Research Statement

Joseph Anderson

My research interests lie in machine learning, geometry, and algorithms. Problems in machine learning and data analysis are abundant, both in theory and applications, but the most important problems are complicated enough that either there is no known solution, or efficient algorithms require a significant amount of ad-hoc analysis and modeling. Furthermore, some solutions are only statistical in nature, only providing asymptotic convergence of an estimator, but no guarantee of efficiency (e.g. only requiring polynomially many samples). Although many of these approaches are theoretically sound and can be simple to implement, some data analysis techniques do not take advantage of the rich structure endowed by the nature of their respective problem framework. I am interested in finding new geometric approaches to machine learning to arrive at novel and unexpected solutions. Furthermore, finding these connections between geometry and statistics provides new ways to express problems in simpler ways, enabling better algorithm design and lowering the barrier of entry for other non-expert scientists and researchers. Such results not only provide new research directions in data analysis, but often reveal new problem domains where classical tools can be used.

The main theme of my work can be expressed with two goals in mind: (1) Use geometric primitives to solve difficult statistical learning problems and (2) Develop new geometric primitives which can be used to simplify learning problems. One of my main contributions is that, perhaps counter-intuitively, several important problems in machine learning can be reduced to a signal separation problem: Independent Component Analysis (ICA). The typical illustration of ICA is the "cocktail party problem" where the observed data is several overlapping audio signals (e.g multiple people speaking in a room) captured by multiple microphones and the task is to separate each one into its own audio track. Although ICA has been well studied by the signal processing community and has a simple formulation, my work demonstrates how one can use it as a powerful algorithmic primitive. Furthermore, my work uses novel algorithmic applications of convex geometry to construct more robust methods for signal separation, in particular providing the first provably efficient algorithm to solve ICA when the data may come from heavy-tailed distributions.

1 Previous work

1.1 Independent Component Analysis as an algorithmic primitive

Independent Component Analysis (ICA) has a rich history, originating in the signal processing community, and is formulated as follows: we are given independent samples of a random variable $X = AS$ where A is a fixed, unknown, $n \times n$ matrix (often referred to as the "mixing matrix") and $S \in \mathbb{R}^n$ is a random vector with mutually independent coordinates. The goal is to recover the matrix A and consequently to be able to "de-mix" the signals and recover S. There is a vast literature on ICA, and it has found many practical applications in communication, medical imaging, finance, and economics.

My work in COLT 2013 ([AGR13]) and COLT 2014 ([ABG$^+$14]) demonstrates reductions to ICA in the algorithmic sense. If one can transform a problem (efficiently, and in a reversible way) into an ICA model, we can use existing methods for ICA to solve the transformed problem, then reverse the initial process to find a solution to the original problem. In this way, we show that with ICA as an *algorithmic primitive* leads to simple, expressive algorithms for some important problems in machine learning, where the reductions are non-trivial and require interesting new connections between statistics and convex geometry.

Learning Gaussian mixture models. As an example of when ICA can be used to solve a seemingly unrelated problem is learning the parameters of Gaussian mixture. A Gaussian mixture model (GMM) can be defined as a weighted combination of multiple Gaussian distributions; they have wide application in topic modeling, economics, finance, image processing, and biology. To *learn* a GMM means to estimate the parameters of each Gaussian distribution which is part of the mixture. This is an important and well-studied learning problem, and the first provably efficient algorithm in high-dimension was discovered by Dasgupta in [Das99]. My work in COLT 2014 ([ABG$^+$14]) demonstrates an algorithm which reduces learning a GMM to solving an ICA model. The main ingredient of this result is a careful way of sampling the data and, by having access to an algorithm for ICA, one can efficiently learn the parameters of a GMM with many more Gaussians than the ambient dimension of the problem. Along with the algorithmic result of a new complexity upper bound for this problem, we prove that the problem is *generically hard* in low dimension; that is, we show limitations on the learnability of a GMM in low-dimension in general, and when the problem will require an exponential number of samples for *any* algorithm to succeed. We extend this to new hardness results for the general ICA problem when the number of signals is larger than the dimension of the samples.

1.2 Algorithmic convex geometry for solving ICA

A standard approach to solve ICA works in two phases: first whiten the data, meaning transform the samples so that they come from a rotated product distribution, then find the rotation. Whitening works as follows: from the model $X = AS$, estimate the covariance $\Sigma = \operatorname{Cov}(X) = \operatorname{Cov}(AS) = A \operatorname{Cov}(S) A^T = AA^T$ by assuming S has independent coordinates. Next, the *whitened* model is given by $\Sigma^{-1/2}X = \Sigma^{-1/2}AS = TS$, where T is an unknown unitary transformation (i.e. a rotation possibly scaling of S). Once the data is whitened, the model can be viewed as only a rotation and scaling of S, and the rotation can be recovered by studying various statistics of the data, for example, the directional fourth moment. Combining these two steps, we can reconstruct A and S up to a permutation and scaling.

