My interests lie at the interface of computer science and statistics. I view questions arising from data analysis through the lens of algorithms and complexity. My main research interests revolve around the application of diffusion, concentration of measure and Riemannian geometry to problems primarily in Machine Learning and Convex Optimization. More broadly, my interests include computational geometry, network algorithms, complexity theory and statistical physics. In the remainder of this document, I shall elaborate on the questions that I have worked on in the past, and the directions that I intend to pursue in the future.

Manifold Learning

Machine learning can be broadly defined as a discipline whose goal is to enable a computer to make inferences from observed data about future observations. There are two key aspects of this field, where the frontier needs to be pushed forward. The first is the question of making inferences from very few observations. This question predates machine learning, and is referred to as "poverty of stimulus" in the work of linguists such as Chomsky. The second is the problem of dealing with data with increasing dimensionality, which has come to prominence through recent applications in vision, bioinformatics and the world wide web.

A hypothesis whose validity would mitigate both of these problems is that high dimensional data tend to lie in the vicinity of a low dimensional embedded manifold. This hypothesis, henceforth referred to as the "manifold hypothesis" has become influential over the past decade and has given rise to the field known as manifold learning. My recent and continuing research addresses two basic questions related to this area. The first is to quantify the sense in which efficient inference is possible under the manifold hypothesis. In work with Partha Niyogi [22], I showed that under the manifold hypothesis, the number of samples needed for the basic task of classification depends only on certain parameters of the manifold, but not on the ambient dimension.

The second is the question of whether the hypothesis that data lie near a manifold be tested efficiently. In a recent paper with Sanjoy Mitter [19], we showed that the number of samples needed to test whether a probability distribution supported in a unit ball in $\mathbb{R}^n$ for large n has a mean-squared distance of less than ϵ from a low dimensional manifold is independent of n, assuming the manifold has bounded curvature and volume. This result gives a bound on the informational complexity of the problem and, in the special case of $k-$means gives rise to the most efficient algorithm currently known in a natural statistical framework. It also leads to an algorithm for fitting piecewise linear curves. I am currently working on obtaining an

algorithm for this task when the intrinsic dimension of the manifold is two or more, and have obtained partial results.

Randomized interior point methods for convex optimization

Interior point methods in the form known today were born with the landmark paper of Karmarkar [8] on the subject. A fundamental insight of this paper and a sequel [9] is the implicit use of a non-Euclidean Riemannian metric that makes the distance of the boundary to any interior point infinite. In recent years, randomized algorithms have had great impact on questions in convex geometry such as computing the volume of a convex set, sampling and optimization. In two recent papers [24, 7], I explored *randomized interior point methods*. The algorithms in [24, 7] and certain long-step variants thereof have been implemented in practice by Huang and Mehrotra [6] in the context of Walk-and-Round Heuristics for Mixed-Integer Linear Programs, and have lead to novel results.

I addressed the question of sampling from the uniform distribution on a convex polytope and linear optimization in work with Kannan [7], and then generalized it to sampling from arbitrary convex sets (equipped with a so-called self-concordant barrier) and convex optimization respectively in [24]. This question has lead to several advances in the theory of Markov chains, in addition to having applications to computing the volume of a polytope and to convex optimization. We designed and analyzed a new random walk on the points of an $n-$dimensional polytope and showed that it mixes in strongly polynomial time if started from a "central" point (this is the first Markov chain algorithm with this property). A novel aspect of this random walk, termed "Dikin walk" is that it is *anisotropic*, in the sense that at different points of the polytope, that transition step roughly consists of picking a random point from an ellipsoid that changes both in shape and size from point to point. This anisotropy leads the walk to take long steps in the interior of the polytope and short steps near the periphery. We also used this result to design an affine interior point algorithm that does a random walk to solve linear programs approximately. This is a natural randomized version of the short-step affine scaling algorithm from linear programming pioneered by Dikin [3]. The fact that our algorithm has a provably polynomial run-time is notable, since such a bound is not known for its deterministic counterpart, the widely used affine scaling algorithm of Dikin.

