

# Theory of Randomized Optimization Heuristics

Anne Auger<sup>\*1</sup>, Carlos M. Fonseca<sup>\*2</sup>, Tobias Friedrich<sup>\*3</sup>,  
Johannes Lengler<sup>\*4</sup>, and Armand Gissler<sup>†5</sup>

- 1 INRIA Saclay – Palaiseau, FR. [anne.auger@inria.fr](mailto:anne.auger@inria.fr)
- 2 University of Coimbra, PT. [cmfonsec@dei.uc.pt](mailto:cmfonsec@dei.uc.pt)
- 3 Hasso-Plattner-Institut, Universität Potsdam, DE. [friedrich@hpi.de](mailto:friedrich@hpi.de)
- 4 ETH Zürich, CH. [johannes.lengler@inf.ethz.ch](mailto:johannes.lengler@inf.ethz.ch)
- 5 Ecole Polytechnique – Palaiseau, FR. [armand.gissler@ens-paris-saclay.fr](mailto:armand.gissler@ens-paris-saclay.fr)

---

## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22081 “Theory of Randomized Optimization Heuristics”.

This seminar is part of a biennial seminar series. This year, we focused on connections between classical topics of the community, such as Evolutionary Algorithms and Strategies (EA, ES), Estimation-of-Distribution Algorithms (EDA) and Evolutionary Multi-Objective Optimization (EMO), and related fields like Stochastic Gradient Descent (SGD) and Bayesian Optimization (BO). The mixture proved to be extremely successful. Already the first talk turned into a two hour long, vivid and productive plenary discussion. The seminar was smaller than previous versions (due to corona regulations), but its intensity more than made up for the smaller size.

**Seminar** February 20–25, 2022 – <http://www.dagstuhl.de/22081>

**2012 ACM Subject Classification** Theory of computation → Bio-inspired optimization; Theory of computation → Evolutionary algorithms; Theory of computation → Theory of randomized search heuristics

**Keywords and phrases** black-box optimization, derivative-free optimization, evolutionary and genetic algorithms, randomized search algorithms, stochastic gradient descent, theoretical computer science

**Digital Object Identifier** 10.4230/DagRep.12.2.87

## 1 Executive Summary

*Anne Auger*  
*Carlos M Fonseca*  
*Tobias Friedrich*  
*Johannes Lengler*

**License**  Creative Commons BY 4.0 International license  
© Anne Auger, Carlos M Fonseca, Tobias Friedrich, Johannes Lengler

This seminar is part of a biennial seminar series. This year, we focused on connections between classical topics of the community, such as Evolutionary Algorithms and Evolutionary Strategies (EA, ES), Estimation-of-Distribution Algorithms (EDA) and Evolutionary Multi-Objective Optimization (EMO), and related fields like Stochastic Gradient Descent (SGD) and Bayesian Optimization (BO). The mixture proved to be extremely successful. Already the first talk turned into a two hours long, vivid and productive plenary discussion. Participants

---

\* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Theory of Randomized Optimization Heuristics, *Dagstuhl Reports*, Vol. 12, Issue 2, pp. 87–102

Editors: Anne Auger, Carlos M Fonseca, Tobias Friedrich, Johannes Lengler, and Armand Gissler



Dagstuhl Reports

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

like Peter Richtarik and Sebastian Stich brought valuable new perspectives from the SGD community, and Mickaël Binois contributed the BO perspective. This yielded some new approaches to long-standing open problems, specifically for a convergence proof of the CMA-ES algorithm on quadratic functions.

Another interesting and fruitful aspect of the seminar was a shift of perspective to search spaces that are under-represented in the community. Traditionally, the search spaces are product spaces, either discrete (especially the  $n$ -dimensional hypercube), or continuous ( $d$ -dimensional Euclidean space). This year we had some intense discussions in plenum and in working groups on other search spaces, triggered especially by Ekhine Irurozki's presentation on permutation spaces.

Naturally, a big part of the seminar was also devoted to classical topics of the community. Highlights included talks by Benjamin Doerr on the first runtime result for the Non-Dominated Sorting Genetic Algorithm (NSGA-II) and by Tobias Glasmachers on Convergence Analysis of the Hessian Estimation Evolution Strategy (HE-ES). The latter is the first convergence proof for a covariance matrix algorithm that does not truncate the condition number of the estimated covariance matrix. Some interesting new topics were also identified in traditional fields, such as whether we can understand better in which situations adaptivity is necessary for efficient optimization by considering  $k$ -adaptive query complexity of optimization benchmarks.

Overall, as organizers we were extremely happy with the mix of core community members and researchers from related fields. The connections with the latter were close enough that scientific discussions could (also) happen on technical levels, which is particularly useful since some low-hanging fruits are available from such interchanges. Importantly, the exchange happened between people who would probably not have met each other outside of the Dagstuhl Seminar.

