Report from Dagstuhl Seminar 14372

Analysis of Algorithms Beyond the Worst Case

Edited by

Maria-Florina Balcan[1], Bodo Manthey[2], Heiko Röglin[3], and Tim Roughgarden[4]

1 **Carnegie Mellon University, US,** `ninamf@cs.cmu.edu`
2 **University of Twente, NL,** `b.manthey@utwente.nl`
3 **Universität Bonn, DE,** `roeglin@cs.uni-bonn.de`
4 **Stanford University, US,** `tim@cs.stanford.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 14372 "Analysis of Algorithms Beyond the Worst Case".

The theory of algorithms has traditionally focused on worst-case analysis. This focus has led to both a deep theory and many beautiful and useful algorithms. However, there are a number of important problems and algorithms for which worst-case analysis does not provide useful or empirically accurate results. This is due to the fact that worst-case inputs are often rather contrived and occur hardly ever in practical applications. Only in recent years a paradigm shift towards a more realistic and robust algorithmic theory has been initiated. The development of a more realistic theory hinges on finding models that measure the performance of an algorithm not only by its worst-case behavior but rather by its behavior on "typical" inputs. In this seminar, we discussed various recent theoretical models and results that go beyond worst-case analysis.

The seminar helped to consolidate the research and to foster collaborations among the researchers working in the different branches of analysis of algorithms beyond the worst case.

Seminar September 7–12, 2014 – http://www.dagstuhl.de/14372
1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity, G.1.6 Optimization, G.2 Discrete Mathematics, G.3 Probability and Statistics, I.2.6 Learning, I.5.3 Clustering
Keywords and phrases analysis of algorithms, probabilistic analysis, smoothed analysis, approximation stability, machine learning
Digital Object Identifier 10.4230/DagRep.4.9.30
Edited in cooperation with Michael Etscheid

1 Executive Summary

Maria-Florina Balcan
Bodo Manthey
Heiko Röglin
Tim Roughgarden

The theory of algorithms has traditionally focused on worst-case analysis. This focus has led to both a deep theory and many beautiful and useful algorithms. There are, however, a number of important problems and algorithms for which worst-case analysis does not provide useful or empirically accurate results. For example, worst-case analysis suggests that

Analysis of Algorithms Beyond the Worst Case, *Dagstuhl Reports*, Vol. 4, Issue 9, pp. 30–49
Editors: Maria-Florina Balcan, Bodo Manthey, Heiko Röglin, and Tim Roughgarden
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

the simplex method is an exponential-time algorithm for linear programming, while in fact it runs in near-linear time on almost all inputs of interest. Worst-case analysis ranks all deterministic caching algorithms equally, while in almost all applications some algorithms (like least-recently-used) are consistently superior to others (like first-in-first-out).

The problem is that worst-case analysis does not take into consideration that worst-case inputs are often rather contrived and occur hardly ever in practical applications. It led to the situation that for many problems the classical theory is not able to classify algorithms meaningfully according to their performance. Even worse, for some important problems it recommends algorithms that perform badly in practice over algorithms that work well in practice only because the artificial worst-case performance of the latter ones is bad.

Only in recent years a paradigm shift towards a more realistic and robust algorithmic theory has been initiated. The development of a more realistic theory hinges on finding models that measure the performance of an algorithm not only by its worst-case behavior but rather by its behavior on typical inputs. However, for an optimization problem at hand it is usually impossible to rigorously define the notion of "typical input" because what such an input looks like depends on the concrete application and on other indistinct parameters. The key to building a rigorous and realistic theory is hence not to define exactly what a typical input looks like, but to identify common properties that are shared by real-world inputs. As soon as such properties are identified, in many cases one can explain why certain heuristics work well in practice while others do not. The next step is then to look for algorithmic means to exploit these properties explicitly in order to obtain improved algorithms in practice that are not tailored to unrealistic worst-case inputs.

Many different models that go beyond classical worst-case theory have been suggested. These models can be divided into two main categories: either they are based on the assumption that inputs are to some extend random (probabilistic analysis) or they consider only inputs that satisfy certain deterministic properties that mitigate the worst case.

Probabilistic Analysis

Average-case analysis is probably the first thought that springs to mind when mentioning probabilistic input models. In such an analysis, one considers the expected performance on random inputs. Starting in the seventies, many algorithms that showed a remarkable performance in practice have been analyzed successfully on random inputs. This includes algorithms for classical optimization problems, such as the traveling salesman problem, or the simplex method for linear programming.

While average-case analysis has been successfully applied to many problems, a major concern is that inputs chosen completely at random have for many problems little in common with inputs arising in practice. Similar as a random TV screen produced by static noise has nothing to do with a typical TV screen, a set of random points does not resemble, for instance, a realistic clustering instance with mostly well-separated clusters.

Smoothed Analysis. To overcome the drawbacks of average-case and worst-case analysis, the notion of smoothed analysis has been suggested by Spielman and Teng in 2001. In this model, inputs are generated in two steps: first, an adversary chooses an arbitrary instance, and then this instance is slightly perturbed at random. The smoothed performance of an algorithm is defined to be the worst expected performance the adversary can achieve. This model can be viewed as a less pessimistic worst-case analysis, in which the randomness rules out pathological worst-case instances that are rarely observed in practice but dominate the worst-case analysis. If the smoothed running time of an algorithm is low (i. e., the algorithm is efficient in expectation on any perturbed instance) and inputs are subject to a small amount of random noise, then it is unlikely to encounter an instance on which the algorithm performs poorly. In practice, random noise can stem from measurement errors, numerical imprecision or rounding errors. It can also model arbitrary influences, which we cannot quantify exactly, but for which there is also no reason to believe that they are adversarial.

The framework of smoothed analysis was originally invented to explain the practical success of the simplex method for linear programming. Spielman and Teng analyzed linear programs in which each coefficient is perturbed by Gaussian noise with standard deviation σ. They showed that the smoothed running time of the simplex method is bounded polynomially in the input size and $1/\sigma$. Hence, even if the amount of randomness is small, the expected running time of the simplex method is polynomially bounded. After its invention smoothed analysis has attracted a great deal of attention and it has been applied in a variety of different contexts, e. g., in multi-objective optimization, local search, clustering, and online algorithms. By now smoothed analysis is widely accepted as a realistic alternative to worst-case analysis.

