Research Statement

Luis Rademacher

1 Summary

My research interests are in theoretical computer science and some related areas, from pure mathematics to applications. I am primarily interested in the foundations of data science and artificial intelligence and this has lead to a focus on problems in convex geometry, machine learning, matrix computations and optimization.

My work in convex geometry aims to understand the "slicing problem", a problem posed by J. Bourgain in 1986 [Bou86] which is one of the main open problems in convex geometry. This importance is based on the fact that it is intimately related to older outstanding problems in convex geometry. In my work [Rad12][1], I disproved a conjecture about Sylvester's (random simplex) problem, a much older problem in convex geometry. A positive answer to the conjecture would have implied a solution to the slicing problem. In my work [Rad15], I investigated properties of convex bodies that are extremal for a variational formulation of the slicing problem.

Within matrix computations, I have focused on the low-rank matrix approximation problem. My most significant contribution in this area has been the introduction of *volume sampling* [DRVW06a, DR10], a way of sampling subsets of columns of a matrix with optimal approximation guarantees. This algorithmic idea has been quite influential in matrix approximation as well as other problems such as the generation of a set of diverse results by a search engine for a given query [KT12]. It has also been used and generalized in more distant research, such as the design of improved algorithms for quadratic integer programming [GS11].

In machine learning, I have been interested in efficient inference of geometric properties of high-dimensional data, understood generally as estimating properties of distributions from samples. In particular, my aim has been to expand the frontier of problems with efficient algorithms and to understand the discrepancy between information-theoretic complexity—the number of samples needed to estimate a property—and algorithmic complexity—the computational cost. One of my noteworthy contributions in this area has been the study of an unusual *blessing of dimensionality* [ABG$^+$14] in the complexity of estimating parameters of Gaussian mixture models, a very common model for data. This work shows in a precise sense that this problem is information-theoretically hard for generic low-dimensional instances and computationally easy for generic high dimensional instances.

[1] All of my papers can be found at `http://web.cse.ohio-state.edu/~lrademac/`.

The scientific community has recognized the relevance of those contributions through the funding of my current and future research via two NSF grants: the NSF Early Career award and a regular grant. These two grants will continue funding the work described in section 3.

2 Research areas

2.1 Convex geometry

Sylvester's problem and the slicing problem. One of my research contributions is my line of work on one of the main open problems in asymptotic convex geometry: the "slicing problem". The problem aims to understand in a precise way the distribution of mass of high-dimensional convex bodies by defining a measure of spread of a convex body, the isotropic constant, and asking whether it has a universal upper bound among all convex bodies and all dimensions. Jean Bourgain posed the slicing problem in 1986 [Bou86], but research on the problem showed deep connections with another famous problem going back to J.J. Sylvester (1865) [Pfi89]. One of my papers [Rad12] deals with a conjecture related to Sylvester's problem: In dimension d, is it true that if K is a convex body contained in another convex body L, then the expected volume of the convex hull of $d+1$ random points in K is less than the expected volume of the convex hull of $d+1$ random points in L? The relevance of this conjecture is that an affirmative answer for all dimensions would imply a solution to the slicing problem [Mec, Rad12]. My work shows that the answer to the conjecture depends on the dimension and it is false for dimensions four and higher, thus disproving the conjecture in general. In another paper, [Rad15], I study the class of shapes that could maximize the isotropic constant. A complete understanding of maximizers would clearly solve the slicing problem as described above. In particular, an outstanding conjecture is that the maximizer in every dimension is the simplex, the higher dimensional generalization of the triangle and the tetrahedron. In my paper I show that if one restricts the search of a maximizer to the (still dense) family of simplicial polytopes, then the only possible maximizer in that class could be the simplex. This provides some evidence to the conjecture.