The seminar took place during the peak of the Omicron wave of Covid19, which made planning very difficult. The key step during preparation phase was a survey among the participants a few weeks before the seminar. We asked how likely it was that they could participate in person, and under which circumstances they would prefer which format (in-person or hybrid). The participants signalled us very clearly that they wanted this event to happen, and that they wanted it to happen in person. We want to thank all participants for their support! Other seminars in the week before and after ours had to be cancelled altogether, and this might also have happened to our seminar if not for the determination of our participants.

The seminar was smaller than previous versions, due to corona regulations. Moreover, some participants had to cancel at the last moment because they were corona-positive, or because they had no reliable child care. Especially the latter point can be frustrating, and we hope that Dagstuhl will be able to resume their support for on-site child care in the future. On the positive side, the intensity of the seminar more than made up for the smaller size, and might even have been due to the smaller number of participants.

Finally, we want to thank Dagstuhl for their great support, both financially and to their great staff. We could always feel that it was their top priority to help us, and we are greatly indebted for the support!

The organizers,

Anne Auger, Carlos M Fonseca, Tobias Friedrich, Johannes Lengler

## 2 Table of Contents

### Executive Summary

*Anne Auger, Carlos M Fonseca, Tobias Friedrich, Johannes Lengler* . . . . . 87

### Overview of Talks

Selection in non-elitist populations: overview and open problems <i>Duc-Cuong Dang</i> . . . . .	91
A First Mathematical Runtime Analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) <i>Benjamin Doerr</i> . . . . .	91
Some Theoretical Thoughts on Permutation-based EAs <i>Benjamin Doerr</i> . . . . .	92
$k$ -Adaptive Black-Box Optimization <i>Carola Doerr</i> . . . . .	92
Convergence Analysis of the Hessian Estimation Evolution Strategy <i>Tobias Glasmachers</i> . . . . .	93
Evolution Strategies Reliably Overcome Saddle Points <i>Tobias Glasmachers</i> . . . . .	93
Introductory Talk on the Theory of Continuous Evolutionary Algorithms <i>Tobias Glasmachers</i> . . . . .	93
Theoretical Aspects of Set-Quality Indicators for Multiobjective Optimization <i>Andreia P. Guerreiro</i> . . . . .	94
Black-Box Permutation Problems and weighted medians <i>Ekhine Irurozki</i> . . . . .	94
Theory of Discrete Randomized Optimization Heuristics – The What, Why, and How <i>Martin S. Krejca</i> . . . . .	95
Failure on Easy Problems <i>Johannes Lengler, Benjamin Doerr, Carola Doerr, and Dirk Sudholt</i> . . . . .	95
Population Diversity Makes Lexicase Selection Fast <i>Johannes Lengler</i> . . . . .	96
Gray-box operator for Vertex Cover <i>Xiaoyue Li</i> . . . . .	96
Randomized Smoothing for Non-Convex Optimization <i>Sebastian U. Stich</i> . . . . .	97
Analyzing the Cost of Randomness in Evolutionary Algorithms <i>Dirk Sudholt and Carlo Kneißl</i> . . . . .	97
On the dynamics of the DE algorithm <i>Ricardo Takahashi</i> . . . . .	98
Mathematical models for Dominance Move: Comparisons and complexity analysis <i>Elizabeth Wanner</i> . . . . .	98

**Working groups**

<i>k</i> -adaptive black-box complexity <i>Carola Doerr and Johannes Lengler</i> . . . . .	98
Theory-Friendly Practical Modelling of Combinatorial Optimisation Problems for ROHs <i>Carlos M. Fonseca</i> . . . . .	99
Adaptation of proof techniques based on the natural SGD interpretation of CMA-ES <i>Tobias Glasmachers</i> . . . . .	100
Lyapunov potential for the linear convergence of CMA-ES <i>Tobias Glasmachers</i> . . . . .	100
Comma and plus selection strategies with noise <i>Johannes Lengler</i> . . . . .	100

**Open problems**

A problem where CMA-ES performs poorly <i>Nikolaus Hansen</i> . . . . .	101
--	-----

<b>Participants</b> . . . . .	102
-------------------------------	-----

## 3 Overview of Talks

### 3.1 Selection in non-elitist populations: overview and open problems

*Duc-Cuong Dang (University of Southampton, GB)*

**License** © Creative Commons BY 4.0 International license  
© Duc-Cuong Dang

**Joint work of** Duc-Cuong Dang, Anton V. Eremeev, Per Kristian Lehre

This talk summarises what we know about the selection mechanisms for non-elitist populations from a theory perspective and open problems. Particularly, we show how selection should be tuned to find an optimum, and which characteristics of the selection one should look for to address multiple optima efficiently. These results allow us to identify problem classes where non-elitist algorithms with the right selection and a proper setting to excel, compared to elitist algorithms or even when truncation selection is used. Tools used to prove these results, these limitations and open problems are discussed.

This talk is based on joint works with Anton V. Eremeev and Per Kristian Lehre [1, 2, 3, 4], and on the fruitful discussions with Pietro S. Oliveto and Tiago Paixao.