Semi-random Models. Semi-random input models can be considered as analogues of smoothed analysis for graph problems and they even predate smoothed analysis by a couple of years. There is a variety of semi-random graph models that go beyond the classical Erdős-Rényi random graphs. In most of these models graphs are generated by a noisy adversary – an adversary whose decisions (whether or not to insert a particular edge) have some small probability of being reversed. Another well-studied class of semi-random models are *planted models*, in which a solution (e. g., an independent set or a partitioning of the vertices in color classes) is chosen and then edges are added randomly or by an adversary of limited power in such a way that the given solution stays a valid solution for the given problem.

Similar to smoothed analysis, semi-random models have been invented in order to better understand the complexity of NP-hard graph problems because Erdős-Rényi random graphs often do not reflect the instances one encounters in practice – many graph problems are quite easy to solve on such random graphs.

Deterministic Input Models

Smoothed analysis and semi-random models are multi-purpose frameworks that do not require much information about how exactly typical inputs for the optimization problem at hand look like. If more information is available, it makes sense to identify structural properties of typical inputs that make them easier to solve than general inputs. There are well known examples of this approach like the TSP, which gets easier (in terms of approximation) when restricted to inputs in which the distances satisfy the triangle inequality. Also in computational geometry it is a very common phenomenon that problems become easier if one assumes that no angles are too small or not too many objects overlap in the same region.

In recent years there has been an increased interest in more sophisticated deterministic input models, in particular for clustering problems. Balcan, Blum, and Gupta introduce and exploit the so-called $(1+\alpha,\varepsilon)$-approximation-stability property of data in the context of clustering. This assumption is motivated by the observation that in many clustering applications there is usually one correct but unknown target clustering, and the goal is to find a clustering that is close to this target clustering and misclassifies only a few objects. On the other hand in the common mathematical formulation of clustering problems a potential function is defined that assigns a value to each clustering. Then a clustering is computed that approximately optimizes the potential function (exact optimization is usually NP-hard). This approach makes sense only if clusterings that approximately optimize the potential function are close to the target clustering. Hence, an implicit assumption underlying this approach is that every clustering that approximately optimizes the objective function is close to the desired target clustering. Balcan et al. made this assumption explicit: they define that a clustering instance satisfies the $(1+\alpha,\varepsilon)$-approximation-stability assumption if in every c-approximation of the potential function at most an ε-fraction of all objects is misclassified compared to the target clustering. Balcan et al. showed that clustering instances with this property are easier to solve than general instances. They have shown specifically how to get ε-close to the target even for values of α for which finding a $1+\alpha$ approximation is NP-hard. Voevodoski et al. have shown that this approach leads to very efficient and accurate algorithms (with improved performance over previous state-of-the-art algorithms) for clustering biological datasets.

Bilu and Linial and later Awasthi, Blum, and Sheffet have considered instances of clustering problems that are *perturbation resilient* in the sense that small perturbations of the metric space do not change the optimal solution. They argue that interesting instances of clustering problems are stable and they prove that the assumption of stability renders clustering polynomial-time solvable. Balcan and Liang further relaxed this assumption to require only that the optimal solution after the perturbations is close to the optimal solution for the unperturbed instance.

These results have triggered a significant amount of work in the past years in the context of clustering and machine learning problems more generally, including subsequent works that proposed new related stability conditions (e. g., the "proximity condition" by Kannan and Kumar). Such works are very good examples demonstrating that identifying properties of real-world inputs can be extremely beneficial for our understanding of algorithmic problems.

Program of the Seminar

The program of the seminar consisted of 23 talks, including the following survey talks:

- Preprocessing of NP-hard problems, Uriel Feige;
- Approximation-stability and Perturbation-stability, Avrim Blum;
- Computational Feasibility of Clustering under Clusterability Assumptions, Shai Ben-David;
- Parametrizing the easiness of machine learning problems, Sanjoy Dasgupta;
- Linear Algebra++: Adventures and Unsupervised Learning, Sanjeev Arora.

The rest of the talks were 30-minute presentations on recent research of the participants. The time between lunch and the afternoon coffee break was mostly left open for individual discussions and collaborations in small groups. One open-problem session was organized.

One of the main goals of the seminar was to foster collaborations among the researchers working in the different branches of analysis of algorithms as sketched above. This is particularly important because at the moment the two communities dealing with probabilistic analysis and deterministic input models are largely disjoint. The feedback provided by the participants shows that the goals of the seminar, namely to circulate new ideas and create new collaborations, were met to a large extent.

The organizers and participants wish to thank the staff and the management of Schloss Dagstuhl for their assistance and support in the arrangement of a very successful and productive event.

2 Table of Contents

Executive Summary

Maria-Florina Balcan, Bodo Manthey, Heiko Röglin, and Tim Roughgarden 30

Overview of Talks

New Measures and Techniques in the Analysis of Online Search Problems
Spyros Angelopoulos 36

Adventures in Linear Algebra++ and Unsupervised Learning
Sanjeev Arora 36

Learning Submodular Functions: An Analysis beyond the Worst Case
Maria-Florina Balcan 36

Computational Efficiency of Clustering Well-Behaved Input Instances
Shai Ben-David 37

Approximation and Perturbation Stability
Avrim Blum 39

Fréchet Distance: Synergies of Algorithms and Lower Bounds
Karl Bringmann 39

Smoothed Analysis of the Successive Shortest Path and Minimum-Mean Cycle Canceling Algorithms
Kamiel Cornelissen 40

The (Parameterized) Complexity of Counting Subgraphs
Radu Curticapean 40

Parametrizing the Easiness of Machine Learning Problems
Sanjoy Dasgupta 41

How to Serve Impatient Users
Matthias Englert 42

Smoothed Analysis of Local Search for the Maximum-Cut Problem
Michael Etscheid 42

Preprocessing of NP-hard Problems
Uriel Feige 43

Average-Case Analysis of Parameterized Problems
Tobias Friedrich 43

On the Smoothed Approximation Performance of the 2-OPT Heuristic
Marvin Künnemann 44

Constant Factor Approximation for Balanced Cut in the PIE Model
Konstantin Makarychev 44