Let $\mathbb{K} \subset \mathbb{R}^n$ be a convex compact set and let $\mathcal{F}$ be a set of convex functions from $\mathbb{K}$ to $\mathbb{R}$. *Online convex optimization* is defined as a repeated T-round game between the player (the algorithm) and Nature. On each round t, the player chooses $x_t \in \mathbb{K}$ and Nature chooses $\ell_t \in \mathcal{F}$. The player suffers loss $\ell_t(x_t)$, observes ℓ_t, and the game

proceeds to the next round. The goal of the player is to choose the sequence of x_t's in such a way as to keep *regret* with respect to any fixed comparator $u \in \mathbb{K}$

$$\text{Regret}_T(u) = \sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(u)$$

as small as possible. In recent work [20] with Alexander Rakhlin, we designed an algorithm for online convex optimization, that does a time-inhomogenous Dikin walk to track a changing distribution. From a technical point of view, the novel aspect is that in order to choose x_t, the player performs a single Dikin step starting from x_{t-1}. This is made possible by the fact that the Dikin walk mixes faster than other known random walks, as one approaches stationarity. I plan to investigate this further because a better understanding of global aspects of the geometry of these manifolds could lead to better average case analysis of interior point methods.

Estimating the surface area of boundaries and randomly sampling hypersurfaces

In [2], we estimate the surface area of an n-dimensional convex body by drawing upon the observation that the amount of heat diffusing outside of a heated body in a short period of time is proportional to its surface area. Computing the surface area of a convex body was mentioned as an open problem in "Geometric Algorithms in Combinatorial Optimization" of Grötschel, Lovász and Schrijver. Our algorithm has a complexity of $O^*(n^4)$ calls to a membership oracle compared to $O^*(n^{8.5})$ for the best previously known algorithm.
In collaboration with Niyogi, I used an extension of this idea to develop an algorithm that samples the surface of a convex body [21]. The underlying intuition is that over a short period of time, heat diffuses almost uniformly out of the surface of a convex set, if the initial distribution is uniform. We also used the same idea to sample a smooth hypersurface that is the boundary of a (not necessarily convex) subset of $\mathbb{R}^n$ if this subset can be sampled uniformly. These algorithms have applications to Goodness-of-Fit tests in statistics.

Spectral clustering and diffusion maps

Several popular approaches to the problem of clustering involve graph partitioning. While these algorithms are very widely used in practice, there are relatively few convergence results regarding their behavior as the number of sample points tends to

infinity. In work with Belkin and Niyogi [18], I have shown that for certain graphs arising naturally from random data, the total weight of all edges cut by a surface converges to its weighted surface area as the number of data points tends to infinity. The weight function is the induced density of the underlying probability distribution on the surface. The proof relates the total weight of edges cut to "the amount of heat that diffuses out," and then analyzes this amount.

Diffusion Maps have gained popularity in recent years as a method of classifying geometric data. Maximum margin classifiers are linear classifiers that maximize the width of the margin that separates positive samples from negative samples which do not penalize errors close to the class boundaries. In recent work [13] with Mahoney, I have analyzed the number of random samples needed to classify the vertices of a graph using gap tolerant classifiers on the diffusion map feature space.

Distributed consensus and network algorithms

A question that has attracted considerable interest has the following general form: *how might a group of linguistic agents arrive at a shared communication system purely through local patterns of interaction, without any global agency enforcing uniformity?* In forthcoming work with Niyogi [23], I have proved several results concerning a model of language evolution in which agents placed at the vertices of a graph communicate and update their beliefs at successive time steps. Our proof involves the analysis of the diffusion of information in a network. In the course of the proof of language convergence, we use random walks and relate language evolution to coalescent processes. In a related question in the area of distributed consensus, each sensor in a network starts out with a real number, and the sensors must communicate to reach an approximate consensus about the arithmetic mean of these values. The traditional algorithm corresponds to a diffusion over a single time scale, and has an averaging time that is controlled by the reciprocal of the spectral gap of the graph. In [16] and [17], I have developed algorithms which use non-convex affine updates (i. e. linear updates with negative coefficients) to accelerate the averaging process. This may be viewed as a multi-scale diffusion, where the different step sizes correspond to diffusion on different scales. In certain graphs such as those with sparse cuts, this outperforms existing algorithms.