2.2 Matrix computations

Low rank matrix approximation Among my main research contributions are efficient algorithms for a certain low rank matrix approximation problem (the column subset selection problem, CSS). My coauthors and I introduced in [DRVW06b] a distribution on subsets of columns of a matrix called *volume sampling* (also known as a fixed-size determinantal point process in related literature) that gives an optimal approximation guarantee for the CSS problem:

Definition 1 (Volume sampling). *Let A be an n-by-m real matrix. Let $1 \leq k \leq m$. The* volume sampling *distribution is the distribution over subsets of k columns of A so that subset $S \subseteq \{1, \dots, m\}$ is picked with probability proportional to the square of the k-dimensional volume of the simplex whose vertices are the origin and the columns of A in S.*

Theorem 2 ([DRVW06b]). *Let A be an n-by-m real matrix. Then*

$$\mathbb{E}\,\|A - \pi_S(A)\|_F^2 \le (k+1)\,\|A - A_k\|_F^2\,, \tag{1}$$

when S is picked according to volume sampling, $\pi_S(A)$ denotes the matrix obtained by projecting all the rows of A onto $\operatorname{span}(a_i : i \in S)$, *and A_k is the matrix of rank k closest to A under the Frobenius norm.*

(The factor "$k+1$" in 1 cannot be improved, see [DRVW06b].)

This paper also introduced an adaptive sampling technique and systematically studied approximation guarantees given by it. Adaptive sampling has been very influential in the matrix approximation literature, such as in the development of optimal row and column-based matrix decompositions in [BW14].

In the first paper, [DRVW06b], *volume sampling* was only used to find (existentially) the optimal approximation factor for CSS. My second paper on this line of research, [DR10], provides an efficient algorithm to sample a representative subset of columns of a matrix according to *volume sampling.*

2.3 Machine learning

In machine learning, I am interested in efficient estimation of geometric properties of high-dimensional data. On this aspect, my aim is to design algorithms that have strong theoretical computational complexity and robustness guarantees and that are also efficient in practice. I am also interested in understanding the boundary between efficiently solvable and intractable problems, both from the information-theoretic sense—the number of samples or data points needed for robust inference—and the algorithmic sense—the computational cost of robust inference.

Efficient algorithms based on the method of moments for high-dimensional estimation, avoiding the curse of dimensionality. Paper [AGR13] is joint work with one of my collaborators and my PhD student J. Anderson. In it we show that given samples from an n-dimensional simplex, one can efficiently estimate the simplex. This paper proposes a simple algorithm based on the efficient enumeration of local maxima of the directional third moment of the sample. This work shows a new aspect of the problem of learning an intersection of halfspaces, as known algorithms for this problem do not have polynomial complexity as a function of the number of halfspaces. Our work shows that when one considers the uniform distribution in the intersection of $n+1$ halfspaces in dimension n the problem is efficiently solvable, a regime that avoids the curse of dimensionality. The simplex learning algorithm is also practical. This line of work also lead to a remarkable discovery: A new and rare "blessing of dimensionality" for the problem of estimating parameters of Gaussian Mixture Models (GMM) (described in [ABG$^+$14], joint work with my collaborators and my PhD students J. Anderson and J. Voss). This problem has a long history (going back at least to Karl Pearson's work in 1894). The unexpected phenomenon is that the amount of data and computation needed to solve the problem decreases sharply as the dimension grows. A technical contribution of our paper is the use of a Poissionization technique to reduce the GMM estimation problem to an instance of the Independent Component Analysis problem, to which existing efficient algorithms can be applied. Poissonization turns a problem with no independence structure into one with independence structure.

The centroid body: algorithms and applications to robust estimation. Paper [AGNR15] (joint work with a collaborator and my PhD students J. Anderson and A. Nandi) introduces the first (as far as I know) algorithmic use of the centroid body, a convex body, going back to the work of Petty [Pet61] in convex geometry, that one can associate to any multivariate distribution with finite first moment. This body can play the role of the covariance matrix for some applications to data analysis with heavy tails—such as financial data. In [AGNR15], the centroid body is used to provide a provably correct algorithm for a standard problem, Independent Component Analysis (ICA), previously not known to be efficiently solvable without higher moments. The paper also uses a "Gaussian damping" reweighting technique that one can apply to an ICA instance to reduce the heavy tailed-case to the (better understood) case with finite higher order moments. The proposed algorithmic ideas are not just of theoretical interest: my preprint "New practical algorithms for heavy-tailed Independent Component Analysis" discusses a practical implementation of them that outperforms the best previously known algorithms in many heavy-tailed regimes.

2.4 Expander graphs and graph sparsification

Motivated by empirical work on reliable information routing in the Internet, in "Expanders via random spanning trees", [FGRV14], my coauthors and I consider the question of whether the union of a few random spanning trees from a given graph has reliability comparable to that of the underlying graph. Spanning trees are traditionally considered for routing because they lead to efficient routing. A mathematical way of modeling the reliability of a graph is the expansion of a graph—a measure of how connected it is. Thus, a formal version of the question is whether one can recover the expansion of every cut of a graph by the union of a few uniformly random spanning trees. We show that in a bounded degree graph, the union of two random spanning trees approximates the size of every cut within a factor of $\log n$, and this is best possible. We also show that the union of two random spanning trees in the complete graph is an expander.

