Matthew Hirn

Assistant Professor, CMSE; Department of Mathematics
mhirn@msu.edu | 517.355.9697 | 619 Red Cedar Rd., Room C303

BIO SKETCH

Matthew Hirn joined the Michigan State University faculty in 2015 as an assistant professor, with joint appointments in the Department of Computational Mathematics, Science and Engineering, and the Department of Mathematics.

Prior to arriving at Michigan State, Hirn held postdoctoral appointments in the Department of Mathematics at Yale University, as part of Ronald Coifman's research group, and in the Département d'Informatique at the École normale supérieure, Paris, as part of Stéphane Mallat's research group. He also held a brief appointment as a visiting assistant professor at Cornell University, where he directed an NSF-funded Research Experience for Undergraduates program on High Dimensional Data Analysis.

Dr. Hirn received his B.A. in mathematics from Cornell University under the supervision of Robert Strichartz, and his Ph.D. in mathematics at the University of Maryland, College Park under the supervision of John Benedetto and Kasso Okoudjou. While at the University of Maryland, he was a member of the Norbert Weiner Center for Harmonic Analysis and Applications.

RESEARCH INTERESTS

Hirn's research interests are at the interface of harmonic analysis and machine learning. Broadly speaking, he develops mathematically provable machine learning algorithms to circumvent prohibitively costly computations in scientific computing, thereby opening new avenues for scientific breakthroughs. He is a passionate supporter of undergraduate and graduate education and research.

Specific interests include: Applied harmonic analysis; manifold learning; smooth extensions and interpolations; deep learning, and applications involving quantum chemistry, *N*-body problems, image analysis, hyperspectral image analysis, flow cytometry, dynamical systems, and fluid mechanics.

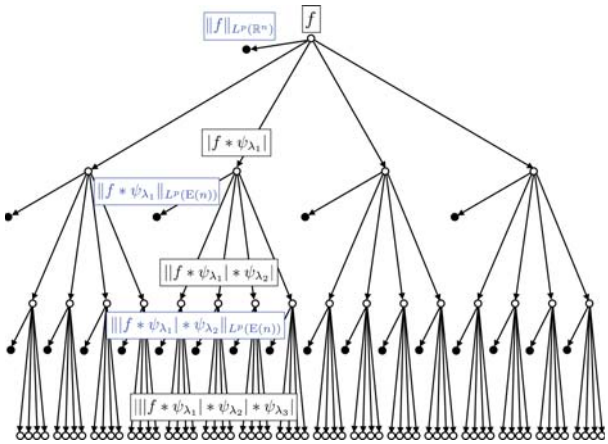

FIGURE 1. Isometry invariant scattering network. The initial function f is transformed through an alternating cascade of linear complex valued wavelet transforms and non-linear modulus operators, yielding an ever expanding set of intermediate functions that are indicated by the empty dots and the black text. At every layer the L^p norm over the Euclidean group $E(n)$ of each function is computed, yielding isometry invariant features; these functionals are denoted with filled dots with blue text. This figure displays two full layers and the intermediate functions of the third layer.

WEBSITE

www.math.msu.edu/~mhirn/

CURRENT RESEARCH FOCUS

I develop mathematically provable machine learning algorithms to circumvent prohibitively costly computations in scientific computing, thereby opening new avenues for scientific breakthroughs. Peta- and exascale computing have enabled for the first time true scale bridging algorithms in numerous computational fields such as astrophysics, climate science, fluid mechanics, and quantum chemistry. These scale-bridging algorithms have lead to massive amounts of high-dimensional distributed data, and interpreting and analyzing this data is a fundamental problem facing science. Machine learning is one approach to extract information from these data sets. However, to be able to apply machine learning algorithms to interpret these multiscale data sets, new paradigms need to be introduced that are capable of learning intrinsic variables of physical systems that encode complex interactions across scales.