In joint work with Harsha, Hayes, Räecke and Radhakrishnan [5], we showed that the asymptotic heat flow is a good way of routing to minimize the sum of the squares of edge loads when there is a single target. In joint work with Lawler [12], I have shown that in graphs where the asymptotic heat flow from one vertex to another has a small ℓ_1 norm, this method of routing performs well under a very general class of

performance measures. This approach works well for some graphs of practical interest such as grids and on all graphs on which a random walk mixes rapidly.

Littlewood-Richardson coefficients and Geometric Complexity Theory

Geometric Complexity Theory is a recent algebro-geometric approach towards the P *vs.* NP question. I collaborated with Mulmuley on "Geometric Complexity Theory V" [14], which is a note in a series of papers explaining this approach. Clebsch-Gordan coefficients have appeared in quantum mechanical computations since Wigner. In this note, we show that assuming certain conjectures from representation theory, for which there is experimental evidence, the positivity of a class of Clebsch-Gordan coefficients for arbitrary semisimple Lie groups can be decided in strongly polynomial time. In [15], I proved that these coefficients are as hard to compute as the number of integer points in a general convex polytope and in particular cannot be done in polynomial time if $P \neq NP$. This answered a question posed by Rassart [26]. The results of [15] forestall the possibility of obtaining an efficient deterministic algorithm to compute Littlewood-Richardson coefficients under a widely believed complexity theoretic hypothesis. The issue of approximation remains open. I am interested in developing randomized algorithms for this task, and have done so in special cases [25].

Schramm-Loewner evolution

Many natural discrete models in physics, such as the Ising model for magnetization and percolation on a hexagonal lattice, converge as the lattice spacing tends to zero to fractals that can be described by a remarkable stochastic differential equation known as Schramm-Loewner evolution (SLE). These fractal curves are parameterized using a quantity known as the halfplane capacity, which is defined below. Let A be a connected subset of the (complex) upper halfplane $\mathbb{H}$ such that $\mathbb{H} \setminus A$ is simply connected. For $y > 0$, consider a Brownian motion $\{B_t\}_{t \geq 0}$ started at $B_0 = \imath y$ where $\imath := \sqrt{-1}$. Let $h(y)$ be the expected height at which it first hits $\mathbb{R} \cup A$. Then the halfplane capacity of A, $\mathrm{hcap}(A)$ is $\lim_{y \to \infty} yh(y)$. In work with Lalley and Lawler, I have shown [11] that for any A, $\mathrm{hcap}(A)$ is comparable to the Lebesgue measure of the set of all points at distance 1 from A, where distance is measured using the usual hyperbolic metric on the upper-half plane.

Future work

Learning from high dimensional data

While modern data sets are typically high dimensional, they can often be successfully modeled as lying on low dimensional manifolds embedded in high dimensional space because they are generated by a dynamical system that has relatively few degrees of freedom. In recent years, the hypothesis that data lie on a low dimensional manifold has gained acceptance, and the class of methods whose theoretical validity is contingent upon this hypothesis has formed a subfield of its own called manifold learning.

Whether data lie near a low-dimensional manifold, is a question that has not received much attention. While the informational complexity of this problem has been determined in my recent work with Mitter [19], it is unclear whether there is an algorithm that would efficiently determine whether there is a manifold with prescribed regularity properties passing through a given set of data. Recent breakthroughs of Charles Fefferman [4] are very relevant to this question. I have had discussions with Fefferman and Mitter on this subject, and we are attempting to express the manifold as the zero section of a vector bundle over a neighborhood of the data. A vector bundle over a space is a collection of smoothly varying vector-spaces (known as fibers) indexed by points of the space, and a section of this vector bundle is a smoothly varying vector field, where each vector belongs to the corresponding fiber.

An example of when data seems to lie near a manifold is the case of human speech, where the waveforms can be viewed to lie in an infinite dimensional space, but the undulations of our vocal chords effectively have few degrees of freedom. The same can be said of the many views of a fixed object from different points in space. Here, the images belong to a high dimensional space, but they can be parameterized by the points (in $3-$space) from which the object was viewed. Many basic questions regarding learning on manifolds remain open. Extending the work of [22] and [18], I am interested in developing a clustering algorithm that finds a cut which minimizes the total amount of heat diffusing out of it along the manifold.