2.5 Optimization

Smoothed complexity of multiobjective optimization. Paper [BGRR14] with my collaborators deals with multiobjective optimization, specifically, optimization problems with multiple linear functions to be optimized. One standard way of handling multiple objectives is to first focus on the set of Pareto optima, that is, feasible solutions for which no objective can be improved without hurting another objective. For example, Pareto optima are used in heuristics for optimization problems [NU69]. It is known that the number of Pareto optima can be exponential in the number of objectives in the worst case. In [BGRR14], my coauthors and I show lower bounds on the number of Pareto optima even when the input is a small random perturbation of a hard instance (namely, in the *smoothed complexity* model), which shows in a precise sense that the problem is generically hard when the number of objectives grows. The bounds are also exponential in the number of objectives. In that paper we also show the first lower bound for a natural multiobjective problem, the multiobjective maximum spanning tree problem with randomly chosen edge weights. The arguments are based on a novel connection between multiobjective optimization and results in discrete geometry and geometric probability about arrangements of hyperplanes.

Infinite dimensional linear and integer programming. Motivated by infinite-horizon optimal planning problems, work [RTV] with collaborators from optimization studies a generalization of packing and covering polyhedra to problems in infinite dimensions. A difficulty in infinite horizon optimal planning problems is that the natural objective can take an infinite value at interesting feasible solutions. A standard but somewhat limited approach to this issue is to reweight the objective function with exponentially decaying coefficients along the infinite time horizon—sometimes economically justified as a *discount factor*. This has the drawback of introducing a strong bias towards earlier time periods of the model. In [RTV] my coauthors and I study a notion of optimality consistent with infinite objectives, whose definition is based on a complementarity condition, namely complementary slackness. In order to implement this idea of using complementary slackness as a notion of optimality, we introduce a primal/dual pairing between infinite packing and covering linear programs that is easy to construct formally, without using algebraic or topological duality. In general, this is only helpful if the proposed pairing is indeed a primal/dual pair in some standard sense. A contribution of [RTV] is to show that for the case of packing/covering linear programs, complementary slackness makes sense for our proposed primal/dual pairing. Specifically, one can define optimality by means of complementary slackness of a pair of solutions and show that such a pair of solutions exists under reasonable assumptions.

3 Future research plans

My broader aims in research can be summarize as follows:

- To develop machine intelligence and understand its limits.
- To understand variational problems over the space of convex bodies.

I will now describe some specific plans to pursue these aims.

Transforming data analysis via new algorithms for feature extraction. Analysis and exploration of data, including classification, inference, and retrieval, are ubiquitous tasks in science and applied fields. Given any such task, a fundamental paradigm is the extraction of features that are relevant. In the design of algorithms for the analysis and exploration of data, feature extraction techniques act as basic building blocks or primitives that model complex behavior when combined. Some of the fundamental feature extraction tools include Principal Component Analysis (PCA), Independent Component Analysis (ICA), and half-space-based learning and classification. Data rarely satisfy the precise assumptions of these models and feature extraction tools, and combining these tools amplifies errors. This motivates the challenging task of designing new algorithms that are robust against noise and that when combined as building blocks the error propagation stays under control.

This line of work will:

1. Raise ICA from a very successful practical tool to an algorithmic primitive with strong theoretical guarantees and applicability to a rich family of problems beyond independence.

2. Find reasonable assumptions and algorithms that allow efficient learning of intersections of half-spaces.

3. Systematically study the following well-motivated refinement of PCA known as the subset selection problem. This refinement aims to select relevant features among the given features of the input data, unlike PCA, which creates new and possibly artificial features.

New feature extraction algorithms enhance the toolbox available to researchers in data-intensive fields such as biology, signal processing and computer vision. They also enable improved data analysis by practitioners in security, marketing, business and government processes, and essentially any field that involves the analysis of feature-rich data. This line of work includes the implementation of the more practical algorithms.

This research is currently being funded by my NSF Early Career Award.