Bilu-Linial Stable Instances of Max Cut and Minimum Multiway Cut
Yury Makarychev 45

Smoothed Analysis of Local Search
Bodo Manthey 45

Analyzing Non-Convex Optimization for Dictionary Learning
Tengyu Ma 46

Efficient Algorithms Beyond Worst-Case Combinatorial Bounds
Matthias Mnich 46

Smoothed Analysis of Multi-Objective Optimization
Heiko Röglin 46

Decompositions of Triangle-Dense Graphs
Tim Roughgarden 47

Utilitarian View of Rankings
Or Sheffet 47

Probabilistic Lipschitzness: A Measure of Data Niceness
Ruth Urner 48

Participants 49

3 Overview of Talks

3.1 New Measures and Techniques in the Analysis of Online Search Problems

Spyros Angelopoulos (UPMC – Paris, FR)

Joint work of Angelopoulos, Spyros; Durr, Christoph; Lopez-Ortiz, Alejandro; Lucarelli, Giorgio; Panagiotou, Konstantinos

We consider the problem of exploring a set of concurrent rays using a single searcher. The rays are disjoint with the exception of a single common point, and in each ray a potential target may be located. The objective is to design efficient search strategies for locating the targets as quickly as possible.

In this talk we will describe some recent measures for the analysis of online search problems and in particular for the star-search problem. The new measures transcend the traditional worst-case analysis and give rise to new algorithmic approaches.

3.2 Adventures in Linear Algebra++ and Unsupervised Learning

Sanjeev Arora (Princeton University, US)

Linear Algebra++ is my name for some extensions to Linear Algebra problems (solving linear equations, rank, finding better basis to represent the data, eigenvalues/eigenvectors) that involve constraints such as sparsity, nonnegativity, and approximation (via ℓ_2 or another norm).

These variants arise in machine learning settings and are often NP-hard, i. e., difficult on worst-case inputs. Hence the only way to design provable algorithms for them is to go beyond worst case analysis.

The talk surveyed some recent work of mine and others that does this, with applications to ML problems such as topic modeling, sparse coding, nonnegative matrix factorization, deep learning, etc.

3.3 Learning Submodular Functions: An Analysis beyond the Worst Case

Maria-Florina Balcan (Carnegie Mellon University – Pittsburgh, US)

Joint work of Balcan, Maria-Florina; Harvey, Nicholas J. A.; Florin Constantin, Satoru Iwata, Lei Wang; Nan Du, Yingyu Liang, and Le Song
Main reference M.-F. Balcan, N. J. A. Harvey, "Learning Submodular Functions," in Proc. of the 43rd Annual ACM Symp. on Theory of Computing (STOC'11), pp. 793–802, ACM, 2011; pre-print available from author's webpage.
URL http://dx.doi.org/10.1145/1993636.1993741
URL http://www.cs.cmu.edu/~ninamf/papers/LearningSubmodular-STOC.pdf

Submodular functions are discrete functions that model laws of diminishing returns and enjoy numerous algorithmic applications that have been used in many areas, including social networks, machine learning, and economics. In this work we use a learning theoretic angle for studying submodular functions. We provide algorithms for learning submodular functions, as well as lower bounds on their learnability. In doing so, we uncover several novel structural results revealing both extremal properties as well as regularities of submodular functions, of interest to many areas.

Our lower bounds highlight the importance of providing an analysis on the learnability of these functions beyond the worst case. For classes of functions that exhibit additional structure in addition to diminishing returns (including probabilistic coverage functions and XOS functions with bounded complexity) we provide stronger guarantees on their learnability together with an interesting application for learning the influence function in a social network.

References

1 Maria Florina Balcan, Nick Harvey. *Learning submodular functions.* STOC 2011.

2 Maria Florina Balcan, Florin Constantin, Satoru Iwata, and Lei Wang. *Learning valuation functions.* COLT 2012.

3 Maria Florina Balcan, Nan Du, Yingyu Liang, and Le Song. *Influence Function Learning in Information Diffusion Networks.* ICML 2014.

3.4 Computational Efficiency of Clustering Well-Behaved Input Instances

Shai Ben-David (University of Waterloo, CA)

The goal of this talk is two-fold. First, I would like to provide a personally biased overview of the research concerning the computational complexity of clustering under data niceness assumptions. Having worked in this area for quite some time now, I feel that while the TCS community appreciates work that may have practical relevance (and clustering is clearly a task that arises in many applications), sometimes in this area there is a significant gap between research motivation and the actual technical results it yields. A secondary aim of this paper is to call the attention of the theoretical research community to some such gaps and encourage further work along directions that might otherwise have seemed resolved.

Computational complexity theory aims to provide tools for the quantification and analysis of the computational resources needed for algorithms to perform computational tasks. Worst-case complexity is by far the best known, most researched and best understood approach to computational complexity theory. In particular, NP-hardness is a worst-case-instance notion. By saying that a task is NP-hard (and assuming P $\neq$ NP), we imply that for every algorithm, there exist infinitely many instances on which it will have to work hard. However, for many problems this measure is unrealistically pessimistic compared to the experience of solving them for practical instances. A problem may be NP-hard and still have algorithms that solve it efficiently for any instance that is likely to occur in practice or any instance for which one cares about finding an optimal solution to.

Several approaches have been proposed to bringing computational complexity theory closer to the actual hardness faced when solving optimization problems on real data. *Average Case Complexity* ([8], [3]), analyzes run time w.r.t. some given probability distribution over the input instances. *Smoothed Analysis* ([9]) examines the running time of a given algorithm by taking the worst case over all inputs of the average runtime of the algorithm over some vicinity of the input. A different approach is to have a notion of "well-behaved-instances", so that on one hand it is reasonable to expect that instances one comes across in applications are so well behaved, and on the other hand there exist algorithms that can solve any well behaved input in polynomial time. Various earlier approaches have addressed computational hardness by defining subsets of relatively-easy instances (most notably, the area of *parameterized complexity* ([7])). [2] and [5] propose general notions of tamed instances that apply across different problems. Both of these papers apply some type of *robustness to perturbations* as the key property of such well behaved instances. Algorithms that efficiently solve NP-hard problems on such perturbation robust instances have been shown to exist for agnostic learning of half-spaces ([4]) and for graph partitioning problems ([6]). [1] formalized a *uniqueness of the optimal solution* criterion as a notion of well behaved clustering instances, which can also be applied to other types of problems. In this note I address the application of such approaches to clustering. I discuss those, as well as other notions of niceness-of-instances that are specific to clustering problems, as a basis for alternatives to worst-case complexity analysis of clustering tasks.