#### References

- 1 D. Corus and D. Dang and A.V. Eremeev and P.K. Lehre *Level-based analysis of genetic algorithms and other search processes*. IEEE Trans. Evol. Comput., 707–719, 2018.
- 2 D. Dang and A.V. Eremeev and P.K. Lehre *Runtime analysis of fitness-proportionate selection on linear*. <http://arxiv.org/abs/1908.08686>, 2019.
- 3 D. Dang and A.V. Eremeev and P.K. Lehre *Non-elitist evolutionary algorithms excel in fitness landscapes with sparse deceptive regions and dense valleys*. In GECCO '21: Genetic and Evolutionary Computation Conference, Lille, France, 1133–1141, July 2021.
- 4 D. Dang and A.V. Eremeev and P.K. Lehre *Escaping local optima with non-elitist evolutionary algorithms*. Thirty-Fifth AAAI Conference on Artificial Intelligence, 12275–12283, 2021.

### 3.2 A First Mathematical Runtime Analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II)

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

**License** © Creative Commons BY 4.0 International license  
© Benjamin Doerr

**Joint work of** Weijie Zheng, Yufei Liu, Benjamin Doerr

**Main reference** Weijie Zheng, Yufei Liu, Benjamin Doerr: “A First Mathematical Runtime Analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II)”, CoRR, Vol. abs/2112.08581, 2021.

**URL** <https://arxiv.org/abs/2112.08581>

In this talk, I want to discuss a recent joint work with Weijie Zheng (SUSTECH) and Yufei Liu (Polytechnique). The non-dominated sorting genetic algorithm II (NSGA-II) is the most intensively used multi-objective evolutionary algorithm (MOEA) in real-world applications. However, in contrast to several simple MOEAs analyzed also via mathematical means, no such study exists for the NSGA-II so far. In this work, we show that mathematical runtime analyses are feasible also for the NSGA-II. As particular results, we prove that with a population size larger than the Pareto front size by a constant factor, the NSGA-II with two classic mutation operators and three different ways to select the parents satisfies the same asymptotic runtime guarantees as the SEMO and GSEMO algorithms on the basic

OneMinMax and LOTZ benchmark functions. However, if the population size is only equal to the size of the Pareto front, then the NSGA-II cannot efficiently compute the full Pareto front (for an exponential number of iterations, the population will always miss a constant fraction of the Pareto front). Our experiments confirm the above findings.

### 3.3 Some Theoretical Thoughts on Permutation-based EAs

*Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR)*

License  Creative Commons BY 4.0 International license  
 © Benjamin Doerr

Joint work of Benjamin Doerr, Yassine Ghannane, Marouane Ibn Brahim

While the theoretical analysis of evolutionary algorithms (EAs) has made significant progress for pseudo-Boolean optimization problems in the last 25 years, only sporadic theoretical results exist on how EAs solve permutation-based problems.

To overcome the lack of permutation-based benchmark problems, we propose a general way to transfer the classic pseudo-Boolean benchmarks into benchmarks defined on sets of permutations. We then conduct a rigorous runtime analysis of the permutation-based  $(1 + 1)$  EA proposed by Scharnow, Tinnefeld, and Wegener (2004) on the analogues of the LEADINGONES and JUMP benchmarks. The latter shows that, different from bit-strings, it is not only the Hamming distance that determines how difficult it is to mutate a permutation  $\sigma$  into another one  $\tau$ , but also the precise cycle structure of  $\sigma\tau^{-1}$ . For this reason, we also regard the more symmetric scramble mutation operator. We observe that it not only leads to simpler proofs, but also reduces the runtime on jump functions with odd jump size by a factor of  $\Theta(n)$ . Finally, we show that heavy-tailed versions of both operators, as in the bit-string case, lead to speed-ups of order  $m^{\Theta(m)}$  on jump functions with jump size  $m$ .

### 3.4 $k$ -Adaptive Black-Box Optimization

*Carola Doerr (Sorbonne University – Paris, FR)*

License  Creative Commons BY 4.0 International license  
 © Carola Doerr

Many black-box optimization techniques have a high degree of adaptiveness. But there are problems for which adaptive sampling has only negligible advantages over non-adaptive sampling, e.g., the famous 2-color Mastermind problem studied by Erdős and Renyi (1963).

In this talk I propose a black-box complexity model that allows us to study the minimal number of queries that are needed to optimize a given problem  $f$ , in dependency of the number of iterations  $k$  that the algorithms are allowed to perform.

### 3.5 Convergence Analysis of the Hessian Estimation Evolution Strategy

*Tobias Glasmachers (Ruhr-Universität Bochum, DE)*

**License** © Creative Commons BY 4.0 International license  
© Tobias Glasmachers

**Joint work of** Tobias Glasmachers, Oswin Krause

**Main reference** Tobias Glasmachers, Oswin Krause: “Convergence Analysis of the Hessian Estimation Evolution Strategy”, *Evol. Comput.*, Vol. 30(1), pp. 27–50, 2022.