Choice of metric in Learning theory

The metric plays a very important role in Learning Theory, and is related to the question of how humans can learn in spite of the poverty of stimulus. In many realistic settings, there is an underlying but unknown metric on the space of examples, such that a nearest neighbor classifier would perform well, if the notion of nearest is

relative to this metric. Kernel methods are based on the thesis that the metric arising from the kernel is more suitable for the purpose of learning than the ambient metric. The field of Semi-supervised Learning is motivated by the assumption that labeled data is expensive but unlabeled data is not, and often, the metric on a neighborhood graph constructed using unlabeled data is taken to be the "true" metric. However the construction of neighborhoods in the graphs typically adhoc or based on human intuition. The construction of canonical metrics which do not rely on human knowledge at all has received scant attention. One direction we are interested in pursuing is to study under the hypothesis that data lies on a manifold, whether the metric obtained as a limit of a geometric flow such as the Ricci flow or mean-curvature flow possesses desirable characteristics from the point of view of Learning. Under these flows, the metric evolves according to gradient descent in the space of metrics, the objective function being a notion of the complexity of the metric such as the Ricci curvature or mean curvature.

Optimization

It is natural to try to use randomized methods when optimizing non-convex functions over a convex set, because of the possibility of escaping local optima which does not exist with natural gradient based deterministic methods. I am working on designing randomized interior point methods that combine the use of the Dikin random walk with the "Go-with-the-winners" algorithm of Aldous and Vazirani [1]. The Dikin walk in [24, 7] and a long-step variant of it was implemented by Huang and Mehrotra [6] in the context of Walk-and-Round Heuristics for Mixed-Integer Linear Programs. For the purpose of The long-step variant has been empirically observed to perform better than the original Dikin walk, but there are no theoretical guarantees on its performance. This is an issue that I would like to address.

Computing the volume of a convex set

Over the past two decades a large body of work has developed on the subject of random sampling on a convex set and computing the volume of a convex body. I have extended the results on sampling polytopes in [7] to any convex set associated with a self-concordant barrier. I am working towards developing faster algorithms to sample and approximate the volume of a large class of convex sets including those convex sets defined by semidefinite constraints.

Scaling limits of concave surfaces

Consider an equilateral triangle subdivided into a fine triangular grid. The Lebesgue measure of all functions defined on the lattice points, whose piecewise linear extensions are concave with fixed boundary conditions appears in several seemingly unrelated areas of mathematics. The flatspaces of the Legendre dual functions form honeycombs, which are tilings of certain domains with hexagonal tiles having different aspect ratios. It is proportional to the probability density that the sum of three Hermitian matrices whose respective eigenvalues are equal to the slopes of the concave functions on the boundary is equal to zero (see Theorem 4 in [10]). It is also the asymptotic value of the Littlewood-Richardson coefficient, suitably scaled and is thus important in Geometric Complexity Theory. I am working with Scott Sheffield towards obtaining an asympototic expression for the logarithm of this measure, which is the "differential entropy" of all concave surfaces on this triangular grid with fixed boundary conditions. We are also interested in the related question of identifying the limiting value of a random concave function with fixed boundary conditions as the mesh size tends to zero, and proving that such a limit exists.

References

[1] D. Aldous and U. Vazirani. *"Go With the Winners" Algorithms,* IEEE Conference on Foundations of Computer Science, (FOCS). 1994.

[2] M. Belkin, H. Narayanan and P. Niyogi. *Heat flow and a faster algorithm to compute the surface area of a convex body.*
IEEE Conference on Foundations of Computer Science (FOCS), October 2006.

[3] I. I. Dikin, *Iterative solutions of problems of linear and quadratic programming,* Soviet Math. Dokl (1967)

[4] C. Fefferman, *Whitney's extension problems and interpolation of data,*
Bulletin of the AMS, April 2009

[5] T. Hayes, P. Harsha, H. Narayanan, J. Radhakrishnan and H. Raecke, *Oblivious Routing to minimize average latency.*
Symposium on discrete algorithms (SODA 2008).