Clustering is a very useful paradigm that is being applied in a wide range of data exploration tasks. The term "clustering" should be thought of as an umbrella notion for a big and varied collection of tasks and algorithmic paradigms. Here, I will focus on clustering tasks that are defined as discrete optimization problems. Most of those optimization problems are NP-hard. I wish to examine whether this hardness remains an issue when we restrict our attention to "clusterable data" – data for which a meaningful clustering exists (one can argue that when there is no cluster structure in a given data set, there is no point in applying a clustering algorithm to it). In other words, we wish to evaluate to what extent current theoretical work supports the "Clustering is difficult only when it does not matter" (CDNM) thesis. A hint to my conclusion – the answer is a resounding NO.

I'll upload the full paper to arXiv soon.

References

1 Maria-Florina Balcan, Avrim Blum, and Anupam Gupta. Approximate clustering without the approximation. In *Proc. of the 20th Annual ACM-SIAM Symp. on Discrete Algorithms, SODA 2009, New York, NY, USA, January 4–6, 2009*, pages 1068–1077, 2009.

2 Shai Ben-David. Alternative measures of computational complexity with applications to agnostic learning. In *Theory and applications of models of computation*, volume 3959 of *Lecture Notes in Comput. Sci.*, pages 231–235. Springer, Berlin, 2006.

3 Shai Ben-David, Benny Chor, Oded Goldreich, and Michael Luby. On the theory of average case complexity (abstract). In *Proc. of the 4th Annual Structure in Complexity Theory Conf., University of Oregon, Eugene, Oregon, USA, June 19–22, 1989*, page 36, 1989.

4 Shai Ben-David and Hans-Ulrich Simon. Efficient learning of linear perceptrons. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000, Denver, CO, USA*, pages 189–195, 2000.

5 Yonatan Bilu and Nathan Linial. Are stable instances easy? In *Innovations in Computer Science – ICS 2010, Tsinghua University, Beijing, China, January 5–7, 2010. Proceedings*, pages 332–341, 2010.

6 Yonatan Bilu and Nathan Linial. Are stable instances easy? *Combinatorics, Probability & Computing*, 21(5):643–660, 2012.

7 R. G. Downey and M. R. Fellows. *Parameterized complexity.* Monographs in Computer Science. Springer-Verlag, New York, 1999.

8 Leonid A. Levin. Average case complete problems. *SIAM J. Comput.*, 15(1):285–286, 1986.

9 Daniel Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In *Proc. of the 33rd Annual ACM Symp. on Theory of Computing*, pages 296–305 (electronic). ACM, New York, 2001.

3.5 Approximation and Perturbation Stability

Avrim Blum (Carnegie Mellon University – Pittsburgh, US)

Joint work of Blum, Avrim; Balcan, Maria-Florina; Gupta, Anupam; Awasthi, Pranjal; Sheffet, Or; Vempala, Santosh

Often in optimization problems, the formal objective is a proxy for a goal of finding a desired "target" answer. For example, in clustering news articles by topic, one would like to find a clustering that agrees with how a human user would cluster them, or in segmenting a picture into objects, one would like to find a segmentation that agrees with the ground truth. If one encodes such a problem as (for instance) a k-means optimization problem, then implicitly one is hoping that the optimal solution to the objective corresponds to a solution that is similar to the desired target answer. This talk describes and discusses approximation-stability and perturbation-resilience, two stability notions motivated by making this and associated assumptions explicit. We present results for instances satisfying these conditions, focusing on clustering but also discussing Nash equilibria and related problems.

Nearly all of this work is joint with Nina Balcan. Major portions were also joint with Pranjal Awasthi, Anupam Gupta, Or Sheffet, and Santosh Vempala. Portions of the talk also discuss work that I was not involved in (e. g., of Nina Balcan and Bruce Liang).

3.6 Fréchet Distance: Synergies of Algorithms and Lower Bounds

Karl Bringmann (ETH Zürich, CH)

Joint work of Bringmann, Karl; Künnemann, Marvin
Main reference K. Bringmann, "Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails," in Proc. of the 55th IEEE Annual Symp. on Foundations of Computer Science (FOCS'14), pp. 661–670, IEEE CS, 2014; pre-print available as arXiv:1404.1448v2 [cs.CG].
URL http://dx.doi.org/10.1109/FOCS.2014.76
URL http://arxiv.org/abs/1404.1448v2

The Fréchet distance is a well-studied and very popular measure of similarity of two curves. Regarding worst-case complexity, all known algorithms to compute the Fréchet distance of two polygonal curves with n vertices have a runtime of $\tilde{\Theta}(n^2)$, omitting logarithmic factors. To obtain a conditional lower bound, we assume the Strong Exponential Time Hypothesis. Under this assumption we show that the Fréchet distance cannot be computed in strongly subquadratic time, i. e., in time $O(n^{2-\delta})$ for any $\delta > 0$. This means that finding faster algorithms for the Fréchet distance is as hard as finding faster algorithms for the satisfiability problem, and the existence of a strongly subquadratic algorithm can be considered unlikely.

To overcome the worst-case quadratic time barrier, restricted classes of curves have been studied that attempt to capture realistic input curves. The most popular such class are c-packed curves, for which the Fréchet distance has a $(1+\varepsilon)$-approximation in time $\tilde{O}(cn/\varepsilon)$ by Driemel, Har-Peled, and Wenk [3]. In dimension $d \geq 5$ we show that this cannot be improved to $O((cn/\sqrt{\varepsilon})^{1-\delta})$ for any $\delta > 0$ unless the Strong Exponential Time Hypothesis fails. Moreover, exploiting properties that prevent stronger lower bounds, we present an improved algorithm with runtime $\tilde{O}(cn/\sqrt{\varepsilon})$, matching our conditional lower bound.

This is partly based on joint work with Marvin Künnemann [2].