Geometry and high-dimensional inference. I am interesting in improving our understanding of the problem of reconstructing structure of probability distributions from sampled data. I will investigate the use of tensor-based and other higher order methods. This line of research lies at the interface of theoretical computer science, machine learning, signal processing and statistics and will have potential impact in all of these fields. In recent years there has been a resurgence of interest in tensor methods in data analysis and inference, particularly in theoretical computer science. These methods will prove useful in a variety of applications in machine learning, signal processing and other fields.

This research will develop algorithms for solving a range of problems including blind source separation, spectral clustering, inference in mixture models and estimating geometry of distributions. It will analyze the complexity of these and related problems. In particular, it will strive to understand the computational efficiency and dependence on the dimension of the space, studying "the curses and blessings of dimensionality". It will also address a somewhat mysterious discrepancy between sample and algorithmic complexity in our understanding of many high dimensional inference problems.

This line of work is currently being funded by my NSF award "AF: Small: Geometry and High-dimensional Inference".

Convex geometry. I am interested in variational approaches to the understanding of extremal problems in convex geometry, such as Mahler's problem and the slicing problem. I am also interested in the use of computer-assisted arguments to understand the extremal shapes for those problems. For example, given the recent understanding of the structure of local optima of Mahler's problem in [NPRZ10], a computer-aided understanding of extremal shapes for Mahler's problem should be approachable in low dimension.

References

[ABG+14] Joseph Anderson, Mikhail Belkin, Navin Goyal, Luis Rademacher, and James R. Voss. The more, the merrier: the blessing of dimensionality for learning large gaussian mixtures. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, Barcelona, Spain, 2014.

[AGNR15] Joseph Anderson, Navin Goyal, Anupama Nandi, and Luis Rademacher. Heavy-tailed Independent Component Analysis. In *FOCS*, Berkeley, CA, 2015.

[AGR13] J. Anderson, N. Goyal, and L. Rademacher. Efficient learning of simplices. *COLT*, 2013.

[BGRR14] Tobias Brunsch, Navin Goyal, Luis Rademacher, and Heiko Röglin. Lower bounds for the average and smoothed number of pareto-optima. *Theory of Computing*, 10:237–256, 2014.

[Bou86] J. Bourgain. On high-dimensional maximal functions associated to convex bodies. *Amer. J. Math.*, 108(6):1467–1476, 1986.

[BW14] Christos Boutsidis and David P. Woodruff. Optimal CUR matrix decompositions. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 353–362, 2014.

[DR10] A. Deshpande and L. Rademacher. Efficient volume sampling for row/column subset selection. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 329 –338, October 2010.

[DRVW06a] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. Matrix approximation and projective clustering via volume sampling. In *SODA*, pages 1117–1126, 2006.

[DRVW06b] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. Matrix approximation and projective clustering via volume sampling. *Theory of Computing*, 2(1):225–247, 2006. Preliminary version in SODA 2006.

[FGRV14] Alan M. Frieze, Navin Goyal, Luis Rademacher, and Santosh Vempala. Expanders via random spanning trees. *SIAM J. Comput.*, 43(2):497–513, 2014.

[GS11] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with PSD objectives. In *FOCS*, 2011.

[KT12] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2-3):123–286, 2012.

[Mec] Mark Meckes. Personal communication.

[NPRZ10] Fedor Nazarov, Fedor Petrov, Dmitry Ryabogin, and Artem Zvavitch. A remark on the Mahler conjecture: local minimality of the unit cube. *Duke Math. J.*, 154(3):419–430, 2010.

[NU69] G. Nemhauser and Z. Ullmann. Discrete dynamic programming and capital allocation. *Management Science*, 15(9):494–505, 1969.

[Pet61] C. M. Petty. Centroid surfaces. *Pacific J. Math.*, 11(4):1535–1547, 1961.

[Pfi89] Richard E. Pfiefer. The historical development of J. J. Sylvester's four point problem. *Mathematics Magazine*, 62(5):309–317, 1989.

[Rad12] Luis Rademacher. On the monotonicity of the expected volume of a random simplex. *Mathematika*, 58(1):77–91, 2012.

[Rad15] Luis Rademacher. A simplicial polytope that maximizes the isotropic constant must be a simplex. *Mathematika*, 2015.

[RTV] Luis Rademacher, Alejandro Toriello, and Juan Pablo Vielma. On packing and covering polyhedra in infinite dimensions. To appear in Operations Research Letters, 2016.