**URL** [https://doi.org/10.1162/evco\\_a\\_00295](https://doi.org/10.1162/evco_a_00295)

I will sketch the convergence proof of a minimal elitist variant of the recently proposed Hessian Estimation Evolution Strategy (HE-ES). The main difference as compared with CMA-ES is that the covariance matrix update yields monotonic convergence of the covariance matrix to the inverse Hessian. This strong stability property allows to prove that the algorithm converges at a linear rate to the minimum of a convex quadratic objective function, where the rate is independent of the problem instance. The same holds for CMA-ES, but we are lacking a proof (since 20 years). The proof works in two steps, both of which employ drift arguments. The first step is to prove convergence of the covariance matrix, which works independent of the evolution of mean and step size. The second step is to reduce the convergence proof to recent powerful results for the (1+1)-ES without CMA.

### 3.6 Evolution Strategies Reliably Overcome Saddle Points

*Tobias Glasmachers (Ruhr-Universität Bochum, DE)*

**License** © Creative Commons BY 4.0 International license  
© Tobias Glasmachers

I will present the (to my knowledge) first result on the behavior of an ES facing saddle points. The (1+1)-ES overcomes even rather difficult saddle points, where it could be subjected to convergence prematurely because the success rate is smaller than  $1/5$ , and in fact arbitrarily close to zero.

### 3.7 Introductory Talk on the Theory of Continuous Evolutionary Algorithms

*Tobias Glasmachers (Ruhr-Universität Bochum, DE)*

**License** © Creative Commons BY 4.0 International license  
© Tobias Glasmachers

I will introduce basic problems (in particular convex quadratic functions) and algorithms (the classic (1+1)-ES and a variant with covariance matrix adaptation (CMA)). I will describe the qualitative behavior of evolution strategies with and without CMA on problems with good and bad conditioning. Some aspects of the behavior are well described by theoretical analysis, while others are open problems. I’ll then give a gist of existing analysis methodologies: Markov chains (and a recent way of proving stability), drift, and the IGO framework.

### 3.8 Theoretical Aspects of Set-Quality Indicators for Multiobjective Optimization

*Andreia P. Guerreiro (INESC-ID – Lisboa, PT)*

**License** © Creative Commons BY 4.0 International license  
© Andreia P. Guerreiro

Set-quality indicators, which map a point set into a scalar value, are a convenient way to assess (the image of) solution sets in multiobjective optimization. Such indicators may comprise in this scalar value the proximity of the set of points to the Pareto front, as well as information regarding the distribution of points in the set. Performance assessment through quality indicators can be viewed as a transformation of the multiobjective optimization problem into a single-objective one, where the goal is to find a point set, frequently bounded in size, that maximizes the quality indicator. Consequently, each indicator is biased towards some point sets. The study of the theoretical properties of quality indicators allows to characterize the indicator-optimal subsets and, therefore, to understand such biases and their implications in performance assessment and in indicator-based evolutionary multiobjective optimization algorithms. Such theoretical aspects will be discussed in this talk.

### 3.9 Black-Box Permutation Problems and weighted medians

*Ekhine Irurozki (Telecom Paris, FR)*

**License** © Creative Commons BY 4.0 International license  
© Ekhine Irurozki

**Main reference** Ekhine Irurozki, Manuel López-Ibáñez: “Unbalanced mallows models for optimizing expensive black-box permutation problems”, in Proc. of the GECCO '21: Genetic and Evolutionary Computation Conference, Lille, France, July 10-14, 2021, pp. 225–233, ACM, 2021.

**URL** <https://doi.org/10.1145/3449639.3459366>

From the paper: Unbalanced Mallows Models for Optimizing Expensive Black-Box Permutation Problems

Expensive black-box combinatorial optimization problems arise in practice when the objective function is evaluated by means of a simulator or a real-world experiment. Since each fitness evaluation is expensive in terms of time or resources, the number of possible evaluations is typically several orders of magnitude smaller than in non-expensive problems. Classical optimization methods are not useful in this scenario. In this talk, we propose and analyze UMM, an estimation-of-distribution (EDA) algorithm based on a Mallows probabilistic model and unbalanced rank aggregation (uBorda). UMM is based on a weighted median for permutations. The core of it is uBorda. Experimental results on black-box versions of LOP and PFSP show that UMM outperforms the solutions obtained by CEGO, a Bayesian optimization algorithm for combinatorial optimization. Nevertheless, a slight modification to CEGO, based on the different interpretations for rankings and orderings, significantly improves its performance, thus producing solutions that are slightly better than those of UMM and dramatically better than the original version. Another benefit of UMM is that its computational complexity increases linearly with both the number of function evaluations and the permutation size, which results in computation times an order of magnitude shorter than CEGO, making it specially useful when both computation time and number of evaluations are limited.

