	Günter Rudolph: Theoretical Aspects of the Averaged Hausdorff Indicator in Biobjective Optimization
	Group Photo
12.15 - 14.00	Lunch Break
14.00 - 15.30	Breakout Session III
	◦ Diversity in randomized optimization
	◦ Information geometric optimization. Topics: how to interpret the full CMA-ES in IGO
15.30 - 16.00	Coffee Break
16.00 - 18.00	*Francisco Chicano*: On the Variable Interaction Graph in Gray-Box Optimization
	Jing Yang: $(1+\lambda)$ Evolutionary Algorithm with Self-adjusting Mutation Rate
	Summary and Discussion of Breakout Sessions

Friday, May 12

9.15 - 10.15	*Jonathan Shapiro*: Max-Min Thompson Sampling for the K-arm dueling bandit problem
	Yohei Akimoto: Optimal Step-size for Weighted Recombination Evolution Strategy
	Bill Langdon: Long-term convergence vs long-term experimental evolution with genetic programming trees
10.15 - 10.45	Coffee Break
10.45 - 12.00	*Martin Krejca*: Estimation of Distribution Algorithms
	Arina Buzdalova: Is it necessary to perform multi-objective optimization when doing multi-objectivization?
	Wrap-up
12.15	Lunch

Participants

- Youhei Akimoto
Shinshu University – Nagano, JP
- Dirk V. Arnold
Dalhousie University – Halifax, CA
- Anne Auger
INRIA RandOpf Team, FR
- Thomas Bäck
Leiden University, NL
- Hans-Georg Beyer
Fachhochschule Vorarlberg – Dornbirn, AT
- Dimo Brockhoff
INRIA Saclay – Île-de-France, FR
- Maxim Buzdalov
ITMO University – St. Petersburg, RU
- Arina Buzdalova
ITMO University – St. Petersburg, RU
- Francisco Chicano
University of Málaga, ES
- Benjamin Doerr
Ecole Polytechnique – Palaiseau, FR
- Carola Doerr
CNRS & UPMC, Paris, FR
- Anton V. Eremeev
Sobolev Institute of Mathematics – Novosibirsk
- Carlos M. Fonseca
University of Coimbra, PT
- Tobias Friedrich
Hasso-Plattner-Institut – Potsdam, DE
- Christian Gießen
Technical University of Denmark – Lyngby, DK
- Tobias Glasmachers
Ruhr-Universität Bochum, DE
- Nikolaus Hansen
INRIA Saclay – Île-de-France, FR
- Matthew W. Hoffman
Google DeepMind – London, GB
- Christian Igel
University of Copenhagen, DK
- Thomas Jansen
Aberystwyth University, GB
- Martin S. Krejca
Hasso-Plattner-Institut – Potsdam, DE
- William B. Langdon
University College London, GB
- Per Kristian Lehre
University of Birmingham, GB
- Johannes Lengler
ETH Zürich, CH
- Frank Neumann
University of Adelaide, AU
- Pietro S. Oliveto
University of Sheffield, GB
- Adam Prugel-Bennett
University of Southampton, GB
- Jonathan E. Rowe
University of Birmingham, GB
- Günter Rudolph
TU Dortmund, DE
- Jonathan L. Shapiro
University of Manchester, GB
- Ofer M. Shir
Tel-Hai College – Upper Galilee, IL
- Dirk Sudholt
University of Sheffield, GB
- Andrew M. Sutton
Hasso-Plattner-Institut – Potsdam, DE
- Olivier Teytaud
Google Switzerland – Zürich, CH
- Lothar Thiele
ETH Zürich, CH
- Heike Trautmann
Universität Münster, DE
- Carsten Witt
Technical University of Denmark – Lyngby, DK
- Jing Yang
Ecole Polytechnique – Palaiseau, FR
- Xin Yao
University of Birmingham, GB
- Christine Zarges
Aberystwyth University, GB