References

1 K. Bringmann. *Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails.* In: Proc. 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS'14), pp. 661–670, 2014.

2 K. Bringmann and M. Künnemann. *Improved approximation for Fréchet distance on c-packed curves matching conditional lower bounds.* Submitted. 2014. arXiv: 1408.1340.

3 A. Driemel, S. Har-Peled, and C. Wenk. *Approximating the Fréchet distance for realistic curves in near linear time.* Discrete & Computational Geometry 48.1 (2012), 95–127.

3.7 Smoothed Analysis of the Successive Shortest Path and Minimum-Mean Cycle Canceling Algorithms

Kamiel Cornelissen (University of Twente, NL)

Joint work of Brunsch, Tobias; Cornelissen, Kamiel; Manthey, Bodo; Röglin, Heiko
Main reference T. Brunsch, K. Cornelissen, B. Manthey, H. Röglin, "Smoothed Analysis of the Successive Shortest Path Algorithm," in Proc. of the 24th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA'13), pp. 1180–1189, SIAM, 2013.
URL http://dx.doi.org/10.1137/1.9781611973105.85

For minimum-cost maximum flow (MCF) algorithms, worst-case running time bounds do not always give a good indication for how the algorithms perform in practice. To model the performance on practical instances, we analyze two MCF algorithms, the successive shortest path (SSP) algorithm and the minimum-mean cycle canceling (MMCC) algorithm, in the framework of smoothed analysis. An adversary can specify the flow network and the edge capacities, but for the edge costs c_e the adversary can only specify density functions f_e of maximum density ϕ, according to which the edge costs are drawn.

We show that in the smoothed setting the SSP algorithm only needs $O(mn\phi(m+n\log n))$ running time, in stark contrast to the worst case running time. We also show almost tight lower bounds.

For the MMCC algorithm we show a smoothed running time of $O(m^2n^2(n\log n+\log\phi))$, which is better than the worst-case running time for dense graphs. We also show a lower bound of $\Omega(m^2n\log\phi)$.

3.8 The (Parameterized) Complexity of Counting Subgraphs

Radu Curticapean (Universität des Saarlandes, DE)

Joint work of Curticapean, Radu; Marx, Dániel
Main reference R. Curticapean, D. Marx, "Complexity of counting subgraphs: only the boundedness of the vertex-cover number counts," in Proc. of the 55th IEEE Annual Symp. on Foundations of Computer Science (FOCS'14), pp. 130–139, IEEE CS, 2014; pre-print available at arXiv:1407.2929v1 [cs.CC].
URL http://dx.doi.org/10.1109/FOCS.2014.22
URL http://arxiv.org/abs/1407.2929v1

We study the parameterized complexity of the following counting version of subgraph isomorphism: Given a "pattern" graph H and a "host" graph G, count the (not necessarily induced) copies of H in G, that is, subgraphs of G that are isomorphic to H.

In the first part, we follow an approach introduced by Marx and Pilipczuk, who studied the decision version of this problem under a set of 10 relevant parameters that can be imposed upon the graphs H and G, e. g., the size of the graphs, the number of connected components, maximum degree, treewidth, genus, and others. For each choice of parameters $k_1, \ldots, k_r$ and $e_1, \ldots, e_s$, they either gave an algorithm that solves subgraph isomorphism in time $f(k_1, \ldots, k_r) \cdot n^{g(e_1, \ldots, e_s)}$ for computable functions f, g – or they showed that such an algorithm would imply P = NP or FPT = W[1].

We are currently transferring this framework to the counting world and ask which parameter combinations admit such algorithms for the problem of counting subgraphs. Note that the set of algorithms can only shrink as counting is always harder than deciding. While this is still work in progress, we have almost obtained a dichotomy for the problem. In particular, we have identified some cases when counting is indeed harder than deciding.

In the second part, we consider only the parameter $|H|$ and ask for $f(|H|) \cdot n^{O(1)}$ algorithms for counting subgraphs. However, in this second problem, the pattern H may only be drawn from a fixed class C of graphs, which gives rise to the problem #Sub(C). There are very simple pattern classes C for which #Sub(C) is already known to be #W[1]-complete, such as the classes of paths, matchings or cycles. On the class of stars however, #Sub(C) actually even admits an $n^O(1)$ algorithm (in particular, this algorithm does not exploit the parameter $|H|$). We want to know for which classes #Sub(C) is fixed-parameter tractable (or even polynomial-time solvable).

It turns out that the only restriction that makes #Sub(C) easy is a bound on the maximum matching size in C. If C admits such a bound c, then #Sub(C) even admits a polynomial-time algorithm with an exponent depending only on c. If C however has unbounded matching size, then #Sub(C) is #W[1]-complete. In particular, we obtain no classes C for which #Sub(C) is FPT, but (probably) not in P, and thus, under the widely-believed assumption that FPT is not equal to #W[1], our dichotomy shows precisely which problems #Sub(C) admit polynomial-time algorithms.

3.9 Parametrizing the Easiness of Machine Learning Problems

Sanjoy Dasgupta (University of California – San Diego, US)

Joint work of Dasgupta, Sanjoy; Chaudhuri, Kamalika; Freund, Yoav; Kpotufe, Samory; Sinha, Kaushik

We consider a variety of basic machine learning tasks whose worst case complexity (computational, or statistical, or both) is exponentially bad. These include:

- Nonparametric classification and regression
- Nearest neighbor search and classification
- Density-based clustering

In each case, the lower bounds are based on constructions that may be considered pathological relative to instances that occur in practice. It has been fruitful to formally describe types of structure that occur in natural instances, and to find algorithms that are able to exploit such structure.

This undertaking is in its infancy, and there are many immediate open problems.

3.10 How to Serve Impatient Users

Matthias Englert (University of Warwick, GB)

License Creative Commons BY 3.0 Unported license
© Matthias Englert
Joint work of Cygan, Marek; Englert, Matthias; Gupta, Anupam; Mucha, Marcin; Sankowski, Piotr
Main reference M. Cygan, M. Englert, A. Gupta, M. Mucha, P. Sankowski, "Catch Them if You Can: How to Serve Impatient Users," in Proc. of the 4th Conf. on Innovations in Theoretical Computer Science (ITCS'13), pp. 485–494, 2013.
URL http://dx.doi.org/10.1145/2422436.2422489

Consider the following problem of serving impatient users: we are given a set of customers we would like to serve. We can serve at most one customer in each time step (getting value v_i for serving customer i). At the end of each time step, each as-yet-unserved customer i leaves the system independently with probability q_i, never to return. What strategy should we use to serve customers to maximize the expected value collected?