### 3.10 Theory of Discrete Randomized Optimization Heuristics – The What, Why, and How

Martin S. Krejca (Sorbonne University – Paris, FR)

License  Creative Commons BY 4.0 International license  
© Martin S. Krejca

Randomized optimization heuristics (ROHs) are algorithms applied to optimization problems where the objective function is only indirectly accessible, that is, it can only be accessed by evaluating solution candidates. Guided by the quality of such candidates, ROHs aim to iteratively generate solutions of better quality. This raises natural questions such as how quickly an ROH improves its solutions, whether it is capable of finding optimal solutions, or if there are any approaches that generate better solutions more quickly. Theoretical analyses aim to answer these questions and more.

In this talk, we provide an introduction to theoretical analyses on ROHs in the *discrete* domain. We introduce the setting that ROHs are applied in, the so-called *black-box* setting, and we discuss how theoretical results aim to answer questions about the real-world application of ROHs. To this end, we introduce common theory benchmark functions, ROHs, as well as the mathematical tools used for their analyses. We conclude by deriving a standard run time result, illustrating how problems in this domain are typically approached.

### 3.11 Failure on Easy Problems

Johannes Lengler (ETH Zürich, CH), Benjamin Doerr (Ecole Polytechnique – Palaiseau, FR), Carola Doerr (Sorbonne University – Paris, FR), and Dirk Sudholt (Universität Passau, DE)

License  Creative Commons BY 4.0 International license  
© Johannes Lengler, Benjamin Doerr, Carola Doerr, and Dirk Sudholt

I will discuss several seemingly easy situations in which evolutionary and genetic algorithms can fail. Some of them are rather surprising, since the hardest regions are not always close to the optimum. In particular, I will mention the detrimental effects of large mutation rates or of large population size for optimizing monotone functions, introduce a simple dynamic setting, and discuss how the self-adapting  $(1, \lambda)$ -EA can fail on OneMax.

#### References

- 1 Benjamin Doerr, Thomas Jansen, Dirk Sudholt, Carola Winzen and Christine Zarges. *Mutation rate matters even when optimizing monotonic functions*. Evolutionary computation, 21(1), 2013, 1-27.
- 2 Johannes Lengler. *A general dichotomy of evolutionary algorithms on monotone functions*. IEEE Transactions on Evolutionary Computation, 24(6), 995–1009, 2019, IEEE
- 3 Johannes Lengler and Jonas Meier. *Large population sizes and crossover help in dynamic environments*. In: International Conference on Parallel Problem Solving from Nature. Springer, Cham, 2020. S. 610-622.
- 4 Mario Alejandro Hevia Fajardo and Dirk Sudholt. *Self-adjusting population sizes for non-elitist evolutionary algorithms: why success rates matter*. In: Proceedings of the Genetic and Evolutionary Computation Conference. 2021. S. 1151-1159.
- 5 Johannes Lengler and Xun Zou. *Exponential slowdown for larger populations: The  $(\mu+1)$ -EA on monotone functions*. Theoretical Computer Science 875 (2021): 28-51.

- 6 Johannes Lengler and Simone Riedi. *Runtime Analysis of the  $(\mu + 1)$ -EA on the Dynamic BinVal Function*. In: European Conference on Evolutionary Computation in Combinatorial Optimization (Part of EvoStar). Springer, Cham, 2021. S. 84-99.

### 3.12 Population Diversity Makes Lexicase Selection Fast

Johannes Lengler (ETH Zürich, CH)

License  Creative Commons BY 4.0 International license

© Johannes Lengler

Joint work of William de Casa, Thomas Helmuth, Johannes Lengler

In Genetic Programming, it is customary to have population sizes in the order of  $\mu = 100 - 1000$ . For the next generation,  $\lambda = 2\mu$  parents are independently selected. In order to select a parent, lexicase selection first removes duplicates in performance space, i.e., individuals which show the same performance on all test cases. Then it picks one test case at random, and removes all candidates which fail on this test case (unless all candidates fail, in which case the test is skipped). This is iterated until only one candidate remains. My two co-authors report that lexicase selection is regarded critically in their community due to its very bad worst-case runtime, but that the runtime in practice is much faster. We investigated why the runtime is fast and found a measure for population diversity such that i) empirically population diversity is large, and ii) a large population diversity guarantees theoretically fast runtimes.

### 3.13 Gray-box operator for Vertex Cover

Xiaoyue Li (Hasso-Plattner-Institut, Universität Potsdam, DE)

License  Creative Commons BY 4.0 International license

© Xiaoyue Li

Joint work of Samuel Baguley, Tobias Friedrich, Timo Kötzing, Xiaoyue Li, Marcus Pappik, Ziena Zeif

Main reference Samuel Baguley, Tobias Friedrich, Timo Kötzing, Xiaoyue Li, Marcus Pappik, Ziena Zeif: “Analysis of a gray-box operator for vertex cover”, in Proc. of the GECCO ’22: Genetic and Evolutionary Computation Conference, Boston, Massachusetts, USA, July 9 – 13, 2022, pp. 1363-1371, ACM, 2022.