Any algorithm which solves ICA using the above approach must efficiently and accurately estimate the second and higher moments of the data. This leads to the question: **Can we solve ICA when only very few moments exist?** In a paper published in FOCS 2015 ([AGNR15]), we answer this question affirmatively by constructing an algorithm which solves ICA when the underlying signals are *heavy-tailed*: we assume only that only the $(1+\epsilon)$ moment exists for the data. Our algorithm is provably efficient: it only requires number of samples polynomial in the dimension and other natural parameters of the problem. The key ingredient is the use of the *centroid body* of the data: a tool from convex geometry that provides and analogue of the covariance matrix when the covariance of the samples is not a well-defined object. This is the first algorithmic use of the centroid body and promises other interesting applications in machine learning beyond only the context of ICA. We then use careful optimization to recover the rotation and hence recover the mixing matrix.

Future work

Geometry does not always make things easier; in fact, it is known that even just estimating a convex body in high-dimension is a hard problem. However, my work leads to new explanations of why some problems which seem hard are actually approachable in high-dimension. It shows this through translating statistical goals into geometric ones that are "distribution-free" in that it is agnostic toward the distribution from which input samples are drawn. The driving question behind my future research is: **what statistical problems can be better understood with geometric tools?**

Removing dependence on moments. I am actively investigating more algorithmic uses of convex geometry in machine learning. There are other constructs that arise in convex geometry, for instance in the book by Gardner ([Gar95]) and work by [Wer04], which open new ways to approach problems particularly in imaging, sampling, and optimization. With the study of these objects comes many interesting open problems, such as investigating isoperimetric inequalities for certain bodies, central-limit theorems for convex bodies, and Mahler's conjecture. In the results toward heavy-tailed ICA we show the first application of the centroid body to machine learning as an algorithmic tool, but the centroid body is a much more general object than just providing a covariance-like measurement of data. One question, for instance, is **what other algorithms could benefit from replacing the covariance matrix of a distribution by the covariance of the centroid body?** Results in this line of work could have impact in many of the natural sciences as well as signal processing and data analysis.

Floating body. One candidate to replace the centroid body in the heavy-tailed setting is the *floating body* of X, which is of interest particularly in assessments of financial risks. The floating body is a generalization of a quantile to multiple dimensions, being defined as the intersection of all half-spaces which contain at least a $1-\alpha$ fraction of the mass of X. The floating body will exist for any distribution, it will always be convex, and additionally it will share the symmetries of the centroid body which make orthogonalization possible. However, we do not currently know of an efficient way to get samples from the floating body, or have formal results for membership in the floating body. This is one goal of my current work, to study the floating body and provide first an information theoretic upper bound for estimating membership, then an algorithmic technique which can be used to improve the heavy-tailed ICA algorithm outlined above.

Geometry of Gaussian mixtures. Another question which connects geometry with machine learning is: **how are two different Gaussian mixtures related, in a geometric sense?** Work by McCann shows that the space of Gaussian distributions is geodesically convex in the manifold of absolutely continuous probability measures with finite second moment; the metric under consideration here is the *Wasserstein Distance*, which is related to the Earth Mover Distance, which is well-known in the machine learning community. A naturally related problem is to determine if the space of Gaussian mixtures has a similar property; if it forms a well-behaved space such as a manifold, one could use tools for optimization over manifolds to learn arbitrary Gaussian Mixtures. The first question I would ask toward a result of this type is: **can we compute and understand the Wasserstein distance between two GMM's?** From there, one could try to determine what kind of optimization, if any, could be applied over the space of Gaussian mixtures with respect to this metric, and whether the space is even convex.

References

[ABG+14] Joseph Anderson, Mikhail Belkin, Navin Goyal, Luis Rademacher, and James Voss. The more, the merrier: the blessing of dimensionality for learning large gaussian mixtures. In *Proceedings of The 27th Conference on Learning Theory*, pages 1135–1164, 2014.

[AGNR15] Joseph Anderson, Navin Goyal, Anupama Nandi, and Luis Rademacher. Heavy-tailed independent component analysis. In *56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

[AGR13] Joseph Anderson, Navin Goyal, and Luis Rademacher. Efficient learning of simplices. In *Conference on Learning Theory*, pages 1020–1045, 2013.

[Das99] S. Dasgupta. Learning Mixture of Gaussians. In *40th Annual Symposium on Foundations of Computer Science*, 1999.

[Gar95] Richard J Gardner. *Geometric tomography*, volume 6. Cambridge University Press Cambridge, 1995.

[Wer04] Elisabeth Werner. Floating bodies and illumination bodies. In *Proceedings of the Conference Integral Geormetry and Convexity, Wuhan.* World Scientific, 2004.