The standard model of competitive analysis can be applied to this problem: picking the customer with maximum value gives us half the value obtained by the optimal algorithm, and using a vertex weighted online matching algorithm gives us $1 - 1/e \approx 0.632$ fraction of the optimum. As is usual in competitive analysis, these approximations compare to the best value achievable by an clairvoyant adversary that knows all the coin tosses of the customers. Can we do better?

We show an upper bound of ≈ 0.648 if we compare our performance to such an clairvoyant algorithm, suggesting we cannot improve our performance substantially. However, these are pessimistic comparisons to a much stronger adversary: what if we compare ourselves to the optimal strategy for this problem, which does not have an unfair advantage? In this case, we can do much better: in particular, we give an algorithm whose expected value is at least 0.7 of that achievable by the optimal algorithm. This improvement is achieved via a novel rounding algorithm, and a non-local analysis.

3.11 Smoothed Analysis of Local Search for the Maximum-Cut Problem

Michael Etscheid (Universität Bonn, DE)

License Creative Commons BY 3.0 Unported license
© Michael Etscheid
Joint work of Etscheid, Michael; Röglin, Heiko
Main reference M. Etscheid, H. Röglin, "Smoothed Analysis of Local Search for the Maximum-Cut Problem," in Proc. of the 25th ACM-SIAM Symp. on Discrete Algorithms (SODA'14), pp. 882–889, SIAM, 2014.
URL http://dx.doi.org/10.1137/1.9781611973402.66

Even though local search heuristics are the method of choice in practice for many well-studied optimization problems, most of them behave poorly in the worst case. This is in particular the case for the Maximum-Cut Problem, for which local search can take an exponential number of steps to terminate and the problem of computing a local optimum is PLS-complete. To narrow the gap between theory and practice, we study local search for the Maximum-Cut Problem in the framework of smoothed analysis in which inputs are subject to a small amount of random noise. We show that the smoothed number of iterations is quasi-polynomial, i. e., it is bounded from above by a polynomial in $n^{\log(n)}$ and ϕ where n denotes the number of nodes and ϕ denotes the perturbation parameter. This shows that worst-case instances are fragile and it is a first step in explaining why they are rarely observed in practice.

3.12 Preprocessing of NP-hard Problems

Uriel Feige (Weizmann Institute, IL)

Joint work of Feige, Uriel; Jozeph, Shlomo
Main reference U. Feige, S. Jozeph, "Universal Factor Graphs," in Proc. of the 39th Int'l Colloquium on Automata, Languages, and Programming (ICALP'12) – Part I, LNCS, Vol. 7931, pp. 339–350, Springer, 2012; pre-print available at arXiv:1204.6484v1 [cs.CC].
URL http://dx.doi.org/10.1007/978-3-642-31594-7_29
URL http://arxiv.org/abs/1204.6484v1

Consider a setting in which an input instance for an NP-hard optimization problem is supplied in two steps. In the first step, one gets to see some partial information about the input instance, referred to as a "preview". Based on this preview, a preprocessing algorithm can spend arbitrary (e. g., exponential) time in preparing some polynomial size "advice" string. In the second step, one gets to see the full input instance. Thereafter, a polynomial time algorithm attempts to solve the instance, and may use for this purpose the advice string prepared by the preprocessing algorithm. For various NP-hard optimization problems we present natural preview functions whose study appears to be well motivated. In certain cases we can prove that preprocessing leads to improved approximation ratios, and in certain cases we can prove limitations on how much preprocessing can help. Many natural questions within this framework remain open, and can serve as fertile ground for future research.

References

1 Uriel Feige, Shlomo Jozeph: *Universal Factor Graphs*. ICALP(1) 2012: 339–350
2 Uriel Feige, Shlomo Jozeph: *Demand Queries with Preprocessing*. ICALP(1) 2014: 477–488.

3.13 Average-Case Analysis of Parameterized Problems

Tobias Friedrich (Universität Jena, DE)

Joint work of Friedrich, Tobias; Nikolaos Fountoulakis; Danny Hermelin; Anton Krohmer

Many real world problems are NP-hard and are therefore generally believed not to be solvable in polynomial time. Additional assumptions on the inputs are necessary to solve such problems efficiently. Two typical approaches are (i) parameterized complexity where we assume that a certain parameter of the instances is small, and (ii) average-case complexity where we assume a certain probability distribution on the inputs. There is a vast literature on both approaches, but none about their intersection. Nevertheless, combining these two approaches seems natural and potentially useful in practice.

Analogous to the classical average-case complexity classes of Levin, we define a hierarchy of parameterized average-case complexity classes. To show the applicability of this theory we prove that the fundamental W[1]-complete problem k-Clique drops to the average-case analog of FPT for Erdős- Rényi random graphs of all densities and scale-free inhomogeneous random graphs.

3.14 On the Smoothed Approximation Performance of the 2-OPT Heuristic

Marvin Künnemann (MPI für Informatik – Saarbrücken, DE)

Joint work of Künnemann, Marvin; Manthey, Bodo

While the 2-OPT heuristic is a very simple and popular approach to solve Euclidean TSP, the theoretical understanding of its practical performance is still quite restricted. As a welcome exception, the smoothed analysis perspective proved useful to explain particularly the heuristic's fast running time observed in practice [1].

In this talk, we focus on the question of how well this paradigm also explains the practical approximation performance of the 2-OPT heuristic. We present new upper and lower bounds and, along the way, discuss aspects of the Gaussian and the ϕ-bounded perturbation models.

This is joint work (in progress) with Bodo Manthey.

References

1 Matthias Englert, Heiko Röglin, and Berthold Vöcking. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP (extended abstract). In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1295–1304. ACM, New York, 2007.