Scattering deep learning for quantum chemistry. Mallat, Poilvert, and I introduced the Scattering Transform for the regression of the potential energy of molecular states in quantum chemistry. We obtained state-of-the-art numerical results on planar organic molecules, achieving errors in line with the best quantum chemical algorithms, but at a fraction of the computational cost. The Scattering Transform is the first machine learning algorithm for quantum chemistry to be fully adapted to the potential energy of molecular states: it is invariant to symmetry group actions on the indices of the atoms and isometry group actions on their positions, and Lipschitz stable to actions of the diffeomorphism group at all length scales. The Scattering Transform has the structure of a deep convolutional network, with multiple cascaded layers, each consisting of a linear and nonlinear transform. Unlike most deep networks, the linear transforms are predefined as wavelet transforms over the isometry group in $\mathbb{R}^3$; this is necessary to circumvent the curse of dimensionality and to ensure the network has the correct physical properties. Each path can be interpreted as encoding a subset of interactions that move across a particular sequence of scales. The specific paths within the network for interpolating the energy are learned from pre-existing data.

Mallat and I continue to study which functional spaces are well approximated by Scattering networks. Within quantum chemistry, we can prove that Coulombic interactions are learned with one layer of the Scattering network; the next energy component to attack is the kinetic energy. We also conjecture that the second layer and beyond learns features analogous to multipole moments. With Bartók and Csányi, we are studying the Scattering network's capability for learning the exchange-correlation energy from Density Functional Theory, which is the most complicated component of the total energy but is a universal functional for all molecular states. We also plan to study the learning capabilities of Scattering networks in the context of Density Matrix Functional Theory.

Manifold learning. I developed with Coifman and Marshall diffusion based manifold learning for dynamic data sampled from a Riemannian manifold with a smooth family of metrics. Diffusion manifold learning algorithms learn hierarchical organizations of data sampled from a smooth, compact Riemannian manifold, in which the finest scale is the manifold geometry. We proved that one can learn from a finite data set a time inhomogeneous Markov chain that in the limit of infinite data converges to the heat kernel of $\partial_t u = \Delta_{g(t)} u$, thus extending seminal results on static manifolds. This result is at the heart of new algorithms that we have developed to learn intrinsic geometric variables of dynamic data sets, such as temporal hyperspectral data. Further geometrical questions loom as we attempt to understand these algorithms further.

Smooth extensions. I proved with Le Gruyer the existence of Quasi Absolutely Minimal Lipschitz Extensions (quasi-AMLEs) between certain classes of metric spaces. In the limit of specified quantities, quasi-AMLEs become AMLEs, and we conjecture that one can pass through this limit to prove existence. The study of AMLEs has a rich history and is linked to analysis, probability, PDEs, and computer vision. Learning algorithms, fundamentally, interpolate a function. Herbert-Voss, McCollum and I developed an efficient algorithm for computing AMLEs for the space $C^{1,1}(\mathbb{R}^n)$. Amongst Whitney type interpolation algorithms, this algorithm is the first to provably compute in $O(N \log N)$ time (N is the number of interpolation points) the order of magnitude of the best Whitney constant to within a dimensionless constant. This computation utilizes a previous result of Le Gruyer in conjunction with a clever application of the Well Separated Pairs Decomposition. An interpolant F with the same norm is computed by replacing Calderon-Zygmund decompositions with an intricate partition due to Wells, which we prove can be reduced to the computation of a convex hull. The query work is merely $O(\log N)$ per query point.