URL <https://doi.org/10.1145/3512290.3528848>

In this flash talk, a gray-box operator tailored for combinatorial optimization problems is presented. The operator is called balanced flip EA, which the mutation operator has been introduced to the specific problem so that the algorithm is extended for better behavior. By applying runtime analysis of both the  $(1+1)$  EA and the balanced EA, a tighter upper bound is provided for balanced flip EA as  $\mathcal{O}(\sqrt[3]{n})$  while the  $(1+1)$  EA is as  $\mathcal{O}(\sqrt{n})$ . Other than the introduction of analysis strategy, the experimental evidence for long bad path to support the partial proof of the lower bound as well as the probability bound for complete bipartite graph has also been discussed. As a result, the flash talk presented the key point that the benefits of gray-box operator should be considered. Especially when it comes to solve the specific problem like combinatorial optimization.

### 3.14 Randomized Smoothing for Non-Convex Optimization

Sebastian U. Stich (CISPA – Saarbrücken, DE)

License © Creative Commons BY 4.0 International license  
© Sebastian U. Stich

Joint work of Ahmad Ajalloeian, Harsh Vardhan, Sebastian Urban Stich

We identify a class of nonconvex functions for which we can show that perturbed gradient descent converges to a global minimum, in contrast to gradient descent without noise that can get stuck in local minima far from a global solution.

We give a brief overview of the used techniques, such as convergence analysis of (stochastic) gradient methods, biased gradients, and randomized smoothing.

Based on joint work with A. Ajalloeian [1] and H. Vardhan [2].

#### References

- 1 A. Ajalloeian and S.U. Stich *On the Convergence of SGD with Biased Gradients*. presented at ICML 2020 Workshop on *Beyond First Order Methods in ML Systems*, arXiv preprint arXiv:2008.00051, 2021.
- 2 H. Vardhan and S.U. Stich *Tackling benign nonconvexity with smoothing and stochastic gradients*. presented at NeurIPS 2021 Workshop on *Optimization for Machine Learning*, arXiv preprint arXiv:2202.0905, 2022.

### 3.15 Analyzing the Cost of Randomness in Evolutionary Algorithms

Dirk Sudholt (Universität Passau, DE) and Carlo Kneißl

License © Creative Commons BY 4.0 International license  
© Dirk Sudholt and Carlo Kneißl

Evolutionary algorithms make countless random decisions during selection, mutation and crossover operations. These random decisions require a steady stream of random numbers, however generating good quality randomness is non-trivial.

We consider the expected number of random bits used throughout a run of an evolutionary algorithm and refer to this as the cost of randomness. We give general bounds on the cost of randomness for mutation-based evolutionary algorithms using 1-bit flips or standard bit mutations using either a naive or a common, more efficient implementation that uses  $\Theta(\log n)$  random bits per mutation. Uniform crossover is a potentially wasteful operator as the number of random bits used equals the Hamming distance of the two parents, which can be up to  $n$ . However, we show for a  $(2+1)$  GA that optimizes OneMax in expected  $\Theta(n \log n)$  evaluations that the total cost of randomness during all crossover operations on OneMax is only  $\Theta(n)$ .

We hope to show that the cost of randomness may be useful as an additional performance measure, to give new insights into search dynamics and to aid in the design of operators that use randomness more carefully.

### 3.16 On the dynamics of the DE algorithm

*Ricardo Takahashi (Federal University of Minas Gerais-Belo Horizonte, BR)*

License  Creative Commons BY 4.0 International license  
© Ricardo Takahashi

This talk presents an ongoing work that aims to develop an analytical study of the Differential Evolution (DE) algorithm behavior. Analytical formulae for the probability of enhancement of and individual in populations of the DE/rand/1/bin and DE/rand/1/exp algorithm versions are developed for the sphere objective function. It is shown that those formulae can be adapted for the study of the algorithm behavior in the optimization of quadratic functions with different relations between the minimal and maximal eigenvalues. In the case of large differences of eigenvalue magnitudes, it is shown that the convergence occurs in different scales, and the formulae approximately hold for the corresponding scale dimension. The known effect of performance degradation as the problem dimension increases is partly explained by the decrease in the probability of enhancement of individuals, as indicated in the formulae. Experimental results show that DE/best algorithm versions, as expected, are able to solve problems of higher dimensions. Further research will be performed in order to examine the convergence dependency with the algorithm parameters.

### 3.17 Mathematical models for Dominance Move: Comparisons and complexity analysis

*Elizabeth Wanner (CEFET – Belo Horizonte, BR)*

License  Creative Commons BY 4.0 International license  
© Elizabeth Wanner

Dominance move (DoM), a binary quality indicator, can be used in multi-objective and many-objective optimization to compare two solution sets. DoM is very intuitive but hard to calculate due to its combinatorial nature. Different mathematical models are presented and analyzed. A computationally fast approximate approach is also discussed. Computational results are promising and an upper bound analysis for the approximation ratio would be useful.