3.15 Constant Factor Approximation for Balanced Cut in the PIE Model

Konstantin Makarychev (Microsoft Corp. – Redmond, US)

Joint work of Makarychev, Konstantin; Makarychev, Yury; Vijayaraghavan, Aravindan
Main reference K. Makarychev, Y. Makarychev, A. Vijayaraghavan, "Constant factor approximation for balanced cut in the PIE model." in Proc. of the 46th Annual ACM Symp. on Theory of Computing (STOC'14), pp. 41–49, ACM, 2014; pre-print available at arXiv:1406.5665v1 [cs.DS].
URL http://dx.doi.org/10.1145/2591796.2591841
URL http://arxiv.org/abs/1406.5665v1

We propose and study a new semi-random semi-adversarial model for Balanced Cut, a planted model with permutation-invariant random edges (PIE). Our model is much more general than planted models considered previously. Consider a set of vertices V partitioned into two clusters L and R of equal size. Let G be an arbitrary graph on V with no edges between L and R. Let E_{random} be a set of edges sampled from an arbitrary permutation-invariant distribution (a distribution that is invariant under permutation of vertices in L and in R). Then we say that $G + E_{\text{random}}$ is a graph with permutation-invariant random edges. We present an approximation algorithm for the Balanced Cut problem that finds a balanced cut of cost $O(|E_{\text{random}}|) + n \cdot \text{polylog}(n)$ in this model.

In the most interesting regime, this is a constant factor approximation with respect to the cost of the planted cut.

3.16 Bilu-Linial Stable Instances of Max Cut and Minimum Multiway Cut

Yury Makarychev (TTIC – Chicago, US)

Joint work of Makarychev, Konstantin; Makarychev, Yury; Vijayaraghavan, Aravindan
Main reference K. Makarychev, Y. Makarychev, A. Vijayaraghavan, "Bilu-Linial Stable Instances of Max Cut and Minimum Multiway Cut," in Proc. of the 25th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA'14), pp. 890–906, SIAM, 2014.
URL http://dx.doi.org/10.1137/1.9781611973402.67

We investigate the notion of stability proposed by Bilu and Linial. We obtain an exact polynomial-time algorithm for γ-stable Max Cut instances with $\gamma > c\sqrt{\log n}\log\log n$ for some absolute constant $c > 0$. Our algorithm is robust: it never returns an incorrect answer; if the instance is γ-stable, it finds the maximum cut; otherwise, it either finds the maximum cut or certifies that the instance is not γ-stable. We prove that there is no robust polynomial-time algorithm for γ-stable instances of Max Cut when γ is less than the best approximation factor for Sparsest Cut with non-uniform demands. That suggests that solving γ-stable instances with $\gamma = o(\sqrt{\log n})$ might be difficult or even impossible.

Our algorithm is based on semidefinite programming. We show that the standard SDP relaxation for Max Cut (with ℓ_2^2 triangle inequalities) is integral if $\gamma > D(n)$, where $D(n)$ is the least distortion with which every n point metric space of negative type embeds into ℓ_1. On the negative side, we show that the SDP relaxation is not integral when $\gamma < D(n/2)$. Moreover, there is no tractable convex relaxation for γ-stable instances of Max Cut when γ is less than the best approximation factor for Sparsest Cut.

Our results significantly improve previously known results. The best previously known algorithm for γ-stable instances of Max Cut required that $\gamma > c\sqrt{n}$ (for some $c > 0$) [1]. No hardness results were known for the problem.

Additionally, we present an exact robust polynomial-time algorithm for 4-stable instances of Minimum Multiway Cut. We also study a relaxed notion of weak stability and present algorithms for weakly stable instances of Max Cut and Minimum Multiway Cut.

References

1 Yonatan Bilu, Amit Daniely, Nati Linial, and Michael Saks. On the practically interesting instances of MAXCUT. In *30th Int'l Symp. on Theoretical Aspects of Computer Science*, vol. 20 of *LIPIcs – Leibniz International Proceedings in Informatics*, pp. 526–537, 2013. http://dx.doi.org/10.4230/LIPIcs.STACS.2013.526

3.17 Smoothed Analysis of Local Search

Bodo Manthey (University of Twente, NL)

Local search algorithms are a powerful tool for finding good solutions for optimization problems that often works remarkable well in practice. Yet many local search algorithms show a very bad performance in the worst case.

Smoothed analysis has been applied successfully in order to explain the performance of local search algorithms.

We survey results and open problems of smoothed analysis applied to local search algorithms.

3.18 Analyzing Non-Convex Optimization for Dictionary Learning

Tengyu Ma (Princeton University, US)

Joint work of Arora, Sanjeev; Ge, Rong; Ma, Tengyu; Moitra, Ankur

Dictionary Learning, or as it is often called sparse coding is a basic algorithmic primitive in many machine learning applications, such as image denoising, edge detection, compression and deep learning. Most of the existing algorithms for dictionary learning minimize a non-convex function by heuristics like alternating minimization, gradient descent or their variants. Despite their success in practice, they are not mathematically rigorous because they could potentially converge to local minima.

In this work we show that, alternating minimization and some other variants indeed converge to global minima provably, under a generative model where samples are of the from $A \cdot x$, and A is a incoherent and low spectral norm ground truth dictionary and x is a stochastic sparse vector. Our framework of analysis is potentially useful for analyzing other alternating minimization type algorithms for problems with hidden variables. This is joint work with Sanjeev Arora, Rong Ge and Ankur Moitra.

3.19 Efficient Algorithms Beyond Worst-Case Combinatorial Bounds

Matthias Mnich (Universität Bonn, DE)

A common trait in combinatorics is to find maximum-size or minimum-size substructures of discrete object, and to measure the size of the structure in terms of the size of the object. These combinatorial results often come with polynomial-time algorithms that show how to find the desired structure of optimal size efficiently. For example, in any 3-SAT formula a fraction of 7/8 of its clauses can always be satisfied by some assignment which can be found in polynomial time.

We address the question how to beat the extremal bounds by a small additive constant k, by algorithms whose run-time depends exponentially only on this k. Such kind of problems are called "parameterizations above guarantee". We present some recent results and work in progress in this area.