[6] K-L. Huang and S. Mehrotra, *An Empirical Evaluation of Walk-and-Round Heuristics for Mixed-Integer Linear Programs.*
`http://www.optimization-online.org/DB_FILE/2010/09/2734.pdf`

[7] R. Kannan and H. Narayanan, *Random walks on polytopes and an affine interior point algorithm.*
Proceedings of 41st ACM symposium on Theory of Computing (STOC 2009).

[8] N. K. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica,* 1984.

[9] N. K. Karmarkar, "On the Riemannian Geometry underlying interior point methods," Mathematical Developments Arising from linear programming. *Contemporary Mathematics, Amer. Math. Soc.,* 1990.

[10] A. Knutson and T. Tao, "Honeycombs and Hermitian Matrices", *Notices of the AMS*, Volume 48, Number 2

[11] S. Lalley, G. Lawler and H. Narayanan, *Geometric Interpretation of Half-Plane Capacity*,
Electronic Communications in Probability,
`http://www.math.washington.edu/~ejpecp/EcpVol14/paper55.abs.html`

[12] G. Lawler and H. Narayanan, *Mixing times and ℓ_p bounds for Oblivious Routing*, Workshop on Analytic Algorithmics and Combinatorics, (ANALCO '09)

[13] M. Mahoney and H. Narayanan, *Learning with spectral kernels and heavy-tailed data,*
`http://arxiv.org/PS_cache/arxiv/pdf/0906/0906.4539v1.pdf`

[14] K. Mulmuley and H. Narayanan, *Geometric Complexity Theory V: On deciding nonvanishing of a generalized Littlewood Richardson coefficient.*
Technical Report TR-2007-05, Computer Science Department, University of Chicago, April 2007

[15] H. Narayanan, *On the complexity of computing Kostka numbers and Littlewood-Richardson coefficients.*
18th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC), June 2006
Journal of Algebraic Combinatorics, volume 24, issue 3, November 2006

[16] H. Narayanan, *Geographic Gossip via affine combinations.* Principles of Distributed Computing (PODC 2007.)

[17] H. Narayanan, *Distributed Averaging in the presence of a Sparse Cut.* Principles of Distributed Computing (PODC 2008.)

[18] H. Narayanan, M. Belkin and P. Niyogi, *On the relation between low density separation, spectral clustering and graph cuts,* 20th Annual Conference on Neural Information Processing Systems (NIPS), December 2006

[19] H. Narayanan, S. Mitter, *Sample Complexity of testing the Manifold Hypothesis.* To appear in 24th Annual Conference on Neural Information Processing Systems (NIPS), December 2010

[20] H. Narayanan, S. Rakhlin, *Random Walk approach to Online Convex Optimization.* To appear in 24th Annual Conference on Neural Information Processing Systems (NIPS), December 2010

[21] H. Narayanan and P. Niyogi, *Sampling Hypersurfaces through Diffusion.* Neural Information Processing Systems (NIPS 2006), December 2006.

[22] H. Narayanan and P. Niyogi, *On the sample complexity of learning smooth cuts on a manifold,* Conference on Learning Theory, June 2009

[23] H. Narayanan and P. Niyogi, *Language evolution, coalescent processes and the consensus problem on a social network,* preprint.

[24] H.Narayanan, *Randomized interior point methods for sampling and optimization,* Submitted to Mathematical Programming, Series A, `http://arxiv.org/PS_cache/arxiv/pdf/0911/0911.3950v2.pdf`

[25] H. Narayanan, *Computing Littlewood-Richardson coefficients: Algorithms and Complexity*, Talk given at the Workshop on Geometric Complexity Theory at the Intractibility Institute, Princeton, June 2010 `http://www.mit.edu/~har/LRtalk.pdf`

[26] E. Rassart. *Geometric approaches to computing Kostka numbers and Littlewood-Richardson coefficients,* Thesis for the Ph.D. degree in Mathematics, Massachusetts Institute of Technology (MIT), 2004