## 4 Working groups

### 4.1 $k$ -adaptive black-box complexity

*Carola Doerr (Sorbonne University – Paris, FR) and Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Carola Doerr and Johannes Lengler

In  $k$ -adaptive black-box complexity, algorithms are only allowed to perform  $k$  iterations of queries through which the optimal solution needs to be learned. In each iteration, the algorithms are allowed to perform an arbitrary number of queries. We are interested in the minimal total number of queries that a  $k$ -adaptive black-box optimization algorithm needs

to perform in order to find an optimal solution. The  $k$ -adaptive black-box complexity models thus interpolate between non-adaptive query complexity ( $k = 1$ ) and fully adaptive query complexity (no restriction on  $k$ ).

We discussed the  $k$ -adaptive black box complexity (BBC) for several benchmark problems, starting with OneMax and Mastermind. For OneMax, the query complexity does not substantially increase (only by a constant factor) even if we restrict to  $k = 1$ . For general Mastermind, this is unclear and poses an interesting research question. We took some first steps in discussing the  $k = 2$  case with  $n$  colors and positions.

For LeadingOnes, the situation is almost opposite than for OneMax: if  $k = n/\alpha$ , then we can find an algorithm with  $2^\alpha n/\alpha$  queries, and we believe that the query complexity is exponentially large in  $\alpha$ , perhaps even  $\Omega(2^\alpha n/\alpha)$ . We spent some time discussing the differences to Permutation-LeadingOnes: it is at least as hard as LeadingOnes, so the lower bounds still apply. Analyzing previous work of Benjamin Doerr and Carola Doerr [1], we have an algorithm that is  $n$ -adaptive and has complexity  $O(n \log n / \log \log n)$ . Perhaps this could be improved to  $O(n \log \log n)$ , but the general picture is that adaptivity needs to be very high to avoid an exponentially high price in terms of complexity.

We finally discussed the situation for HiddenSubset, which seems to be an example for a problem that requires an intermediate level of adaptivity. Known algorithms are  $(\log n)$ -adaptive, with optimal runtime  $O(n \log n)$ . It might be interesting to study smaller  $k$ , and we conjecture that this would increase the complexity substantially.

## References

- 1 Benjamin Doerr and Carola Winzen *Black-Box Complexity: Breaking the  $O(n \log n)$  Barrier of LeadingOnes*. presented at Artificial Evolution – 10th International Conference, Evolution Artificielle, EA 2011, Angers, France, October 24-26, 2011, Revised Selected Papers, 205–216, [https://doi.org/10.1007/978-3-642-35533-2\\_18](https://doi.org/10.1007/978-3-642-35533-2_18), 2011.
- 2 Jin-Kao Hao and Pierrick Legrand and Pierre Collet and Nicolas Monmarché and Evelyne Lutton and Marc Schoenauer, *Artificial Evolution – 10th International Conference, Evolution Artificielle, EA 2011, Angers, France, October 24-26, 2011, Revised Selected Papers, Lecture Notes in Computer Science, 7401, 2012.*

## 4.2 Theory-Friendly Practical Modelling of Combinatorial Optimisation Problems for ROHs

Carlos M. Fonseca (University of Coimbra, PT)

License © Creative Commons BY 4.0 International license  
© Carlos M. Fonseca

Problem modelling is often overlooked as the crucial step preceding the practical application of randomised optimisation heuristics (ROHs). In order to make ROHs more accessible in the real world, there is a need to specify how optimisation algorithms interface with the problem instances of interest. In addition, guidance should be provided to practitioners on how problem-specific information can be exposed to the algorithms through such an interface. In other words, such an interface definition should support an associated problem modelling methodology.

From a different perspective, such an interface specification must also support the development of practical algorithms that can be applied directly to any given problem implementing that interface. A brief presentation of two Application Programming Interfaces

(APIs) for combinatorial optimisation under development, and of the design principles behind them, was the starting point for a discussion on the potential of such an approach and how theory-friendly it might be.

### 4.3 Adaptation of proof techniques based on the natural SGD interpretation of CMA-ES

*Tobias Glasmachers (Ruhr-Universität Bochum, DE)*

License  Creative Commons BY 4.0 International license  
© Tobias Glasmachers

As a minimal example, we went through IGO framework and NES algorithm, using 1D example  $f(x) = x^2$  and Gaussians. We discussed the quantile rescaling technique in detail, and whether it may or may not be understood as a static fitness transformation. We also discussed the role of rescaling the gradient with the Fisher information matrix for achieving linear convergence. We identified potential connections to variance-reduced gradient descent techniques, as well as to the sampling literature. The general impression was that the natural SGD setting with quantile rescaling is too far from the standard analysis of SGD to allow for a straightforward transfer of proof techniques.

### 4.4 Lyapunov potential for the linear convergence of CMA-ES

*Tobias Glasmachers (Ruhr-Universität Bochum, DE)*

License  Creative Commons BY 4.0 International license  
© Tobias Glasmachers

We started with a “naive” potential, which is a straightforward extension of the (1+1)-ES potential. We then focused on Armand’s problem and understood that it is a model of the initial phase of optimizing the discus problem. We tried to fix the  $\log(\|m\|)$  term to achieve additive drift in this situation, without much success. Therefore, we shifted the discussion to how a multi-stage drift for that situation may work.