3.20 Smoothed Analysis of Multi-Objective Optimization

Heiko Röglin (Universität Bonn, DE)

Joint work of Röglin, Heiko; Brunsch, Tobias
Main reference T. Brunsch, H. Röglin, "Improved Smoothed Analysis of Multiobjective Optimization," in Proc. of the 44th Annual ACM Symp./ on Theory of Computing (STOC'12), pp. 407–426, ACM, 2012.
URL http://dx.doi.org/10.1145/2213977.2214016

For most multi-objective optimization problems, the number of Pareto-optimal solutions is usually small in experiments even though in the worst case instances with an exponential number of Pareto-optimal solutions exist. In order to explain this discrepancy, the number of Pareto-optimal solutions has been studied in the model of smoothed analysis in a series of papers over the last decade. In this talk we survey some recent upper and lower bounds on the smoothed number of Pareto-optimal solutions.

3.21 Decompositions of Triangle-Dense Graphs

Tim Roughgarden (Stanford University, US)

Joint work of Gupta, Rishi; Roughgarden, Tim; Seshadhri, C.
Main reference R. Gupta, T. Roughgarden, C. Seshadhri, "Decompositions of Triangle-Dense Graphs," to appear in Proc. of ITCS 2015; pre-print available from author's webpage.
URL http://theory.stanford.edu/~tim/papers/triangle.pdf

High triangle density – the graph property stating that a constant fraction of two-hop paths belong to a triangle – is a common signature of social networks. This paper studies triangle-dense graphs from a structural perspective. We prove constructively that significant portions of a triangle-dense graph are contained in a disjoint union of dense, radius 2 subgraphs. This result quantifies the extent to which triangle-dense graphs resemble unions of cliques. We also show that our algorithm recovers planted clusterings in approximation-stable k-median instances.

3.22 Utilitarian View of Rankings

Or Sheffet (Harvard University, US)

Joint work of Boutillier, Craig; Caragiannis, Ioannis; Haber, Simi; Lu, Tyler; Procaccia, Ariel; Sheffet, Or
Main reference C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, O. Sheffet, "Optimal social choice functions: a utilitarian view," in Proc. of the 13th ACM Conf. on Electronic Commerce (EC'12), pp. 197–214, ACM, 2012; pre-print available from author's webpage.
URL http://dx.doi.org/10.1145/2229012.2229030
URL http://www.cs.cmu.edu/~arielpro/papers/optvoting.full.pdf

In the extensively studied problem of Social Choice (or rank aggregation) n individuals are picking together one alternative out of m possible alternatives. Each individual has a preference among all m alternatives (a total ranking), and the social choice function takes as input these n rankings and outputs a single chosen alternative, called the winner. In this talk we do not focus on the age-old question of truthful rank aggregation, but rather attempt to define the high-level idea that the winner is supposed to be the most favorable alternative for the entire population.

Inspired but newly formulated ideas as to the role of clustering [1], we view social choice as a proxy for maximizing social welfare. Our premise is that agents have (possibly implicit or latent) utility functions, and the goal of a social choice function is to maximize the social welfare – i. e., (possibly weighted) sum of agent utilities – of the selected alternative. We study the model both under worst-case and non-worst case assumptions. We will also discuss current, open ended, work as to maximizing utility of a matching / stable matching.

References

1 Maria-Florina Balcan, Avrim Blum, and Anupam Gupta. Approximate clustering without the approximation. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1068–1077. SIAM, Philadelphia, PA, 2009.

3.23 Probabilistic Lipschitzness: A Measure of Data Niceness

Ruth Urner (Carnegie Mellon University – Pittsburgh, US)

Main reference R. Urner, "Learning with non-Standard Supervision," Ph. D. Thesis, University of Waterloo, 2013.
URL https://uwspace.uwaterloo.ca/handle/10012/7925

Machine learning theory mostly analyses the learning performance of learning in the worst case over all data-generating distributions. However, this framework is often too pessimistic. In particular, for learning settings that employ unlabeled data, such as semi-supervised and active learning, provable savings in label complexity are impossible without assumptions of niceness on the data generating distribution. We propose Probabilistic Lipschitzness (PL), to model marginal-label relatedness of a distribution. This notion is particularly useful for modeling niceness of distributions with deterministic labeling functions. We present convergence rates for Nearest Neighbor learning under PL. We further summarize reductions in labeled sample complexity for learning with unlabeled data (semi-supervised and active learning) under PL.

References

1 Ruth Urner. *Learning with non-Standard Supervision*. Ph. D. Thesis, University of Waterloo, 2013.

2 Ruth Urner and Shai Ben-David and Shai Shalev-Shwartz. *Unlabeled data can Speed up Prediction Time*. Proceedings of the 28th International Conference on Machine Learning (ICML), 2011.

3 Ruth Urner, Sharon Wullf and Shai Ben-David. *PLAL: Cluster-based active learning*. Proceedings of the 26th Conference on Learning Theory (COLT), 2013.

Participants

- Spyros Angelopoulos
CNRS & UPMC, FR
- Sanjeev Arora
Princeton University, US
- Maria-Florina Balcan
Carnegie Mellon University, US
- Shai Ben-David
University of Waterloo, CA
- Markus Bläser
Universität des Saarlandes, DE
- Avrim Blum
Carnegie Mellon University, US
- Karl Bringmann
ETH Zürich, CH
- Kamiel Cornelissen
University of Twente, NL
- Radu Curticapean
Universität des Saarlandes, DE
- Sanjoy Dasgupta
University of California –
San Diego, US
- Matthias Englert
University of Warwick, GB
- Michael Etscheid
Universität Bonn, DE
- Uriel Feige
Weizmann Institute, IL
- Tobias Friedrich
Universität Jena, DE
- Anupam Gupta
Carnegie Mellon University, US
- Marvin Künnemann
MPI für Informatik –
Saarbrücken, DE
- Tengyu Ma
Princeton University, US
- Konstantin Makarychev
Microsoft Res. – Redmond, US
- Yury Makarychev
TTIC – Chicago, US
- Bodo Manthey
University of Twente, NL
- Matthias Mnich
Universität Bonn, DE
- Heiko Röglin
Universität Bonn, DE
- Tim Roughgarden
Stanford University, US
- Or Sheffet
Harvard University, US
- Christian Sohler
TU Dortmund, DE
- Ruth Urner
Carnegie Mellon University, US
- Sergei Vassilvitskii
Google Inc. –
Mountain View, US