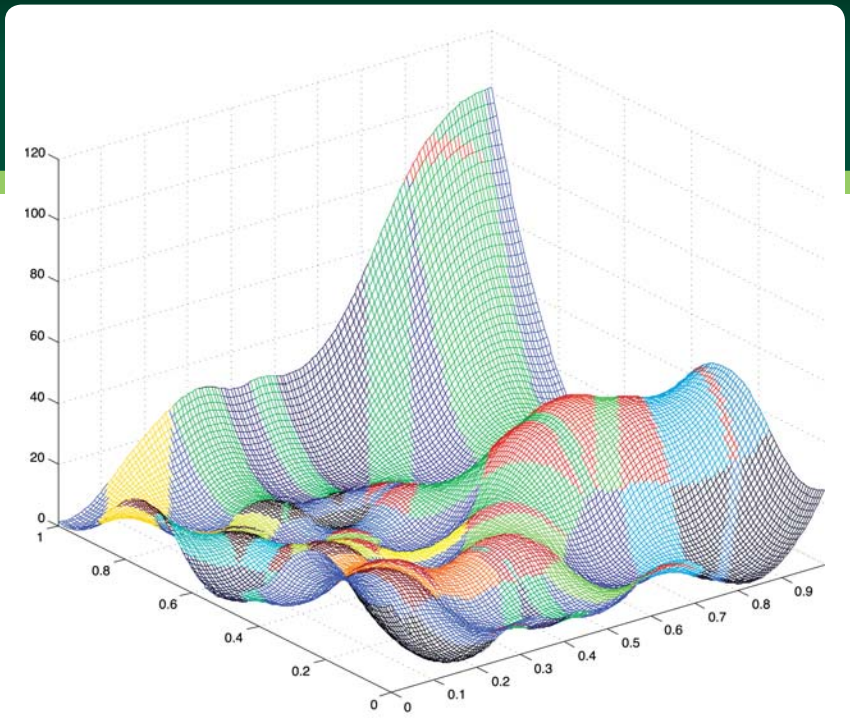

FIGURE 2. Smooth interpolation. The function $F \in C^{1,1}(\mathbb{R}^2)$ displayed above is the interpolating function of a finite set of data consisting of locations $\{x_i\}_{i=1}^N \subset \mathbb{R}^2$, function values $\{a_i\}_{i=1}^N \subset \mathbb{R}$, and gradients $\{b_i\}_{i=1}^N \subset \mathbb{R}^2$, so that $F(x_i) = a_i$ and $\nabla F(x_i) = b_i$ for each $i = 1, \ldots, N$. Additionally, the value of the semi-norm $\mathrm{Lip}(\nabla F)$ is within a factor of 20 of the minimum possible value. The "patchwork quilt" pattern denotes the partition of $\mathbb{R}^2$ on which local pieces of the interpolant F are constructed. The one time cost of computing F is $O(N \log N)$ and the query work for each new point is $O(\log N)$.

Moving forward, Le Gruyer and I have partial results on the best Whitney constant for $C^{1,1}(\mathbb{R}^n, \mathbb{R}^m)$; I am also studying the analogous problem for $C^{2,1}(\mathbb{R}^n)$. Positive results along these lines, in addition to their theoretical impact, could yield further algorithmic developments as in my work with Herbert-Voss and McCollum. We also continue to study questions related to AMLEs, including the aforementioned conjecture and developing a stronger notion for $C^{1,1}(\mathbb{R}^n)$ that can yield uniqueness.

RECENT PUBLICATIONS

M.J. Hirn, S. Mallat, N. Poilvert, "Quantum Energy Regression using Scattering Transforms." 2015. arXiv:1502.02077.

A. Herbert-Voss, M.J. Hirn, F. McCollum, "Computing minimal interpolants in $C^{1,1}(\mathbb{R}^d)$," arXiv:1411.5668 (2015).

M.J. Hirn, E. Le Gruyer. "A general theorem of existence of quasi absolutely minimal Lipschitz extensions." *Mathematische Annalen*, vol. 359, no. 3–4, pp. 595–628, arXiv:1211.5700 (2014).

R.R. Coifman, M.J. Hirn, "Diffusion maps for changing data," *Applied and Computational Harmonic Analysis*, vol. 36, no. 1, pp. 79–107, arXiv:1209.0245 (2014).