### 4.5 Comma and plus selection strategies with noise

*Johannes Lengler (ETH Zürich, CH)*

License  Creative Commons BY 4.0 International license  
© Johannes Lengler

**Main reference** Benjamin Doerr: “Does Comma Selection Help to Cope with Local Optima?”, *Algorithmica*, Vol. 84(6), pp. 1659–1693, 2022.

**URL** <https://doi.org/10.1007/s00453-021-00896-7>

We started gathering examples of problems on which comma selection performs differently to plus selection, focusing on theoretically proven examples. Although the topic was originally set to deal with noise, we considered deterministic problems for the most part.

It is known that comma selection in a  $(\mu, \lambda)$  EA does not perform better than plus selection on JUMP functions, and this holds for arbitrary population sizes  $\mu$  and  $\lambda$ . On the CLIFF function, comma strategies perform well, enabling a  $(1, \lambda)$  EA to jump down the cliff if

all offspring are down the cliff and the algorithm moves towards the global optimum without jumping back up the cliff. For optimal fixed  $\lambda$ , the best known expected runtime is essentially  $n^{3.97\dots}$  (up to sub-polynomial factors). Self-adjusting  $\lambda$  can reduce this to  $O(n \log n)$  if a reset mechanism is used that resets  $\lambda$  to 1 if it exceeds a given threshold. The effect is similar to using hyperheuristics or ageing where occasionally non-elitist steps are accepted. The structure of the function is quite benign as the gradient past the cliff points towards the global optimum.

It is also known that for large values of  $\lambda$ , a comma strategy will behave like a plus strategy because there is a high probability of cloning the parent. If  $\lambda$  is too small,  $\lambda \leq (1 - \varepsilon) \log_{\frac{e}{e-1}} n$ , optimising any fitness function with a unique optimum takes exponential time. So there is a narrow region for values of  $\lambda$  where one can see an advantage of comma selection over plus selection.

We also discussed elitist versus non-elitist algorithms more generally (e.g. tournament selection, self-adaptation and island models). For all algorithms discussed, the crucial issue came down to balancing exploration and exploitation: being able to escape from local optima while also being able to climb hills.

We came to the conclusion that there is a gap between theory and practice as we're lacking convincing examples (apart from CLIFF) where comma selection provably helps, whereas in practice comma strategies seem to be quite popular to escape from local optima. Part of this gap might be caused by theory traditionally aiming to find the exact global optimum, whereas in practice one is usually content with a good approximation of the optimal fitness. Using a fixed-target perspective for less ambitious targets might give a different picture as much smaller values of  $\lambda$  might be sufficient.

## 5 Open problems

### 5.1 A problem where CMA-ES performs poorly

*Nikolaus Hansen (INRIA Saclay – Palaiseau, FR)*

License  Creative Commons BY 4.0 International license  
© Nikolaus Hansen

We investigate a 6-dimensional curve fitting problem, where CMA-ES takes a long time to approach the global optimum. The problem is simple and visually intuitive. Parameters move very slowly in the same direction towards the optimum for a long time, in which case we would expect the covariance matrix to speed up the movements. However, the sample distribution appears to be stable with a condition number of about  $10^6$ . The leading hypothesis for the reason is that the problem has a very narrow but slightly bent ridge. The bend prevents a faster approach to the optimum.

## Participants

- Mickaël Binois  
INRIA – Sophia Antipolis, FR
- Alexandre Chotard  
Calais University, FR
- Duc-Cuong Dang  
University of Southampton, GB
- Benjamin Doerr  
Ecole Polytechnique –  
Palaiseau, FR
- Carola Doerr  
Sorbonne University – Paris, FR
- Carlos M. Fonseca  
University of Coimbra, PT
- Armand Gissler  
Ecole Polytechnique –  
Palaiseau, FR
- Tobias Glasmachers  
Ruhr-Universität Bochum, DE
- Andreia P. Guerreiro  
INESC-ID – Lisboa, PT
- Nikolaus Hansen  
INRIA Saclay –  
Palaiseau, FR
- Ekhnine Irurozki  
Telecom Paris, FR
- Martin S. Krejca  
Sorbonne University – Paris, FR
- Johannes Lengler  
ETH Zürich, CH
- Xiaoyue Li  
Hasso-Plattner-Institut,  
Universität Potsdam, DE
- Peter Richtarik  
KAUST – Thuwal, SA
- Günter Rudolph  
TU Dortmund, DE
- Sebastian U. Stich  
CISPA – Saarbrücken, DE
- Dirk Sudholt  
Universität Passau, DE
- Andrew M. Sutton  
University of Minnesota –  
Duluth, US
- Ricardo Takahashi  
Federal University of Minas  
Gerais-Belo Horizonte, BR
- Elizabeth Wanner  
CEFET – Belo Horizonte, BR

